scientific reports

OPEN

Characterization of the human TARDBP gene promoter

Marco Baralle[1] & Maurizio Romano[2]✉

The expression of TDP-43, the main component of neuronal intracellular inclusions across a broad spectrum of ALS and FTD disorders, is developmentally regulated and studies in vivo have shown that TDP-43 overexpression can be toxic, even before observation of pathological aggregates. Starting from these observations, the regulation of its expression at transcriptional level might represent a further key element for the pathogenesis of neurodegenerative diseases. Therefore, we have characterized the human TARDBP promoter, in order to study the transcriptional mechanisms of expression. Mapping of cis-acting elements by luciferase assays in different cell outlined that the activity of the promoter seems to be higher in SH-SY5Y, Neuro2A, and HeLa than in HEK293. In addition, we tested effects of two SNPs found in the promoter region of ALS patients and observed no significant effect on transcription levels in all tested cell lines. Lastly, while TDP-43 overexpression did not affect significantly the activity of its promoter (suggesting that TDP-43 does not influence its own transcription), the presence of the 5′UTR sequence and of intron-1 splicing seem to impact positively on TDP-43 expression without affecting transcript stability. In conclusion, we have identified the region spanning nucleotides 451–230 upstream from the transcription start site as the minimal region with a significant transcription activity. These results lay an important foundation for exploring the regulation of the TARDBP gene transcription by exogenous and endogenous stimuli and the implication of transcriptional mechanisms in the pathogenesis of TDP-43 proteinopathies.

TAR DNA-Binding Protein 43 (TDP-43) is an ubiquitously expressed and highly conserved nuclear protein encoded by the human TARDBP gene, and it is involved in several cellular mechanisms, including transcription, pre-mRNA processing, splicing and translation[1,2]. After initial implication in the pathogenesis of monosymptomatic forms of cystic fibrosis[3–7], it has been discovered a direct association between TDP-43 abnormalities and neurodegenerative diseases, and in particular with Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Degeneration (FTD)[8,9] was discovered. Over the last years, TDP-43 immunoreactive inclusions have also been reported in 70% of patients with hippocampal sclerosis, 30% of patients with Alzheimer's disease, 33% of patients with Pick's disease and in a subset of patients with Lewy-body related disease[10–17]. Whereas many efforts have been expended in characterizing TDP-43 pathophysiological functions, a contentious debate ensued and is still ongoing over the possible mechanisms whereby TDP-43 triggers neurodegeneration. It has been proposed both that cytoplasmic aggregates might be neurotoxic ("gain of function" hypothesis) and that the sequestration of the factors in aggregates might induce a functional deficiency resulting in alteration of the TDP-43-regulated processes ("loss of function" hypothesis)[18–22]. Of course, the two hypotheses are not necessarily mutually exclusive, and studies with animal models have shown that TDP-43 overexpression can be neurotoxic, even without presence of inclusions in worms[23], flies[24–26], mice[27–33], and monkeys[34]. Importantly, mice expressing the mutant A315T TDP-43 mice show a dose-dependent degeneration of cortical and spinal motor neurons, without evidence of cytoplasmic TDP-43 inclusions[32]. These studies suggest that even an increase in TDP-43 can contribute to neurodegeneration without occurrence of aggregation. In addition, elevated levels of TDP-43 has been found in the cerebrospinal fluid[35–37] as well as in plasma of ALS[38–40], FTD and Alzheimer[41,42] patients, and in FTD patients carrying C9orf72 expansion or GRN mutations, plasma phosphorylated TDP-43 levels have been found to be higher than in healthy controls[43]. On the other hand, it is now well established TDP-43 levels are tightly controlled: in fact, it has been demonstrated that TDP-43 can regulate its own expression through a negative feedback loop[44–51]. Altogether these observations suggest that regulation of TDP-43 expression at transcriptional level might also be implicated in the pathogenesis of neurodegenerative diseases.

The study of the TARDBP gene promoter is therefore useful to shed further light on the pathogenesis of TDP-43 proteinopathies. Interestingly, two SNPs were found within the promoter sequence of TARDBP gene in ALS patients[52]. However, no functional studies have been so far carried out.

[1]International Centre for Genetic Engineering and Biotechnology (ICGEB), Area Science Park, Padriciano, Trieste, Italy. [2]Department of Life Sciences, University of Trieste, Via A. Valerio 28, 34127 Trieste, Italy. ✉email: mromano@ units.it

A

```
-1316  aaggaaggtgggaagggagggaaggggggggaaagggagggaagggaggaa   50
-1266  aggaagggaaggaaaggaaagaaaggaaagaaaagatagaaaagaaaggt   100
-1216  ggctcgtgcctgtaatcccagcactttggggagctgaggcaggcggatca   150
-1166  cttgaggccatgagttcgagatcagcctggccaacatggccaaaccctgt   200
-1116  ctctactaaaaatacaaaaattagccgggcatggtgtctttgcacgccta   250
-1066  taatcccagctatcccttgaacctaggagttggaagtggcagtgagctgc   300
-1016  agattgcgccactgcactccagcctgggggacaagagcgagactccgtct   350
 -966  caaaaaaaaaaaaaaaaaaagtgaaacaaaaagtgtgccttaaaatagat   400
 -916  gaagtactgtattattattattattaaagggctggagccagaacaagtta   450
 -866  gcattttgtgcctcgagattcaaatcccaggtctaccatttattccttgg   500
 -816  gtgacatttttttggggggagggggctggaggacagggtctcactctgtcg   550
 -766  cctagactggagtgaagtggcgcgatctcggctcaccgcaacctccgcct   600
 -716  accgcgttcaagcaattctcctgcttcaggctcccgagtagatgggacta   650
 -666  caggcgcgcgccactaccgcccggctaatccttgcgtcactttaagcaag   700
 -616  acacttgatctctctgtgcctcagtttcctcacccgcagaatgaatatca   750
 -566  atagctcctgctacacaaaactgtttgggctattaaatatgtgaatactc   800
 -516  agaaagtacctggcacaaagtagccttatcattgctattatttagaacaa   850
 -466  agttattgtttccaggtttggaggtgggctggcggagggaggagaagacg   900
 -416  cactaggttgcctttcagaaaaggggtccaggccttatcctgcctcccag   950
 -366  gtggactcggcccgggatcccgagccccgcccctccaggccccgcccg   1000
 -316  ccccgccctcgctggccgagactctgtgctactctcgcgcgctttgaaag   1050
 -266  tgcgcgagctccgaccgagacctgggccggcggttggtcggccgaaacca   1100
 -216  tctgctacgttacgtggggggaggtcagctcctatacagccacactttt   1150
 -166  tcaatcttcagcttttcaggcctaggtcctggttttattttagctttcag   1200
 -116  agcttctctgtaggcttgaatttccaggaggcagcccgagtccctgggga   1250
  -66  gagggtcatcgtggctctcgtcggtgccttccacctctggacaggcaacc   1300
  -16  ggtgggagaggacgcc                                    1316
   +1  GGTGGGCGGGGGGAGGAGGCGGCCCTAGCGCCATTTTGTGGGAGCGAAGC   1366
  +50  GGTGGCTGGGCTGCGCTTGGGTCCGTCGCTGCTTCGGTGTCCCTGTCGGG   1466
 +100  CTTCCCAGCAGCGGCCTAGCGG                              1588
```

B
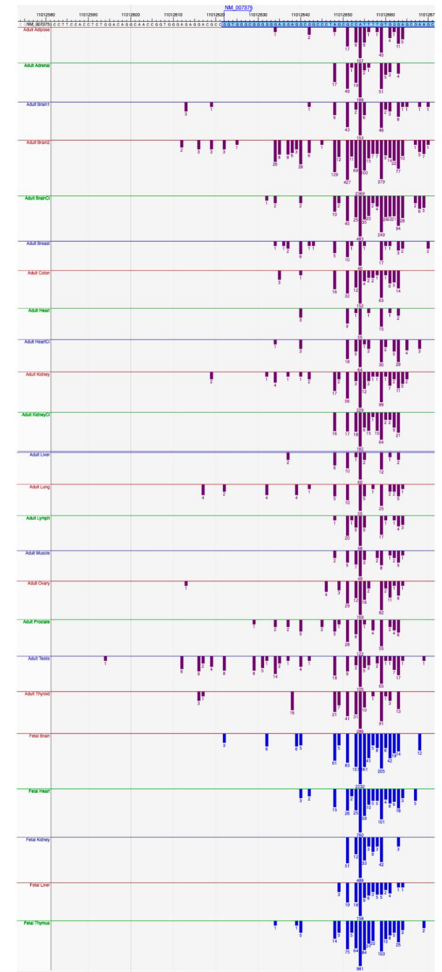
Figure 1. Sequence of the human TARDBP promoter and map of the TARDBP TSSs in adult and fetal tissues. (**A**) Proximal sequences (− 1316/ + 122) of the human TARDBP promoter are shown. Transcript reference: NM_007375.4. The promoter sequences are numbered (left side) relative to the transcription start site (TSS, + 1). (**B**) Graphical overview of the multiple TARDBP TSSs mapped in adult and fetal tissues presented in the DataBase of Transcriptional Start Sites (DBTSS). The top turquoise (5′-region) box correspond to the RefSeq NM_007375. Red boxes represent DBTSS clones of Adult tissues, while blue boxes represent DBTSS clones of Fetal tissues.

In this work, we have characterized the cis-acting elements important for TARDBP gene regulation and evaluated its cell-type specificity. Subsequently, we have characterized the possible impact of some cis-acting elements (SNPs and 5′UTR), trans-acting factors (TDP-43 itself) and events (splicing) on the transcriptional activity of TARDBP promoter.

Results

Bioinformatic analysis of the *TARDBP* promoter.
The characterization of *TARDBP* gene sequence (Ensembl gene ID: ENSG00000120948) was initially performed by a bioinformatic analysis using the UCSC Genome Browser (GRCh37/hg19). The *TARDBP* gene, mapped to chromosome 1p36.22 region, spans 12870 bp from position 11,072,679 to 11,085,548 on the forward DNA strand. Figure 1 shows the TARDBP promoter sequence (starting from the 1316 nucleotide upstream from the TSS of NM_007375.3 TDP-43 transcript). Inspection of UCSC TARDBP transcripts matched with those retrieved from the Database of Transcriptional Start Site (DBTSS, http://dbtss.hgc.jp [53,54]), most of transcription start sites (TSS) mapped at the corresponding positions indicated in the Reference Sequence NM_007375 (Fig. 1A,B). Apparently, there are no tissue-specific TSSs, since only one mRNA species seems to be predominant in Adult and Fetal Tissues, as well as in different tissues (Fig. 1B).

Subsequently, by querying the Eukaryotic Promoter Database[55], different putative TATA-box motifs ([p-value = 0.01]: − 930, − 865, − 606, − 571, − 513, − 309, upstream of TSS shown in Fig. 1A), CG-box sequences (CCAAT-box [p-value = 0.01]: − 744, -677, − 572, − 61, upstream of TSS shown in Fig. 1A) were identified, but deeper inspection revealed that the retrieved sequences were not consistent either with the canonical human
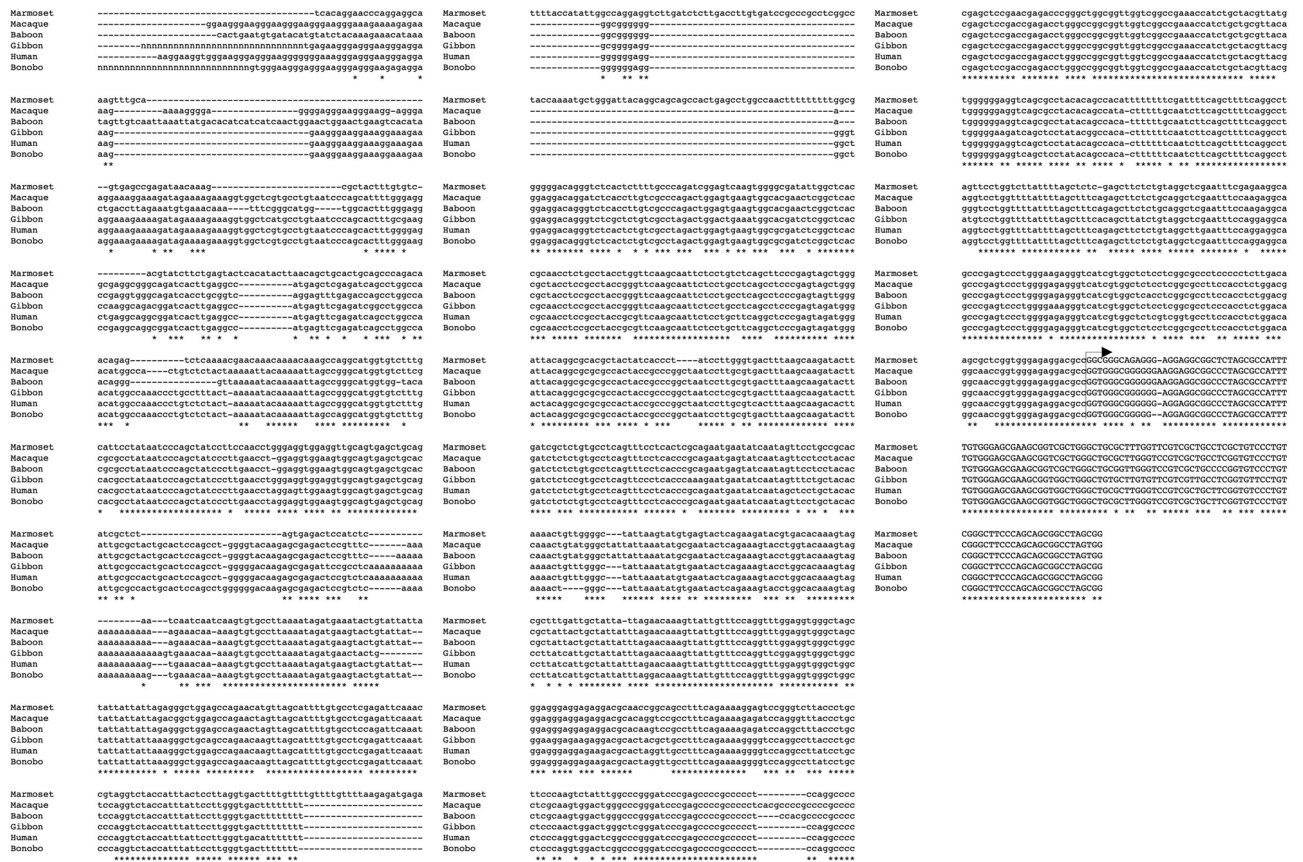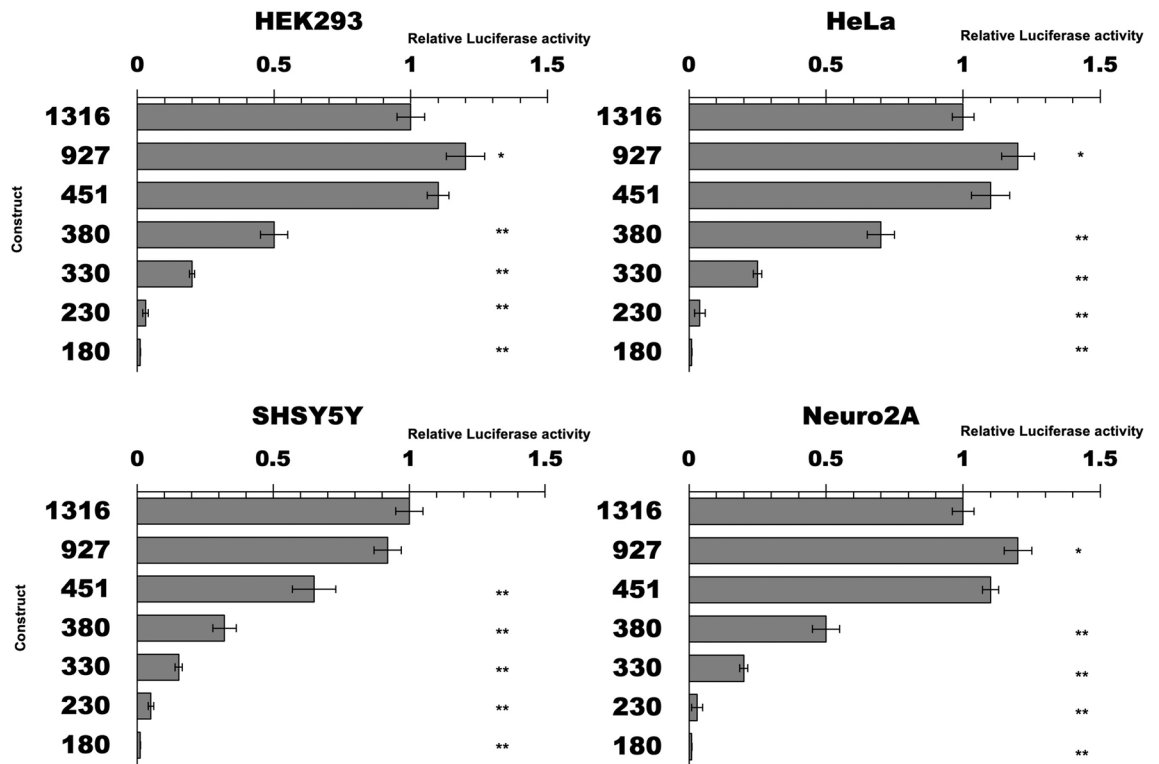
Figure 2. Genomic alignment of putative promoter regions from Homo sapiens, Pan troglodytes and Callithrix jacchus TARDBP genes. The alignment of the 1316nt—sequence upstream and 78 bp downstream of the transcription starting site of Human (Homo sapiens, ENSG00000120948, NM_007375.3) was carried out versus: Marmoset (Callithrix jacchus, ENSCJAG00000002381), Macaque (Macaca mulatta, ENSMMUG00000007456), Olive baboon (Papio anubis, ENSPANG00000017459), Gibbon (Nomascus leucogenys, ENSNLEG00000009512) and Bonobo (Pan paniscus, ENSPPAG00000039086) TARDBP transcripts, by using the MUSCLE alignment program (http://www.ebi.ac.uk/Tools/muscle/index.html).

promoter motifs or with the main mapped TSS (DBTSS, http://dbtss.hgc.jp, Fig. 1). In addition, a further analysis performed with another promoter search tool (GPMiner[56]) failed to detect the main core promoter elements. These observations suggest that the expression of the human TARDBP gene is driven by a TATA-less promoter.

Conservation of TARDBP promoter sequence throughout evolution. In order to study the evolutionary conservation of TARDBP promoter, the 1316 nucleotides upstream of the TSS of TARDBP transcripts from human and other species were retrieved using GenBank and Ensembl Genome Browser (http://www.ensembl.org/). Alignments were performed using MUSCLE software[57]. Among primates, we compared the genomic sequences of the putative TARDBP promoter region from Hominoidea (*Homo sapiens*—Human, *Pan paniscus*—Bonobo, and *Nomascus leucogenys*—Gibbon), from of New World monkey (*Callithrix jacchus*—Common marmoset) and from Old World monkey (*Macaca mulatta*—Rhesus macaque; *Papio anubis*—Olive baboon). In general, this comparison highlighted that the relative degree of similarity among all three species increases significantly with respect to the proximity of the TSS. Then, in the region spanning from nucleotide -1316 to -1000 upstream from TSS, the hominoid sequences seem to be closer to each other than with that of New and Old World monkeys. On the other hand, in the region spanning the 1000 nucleotides upstream from TSS, the sequences from all primates show a high degree of identity (Fig. 2). On the other hand, the alignment of the human, mouse and rat putative TARDBP promoter sequences shows a low degree of global similarity (Fig. 3A). Nevertheless, the region of the rodent gene promoters spanning approximately 500nt upstream from the TSS shows a higher level of similarity, suggesting that proximal region of the putative TARDBP promoter region might encompass regulatory elements conserved across primates and rodents. In fact, pairwise sequence alignment between the 1316nt-sequence upstream of TSS of the human/mouse and human/rat TARDBP orthologous genes outlined a 62%/63% of local similarity/identity (with 8% and 5% gaps, respectively) between the sequences of the two species, in 300/500nt-region upstream of the respective TSSs (Fig. 3B). In consideration of the observations pinpointing how age regulated variations in TDP-43 expression seem to be evolutionary conserved at least in mouse and fly[58,59], we tried to perform pairwise sequence alignment between the 1316nt-sequence upstream of TSS of the mouse and fruit fly TARDBP ortholog genes. The local alignment of two the

Figure 3. Global and local alignment of putative promoter regions from Homo sapiens, Mus musculus, Rattus norvegicus and Drosophila melanogaster TARDBP ortholog genes. (**A**) The alignment of the 1316nt—sequence upstream of the transcription starting site of Human (RefSeq: NM_007375.3), Mouse (ENSMUST00000084125) and Rat (ENSRNOT00000049822) TARDBP transcripts was generated with MUSCLE alignment software (http://www.ebi.ac.uk/Tools/muscle/index.html). (**B**) Pairwise Sequence Alignment was used to identify regions of identity/similarity between the promoter sequences of the Human/Mouse and Human/Rat TARDBP orthologs. The local similarities between the sequences were identified by using the EMBOSS Matcher Local alignment tool (https://www.ebi.ac.uk/Tools/psa/).

sequences returned a low % of global identity/similarity (% Identity/Similarity: 41.4%, % Gaps: 46.8%), without outlining regions evolutionary conserved between these two species (data not shown).

Characterization of cis-regulatory sequences of TARDBP promoter.   In order to identify the minimal functional sequence responsible for the transcriptional activity of the promoter of the human TARDBP gene, a luciferase assay was set up with deletion-fragments of the putative TARDBP promoter constructs cloned into the pGL4 luciferase reporter vector. Different segments of the putative regulatory region of the promoter were amplified through a PCR reaction, and eight constructs differing in length were produced: the sequences encompassed between 27 nucleotides downstream from the TSS (+ 1nt, RefSeq NM_007375.3) to 1316, 927, 451, 380, 320, 280, 230 and 180 nucleotides upstream from the TSS (Fig. 4).

The promoter fragments were subcloned into the pGL4.11 vector, transfected in cell lines of different tissue origin (HEK293, HeLa, Neuro2A and SH-SY5Y) and the luciferase activity of each construct was normalized versus the 1316 construct (i.e., 1316 = 1).

Intracellular analysis showed that most of the promoter activity was retained in the fragments spanning from 1316 to 451 nucleotides upstream from the TSS within all tested cell lines (Fig. 4). Further deletions resulted in dramatic reduction of activity (380 and 330 constructs), with almost complete loss of activity when the promoter was shortened to 230 and 180 nucleotides upstream from the TSS, in all tested cell lines. These results show that the minimal promoter region encompasses the 451 nucleotides upstream from the TSS of RefSeq NM_007375.

We also tried to detect any kind of tissue-specific promoter activity. To this aim, we compared the activity of the 1316, 927 and 451 promoter fragments among SH-SY5Y, Neuro2A, HeLa, and HEK293 cell lines, considering the HEK293 cells as the unitary reference. Indeed, the constructs showed approximately 6–8 × higher activity
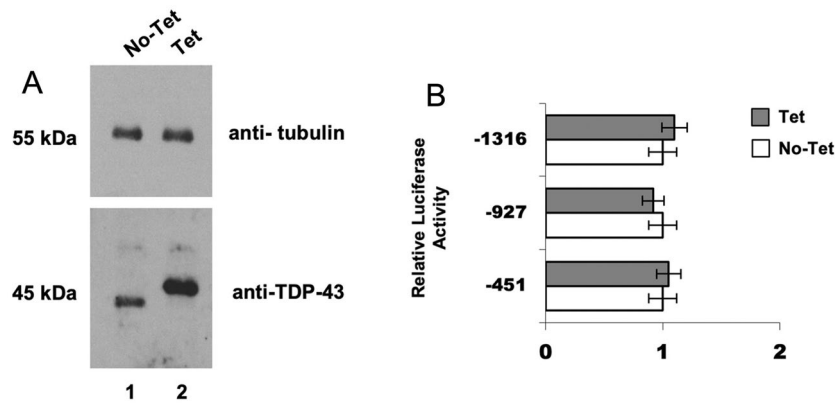
Figure 4. Deletion analysis of the human TARDBP promoter. Firefly luciferase activity of TARDBP promoter deletants in different cell lines (HEK293, HeLa, SH-SY5Y and Neuro2A) was measured in the cell lysate and values were normalized against Renilla. Activity of deletants is expressed as-fold against the cells transfected with the 1316 construct ($=1$). Bars indicate the mean value of three independent assays. Error bars indicate standard deviation: $*p < 0.05$, and $**p < 0.01$ (One way ANOVA with Tukey test).
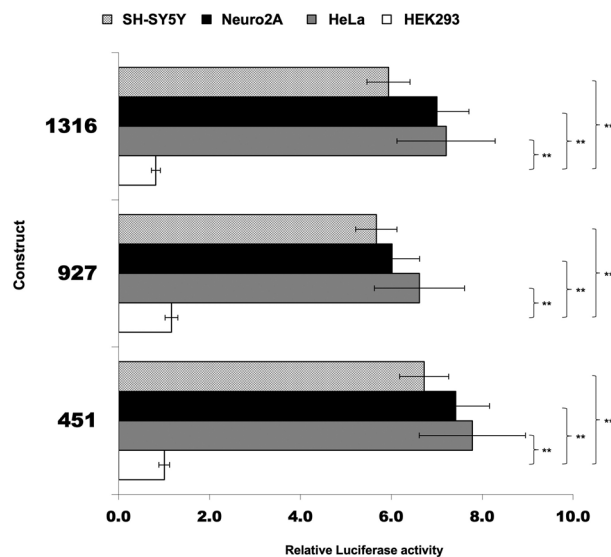
Figure 5. Comparison of the human TARDBP promoter activity in different cell lines. Firefly luciferase activity of 1316, 927 and 451 constructs was tested (normalized against Renilla) in different cell lines (HEK293, HeLa, SH-SY5Y and Neuro2A). Activity is expressed as-fold against transfections in HEK293 cells ($=1$). Data are mean ($\pm$SD) of three independent assays: $*p < 0.05$, and $**p < 0.01$ (One way ANOVA with Tukey test).

Figure 6. Effects of ALS-related SNPs on TARDBP promoter activity. (A) Schematic representation of the human TARDBP promoter variants. The diagram depicts the wild type and the mutated promoter sequences used for the luciferase assay. The mutated nucleotides (and their positions) are indicated. (B) Luciferase activity of TARDBP promoter variants in different cell lines (SH-SY5T, HEK293, HeLa and Neuro 2A). Luciferase activity (normalized against Renilla) is expressed as-fold against transfections of the haplotype [4439:T; 4901:T] (=1). Data are mean (±SD) of three independent assays/cell line. The mean values did not differ to a statistically significant extent (p ≥ 0.05, one way ANOVA with Tukey test).

both in two neuronal cell lines (SH-SY5Y, Neuro2A) and in HeLa cervical carcinoma cell line than in HEK293 embryonic kidney cells (Fig. 5).

Transcriptional effects of SNPs found within the TARDBP promoter of ALS patients. The analysis of pathogenic mutations found in the *TARDBP* gene sequence of 46 Australian patients of European descent affected by sporadic Amyotrophic Lateral Sclerosis (sALS) has revealed two promoter variants (c.1-562t>c and c.1-100t>c) with a different frequency in patients than in controls (115 neurologically normal people or Hap-Map European and Sub-Saharan African cohorts)[52]. The c.1-562t>c single nucleotide substitution (rs9430335; NG_008734.1:g.4439CT; NM_007375.3:c.-696C>T) was found in homozygosity (C/C) at higher frequency and in heterozygosity (T/C) at lower frequency in sALS patients, as compared to controls (0.2 vs 0.06 and 0.1 vs 0.3, respectively). On the other hand, the c.1-100t>c polymorphism (rs968545; NG_008734.1:g.4901T>C; NM_007375.3:c.-234T>C) was present only in heterozygosity (T/C) at higher frequency (0.2 vs 0.1) than in controls[52].

Therefore, we sought to test the impact of these SNPs on the transcription of TDP-43, in order to find a possible functional correlation with the manifestation of ALS. In order to analyze the SNP effects alone or in combination, we created three variants of our 927 original pGL4 construct (4439:T; 4901:T), so reproducing all the possible alleles (4439:T; 4901:C), (4439:C; 4901:C) and (4439:C; 4901:T) (Fig. 6A). The constructs were transfected in SH-SY5Y, HEK293, HeLa and Neuro2A cell lines. The luciferase activity of the three mutant constructs was compared to that of control (4439:T; 4901:T), used as unitary reference (Fig. 6B): no statistically significant differences were observed in the transcriptional activity of the three variants in all the tested cell lines (Fig. 6B).

TDP-43 does not influence its own transcription. After characterizing the sequence of the promoter of TARDBP gene and its activity in different cell lines, we wished to evaluate the ability of TDP -43 to influence, directly or indirectly, the synthesis of its own transcript. For transfections of these constructs, a stable HEK293-

Figure 7. Effects of TDP-43 overexpression on TARDBP promoter activity. (**A**) Immunoblot analysis of TDP-43 expression after 48 h-Tetracycline induction (Tet) of the HEK293-Flp-In-TDP-43 wild type stable cell line. Both endogenous (lane 1) and Flag-tagged (lane 2) proteins were visualized using anti-TDP-43 polyclonal antibody (ProteinTech). As expected, overexpression of Flag-tagged wild-type TDP-43 silenced the expression of the endogenous protein by triggering autoregulatory loop. Beta-Tubulin was used as loading control to normalize the levels of detected proteins. (**B**) Effects of TDP-43 overexpression on TARDBP promoter activity. The luciferase activity of three different constructs (1316, 927 and 451) transfected in the HEK293-Flp-In-TDP-43 wild type stable cell line was assayed without (No-Tet) or after (Tet) Tetracycline induction for 48 h. No statistically significant differences were found. Data are mean ($\pm$ SD) of three independent assays. Significance values refer to comparisons against control transfections (No-Tet) by using one way ANOVA with Tukey test.

flp-in cell line carrying a single copy of flag-tagged wild type human TDP-43 (HEK-TDP43wt) was used. This cell line was generated in our laboratory[44,60–62] and was used as it allows a more homogenous TDP-43 expression from a Tetracycline inducible expression vector.

The HEK-TDP43wt cells were transfected with the constructs 451, 927, and 1316 (along with pRL-TK plasmid). Subsequently, TDP-43 expression was driven by Tetracycline induction (48 h) and its levels were probed by Western blotting (Fig. 7A). Uncropped images for western blots shown in Fig. 7A are shown in Supplementary Fig. 1. As expected, Tetracycline induction of the Flag-tagged wild-type TDP-43 caused reduction of endogenous protein expression by triggering the known autoregulatory loop (Fig. 7A, lane 2). However, analysis of the luciferase activity 48 h later did not show significant differences in promoter activity upon TDP-43 tetracycline-induced overexpression (Fig. 7B). Similarly, no significant differences were observed after transient transfection of SH-SY5Y with the constructs 451, 927, and 1316 (with Renilla reporter) after TDP-43 transient overexpression (data not shown). These results suggest that TDP-43 does not influence its own transcription.

The TARDBP 5′UTR and intron 1 splicing positively impact the luciferase expression.

The 5′UTR region of TARDBP gene encompasses exon 1 (102 bp) and the first 12 bp of exon 2, separated by intron 1 (972 bp). In order to explore the presence of additional elements able to modulate TDP-43 expression, the functional impact of the 5′ UTR (construct 451 + Ex1Ex2) and intron 1 (construct 451 + Ex1-IVS1-Ex2) of *TARDBP* was analysed by generating variants of the 451 plasmid (Fig. 8A). A first variant was containing the TARDBP 5′UTR (exon 1, 102 bp and exon 2, 12 bp) correctly spliced (451 + Ex1Ex2). The second construct was created by inserting the region encompassing exon1 (102 bp), intron 1 (972 bp) and the first 12 bp of exon 2 of TARDBP gene in between the 451 promoter and the luciferase ATG codon (451 + Ex1-IVS1-Ex2 wt). The third construct was a mutant of latter construct where the 3′ splice site of intron 1 was disrupted (451 + Ex1-IVS1-Ex2 mut).

In SH-SY5Y cells, the presence of a correctly sized amplicon was observed only after transfection of the 451, 451 + Ex1Ex2 and 451 Ex1-IVS1-Ex2 wt constructs (Fig. 8B). Uncropped images for agarose gel shown in Fig. 8B are shown in Supplementary Fig. 2. Quantitation of the mRNA species was determined by Real Time PCR (Fig. 8C). By using the 451 construct as the reference (= 1), the relative expression of the 451 + Ex1Ex2 (containing the pre-arranged and correctly spliced 5′UTR region) was 2.6 × that of the reference, while the 451 + Ex1-IVS1-Ex2 wt construct showed a 25 × increment in relative expression compared to the -451 construct). Conversely, the mRNA levels of the 451 + Ex1-IVS1-Ex2 mut construct were negligible (0.1 × vs 451 control). In order to test if mutant transcript is unstable due to low efficiency of splicing, the RNA level of 451 + Ex1-IVS1-Ex2 wt and 451 + Ex1-IVS1-Ex2 mut constructs were monitored with and without addition of a transcriptional inhibitor (Actinomycin D) that causes complete transcriptional arrest[63]. Considering the higher mRNA levels derived from the intron-containing construct (451 + Ex1-IVS1-Ex2 wt), to prevent the possibility that the decay machinery might be saturated by the spliced mRNAs, we used proportionately less of this construct, while keeping constant the amount of total transfected DNA. While it was not possible to determine the half-life of the 451 + Ex1-IVS1-Ex2 mut mRNA (whose expression was negligible), the mRNA species of both the 451 + Ex1-IVS1-Ex2 wt and the 451 + Ex1Ex2 constructs exhibited half-lives of ~ 7 h, indicating no significant differences in mRNA stability dependent on splicing (Fig. 8D). Lastly, the relative luciferase activities positively correlate with the occurrence of splicing. In fact, the same constructs were transfected in SH-SY5Y cells and their luciferase activity was measured (Fig. 9): the activity of the 451 + Ex1Ex2 wt (containing the pre-arranged and correctly spliced 5′UTR
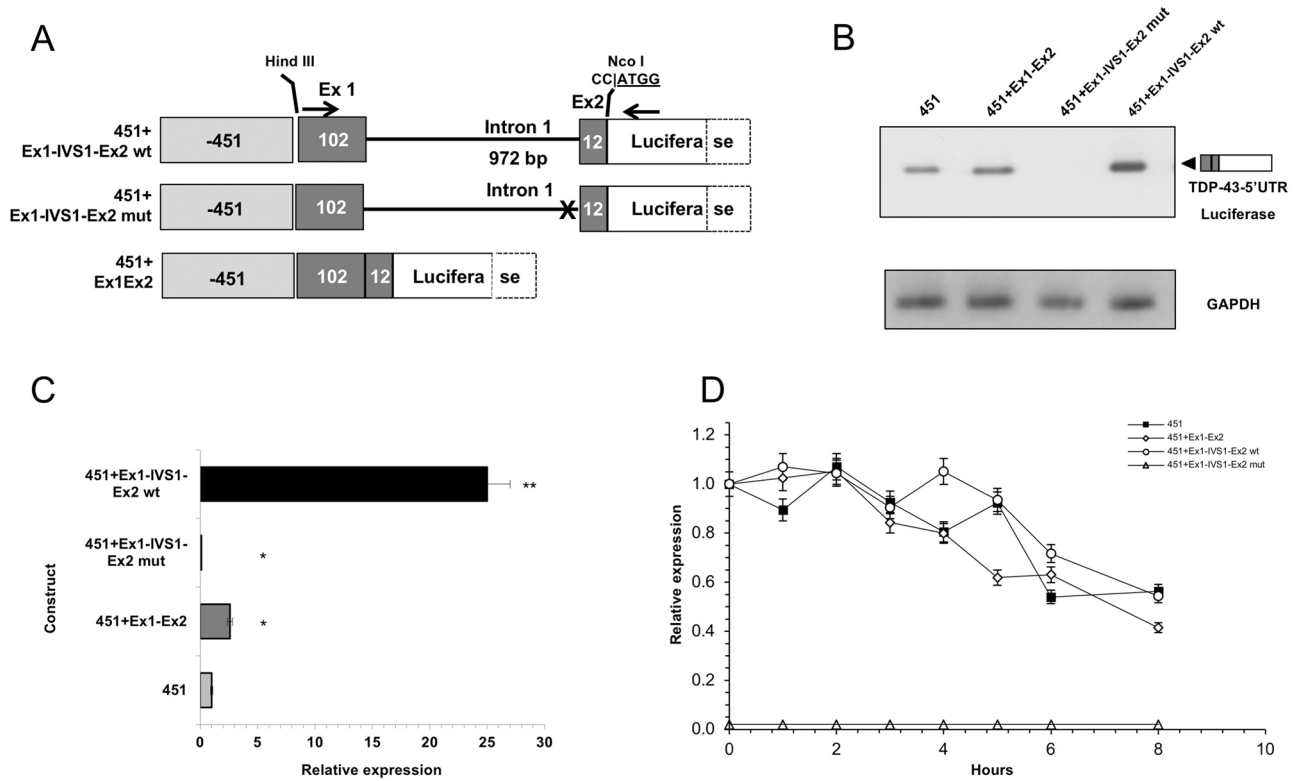
Figure 8. Effects of 5′UTR and intron 1 on the transcriptional activity of the human TARDBP promoter. (**A**) Diagrams of the 5′UTR TDP-43 constructs generated using the pGL-451 vector. In the 451 + Ex1-IVS1-Ex2 wt construct, the 1086 bp genomic region of the human TARDBP gene spanning the 5′UTR, including Exon 1 (102 bp), Intron 1 (972 bp) to the first 12 nt of Exon 2 was (Hind III-Nco I) cloned in between the 451nt-promoter and the firefly luciferase ORF (Luciferase). In the 451 + Ex1-IVS1-Ex2 mut construct, the 3′ splicing site of Intron 1 was deleted. In the 451 + Ex1Ex2 construct, the fragment encompassing Exon 1 and Exon 2 (without Intron 1) was cloned in the 451 vector. (**B**) RT-PCR of the 5′UTR TDP-43 mRNA species (upper panel). Amplification of GAPDH (lower panel) was used as the endogenous control in the quantitative analysis of RT-PCR. (**C**) Quantitation of the 5′UTR TDP-43 mRNA species by Real Time PCR. GAPDH was used to confirm normalization of total RNA levels. Co-transfected renilla orf was used to normalize the luciferase qPCR. The 451 construct was used as reference (= 1). (**D**) Effects of transcriptional inhibition on mRNA stability. After 36 h post-transfection of the luciferase (451 + Ex1-IVS1-Ex2 wt and 451 + Ex1-IVS1-Ex2 mut) constructs, SH-SY5Y cells were treated with actinomycin D 5 μg/ml. The relative levels of the indicated mRNAs (wt and mut) were assessed at the designated time points, following shutoff of transcription using qRT-PCR to determine mRNA half-lives. The average half-lives are reported with SD from two independent experiments. Significance values refer to comparisons against t = 0 h, wt: *$p < 0.05$ (one way ANOVA with Tukey test). The mean values for 451 + Ex1-IVS1-Ex2 mut construct were negligible and did not differ to a statistically significant extent ($p \geq 0.05$).
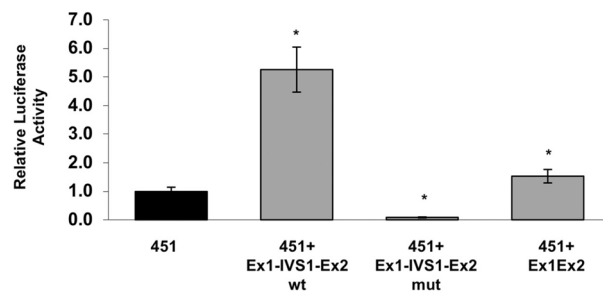
Figure 9. Effects of 5′UTR and intron 1 on the luciferase activity of the TARDBP promoter constructs. The luciferase activity of 5′UTR TDP-43 constructs depicted in Fig. 8A was assayed by transient cotransfections (luciferase and renilla) in SH-SY5Y cells. After 48 h, cells were assayed for firefly luciferase expression (normalized against renilla). The empty expression vector was tested and produced background levels of luciferase activity (data not shown). Data are mean (± SD) of three independent assays. Significance values refer to comparisons against control transfections (451): *$p < 0.05$ (one way ANOVA with Tukey test).

region) was $1.5\times$ that of the control. The 451 + Ex1-IVS1-Ex2 wt construct showed a $5\times$ increment in luciferase activity (as compared to the control, -451 construct). On the other hand, as expected, the 451 + Ex1-IVS1-Ex2 mut construct showed insignificant activity when compared with the other constructs (Fig. 9). Altogether these results suggest that the presence of the 5′UTR as well as the correct splicing event of intron 1 are elements able to modulate the luciferase expression (and, potentially, of TDP-43) at transcriptional level.

Discussion

We have experimentally characterized the promoter region of the human TARDBP gene. A previous prediction indicated that the core regions of the TARDBP promoter could be located between 500 nucleotides upstream of exon 1 and 66 nucleotides downstream of exon 1 and in two regions of intron 1 (212 and 613 nucleotides)[52]. Our predictions failed to identify the canonical TATA box and CAAT box transcriptional regulatory elements and interrogation of the database DBTSS (cataloging the positions of TSSs in the genome) outlined how, in most TDP-43 transcripts, transcription starts from the TSS indicated in the main transcript isoform NM_007375. Regarding the spatio-temporal expression pattern, TDP-43 is a developmentally regulated protein[58,64], while its distribution in adult tissues seems to be ubiquitous[6,58,64]. These observations are consistent with the hypotheses supported by recent studies reporting that TATA box is present in only a minority of promoters[65–67], often with tissue-specificity such as liver and muscle[68–70].

From an evolutionary perspective, it is interesting to note that, among primates, the greatest degree of evolutionary conservation encompasses the 600nts upstream of the TSS. On the other hand, the alignment of human and rodents sequences highlights the 200nts upstream of the TSS as the region of higher homology. These observations suggested that the region in-between plays an important role in directing transcription of the TARDBP gene. Indeed, this hypothesis is consistent with the experimental mapping performed in this study that has outlined the region spanning nucleotides 451–230 as necessary for the minimal promoter activity in all the tested cell lines (Fig. 4).

The comparison of the transcriptional activity in different cell lines highlights how the TARDBP promoter activity is high in cell lines of neuronal origin as well as in HeLa, as compared to the human embryonic kidney 293 cells (Fig. 5). These observations show that the promoter is active in human and mouse cell lines of different tissue origin (as expected) and suggest that apparent relative lower activity observed in HEK293 cells might rely on biochemical and/or genomic factors associated to this specific cell line. In addition, it should be also considered that, in comparison with other tested cell lines, HEK293 cell line has no evident tissue-specific gene expression and show expression of markers of several tissues[71].

It is well established that neuronal TDP-43 cytoplasmic inclusions are present in several different neurodegenerative diseases, such as FTD, ALS and Alzheimer and that the sequestration of this protein in aggregates may create gain- and loss-of-function events causing (directly or indirectly) cellular toxicity or alteration of the TDP-43-regulated gene expression. Understanding the factors that control TDP-43 expression within cells will certainly provide a better insight into disease origin and progression. Indeed, it has been observed that an increase of TDP-43 protein levels can activate the aggregation process in both cytoplasmic and nuclear compartments[20,60]. In addition, studies with animal models have shown that TDP-43 overexpression can be neurotoxic, even without the presence of aggregates[30,32,33,72,73]. Indeed, although TDP-43 can regulate its own expression through a negative feedback loop[44], some observations suggest that regulation of its expression might may also occur at transcriptional level. In particular, TDP-43 expression seems to be developmentally regulated[64], but also TDP-43 expression has been shown to decrease with age in a variety of organisms such as fruit fly[59] and mouse[58].

Looking for potential modifiers of the TARDBP promoter activity, we have focused the attention on two SNPs (c.1-562t>c, rs9430335 and c.1-100t>c, rs968545) previously identified within the TARDBP promoter of SALS patients. Although the variants do not change any transcription factor binding site, we have functionally tested their transcriptional impact. Although, in our system, we did not find statistically significant differences in the promoter activity of each haplotype, we can hypothesize that the potential regulatory SNPs do not lead to variation of TARDBP promoter activity directly, but cannot exclude that other possibilities of how these may affect transcription levels do exist. The variants might affect transcription at endogenous copy numbers, but not for overexpressed exogenous constructs. In addition regulatory elements, outside the promoter region included in our reporter plasmid and might interplay with the variants and modulate the activity of the promoter.

Subsequently, since TDP-43 autoregulates its own expression through a negative feedback mechanism via direct interaction with its own 3′UTR[44,48,74,75], we verified if there are transcriptional effects of TDP-43 on its own promoter activity and it does not modulate its own transcription.

Then, considering the known ability of the 5′UTRs to affect mainly protein translation efficiency[76], we next sought to test the possible effects of the TARDBP 5′UTR and intron 1 splicing on the expression of the associated reporter gene (Fig. 8).

On one hand, the observation of a significant increase in luciferase activity associated with the presence of the TARDBP "pre-arranged" 5′UTR (construct 451 + Ex1Ex2) indicated that this element can positively influence gene expression. Consistently, recent studies have shown that approximately 35% of human 5′UTRs are annotated as harboring introns[77] and recent observations suggest that introns in UTRs may have specific regulatory functions by affecting the rate of transcription rather than transcript stability. On the other hand, our experiments aimed at testing the relative mRNA decay rates of the luciferase constructs indicate that these elements have little or no impact on the stability of the mature mRNA, and therefore, they might affect gene expression through modulation of other steps of RNA metabolisms, such as transcription, mRNA levels, mRNA export, translational efficiency[78–81].

In conclusion, growing lines of evidence suggest that the pathophysiology of ALS, FTD and other neurodegenerative diseases might be the consequence of the convergence of multiple risk factors acting also on changes

of TDP-43 levels. Indeed, TDP-43 expression can be modulated during development[64], in a cell specific fashion, and, through aging, these fluctuations are evolutionary conserved at least in mice and flies[58,59]. Notwithstanding there are no clear data of age-related changes in TDP-43 in human[82,83], altogether these observations suggest that a better understanding of the mechanisms and factors involved in controlling the activity of the TARDBP promoter and, consequently, TDP-43 expression will give clues to uncover novel processes implicated in the onset and progression of the TDP-43 proteinopathies.

Methods

Constructs.
The 1316 bp region upstream of the TARDBP TSS was amplified by PCR and cloned in pGEM-T easy vector. The 1316 bp insert was digested with the restriction enzymes KpnI and HindIII (these sites were designed into the primers used in PCR) and subcloned into the pGL4.1 vector (Promega, Madison, WI, USA) digested with the same restriction enzymes. Other 7 deletion-mutants of TDP-43 promoter (927, 451, 380, 330, 280, 230, 180 bp upstream of the TSS) were then generated by PCR using gene-specific forward primers and the pGL4-1316 construct as template. Deletion-mutant inserts were KpnI-HindIII-subcloned in the pGL4.1 vector. TARDBP promoter SNPs were generated by PCR mutagenesis. The identity of all constructs was verified by DNA sequencing. Primer sequences are available.

Transfections and luciferase assays.
The cell lines used for transfections were the following: HEK293, HeLa, Neuro2A, and SH-SY5Y. Each construct was transfected in duplicate, 600 ng of DNA was diluted in a final volume of 100 µl with Opti-MEM I Reduced Serum Medium (ThermoFisher, Waltham, MA, USA). To normalize for transfection efficiency, the cells were co-transfected with 20 ng/well of pRL-TK (Promega, Madison, WI, USA) plasmid along with the promoter constructs. In parallel, 98 µl of Opti-MEM were added to 2 µl Lipofectamine 2000 Reagent (ThermoFisher, Waltham, MA, USA). After 5 min. at room temperature, the two solutions were combined and mixed, and 20 min later, the growth medium was removed from the cells and replaced with fresh DMEM containing 5% FBS. Then, the transfection mix was added to each well.

Total cell lysate was prepared from cells 48 h post transfection and firefly luciferase activity was assayed using the Beetle-Juice KIT (P.J.K. GmbH, Kleinblittersdorf, Germany) and a Turner Design 20/20 Luminometer (Turner BioSystems, Sunnyvale, CA, USA). Control transfections with pGL4.1 empty vectors were used as negative controls. Data presented are mean ± Standard deviation (SD) of at least three independent experiments and given as fold expression overexpression indicated in the Figures set arbitrarily at 1. Statistical analysis was performed using Student's T or ANOVA tests. In order to verify the possible influence of TDP-43 on its own promoter, we used the previously generated human stable cell line Flp-In HEK293 -TDP-43 wild type, carrying the human TDP-43 transgene locus-specific-integrated, whose expression is driven by an inducible version of the Tetracycline (Tet) promoter[44].

One day before transfection, Flp-In HEK293 -TDP-43 wild type (wt) cells were seeded at 60% confluence. Lipofectamine reagent was used for transfection with 300 ng of the reporter plasmids and 20 ng of pRL-TK vector. After 12 h, the medium was changed, and the cells were incubated further for 24 h. Flag-tagged TDP-43 expression was induced with 1 µg/ml tetracycline. The luciferase/renilla activities were then assayed and statistical analyses were performed using Anova one-way followed by Tukey's test.

Gene expression quantitation.
In order to quantify the mRNA levels of the 5′UTR TDP-43 constructs, co-transfections of luciferase constructs and pRL-TK were carried out as described in the section "Transfections and Luciferase assays".

Relative expression of constructs 451, 451 + Ex1Ex2, 451 Ex1-IVS1-Ex2 wt and 451 Ex1-IVS1-Ex2 mut constructs was performed by semi-quantitative RT–PCR and by quantitate Real Time PCR. Forty eight hours after transfection, cultures were terminated and RNA was extracted with Trifast reagent (Euroclone, Milan, Italy), according to manufacturer's instruction. In the semi-quantitative RT–PCRs, GAPDH was used as internal control. The sequences of primers were (5′–3′): GAPDH_Ex1s, 5′-CGCTCTCTGCTCCTCCTGTT-3′; GAPDH_Ex2as, 5′-CCATGGTGTCTGAGCGATGT-3′; TDP-43 Ex1 + 9 (277 bp)_s, 5′-AGCTTGCGCCATTTTGTGGGAGCGA-3′; Luc ATG + 155(277 bp)_as, 5′-GTAATGTCCACCTCGATATGTGCGT-3′. Optical density of amplicons was calculated using the ImageJ image processing program[84] and the expression levels were numerically expressed as the ratio of applicant density of the target amplicons over that of GAPDH.

Quantitation of the 5′UTR TDP-43 mRNAs by Real Time PCR, co-transfections of luciferase constructs and pRL-TK were carried out as described in the section "Transfections and Luciferase assays". The extracted RNA (1 µg) was DNAse-treated and reverse-transcribed using the M-MLV reverse transcriptase (Thermo Fisher Scientific, Waltham, MA, USA), and oligo-dT (0.2 µg/reaction as primer), in a final volume of 40 µl, essentially as indicated in the manufacturer's instructions. The CFX96 real-time PCR detection system (Bio-Rad Laboratories, Redmond, WA, USA) was used for all qPCRs. SYBR green assays were optimized (annealing temperature, and incubation time) with the iQ SYBR Green Supermix (Bio-Rad), according to the manufacturer's instructions, using tenfold serial dilutions of cDNAs. The amplification mixture consisted of 0.2 µM primers, 5 µl of SYBR Green Supermix, and 4 µl template DNA in a total volume of 10 µl. Samples were amplified with the following program: initial denaturation at 95 °C for 2 min, followed by 40 cycles of denaturation for 15 s at 95 °C and annealing/elongation for 30 s at 60 °C. The expression of the co-transfected renilla orf was used to normalize the luciferase qPCR data.

Primer sequences were: Luciferase (62 bp), forward 5′-CACAAAGCCATGAAGCG-3′ and reverse 5′-ATATGTGCGTCGGTAAAG-3′; Renilla (140 bp), forward 5′-CAGAACAAAGGAAACGGATGA-3′ and reverse 5′-CGCGTTACCATGTAAAAAA-3′.

Melting curves were examined to determine the specificity of each product and relative expression data were calculated with the ΔΔCT method[85].

For experiments with actinomycin D, SH-SY5Y cells were transfected with 3000 ng of no-intron constructs (451, 451 + Ex1Ex2) or of the 451 Ex1-IVS1-Ex2 mut plasmid or 600 ng of the 451 Ex1-IVS1-Ex2 mut construct in a p100 plate. Total DNA was normalized to 3000 ng by adding an appropriate amount of pUC18 vector. Twelve hours later, cells deriving from each transfection were harvested and plated in different wells (one for each actinomycin incubation time point). Twenty-four hour after plating, actinomycin D (5 μg/ml) was added. The relative levels of the indicated mRNAs were assessed at the designated time points, following shutoff of transcription using qRT-PCR. Samples were collected at the indicated times using TRIzol. Total RNA from these samples was analyzed via qRT-PCR using the abovementioned Luciferase primers and normalization was carried out versus the housekeeping gene RPII (RPII_f: 5′-GCCCACGTCCAATGACAT-3′; RPII_r: 5′-GTGCGGCTGCTTCCA TAA-3′). Relative transcript abundances were determined using the ΔΔCt method[85].

Western blot.   Cells extracts were prepared in PBS containing 1×protease inhibitor (Roche Diagnostics, Indianapolis, IN, USA). Proteins were separated by SDS-PAGE and transferred to nitrocellulose (Cytiva, formerly GE Healthcare, Chicago, IL, USA) and protein detection was carried out with standard Western blotting techniques. After transfer, membranes were incubated for 10 h in blocking solution (5% nonfat dry milk in PBS containing 0.1% Tween-20, T-PBS) to prevent non-specific binding. Subsequently, membranes were incubated for 1 h at room temperature with specific primary antibodies diluted in blocking solution. Expression levels of both endogenous and flag-tagged human wild-type TDP-43 was monitored by using a commercially available polyclonal anti-TDP-43 antibody (Proteintech, Rosemont, IL, USA, 10782-2-AP). Endogenous tubulin was used as a loading control, using an in-house made mouse monoclonal antibody. Immunoblots were developed by using the ECL Star Enhanced Chemiluminescent Substrate (EuroClone, Pero, Milan, Italy).

References

1. Ratti, A. & Buratti, E. Physiological functions and pathobiology of TDP-43 and FUS/TLS proteins. *J. Neurochem.* **138**, 95–111 (2016).
2. Guo, L. & Shorter, J. Biology and pathobiology of TDP-43 and emergent therapeutic strategies. *Cold Spring Harb. Perspect. Med.* **7**, 024554 (2017).
3. Niksic, M., Romano, M., Buratti, E., Pagani, F. & Baralle, F. E. Functional analysis of cis-acting elements regulating the alternative splicing of human CFTR exon 9. *Hum. Mol. Genet.* **8**, 2339–2349 (1999).
4. Pagani, F. *et al.* Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. *J. Biol. Chem.* **275**, 21041–21047 (2000).
5. Buratti, E. & Baralle, F. E. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J. Biol. Chem.* **276**, 36337–36343 (2001).
6. Buratti, E. *et al.* Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *Embo J.* **20**, 1774–1784 (2001).
7. Buratti, E., Brindisi, A., Pagani, F. & Baralle, F. E. Nuclear factor TDP-43 binds to the polymorphic TG repeats in CFTR intron 8 and causes skipping of exon 9: A functional link with disease penetrance. *Am. J. Hum. Genet.* **74**, 1322–1325 (2004).
8. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
9. Arai, T. *et al.* TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem. Biophys. Res. Commun.* **351**, 602–611 (2006).
10. Probst, A., Taylor, K. I. & Tolnay, M. Hippocampal sclerosis dementia: A reappraisal. *Acta Neuropathol.* **114**, 335–345 (2007).
11. Amador-Ortiz, C. *et al.* TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer's disease. *Ann. Neurol.* **61**, 435–445 (2007).
12. Hasegawa, M. *et al.* TDP-43 is deposited in the Guam parkinsonism-dementia complex brains. *Brain* **130**, 1386–1394 (2007).
13. Higashi, S. *et al.* Concurrence of TDP-43, tau and alpha-synuclein pathology in brains of Alzheimer's disease and dementia with Lewy bodies. *Brain Res.* **1184**, 284–294 (2007).
14. Freeman, S. H., Spires-Jones, T., Hyman, B. T., Growdon, J. H. & Frosch, M. P. TAR-DNA binding protein 43 in pick disease. *J. Neuropathol. Exp. Neurol.* **67**, 62–67 (2008).
15. Uryu, K. *et al.* Concomitant TAR-DNA-binding protein 43 pathology is present in Alzheimer disease and corticobasal degeneration but not in other tauopathies. *J. Neuropathol. Exp. Neurol.* **67**, 555–564 (2008).
16. Mackenzie, I. R., Foti, D., Woulfe, J. & Hurwitz, T. A. Atypical frontotemporal lobar degeneration with ubiquitin-positive, TDP-43-negative neuronal inclusions. *Brain* **131**, 1282–1293 (2008).
17. Bigio, E. H. *et al.* TDP-43 pathology in primary progressive aphasia and frontotemporal dementia with pathologic Alzheimer disease. *Acta Neuropathol.* **120**, 43–54 (2010).
18. Xu, Z. S. Does a loss of TDP-43 function cause neurodegeneration?. *Mol. Neurodegener.* **7**, 27 (2012).
19. Gao, J., Wang, L., Huntley, M. L., Perry, G. & Wang, X. Pathomechanisms of TDP-43 in neurodegeneration. *J. Neurochem.* **146**, 7–20 (2018).
20. Budini, M., Romano, V., Quadri, Z., Buratti, E. & Baralle, F. E. TDP-43 loss of cellular function through aggregation requires additional structural determinants beyond its C-terminal Q/N prion-like domain. *Hum. Mol. Genet.* **24**, 9–20 (2015).
21. Vanden Broeck, L., Callaerts, P. & Dermaut, B. TDP-43-mediated neurodegeneration: Towards a loss-of-function hypothesis?. *Trends Mol. Med.* **20**, 66–71 (2014).
22. Iguchi, Y. *et al.* Loss of TDP-43 causes age-dependent progressive motor neuron degeneration. *Brain* **136**, 1371–1382 (2013).
23. Ash, P. E. A. *et al.* Neurotoxic effects of TDP-43 overexpression in *C. elegans. Hum. Mol. Genet.* **19**, 3206–3218 (2010).
24. Li, Y. *et al.* A Drosophila model for TDP-43 proteinopathy. *Proc. Natl. Acad. Sci. USA* **107**, 3169–3174 (2010).
25. Feiguin, F. *et al.* Depletion of TDP-43 affects Drosophila motoneurons terminal synapsis and locomotive behavior. *FEBS Lett.* **583**, 1586–1592 (2009).

26. Godena, V. K. *et al.* TDP-43 regulates Drosophila neuromuscular junctions growth by modulating Futsch/MAP1B levels and synaptic microtubules organization. *PLoS ONE* **6**, e17808 (2011).
27. Xu, Y.-F. *et al.* Wild-type human TDP-43 expression causes TDP-43 phosphorylation, mitochondrial aggregation, motor deficits, and early mortality in transgenic mice. *J. Neurosci. Off. J. Soc. Neurosci.* **30**, 10851–10859 (2010).
28. Igaz, L. M. *et al.* Dysregulation of the ALS-associated gene TDP-43 leads to neuronal death and degeneration in mice. *J. Clin. Invest.* **121**, 726–738 (2011).
29. Stallings, N. R., Puttaparthi, K., Luther, C. M., Burns, D. K. & Elliott, J. L. Progressive motor weakness in transgenic mice expressing human TDP-43. *Neurobiol. Dis.* **40**, 404–414 (2010).
30. Wils, H. *et al.* TDP-43 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of ALS and frontotemporal lobar degeneration. *Proc. Natl. Acad. Sci. USA* **107**, 3858–3863 (2010).
31. Tsao, W. *et al.* Rodent models of TDP-43: Recent advances. *Brain Res.* **1462**, 26–39 (2012).
32. Wegorzewska, I., Bell, S., Cairns, N. J., Miller, T. M. & Baloh, R. H. TDP-43 mutant transgenic mice develop features of ALS and frontotemporal lobar degeneration. *Proc. Natl. Acad. Sci.* **106**, 18809–18814 (2009).
33. Shan, X., Chiang, P. M., Price, D. L. & Wong, P. C. Altered distributions of Gemini of coiled bodies and mitochondria in motor neurons of TDP-43 transgenic mice. *Proc. Natl. Acad. Sci. USA* **107**, 16325–16330 (2010).
34. Uchida, A. *et al.* Non-human primate model of amyotrophic lateral sclerosis with cytoplasmic mislocalization of TDP-43. *Brain J. Neurol.* **135**, 833–846 (2012).
35. Kasai, T. *et al.* Increased TDP-43 protein in cerebrospinal fluid of patients with amyotrophic lateral sclerosis. *Acta Neuropathol.* **117**, 55–62 (2009).
36. Noto, Y.-I. *et al.* Elevated CSF TDP-43 levels in amyotrophic lateral sclerosis: specificity, sensitivity, and a possible prognostic value. *Amyotroph. Lateral Scler.* **12**, 140–143 (2011).
37. Majumder, V., Gregory, J. M., Barria, M. A., Green, A. & Pal, S. TDP-43 as a potential biomarker for amyotrophic lateral sclerosis: A systematic review and meta-analysis. *BMC Neurol.* **18**, 90 (2018).
38. Mougeot, J.-L.C., Li, Z., Price, A. E., Wright, F. A. & Brooks, B. R. Microarray analysis of peripheral blood lymphocytes from ALS patients and the SAFE detection of the KEGG ALS pathway. *BMC Med. Genom.* **4**, 74 (2011).
39. Verstraete, E. *et al.* TDP-43 plasma levels are higher in amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler.* **13**, 446–451 (2012).
40. Swarup, V. *et al.* Deregulation of TDP-43 in amyotrophic lateral sclerosis triggers nuclear factor κB–mediated pathogenic pathways. *J. Exp. Med.* **208**, 2429–2447 (2011).
41. Foulds, P. *et al.* TDP-43 protein in plasma may index TDP-43 brain pathology in Alzheimer's disease and frontotemporal lobar degeneration. *Acta Neuropathol.* **116**, 141–146 (2008).
42. Zhang, N., Gu, D., Meng, M. & Gordon, M. L. TDP-43 is elevated in plasma neuronal-derived exosomes of patients with Alzheimer's disease. *Front. Aging Neurosci.* **12**, 166 (2020).
43. Suárez-Calvet, M. *et al.* Plasma phosphorylated TDP-43 levels are elevated in patients with frontotemporal dementia carrying a C9orf72 repeat expansion or a GRN mutation. *J. Neurol. Neurosurg. Psychiatry* **85**, 684–691 (2014).
44. Ayala, Y. M. *et al.* TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J.* **30**, 277–288 (2011).
45. Buratti, E. & Baralle, F. E. TDP-43: new aspects of autoregulation mechanisms in RNA binding proteins and their connection with human disease. *FEBS J.* **278**, 3530–3538 (2011).
46. Avendano-Vazquez, S. E. *et al.* Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. *Genes Dev.* **26**, 1679–1684 (2012).
47. Bembich, S. *et al.* Predominance of spliceosomal complex formation over polyadenylation site selection in TDP-43 autoregulation. *Nucleic Acids Res.* **42**, 3362–3371 (2014).
48. White, M. A. *et al.* TDP-43 gains function due to perturbed autoregulation in a Tardbp knock-in mouse model of ALS-FTD. *Nat. Neurosci.* **21**, 552–563 (2018).
49. Sugai, A. *et al.* Non-genetically modified models exhibit TARDBP mRNA increase due to perturbed TDP-43 autoregulation. *Neurobiol. Dis.* **130**, 104534 (2019).
50. Weskamp, K. *et al.* Shortened TDP43 isoforms upregulated by neuronal hyperactivity drive TDP43 pathology in ALS. *J. Clin. Invest.* **130**, 1139–1155 (2020).
51. Hawley, Z. C. E., Campos-Melo, D. & Strong, M. J. Evidence of A negative feedback network between TDP-43 and miRNAs dependent on TDP-43 nuclear localization. *J. Mol. Biol.* **432**, 166695 (2020).
52. Luquin, N., Yu, B., Saunderson, R. B., Trent, R. J. & Pamphlett, R. Genetic variants in the promoter of TARDBP in sporadic amyotrophic lateral sclerosis. *Neuromuscul. Disord.* **19**, 696–700 (2009).
53. Suzuki, A. *et al.* DBTSS/DBKERO for integrated analysis of transcriptional regulation. *Nucleic Acids Res.* **46**, D229–D238 (2018).
54. Suzuki, A. *et al.* DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. *Nucleic Acids Res.* **43**, D87–D91 (2015).
55. Dreos, R., Ambrosini, G., Groux, R., CavinPérier, R. & Bucher, P. The eukaryotic promoter database in its 30th year: Focus on non-vertebrate organisms. *Nucleic Acids Res.* **45**, D51–D55 (2017).
56. Lee, T.-Y., Chang, W.-C., Hsu, J.B.-K., Chang, T.-H. & Shien, D.-M. GPMiner: an integrated system for mining combinatorial cis-regulatory elements in mammalian gene group. *BMC Genom.* **13**, S3 (2012).
57. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
58. Huang, C., Xia, P. Y. & Zhou, H. Sustained expression of TDP-43 and FUS in motor neurons in Rodent's lifetime. *Int. J. Biol. Sci.* **1**, 396–406. https://doi.org/10.7150/ijbs.6.396 (2010).
59. Cragnaz, L. *et al.* An age-related reduction of brain TBPH/TDP-43 levels precedes the onset of locomotion defects in a Drosophila ALS model. *Neuroscience* **311**, 415–421 (2015).
60. Budini, M. *et al.* Cellular model of TAR DNA-binding protein 43 (TDP-43) aggregation based on its C-terminal Gln/Asn-rich region. *J. Biol. Chem.* **287**, 7512–7525 (2012).
61. De Conti, L. *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).
62. Prpar Mihevc, S., Baralle, M., Buratti, E. & Rogelj, B. TDP-43 aggregation mirrors TDP-43 knockdown, affecting the expression levels of a common set of proteins. *Sci. Rep.* **6**, 33996 (2016).
63. Bensaude, O. Inhibiting eukaryotic transcription. Which compound to choose? How to evaluate its activity?. *Transcription* **2**, 103–108 (2011).
64. Sephton, C. F. *et al.* TDP-43 is a developmentally regulated protein essential for early embryonic development. *J. Biol. Chem.* **285**, 6826–6834 (2009).
65. FitzGerald, P. C., Shlyakhtenko, A., Mir, A. A. & Vinson, C. Clustering of DNA sequences in human promoters. *Genome Res.* **14**, 1562–1574 (2004).
66. Frith, M. C. *et al.* A code for transcription initiation in mammalian genomes. *Genome Res.* **18**, 1–12 (2008).
67. Bajic, V. B. *et al.* Mice and men: Their promoter properties. *PLoS Genet.* **2**, e54 (2006).
68. Yamashita, R., Suzuki, Y., Sugano, S. & Nakai, K. Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity. *Gene* **350**, 129–136 (2005).
69. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**, 626–635 (2006).

70. Valen, E. *et al.* Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Res.* **19**, 255–265 (2009).
71. Stepanenko, A. A. & Dmitrenko, V. V. HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene* **569**, 182–190 (2015).
72. Stallings, N. R., Puttaparthi, K., Luther, C. M., Burns, D. K. & Elliott, J. L. Progressive motor weakness in transgenic mice expressing human TDP-43. *Neurobiol. Dis.* https://doi.org/10.1016/j.nbd.2010.06.017 (2010).
73. Xu, Y.-F. *et al.* Wild-type human TDP-43 expression causes TDP-43 phosphorylation, mitochondrial aggregation, motor deficits, and early mortality in transgenic mice. *J. Neurosci.* **30**, 10851–10859 (2010).
74. Avendaño-Vázquez, S. E. *et al.* Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. *Genes Dev.* **26**, 1679–1684 (2012).
75. Sugai, A. *et al.* Robustness and vulnerability of the autoregulatory system that maintains nuclear TDP-43 levels: A trade-off hypothesis for ALS pathology based on in silico data. *Front. Neurosci.* **12**, 28 (2018).
76. Araujo, P. R. *et al.* Before it gets started: Regulating translation at the 5′ UTR. *Comp. Funct. Genomics* **2012**, 475731 (2012).
77. Cenik, C., Derti, A., Mellor, J. C., Berriz, G. F. & Roth, F. P. Genome-wide functional analysis of human 5′ untranslated region introns. *Genome Biol.* **11**, R29 (2010).
78. Buchman, A. R. & Berg, P. Comparison of intron-dependent and intron-independent gene expression. *Mol. Cell. Biol.* **8**, 4395–4405 (1988).
79. Bourdon, V., Harvey, A. & Lonsdale, D. M. Introns and their positions affect the translational activity of mRNA in plant cells. *EMBO Rep.* **2**, 394–398 (2001).
80. Fong, Y. W. & Zhou, Q. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414**, 929–933 (2001).
81. Le Hir, H., Gatfield, D., Izaurralde, E. & Moore, M. J. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20**, 4987–4997 (2001).
82. Uchino, A. *et al.* Incidence and extent of TDP-43 accumulation in aging human brain. *Acta Neuropathol. Commun.* **3**, 1 (2015).
83. Wilson, A. C., Dugger, B. N., Dickson, D. W. & Wang, D.-S. TDP-43 in aging and Alzheimer's disease: A review. *Int. J. Clin. Exp. Pathol.* **4**, 147–155 (2011).
84. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
85. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods* **25**, 402–408 (2001).

Acknowledgements

Author contributions

Study conception and design: M.R.; Acquisition of data: M.B. and M.R.; Analysis and interpretation of data: M.B. and M.R.; Drafting of manuscript: M.R. and M.B.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-89973-z.

Correspondence and requests for materials should be addressed to M.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.