1 Supplementary Notes

1.1 Sequences of some of the TDP43 targets in Section 2.4

CLIP34nt: GAGAGAGCGCGUGCAGAGACUUGGUGGUGCAUAA

CLIP6: UUGUGGUGUGCUUUGCAGGAGGACU

CLIP34nt_UG6: GAGAGAGCGCGUGUGUGUGUGGUGGUGCAUAA

1.2 Sequences of wild-type and mutant 3' UTRs of ERBB2 mRNA in section 2.7

WT-URE/WT-331b:

GGGCGAAUUGGAGCUCCACCGCGGUGGCGGCCGCUCUAGAAGUGCUUUUCUGUUUAGU-
UUUUACUUUUUUUGUUUUGUUUUUUUAAAGAUGAAAUAAAGACCCAGGGGGGGCCCGGUAC

MT-URE/WT-331b:

GGGCGAAUUGGAGCUCCACCGCGGUGGCGGCCGCUCUAGAAGUGCUUUUCUGUUUAGU-
UUUUACUGUGUGUGUGGUGUGUGUGUAAAGAUGAAAUAAAGACCCAGGGGGGGCCCGGUAC

WT-URE/MT-331b:

GGGCGAAUUGGAGCUCCACCGCGGUGGCGGCCGCUCUAGAAGUGCUUUUCUGUUUAGU-
UUUUACUUUUUUUGUUUUGUUUUUUUAAAGAUGAAAUAAAGAGCGACGCGGGGCCCGGUAC

MT-URE/MT-331b:

GGGCGAAUUGGAGCUCCACCGCGGUGGCGGCCGCUCUAGAAGUGCUUUUCUGUUUAGU-
UUUUACUGUGUGUGUGGUGUGUGUGUAAAGAUGAAAUAAAGAGCGACGCGGGGCCCGGUAC

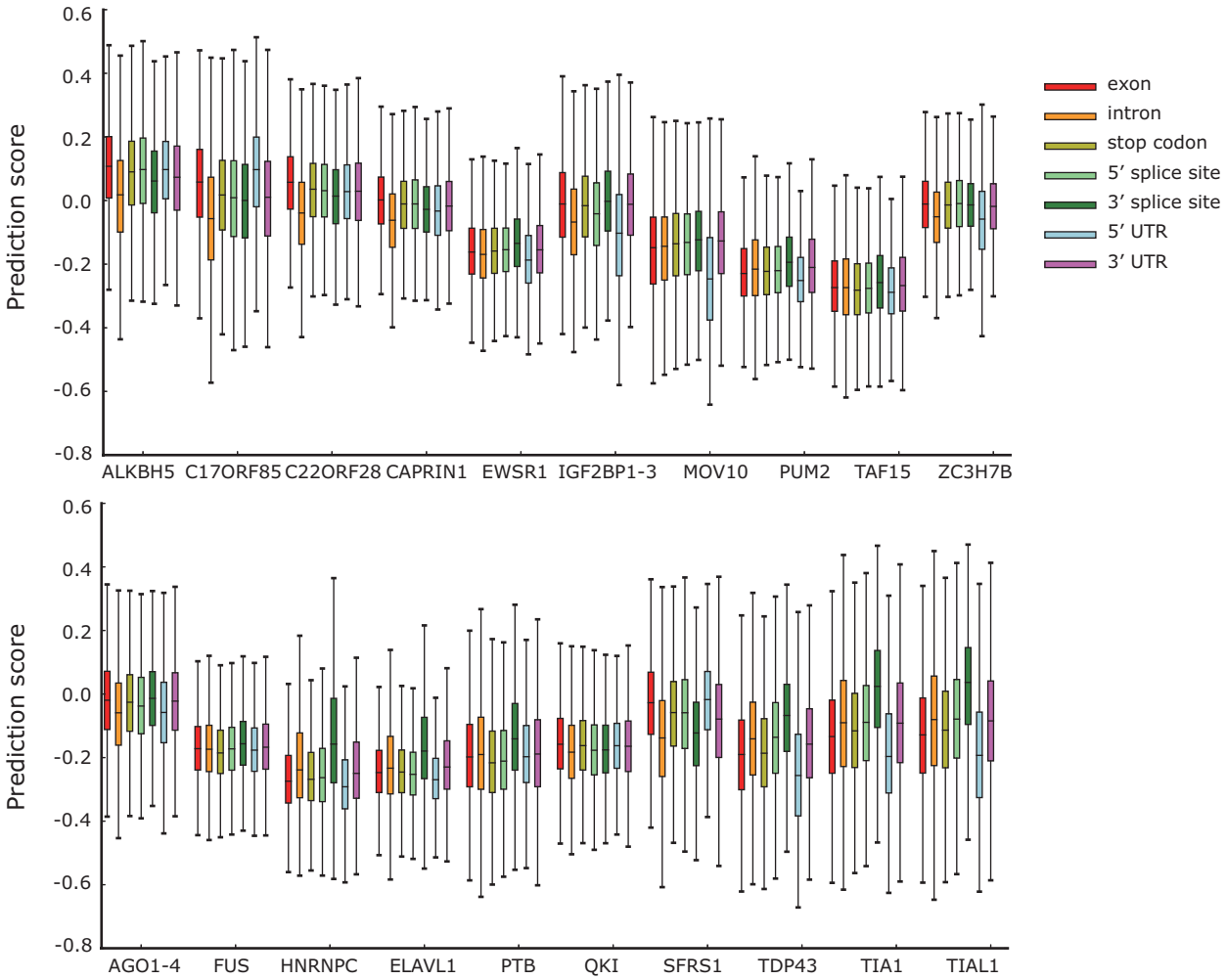
1.3 The distributions of the predicted binding scores across the transcriptome reflect the functions of different RBPs

The distributions of the prediction scores of DeBooster over different characterized genomic regions reflected the functions of individual RBPs (Supplementary Fig 1). In particular, C17ORF85 (also known as NCBP3) had high prediction scores in the 5' UTRs, which was consistent with the known fact that C17ORF85 binds to the m7G caps of mRNAs, and is actively involved in mRNA transport [1]. In addition, EWRS1, FUS and TAF15 all belong to the FET family and share similar patterns of binding preferences. For example, their binding sites were enriched near the 3' splice sites, which agreed with the previous study [2]. Consistent with the previous results derived from PAR-CLIP experiments [3], our results also showed that both PUM2 and IGF2BP1-3 binding sites were enriched in the 3' UTRs. This observation aligned with the previous evidences that PUM2 binds to the 3' UTRs of mRNAs and regulates the miRNA-mediated mRNA degradation [4], and the IGF2BP proteins play an important role in the regulation of mRNA transport [5]. Also, as MOV10 mainly functions as an RNA helicase regulating mRNA stability [6], it was not suprising to see that its binding targets were enriched in the 3' UTRs while depleted in the 5' UTRs. The argonaute proteins (AGO1-4) have important regulatory functions in miRNA processing and miRNA-mediated gene silencing [7]. Previous experiments also found that these argonaute proteins bind to the splice sites, especially the 3' splice sites, and actively participate in splicing regulation [8]. These known functions of the argonaute proteins can also be reflected in our comparative studies, in which both 3' UTRs and the regions near the 3' splice sites displayed relatively higher prediction scores than other regions. TDP43 plays an important role in multiple aspects of gene regulation, such as DNA/RNA binding and splicing [9]. A previous study [10] showed that TDP43 has a relatively large proportion of intronic binding sites and a small portion of exonic binding targets. Such a result was also consistent with our observation. ELAVL1, as called HUR, is a well-known RBP that increases mRNA stability and regulates alternative splicing [11]. Both previous PAR-CLIP experiments [12] and our comparative studies of the prediction scores conformed that the ELAVL1 binding sites are significantly enriched in the 3' UTRs and near the 3' splice sites. The splicing regulator SFRS1 is a proto-oncogene whose overexpression can be involved in various types of cancers [13]. Our comparison showed that SFRS1 prefers binding to exons and the 5' UTRs. Probably this phenomenon can be explained by the known fact that SFRS1 actively binds to

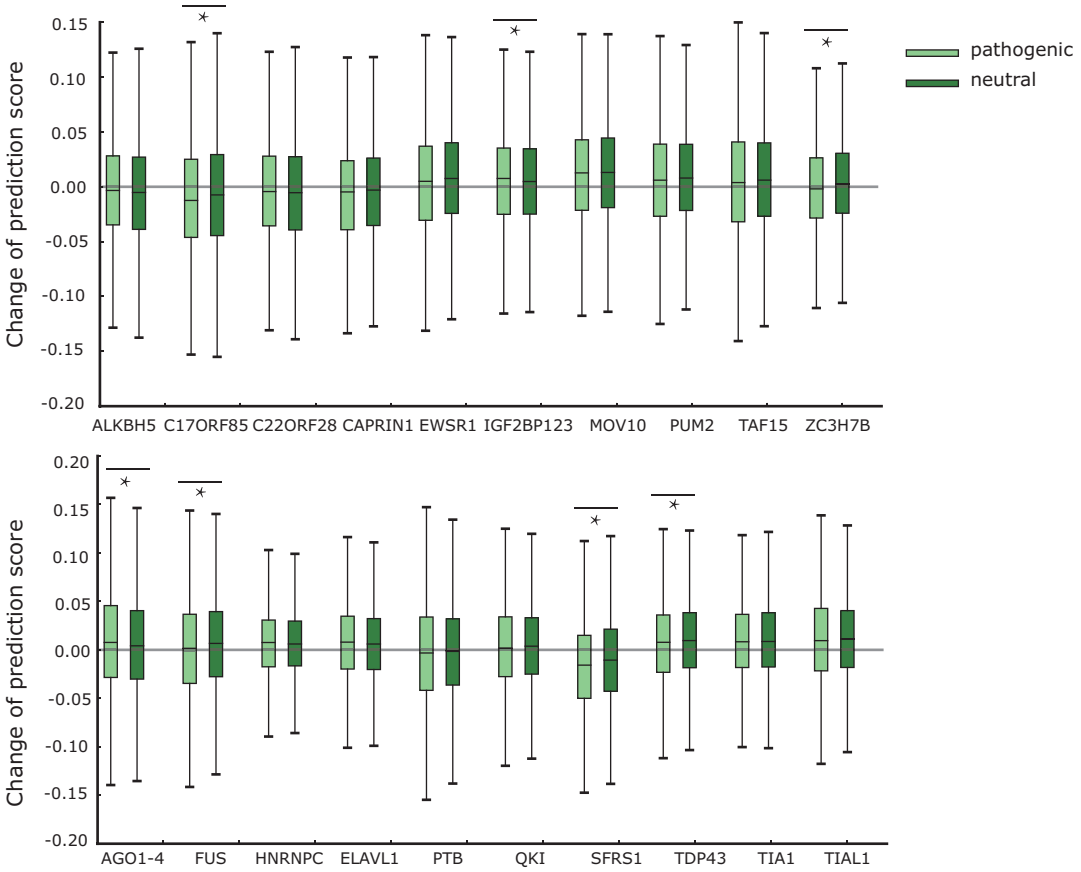
the exonic splicing enhancers (ESEs) or exonic splicing silencers (ESSes) [14] and may also play a critical role in translational regulation [15–17].

Although the major results of our comparative analysis were in agreement with previous known functions of RBPs, there were a few places that our studies did not exactly match the previous results derived from CLIP-seq experiments. For example, we observed an enrichment of TIA1 binding near the 3' splice sites, while the previous iCLIP results [18] showed that a large fraction of TIA1 targets were located around the 5' splice sites. This discrepancy was probably due to the noise (e.g., false positives or false negatives) from experimental data. Nevertheless, the preferred binding regions of individual RBPs identified by DeBooster were mostly consistent with the previously knwon evidences or studies about the functions of these RBPs.

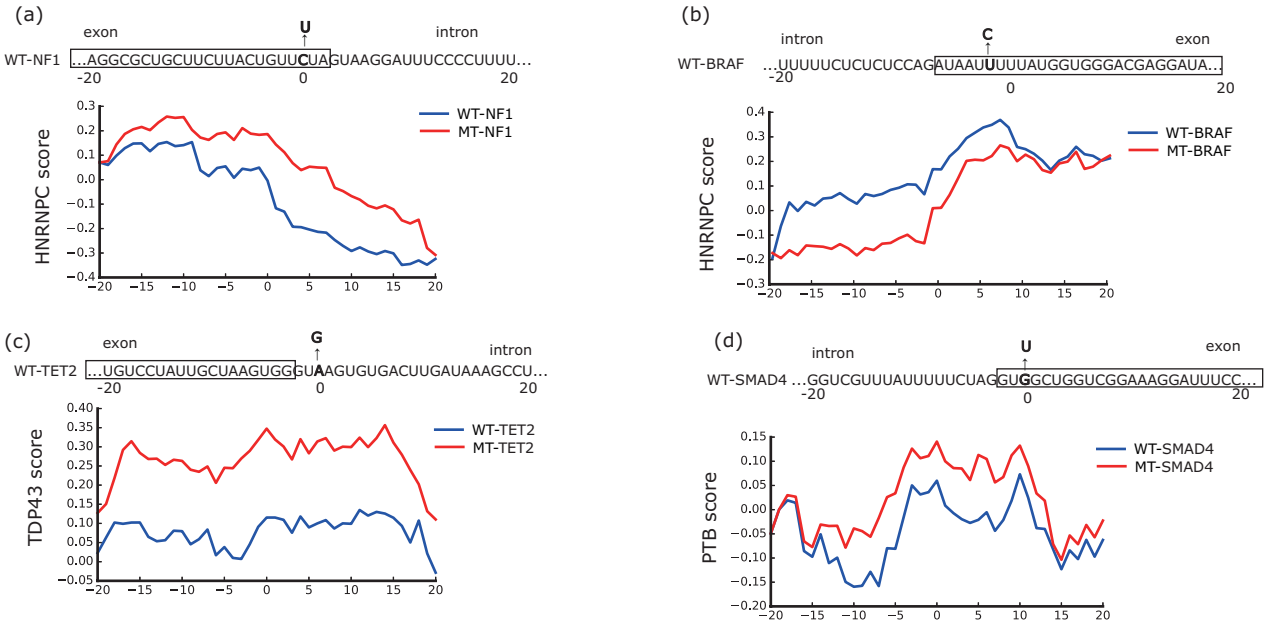
2 Supplementary Figures



Supplementary Figure 1 The distributions of the binding scores predicted by DeBooster in different characterized genomic regions for individual RBPs.



Supplementary Figure 2 The changes of the predicted binding scores predicted by DeBooster for the pathogenic and neutral mutations that were randomly selected from the COSMIC records (Methods). *: $p<0.001$, Student's t test.



Supplementary Figure 3 Examples of the predicted effects on the potentially disease-causing mutations near the splice sites.

References

[1] Anna Gebhardt, Matthias Habjan, Christian B Benda, Arno Meiler, Darya A Haas, Marco Y Hein, Angelika Mann, Matthias Mann, Bianca Habermann, and Andreas Pichlmair. mRNA export through an additional cap-binding complex consisting of NCBP1 and NCBP3. *Nature Communications*, 6, 2015.

[2] Erik G Larsson, Simon Runge, Jeffrey D Nusbaum, Sujitha Duggimpudi, Thalia A Farazi, Markus Hafner, Arndt Borkhardt, Chris Sander, and Thomas Tuschl. RNA targets of wild-type and mutant FET family proteins. *Nature Structural & Molecular Biology*, 18(12):1428, 2011.

[3] Markus Hafner, Markus Landthaler, Lukas Burger, Mohsen Khorshid, Jean Hausser, Philipp Berninger, Andrea Rothballer, Manuel Ascano, Annacarina Jungkamp, Mathias Munschauer, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 141(1):129–141, 2010.

[4] Antonio Galgano, M Forrer, Mihaela Zavolan, and Andre P Gerber. Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. *PLOS ONE*, 3(9), 2008.

[5] Jessica L Bell, Kristin Wachter, Britta Muhleck, Nikolaos Pazaitis, Marcel Kohn, Marcell Lederer, and Stefan Huttelmaier. Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs): post-transcriptional drivers of cancer progression? *Cellular and Molecular Life Sciences*, 70(15):2657–2675, 2012.

[6] Lea Haarup Gregersen, Markus Schueler, Mathias Munschauer, Guido Mastrobuoni, Wei Chen, Stefan Kempa, Christoph Dieterich, and Markus Landthaler. MOV10 is a 5' to 3' RNA helicase contributing to UPF1 mRNA target degradation by translocation along 3' UTRs. *Molecular Cell*, 54(4):573–585, 2014.

[7] Gyorgy Hutvagner and Martin J Simard. Argonaute proteins: key players in RNA silencing. *Nature Reviews Molecular Cell Biology*, 9(1):22, 2008.

[8] Maya Ameyarzazoua, Christophe Rachez, Mouloud Souidi, P Robin, Lauriane Fritsch, Robert Young, Nadya Morozova, Romain Fenouil, Nicolas Descostes, Jeanchristophe Andrau, et al. Argonaute proteins couple chromatin silencing to alternative splicing. *Nature Structural & Molecular Biology*, 19(10):998, 2012.

[9] Emanuele Buratti and Francisco E Baralle. The multiple roles of TDP-43 in pre-mRNA processing and gene expression regulation. *RNA Biology*, 7(4):420, 2010.

[10] Clotilde Lagiertourenne and Magdalini Polymenidou. TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. *Human Molecular Genetics*, 19, 2010.

[11] Neelanjan Mukherjee, David L Corcoran, Jeffrey D Nusbaum, Stoyan Georgiev, Markus Hafner, Manuel Ascano, Thomas Tuschl, Uwe Ohler, and Jack D Keene. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Molecular Cell*, 43(3):327–339, 2011.

[12] Svetlana Lebedeva, Marvin Jens, Kathrin Theil, Bjorn Schwanhausser, Matthias Selbach, Markus Landthaler, and Nikolaus Rajewsky. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Molecular Cell*, 43(3):340–352, 2011.

[13] Rotem Karni, Elisa De Stanchina, Scott W Lowe, Rahul Sinha, David Mu, and Adrian R Krainer. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nature Structural & Molecular Biology*, 14(3):185, 2007.

[14] Jeremy R Sanford, Xin Wang, Matthew Mort, Natalia Vanduyn, D N Cooper, Sean D Mooney, Howard J Edenberg, and Yunlong Liu. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Research*, 19(3):381, 2009.

[15] Gracjan Michlewski, Jeremy R Sanford, and Javier F Caceres. The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Molecular Cell*, 30(2):179–189, 2008.

[16] Jeremy R Sanford, Nicola K Gray, Karsten Beckmann, and Javier F Caceres. A novel role for shuttling SR proteins in mRNA translation. *Genes & Development*, 18(7):755–768, 2004.

[17] J S Kim, R Y Park, Jk Chen, Soojin Jeong, and T Ohn. Splicing factor SRSF3 represses the translation of programmed cell death 4 mRNA by associating with the 5'-UTR region. *Cell Death & Differentiation*, 21(3):481, 2014.

[18] Melis Kayikci, Michael Briese, Kathi Zarnack, Nicholas M Luscombe, Gregor Rot, Bla Zupan, Toma Curk, and Jernej Ule. iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLOS Biology*, 8(10), 2010.