Supplementary Notes and Figures for "A deep boosting based approach for capturing the sequence binding preferences of RNA-binding proteins from high-throughput CLIP-seq data"

S. Li[†], F. Dong[†], Y. Wu[†], S. Zhang, C. Zhang, X. Liu, T. Jiang and J. Zeng[*]

S1   Supplementary Notes

S1.1   Performance comparisons between the DeBooster models with different feature encoding schemes

We added two more computational experiments to investigate the individual contributions of the new feature vector and the deep boosting model to our final prediction. In the first experiment, we used the same feature vector but switched to a traditional Adaboost model to make prediction (denoted as "Adaboost"). In the second experiment, we dropped the word order feature encoding scheme but used the same deep boosting method to predict binding sites (denoted as "Fewer-feature"). The results of these two computational experiments are shown in Table S1 in which the prediction results of both original DeBooster model and previous DBN model are also shown for comparisons. In these two experiments, both "Adaboost" and "Fewer-feature" models achieved similar performance to that of the previous DBN method, but did not perform as well as DeBooster. Therefore, both the deep boosting method and the new feature vector significantly contribute the final excellent prediction results of DeBooster.

Next, to investigate whether the performance of DeBooster can be improved if the structural features were also added, we first predicted the secondary structures from the RNA sequence and then used the bag-of-word models with the word lengths of 1, 2 and 3 to encode the structural features. We then used the deep boosting model to train from both sequence and structural features (i.e., the combination of both original sequence features and new structural features). We found that adding the structural features did not improve greatly the performance (Table S1). Thus, we chose to use only the sequence features in our model.

S1.2   Comparison to GraphProt with respect to the generated sequence motifs

We also compared the sequence logos generated by DeBooster and GraphProt (Fig S4). The RBPs shown here are the same as the those shown in the GraphProt paper [1]. Since there may be many different reference motifs that were reported previously in the literature, to be fair, here we chose the same literature sources as in the GraphProt paper [2–9]. Overall, DeBooster seemed to generate better motifs than GraphProt in various cases. For example, the SFRS1 motif generated by DeBooster has a C at the end of the AG rich sequence, which was consistent with the literature [2] but was not captured by GraphProt. For TDP43, the Debooster result showed an obvious UG repeat pattern [5], while the sequence logo generated by GraphPort seemed to be dominated by U. In addition, the UGUANAUA motif [6] for PUM2

---

[*]To whom correspondence should be addressed. Tel: +86 10 62781693; Fax: +86 10 62797331; Email: zengjy321@tsinghua.edu.cn. [†]These authors contributed equally to the paper.

was more recognizable by DeBooster than GraphProt. Furthermore, the ACUAAC motif [6] for QKI was more clearly shown in the DeBooster result. This comparison further demonstrated the superiority of DeBooster over GraphProt in detecting the sequence features of RBP binding sites. Comparisons for the sequence motifs for all the 24 training datasets are also shown in Fig S5.

S1.3 Comparison to GraphProt with respect to the recovery of false negatives

We also tested the performance of GraphProt on the task of recovering false negatives. Table S2 showed the AUROC scores of GraphProt when using different combinations of training and testing datasets. Compared to Fig 3b in the main text, we found obvious that DeBooster outperformed GraphProt on this task. For each pair of training and testing datasets, all results predicted by DeBooster showed the AUROC scores larger than 0.9, while only two of the 12 GraphProt results were above 0.9.

S1.4 Comparison to GraphProt with respect to the analysis of the regulation of mRNA degradation

We also tested GraphProt using the same sequences as in the analysis of the regulation of mRNA degradation in the main text. Fig S6 shows the test results, which corresponds to Figs 5b-5d in the main text. Here the UTRs were grouped according to the fold changes of their mRNA half-lives after MOV10 knock-down. In Fig 5b and Fig S6a, we calculated the fraction of the UTRs that contained at least one predicted MOV10 binding site in each group. As the fold-changes of mRNA half-lives decreased, the fraction of UTRs with the predicted MOV10 binding also decreased in both DeBooster and GraphProt results. In Fig 5c and Fig S6b, we calculated the sum of positive DeBooster scores or GraphProt margins for each group, and found a similar decreasing trend in both results. In particular, if we grouped the UTRs according to the sum of positive DeBooster scores or GraphProt margins, their fold-changes of mRNA half-lives were also significantly different (Figs 5d and S6c). Thus, DeBooster and GraphProt both performed well and yielded similar results on this task.

On the other hand, in practice DeBooster runs much faster than GraphProt on such large-scale data analysis tasks, as DeBooster takes only sequence features as input while GraphProt requires the prediction of RNA secondary structure, which is often a time-consuming process. We also conducted a simple test to compare the running time of both DeBooster and GraphProt. In particular, we ran both methods to predict 10,000 samples using the same machine. DeBooster took only 13 seconds to complete the prediction, while GraphProt took about one hour and 11 minutes. Thus, there is still an advantage to use DeBooster to perform a prediction task on millions of samples.

S1.5 Comparison to GraphProt with respect to the analysis of ADAR1 binding patterns

We also compared the performance of DeBooster and GraphProt on predicting different types of ADAR1 binding sites (Fig S7). Tests were performed using the transcriptome data only from chromosome 17, as it would take too long for GraphProt to predict all the samples from the whole human transcriptome. For the binding-editing distances of the new predicted sites, both DeBooster and GraphProt showed significant difference between the three models, although GraphProt predictions displayed overall larger binding-editing distances. However, the motifs generated by GraphProt for the three models seemed alike. The sequence motifs generated by GraphProt were all homologous GC rich elements (Fig S7c). There may be several reasons to

	DBN	DeBooster	Adaboost	Fewer-feature	With structure
ALKBH5	0.714	0.752	0.718	0.733	0.736
C17ORF85	0.820	0.866	0.838	0.848	0.847
C22ORF28	0.792	0.840	0.802	0.830	0.828
CAPRIN1	0.834	0.903	0.849	0.873	0.878
AGO2	0.809	0.863	0.819	0.845	0.850
ELAVL1	0.966	0.972	0.960	0.965	0.967
SFRS1	0.931	0.950	0.926	0.937	0.940
HNRNPC	0.962	0.958	0.955	0.951	0.955
TDP43	0.876	0.909	0.889	0.895	0.906
TIA1	0.891	0.902	0.889	0.889	0.900
TIAL1	0.870	0.886	0.863	0.868	0.881
AGO1-4	0.881	0.908	0.867	0.882	0.893
ELAVL1(B)	0.961	0.976	0.961	0.957	0.962
ELAVL1(A)	0.966	0.972	0.961	0.963	0.967
EWSR1	0.966	0.969	0.961	0.953	0.957
FUS	0.980	0.982	0.978	0.964	0.967
ELAVL1(C)	0.994	0.995	0.993	0.979	0.983
IGF2BP1-3	0.879	0.883	0.857	0.870	0.878
MOV10	0.854	0.918	0.876	0.888	0.894
PUM2	0.971	0.973	0.960	0.961	0.967
QKI	0.983	0.980	0.975	0.976	0.979
TAF15	0.983	0.988	0.980	0.978	0.983
PTB	0.983	0.943	0.939	0.939	0.940
ZC3H7B	0.796	0.897	0.812	0.852	0.856

Table S1: The area under receiver operator characteristic curve (AUROC) scores of DBN, DeBooster, Adaboost, DeBooster with fewer features, and DeBooster with structure features, respectively.

Train & Test datasets	ELAVL1	ELAVL1(A)	ELAVL1(B)	ELAVL1(C)
ELAVL1	-	0.840	0.859	0.905
ELAVL1(A)	0.850	-	0.843	0.892
ELAVL1(B)	0.869	0.843	-	0.905
ELAVL1(C)	0.848	0.831	0.837	-

Table S2: Cross-dataset AUROC scores calculated using GraphProt predictions.

explain why the three GraphProt models showed almost identical motifs: (i)As shown by the previous validation tests, DeBooster may generate better motifs than GraphProt. So maybe GraphProt was just not able to learn the difference between the motifs of the three models. (ii) In the training datasets shown in Fig 2a in the main text, all binding sites were less than 75 nt, while the ADAR1 binding sites were much longer (average 190 nt). So maybe it was a harder task for GraphProt to learn the motifs in such long sequences. (iii) Maybe there was actually no difference between the motifs of the three models, and DeBooster simply learned some bias. We argued that this was not likely the case, because the binding-editing distances of the predicted sites recovered by the three models displayed significant difference, which indicated that there may exist some intrinsically different binding features between these three groups of binding sites.

S1.6 The influence of mutations inside and outside the RBP binding sites

If a mutation occurs outside the binding region, it should generally have nearly no effect on RBP binding, as it does not directly change the nucleotide sequence of the binding sites. On the other hand, in some cases, it may also have some slight influence, because it may cause the changes of local RNA secondary structure. In our DeBooster model, if a mutation nearby (but not inside) the binding sites appears in the flanking 150 nt regions, it may also change the prediction score. Nevertheless, such changes should be generally smaller compared to those caused by mutations inside the binding sites.

In the validation tests shown in the main text, some RBPs' binding sites spanned the whole 40 nt window, but others had shorter binding regions that was less than 40 nt. In particular, the average lengths of binding sites for QKI, FUS, HNRNPC and PTB, whose average lengths of binding sites were 31 nt, 25 nt, 39 nt and 26 nt, respectively. So several positions near the sides of the 40 nt windows shown in Fig 9f and Figs S3a, S3d for these RBPs actually corresponded to the prediction scores of binding sites for those mutations in the flaking regions. The changes of the prediction scores were actually relatively smaller in these positions, which indicated that the mutations in the flanking regions generally have less effect on the changes of the prediction scores.

We further investigated the p values derived from Student's t tests by examining the difference between the effects of pathogenic and neutral mutations on the changes of the DeBooster prediction scores, as in the previous statistical tests that we conducted in Fig 8. Fig S8 shows the analysis results on those mutations near $5'$ and $3'$ splice sites for seven RBPs. Among these seven proteins, QKI, FUS and PTB have average lengths of binding sites less than 40 nt. Correspondingly, we also observed obvious decrease in the p values near the edges of binding sites for these three RBPs. For other four RBPs (i.e., SFRS1, TIA1, TDP43 and HNRNPC) whose binding sites cover almost the whole 40 nt window (HNRNPC binding sites cover 39 nt), the p values did not display such a change pattern. All these results indicated that the mutations occurring outside the RBP binding sites generally have less influence on the predicted binding scores than those within the binding regions.

S1.7 The distributions of the predicted binding scores in different characterized genomic regions across the transcriptome

We performed a comparative analysis of the binding scores predicted by DeBooster for individual RBPs in different characterized genomic regions across the transcriptome. In particular, we used RefSeq [10] and UCSC [11] genome annotations for human release hg19. For all different types of genomic regions, including exons, introns and UTRs, we randomly selected 10,000 records whose lengths were equal to the average length of all RBP binding sites in training data. For each record of splice sites and stop codons, we also extended the selected sites both

upstream and downstream such that the splice site or stop codon was located at the center of each sample. After that, we used the trained DeBooster model to infer the prediction scores of individual selected regions.

The transcriptome-wide analysis showed that the distributions of the prediction scores of DeBooster over different characterized genomic regions may reflect the functions of individual RBPs (Fig S1). 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 [12]. 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 3′ splice sites, which agreed with the previous study [13]. Consistent with the previous results derived from PAR-CLIP experiments [6], 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 [14], and the IGF2BP proteins play an important role in the regulation of mRNA transport [15]. Also, as MOV10 mainly functions as an RNA helicase regulating mRNA stability [16], it was not surprising 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 [17]. Previous experiments also found that these argonaute proteins bind to splice sites, especially 3′ splice sites, and actively participate in splicing regulation [18]. These known functions of the argonaute proteins can also be reflected in our comparative studies, in which both 3′ UTRs and regions near 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 [19]. A previous study [20] 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 [21]. Both previous PAR-CLIP experiments [22] and our comparative studies 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 [23]. 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 exonic splicing enhancers (ESEs) or exonic splicing silencers (ESSes) [24] and may also play a critical role in translational regulation [25–27].

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 3′ splice sites, while the previous iCLIP results [28] showed that a large fraction of TIA1 targets were located around 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 known evidences or studies about the functions of these RBPs.

S1.8   Sequences of three TDP43 targets described in the main text

CLIP34nt: GAGAGAGCGCGUGCAGAGACUUGGUGGUGCAUAA
CLIP6: UUGUGGUGUGCUUUGCAGGAGGACU
CLIP34nt_UG6: GAGAGAGCGCGUGUGUGUGUGUGGUGGUGCAUAA

S1.9 Sequences of both wild-type and mutant 3′ UTRs of mRNA ERBB2 described in the main text

WT-URE/WT-331b: GGGCGAAUUGGAGCUCCACCGCGGUGGCGGCCGCUCUAGAA-
GUGCUUUUCUGUUUAGUUUUUACUUUUUUUGUUUUGUUUUUUUAAAGAUGAAAUAA-
AGACCCAGGGGGGGCCCGGUAC

MT-URE/WT-331b: GGGCGAAUUGGAGCUCCACCGCGGUGGCGGCCGCUCUAGAA-
GUGCUUUUCUGUUUAGUUUUUACUGUGUGUGUGGUGUGUGUGUAAAGAUGAAAUAA-
AGACCCAGGGGGGGCCCGGUAC

WT-URE/MT-331b: GGGCGAAUUGGAGCUCCACCGCGGUGGCGGCCGCUCUAGAA-
GUGCUUUUCUGUUUAGUUUUUACUUUUUUUGUUUUGUUUUUUUAAAGAUGAAAUAA-
AGAGCGACGCGGGGCCCGGUAC

MT-URE/MT-331b: GGGCGAAUUGGAGCUCCACCGCGGUGGCGGCCGCUCUAGAA-
GUGCUUUUCUGUUUAGUUUUUACUGUGUGUGUGGUGUGUGUGUAAAGAUGAAAUAA-
AGAGCGACGCGGGGCCCGGUAC

S2 Supplementary Figures
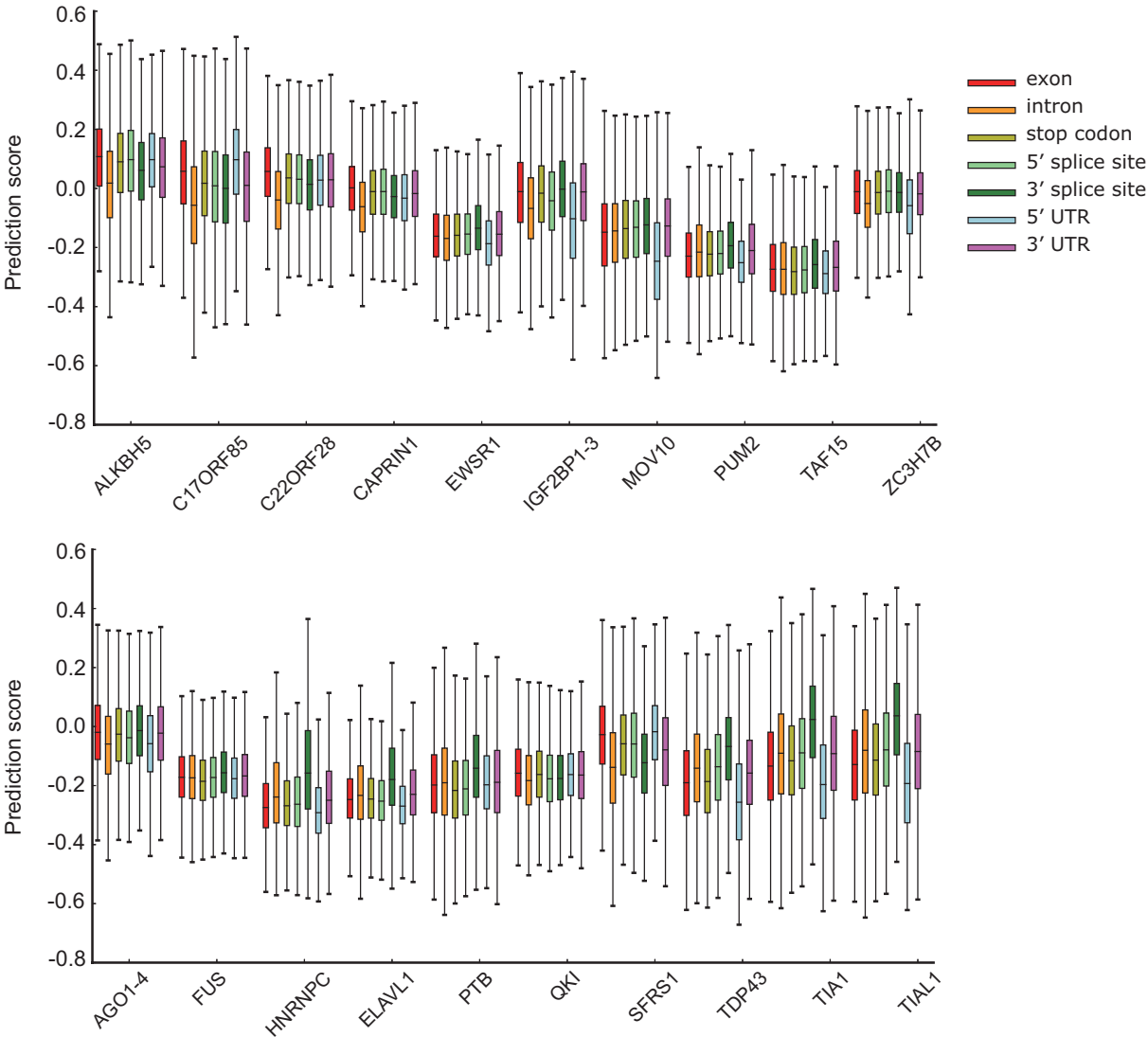
Figure S1: The distributions of the binding scores predicted by DeBooster in different characterized genomic regions for individual RBPs.
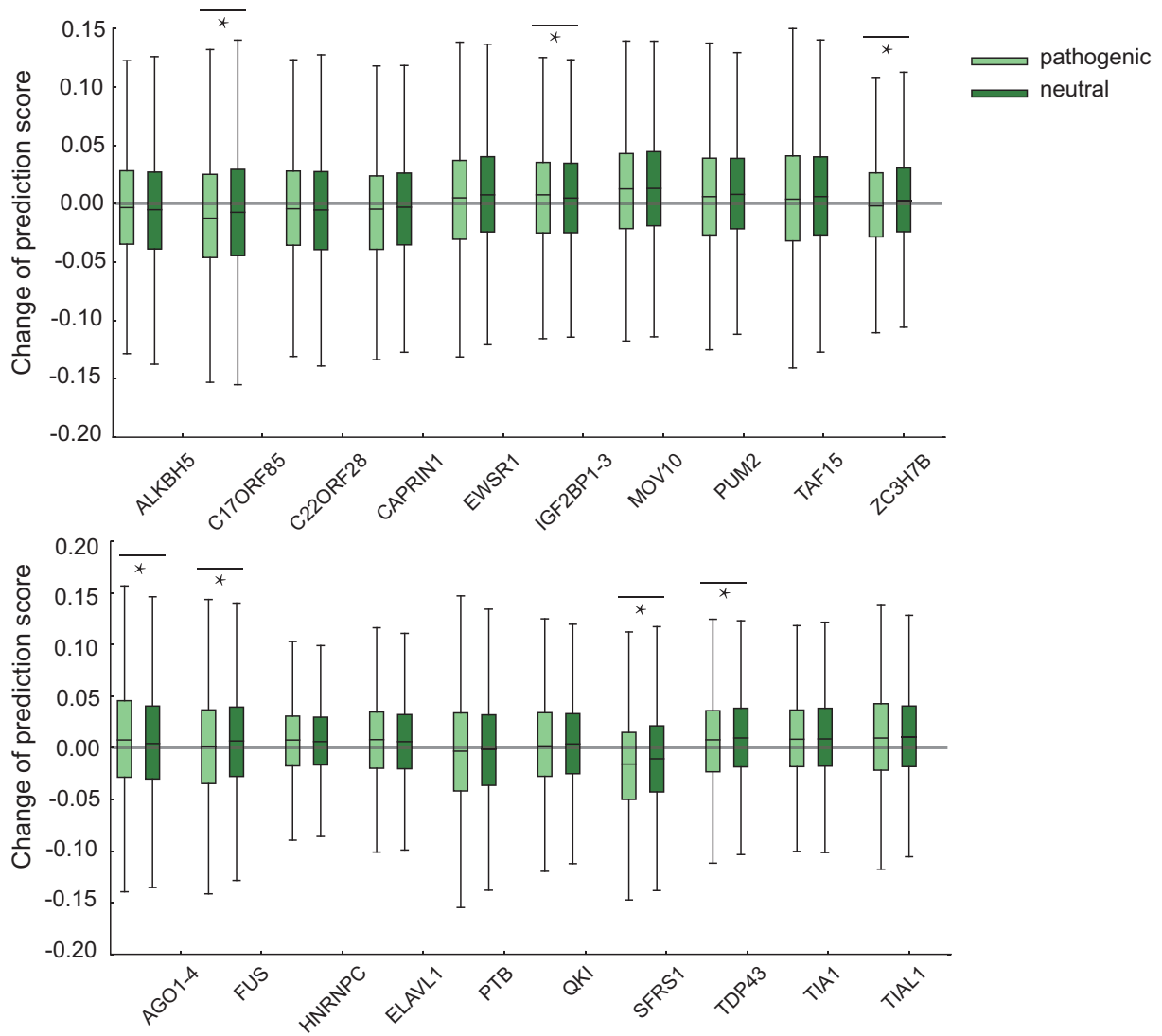
Figure S2: The changes of RBP binding scores predicted by DeBooster for the pathogenic and neutral mutations (10,000 each) that were randomly selected from the COSMIC records. *: p<0.001, Student's t test.
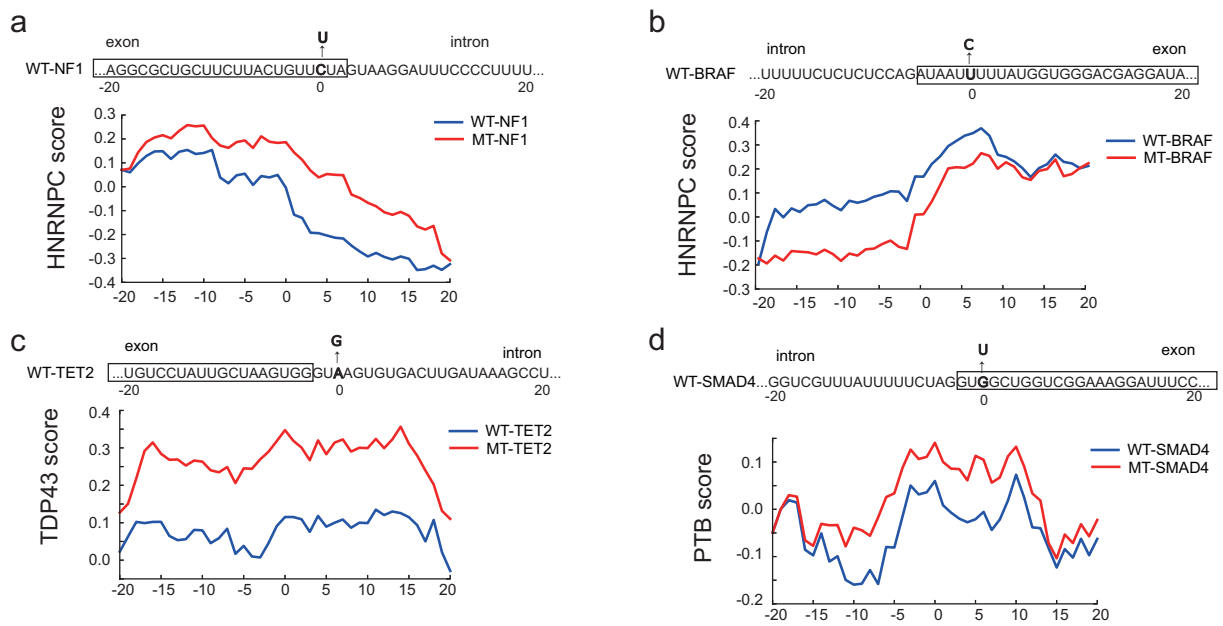
Figure S3: Additional examples on the predicted effects on the potentially disease-causing mutations near splice sites. **(a, b)** The exonic mutations of the HNRNPC binding targets near splice sites for genes *NF1* and *BRAF*, respectively. **(c)** The intronic mutation of the TDP43 binding targets near a splice site for gene *TET2*. **(d)** The exonic mutation of the PTB binding site near a splice site for gene *SMAD4*. Abbreviation: WT, wild-type; MT, mutant; URE, U-rich element.

Figure S4: Comparisons between the sequence motifs of several RBPs generated by both DeBooster and GraphProt. The known motifs from the literature are also shown for comparisons

RBP	DeBooster motif	GraphProt motif	RBP	DeBooster motif	GraphProt motif
ALKBH5			ELAVL1(B)		
C17ORF85			ELAVL1(A)		
C22ORF28			EWSR1		
CAPRIN1			FUS		
AGO2			ELAVL1(C)		
ELAVL1			IGF2BP1-3		
SFRS1			MOV10		
HNRNPC			PUM2		
TDP43			QKI		
TIA1			TAF15		
TIAL1			PTB		
AGO1-4			ZC3H7B		

Figure S5: Motifs generated by both DeBooster and GraphProt for all 24 training datasets.

11
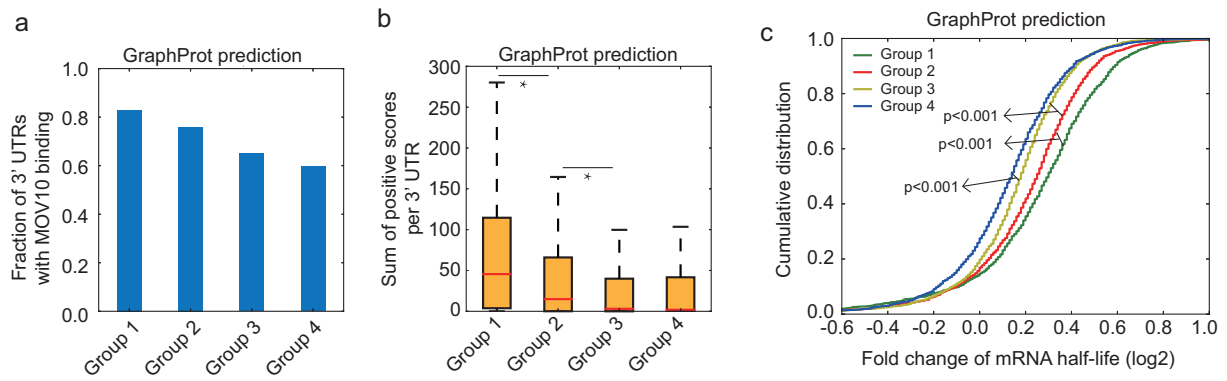
Figure S6: Performance of GraphProt on predicting the binding effects of MOV10 on mRNA degradation. (a) Fractions of 3′ UTRs with the MOV10 binding sites predicted by GraphProt in the four groups, in which UTRs were evenly separated according to the fold changes of their mRNA half-lives after MOV10 knock-down. Groups 1, 2, 3 and 4 corresponded to top 25%, 25%-50%, 50%-75% and bottom 25% fold changes, respectively. (b) The sum of the positive GraphProt margins per UTR for the four groups, which were classified in the same manner as in (a). *: p value < 0.001, Wilcoxon rank sum test. (c) The cumulative distribution on the fold changes of mRNA half-lives for four groups of UTRs, classified and ranked according to the GraphProt prediction margins in a descending order. That is, Groups 1, 2, 3 and 4 corresponded to the UTRs with top 25%, 25%-50%, 50%-75% and bottom 25% of the GraphProt predicted margins, respectively. The p values were computed using the Wilcoxon rank sum test.
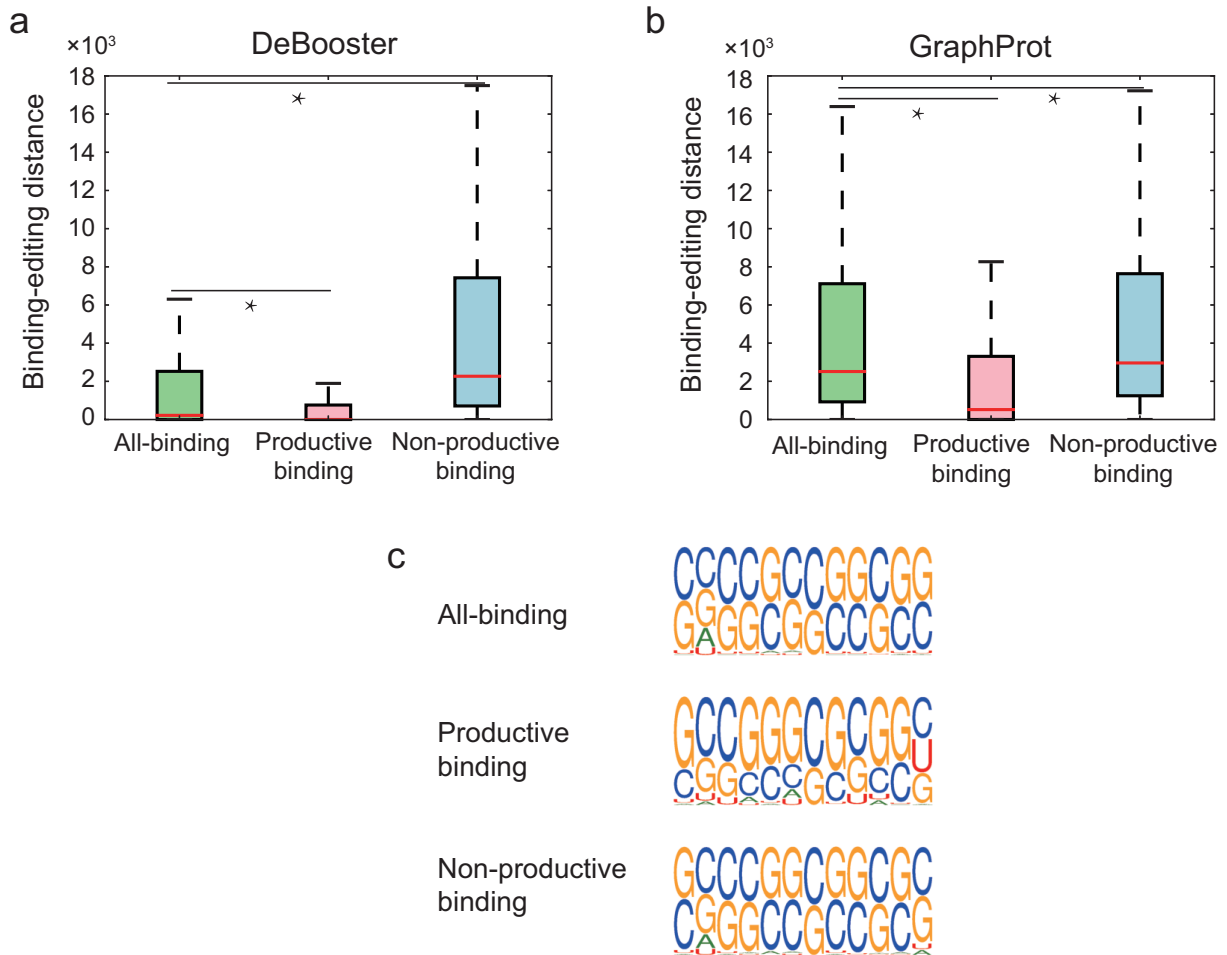
Figure S7: The results of DeBooster and GraphProt models on chromosome17, trained on three different training datasets, which represented all CLIP-seq measured ADAR1 binding sites, productive ADAR1 binding sites (i.e, triggering A-to-I editing), and non-productive ADAR1 binding sites (i.e, without triggering A-to-I editing), respectively. (a,b) The boxplots of the binding-editing distances, which were defined as the genomic distances between the ADAR1 binding sites and the closet editing sites, for three different DeBooster and GraphProt models, respectively. *: p value < 0.001, Wilcoxon rank sum test. (c) The sequence motifs generated by the three different GraphProt models.
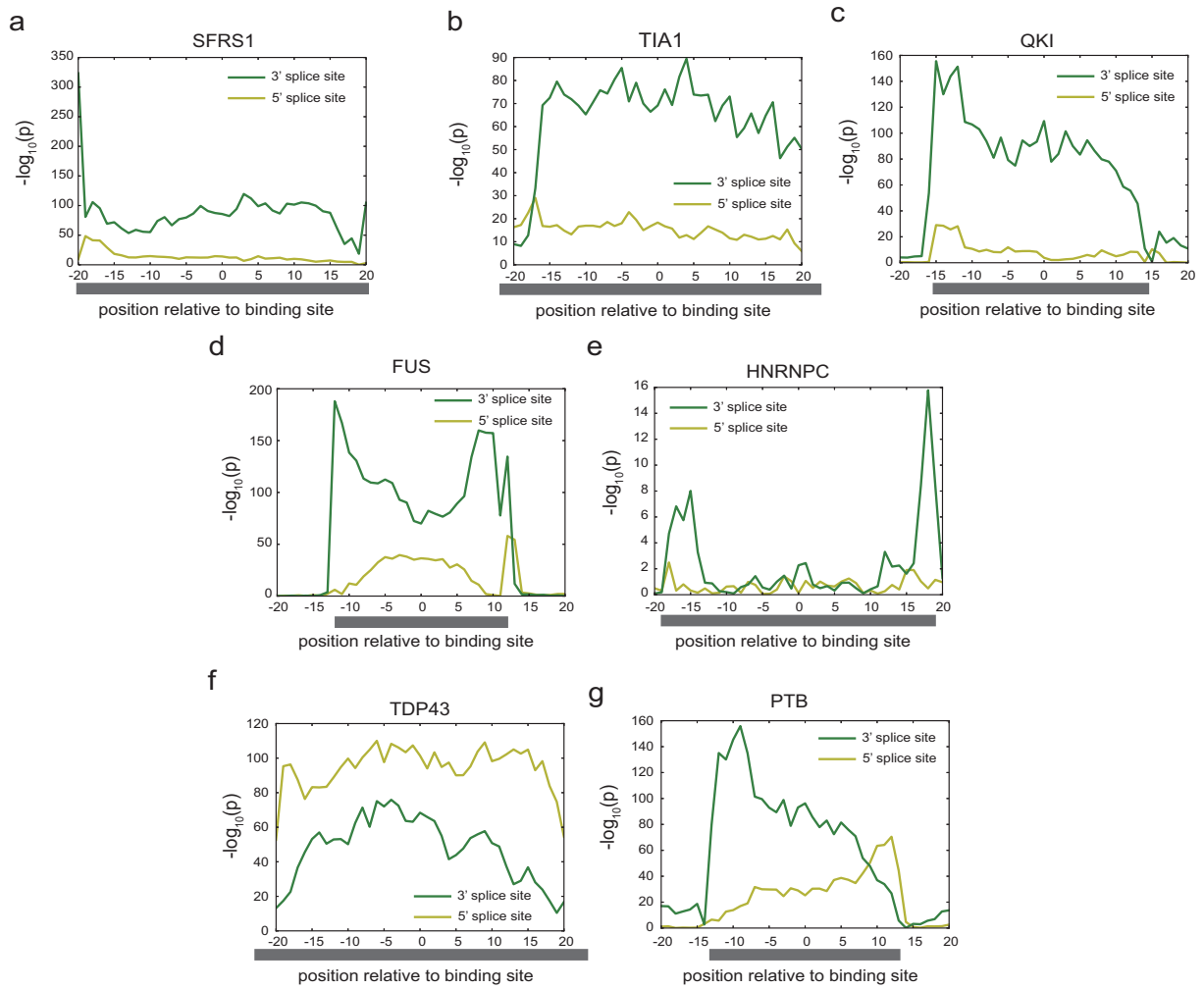
Figure S8: The p values along the 40 nt window around the centers of RBP binding sites, calculated using the changes of prediction scores between pathogenic and neutral mutations near the 5′ or 3′ splice sites. The black bars under the subfigures represent the locations occupied by individual RBPs. (a-g) correspond to the results of SFRS1, TIA1, QKI, FUS, HNRNPC and PTB, respectively. The Student's test was used to compute the p values.

Supplementary References

[1] Daniel Maticzka, Sita J Lange, Fabrizio Costa, and Rolf Backofen. GraphProt: modeling binding preferences of RNA-binding proteins. *Genome Biology*, 15(1):1–18, 2014.

[2] Roland Tacke, Yan Chen, and James L Manley. Sequence-specific rna binding by an SR protein requires RS domain phosphorylation: Creation of an SRp40-specific splicing enhancer. *Proceedings of the National Academy of Sciences of the United States of America*, 94(4):1148–1153, 1997.

[3] Fenbiao Gao, C C Carson, Todd Levine, and Jack D Keene. Selection of a subset of mrnas from combinatorial 3' untranslated region libraries using neuronal rna-binding protein Hel-N1. *Proceedings of the National Academy of Sciences of the United States of America*, 91(23):11207–11211, 1994.

[4] Ismael Perez, C H Lin, J G Mcafee, and James G Patton. Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. *RNA*, 3(7):764–778, 1997.

[5] Claudia Colombrita, Elisa Onesto, Francesca Megiorni, Antonio Pizzuti, Francisco E Baralle, Emanuele Buratti, Vincenzo Silani, and Antonia Ratti. TDP-43 and FUS rna-binding proteins bind distinct sets of cytoplasmic messenger rnas and differently regulate their post-transcriptional fate in motoneuron-like cells. *Journal of Biological Chemistry*, 287(19):15635–15647, 2012.

[6] 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.

[7] Jessica I Hoell, Erik Larsson, Simon Runge, Jeffrey D Nusbaum, Sujitha Duggimpudi, Thalia A Farazi, Markus Hafner, Arndt Borkhardt, C Sander, and Thomas Tuschl. RNA targets of wild-type and mutant FET family proteins. *Nature Structural & Molecular Biology*, 18(12):1428–1431, 2011.

[8] Julian Konig, Kathi Zarnack, Gregor Rot, Toma Curk, Melis Kayikci, Bla Zupan, Daniel J Turner, Nicholas M Luscombe, and Jernej Ule. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Structural & Molecular Biology*, 17(7):909–915, 2010.

[9] Laura M Dember, Nancy D Kim, Karenqianye Liu, and Paul Anderson. Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities. *Journal of Biological Chemistry*, 271(5):2783–2788, 1996.

[10] Kim D Pruitt, Susan M Hiatt, Francoise Thibaudnissen, Alexander Astashyn, Olga Ermolaeva, Catherine M Farrell, Jennifer Hart, Melissa J Landrum, Kelly M Mcgarvey, Nuala A O Leary, et al. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Research*, 42, 2014.

[11] Brian J Raney, Timothy R Dreszer, Galt P Barber, Hiram Clawson, Pauline A Fujita, Ngan Nguyen, Benedict Paten, Ann S Zweig, Donna Karolchik, and W James Kent. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC genome browser. *Bioinformatics*, 30(7):1003–1005, 2013.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

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

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.