RNA/DNA Binding Protein TDP43 Regulates DNA Mismatch Repair Genes with Implications for Genome Stability

Vincent E. Provasek[1,2], Albino Bacolla[3], Suganya Rangaswamy.[1], Joy Mitra[1], Manohar Kodavati[1], Issa O. Yusuf[4], Vikas H. Malojirao[1], Velmarini Vasquez[1], Gavin W. Britz[1,5], Guo-Min Li[6], Zuoshang Xu [4], Sankar Mitra[1], Ralph M. Garruto[7], John A. Tainer[3] and Muralidhar L. Hegde[1,8]*

[1]Division of DNA Repair Research within the Center for Neuroregeneration, Department of Neurosurgery, Houston Methodist Research Institute, Houston, TX 77030, USA.
[2]School of Medicine, Texas A&M University, College Station, TX 77843, USA.
[3]Department of Molecular and Cellular Oncology, Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[4]Department of Biochemistry and Molecular Biotechnology, University of Massachusetts Chan Medical School, Worcester, MA, 01655, USA
[5]Department of Neurosurgery and Department of Neuroscience, Weill Cornell Medical College, New York, NY 10065, USA.
[6]Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA.
[7]Department of Biological Sciences, Binghamton University, State University of New York, Binghamton, NY 13902.
[8]Department of Neuroscience, Weill Cornell Medical College, New York, NY 10065, USA.

[*] Author to whom correspondence should be addressed.

Abstract

TDP43 is an RNA/DNA binding protein increasingly recognized for its role in neurodegenerative conditions including amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). As characterized by its aberrant nuclear export and cytoplasmic aggregation, TDP43 proteinopathy is a hallmark feature in over 95% of ALS/FTD cases, leading to the formation of detrimental cytosolic aggregates and a reduction in nuclear functionality within neurons. Building on our prior work linking TDP43 proteinopathy to the accumulation of DNA double-strand breaks (DSBs) in neurons, the present investigation uncovers a novel regulatory relationship between TDP43 and DNA mismatch repair (MMR) gene expressions. Here, we show that TDP43 depletion or overexpression directly affects the expression of key MMR genes. Alterations include MLH1, MSH2, MSH3, MSH6, and PMS2 levels across various primary cell lines, independent of their proliferative status. Our results specifically establish that TDP43 selectively influences the expression of MLH1 and MSH6 by influencing their alternative transcript splicing patterns and stability. We furthermore find aberrant MMR gene expression is linked to TDP43 proteinopathy in two distinct ALS mouse models and post-mortem brain and spinal cord tissues of ALS patients. Notably, MMR depletion resulted in the partial rescue of TDP43 proteinopathy-induced DNA damage and signaling. Moreover, bioinformatics analysis of the TCGA cancer database reveals significant associations between TDP43 expression, MMR gene expression, and mutational burden across multiple cancers. Collectively, our findings implicate TDP43 as a critical regulator

of the MMR pathway and unveil its broad impact on the etiology of both neurodegenerative and neoplastic pathologies.

Keywords: TDP43, DNA mismatch repair (MMR), DNA double-strand breaks (DSBs), R-loop, Neurodegeneration, Amyotrophic lateral sclerosis (ALS).

Introduction

Transactive response DNA binding protein of 43 kDa (TDP-43) is a nuclear protein involved in the regulation of expression of a wide range of genes, including those associated with DNA damage response and repair pathways. This highly conserved RNA/DNA-binding protein is furthermore linked to multiple pathological processes, including neurodegeneration and cancer (1-3). TDP43 is expressed in a variety of tissues, with particularly high levels in the brain and spinal cord and is essential for normal development and function of the nervous system (4). RNA processing is a complex and tightly regulated process that involves numerous RNA-binding proteins to facilitate appropriate RNA processing. TDP43 plays a central role in several aspects of RNA processing, including alternative splicing, mRNA stability, and transport (5-7). For example, TDP43 binds to the 3'-untranslated region (UTR) of target mRNAs to regulate their stability and localization (8, 9). Additionally, TDP43 interacts with various splicing factors to coordinate efficient intron removal from target transcripts (10). The precise mechanisms by which TDP43 regulates RNA processing are not completely understood, yet several studies have provided important insights into this process. For example, TDP43 may act as a scaffold protein that brings together different components of the RNA processing machinery, thereby, facilitating their interactions and coordination (8). In addition, TDP43 may facilitate the recruitment of other regulatory proteins to target mRNA influencing either their stabilization or degradation (5, 11, 12). The specific mechanisms that functionally link TDP43, and DNA repair have also been shown in the context of neurodegenerative disease. We have reported that TDP43 is a critical scaffolding factor required for the DNA double-strand break (DSB) repair (13). However, given the extensive role of TDP43 in RNA metabolism, we questioned whether it may contribute to DNA repair in a more global aspect via its effects on the processing of DNA repair gene transcripts. Here, we discover a previously unknown role of TDP43 in modulating the expression of MMR genes.

MMR is a critical DNA repair pathway responsible for at least a 10-fold increase in the replication fidelity of replicating cells (14). Deficient MMR (dMMR) is classically associated with neoplastic pathologies, most notably Lynch Syndrome (15-17). The MMR family consists of six core members, MLH1, MSH2, MSH3, MSH6, PMS2, and PMS1. These components function as heterodimers that recognize mispaired or otherwise aberrant nucleic acid substrates and initiate their excision and replacement. The recognition heterodimers, MutSα (MSH2-MSH6) and MutSβ (MSH2-MSH3) interact with the MutLα (MLH1-PMS2) heterodimer to initiate the repair process; this is followed by several other DNA repair factors, including RPA, PCNA, exonuclease 1 and DNA polymerase δ, to excise and re-synthesized damaged sequences (18, 19). Importantly, defects in any one of these factors, particularly MutSα or MutLα, is sufficient to disrupt the entire MMR machinery and thus lead to increased mutational load and tumorigenesis.

While the consequence of dMMR on cancer development are well known, there are other functions of MMR that are less understood. One of these functions involves the role of MMR at the intersection between DNA repair and the DNA damage response (DDR) signaling pathway. The relationship between MMR and DDR has been reviewed elsewhere (20-24). However, the importance of this interaction is evinced by the requirement of MMR for chemotherapy-induced (CTX) cancer cell killing, particularly following select platinum-based, and DNA alkylator agents.

Reports of CTX resistance in cancers of the colon, endometrium, and CNS, among others, demonstrate a correlation with dMMR and subsequently reduced cell killing (25-29). At the center of this apparent paradox is the function of MMR-induced DDR activation. Specifically, it was shown that without an active MMR pathway, DNA damage-induced DDR activation did not occur, and cells did not undergo subsequent apoptosis (30). Overabundance of MMR has likewise been attributed to negative effects on genome integrity. There are limited reports showing that overexpression of certain MMR factors (e.g., MLH1 or PMS2) induce a hypermutable phenotype in vivo (31, 32). Moreover, others have shown that overexpression of MSH2, irrespective of its repair capability, is sufficient to induce cell killing, possibly through direct activation of apoptosis via specific domains within the protein (33). Perhaps it should come as no surprise that the MMR repair system could indeed act in the pro-instability manner, given its documented role in facilitating somatic hypermutation during immunoglobulin class switch recombination (22, 24, 34-36). In any case, the ability of the MMR system to either repair or induce mutations clearly depends on cellular context and expression regulation.

The question of MMR activity and regulation takes on renewed importance in the neurodegeneration field, where less is known regarding the role of MMR in the nondividing cell. Early reports have suggested that MMR activity in the context of non-dividing cells conferred a mutagenic effect, possibly due to a loss of strand discrimination signals (37, 38). In neurons, the function of MMR is even less clear. At least one early report indicated neuronal MMR would be limited to the repair of deamination products, including 5-methyl cytosine to thymine, or cytosine to uracil (39). More recently, however, there has been renewed interest in the role of MMR in neurons, especially within the context of neurodegeneration. Recent reports have indicated that MMR may contribute to the expansion, and downstream clinical severity, of trinucleotide repeat (TNR) mutations like those observed in Huntington's Disease (HD) (40-45). Taken together, there remain many significant questions regarding MMR function, particularly in nondividing cells. One clear theme, however, is that the expression of MMR proteins appears to exist in a delicate balance that, if disturbed, may have important consequences to cellular function, genomic integrity, and degenerative disease.

In this study, we shed light on at least one mechanism maintaining this balance by reporting for the first time that TDP43 regulates the expression of MMR genes. We showed that TDP43 exerts this role across multiple cell lines and animal models. We furthermore traced the mechanism of this effect by showing that TDP43 acts to regulate the RNA handling of at least two key MMR transcripts: *MLH1* and *MSH6*. We then tested whether this phenomenon also occurs in disease conditions associated with TDP43: neurodegeneration and cancer. By using animal models of TDP43-associated ALS (TDP43-ALS) and CNS tissues of Guamanian ALS patients, we discovered that the expression of MMR genes is indeed altered at the protein level. Moreover, we showed that similar changes observed using in vitro models of TDP43 pathology also increased the levels of DNA damage markers and impaired cellular response to oxidative DNA damage. Finally, we showed that TDP43-mediated expression of MMR genes evidently also extends to various cancers. By using extensive bioinformatic analysis, we found informative correlations between TDP43 and MMR gene expression, with concomitant associations with patient survival. Collectively these findings link TDP43 to active regulation of MMR protein levels and consequently TDP43 dysregulation to implicated roles in the pathogenesis of the two most important age-related disease conditions of our time: neurodegeneration and cancer.

Results

TDP43 Regulates the Expression of DNA Mismatch Repair Genes.
Exploratory investigation of TDP43's effects on mRNA expression of DNA repair genes was first conducted using an RT2-Profiler assay (Qiagen) with HEK293 cells, treated with either control or TDP43 directed siRNAs. When TDP43 expression was decreased to approximately 50% of control (**Fig. 1*A***), we observed significant changes in the expression of multiple DNA repair protein families (**Fig. 1 *B*** and ***C***, *SI Appendix*, **Table S1**). Of those families identified, MMR factors MLH1, MSH3, MSH6, and PMS2 exhibited a greater than two-fold decrease in their gene expressions. These results were confirmed by qRT-PCR (**Fig. 1*D*** and *SI Appendix,* **Table S2)** and western blotting (WB) analysis (**Fig. 1*E*)** using HEK293 cultures with TDP43 knockdown (KD) (**Fig. 1*D-E***). We next questioned how MMR expression would change when TDP-43 is overexpressed. Notably, moderate overexpression (OE) of WT-TDP43 in differentiated neuron-like SH-SY5Y cells was sufficient to boost the expression of key MMR factors markedly beyond their basal levels (**Fig. 1*F***).

We next assessed the expression levels of MMR proteins in both dividing and non-dividing cells, using neural progenitor cells and differentiated neurons. To do so, we induced and differentiated induced pluripotent stem cells (iPSC) to neural progenitor stem cells (NPSC) to iMNs, as shown in **Fig. 2*A***. WB analysis of MMR factors in cycling NPSCs and differentiated iMNs showed that levels of MLH1, MSH2 and MSH3 were significantly reduced in the iMN stage compared to those in the NPSC stage (**Fig. 2*B***). Interestingly, we also observed in protein isolates of cultured cells taken at the iPSC, NPSC, and iMN stages of differentiation showed decreasing levels of MMR expression. These results underscore the significant impact of TDP43 levels on MMR protein expressions, underscoring its potential role in regulating DNA repair regardless of the cell cycle status.

Given these results, we questioned whether TDP43 KD also decreased expression of MMR genes in iPSC-derived iMNs. Using WB (**Fig. 2*C***) and immunofluorescence (IF) microscopy (**Fig. 2*D***) analyses, we observed that lentiviral-delivered shRNA TDP43 KD in cultured motor neurons indeed caused decreased expression of MLH1, MSH3, MSH6 and PMS2. This further clarified that TDP43 is necessary and sufficient for regulating the expression of MMR factors in non-dividing neurons and that any disruption in the levels of TDP43 could have knock-on effects with respect to the expression of MMR genes. We next turned our attention to identifying potential mechanisms underlying these observations.

TDP43 Depletion and Mutations Alter Transcriptional Processing of MMR Transcripts.
TDP43 is an RNA/DNA binding protein with a well described role in mediating myriad RNA transactions across cells and tissues. Briefly, TDP43 works in conjunction with other nuclear protein factors, including the spliceosome, to ensure efficient removal of intronic sequences from pre-mRNAs to produce mature mRNAs transcripts (Fig. **3*A***) (46). Disruption of this process can lead to the retention of intronic sequences: which can hamper transcript nuclear export, permit cytosolic degradation, or permit synthesis of alternative protein products. Moreover, perturbations of this process have been reported to contribute significantly to neuropathology of multiple diseases, especially ALS and FTD (1, 9, 47-49).

Given these functions, we questioned whether TDP43 might regulate MMR expression by affecting the processing of MMR transcripts. To this end, we first probed the splicing patterns of two MMR transcripts using endpoint PCR analysis of total RNA extracts from HEK293 cells with TDP43 KD (**Fig. 3*B-E*** and *SI Appendix,* **Table S3**). Primers were designed to amplify selected

Exon-Intron-Exon regions of each gene such that the size of the PCR product would reflect the presence (as in pre-mRNA) or absence (as in mature mRNA) of intronic DNA sequences (*SI Appendix,* **Table S2**). To inform our selection of transcript regions, we employed two complementary *in silico* approaches: RBPSuite (50) and RBPMap (v1.2) (51). These tools aided in identifying the segments of each MMR transcript where TDP43 is expected to bind (**Fig. 3 *C*** and ***D***). Leveraging the insights from both platforms, we tailored primers to amplify Exon-Intron-Exon regions encompassing the predicted TDP43 binding sites. Based on these predictions, we chose to focus on the identified regions between Exon 17 and Exon 18 of MLH1 and Exon 8 and Exon 9 of MSH6 as potential TDP43 splicing targets. We then performed siRNA TDP43 KD and measured the resulting mRNAs. In the case of MHL1 and MSH6, TDP43 KD caused greater than 50% decrease in the fully spliced mRNA product and a concomitant increase of at least one predicted un-spliced variant. Evaluations of other MMR transcripts like, MSH3 and PMS2, displayed either no changes in splicing or sporadic decreases in the fully spliced transcript, respectively (*SI Appendix,* **Fig. S1*A***). We next investigated the effect of ALS-linked TDP43 mutations M337V, Q331K and A315T on MMR transcript processing. These mutations have been reported to induce TDP43 proteinopathy with certain transcriptomic effects (9). In these experiments We used doxycycline (Dox)-inducible TDP43 WT or mutant expressing SH-SY5Y lines at levels not exceeding 2-fold over the endogenous TDP43 level (*SI Appendix,* **Fig. S1*B***) (1, 2, 13). All mutants influenced the pre-mRNA processing of *MLH1* and *MSH6* genes in a manner similar to that in TDP43 KD cells, whereas *PMS2* and *MSH3* transcripts were unaffected. Finally, we conducted a similar splicing experiment using human spinal cord autopsy specimens from Guamanian ALS patients and age-matched controls (*SI Appendix* **S2*D***). These patient samples had been previously validated for the presence of TDP43 pathology and associated DNA damage accumulation (13). The data revealed an accumulation of MLH1 pre-mRNA in patient tissues relative to the controls, while PMS2 splicing remained consistent. In summary, perturbations of TDP43 in human cells specifically induce alternative splicing of two key MMR transcripts, MLH1 and MSH6.

TDP43 Regulates the Stability of MLH1 and MSH6 Transcripts.

TDP43 is shown to modulate the stability of the processed transcripts (52, 53). To identify this effect on MLH1 and MSH6 transcripts, we measured the influence of TDP43 on the stability of these transcripts. To accomplish this, we utilized a modified version of a previously described transcript stability protocol (54). Briefly, we used RT-PCR to measure the relative abundance of each transcript in total RNA samples isolated from nondividing neuronal cell cultures treated with scrambled control or TDP43 siRNAs at set timepoints after exposure to α-amanitin, a potent RNA polymerase II inhibitor (Schematically shown in **Fig. 4*A***. We then fit the fold change data over time to a one-phase decay curve, which was then used to compute the approximate half-life of the transcripts. Surprisingly, we observed the rate of MLH1 and MSH6 transcript degradation increased compared to controls during the first twelve hours of treatment (**Fig. 4 *B*** and ***C***). The loss of TDP43 resulted in a 38% decrease in the half-life of MLH1 transcripts and a 56% decrease in the half-life of MSH6 transcripts (**Fig. 4*D***). Plots of mRNA transcripts of MLH1, MSH2, MSH3 and MSH6 relative to their respective fold expressions (*SI Appendix* **Fig. S2*A-D***) were used to calculate the transcript half-life. Given the autoregulatory function of TDP43 on its own expression, we also measured the abundance of TDP43 transcripts over time (55). Expectedly, we observed a 23% increase in the half-life of TDP43 transcripts in its KD cell cultures (**Fig. 4*E***). Taken together, these data indicate TDP43 promotes the stability and persistence of MLH1 and

MSH6 transcripts. This suggests any condition which disrupts TDP43 homeostasis may also have direct effects on MMR expression.

TDP43 Pathology Disrupts the MMR Protein Expression Levels *in vivo*.

To determine whether MMR expression was altered within the context of neurodegenerative disease, we utilized two distinct murine models of TDP43-ALS. (57)The first murine model utilized a CRISPR-FLeX system to allow tetracycline inducible expression of a Tdp43 mutant lacking its nuclear localization sequence (Tdp43ΔNLS) under the ubiquitin-C (UBC) promoter (**Fig. 5*A***) (56). This model demonstrates mislocalized Tdp43 aggregates throughout the body, including the CNS. In this model, total protein lysates of cortical brain samples showed significant increases in expression of Msh2 and Msh3, with trends toward increased expression of Msh6 (**Fig. 5 *B*** and ***C***). Unexpectedly, we did not observe clear increases in Mlh1 despite increased levels of TDP43. We attribute this increase in Tdp43 expression to a disruption in the autoregulatory function of Tdp43, where mislocalization of Tdp43 within the cytosol results in a relative loss of nuclear Tdp43 with subsequent loss of negative autoregulation. We suspect that the resulting increase in Tdp43 expression is endogenous in origin and may work to bind and facilitate MMR expression. To further test this hypothesis and confirm whether increased murine Tdp43 expression was sufficient to increase MMR expression, we utilized a second model of Tdp43-ALS with moderate overexpression of murine Tdp43 under the major prion protein (PrP) promoter (**Fig. 5 *D***). In this model, we observed minor sex-specific differences in MMR expression. Both Mlh1 and Msh3 were upregulated in males, whereas females showed upregulation of Mlh1, Msh2, and Msh3 (**Fig. 5*E-G*** and *SI Appendix,* **Fig. S3**).

Given these results in mouse models of TDP43 ALS, we next questioned whether MMR protein expression was altered in the post-mortem CNS tissues of Guamanian ALS affected patients. When we tested expression of MMR in the in-soluble protein fraction of total protein isolates, we observed consistent increases in detection of MLH1 protein in ALS samples relative to controls (**Fig. 5 *H*** and ***I*** and *SI Appendix,* **Table S4**). Taken together, these data indicate that TDP43 can modulate the level of key MMR factors *in vivo*.

Depletion of MMR Factors Rescues TDP43 Pathology-Associated DNA Damage.

The roles of MMR toward DDR signaling and apoptosis have been previously described (21, 23, 30, 58). To determine if loss of TDP43 confers similar functional effects on DNA damage-induced cell killing as direct MMR KD, we employed the MTT assay. HEK293 cell cultures were treated with 10 μM 6-thioguanine (6-TG) for 24 h followed by a 24 h recovery period prior to adding MTT reagent to culture media. MMR-deficient cultures were achieved via simultaneous siRNA KD of MLH1 and MSH2 (siMMR). In line with previous reports, MMR deficiency prevented 6-TG mediated cell killing while combined MLH1 and MSH2 overexpression (OE-MMR) had the opposite effect. Importantly, TDP43 KD also mitigated 6-TG cell killing while TDP43 overexpression enhanced cell killing compared to controls (**Fig. 6*A***). These results indicate that TDP43-mediated changes in MMR expression may also influence cellular response to DNA damage.

We next questioned whether MMR may contribute to TDP43 proteinopathy-induced DNA damage. To this end, we utilized a cell model of TDP43-ALS. Cultures of SH-SH5Y cells carrying a doxycycline-inducible plasmid expressing TDP43ΔNLS were used to recapitulate TDP43 proteinopathy. After differentiation, cultures were induced for five days to yield significant cellular pathology, increased basal levels of DNA damage, and DDR activation. To identify potential MMR-mediated DNA damage, we utilized alkaline single cell gel

electrophoresis (Comet Assay) as a sensitive measure of SSBs and DSBs. Interestingly, we observed significant increases in DNA damage in TDP43ΔNLS cell lines, which was partially mitigated by KD of MSH2 (**Fig. 6*B***). WB analysis was also performed using samples of these cultures to assess γH2AX levels as a marker of DNA damage (**Fig. 6 *C*** and ***D***). Results from this experiment showed TDP43ΔNLS expression alone caused significant increases in DNA damage marker γH2AX and evidence of pathologic TDP43 aggregates, as identified by bands at 43 kDa and 35 kDa. Expectedly, MSH2 expression was also upregulated. Importantly, MSH2 KD appeared to diminish TDP43 proteinopathy-induced DNA damage, as indicated by lower γH2AX expression (**Fig. 6 *C*** and ***D***).

Given the effect of MSH2 KD on basal levels of TDP43 proteinopathy-induced DNA damage, we hypothesized the MMR system would also affect DNA repair capacity. To this end, we tested the ability of TDP43ΔNLS cells to recover from low level DNA damage induced by glucose oxidase (GO) (**Fig. 6 *E*** and ***F*** and *SI Appendix,* **Fig. S4**). Cell cultures treated with 100 nM GO for 45 min were allowed to recover for 6 h after treatment. WB analysis of these cell extracts showed not only an increase in markers of DNA damage and DDR signaling immediately after GO challenge, but also that such increases persisted in TDP43ΔNLS-expressing cells compared to controls after the recovery period. Again, KD of MSH2 appeared to improve cellular recovery from GO- induced damage, as shown by decreased levels of cleaved caspase 3 and phosphorylated ATM. These results align with previous reports describing the role of MMR in activating the DDR and underscore the regulatory role of TDP43 proteinopathy in this process.

TDP43 Expression Correlates with the Expression of MMR Genes in Cancer.

Finally, we questioned whether TDP43-mediated regulation also extended to the cancer disease context. To learn if TDP43 expression was linked to that of MMR genes in vivo, we examined transcriptomic profiles of tumors by analyzing RNA-seq data from TCGA database. First, we compared mRNA abundance of *TARDBP* between tumor types and matched controls for which at least 10 normal samples were available (15 tumor-control pairs total), which showed that *TARDB* was overexpressed in 60% (9 total) of cases, not changed in 3 cases and downregulated in 3 cases (**Fig. 7*A***).

Having found that TDP43 expression is generally dysregulated in cancer, we then assessed if the expression of TDP43 would correlate with patient survival by dividing the patients for each tumor type into two groups, a g_high group where TDP43 mRNA levels were above the population mean, and a g_low group, below the mean. A comprehensive analysis of hazard ratios in all 33 TCGA tumor types revealed that in four types of tumor, i.e. sarcoma, liver carcinoma, adrenocortical carcinoma and low grade glioma, patients with high TDP43 expression levels (the g_high group) would be associated high risk of poor outcome (*SI Appendix,* **Fig. S5*A***). Kaplan-Meier survival curves of this data indicated that patients with sarcoma incurred the most dramatic association between gene expression and survival, with the g_high group incurring an ~2-fold increase in poor prognosis relative to the glow group (*SI Appendix* **Fig. S5*B***). Therefore, assuming that *TARDB* expression correlates with TDP43 levels, our data suggest that TDP43 is an informative biomarker for some types of tumor.

Next, we developed a protocol to identify all genes coexpressed with TDP43 in cancer. From an initial screen of 677,457 linear regressions between RNA-seq data of *TARDBP* and ~20,000 genes in each of the 33 TCGA tumor types, ~79,000 regressions passed our threshold set at a p-value

<2.5E-6 (accounting for Bonferroni correction for multiple comparisons) for significant coexpression, 57% being positive and 43% being negative associations (**Fig. 7*B***). After filtering for genes positively coexpressed with *TARDBP* in at least 10 tumor types, and excluding those residing on the same chromosomal arm as *TARDBP*, our algorithm returned a total of 767 highly specific genes. The strongest correlation was observed for *PNN* in acute myeloid leukemia, which displayed a regression coefficient of 0.81 and a p-value of 1.8E-41 (*SI Appendix,* **Fig. S5*C***). Importantly, for the MMR genes, MSH6 was co-expressed with TDP43 in 20 tumor types, followed by MSH2 in 15 tumor types and PMS1 in 11 tumor types (**Fig. 7*C***); interestingly, MSH3, whose expression correlated with that of TDP43 in only a few tumor types using our threshold, displayed the strongest association among MMR genes, with a regression coefficient of 0.78 and a p-value of 2.2E-17 in uveal melanoma (*SI Appendix,* **Fig. S5*D***). From this analysis, we conclude that MSH2 and MSH6, which code for the MutSα complex for the repair of small loops and base mismatches, are strongly coexpressed with TDP43 in cancer.

Given the role of the MMR pathway in repairing DNA, and the coexpression of *TARDBP* with some MMR components, we asked if there would be differences in mutational loads in cancer between the g_high and g_low groups, for both *TARDBP* and MMR genes. For *TARDBP* the strongest differences were observed in invasive breast cancer, lung adenocarcinoma, bladder urothelial carcinoma, stomach adenocarcinoma, low grade glioma and liver hepatocellular carcinoma, in which the g_high group incurred more mutations than the g_low group. In two additional tumors of the kidney and endometrium the g_low group displayed fewer mutations than g_high (*SI Appendix,* **Fig. S5**). For the MMR genes, hierarchical clustering placed *MSH6* in the same branch as *TARDBP* and near *MSH2*, with tumors of the breast and lung characterized by strong increases in mutational burden in patients overexpressing *TARDBP*, *MSH6* and *MSH2* (**Fig. 7*D***). As anticipated, *TARDBP* was strongly coexpressed with *MSH6* and *MSH2* in invasive breast cancer and lung adenocarcinoma, with regression coefficients ranging from 0.35 to 0.47 and p-values from 7.56E-17 to 6.78E-62 (**Fig. 7*E***).

Finally, to further investigate the type of genes coexpressed with TDP43 (the 767 HS set, **Fig. 7*B***), we conducted a gene functional annotation analysis, which relies on the extent of overlap between a given set of genes and curated databases. The 767 HS gene set was strongly overrepresented in genes whose products undergo SUMOylation (**Fig. 8*A***), a post-translational modification that plays a crucial role in nuclear transport, DNA replication and repair, mitosis, and signal transduction, and which is associated with poor outcome in cancer (59). As predicted, other types of genes overrepresented in the 767 HS set were genes involved in the cell cycle, mitosis, splicing, kinetochore formation, DNA damage and repair, and DNA replication (**Fig. 8*A***). Thus, our analysis uncovered a link between TDP43 overexpression and the concomitant overexpression of genes whose processes are at the core of tumor growth. With a z-score of 2.75, the association between TDP43 and MMR was among the strongest also in normal human tissues, ranking fifth, as predicted by KEGG pathways, only preceded by splicing, RNA transport, DNA replication and cell cycle. As per individual genes, at a z-score >5 for predicted biological processes, both for *MSH2*, *MSH6* and *MLH1*, and for most of the top genes coexpressed with TDP43, the strongest association was with DNA unwinding involved in DNA replication (**Fig. 8*B***). In summary, our in-silico analysis confirms a strong association of TDP43 with the MMR pathway, which is seen to occur in the context of cell cycle regulation and protection of genome fidelity in normal cell, but which represent processes that are strongly activated in cancer to achieve uncontrolled cell division and high genome instability.

Discussion

Two major age-related disease areas with critical grand challenges for biology and medicine are the etiology and intervention for neurodegeneration and cancer. MMR balance has been implicated in both processes. This observation raises the question as to how MMR balance is dynamically regulated in cells. As TDP43 has been linked to DDR gene expression and found to have a direct role in DSB repair (13), we reasoned that TDP43 may be a key factor for actively regulating MMR balance. By employing the RT^2-Profiler assay to analyze HEK293 RNA extracts, we discovered a consistent decrease in the expression of several MMR genes following TDP43 KD. This phenomenon was not limited to a specific cell type; both NPCs and iMNs also exhibited suppressed MMR function upon TDP43 KD. Having established this clear phenotype, we sought to uncover the potential mechanisms.

TDP43 is widely recognized for its involvement in RNA processing. Likewise, its pathological mislocalization has been linked to detrimental transcriptomic alterations in both experimental and post-mortem ALS tissues (5, 47, 60-63). Investigating TDP43's effect on MMR transcripts, we utilized *in silico* tools and manual identification of UG-rich repeat regions to predict TDP43 binding sites on *MLH1* and *MSH6* transcripts (64, 65). We chose these genes for their clear binding potential and consistent co-expression with TDP43. We then employed end-point RT-PCR with primers designed to amplify the identified regions to ascertain whether loss of TDP43 led to the inclusion of intron sequences. Based on the sizes of amplicon products, we conclude that TDP43 KD negatively affects transcriptional processing of these genes. Furthermore, we observed this impact also occurred in cell models expressing different ALS-associated TDP43 mutations, as well as from RNA samples isolated from Guamanian ALS tissues. Significantly, these results align with previous reports identifying TDP43-dependent splicing of ALS disease-related genes (5, 47, 49, 60, 62, 63). Delving deeper into the consequences of mis-splicing and considering the potential of decreased transcript stability resulting from it, we tested the effect of TDP43 on the stability of MLH1 and MSH6 transcripts (66). Utilizing α-amanitin as an RNA polymerase II inhibitor, we conducted an RNA stability experiment that measured transcript abundance over time. Using this approach, we confirmed that TDP43 KD caused an accelerated decrease in MLH1 and MSH6 mRNA levels, suggesting a mechanism where TDP43 modulates MMR expression through transcript handling. Conversely, we observed the opposite effect on levels of TDP43 itself in accordance with its autoregulatory function (62).

Notably, our results highlighted the bidirectional nature of TDP43's influence on MMR expression. Differentiated cells showed a decrease in MMR protein levels, which was reversible by WT TDP43 overexpression. Interestingly, this overexpression not only declined MMR expression in nondividing cell, but also increased MMR expression to levels higher than that observed in dividing controls. Significantly, we also observed that MMR expression decreased with each stage of cellular differentiation of iPSCs into NSCs and then into motor neurons. To determine whether this effect was relevant to neurodegenerative disease, we measured MMR expression in brain tissue of two separate ALS mouse models. The first utilized low overexpression of murine Tdp43, while the second utilized a Tdp43 ΔNLS knock-in mutation without a change in copy number (57). Both models exhibited increased MMR expression, albeit with slight differences. The fact that our Tdp43 ΔNLS mouse model exhibited this phenotype was intriguing. We suspect mislocalization of TDP43 in these cells induces a relative increase in TDP43 via its autoregulatory feedback mechanism (62). This, in turn, increases MMR expression similar to that observed in the TDP43 overexpression model. We also repeated these tests in a third, more aggressive mouse model ALS that expresses very high levels of human TDP43 and severe early onset of disease. Unexpectedly,

we did not observe the widespread increase in MMR expression. We reason that this may be an artifact of the sudden and severe cellular pathology exhibited by this model that causes rapid apoptosis of affected cells. Finally, we measured the expression of MMR protein from CNS tissues of Guamanian ALS-affected patients. Despite limited technical hindrances, we reliably show that at least one MMR protein, MLH1, is significantly higher in the insoluble protein fraction of ALS isolates compared to controls. Taken together, these observations highlight the strong influence of TDP43 on MMR expression as it reverses natural MMR downregulation during cellular differentiation while increasing expression in models and tissues of TDP43 ALS.

When considering how TDP43 may effect these changes on MMR expression, two potential mechanisms may be considered. First, our findings showing that TDP43 modulates transcriptional handling of MLH1 and MSH6 suggest other MMR factors may be regulated in a similar manner. While we did not explicitly study how TDP43 influences transcript handling of every MMR gene, TDP43 is known to interact with hundreds of different transcripts (47); the true scope of its influence is likely much larger. Therefore, it is plausible that TDP43 may also directly affect other MMR transcripts in a manner not yet described. Second, TDP43 overexpression likely causes network level adjustments to the cellular transcriptome leading to a multitude of changes, of which increased MMR expression is a part. Our bioinformatics experiments showing the functional annotation of genes enriched by TDP43 overexpression support this notion. Several of the identified proteins were closely linked to cell cycle control, mitosis, splicing, kinetochore formation, DNA damage/repair, and DNA replication. This aligns with other reports on TDP43's contribution toward maintaining cancer cell stemness and cellular differentiation (67, 68). There are also implications for neurodegeneration. Specifically, Alzheimer's disease-affected neurons adopt an altered transcriptomic profile characterized by markers of cell cycle alteration and de-differentiation from the neuronal state (69). In both cases, TDP43 pathology is associated with disease-relevant alterations in the cellular transcriptome. Exhaustive dissection of the myriad consequences of these changes is still lacking. However, our data indicates MMR dysregulation may be a part of this complex system of events.

To better understand how MMR dysregulation may contribute to TDP43-linked disease processes, we utilized a previously reported neuronal cell model of mislocalized TDP43 expression (56). In this model, as in our Tdp43 ΔNLS mouse model, expression of mislocalized TDP43 increased MMR expression. To test the contribution of MMR, we chose to isolate its effect via KD of MSH2; this effectively diminishes MMR damage recognition capability and negatively affects expression and activity of other MMR factors (70-72). Using the sensitive technique of alkaline single cell gel electrophoresis, we observed MSH2 KD partially rescued Tdp43 ΔNLS induced DNA damage. Furthermore, we also showed MSH2 KD improved cellular capacity to repair oxidative DNA damage by glucose oxidase with concomitant decreases in markers of DDR signaling activation following a 6 h recovery period.

These findings may have important implications for both cancer and neurodegenerative pathologies. From a neuronal perspective, our results showing MMR enhancing Tdp43 ΔNLS-induced DNA damage, DDR signaling, and apoptosis are novel. They also align with existing reports of increased DDR signaling and DNA damage observed in neurodegenerative disease (73-75). Exactly how MMR exerts this effect is poorly understood. There are multiple reports suggesting MMR activity in nondividing cells is likely mutagenic, possibly due to a lack of strand discrimination signal (37, 38). This alone could be a direct source of increased DNA damage and DDR signaling. The response of MMR proteins to unique nucleic acid structures associated with neurodegenerative disease, such as R-Loops and trinucleotide repeats (TNR) mutations, is also

poorly understood. Exciting new work has linked MMR to the progression of TNR diseases such as Huntington's Disease (HD), Myotonic Dystrophy (MD), and Fragile X syndrome (FXS), among others (40, 42-45, 76, 77). MLH1, for example, was shown to play a direct role in the expansion of certain TNR mutations, leading to disease progression in a mouse model of HD (41). R-loops are another unique structure associated with ALS and other neurodegenerative diseases. These hybrid structures of RNA and DNA exist transiently at sites of heavy transcription and have been associated with TDP43 dysfunction, TNR expansion, and C9ORF72 ALS progression (78-83). Interestingly, MMR activity has also been linked to sites of heavy transcription (17, 84, 85), highlighting the importance of understanding how MMR may interact with R-Loop structures.

From a cancer perspective, loss of MMR is a well-known step along the path of carcinogenesis, but its loss is equally important for chemotherapeutic cancer cell killing. This second point is based on MMR-mediated DDR and apoptotic signaling (20, 21, 27, 30). It has been shown in multiple cancers of the endometrium, colon, and brain that dMMR impairs the cell killing ability of chemotherapy, especially for platinum-based, and alkylator agents (25, 26, 28, 29). Surprisingly, overexpressing MMR in dividing cells has also been linked to a mutator phenotype and DDR activation (31-33, 37, 86). For example, *in vitro* overexpression of either MLH1 or PMS2, disrupted normal MMR activity leading to increased mutations (31, 32). Our computational analyses revealed a strong association of TDP43 with the MMR pathway with clear trends of TDP43 and MMR co-expression. We furthermore found that TDP43 is often upregulated in cancer and confers a poor prognosis. These observations align with previous reports associating TDP43 with enhanced tumor progression and treatment resistance (87, 88). Taken together, our results reveal new mechanistic links for TDP43 with MMR and cancer pathology.

Our study is not without limitations. There remains a need to fully characterize how TDP43 affects the transcriptional handling of all MMR factors. Additionally, our exploration of MMR's contribution to DNA damage and TDP43 pathology was but the first step toward understanding the full consequence of this molecular relationship. Finally, our exploration of the TDP43-MMR relationship in cancer would benefit from in vitro testing of hypotheses generated from in silico analysis.

Currently, our collective data uncover a fundamental finding of TDP43-mediated control of MMR expression that extends across multiple cell models, animal models, and disease conditions. Our data, in combination with previous reports, present TDP43-mediated regulation of MMR as part of a shared cellular program with critical implications in both neurodegeneration and cancer. We anticipate that this discovery will serve as a foundation for future work testing how this relationship may potentially be leveraged to treat disease processes ranging from TNR mutations in HD to improved targeting of cancer stem cells. As the medical establishment must increasingly confront the harsh realities of caring for an aging population, the importance of understanding the underlying molecular mechanisms shared by these two major age-related disease classes with potential for positive interventions should not be understated.

Methods

Cell Culture

Human neuroblastoma SH-SY5Y cells (ATCC) or HEK293 (ATCC) cells were cultured in DMEM/F12 media (Gibco) or DMEM High-glucose (Gibco) with 10% Fetal Bovine Serum (Sigma) and 1% Penicillin-Streptomycin (Sigma) at 37 °C with 5% CO_2. SH-SY5Y cells were differentiated with Retinoic acid (RA) (10 μM) and BDNF (50 ng/mL) in DMEM/F12 and 1%

FBS for 4 to 6 days. WT human neural stem cells were (Jackson Labs) maintained on Geltrex LDEV-Free in Stem Pro medium (Gibco). Terminally differentiated neurons were derived from NPCs (Jackson Lab) WT (Cat # JIPSC1000) and TDP43-Q331K (Cat # JIPSC1066), according to established methods (89). Briefly, NSCs were plated at 2.5–5 $\times 10^4$ cells/cm^2 in 60 cm dish coated with Poly-D-Lysine and Laminin and initially maintained using StemPro media (Thermo Fisher) before transitioning to Neural Differentiation Media consisting of Neurobasal Medium (Gibco), 2% B-27 Serum-Free Supplement (Gibco), 2 mM GlutaMAX-I Supplement (Gibco). On day 7, dibutyryl cAMP (Sigma Cat #0627) was added at 10 μM concentration for 3 days to complete neural differentiation. Human iPSC to motor neuron differentiation. Control iPSC clones KYOU-DXR0109B (201B7) (ATCC) were initially plated on a 60-cm dish, then passaged to a T-25 flask with neuronal basic medium (mixture of 50% Neurobasal medium and 50% DMEM/F12 medium, with N2 and B27 supplements without vitamin A), following collagenase type IV digestion. After 2 days incubating with 5 μM ROCK Inhibitor (Y-27632, RI, from Merck Millipore), 40 μM TGF-β inhibitor (SB 431524, SB, Tocris Bioscience), 0.2 μM bone morphogenetic protein inhibitor (LDN-193189, LDN, from Stemgent), and 3 μM GSK-3 inhibitor (CHIR99021, CHIR, from Tocris Bioscience), suspended cell spheres were then incubated with a neuronal basic medium containing 0.1 μM retinoic acid (RA, from Sigma) and 500 nM Smoothened Agonist (SAG, from Merck Millipore) for 4 days. Cells were then incubated for 2 days in a neuronal basic medium containing RA, SAG, 10 ng/ml Brain-derived neurotrophic factor (BDNF, from Peprotech), and 10 ng/ml Glial cell-derived neurotrophic factor (GDNF, from Peprotech). Cell spheres were then dissociated with a neuronal basic medium containing trypsin (0.025%)/DNase in the water bath for 20 min at 37 °C. Then they were pipetted into single cells with the medium containing trypsin inhibitor (1.2 mg/mL). After cell counting, a defined number of cells were seeded into 20 μg/mL Laminin (Life technologies) -coated dishes or chamber slides and incubated for 5 days in a neuronal basic medium containing RA, SAG, BDNF, GDNF, and 10 μM DAPT, then incubated for 2 days in a neuronal basic medium containing BDNF, GDNF, and 20 μM Inhibitor of γ-secretase (DAPT, from Tocris Bioscience). For motor neuron maturation, cells were kept for over 7 days in a medium containing BDNF, GDNF, and 10 ng/ml ciliary neurotrophic factor (CNTF, from Peprotech).

Transfection, Plasmids, and siRNAs

SH-SY5Y and HEK293/T cells were transfected with siRNA using Lipofectamine 3000 (Thermo Fisher) per the manufacturer's instructions. NSCs were transfected using Lipofectamine Stem Reagent (Thermo Fisher) per the manufacturer's instructions. Gene KD was achieved using Silencer Select siRNAs (Thermo Fisher) for Control (Cat #4390843), TDP43 (Cat #4390771), MLH1 (Cat #AM51331), and MSH2 (Cat #AM16708).

Protein Isolation and Western blotting (WB)

Total cell lysates from SH-SY5Y, HEK293, NP cell lines, mouse cortical tissues, and human CNS tissues were extracted with RIPA lysis buffer (Millipore) containing the protease/phosphatase inhibitor cocktail (Thermo Fisher). Mouse and human tissue were first flash-frozen on liquid nitrogen and homogenized by mortar and pestle. After mixing with RIPA lysis buffer, total homogenates were probe sonicated at low amplitude, 3 times for 5 sec each, on ice. After centrifugation at 16000xg for 15 min at 4°C, supernatants containing the soluble protein fraction were decanted and quantified using Bradford assay. The remaining pellet containing the insoluble protein fraction was resuspended in 2% SDS in PBS buffer and probe sonicated twice before gel electrophoresis. Protein isolates were mixed with protein loading buffer containing 100 mM DTT

reducing agent and denatured for 5 min at 95°C before loading onto 4–12% Bis-Tris NuPAGE precast gels (Thermo Fisher) for electrophoresis. Following transfer to nitrocellulose membrane and incubating with primary and secondary antibodies, the protein signal was detected using the LI-COR XF imaging system and analyzed using Empiria Imaging Software (LI-COR). Where applicable, total protein stains of transferred blots were performed using Revert Total Protein Stain (LI-COR) and used to normalize target protein expression. Primary antibodies used: MLH1 (Cell Signaling #4256), MSH2 (Cell Signaling #2017), MSH3 (Protein Tech #22393-1-AP), MSH6 (Protein Tech #18120-1-AP), PMS2 (ABClonal #A6947), GAPDH (Cell Signaling #2118), γH2AX (Cell Signaling #2577), TDP43 (Protein Tech #10782-2-AP), FLAG (Sigma #F1804), cleaved-caspase-3 (Cell Signaling #9664), and PARP-1 (Cell Signaling #9542).

Cell Death/Survival Assay

MTT assay: Treated or untreated HEK cells exposed to 6-thioguanine (10 μM) or methylmethanesulfonate (1 mM) were seeded in 96-well ELISA plates (Corning) in triplicates. Cultures were incubated 48 h post post-treatment. MTT assay was performed following the manufacturer's protocol (TREVIGEN). Briefly, 10 μL of 5 mg/mL MTT reagent (Sigma) was added to each well and incubated for 24 h, then 100 μL of 10% SDS in PBS detergent (Thermo) was added. Reagents on plates were gently mixed by rotational agitation for 10 min before measuring the absorbance at 570 nm using a microplate reader (Bio-Rad, 680 XR).

Plasmid DNA and Inducible Cell Lines

Doxycycline-inducible TDP43 WT, ΔNLS, Q331K, A315T, or M333V model was generated in the SH-SY5Y cell line as described elsewhere (2, 13). Briefly, FLAG-tagged TDP43 WT or mutant coding DNA sequence (CDS) was PCR amplified and cloned into a doxycycline-inducible pCW vector (Addgene #50661) at NheI (NEB) and BamHI (NEB) sites.

Mouse Models

The endogenous Tdp43ΔNLS mouse model was generated by CRISPR/Cas9-mediated knock-in of NLS deleted Exon3 of mouse *Tardbp* gene flanked by loxP sequences within the intronic region of the gene. Conditional expression of Cre recombinase driven by the Mnx1 promoter (motor neuron-specific) in the bigenic Cre::Tdp43ΔNLS mice induced the neuron-specific expression of murine Tdp43ΔNLS, resulting in the manifestation of human ALSTDP-43-type proteinopathy in the central nervous system (https://doi.org/10.21203/rs.3.rs-3879966/v1). Brain tissue samples from a moderately overexpressing murine Tdp43 ALS mouse model were kindly provided by the laboratory of co-author Dr. Zuoshang Xu (57). Mice ages ranged from 94 to 396 days. Animals were anesthetized using isoflurane administered through a SomnoSuite Low Flow Anesthesia System (Kent Scientific), then transcardially perfused with ice-cold phosphate-buffered saline (PBS, Thermo). Brain tissues were dissected from the cranium and snap-frozen over liquid nitrogen. Brain tissues from five subjects of ages ranging from 94 d to 296 d (use d or days throughout) from each of the following groups were used for analysis for a total of 20 mice, including the control-male, control-female, transgenic-male, and transgenic-female groups.

Protein Isolation from Human Decedent Tissues

Autopsied cortex and spinal cord tissue samples of Guamanian ALS and age-matched controls from the same ethnic background were acquired from the Binghamton University Biospecimen Archive. The clinical characteristics of the decedents are described in Supplementary Table S4.

Snap-frozen cortical tissue specimens were first ground into powder using liquid nitrogen in a mortar-pestle. For total protein extraction, about 20 mg of tissue powder from each sample was lysed using 1X RIPA buffer added with cocktail protease and phosphatase inhibitors (Thermo). Samples were incubated on ice for 30 min, and subsequently treated with 10 sec-pulse sonication three times at 5 amplitudes on ice. Samples were then centrifuged at 13000 rpm at 4°C for 15 min three times to remove most of the brain fats. Finally, the soluble and insoluble fractions of each sample were collected in separate tubes. The insoluble pellet was then resuspended in 2% SDS buffer (2% SDS, 50 mM Tris-HCl, pH 7.4). Protein concentration was determined using the Bradford assay (Sigma), according to the manufacturer's instructions.

Comet Assay

Alkaline comet assay was performed using SH-SY5Y cells treated with siControl or siTDP43 for 96 h as well as TDP43ΔNLS or WT cells induced with doxycycline for 72 h, according to the manufacturer's protocol (TREVIGEN). Briefly, cells were mixed with low-melt agarose (TREVIGEN) and placed on comet assay slides (TREVIGEN), followed by overnight incubation in lysis buffer at 4°C in the dark and single cell electrophoresis at 21 V for 30 min in alkaline electrophoresis buffer (TREVIGEN) under manufacturer-recommended conditions. Slides were finally subjected to staining with SYBR® Gold for visualization of genomic DNA under AXIO Observer inverted microscope (Carl Zeiss).

Immunofluorescence (IF)

IPSC-derived iMNs were cultured in 8-well chamber slides (Millicell EZ slides, Millipore) and fixed with 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS) for 30 min, then permeabilized with 0.2% Tween-20 in PBS for 15 min at room temperature. Samples were blocked with 3% BSA in PBS for 1 h. Primary antibody incubation was done with anti-TDP43 (Protein Tech #10782-2-AP) and anti-PMS2 (ABClonal #A6947) antibodies overnight at 4°C in 1% BSA solution, followed by washing three times prior to staining with fluorophore-tagged corresponding secondary antibodies for 1 h at room temperature in 1% BSA solution. Slides were then washed three times and nuclei were counterstained with DAPI. Images were captured with the AXIO Observer fluorescence microscope (Carl Zeiss).

RT^2-Profiler Assay

Complementary DNA (cDNA) was synthesized using total RNA isolated from siTDP43 or siControl-treated HEK293 cells. Samples were analyzed by SYBR green-based quantitative real-time (qRT) RT^2 PCR arrays for evaluating expressions of a pre-defined panel of DNA damage response (DDR) genes (Qiagen, Cat#: PAHS-042Z). The 96-well RT^2 profiler plate contained primers for 84 DDR genes, 5 housekeeping genes, and 3 negative controls (*SI Appendix* **Fig. S1** ***A***).

qRT-PCR and End-Point PCR

Approximately $2x10^6$ cells were collected and centrifuged at 4°C at 2000 rpm for 3 min for extraction of RNA. Total RNA was isolated using Trizol (Thermo Scientific, Waltham, MA, USA) according to manufacturer instructions. RNA purity and quantification were determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, US). Total RNA was reverse transcribed into cDNA using the Super Script Vilo Kit (Thermo Scientific, Waltham, MA, USA). The qRT-PCR amplification was performed in triplicate using the ABI 7500 (Applied Biosystems, Foster City, CA, USA) using the PowerUp SYBER Green Master Mix (Thermo

Scientific, Waltham, MA, USA). *HPRT* and *GAPDH* housekeeping genes were utilized as internal controls. Primers were purchased from Sigma and are shown in *SI Appendix* **Table S2** and **S3**. The relative expression of each validated gene was determined using the $2^{-\Delta\Delta Ct}$ method. Two-sided student's t-test was performed and results with $p<0.05$ (*) were statistically significant.

Transcript Splicing Analysis

cDNA prepared from total RNA isolates of control and TDP43 KD HEK293 cells were used for MLH1 and MSH6 splicing analysis using End-Point PCR. To determine whether intronic sequences were included in the transcripts of TDP43 KD cells, primers were designed to anneal to two consecutive exon sequences surrounding one intronic sequence. The intervening cDNA sequence was then amplified using LongAmp Taq polymerase (NEB). Primers for each gene region are shown in Supplementary Table S3. Amplicons were then subject to electrophoresis on 2% agarose gel (Sigma) and stained with ethidium bromide (Fisher Scientific) before visualization using the LI-COR XF Imaging System.

RNA Stability Assay

The stability of target RNA transcripts was assessed as previously described (90, 91). SH-SY5Y cells underwent two treatments with Silencer Select siRNA against TDP43 (Thermo #4392420) on days 6 and 7 after seeding cells on the plate. On day 8, control cells were harvested, while the experimental cells were pre-treated with 10 μM of α-amanitin (Sigma) prior to harvesting at 8 h, 12 h, and 24 h timepoints. cDNA was prepared from total RNA and quantified by qRT-PCR. Normalization of C_T values was performed using 18s rRNA as the internal control. Fold-change at each timepoint was calculated, relative to the control, and values were fitted to a one-phase decay plot to determine the half-life of each transcript under consideration using GraphPad Prism software.

Bioinformatic Analysis

TCGA RNA-seq transcriptomic data and clinical patient data were obtained using the R package TCGA-assembler v. 2 (92) and processed with custom scripts to obtain gene expression data in both tumors and matched controls, as well as Kaplan-Meier survival curves (93). The tumor abbreviations used were: ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; DLBC, lymphoid neoplasm diffuse large B-cell lymphoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LAML, acute myeloid leukemia; LGG, brain lower grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors, THCA, thyroid carcinoma; THYM, thymoma, UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma.

Data were plotted using the R libraries "ggplot2", "ggpubr", "extrafont", "dplyr", "survival", and "survminer" and SigmaPlot (https://systatsoftware.com/sigmaplot/), and further edited in Canvas (https://www.canvasgfx.com/products/canvas-x-draw). Functional annotations of gene sets were performed with DAVID (https://david.ncifcrf.gov/); ARCHS4 (https://maayanlab.cloud/archs4/)

was used to retrieve "Predicted biological processes (GO)", "Predicted pathways (KEGG)", and "Most similar genes based on co-expression" from large sets of RNA-seq data for *TARDB* and key MMR-related genes. For correlations between gene expression and mutational burden in TCGA, for each gene and each tumor type, patients were divided into two groups, a g_high group with RNA-seq data above population mean, and a g_low group with RNA-seq data below population mean. TCGA patients codes for the g_high and g_low groups were then intersected with the same patients codes from the Catalogue Of Somatic Mutations In Cancer (COSMIC, https://www.sanger.ac.uk/tool/cosmic/; CosmicMutantExport.tsv file, version 92) to retrieve the curated list of simple somatic mutations, single bases substitutions and small insertions and deletions, in the tumor samples, and the differences in mutational loads (log10 of number of mutations) between g_high and g_low were assessed with t tests. Matching COSMIC patient codes were available for 24 of the 33 TCGA tumor types. Heatmap of statistical differences in mutational loads was generated with Heatmapper (http://heatmapper.ca). Values used for the heatmap were the -log10 p-values for h_high>g_low, and log10 p-values for g_high<g_low. Scores were reported as z-score, i.e. the number of standard deviations each value distanced from the mean across the 24 tumor types. C++ codes for t-tests and linear regressions were used in pipeline custom scripts.

Acknowledgements

This research is primarily supported by the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institute of Aging (NIA) of the National Institutes of Health (NIH) under award number RF1NS112719 to M.L.H. Research in the Hegde laboratory is also supported by the NIH awards R01NS088645, R03AG064266, and R01NS094535, as well as the Sherman Foundation Parkinson's Disease Research Challenge Fund and the Houston Methodist Research Institute's internal funds. M.L.H also acknowledges Everett E. and Randee K. Bernal for their support via Centennial Endowed Chair of Neurological Institute. Efforts by J.A.T and A.B are supported by NIH P01 CA092584 and R35 CA220430, Cancer Prevention and Research Institute of Texas RP180813 and Robert A. Welch Chemistry Chair (G-0010). This work used HPC resources at the Texas Advanced Computing Center (TACC) at The University of Texas at Austin (URL: http://www.tacc.utexas.edu). The authors thank Dr. Anna Dodson at Houston Methodist Research Institute (Houston, TX) for assistance with document editing.

Author Contributions

V.E.P. performed experiments with assistance from S.R., J.M., M.K., V.M and V.V. Additionally, V.E.P. conducted data analysis with interpretation and wrote the manuscript. S.M., R.G., Z.X., and J.A.T. provided valuable tissues, intellectual insights and manuscript feedback. A.B. performed bioinformatics analysis and wrote results for Figs. 6 and 7. M.L.H. supervised the study and prepared the final manuscript. All authors participated in the discussion and provided valuable feedback on the manuscript.

Competing Interests Statement

The authors declare no competing or financial interests.

Data Availability Statement

Raw data for experiments in this study can be provided upon reasonable request to the corresponding author. A copy of normalized TCGA RNA-seq data is available at https://doi.org/10.5281/zenodo.7885656. Source codes used for the TCGA bioinformatic analyses are available at https://doi.org/10.5281/zenodo.7874703. C++ codes for t-test and linear regression are available at https://github.com/abacolla.

References

1. E. N. Guerrero *et al.*, TDP-43/FUS in motor neuron disease: Complexity and challenges. *Prog Neurobiol* **145-146**, 78-97 (2016).
2. E. N. Guerrero *et al.*, Amyotrophic lateral sclerosis-associated TDP-43 mutation Q331K prevents nuclear translocation of XRCC4-DNA ligase 4 complex and is linked to genome damage-mediated neuronal apoptosis. *Hum. Mol. Genet.* **28**, 2459-2476 (2019).
3. C. Colombrita *et al.*, RNA-binding proteins and RNA metabolism: a new scenario in the pathogenesis of Amyotrophic lateral sclerosis. *Arch Ital Biol* **149**, 83-99 (2011).
4. M. Uhlén *et al.*, Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
5. M. Hallegger *et al.*, TDP-43 condensation properties specify its RNA-binding and regulatory repertoire. *Cell* **184**, 4680-4696.e4622 (2021).
6. X. Ma *et al.*, The Regulatory Role of RNA Metabolism Regulator TDP-43 in Human Cancer. *Front Oncol* **11**, 755096 (2021).
7. L. De Conti *et al.*, TDP-43 affects splicing profiles and isoform production of genes involved in the apoptotic and mitotic cellular pathways. *Nucleic Acids Res* **43**, 8990-9005 (2015).
8. C. A. Droppelmann, D. Campos-Melo, M. Ishtiaq, K. Volkening, M. J. Strong, RNA metabolism in ALS: when normal processes become pathological. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* **15**, 321-336 (2014).
9. B. D. Freibaum, R. K. Chitta, A. A. High, J. P. Taylor, Global analysis of TDP-43 interacting proteins reveals strong association with RNA splicing and translation machinery. *J. Proteome Res.* **9**, 1104-1120 (2010).
10. J. Mandrioli, L. Mediani, S. Alberti, S. Carra (2020) ALS and FTD: Where RNA metabolism meets protein quality control. in *Semin. Cell Dev. Biol.* (Elsevier), pp 183-192.
11. A. L. Brown *et al.*, TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. *Nature* **603**, 131-137 (2022).
12. M. W. Baughn *et al.*, Mechanism of STMN2 cryptic splice-polyadenylation and its correction for TDP-43 proteinopathies. *Science* **379**, 1140-1149 (2023).
13. J. Mitra *et al.*, Motor neuron disease-associated loss of nuclear TDP-43 is linked to DNA double-strand break repair defects. *Proc. Nat. Acad. Sci. U.S.A.* **116**, 4696-4705 (2019).
14. T. A. Kunkel, D. A. Erie, Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu Rev Genet* **49**, 291-313 (2015).
15. R. Fishel, Mismatch repair. *J Biol Chem* **290**, 26395-26403 (2015).
16. N. Pecina-Slaus, A. Kafka, I. Salamon, A. Bukovac, Mismatch Repair Pathway, Genome Stability and Cancer. *Front Mol Biosci* **7**, 122 (2020).
17. Y. Huang, G. M. Li, DNA mismatch repair in the chromatin context: Mechanisms and therapeutic potential. *DNA Repair (Amst)* **93**, 102918 (2020).

18. J. Ortega, G. S. Lee, L. Gu, W. Yang, G. M. Li, Mispair-bound human MutS-MutL complex triggers DNA incisions and activates mismatch repair. *Cell Res.* **31**, 542-553 (2021).
19. Y. Zhang *et al.*, Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell* **122**, 693-705 (2005).
20. Z. Li, A. H. Pearlman, P. Hsieh, DNA mismatch repair and the DNA damage response. *DNA Repair (Amst)* **38**, 94-101 (2016).
21. D. Gupta, C. D. Heinen, The mismatch repair-dependent DNA damage response: Mechanisms and implications. *DNA Repair (Amst)* **78**, 60-69 (2019).
22. M. A. Edelbrock, S. Kaliyaperumal, K. J. Williams, Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities. *Mutat Res* **743-744**, 53-66 (2013).
23. G. M. Li, The role of mismatch repair in DNA damage-induced apoptosis. *Oncol Res* **11**, 393-400 (1999).
24. G. M. Li, Mechanisms and functions of DNA mismatch repair. *Cell Res* **18**, 85-98 (2008).
25. B. L. Adamsen, K. L. Kravik, P. M. De Angelis, DNA damage signaling in response to 5-fluorouracil in three colorectal cancer cell lines with different mismatch repair and TP53 status. *Int J Oncol* **39**, 673-682 (2011).
26. L. Tentori *et al.*, Influence of MLH1 on colon cancer sensitivity to poly(ADP-ribose) polymerase inhibitor combined with irinotecan. *Int J Oncol* **43**, 210-218 (2013).
27. A. Sawant, A. Kothandapani, A. Zhitkovich, R. W. Sobol, S. M. Patrick, Role of mismatch repair proteins in the processing of cisplatin interstrand cross-links. *DNA repair* **35**, 126-136 (2015).
28. Y. Li *et al.*, MLH1 enhances the sensitivity of human endometrial carcinoma cells to cisplatin by activating the MLH1/c-Abl apoptosis signaling pathway. *BMC Cancer* **18**, 1294 (2018).
29. J. Stritzelberger, L. Distel, R. Buslei, R. Fietkau, F. Putz, Acquired temozolomide resistance in human glioblastoma cell line U251 is caused by mismatch repair deficiency and can be overcome by lomustine. *Clin Transl Oncol* **20**, 508-516 (2018).
30. D. Fink *et al.*, The role of DNA mismatch repair in platinum drug resistance. *Cancer Res* **56**, 4881-4886 (1996).
31. S. L. Gibson *et al.*, Overexpression of the DNA mismatch repair factor, PMS2, confers hypermutability and DNA damage tolerance. *Cancer Lett* **244**, 195-202 (2006).
32. P. V. Shcherbakova, T. A. Kunkel, Mutator Phenotypes Conferred by MLH1Overexpression and by Heterozygosity for mlh1Mutations. *Mol. Cell. Biol.* **19**, 3177-3183 (1999).
33. H. Zhang *et al.*, Apoptosis induced by overexpression of hMSH2 or hMLH1. *Cancer Res* **59**, 3021-3027 (1999).
34. R. Chahwan, W. Edelmann, M. D. Scharff, S. Roa, Mismatch-mediated error prone repair at the immunoglobulin genes. *Biomed Pharmacother* **65**, 529-536 (2011).
35. P. Hsieh, Molecular mechanisms of DNA mismatch repair. *Mutat Res* **486**, 71-87 (2001).
36. J. Pena-Diaz, J. Jiricny, Mammalian mismatch repair: error-free or error-prone? *Trends Biochem Sci* **37**, 206-214 (2012).
37. G. P. Rodriguez *et al.*, Mismatch repair-dependent mutagenesis in nondividing cells. *Proc Natl Acad Sci U S A* **109**, 6153-6158 (2012).

38. G. F. Crouse, Non-canonical actions of mismatch repair. *DNA Repair (Amst)* **38**, 102-109 (2016).
39. P. J. Brooks, C. Marietta, D. Goldman, DNA mismatch repair and DNA methylation in adult brain neurons. *J Neurosci* **16**, 939-945 (1996).
40. N. Keogh, K. Y. Chan, G. M. Li, R. S. Lahue, MutSbeta abundance and Msh3 ATP hydrolysis activity are important drivers of CTG*CAG repeat expansions. *Nucleic Acids Res* **45**, 10068-10078 (2017).
41. R. M. Pinto *et al.*, Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. *PLoS Genet* **9**, e1003930 (2013).
42. R. A. Lokanga, X. N. Zhao, K. Usdin, The mismatch repair protein MSH2 is rate limiting for repeat expansion in a fragile X premutation mouse model. *Hum Mutat* **35**, 129-136 (2014).
43. R. R. Iyer, A. Pluciennik, DNA Mismatch Repair and its Role in Huntington's Disease. *J Huntingtons Dis* **10**, 75-94 (2021).
44. A. Halabi, S. Ditch, J. Wang, E. Grabczyk, DNA mismatch repair complex MutSbeta promotes GAA.TTC repeat expansion in human cells. *J Biol Chem* **287**, 29958-29967 (2012).
45. M. H. M. Schmidt, C. E. Pearson, Disease-associated repeat instability and mismatch repair. *DNA Repair (Amst)* **38**, 117-126 (2016).
46. J. C. Rengifo-Gonzalez *et al.*, The cooperative binding of TDP-43 to GU-rich RNA repeats antagonizes TDP-43 aggregation. *eLife* **10** (2021).
47. M. C. Cao, E. L. Scotter, Transcriptional targets of amyotrophic lateral sclerosis/frontotemporal dementia protein TDP-43 - meta-analysis and interactive graphical database. *Dis Model Mech* **15** (2022).
48. C. C. Chou *et al.*, TDP-43 pathology disrupts nuclear pore complexes and nucleocytoplasmic transport in ALS/FTD. *Nature neuroscience* **21**, 228-239 (2018).
49. F. C. Fiesel, S. S. Weber, J. Supper, A. Zell, P. J. Kahle, TDP-43 regulates global translational yield by splicing of exon junction complex component SKAR. *Nucleic Acids Res* **40**, 2668-2682 (2012).
50. X. Pan, Y. Fang, X. Li, Y. Yang, H. B. Shen, RBPsuite: RNA-protein binding sites prediction suite based on deep learning. *BMC Genomics* **21**, 884 (2020).
51. I. Paz, I. Kosti, M. Ares, Jr., M. Cline, Y. Mandel-Gutfreund, RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res* **42**, W361-367 (2014).
52. H. Sidibé *et al.*, TDP-43 stabilizes G3BP1 mRNA: relevance to amyotrophic lateral sclerosis/frontotemporal dementia. *Brain* **144**, 3461-3476 (2021).
53. K. Izumikawa *et al.*, TDP-43 stabilises the processing intermediates of mitochondrial transcripts. *Sci Rep* **7**, 7709 (2017).
54. J.-E. Deshaies *et al.*, TDP-43 regulates the alternative splicing of hnRNP A1 to yield an aggregation-prone variant in amyotrophic lateral sclerosis. *Brain* **141**, 1320-1333 (2018).
55. Y. M. Ayala *et al.*, TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J.* **30**, 277-288 (2011).
56. Mitra J., Kodavati M.K., Guerrero E.N., Rao K.S., Garruto R.M., and Hegde M.L., Endogenous TDP-43 mislocalization in a novel knock-in mouse model reveals DNA

repair impairment, inflammation, and neuronal senescence. *Research square* https://www.researchsquare.com/article/rs-3879966/v1 (2024).

57. C. Yang *et al.*, Low-level overexpression of wild type TDP-43 causes late-onset, progressive neurodegeneration and paralysis in mice. *PloS one* **17**, e0255710 (2022).

58. T. Yan, S. E. Berry, A. B. Desai, T. J. Kinsella, DNA mismatch repair (MMR) mediates 6-thioguanine genotoxicity by introducing single-strand breaks to signal a G2-M arrest in MMR-proficient RKO cells. *Clinical Cancer Research* **9**, 2327-2334 (2003).

59. A. Biederstadt *et al.*, SUMO pathway inhibition targets an aggressive pancreatic cancer subtype. *Gut* **69**, 1472-1482 (2020).

60. M. W. Baughn *et al.*, Mechanism of STMN2 cryptic splice-polyadenylation and its correction for TDP-43 proteinopathies. *Science* **379**, 1140-1149 (2023).

61. J. Humphrey *et al.*, Integrative transcriptomic analysis of the amyotrophic lateral sclerosis spinal cord implicates glial activation and suggests new risk genes. *Nat Neurosci* **26**, 150-162 (2023).

62. M. Polymenidou *et al.*, Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat Neurosci* **14**, 459-468 (2011).

63. J. R. Tollervey *et al.*, Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat Neurosci* **14**, 452-458 (2011).

64. Y. M. Ayala *et al.*, Human, Drosophila, and C. elegans TDP43: nucleic acid binding properties and splicing regulatory function. *Journal of molecular biology* **348**, 575-588 (2005).

65. P. J. Lukavsky *et al.*, Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* **20**, 1443-1449 (2013).

66. D. Nikom, S. Zheng, Alternative splicing in neurodegenerative disease and the promise of RNA therapies. *Nat Rev Neurosci* **24**, 457-473 (2023).

67. L. Guo *et al.*, TDP43 promotes stemness of breast cancer stem cells through CD44 variant splicing isoforms. *Cell Death Dis* **13**, 428 (2022).

68. X. Tang *et al.*, Activation of PPAR-β/δ Attenuates Brain Injury by Suppressing Inflammation and Apoptosis in a Collagenase-Induced Intracerebral Hemorrhage Mouse Model. *Neurochem. Res.* **45**, 837-850 (2020).

69. J. Mertens *et al.*, Age-dependent instability of mature neuronal fate in induced neurons from Alzheimer's patients. *Cell Stem Cell* **28**, 1533-1548 e1536 (2021).

70. P. Friedhoff, P. Li, J. Gotthardt, Protein-protein interactions in DNA mismatch repair. *DNA Repair (Amst)* **38**, 50-57 (2016).

71. S. Hassen *et al.*, Interdependence of DNA mismatch repair proteins MLH1 and MSH2 in apoptosis in human colorectal carcinoma cell lines. *Mol Cell Biochem* **412**, 297-305 (2016).

72. R. R. Iyer, A. Pluciennik, V. Burdett, P. L. Modrich, DNA mismatch repair: functions and mechanisms. *Chem Rev* **106**, 302-323 (2006).

73. S. J. Hill *et al.*, Two familial ALS proteins function in prevention/repair of transcription-associated DNA damage. *Proc Natl Acad Sci U S A* **113**, E7701-E7709 (2016).

74. J. R. Kok, N. M. Palminha, C. Dos Santos Souza, S. F. El-Khamisy, L. Ferraiuolo, DNA damage as a mechanism of neurodegeneration in ALS and a contributor to astrocyte toxicity. *Cell Mol Life Sci* **78**, 5707-5729 (2021).

75. H. Wang, M. Kodavati, G. W. Britz, M. L. Hegde, DNA Damage and Repair Deficiency in ALS/FTD-Associated Neurodegeneration: From Molecular Mechanisms to Therapeutic Implication. *Front. Mol. Neurosci.* **14**, 784361 (2021).
76. N. Chatterjee, Y. Lin, J. H. Wilson, Mismatch repair enhances convergent transcription-induced cell death at trinucleotide repeats by activating ATR. *DNA Repair (Amst)* **42**, 26-32 (2016).
77. V. Ezzatizadeh *et al.*, The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model. *Neurobiol Dis* **46**, 165-171 (2012).
78. J. Cuartas, L. Gangwani, R-loop Mediated DNA Damage and Impaired DNA Repair in Spinal Muscular Atrophy. *Front Cell Neurosci* **16**, 826608 (2022).
79. C. H. Freudenreich, R-loops: targets for nuclease cleavage and repeat instability. *Curr Genet* **64**, 789-794 (2018).
80. E. E. Laverde *et al.*, R-loops promote trinucleotide repeat deletion through DNA base excision repair enzymatic activities. *J Biol Chem* **295**, 13902-13913 (2020).
81. K. Reddy *et al.*, Processing of double-R-loops in (CAG).(CTG) and C9orf72 (GGGGCC).(GGCCCC) repeats causes instability. *Nucleic Acids Res* **42**, 10473-10487 (2014).
82. X. A. Su, C. H. Freudenreich, Cytosine deamination and base excision repair cause R-loop-induced CAG repeat fragility and instability in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **114**, E8392-E8401 (2017).
83. M. Wood *et al.*, TDP-43 dysfunction results in R-loop accumulation and DNA replication defects. *J Cell Sci* **133** (2020).
84. Y. Huang, L. Gu, G. M. Li, H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation. *J Biol Chem* **293**, 7811-7823 (2018).
85. Y. Huang, G. M. Li, DNA mismatch repair preferentially safeguards actively transcribed genes. *DNA Repair (Amst)* **71**, 82-86 (2018).
86. S. T. Bak, D. Sakellariou, J. Pena-Diaz, The dual nature of mismatch repair as antimutator and mutator: for better or for worse. *Front Genet* **5**, 287 (2014).
87. H. Ke *et al.*, Loss of TDP43 inhibits progression of triple-negative breast cancer in coordination with SRSF3. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3426-e3435 (2018).
88. H. L. Zhao *et al.*, Long noncoding RNA MIAT promotes the growth and metastasis of non-small cell lung cancer by upregulating TDP43. *Eur Rev Med Pharmacol Sci* **24**, 7209 (2020).
89. H. Wang *et al.*, Mutant FUS causes DNA ligation defects to inhibit oxidative damage repair in Amyotrophic Lateral Sclerosis. *Nat. Commun.* **9**, 3683 (2018).
90. J. E. Deshaies *et al.*, TDP-43 regulates the alternative splicing of hnRNP A1 to yield an aggregation-prone variant in amyotrophic lateral sclerosis. *Brain* **141**, 1320-1333 (2018).
91. M. Ratnadiwakara, M. L. Änkö, mRNA Stability Assay Using transcription inhibition by Actinomycin D in Mouse Pluripotent Stem Cells. *Bio Protoc* **8**, e3072 (2018).
92. L. Wei *et al.*, TCGA-assembler 2: software pipeline for retrieval and processing of TCGA/CPTAC data. *Bioinformatics* **34**, 1615-1617 (2018).
93. A. Bacolla, J. A. Tainer, Robust Computational Approaches to Defining Insights on the Interface of DNA Repair with Replication and Transcription in Cancer. *Methods Mol Biol* **2444**, 1-13 (2022).

Figure Legends

Figure 1

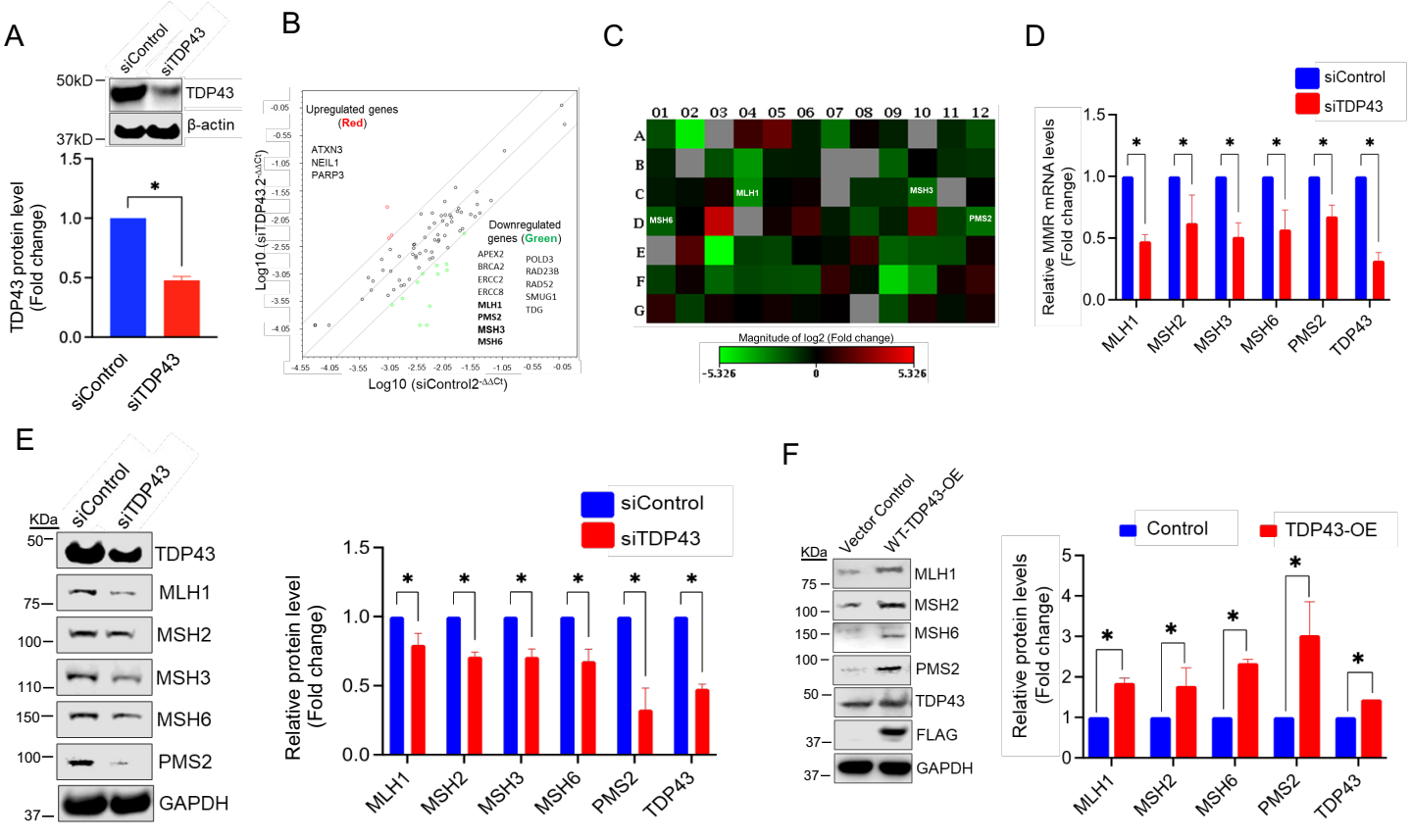

Fig. 1. **TDP43 Modulates the Expression of DNA Mismatch Repair (MMR) factors**.
(**A**) Western blot analysis of TDP43 knockdown (KD) in siControl or siTDP43 treated HEK293 cells. β-actin served as the loading control. Quantitation of relative TDP43 protein levels is shown in the bottom. (**B**) Scatter plot of RT2-Profiler DNA Repair Expression Array representing genes in siControl or siTDP43-treated cells that display >2-fold difference in mRNA expressions compared with that of control cells. Red, green, and black circles indicate upregulated genes, downregulated genes, and nonregulated genes, respectively. (**C**) Heatmap highlighting altered expressions of MMR and other DNA repair genes in siControl versus siTDP43 cells. Red, green, and black squares indicate upregulated genes, downregulated genes, and unaffected genes, respectively. (**D**) Histogram showing the relative mRNA transcript levels of MMR factors in siControl or siTDP43-treated HEK293 cells. (**E**) Western blot (WB) images illustrate the comparison of MMR protein expressions between the siControl and siTDP43-treated HEK293 cells. And the quantitation of relative protein levels in fold changes. (**F**) WB analysis of the expression of indicated MMR factors with or without TDP43 OE in SH-SY5Y cells. Quantitation of relative protein levels in fold changes, normalized to the GAPDH level shown in histogram (right). Significance values (p-values) are as follows: $p > 0.5$ (ns), $p < 0.5$ (*). Error bars indicate mean ± SEM, from three independent biological replicates.

Figure 2

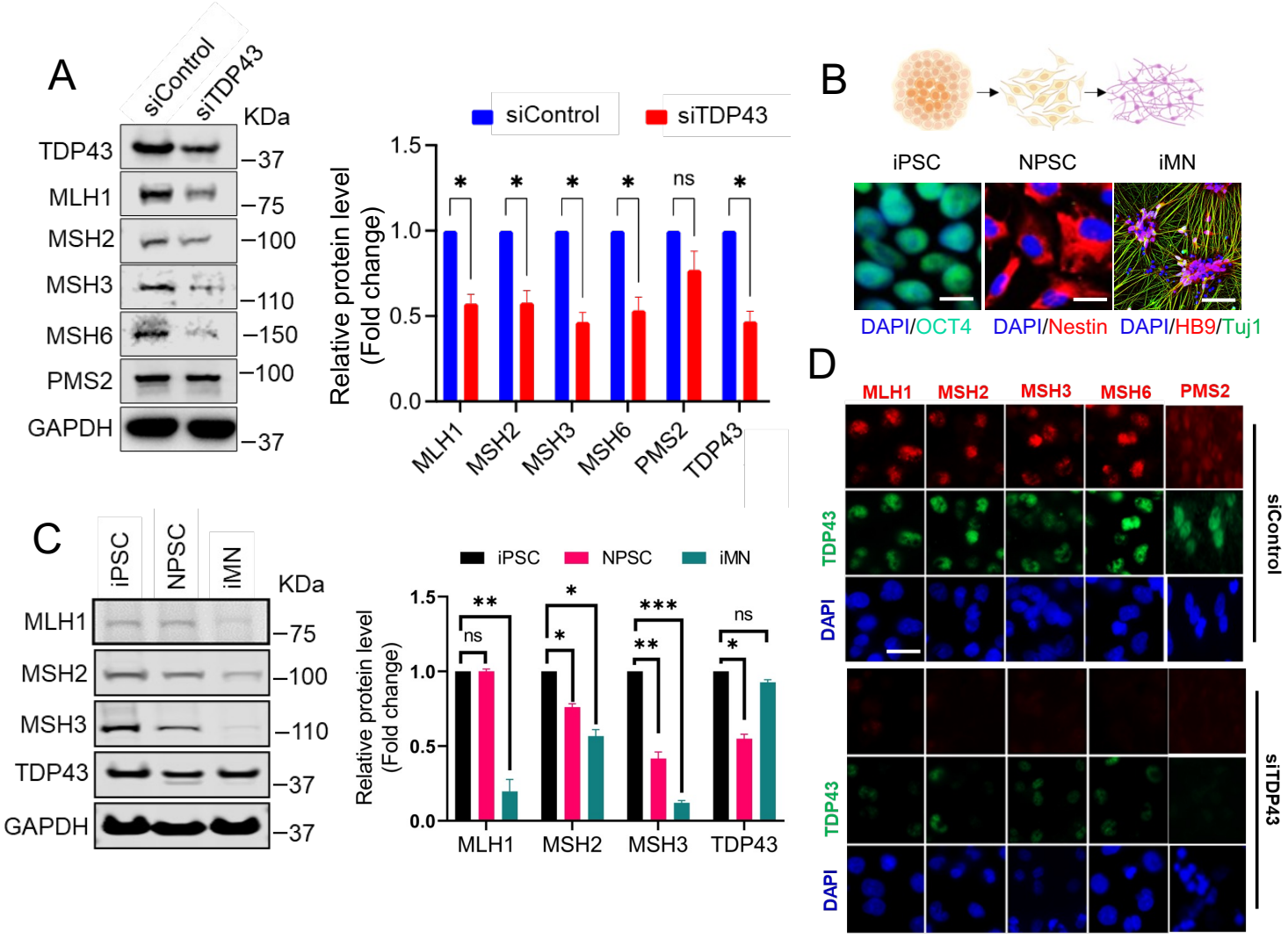

Fig. 2. Differential expressions of MMR genes in iPSC-derived neuronal cells and their regulation by TDP43. (**A**) WB images and histogram illustrating a quantitative comparison of MMR expression in control and siTDP43-treated human pluripotent stem cell (iPSC)-derived neural lineage progenitor stem cells (NPSCs). Quantitation of protein levels normalized to that of GAPDH. (**B**) Schematic of iPSC, NPSC, and terminally differentiated motor neurons (iMN) utilized in this study. (**C**) WB images show the expression of select MMR factors across the following states of cellular differentiation: iPSC, NPSC, iMN. **D**) Immunofluorescence (IF) images revealing the expression of TDP43 (green); and MLH1 (red), MSH2 (red), MSH3 (red), MSH6 (red) and PMS2 (red), and nuclear DNA (DAPI-blue) after siTDP43-mediated KD of TDP43 in iPSC-derived iMNs. Significance values (p-values) are as follows $p > 0.5$ (ns), $p < 0.5$ (*), $p < 0.01$ (**), $p < 0.001$ (***). Error bars indicate mean ± SEM from three independent experiments.

Figure 3

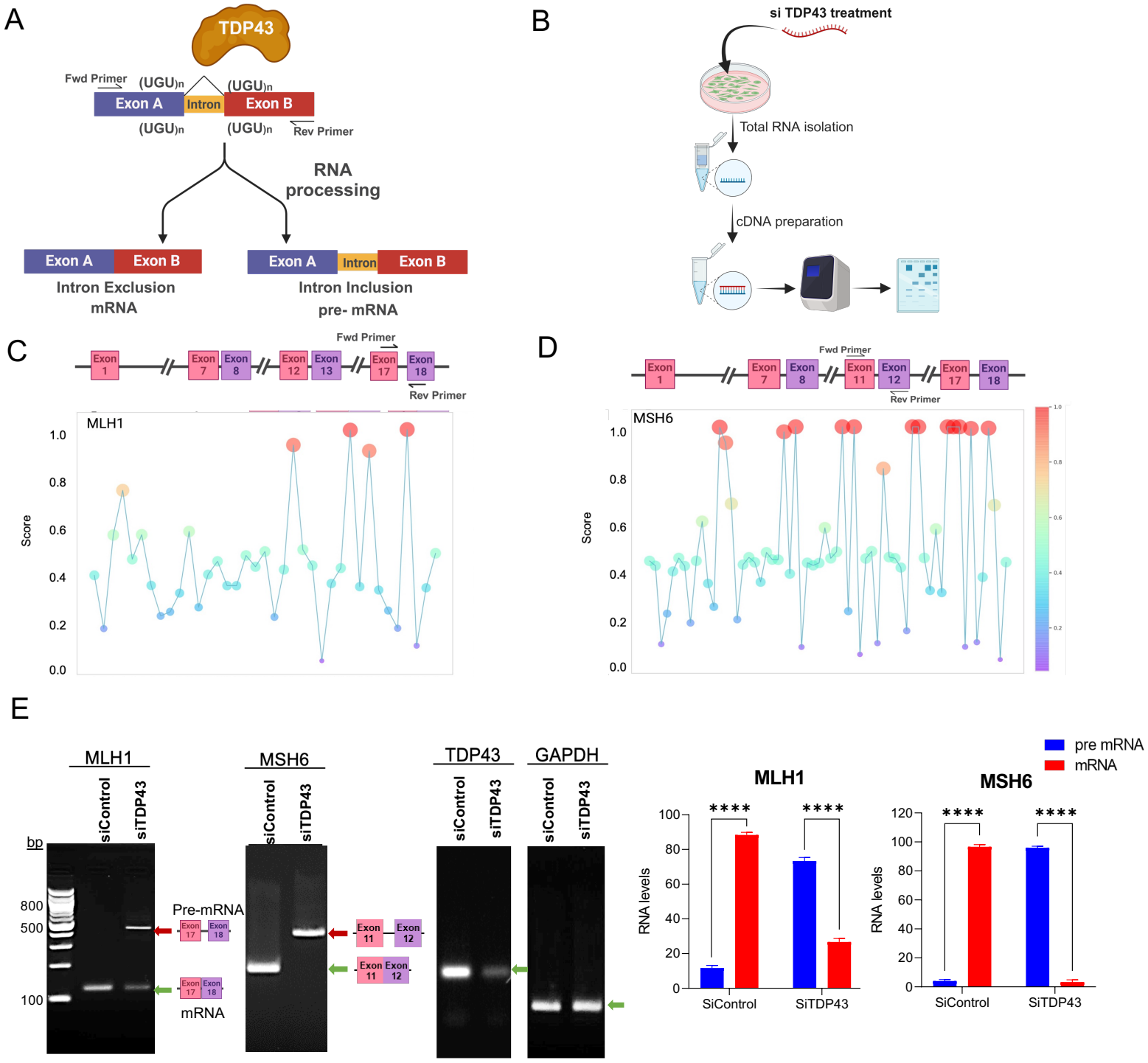

Fig. 3. TDP43 depletion alters transcriptional processing of specific MMR factors.
(**A-B**) A conceptual representation illustrating TDP43's anticipated role in pre-mRNA processing and a schematic of experimental design/ workflow employed to characterize TDP43-induced splicing of MMR transcripts by quantitating relative pre-mRNA versus mRNA transcript levels. (**C-D**) *In silico* computational analysis (RBPSuite) MLH1 and MSH6 sequences identified multiple putative TDP43 binding sites across each transcript. These identified sequences guided the primer design, to amplify exon-intron-exon junctions for post-cDNA synthesis amplification. (**E**) PCR amplification of selected regions of the MLH1 amplifying the region between Exon 17—Exon 18, and MSH6 amplifying the region between Exon11—Exon12 transcripts derived from cDNA prepared from HEK293 cultures treated with either siControl or siTDP43, histograms showing fold change of band intensity representing pre mRNA and fully spliced mRNA. Significance values (p-values) are as follows: p >0.5 (ns), p <0.0001 (****). Error bars indicate mean ± SEM from three independent experiments.

Figure 4

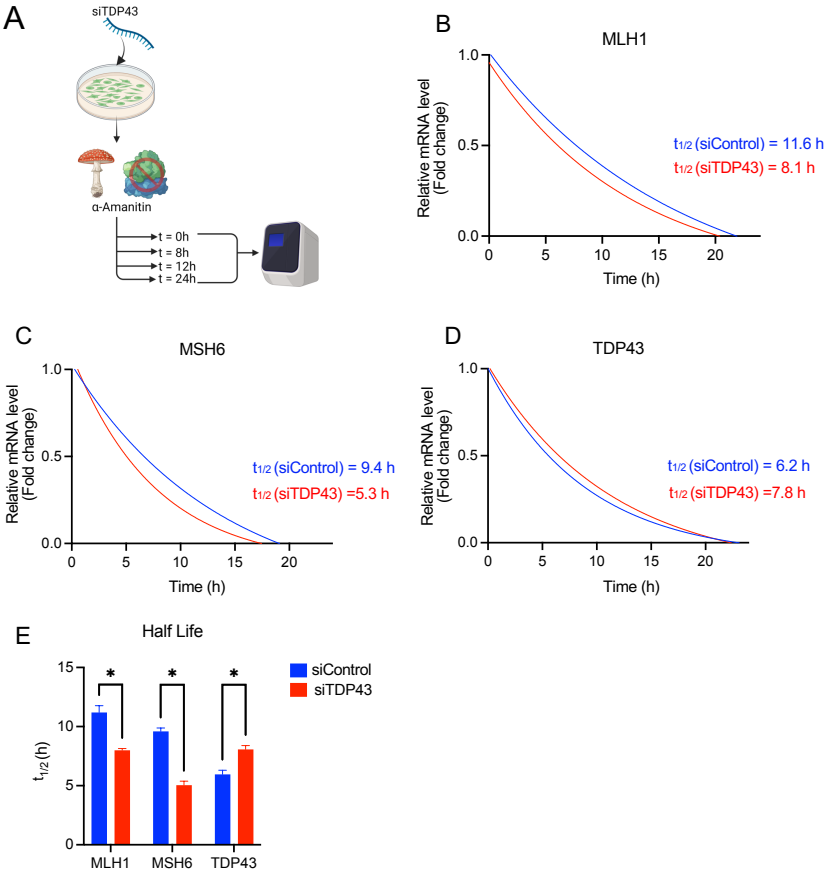

Fig. 4. TDP43 Regulates the Stability of MLH1 and MSH6 Transcripts *in vitro*.
(**A**) A schematic showing the experimental procedure for measuring the transcript half-life using nondividing neuronal cultures and the RNA polymerase II inhibitor α-amanitin. (**B-D**) Representative one-phase decay plots derived from mRNA transcript fold change expression between control and TDP43 siRNA treated cell cultures (**E**) Computed half-life ($t_{1/2}$) values are provided for each experimental group. Significance values (p-values) are as follows: $p > 0.5$ (ns), $p < 0.5$ (*). Error bars indicate mean ± SEM from three independent experiments.

Figure 5

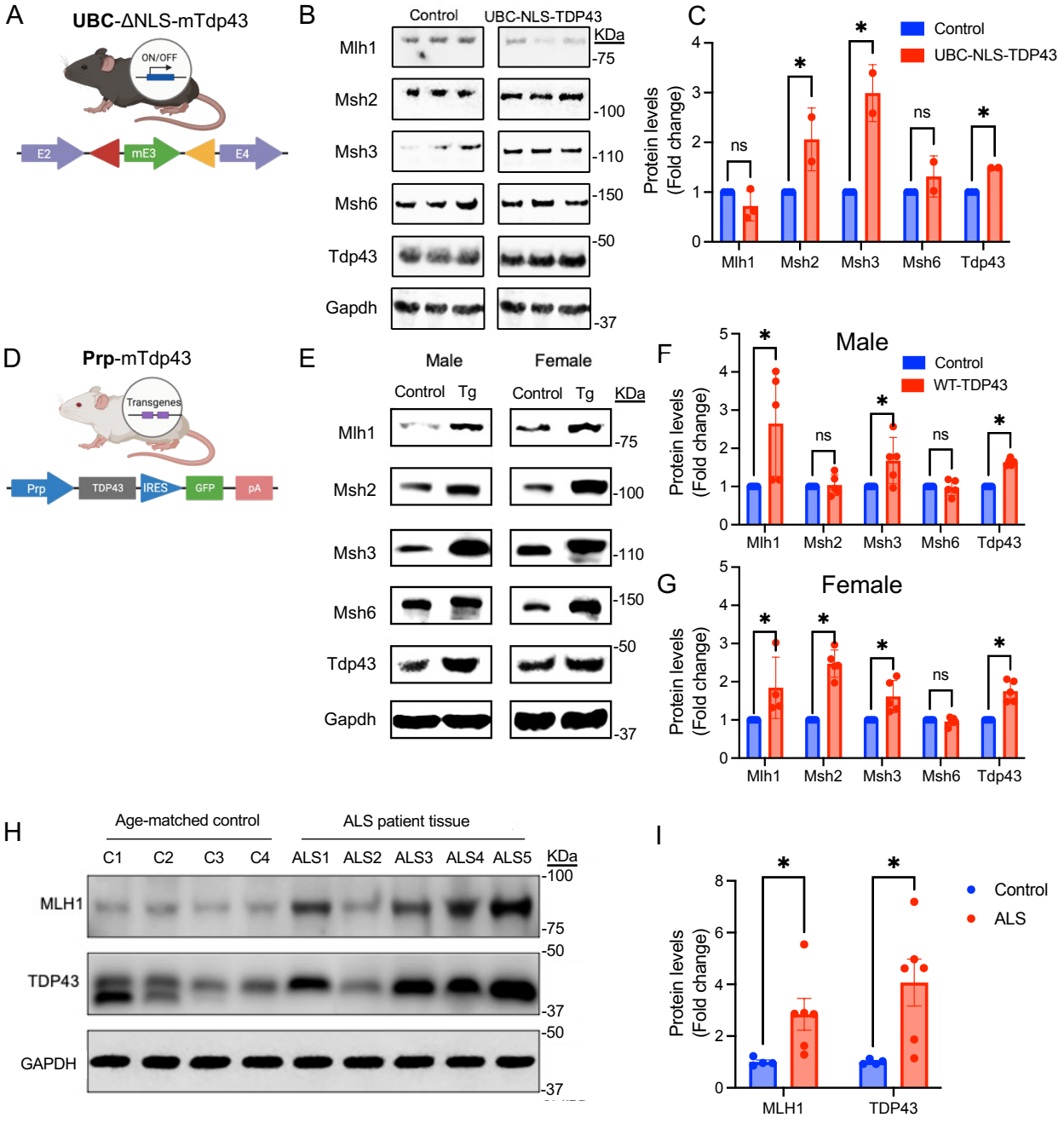
A
UBC-ΔNLS-mTdp43
ON/OFF
E2
mE3
E4
B
Control
UBC-NLS-TDP43
KDa
Mlh1
Msh2
Msh3
Msh6
Tdp43
Gapdh
-75
-100
-110
-150
-50
-37
C
Control
UBC-NLS-TDP43
Protein levels (Fold change)
ns
*
Mlh1
Msh2
Msh3
Msh6
Tdp43
D
Prp-mTdp43
Transgenes
Prp
TDP43
IRES
GFP
pA
E
Male
Female
Control
Tg
F
Male
WT-TDP43
G
Female
H
Age-matched control
ALS patient tissue
C1
C2
C3
C4
ALS1
ALS2
ALS3
ALS4
ALS5
MLH1
TDP43
GAPDH
I
ALS

Fig. 5. Mouse Models of TDP43-ALS Display Altered MMR Expression Phenotype.
(**A**) Schematic of the construct used to generate Tdp43-ΔNLS expression under the UBC promoter in C57BL6 mice. (**B**) WB images of cortical brain tissue lysate from three control and three transgenic (Tg) mice are shown. (**C**) Histograms showing fold change in immunoblot band intensity of averaged Tg samples (red) relative to control (blue) indicate significant increases in Msh2, Msh3, and Msh6 levels in Tdp43-ΔNLS mice. (**D**) Schematic of the construct used to generate moderate overexpression WT murine Tdp43 (mTDP43) under the PrP promoter in FVB/NJ mice. (**E**) Representative WB images of cortical brain tissue lysate from control and Tg mice are shown for male and female mice. (**F-G**) Histograms of fold change of WB band intensity of averaged Tg samples, relative to the control indicates significant increases in expressions of Mlh1, Msh2, and Msh3. (**H-I**) WB and quantitative histogram of insoluble protein isolates from control and Guam ALS affected human post-mortem CNS tissue. Significance values (p-values) are as follows: $p > 0.5$ (ns), $p < 0.5$ (*). Error bars indicate mean ± SEM from at least three independent experiments.

Figure 6

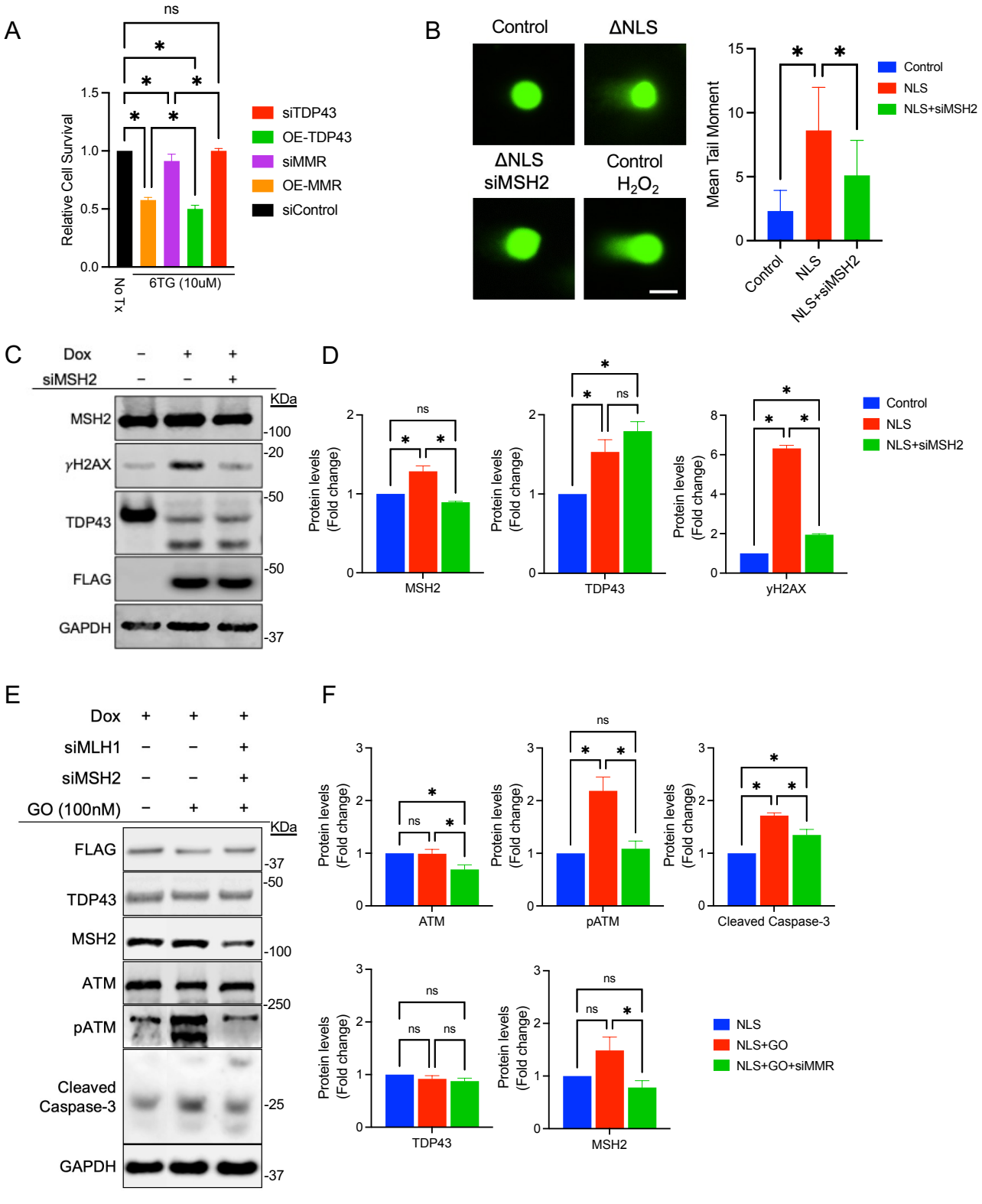
A
ns
*
Relative Cell Survival
1.5
1.0
0.5
0.0
No Tx
6TG (10uM)
siTDP43
OE-TDP43
siMMR
OE-MMR
siControl
B
Control
ΔNLS
ΔNLS siMSH2
Control H2O2
Mean Tail Moment
15
10
5
0
Control
NLS
NLS+siMSH2
C
Dox
siMSH2
KDa
MSH2
γH2AX
TDP43
FLAG
GAPDH
-100
-20
-50
-50
-37
D
Protein levels (Fold change)
MSH2
TDP43
yH2AX
Control
NLS
NLS+siMSH2
E
Dox
siMLH1
siMSH2
GO (100nM)
KDa
FLAG
TDP43
MSH2
ATM
pATM
Cleaved Caspase-3
GAPDH
-37
-50
-100
-250
-25
-37
F
ATM
pATM
Cleaved Caspase-3
TDP43
MSH2
NLS
NLS+GO
NLS+GO+siMMR

Fig. 6. MMR Expression Modulates TDP43 Proteinopathy-induced DNA Damage.
(**A**) MTT assay showing cellular viability of HEK293 cells treated with 10 mM 6-thioguanine (6-TG) relative to non-treated controls. Experimental groups included scrambled siRNA (siControl), siRNA against TDP43 (siTDP43), and MLH1 and MSH2 (siMMR), overexpression of wildtype TDP43 (OE-TDP43), and MLH1 and MSH2 (OE-MMR). (**B**) Alkaline single-cell gel electrophoresis (comet assay) in TDP43-ΔNLS (NLS) shows increased levels of DNA damage relative to controls, which was partially rescued by siMSH2. Histogram shows the quantitation of the mean comet tail moment across each condition. (**C-D**) WB analysis with quantitation histograms of NLS-expressing cells shows an increased level of DNA damage marker γH2AX that decreases following siMSH2 treatment. (**E-F**) WB analysis with quantitation histograms of NLS-expressing cells 6 h after a 45 min exposure to glucose oxidase (GOx). GOx treatment significantly increased the expression of phosphorylated ATM (pATM) and cleaved-caspase 3 which were partially rescued by siMSH2. Significance values (p-values) are as follows: $p > 0.5$ (ns), $p < 0.5$ (*). Error bars indicate mean $\pm$ SEM from three independent experiments.

Figure 7

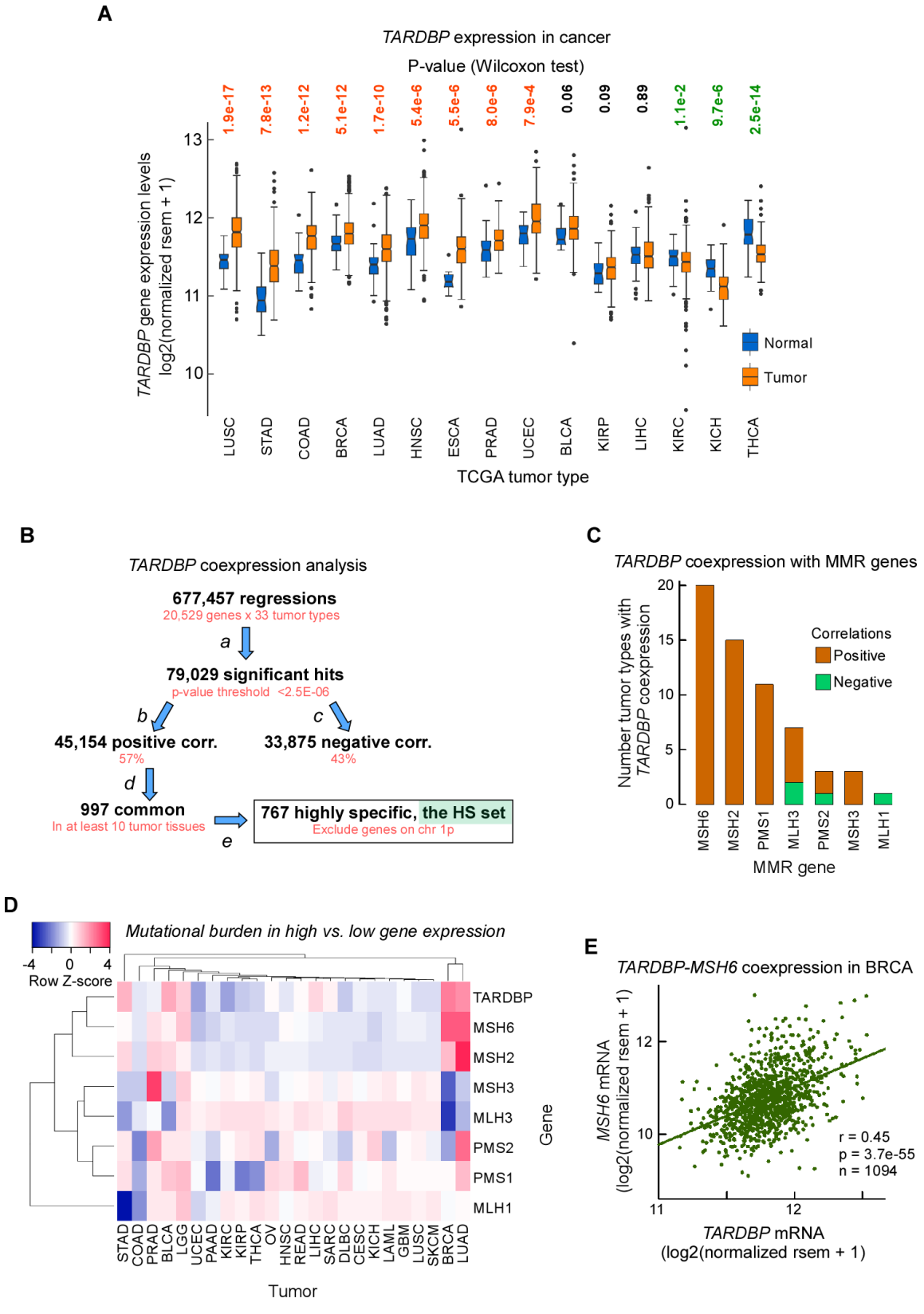
A
TARDBP expression in cancer
P-value (Wilcoxon test)
1.9e-17
7.8e-13
1.2e-12
5.1e-12
1.7e-10
5.4e-6
5.5e-6
8.0e-6
7.9e-4
0.06
0.09
0.89
1.1e-2
9.7e-6
2.5e-14
TARDBP gene expression levels
log2(normalized rsem + 1)
Normal
Tumor
LUSC
STAD
COAD
BRCA
LUAD
HNSC
ESCA
PRAD
UCEC
BLCA
KIRP
LIHC
KIRC
KICH
THCA
TCGA tumor type
B
TARDBP coexpression analysis
677,457 regressions
20,529 genes x 33 tumor types
a
79,029 significant hits
p-value threshold <2.5E-06
b
c
45,154 positive corr.
57%
33,875 negative corr.
43%
d
997 common
In at least 10 tumor tissues
e
767 highly specific, the HS set
Exclude genes on chr 1p
C
TARDBP coexpression with MMR genes
Number tumor types with TARDBP coexpression
Correlations
Positive
Negative
MSH6
MSH2
PMS1
MLH3
PMS2
MSH3
MLH1
MMR gene
D
Mutational burden in high vs. low gene expression
Row Z-score
TARDBP
MSH6
MSH2
MSH3
MLH3
PMS2
PMS1
MLH1
Gene
STAD
COAD
PRAD
BLCA
LGG
UCEC
PAAD
KIRC
KIRP
THCA
OV
HNSC
READ
LIHC
SARC
DLBC
CESC
KICH
LAML
GBM
LUSC
SKCM
BRCA
LUAD
Tumor
E
TARDBP-MSH6 coexpression in BRCA
MSH6 mRNA
(log2(normalized rsem + 1)
r = 0.45
p = 3.7e-55
n = 1094
TARDBP mRNA
(log2(normalized rsem + 1)

Fig. 7. TARDB is Overexpressed in Multiple Tumor Types and Coexpresses with MSH6/2.

(**A**) Box plots of *TARDBP* gene expression in 15 TCGA tumor types and matched controls. mRNA levels were normalized rsem data from RNA-seq analyses. Only tumor types with at least 10 control samples were included. Box plots display the interquartile range (IQR) from Q1 to Q3 (25–75% percentiles), median (center line), whiskers extending from Q1–1.5*IQR to Q3 + 1.5*IQR and outliers (dots). (**B**) Workflow to assess genes coexpressed with *TARDBP* in cancer. *Top*, initial screen of linear regressions of rsem normalized RNA-seq data between *TARDBP* and each of the mapped transcripts in 33 TCGA tumor types. a) filter correlations by regression coefficients with a p-value <2.5E-6, which represents an approximate Bonferroni correction for multiple comparisons (p-value/N). b and c) separate positive correlations from negative correlations. d) select genes whose correlation with *TARDBP* is observed in at least 10 different tumor types. e) filter out genes residing on the same chromosomal arm as *TARDBP* since coexpression may arise from chromosomal amplification. (**C**) Bar plot of the number of TCGA tumor types for which a significant correlation for the coexpression between *TARDBP* and MMR genes was found, from panel B (b,c). (**D**) Heat map showing mutation burden (single base substitutions) in high versus low expression of listed genes according to tumor type. (**E**) Dot plot and linear regression for the MMR gene (*MSH6*) most strongly coexpresed with *TARDBP* and tumor type (BRCA), from panel B (e).

Figure 8

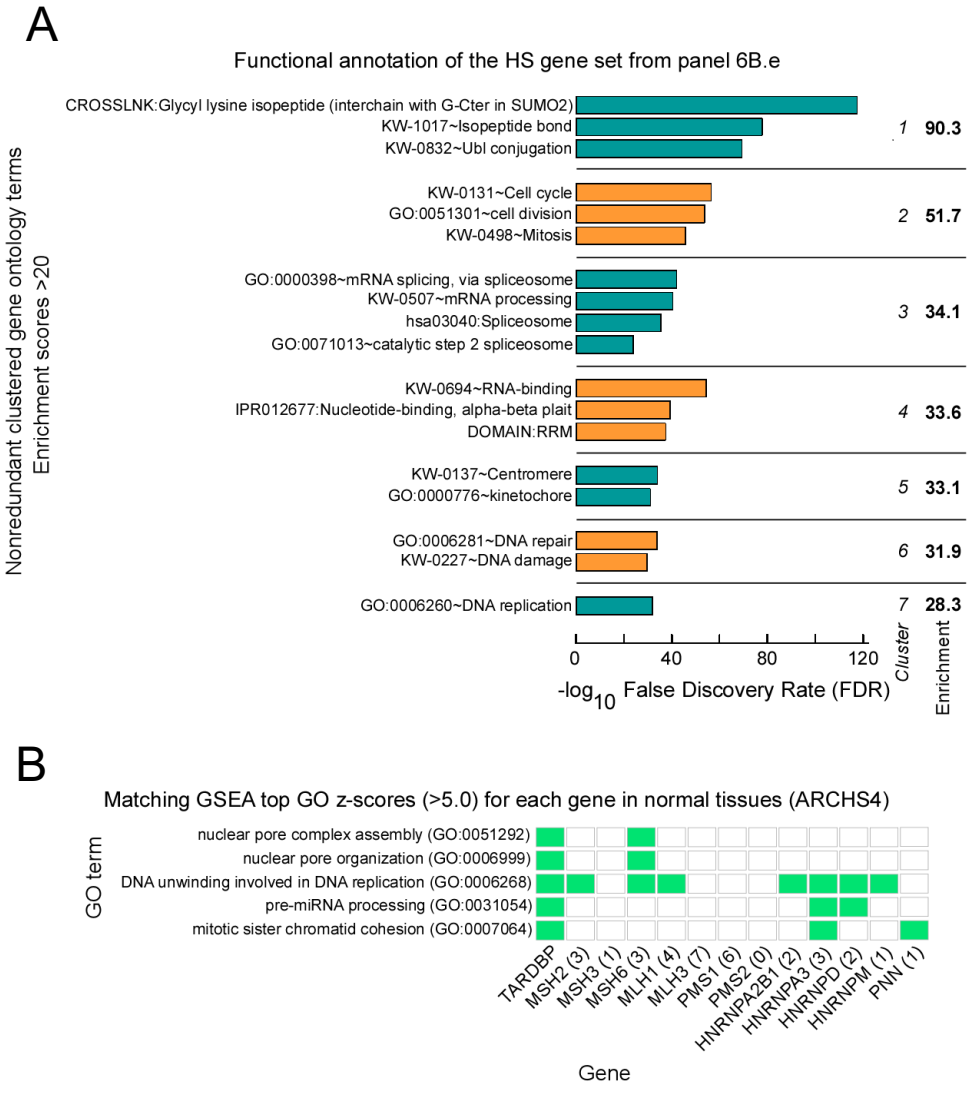

Fig. 8. Genes Coexpressed with *TARDBP* are at the Core of Cell Cycle and DNA Repair.

(**A**) Gene Set Enrichment Analysis (GSEA) of genes highly coexpressed with *TARDBP*. Only enrichment scores >20 and the ontology terms with the strongest enrichment score for redundant ontology terms were displayed. (**B**) Check box for predicted biological processes (GO) with z-scores >5 for *TARDBP*, and *TARDBP*-matching GO terms with z-scores >5 for MMR genes and HNRNP genes highly coexpressed with *TARDBP* at levels above that of the *PNN* gene (Pearson's coefficient >0.63 in human normal tissues). In parenthesis the number of GO terms with z-scores >5.