Clusters of deep intronic RbFox motifs embedded in large assembly of splicing regulators sequences regulate alternative splicing

Francesco Tomassoni-Ardori[1], Mary Ellen Palko[1], Melissa Galloux[2], and Lino Tessarollo[1*].

[1] Neural Development Section, Mouse Cancer Genetics Program, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD, 21702, USA;
[2]Independent bioinformatician, Marseille, France

*Corresponding author:

Lino Tessarollo, Ph.D.
Mouse Cancer Genetics Program, CCR
NCI-Frederick, 7th Street,
Bldg. 560,
Frederick, MD, 21702, USA
Phone: 301-846-1202
Fax: 301-846-7017

Email: tessarol@mail.nih.gov

Short title: Clusters of deep intronic RbFox motifs regulate alternative splicing

Abstract

The RbFox RNA binding proteins regulate alternative splicing of genes governing mammalian development and organ function. They bind to the RNA sequence (U)GCAUG with high affinity but also non-canonical secondary motifs in a concentration dependent manner. However, the hierarchical requirement of RbFox motifs, which are widespread in the genome, is still unclear. Here we show that deep intronic, tightly clustered RbFox1 motifs cooperate and are important regulators of alternative exons splicing. Bioinformatic analysis revealed that (U)GCAUG-clusters are widely present in both mouse and human genes and are embedded in sequences binding the large assembly of splicing regulators (LASR). Integrative data analysis from eCLIP and RNAseq experiments showed a global increase in RNA isoform modulation of genes with Rbfox1 eCLIP-peaks associated with these clusters. Experimentally, by employing recombineering mutagenesis in a bacterial artificial chromosome containing the NTrk2 mouse region subjected to alternative splicing we showed that tightly clustered (U)GCAUG motifs in the middle of 50 Kb introns are necessary for RbFox1 regulation of NTrk2 gene isoforms expression. Moreover, clustered (U)GCAUG-motifs promote the recruitment of RbFox1 proteins to form a Rbfox1/LASR complex required for splicing. These data suggest that clustered, distal intronic Rbfox-binding motifs embedded in LASR binding sequences are important determinants of RbFox1 function in the mammalian genome and provide a target for identification of pathogenic mutations.

Key words: RbFox1, RbFox2, RbFox3, A2bp1; A2BP1, Ntrk2, BAC recombineering, RNA Binding Proteins, RBP, introns, alternative splicing

Introduction

In eukaryotes, alternative splicing of precursor mRNAs (pre-mRNAs) is a critical process involving the removal of introns and the inclusion or skipping of specific exons. In humans, more than 90% of genes are estimated to be alternatively spliced, contributing to the generation of gene-isoforms and, consequently, genome complexity. Alternatively spliced exons are flanked by longer introns compared with those flanking constitutively spliced exons suggesting that intron length may harbor increased genetic content for regulation of alternative exon choice (Conboy, 2021; Hollander et al., 2016). However, deciphering how the content of long introns influences splicing has been difficult because many RNA-binding proteins (RBPs) can act directly or indirectly to modulate pre-mRNA splicing processes by binding to recognition motifs depending on affinity and context, and because specific sequences may lead to mRNA looping which also affects splicing (Gehring and Roignant, 2021; Ule and Blencowe, 2019). It has also been reported that alternatively spliced introns, which are usually long, are removed last by the spliceosome although the regulatory mechanism is poorly defined (Choquet et al., 2023). Possibly, some delay may be caused by a sequential removal of small intron chunks through a stochastic recursive splicing process that is unique to longer introns (>10 Kb), whereas shorter introns (<1.5 kb) are excised in one step as complete units (Wan et al., 2021). Nevertheless, experimentally addressing alternative splicing involving long introns has been challenging because of the lack of suitable in vitro and in vivo systems to perform such studies. *Ntrk2* belongs to the class of genes with large introns and generating different isoforms by alternative splicing. It spans about 350 Kb of the genome and by alternative splicing generates a full-length tyrosine kinase receptor (TrkB.FL) and a truncated isoform lacking kinase activity (TrkB.T1). Importantly, it includes long introns (about 50 Kb) surrounding the alternative spliced exons

(Kumanogoh et al., 2008) and the size of these introns suggests significant coding information for precise spatio-temporal expression of the TrkB isoforms in response to the many developmental and environmental cues (Zhang and Huang, 2006). Almost nothing is known about the regulatory mechanisms of TrkB isoforms expression, but proper expression of these receptors is crucial for normal development and function of the mammalian brain as their dysregulation leads to neurodevelopmental and psychiatric disorders (Tessarollo and Yanpallewar, 2022). We have reported that the RBP RbFox1 regulates the expression levels of receptor isoforms generated by the TrkB locus through a still unknown mechanism (Tomassoni-Ardori et al., 2019). Gene targeting experiments in mouse and analysis of Rbfox1 copy number variations in human has shown that Rbfox1 is a critical player in neuronal excitability and, like TrkB, has been associated with multiple psychiatric disorders and complex pathologies of the central nervous system (CNS) including epilepsy, autism spectrum disorders, Alzheimer and Parkinson diseases (Gehman et al., 2011; Lal et al., 2013; Lee et al., 2016; Lin et al., 2016; O'Leary et al., 2022; Raghavan et al., 2020; Wen et al., 2015; Weyn-Vanhentenryck et al., 2014). In mammals, Rbfox1 belongs to a family of RNA-binding proteins (RBPs) involved in alternative splicing regulation which includes three members: Rbfox1, Rbfox2, and Rbfox3 (also known as NeuN, widely used as a neuronal marker). The expression of Rbfox1 is restricted to neurons, heart, and skeletal muscle where it participates in the regulation of pre-mRNA splicing in the nucleus by binding predominantly to intronic regions (Conboy, 2017; Kuroyanagi, 2009). Indeed, CLIP experiments in cell lines and in brain have shown Rbfox1 is associated mostly with intronic sequences in the nucleus (Damianov et al., 2016; Lovci et al., 2013; Weyn-Vanhentenryck et al., 2014). Moreover, Rbfox1-isoforms preferentially expressed in the cytoplasm can also regulate mRNA stability and expression of target transcripts by binding to

their 3'UTR sequence in the cytoplasmic compartment (Lee et al., 2016). All the Rbfox family members share an almost identical and highly conserved RNA recognition motif (RRM) capable of binding with high affinity to the penta/hexa-ribonucleotide (U)GCAUG (Jin et al., 2003; Kuroyanagi, 2009; Nakahata and Kawamoto, 2005). However, RbFox proteins can also bind to secondary motifs *in vitro* and *in vivo* enabling RbFox concentration-dependent regulation of exon inclusion in neuronal differentiation and diversification (Begg et al., 2020).

To investigate how RbFox1 regulates inclusion or exclusion of the TrkB.T1 encoding exon surrounded by 50 Kb introns, we exploited the Bacterial Artificial Chromosome (BAC) manipulation technology to generate a ~165 Kb *Ntrk2* minigene including the intact intron/exon genomic structure of the TrkB locus (TrkB-BAC minigene) undergoing alternative splicing (Liu et al., 2003). By enhanced RNA Cross-Linking Immuno-Precipitation (eCLIP) experiments and BAC recombineering mutagenesis in cells containing mutant BACs we found that Rbfox1 modulates the expression of the TrkB.FL and TrkB.T1 isoforms by binding to tightly clustered RbFox binding sites located up to 25 Kb from the alternative spliced exon. Moreover, we show that these clustered penta/hexa-nucleotides (T)GCAUG are widely present in distal intronic regions of the mouse and human genome, coinciding with eCLIP peaks and positively correlating with RbFox1 regulation of expression of alternatively spliced genes. Our data suggest that clustered Rbfox motifs located in far intronic regions are a general feature of the mammalian genome and are important regulators of gene-isoform expression. In addition, the Bacterial Artificial Chromosome (BAC) manipulation technology provides a new tool to study mechanisms regulating large introns splicing.

Results

A Bacterial Artificial Chromosome (BAC) system to study deep intronic regions mediating Rbfox1 alternative splicing function.

We have previously reported that RbFox1 regulates the expression of the receptor isoforms encoded by the *Ntrk2* gene (TrkB) (Tomassoni-Ardori et al., 2019). This gene spans about 350 Kb across the mouse genome, generates two major receptor isoforms and includes large introns of about 50 Kb in the region where alternative splicing occurs. TrkB isoforms have identical extracellular and transmembrane domain and either a short intracellular region (TrkB.T1) or a catalytic tyrosine kinase domain (TrkB.FL) (Fig. 1A). RbFox1 regulation of TrkB isoform expression is not mediated by the isoforms unique 3'UTR, as reported for other genes (Lee et al., 2016; Tomassoni-Ardori et al., 2019). Therefore, to investigate whether other intronic sequences regulate alternative splicing, using Bacterial Artificial Chromosome (BAC) manipulation technology (Liu et al., 2003; Sharan et al., 2009) we modified a *Ntrk2* BAC containing the genomic region subjected to alternative splicing and including the extracellular juxtamembrane and the transmembrane coding exons, the TrkB.T1-specific exon and the first two exons encoding the tyrosine kinase domain (Fig. 1A). The BAC is about 164 Kb and includes the two large 50 Kb introns surrounding the alternatively spliced TrkB.T1 exon (Fig. 1A). A cDNA fragment encoding the missing part of the extracellular domain preceded by a synthetic CAG promoter (Alexopoulou et al., 2008) was fused in frame to the upstream extracellular juxtamembrane exon of the BAC; downstream, the missing tyrosine kinase domain coding cDNA was fused in frame to the second tyrosine kinase exon of the BAC (Fig. 1A) (Kumanogoh et al., 2008). These additions allowed for expression of the complete TrkB.T1 and TrkB.FL

receptor isoforms (TrkB-BAC minigene). Importantly, it preserved the intact, endogenous genomic structure of a large portion of the gene where alternative splicing occurs.

Next, we introduced the TrkB-minigene into T-REx-293 cells, a human HEK293 cell line engineered to express a specific gene, in our case RbFox1, under the control of the tetracycline inducible system by Doxycycline (Dox) (Tomassoni-Ardori et al., 2019; Yao et al., 1998). Importantly, T-REx-293 cells do not express TrkB receptors endogenously greatly facilitating the interpretation of the expression data obtained from the minigene in the absence (no Dox) or presence (+Dox) of Rbfox1. Two independent T-REx-293;Rb-Fox1clones with the TrkB minigene confirmed expression of both TrkB.T1 and TrkB.FL. Strikingly, in both clones, Rbfox1 expression changed the expression levels of the TrkB isoforms. Specifically, Rbfox1 significantly increased TrkB.FL isoform at both protein and mRNA levels while TrkB.T1 decreased, although to a lesser degree (Fig. 1 B-C). These data suggest that the TrkB-minigene contains elements mediating Rbfox1 function on TrkB isoform alternative splicing.

To investigate whether the RNA-binding function of Rbfox1 is required to regulate alternative splicing of the TrkB minigene, we tested a mutant Rbfox1 with impaired RNA-binding activity (F158A) (Hakim et al., 2010; Jin et al., 2003; Tomassoni-Ardori et al., 2019). As shown in Fig. 1D-F, transfection of the Rbfox1-F158A into cells with the minigene failed to modulate the expression of TrkB isoforms both at the mRNA and the protein level, suggesting that a functional RNA-binding domain is essential for Rbfox1 activity.

Tightly clustered RbFox1 binding sequences in deep intronic regions of the TrkB minigene overlap with eCLIP peaks.

The requirement of a functional RbFox1 RNA-binding domain to regulate TrkB isoform expression suggests direct binding to TrkB mRNA. Therefore, to identify RbFox1 binding sites in the TrkB pre-mRNA expressed by the minigene we performed Enhanced Cross-Linking Immuno-Precipitation (eCLIP) experiments in the presence or absence of Rbfox1 (Van Nostrand et al., 2016). Immuno-precipitation with a RbFox1-specific monoclonal antibody and RNA sequencing identified seven statistically significant eCLIP peaks across the minigene when RbFox1 was expressed. Interestingly, all peaks were located in distal intronic regions and outside the TrkB.T1 3'UTR confirming our previous result that RbFox1 does not regulate TrkB splicing through the 3'UTR ((Tomassoni-Ardori et al., 2019); Fig. 2; Suppl. Fig 1). Curiously, two of the seven eCLIP peaks mapped on sequences about 300-nucleotides long characterized by the presence of multiple closely clustered (T)GCATG-motifs (Fig. 2 B, C). These clusters, defined as 'Cluster 1' and 'Cluster 2', are in the middle of the two large 50 Kb introns flanking the TrkB.T1-coding exon and display, respectively, nine (T)GCATG-motifs, for the intron upstream, and seven, for the intron downstream (Fig. 2B, C).

Clusters of Rbfox1 binding sites are widespread in the mouse and human genome.

The finding that eCLIP peaks correspond to (T)GCATG-clusters and may impact on the regulation of TrkB isoforms expression levels prompted us to investigate whether the presence of (T)GCATG-clusters is widespread in the genome and is part of the RbFox1 regulatory function on gene expression. To screen the mouse genome, we first defined 'clusters of motifs' as genomic sites with at least four or more (T)GCATG elements confined within 500 nucleotides (Fig. 3A). Overlapping clusters were considered as one, large, extended cluster.

Bioinformatic analysis revealed that the mouse genome has about twenty-one thousand (T)GCATG-clusters present in about 8,300 genes and most of the clusters are located in intronic regions (Fig. 3B, C) (Suppl. Table 1). Analysis of Rb-Fox1 High-Throughput Sequencing of RNA isolated by Crosslinking Immunoprecipitation (HITS-CLIP) data from the mouse brain showed that about 8300 genes have HITS-CLIP peaks and about half of them (4186) have (T)GCATG-clusters (Fig. 3C) (Weyn-Vanhentenryck et al., 2014). Importantly, 20% of genes with HITS-CLIP peaks, have peaks mapping to the (T)GCATG-clusters (1687 out of 8356) (Fig. 3D). These data suggest that (T)GCATG-clusters could be significant determinants of Rbfox1 binding to mouse pre-mRNA transcripts.

We next tested whether (T)GCATG-clusters are also present in the human genome by using the same strategy used to scan the mouse genome. Curiously, we found that the human genome has about the same number of (T)GCATG-clusters as found in the mouse genome (19977 versus 21005) although 15% more genes in humans contain clusters (9826 in human versus 8348 in mouse) (Fig. 3E, F; Suppl. Table 2). Since CLIP data is not available from human brain, we used the eCLIP data obtained from the T-REx-293;RbFox1 human cells containing the TrkB minigene to investigate the relationship between Rbfox1 binding (eCLIP peaks) to (T)GCATG-clusters. This analysis confirmed that the 'canonical' Rbfox1 motif (T)GCATG is the most enriched motif among all the eCLIP peaks and that, as in mouse, the distribution of RbFox1 peaks is mostly present in intronic regions (Lee et al., 2016; Lovci et al., 2013; Weyn-Vanhentenryck et al., 2014) (Suppl. Fig. 2). Although HEK293 are non-neuronal cells, we found 4903 genes containing Rbfox1 eCLIP peaks, of which 2908 (~60%) included (T)GCATG-clusters (Fig. 3F). Importantly, 1237 of the 4903 genes with eCLIP peaks, had eCLIP peaks mapping to (T)GCATG-clusters, a percentage similar to that observed in mouse (25% in HEK293 cells

versus 20% in mouse brain) (Fig. 3G). Altogether, these data show that Rbfox1 directly binds (T)GCATG-clusters of both human and mouse transcripts.

Moreover, since Rbfox1 is expressed predominantly in brain, heart and skeletal muscles, we analyzed (T)GCATG-clusters presence in all genes associated with diseases in these tissues by exploiting the catalog of human genome-wide association studies (GWAS Catalog). Supplementary Table 1 lists the genetic coordinates for all the (T)GCATG-clusters mapping on genes associated with muscle disorders, brain and heart diseases. Overall, we found that about 40% of the genes involved in these pathologies have (T)GCATG-clusters including brain disease 36.5%, muscle disorders 40.8% and heart disease 39.5%.

The presence of Rbfox1 binding site clusters in human alternatively spliced genes correlates with enhanced activity of Rbfox1 on gene isoform expression.

To investigate genome wide whether eCLIP peaks on (T)GCATG-clusters are significant in regulating RbFox1activity on gene isoform expression we performed RNAseq isoform expression analysis in HEK293-RbFox1 cells in the absence (No Dox) or presence (+Dox) of Rbfox1, as in the eCLIP experiments. Cross-analysis between eCLIP and RNAseq data showed that, among the genes with eCLIP peaks, there is a higher percentage of genes with differential expression of isoforms when eCLIP peaks are present on the (T)GCATG-clusters compared to the group of genes with peaks outside the clusters or with peaks but lacking (T)GCATG-clusters (Fig. 3H-I). Importantly, this difference is maintained irrespective of the level of fold change in isoform expression up to a threshold of 100% fold-change (Fig. 3H). Moreover, the cumulative average of $\log_2$ fold change in expression was also significantly higher for gene isoforms with

eCLIP peaks on clusters (1.81±0.063 Log_2FC) compared to those with eCLIP outside of the clusters or with eCLIP peaks but lacking the (T)GCATG-clusters (1.56±0.055 and 1.39±0.049 Log_2FC respectively; Fig. 3I). These data suggest that the presence of (T)GCATG-clusters in the genome is an important determinant of RbFox1 activity on gene isoform expression.

Deletion of deep intronic (T)GCATG-clusters eliminates Rbfox1 splicing function on TrkB.

To directly test the role of the (T)GCATG-clusters in mediating Rbfox1 function on TrkB isoforms expression we used a recombineering deletion approach in the TrkB-BAC minigene (Liu et al., 2003). Deletion of the clusters was achieved by removing, respectively, only about 600 and 400 nucleotides from each of the 50 Kb introns upstream and downstream of the TrkB.T1 exon (Suppl. Fig 3). Importantly these short sequences were, respectively, about 25 and 15 Kb upstream and downstream of the TrkB.T1 exon. Removal of each individual cluster independently did not significantly impair Rbfox1 function on TrkB isoform expression when compared to the control BAC (Fig. 1C; Fig. 4A-F). Specifically, the TrkB.FL isoform is again strongly upregulated by RbFox1 both at the RNA and protein level in both minigenes with either Cluster 1 or Cluster 2 deleted. The effect on TrkB.T1 levels was less obvious since deleting Cluster 2 did not change its levels compared to the wild-type BAC, while Cluster 1 deletion appeared to render TrkB.T1 levels insensitive to RbFox1 (Fig. 4A-F). Most importantly, deletion of both clusters together completely abolished any effect of Rbfox1 (Fig. 4 G-I) on the expression levels of both TrkB.FL and TrkB.T1. This result suggest that Rbfox1 regulates gene isoform expression levels by interacting with 'deep' intronic, tightly clustered RNA binding motifs in the immature RNA transcripts.

Lastly, to test whether deletion of the two clusters changed the Rbfox1 binding landscape on the BAC, we performed eCLIP experiments in one of the cell lines harboring the deletion of both Cluster 1 and Cluster 2. Interestingly, comparison between the eCLIP peaks of the WT and the mutant minigene showed that only the peaks corresponding to Cluster1 and 2 are missing in the mutant while all other eCLIP peaks were unaltered (Suppl. Fig. 4). These data suggest that removal of the two clusters does not affect RbFox1 binding to the other existing mRNA sites and does not generate new alternative eCLIP peaks.

(T)GCATG-clusters are embedded in sequences binding the large assembly of splicing regulators (LASR).

Analysis of the RbFox1 cluster sites in the TrkB minigene showed abundant repetitive elements, including poly-GT sequences (Fig. 2C), which have been found to be present in the proximity of CLIP-peaks of Rbfox proteins (Damianov et al., 2016). Because these sequences are associated with the large assembly of splicing regulators (LASR), a protein complex containing hnRNP M, hnRNP H/F, hnRNP C, Matrin3, hnRNPUL2, NF110/NFAR-2, NF45, and DDX5 (Damianov et al., 2016; Peyda et al., 2025; Ying et al., 2017), we tested for enrichment of short repetitive sequences within the (T)GCATG-clusters in both mouse and human (Fig. 5A, C and B, D respectively). By analyzing all Rbfox clusters with at least five or more (T)GCATG-motifs, we found that motifs for the binding of hnRNP-M (GU-rich pentamers) and hnRNP-H (polyG-rich pentamers), two Rbfox1 interacting partners and members of the LASR, are significantly enriched within the (T)GCATG-clusters when compared to random control sequences with an equal length to the median of the Rbfox-clusters. On the contrary, polyU-rich pentamers binding

hnRNP-C, have reduced distribution within the (T)GCATG-clusters as compared to random control sequences (Fig. 5A, B). A similar analysis was also conducted on 'TG/CA' sequences that are often found in repetitive patterns within introns of genes and can act as splicing regulatory elements (Gabellini, 2001; Hui et al., 2003; Sharma et al., 2005). By analyzing short tandem repeats $(TG/CA)_6$ including a total of 64 different sequences derived from all the possible $(TG/CA)_6$ combinations we unveiled a significant enrichment of $(TG/CA)_6$ repeats within the Rbfox-clusters, suggesting that clustered (T)GCATG-motifs are embedded in sequences rich in short repeats binding interacting partners involved in splicing regulation (Fig. 5C, D).

Rbfox1 proteins also interact directly with LASR through their C-terminal domain (CTD). The CTD contains a low-complexity sequence rich in tyrosines which are required for assembly of Rbfox/LASR into higher-order complexes and promote RbFox1 ability to activate splicing (Ying et al., 2017). Mutation of the tyrosine residues in the CTD region impairs RbFox1 splicing activity even if it retains the ability of interacting with LASR (Ying et al., 2017). We hypothesized that the presence of (T)GCATG-clusters could favor the recruitment of Rbfox1 and the formation of higher-order assembly with LASR by locally increasing the concentration of Rbfox1-proteins. To test this hypothesis, we investigated whether a splicing-defective mutant Rbfox1 (Rbfox1-Y10) containing 10 tyrosine to serine residue mutations in the CTD could retain some splicing activity on the TrkB minigene transcripts (Ying et al., 2017). We argued that the presence of (T)GCATG-clusters is sufficient to increase the local concentration of RbFox1 proteins in intronic regions to allow, at least partially, the formation of functional higher-order assembly of Rbfox/LASR. Remarkably, we found that the Rbfox1-Y10 mutant retains splicing activity in the TrkB-BAC minigene system, although at reduced levels compared to the wild-type

Rbfox1 control (Fig. 5E, F). As expected, the Rbfox1-Y10 mutant did not have any effect on transcripts splicing generated from the TrkB BAC cluster mutant (Suppl. Fig 5). Altogether, these data suggest that the RbFox1 CTD domain requirement to mediate higher-order assembly of Rbfox/LASR for splicing activation by Rbfox is not absolute. Instead, the Rbfox/LASR splicing function also depends on deep intronic clustered (U)GCAUG-motifs which promote local recruitment of multiple RbFox1 proteins to form Rbfox1/LASR higher-order assemblies required for splicing (Fig. 6).

Discussion

Alternative splicing of pre-mRNA is a fundamental genetic process expanding gene function diversity. Several mechanisms regulate the splicing networks during normal differentiation and development including, for example, spatio-temporal expression of specific RBP, regulation of the stoichiometry of different proteins part of the spliceosome, the distribution of RBP motifs present in the pre-mRNA, and RNA sequences that form secondary structures bridging across individual introns to include or skip exons. The size of introns involved in alternative splicing can vary greatly from a few base pairs to hundreds of Kb. However, the biological significance of this variability is unknown. Likewise, it is also unclear how deep intronic regions can influence alternative splicing. Here, by studying the role of RbFox1 in regulating alternative splicing of the *Ntrk2* gene we have found that deep intronic, tightly clustered RbFox1 binding motifs are required for the normal expression of Ntrk2 receptor isoforms. Importantly, similar motifs are widespread in the mouse and human genome and when present in the context of sequences binding the LASR, mediate RbFox1 regulation of gene expression isoforms.

By using bacterial genetic engineering technology, we have generated an uncommonly large minigene (about 164 Kb) including native 50Kb introns and expressing both major isoforms of the *Ntrk2* gene. This system was used to show that RbFox1 regulates the expression of a TrkB alternatively spliced exon through clusters of RbFox binding sites as far away as 25 Kb upstream and downstream of the exon itself. The presence of such tightly clustered RbFox protein binding sites is not limited to the *Ntrk2* gene as they are widespread in both the mouse and human genome and more than 75% of them are located in deep intronic regions. One of the most important features of the TrkB-BAC minigene is the inclusion of the intact genomic structure encompassing two 50 Kb introns surrounding the TrkB.T1-coding exon subjected to alternative splicing. This system provides a tool to directly test mechanisms of alternative splicing regulation across large introns and helps in understanding their significance. Large introns are particularly abundant in genes regulating complex and diverse biological functions. For example, many genes expressed in brain, which require a precise spatio-temporal regulation of expression of different isoforms in response to developmental cues as well as environmental stimuli, have large introns. In the case of TrkB and its ligand BDNF, while BDNF is under the control of 8 different promoters, which provide great versatility in regulating its expression, TrkB, which has only one promoter, provides signaling diversity by generating spliced isoforms (Luberg et al., 2010). Thus, the BAC system provides a unique tool to understand splicing of large introns and the context of the sequences regulating them.

Combining eCLIP data and BAC mutagenesis, we found that deletion of the two clusters surrounding the TrkB.T1 encoding exon, although 50 Kb apart, completely disrupts RbFox1 activity on TrkB isoforms regulation and no changes in eCLIP sites were detected after deletion. This was surprising because there are about 120 additional RbFox binding sites distributed along

the introns surrounding the TrkB.T1-encoding exon. However, the clusters with eCLIP peaks are part of regions rich in elements binding other LASR, suggesting that RbFox binding sites are necessary but not sufficient for splicing (Ying et al., 2017). These data support a model in which multiple (U)GCAUG-motifs are required to initiate the formation of high density RbFox1/LASR complexes which drive splicing. Indeed, mutation of the tyrosines in the CTD domain which allows for RbFox1 physical aggregation and formation of higher-order assembly of Rbfox/LASR only partially abrogate the splicing phenotype in the TrkB-BAC. This suggests that clusters of RbFox binding sites are sufficient to initiate, at least in part, the catalysis leading to the formation of higher-order assembly of Rbfox/LASR (Damianov et al., 2016; Peyda et al., 2025). Moreover, the data further supports the notion that the context of the RbFox clusters is critical in determining RbFox1 activity. The finding that removal of only one cluster either upstream or downstream does not cause a significant phenotype, was somewhat puzzling. Intronic sequences responsible for RNA looping could be important determinants in bringing the RbFox1/LASR complex to the appropriate location for splicing (Lovci et al., 2013). Analysis of the 50 Kb sequence between the two clusters may help understand whether RNA bridges that allow for exclusion of the TrkB.T1 exon between the two clusters are possible and may be sufficient to lead to an upregulation of the TrkB.FL isoform (Lovci et al., 2013).

The TrkB-BAC minigene also provides a powerful tool for the functional screening of RbFox1 gene variants isolated from patients with developmental brain abnormalities as it uncovers mutations affecting RbFox1 splicing function (Li et al., 2023). Although its use has so far been limited to the functional testing of RbFox1 variants with mutations within the RNA binding domain, the finding that it can expose functional defects of RbFox1 with mutations in the tyrosine rich CTD domain suggests a use to test RbFox1 variants with mutations in other

domains that may be critical for its interaction with other RBP or proteins of the spliceosome complex (Fig. 5E, F, Suppl. Fig. 5).

In summary, we have identified clusters of RbFox1 binding sites embedded in repetitive sequences binding the large assembly of splicing regulators. These clusters are important determinants of RbFox1 splicing function and are widely distributed in the mammalian genome, particularly in very deep intronic regions of alternatively spliced genes. RbFox2 also regulate TrkB isoform splicing, at least in vitro, binding to the same (U)GCAUG sequence, suggesting that this TrkB-BAC minigene could be used to test the functional significance of RbFox protein variants isolated from patients with other pathologies (Tomassoni-Ardori et al., 2019; Verma et al., 2016). Moreover, the relevance of RbFox binding clusters in regulating gene isoform expression suggests that these clusters should be added to the list of investigated sequences in the genome for natural mutations in deep intron elements of gene causing brain, heart and skeletal muscle diseases in humans (Suppl. Table 1, (Li et al., 2023)).

Lastly, our study suggests that BAC-based minigenes up to 200 Kb with unaltered genomic structure could be used to identify and explore, in detail, the function of new elements in deep distal intron regions bound by other RBPs mediating splicing and/or distal interactions (Lunde et al., 2007; Van Nostrand et al., 2020a; Van Nostrand et al., 2020b; Zhang et al., 2013).

Methods

TrkB-BAC Minigene

The mouse BAC clone RP23-424E11 (Bacpac Genomics) containing 194 kb of the Ntrk2 (TrkB) genomic locus from exon 11 (98 bp; encoding the transmembrane region) to exon 14 (131 bp; corresponding to the second exon of the kinase domain), was modified using the BAC manipulation system as described in (Sharan et al., 2009) to create a functioning minigene that, by alternative splicing, expresses both the truncated (TrkB.T1) and full-length (TrkB.FL) transcript. Briefly, the synthetic CAG promoter and the TrkB extracellular domain cDNA sequence from the 'start' codon (ATG) to the sequence corresponding to mouse exon 10 (36 bp) was placed in frame with exon 11; the cDNA sequence encoding the TrkB kinase domain from the sequence corresponding to exon 15 (173 bp) to the 'stop' codon was placed in frame with exon 14 (Fig.1A), followed by a pGKneo cassette for selection with G418 Sulfate 300 ug/ml (ThermoFisher Scientific 10131035) to generate subsequent clonal cell lines. The BAC containing the TrkB minigene was transfected into a T-Rex 293 doxycycline-inducible Rbfox1 expressing cell line as previously described (Tomassoni-Ardori et al., 2019) using Targefect per manufacturer instructions (Targeting Systems, Targefect-BAC). Rbfox1 expression was induced by Doxycycline (D3447, Millipore-Sigma) 0.5 mg/ml.

Reagents for BAC recombineering technology are available at:

https://frederick.cancer.gov/resources/repositories/Brb/#/recombineeringInformation

Western Blot Analysis

Cells were lysed using RIPA lysis buffer (20-188, Millipore-Sigma) and incubated 30 min at 4°C before centrifugation at 13,000 rpm at 4°C. Supernatants were collected and transferred into new tubes. Protein concentrations was quantified using a BCA assay (23225, ThermoFisher Scientific) and samples were prepared with equal amounts of total protein before adding Laemmli sample buffer 2X (S3401, Sigma-Aldrich). Samples were heated at 95°C for 5 min (protein denaturation step) before loading onto a 4–12% NuPAGE precast gel for Western analysis (ThermoFisher Scientific). After the transfer to PVDF membranes (LC2005, ThermoFisher Scientific), blots were blocked in 5% non-fat milk in TBS-Tween (0.1%) and incubated overnight at 4°C with the specific antibodies. Primary antibodies were anti-panTrk C15 (used for detecting TrkB.FL - against the intracellular kinase-domain of Trk and therefore recognizing all Trk receptors; sc-139; Santa Cruz), anti-TrkB.T1 C13 (sc-119; Santa Cruz), anti-Rbfox1 (MABE985, clone 1D10, Millipore-Sigma), anti GAPDH (MAB374; Millipore). After incubation with the appropriate horseradish peroxidase (HRP)-conjugated secondary antibodies (Millipore), membranes were incubated with enhanced chemiluminescent substrate (34076, ThermoFisher Scientific) for detection of HRP enzyme activity and visualized with a Syngene gel documentation system (GeneSys).

To generate the mutant Rbfox1-Y10, the mouse Rbfox1 cDNA was modified in the C-terminal region converting 10 successive tyrosine residues to serine starting from tyrosine at position 299 and ending with tyrosine 341 as described by Ying and colleagues (Ying et al., 2017). WT and mutant Rbfox1-Y10 cDNA were cloned into the mammalian expression vector pcDNA3.1-neo for expression studies. Cells were transiently transfected using X-tremeGENE 9 DNA Transfection Reagent (6365779001, Millipore-Sigma) and screened by Western analysis 48h post-transfection.

QPCR Analysis

Total RNA was extracted from cells using the Qiagen RNeasy Mini kit (Cat.no 74104) according to the manufacturer's instruction. cDNA was then generated using SuperScript III First-Strand Synthesis System (Cat. No 18080–051, ThermoFisher Scientific). Real time PCR was performed using BioRad iTaq Universal SYBR-green Supermix (Cat.No. 172–5120) in a MX3000P (Agilent Technologies) apparatus using the following program: 95°C for 3 min; 95°C 10 s, 60°C 20 s for 40 cycles; 95°C 1 min and down to 55°C (gradient of 1°C) for 41 cycles (melting curve step). Delta Ct values were obtained using GAPDH as reference gene.

Student t-test was applied for statistical significance assessment.

Primers used:

TrkB common forward: 5'-AGCAATCGGGAGCATCTCT-3'

TrkB.FL reverse: 5'-CTGGCAGAGTCATCGTCGT-3'

TrkB.T1 reverse: 5'-TACCCATCCAGTGGGATCTT-3'

GAPDH forward: 5'-TGCGACTTCAACAGCAACTC-3'

GAPDH reverse: 5'-ATGTAGGCCATGAGGTCCAC-3'

Rbfox1 forward: 5'- TGGCCCCAGTTCACTTGTAT-3'

Rbfox1 reverse: 5'- GCAGCCCTGAAGGTGTTGTA-3'

Enhanced Cross-Linking Immuno-Precipitation (eCLIP)

eCLIP studies were performed by Eclipse Bioinnovations Inc (San Diego, www.eclipsebio.com) according to the published single-end seCLIP protocol (Van Nostrand et al., 2017) with the following modifications. Approximately 20 million FIp-In-T-Rex-293T cells were UV

crosslinked at 400 mJoules/cm2 with 254 nm radiation. The cell pellet was lysed using 1mL of eCLIP lysis mix and subjected to two rounds of sonication for 4 minutes, with 30 second ON/OFF at 75% amplitude. 1% of the lysate was treated with Proteinase K to digest the RNA bound proteins. The RNA was then isolated (Zymo), and measured on the TapeStation to assess quality and quantity. The lysate volume equating to 100ug of total RNA was enzymatically digested with a 1:10 dilution of RNase-1 (Ambion) and used as starting material. Validated antibodies were then pre-coupled to Anti-Rabbit IgG Dynabeads (ThermoFisher), added to the digested lysate, and incubated overnight at 4°C. Prior to immunoprecipitation, 2% of the sample was taken as the paired input sample, with the remainder magnetically separated and washed with eCLIP high stringency wash buffers. IP and input samples were cut from the membrane from approximately 75 kDa and above. RNA adapter ligation, IP-Western, reverse transcription, DNA adapter ligation, and PCR amplification were performed as previously described. The eCLIP cDNA adapter contains a sequence of 10 random nucleotides at the 5′ end. This random sequence serves as a unique molecular identifier (UMI) after sequencing primers are ligated to the 3′ end of cDNA molecules (Kivioja et al., 2011). Therefore, eCLIP reads begin with the UMI and, in the first step of the analysis, UMIs were pruned from read sequences using umi_tools (v0.5.1) (Smith et al., 2017). UMI sequences were saved by incorporating them into the read names in the FASTQ files to be utilized in subsequent analysis steps. Next, 3′-adapters were trimmed from reads using cutadapt (v2.7) (Kechin et al., 2017), and reads shorter than 18 bp in length were removed. Reads were then mapped to a database of human repetitive elements and rRNA sequences compiled from Dfam (Hubley et al., 2016) and Genbank (Benson et al., 2013). All non-repeat mapped reads were mapped to the human genome (hg38) and the custom mouse Ntrk2 minigene using STAR (v2.6.0c) (Dobin et al., 2013). PCR duplicates were removed using

umi_tools (v0.5.1) by utilizing UMI sequences from the read names and mapping positions. Peaks were identified within eCLIP samples using the peak caller CLIPper (https://github.com/YeoLab/clipper) (Lovci et al., 2013). For each peak, IP versus input fold enrichment values were calculated as a ratio of counts of reads overlapping the peak region in the IP and the input samples (read counts in each sample were normalized against the total number of reads in the sample after PCR duplicate removal). A p-value was calculated for each peak by the Yates' Chi-Square test, or Fisher Exact Test if the observed or expected read number was below 5. Comparison of different sample conditions was evaluated in the same manner as IP versus input enrichment; for each peak called in IP libraries of one sample type we calculated enrichment and p-values relative to normalized counts of reads overlapping these peaks in another sample type. Peaks were annotated using transcript information from GENCODE (Frankish et al., 2019) with the following priority hierarchy to define the final annotation of overlapping features: protein coding transcript (CDS, UTRs, intron), followed by non-coding transcripts (exon, intron).

The eCLIP data generated for this study have been deposited in NCBI's Gene Expression Omnibus and is accessible through GEO Series accession number GSE263172.

RNA seq analysis

The mRNA-Seq samples were pooled and sequenced using NovaSeq 6000 S1 Illumina® Stranded mRNA Prep and paired-end sequencing files were generated. The quality of the reads was assessed using FastQC (Andrews, 2010) . There were 160-179 million reads generated and more than 91% of bases were above the quality score of Q30. Low-quality bases and adapters were trimmed using Cutadapt (Martin, 2011). Remaining reads were mapped to the human reference genome (hg38) using STAR alignment tool (Dobin et al., 2013). The average mapping rate of all samples was

88%. Unique alignment was above 80%. Library complexity was measured in terms of unique fragments in the mapped reads using Picard's (Institute, Accessed: 2018/02/21; version 2.17.8) MarkDuplicate utility. In addition, the gene and isoform-expression estimates were quantified for all samples using RSEM (Li and Dewey, 2011). Differentially expressed transcripts at gene and isoform level were identified using DESEq2 (Love et al., 2014) and limma (Ritchie et al., 2015) package in R, respectively. The gene-ontology (GO) analysis for differentially expressed genes was carried out using WebGestalt (Liao et al., 2019) . The RNA-seq data generated for this study have been deposited in NCBI's Gene Expression Omnibus and is accessible through GEO Series accession number GSE263173.

The difference in proportion of differentially expressed isoforms between each condition was assessed using a Chi-Squared test with significance at $P < 0.01$. The difference in expression level of differentially expressed isoforms between each condition was assessed using two-sided Student's t test with significance at $P < 0.01$.

Identification and analysis of (T)GCATG clusters

The UTR exon records were added to the GRCh38 RefSeq annotation file (available on the NCBI Human Genome Resources portal) using the Python script created by David Managadze (https://github.com/yfu/tools/blob/master/add_utrs_to_gff.py). This file was parsed using a custom Python script to find any cluster containing at least 4 (T)GCATG sites in a window of 500 bp. For each identified cluster, additional (T)GCATG sequences less than 500 bp away from the cluster was considered as an extension of the cluster and was added to it. The genomic location, type of location (exon, distal intron, proximal intron, 5'UTR, 3'UTR or mixed), and the number of (T)GCATG sequences were collected for each cluster. The overlap between the genes containing eCLIP peaks and the genes containing clusters was analyzed using a custom R script.

Motifs for the binding of hnRNP-M (fourteen GU-rich pentamers: 'TGTTG','GTGTT',TTGTG','GTTGT','TGTGT','TGGTT','TTGGT','GTGTG','GGTGT','TGTGG' ,'GTGGT','GTTGG', 'TGGTG', 'GGTTG'), hnRNP-H (seven polyG-rich pentamers: 'GGGGT', 'GGGGG', 'CGGGG', 'AGGGG', 'TGGGG', 'GGGGC', 'GGGGA'), hnRNP-C (seven polyU-rich pentamers: 'ATTTT', 'GTTTT', 'CTTTT', 'TTTTA', 'TTTTC', 'TTTTG', 'TTTTT') and all possible combinations of TG and CA repeats containing at least 6 repeats $(TG/CA)_6$ were counted in the Rbfox-clusters with at least five (T)GCATG-motifs, extended to 50 bp in both directions, using a custom Python script. The same motifs were counted also in 10,000 random sequences with a length equal to the median length of the Rbfox clusters. The frequency of each motif present in the Rbfox-clusters was then compared to the frequency found in the 10,000 random sequences with a length equal to the median length of the Rbfox clusters using a chi-square test. P-values were FDR-corrected, and plots were generated using ggplot2 package (version 3.4.4) in R.

The human genes associated with brain diseases (9883 genes), muscle disorders (770 genes) and heart diseases (3003 genes) that were generated by the NHGRI-EBI catalog of human genome-wide association studies (GWAS Catalog), were analyzed for the presence of (T)GCATG-clusters and displayed as additional tabs in Supplementary Table 1.

Acknowledgments

We would like to thank Eileen Southon and Jodi Becker for critical reading of the manuscript, Thomas Gonatopoulos and Zhi-Ming Zheng for suggestions and input on the study, Bao Tran, Yongmei Zhao and Jyoti Shetty of the Sequencing Facility of the Frederick National Laboratory for Cancer Research for their help with the RNAseq experiments, Kylie Shen and Heather Foster from Eclipse Bioinnovations for the eCLIP analysis. This work was supported by the Center for Cancer Research of the Intramural Research Program of National Cancer Institute, NIH.

AUTHOR CONTRIBUTIONS

MG conducted the bioinformatics analyses. FTA and MEP conducted biological experiments. FTA and LT designed the study and wrote the manuscript with input from all authors.

References

Alexopoulou, A.N., J.R. Couchman, and J.R. Whiteford. 2008. The CMV early enhancer/chicken beta actin (CAG) promoter can be used to drive transgene expression during the differentiation of murine embryonic stem cells into vascular progenitors. *BMC Cell Biol*. 9:2.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.

Begg, B.E., M. Jens, P.Y. Wang, C.M. Minor, and C.B. Burge. 2020. Concentration-dependent splicing is enabled by Rbfox motifs of intermediate affinity. *Nat Struct Mol Biol*. 27:901-912.

Benson, D.A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, and E.W. Sayers. 2013. GenBank. *Nucleic Acids Res*. 41:D36-42.

Choquet, K., A.R. Baxter-Koenigs, S.L. Dulk, B.M. Smalec, S. Rouskin, and L.S. Churchman. 2023. Pre-mRNA splicing order is predetermined and maintains splicing fidelity across multi-intronic transcripts. *Nat Struct Mol Biol*. 30:1064-1076.

Conboy, J.G. 2017. Developmental regulation of RNA processing by Rbfox proteins. *Wiley Interdiscip Rev RNA*. 8.

Conboy, J.G. 2021. Unannotated splicing regulatory elements in deep intron space. *Wiley Interdiscip Rev RNA*. 12:e1656.

Damianov, A., Y. Ying, C.H. Lin, J.A. Lee, D. Tran, A.A. Vashisht, E. Bahrami-Samani, Y. Xing, K.C. Martin, J.A. Wohlschlegel, and D.L. Black. 2016. Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. *Cell*. 165:606-619.

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 29:15-21.

Frankish, A., M. Diekhans, A.M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J.M. Mudge, C. Sisu, J. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, S. Carbonell Sala, J. Chrast, F. Cunningham, T. Di Domenico, S. Donaldson, I.T. Fiddes, C. Garcia Giron, J.M. Gonzalez, T. Grego, M. Hardy, T.

Hourlier, T. Hunt, O.G. Izuogu, J. Lagarde, F.J. Martin, L. Martinez, S. Mohanan, P. Muir, F.C.P. Navarro, A. Parker, B. Pei, F. Pozo, M. Ruffier, B.M. Schmitt, E. Stapleton, M.M. Suner, I. Sycheva, B. Uszczynska-Ratajczak, J. Xu, A. Yates, D. Zerbino, Y. Zhang, B. Aken, J.S. Choudhary, M. Gerstein, R. Guigo, T.J.P. Hubbard, M. Kellis, B. Paten, A. Reymond, M.L. Tress, and P. Flicek. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. 47:D766-D773.

Gabellini, N. 2001. A polymorphic GT repeat from the human cardiac Na+Ca2+ exchanger intron 2 activates splicing. *Eur J Biochem*. 268:1076-1083.

Gehman, L.T., P. Stoilov, J. Maguire, A. Damianov, C.H. Lin, L. Shiue, M. Ares, Jr., I. Mody, and D.L. Black. 2011. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat Genet*. 43:706-711.

Gehring, N.H., and J.Y. Roignant. 2021. Anything but Ordinary - Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends Genet*. 37:355-372.

Hakim, N.H., T. Kounishi, A.H. Alam, T. Tsukahara, and H. Suzuki. 2010. Alternative splicing of Mef2c promoted by Fox-1 during neural differentiation in P19 cells. *Genes Cells*. 15:255-267.

Hollander, D., S. Naftelberg, G. Lev-Maor, A.R. Kornblihtt, and G. Ast. 2016. How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing? *Trends Genet*. 32:596-606.

Hubley, R., R.D. Finn, J. Clements, S.R. Eddy, T.A. Jones, W. Bao, A.F. Smit, and T.J. Wheeler. 2016. The Dfam database of repetitive DNA families. *Nucleic Acids Res*. 44:D81-89.

Hui, J., K. Stangl, W.S. Lane, and A. Bindereif. 2003. HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat Struct Biol*. 10:33-37.

Institute, B. Accessed: 2018/02/21; version 2.17.8. Picard Tools. *Broad Institute, GitHub repository*.

Jin, Y., H. Suzuki, S. Maegawa, H. Endo, S. Sugano, K. Hashimoto, K. Yasuda, and K. Inoue. 2003. A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J*. 22:905-912.

Kechin, A., U. Boyarskikh, A. Kel, and M. Filipenko. 2017. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *J Comput Biol*. 24:1138-1143.

Kivioja, T., A. Vaharautio, K. Karlsson, M. Bonke, M. Enge, S. Linnarsson, and J. Taipale. 2011. Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods*. 9:72-74.

Kumanogoh, H., J. Asami, S. Nakamura, and T. Inoue. 2008. Balanced expression of various TrkB receptor isoforms from the Ntrk2 gene locus in the mouse nervous system. *Mol Cell Neurosci*. 39:465-477.

Kuroyanagi, H. 2009. Fox-1 family of RNA-binding proteins. *Cell Mol Life Sci*. 66:3895-3907.

Lal, D., H. Trucks, R.S. Moller, H. Hjalgrim, B.P. Koeleman, C.G. de Kovel, F. Visscher, Y.G. Weber, H. Lerche, F. Becker, C.J. Schankin, B.A. Neubauer, R. Surges, W.S. Kunz, F. Zimprich, A. Franke, T. Illig, J.S. Ried, C. Leu, P. Nurnberg, T. Sander, E.M. Consortium, and E. Consortium. 2013. Rare exonic deletions of the RBFOX1 gene increase risk of idiopathic generalized epilepsy. *Epilepsia*. 54:265-271.

Lee, J.A., A. Damianov, C.H. Lin, M. Fontes, N.N. Parikshak, E.S. Anderson, D.H. Geschwind, D.L. Black, and K.C. Martin. 2016. Cytoplasmic Rbfox1 Regulates the Expression of Synaptic and Autism-Related Genes. *Neuron*. 89:113-128.

Li, B., and C.N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 12:323.

Li, D., Q. Wang, A. Bayat, M.R. Battig, Y. Zhou, D.G. Bosch, G. van Haaften, L. Granger, A.K. Petersen, L.A. Perez-Jurado, G. Aznar-Lain, A. Aneja, M. Hancarova, S. Bendova, M. Schwarz, R. Kremlikova Pourova, Z. Sedlacek, B.A. Keena, M.E. March, C. Hou, N. O'Connor, E.J. Bhoj, M.H. Harr, G. Lemire, K.M. Boycott, M.C. Towne, M. Li, M. Tarnopolsky, L. Brady, M.J. Parker, H. Faghfoury, L.K. Parsley, E. Agolini, M.L. Dentici, A. Novelli, M.S. Wright, R. Palmquist, K. Lai, M. Scala, P. Striano, M. Iacomino, F. Zara, A. Cooper, T.J. Maarup, M. Byler, R.R. Lebel, T.B. Balci, R.J. Louie, M.J. Lyons, J. Douglas, C.B. Nowak, A. Afenjar, J. Hoyer, B. Keren, S.M. Maas, M.M. Motazacker, J.A. Martinez-

Agosto, A.M. Rabani, E.M. McCormick, M. Falk, S.M. Ruggiero, I. Helbig, R.S. Moller, L. Tessarollo, F. Tomassoni-Ardori, M.E. Palko, T.C. Hsieh, P.M. Krawitz, M. Ganapathi, B.D. Gelb, V. Jobanputra, A. Wilson, J. Greally, S. Jacquemont, K. Jizi, B. Ange-Line, C. Quelin, V.K. Misra, E. Chick, C. Romano, D. Greco, A. Arena, M. Morleo, V. Nigro, R. Seyama, Y. Uchiyama, N. Matsumoto, R. Taira, K. Tashiro, Y. Sakai, G. Yigit, B. Wollnik, M. Wagner, B. Kutsche, A.C. Hurst, M.L. Thompson, R.J. Schmidt, L.M. Randolph, R.C. Spillmann, V. Shashi, et al. 2023. Spliceosome malfunction causes neurodevelopmental disorders with overlapping features. *J Clin Invest*.

Liao, Y., J. Wang, E.J. Jaehnig, Z. Shi, and B. Zhang. 2019. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res*. 47:W199-W205.

Lin, L., J. Goke, E. Cukuroglu, M.R. Dranias, A.M. VanDongen, and L.W. Stanton. 2016. Molecular Features Underlying Neurodegeneration Identified through In Vitro Modeling of Genetically Diverse Parkinson's Disease Patients. *Cell Rep*. 15:2411-2426.

Liu, P., N.A. Jenkins, and N.G. Copeland. 2003. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res*. 13:476-484.

Lovci, M.T., D. Ghanem, H. Marr, J. Arnold, S. Gee, M. Parra, T.Y. Liang, T.J. Stark, L.T. Gehman, S. Hoon, K.B. Massirer, G.A. Pratt, D.L. Black, J.W. Gray, J.G. Conboy, and G.W. Yeo. 2013. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat Struct Mol Biol*. 20:1434-1442.

Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15:550.

Luberg, K., J. Wong, C.S. Weickert, and T. Timmusk. 2010. Human TrkB gene: novel alternative transcripts, protein isoforms and expression pattern in the prefrontal cerebral cortex during postnatal development. *J Neurochem*. 113:952-964.

Lunde, B.M., C. Moore, and G. Varani. 2007. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*. 8:479-490.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011*. 17:3.

Nakahata, S., and S. Kawamoto. 2005. Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. *Nucleic Acids Res*. 33:2078-2089.

O'Leary, A., N. Fernandez-Castillo, G. Gan, Y. Yang, A.Y. Yotova, T.M. Kranz, L. Grunewald, F. Freudenberg, E. Anton-Galindo, J. Cabana-Dominguez, A. Harneit, J.I. Schweiger, K. Schwarz, R. Ma, J. Chen, E. Schwarz, M. Rietschel, H. Tost, A. Meyer-Lindenberg, C.A. Pane-Farre, T. Kircher, A.O. Hamm, D. Burguera, N.R. Mota, B. Franke, S. Schweiger, J. Winter, A. Heinz, S. Erk, N. Romanczuk-Seiferth, H. Walter, A. Strohle, L. Fehm, T. Fydrich, U. Lueken, H. Weber, T. Lang, A.L. Gerlach, M.M. Nothen, G.W. Alpers, V. Arolt, S. Witt, J. Richter, B. Straube, B. Cormand, D.A. Slattery, and A. Reif. 2022. Behavioural and functional evidence revealing the role of RBFOX1 variation in multiple psychiatric disorders and traits. *Mol Psychiatry*. 27:4464-4473.

Peyda, P., C.H. Lin, K. Onwuzurike, and D.L. Black. 2025. The Rbfox1/LASR complex controls alternative pre-mRNA splicing by recognition of multipart RNA regulatory modules. *Genes Dev*. 39:364-383.

Raghavan, N.S., L. Dumitrescu, E. Mormino, E.R. Mahoney, A.J. Lee, Y. Gao, M. Bilgel, D. Goldstein, T. Harrison, C.D. Engelman, A.J. Saykin, C.D. Whelan, J.Z. Liu, W. Jagust, M. Albert, S.C. Johnson, H.S. Yang, K. Johnson, P. Aisen, S.M. Resnick, R. Sperling, P.L. De Jager, J. Schneider, D.A. Bennett, M. Schrag, B. Vardarajan, T.J. Hohman, R. Mayeux, and I. Alzheimer's Disease Neuroimaging. 2020. Association Between Common Variants in RBFOX1, an RNA-Binding Protein, and Brain Amyloidosis in Early and Preclinical Alzheimer Disease. *JAMA Neurol*. 77:1288-1298.

Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 43:e47.

Sharan, S.K., L.C. Thomason, S.G. Kuznetsov, and D.L. Court. 2009. Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc*. 4:206-223.

Sharma, V.K., S.K. Brahmachari, and S. Ramachandran. 2005. (TG/CA)n repeats in human gene families: abundance and selective patterns of distribution according to function and gene length. *BMC Genomics*. 6:83.

Smith, T., A. Heger, and I. Sudbery. 2017. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res*. 27:491-499.

Tessarollo, L., and S. Yanpallewar. 2022. TrkB Truncated Isoform Receptors as Transducers and Determinants of BDNF Functions. *Front Neurosci*. 16:847572.

Tomassoni-Ardori, F., G. Fulgenzi, J. Becker, C. Barrick, M.E. Palko, S. Kuhn, V. Koparde, M. Cam, S. Yanpallewar, S. Oberdoerffer, and L. Tessarollo. 2019. Rbfox1 up-regulation impairs BDNF-dependent hippocampal LTP by dysregulating TrkB isoform expression levels. *Elife*. 8.

Ule, J., and B.J. Blencowe. 2019. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol Cell*. 76:329-345.

Van Nostrand, E.L., P. Freese, G.A. Pratt, X. Wang, X. Wei, R. Xiao, S.M. Blue, J.Y. Chen, N.A.L. Cody, D. Dominguez, S. Olson, B. Sundararaman, L. Zhan, C. Bazile, L.P.B. Bouvrette, J. Bergalet, M.O. Duff, K.E. Garcia, C. Gelboin-Burkhart, M. Hochman, N.J. Lambert, H. Li, M.P. McGurk, T.B. Nguyen, T. Palden, I. Rabano, S. Sathe, R. Stanton, A. Su, R. Wang, B.A. Yee, B. Zhou, A.L. Louie, S. Aigner, X.D. Fu, E. Lecuyer, C.B. Burge, B.R. Graveley, and G.W. Yeo. 2020a. A large-scale binding and functional map of human RNA-binding proteins. *Nature*. 583:711-719.

Van Nostrand, E.L., T.B. Nguyen, C. Gelboin-Burkhart, R. Wang, S.M. Blue, G.A. Pratt, A.L. Louie, and G.W. Yeo. 2017. Robust, Cost-Effective Profiling of RNA Binding Protein Targets with Single-end Enhanced Crosslinking and Immunoprecipitation (seCLIP). *Methods Mol Biol*. 1648:177-200.

Van Nostrand, E.L., G.A. Pratt, A.A. Shishkin, C. Gelboin-Burkhart, M.Y. Fang, B. Sundararaman, S.M. Blue, T.B. Nguyen, C. Surka, K. Elkins, R. Stanton, F. Rigo, M. Guttman, and G.W. Yeo. 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods*. 13:508-514.

Van Nostrand, E.L., G.A. Pratt, B.A. Yee, E.C. Wheeler, S.M. Blue, J. Mueller, S.S. Park, K.E. Garcia, C. Gelboin-Burkhart, T.B. Nguyen, I. Rabano, R. Stanton, B. Sundararaman, R. Wang, X.D. Fu, B.R. Graveley, and G.W. Yeo. 2020b. Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol*. 21:90.

Verma, S.K., V. Deshmukh, C.A. Nutter, E. Jaworski, W. Jin, L. Wadhwa, J. Abata, M. Ricci, J. Lincoln, J.F. Martin, G.W. Yeo, and M.N. Kuyumcu-Martinez. 2016. Rbfox2 function in RNA metabolism is impaired in hypoplastic left heart syndrome patient hearts. *Sci Rep*. 6:30896.

Wan, Y., D.G. Anastasakis, J. Rodriguez, M. Palangat, P. Gudla, G. Zaki, M. Tandon, G. Pegoraro, C.C. Chow, M. Hafner, and D.R. Larson. 2021. Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell*. 184:2878-2895 e2820.

Wen, M., Y. Yan, N. Yan, X.S. Chen, S.Y. Liu, and Z.H. Feng. 2015. Upregulation of RBFOX1 in the malformed cortex of patients with intractable epilepsy and in cultured rat neurons. *Int J Mol Med*. 35:597-606.

Weyn-Vanhentenryck, S.M., A. Mele, Q. Yan, S. Sun, N. Farny, Z. Zhang, C. Xue, M. Herre, P.A. Silver, M.Q. Zhang, A.R. Krainer, R.B. Darnell, and C. Zhang. 2014. HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep*. 6:1139-1152.

Yao, F., T. Svensjo, T. Winkler, M. Lu, C. Eriksson, and E. Eriksson. 1998. Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells. *Hum Gene Ther*. 9:1939-1950.

Ying, Y., X.J. Wang, C.K. Vuong, C.H. Lin, A. Damianov, and D.L. Black. 2017. Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. *Cell*. 170:312-323 e310.

Zhang, C., K.Y. Lee, M.S. Swanson, and R.B. Darnell. 2013. Prediction of clustered RNA-binding protein motif sites in the mammalian genome. *Nucleic Acids Res*. 41:6793-6807.

Zhang, J., and E.J. Huang. 2006. Dynamic expression of neurotrophic factor receptors in postnatal spinal motoneurons and in mouse model of ALS. *J Neurobiol*. 66:882-895.

Figure legends

Figure 1. Generation of a TrkB-BAC minigene system to study Rbfox1 function on TrkB isoforms expression levels.

(A) Schematic representation of the murine *NTrk2* (TrkB) gene and of the TrkB-BAC minigene. Exon length is indicated in base pairs (bp) while intron length is in kilo-bases (Kb). White and black stars indicate the start and stop codons, respectively. Exons in black encode for the extra-cellular (ED) and the transmembrane (TM) domain and are common to both the truncated (TrkB.T1) and full length (TrkB.FL) isoforms. In green (T1) is the exon unique to the TrkB.T1 isoform. In orange are the exons specific to the TrkB.FL isoform and encoding the tyrosine kinase domain (TK). The TrkB-BAC minigene includes a murine 164 Kb genomic fragment with an upstream synthetic CAG promoter and a cDNA fragment encoding the extracellular domain fused in frame to a juxtamembrane domain exon, and, downstream, a cDNA fragment encoding the TK domain and a neomycin resistance cassette (NEO) used for selection. **(B)** Quantitative PCR analysis of TrkB.FL, TrkB.T1 and Rbfox1 expression from two independent HEK293 clones expressing the TrkB-BAC minigene (77-5 and 77-6 cells) in the absence (No Dox) or presence (+ Dox) of doxycycline (0.5 µg/ml for 48h); n=6 ±SEM (n=3 for each clone). **(C)** Western blot analysis of lysates from parental T-REx 293 cells (T-REx), T-REx 293 cells with knocked-in Rbfox1 under doxycycline control (T-REx-Fox1), and the two clones expressing the TrkB-BAC minigene (77-5 and 77-6) with or without doxycycline (Dox: 0.5 µg/ml for 48h) and probed with an antibody recognizing TrkB.FL or TrkB.T1 specifically. An antibody against RbFox1 was used to verify RbFox1 induction and Gapdh was used as a loading control. Note that the parental T-REx cell line does not express RbFox1 while the engineered cell line and the

77-5 and 77-6 minigene-expressing clones express RbFox1 only in the presence of doxycycline. **(D)** Schematic representation of the RbFox1 protein showing the location of the RNA recognition motif (RRM in red) and the position of the F158A mutation within the RRM motif. **(E)** Quantitative PCR analysis of TrkB.FL, TrkB.T1 and Rbfox1 expression levels from two independent clones with the TrkB-BAC minigene not transfected (NT) or transfected with the Rbfox1-F158A mutant expressing plasmid; n=6 ±SEM (n=3 for each clone). **(F)** Western blot analysis as in (C) from two independent clones with the TrkB-BAC, untransfected (NT) or transfected with a plasmid expressing wild-type Rbfox1 (Rbfox1) or the mutant Rbfox1-F158A (F158A).

Figure 2. Rbfox1 eCLIP peaks in a TrkB-BAC minigene coincides with clusters of (T)GCATG motifs.

(A) eCLIP analysis of RbFox1 in HEK293 cells with the TrkB-BAC. eCLIP peaks were derived by subtracting the signal obtained in the absence of RbFox1 (-Dox), considered as background, from the signal from the same cells (line 77-5 from Fig. 2) expressing RbFox1. In green are seven statistically significant eCLIP peaks, all in distal intronic regions. Numbers under each eCLIP peak indicate the p-value ($-\log_{10}$). **(B)** Enlargement of the areas containing two eCLIP peaks (green) relative to the position of RbFox1 (T)GCATG binding motifs (red). **(C)** Sequence of the DNA region including the two (T)GCATG-Rbfox1 motif (red) clusters shown in (B). Note that both clusters of (T)GCATG-motifs (cluster 1 and cluster 2) are located in distal intronic regions flanking the specific TrkB.T1 exon.

Figure 3. (U)GCATG-clusters are widespread in the mammalian genome and are associated to RbFox1 isoform gene expression regulation.

(A) (T)GCATG clusters across the mouse and human genome were defined by the presence of at least four or more motifs within a genomic space of 500 bp. (B) Donut chart showing the distribution in mouse of the 21005 clusters of motifs between exons, 5'UTR, 3'UTR, proximal introns (within a distance of 500 bp from exons), distal introns and mixed (located in between those categories). (C) Venn diagram showing the overlap between the total number of genes with (T)GCATG clusters (8348) and the genes with significant Rbfox1 HITS-CLIP peaks (8356 genes) in mouse brain [from (Weyn-Vanhentenryck et al., 2014)]. (D) Donut chart showing the number of genes displaying significant HITS-CLIP peaks on (T)GCATG clusters (1687 genes; 20%) and those without HITS-CLIP peaks on clusters (6669 genes; 80%) among the total number of genes bound by Rbfox1 in mouse brain (8356 genes). **(E)** Donut chart showing the distribution of the 19977 (T)GCATG clusters identified in the human genome between exons, 5'UTR, 3'UTR, proximal introns (within a distance of 500 bp from exons), distal introns and mixed (located in between those categories). **(F)** Venn diagram showing that 2908 genes overlap between the total number of human genes with (T)GCATG clusters (9826) and genes with Rbfox1 eCLIP peaks (4903 genes) identified in the T-REx-293 cell analysis. **(G)** Donut chart showing the number of human genes displaying significant eCLIP peaks on (T)GCATG clusters (1237 genes; 25%) and the number of genes without eCLIP peaks on (T)GCATG clusters (3666 total genes; 75%). **(H)** Chart depicting the relationship between the number of genes (%) with a change in isoform expression and the % change in expression of the isoforms (% Fold Change isoform expression) in the set of genes with and without eCLIP peaks on (T)GCATG clusters. **(I)** Graph showing the overall average of fold change isoform expression (Log_2FC Average) for the set of genes with eCLIP peaks on (T)GCATG clusters and eCLIP peaks outside of the clusters, or eCLIP peaks in the absence of clusters. ** $p \leq 0.01$; *** p-value$\leq$ 0.00001.

Figure 4. Deletion of deep intronic (T)GCATG clusters impairs RbFox1 regulation of TrkB isoform expression.

(A) Schematic representation of the location (red X) of the Cluster 1 deletion in the TrkB-BAC minigene analyzed in (B-C). (B) Western blot analysis of TrkB.FL, TrkB.T1 and RbFox1 protein expression levels from two independent cell lines (1-1 and 1-2) with the TrkB-BAC minigene with cluster 1 deleted, in the absence or presence of doxycycline (No Dox or +Dox 0.5 µg/ml for 48h); RbFox1- and Gapdh-specific antibodies were used, respectively, to verify doxycycline induction of Rbfox1 and as a control of protein loading. (C) Quantitative PCR analysis of *TrkB.FL, TrkB.T1* and *Rbfox1* expression levels in the absence or presence of doxycycline of clones as in (B); n=6 ±SEM (n=3 for each clone). (D) Schematic representation of the location of the Cluster 2 deletion in the TrkB-BAC minigene analyzed in (E-F). (E) Western blot analysis as in (B) of two cell lines (27-51 and 27-53) expressing the TrkB-BAC minigene with cluster 2 deleted. (F) Quantitative PCR analysis as in (C) of the two independent cell lines analyzed in (E); n=6 ±SEM (n=3 for each clone). (G) Schematic representation of the location of Cluster 1 and 2 deleted in the TrkB-BAC minigene analyzed in (H, I). (H) Western blot analysis as in (B) of the two cell lines (32-1 and 32-2) expressing the TrkB-BAC minigene with cluster 1 and 2 deleted. (I) Quantitative PCR analysis as in (C) of the two independent cell lines analyzed in (H). n=6 ±SEM (n=3 for each clone).

Figure 5. (T)GCATG clusters are embedded in sequences binding the LASR

(A) Volcano plot illustrating the enrichment of fourteen GU-rich pentamer motifs of hnRNP-M (blue dots)
['TGTTG','GTGTT',TTGTG','GTTGT','TGTGT','TGGTT','TTGGT','GTGTG','GGTGT','TGTGG','GTGGT','GTTGG', 'TGGTG', 'GGTTG'], seven polyG-rich pentamers motifs of hnRNP-H

(gray dots) ['GGGGT', 'GGGGG', 'CGGGG', 'AGGGG', 'TGGGG', 'GGGGC', 'GGGGA'] and seven polyU-rich pentamers of hnRNP-C (green dots) ['ATTTT', 'GTTTT', 'CTTTT', 'TTTTA', 'TTTTC', 'TTTTG', 'TTTTT'] within (T)GCAUG-clusters containing at least 5 or more (T)GCATG motifs in the mouse genome. -Log_{10}(p-value) is indicated on the Y axis while Log_2(Fold Change) is indicated on the X axis. P-values greater than 0.01 were considered as statistically significant. **(B)** Volcano plot illustrating the enrichment of motifs of hnRNP-M (blue dots), hnRNP-H (gray dots) and hnRNP-C (green dots) in the human genome analyzed as in (A). **(C)** Volcano plot illustrating the enrichment of $(TG/CA)_6$ repeats (64 different combinations) within (T)GCAUG-clusters containing at least 5 or more (T)GCATG motifs in the mouse genome. -Log_{10}(p-value) is indicated on the Y axis while Log_2(Fold Change) is indicated on the X axis. P-values greater than 0.01 were considered as statistically significant. **(D)** Volcano plot illustrating the enrichment of $(TG/CA)_6$ repeats (64 different combinations) within (T)GCAUG-clusters in the human genome analyzed as in (C). **(E)** Western blot analysis of lysates from HEK293 cells with the WT TrkB-BAC minigene 48h after transfection with a control (Rbfox1) or an Rbfox1 cDNA with mutations in 10 tyrosine residues in the CTD region (Ying et al., 2017) showing that the mutant Rbfox1-Y10 partially retains the ability to regulate TrkB isoform expression through (T)GCATG-clusters. Non-transfected cells were used as control (NT). TrkB.FL, Rbfox1 and Gapdh protein levels were analyzed as in Fig 1. **(F)** Immunoblot quantification analysis of TrkB.FL protein levels in HEK293 cells with the WT TrkB-BAC minigene 48h after transfection with Rbfox1 or Rbfox1-Y10 relative to not transfected cells (NT) as in (E); n=3±SEM (One-way ANOVA).

Figure 6. Model summary of how RbFox1 regulates pre-mRNA splicing in conjunction with LASR through binding of (U)GCAUG-clusters. (U)GCAUG-clusters participate in the

pre-mRNA splicing regulation by facilitating local recruitment and accumulation of Rbfox1 proteins to pre-mRNA sites where they form large RbFox1 protein aggregates and higher-order assembly with the LASR (normal splicing activity in green box). Conversely, the Rbfox1-Y10 mutant is unable to form large Rbfox1-protein aggregates but can still be recruited to the (U)GCAUG-clusters and mediate a reduced LASR assembly leading to limited splicing efficiency (reduced splicing activity in orange box). Inactivation of the RbFox1 RNA-binding motif (RRM) or deleting (U)GCAUG-clusters leads to a complete loss of RbFox1 splicing functions (no splicing activity in red boxes).

Supplemental Figure Legends

Supplemental Figure 1. Location, statistics, and fold change enrichment of the seven eCLIP peaks found across the minigene sequence shown in Figure 2.

Supplementary Figure 2. Analysis of Rbfox1-eCLIP peaks in HEK293 cells with induced Rbfox1 expression (+Dox) compared to uninduced cells (No Dox).

(A) Top five enriched motifs identified in CLIP-seq peaks by the HOMER motif analysis. (B) Pie chart depicting the relative frequency of eCLIP peaks that map to each specific gene region (with a peak Log_2 fold enrichment ≥ 3 and p-value ≤ 0.001). (C, D) Peak Metagene Plot, depicting the average number of peaks mapped to the specific genomic regions indicated in B. The number of peaks was calculated for each gene region followed by normalization with the length of the regions. The average number of peaks was then calculated for a set number of positions along the regions.

Supplementary Figure 3. Strategy to delete (T)GCATG-clusters from the TrkB-BAC minigene.

(A) Schematic representation of the TrkB-BAC minigene showing the position of the PCR-primers designed to detect the deletion of (T)GCATG-cluster 1 and (T)GCATG-cluster 2 (indicated by red X) analyzed in (C). (B) magnification of Cluster 1 and 2 areas indicating the location and sequence of the primers used for the analysis. (C) PCR analysis of genomic DNA from HEK293 cells used as negative control, cell lines expressing the 'wild-type' TrkB-BAC minigene (77-5 and 77-6 cells), cell lines expressing the TrkB-BAC minigene with cluster 1 deletion (1-1 and 1-2 cells), cell lines expressing the TrkB-BAC minigene with cluster 2 deletion

(27-51 and 27-53 cells) and cell lines expressing the TrkB-BAC minigene with both cluster 1 and 2 deletion (32-1 and 32-2 cells). The PCR detecting cluster 1 deletion shows an amplicon of 891 bp (wild-type minigene sequence) and an amplicon of 279 bp (deletion of cluster 1). The PCR detecting cluster 2 deletion shows an amplicon of 746 bp (wild-type minigene sequence) and an amplicon of 322 bp (deletion of cluster 2).

Supplementary Figure 4. Deletion of RbFox1 binding sites in Cluster 1 and 2 leads to the loss of only Cluster 1 and 2-specific eCLIP peaks.

(A) Schematic representation of the TrkB-BAC minigene indicating the location of cluster1 and 2 deletion (red x) with the RbFox1 (+ Dox) eCLIP analysis of wild type TrkB-BAC expressing cells (line 77-5) compared to TrkB-BAC cells with cluster 1 and 2 deletions (line 32-1). Note the presence of only two statistically significant eCLIP areas in the intronic regions (black marks) corresponding to cluster 1 and 2. (B) Enlargement of the areas containing the two eCLIP peaks (indicated by shadowed arrows) corresponding to the (T)GCATG clusters (red marks). (C) Location, statistics, and fold change enrichment of all eCLIP peaks located in cluster 1 and cluster 2 in (B).

Supplementary Figure 5. Mutant Rbfox1-Y10 does not regulate TrkB isoform expression in the TrkB-BAC minigene lacking Rbfox1 binding cluster 1 and 2.

(A) Schematic representation of the TrkB-BAC minigene indicating the location of cluster1 and 2 deletion (red x). (B) Western blot analysis of lysates from HEK293 cells with the mutant cluster 1 and 2 deletion TrkB-BAC minigene 48h after transfection with a control (RbFox1) or an RbFox1 cDNA with mutations in 10 tyrosine residues in the CTD region (Ying et al., 2017)

(Rbfox1-Y10). Non-transfected cells were used as control (NT). TrkB.FL, Rbfox1 and Gapdh protein levels were analyzed as in Fig. 1.

Fig. 1

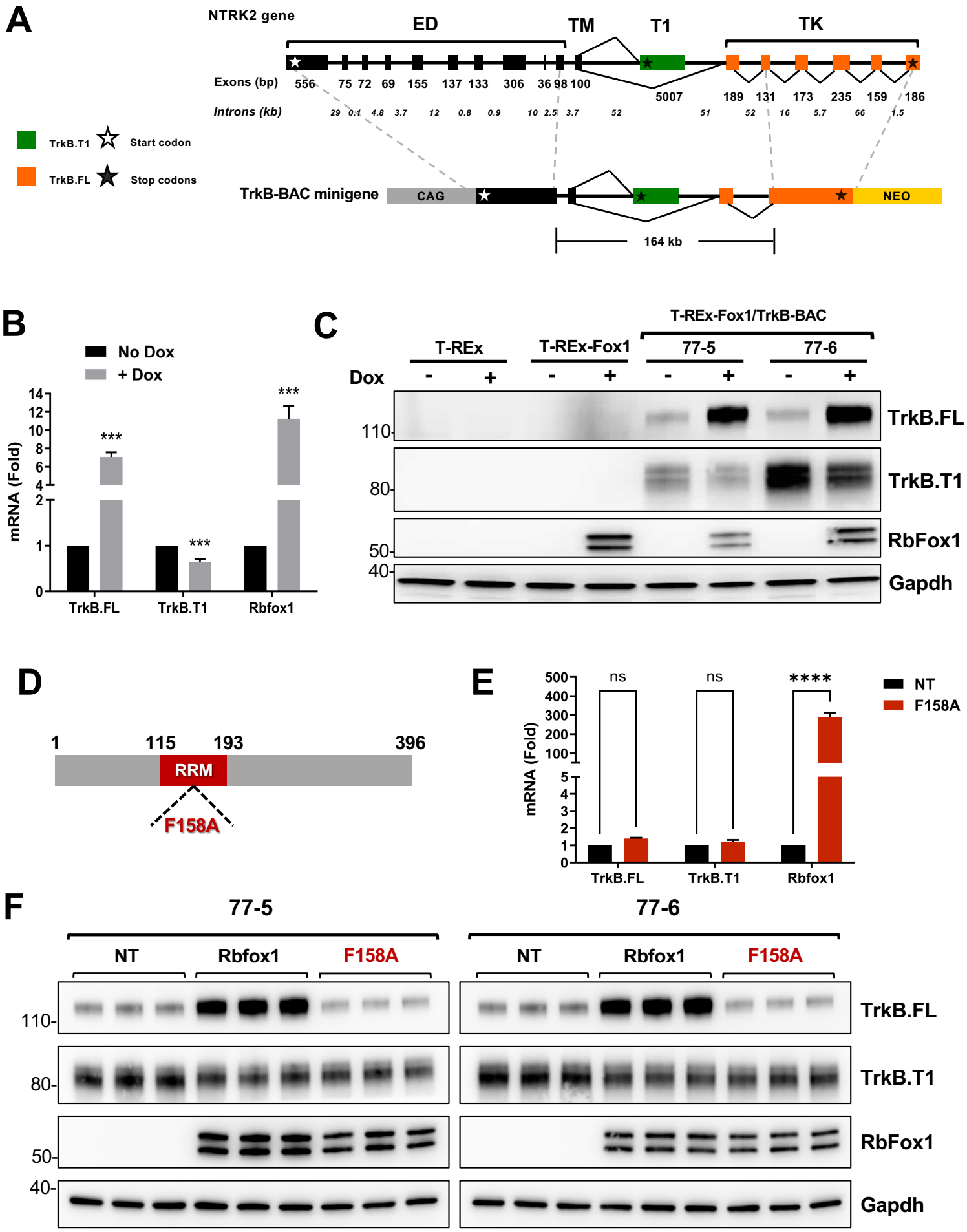
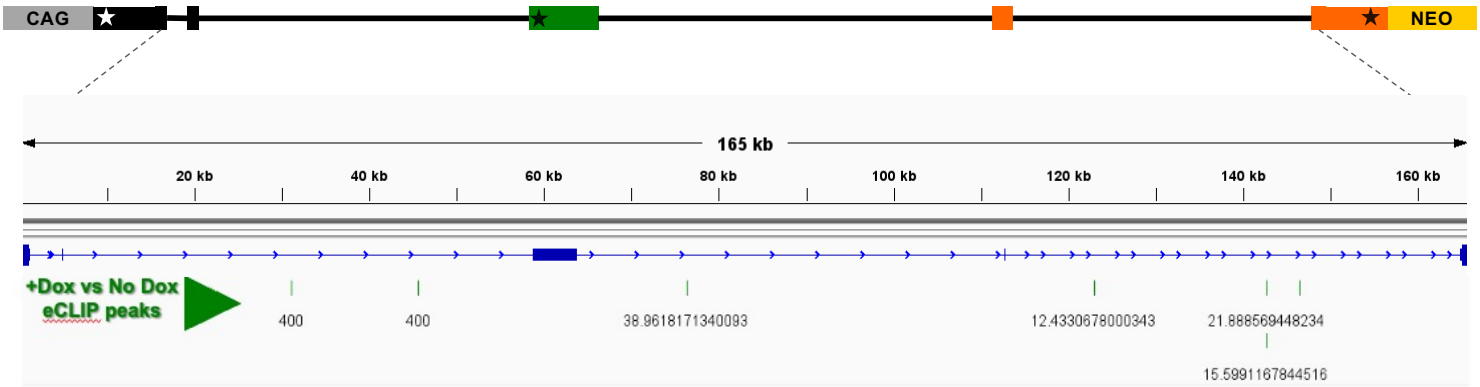
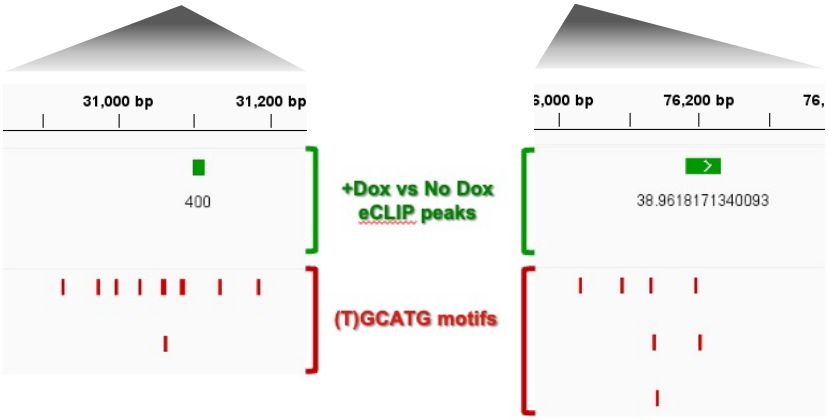

A
NTRK2 gene
ED
TM
T1
TK
Exons (bp)
556 75 72 69 155 137 133 306 36 98 100 5007 189 131 173 235 159 186
Introns (kb)
29 0.1 4.8 3.7 12 0.8 0.9 10 2.5 3.7 52 51 52 16 5.7 66 1.5
TrkB.T1
Start codon
TrkB.FL
Stop codons
TrkB-BAC minigene
CAG
NEO
164 kb
B
No Dox
+ Dox
mRNA (Fold)
TrkB.FL
TrkB.T1
Rbfox1
C
T-REx-Fox1/TrkB-BAC
T-REx
T-REx-Fox1
77-5
77-6
Dox
TrkB.FL
TrkB.T1
RbFox1
Gapdh
D
1
115
193
396
RRM
F158A
E
ns
ns
****
NT
F158A
mRNA (Fold)
TrkB.FL
TrkB.T1
Rbfox1
F
77-5
77-6
NT
Rbfox1
F158A
TrkB.FL
TrkB.T1
RbFox1
Gapdh

Fig. 2

A



B



C

...CTAAGAAAATTCCTTTCAAACGCGTTGTGCG
CCTTGACCATATTCCTCCTACACCTTCCTCTTC
TGTCCAACCCTTCCTGTGCTCGGCTACCCCTT
CCAGGCCTCCCTCCTGTTCTCCTGTCACAAAC
ACGTGTG**TGCATG**TGTGTGTGTACACATGTGT
GTATGTGTATATACATATGTG**TGCATG**TGTATG
TAGGCGTATATGA**GCATG**TGTGTATATATATAT
GTGTGTGTGTG**TGCATG**TGTGTATGTATATAT
GCATATGTG**TGCATGCATG**TGTGTAGGTGTGT
ATGA**GCATG**TGTAAATACATATGTGTGCATCT
GTGTGCACATGTGTGTATGTG**TGCATG**TGTGT
GTGTGTAGGTGAGTATGAGCATTTATATATGCA
TATGAG**TGCATG**TGTGTATCTGTGTGCATACA
TGTGCACATGTGTGTGTTATAATTTTAGCCTTT
GTGACTGGATAAACCCCCACTTATCCATCTTC
ACCAGTGAGTCATCTGTAGACAGACCCTTATG
GCTTGGTCTACAAGTTGCTTATCAAAATGAACA
CAACAGCAAAATCTGGACAATTCCTGACCTAT
GGGTGCCTTAAACTTGGTAGTTTAGGAAAAC...

...TTCTTGTCATGGTTTTTAGGTGTGTGT
GTGTGTATGCCTGGGCAGGAGTGTGAG
TATGTGAAAATGTATATACAAATGTGTGA
GTGTGTGTGTGCATAATGTGTGCAAGTA
TATGTGAGAGTCTTGTGTGTGTGCCTGT
GTGTGTGTATCCTTGGATATCTATGTGT
GCCTAGGTATG**TGCATG**AGTGTGTGTAA
GTATGTGTATGTTTGTTTGTATGTGTTGT
GTGAATGTATATATA**TGCATG**TGTGCAA
CTATGTGTGTGAGTATTGTGCACGTGTG
TGCATGCATGCATGTGTAAAGTATGTGT
GAGAGTATTATGTGTGACAGTGTGTGTG
ATAGTGTG**TGCATG**T**TGCATG**AGTGTGT
GTATCTGTGTGAAAGTGTGTGTTTATAC
ACATGCTCGGTGTGAGTGTGTGTATGTG
TGAT...

Fig. 3

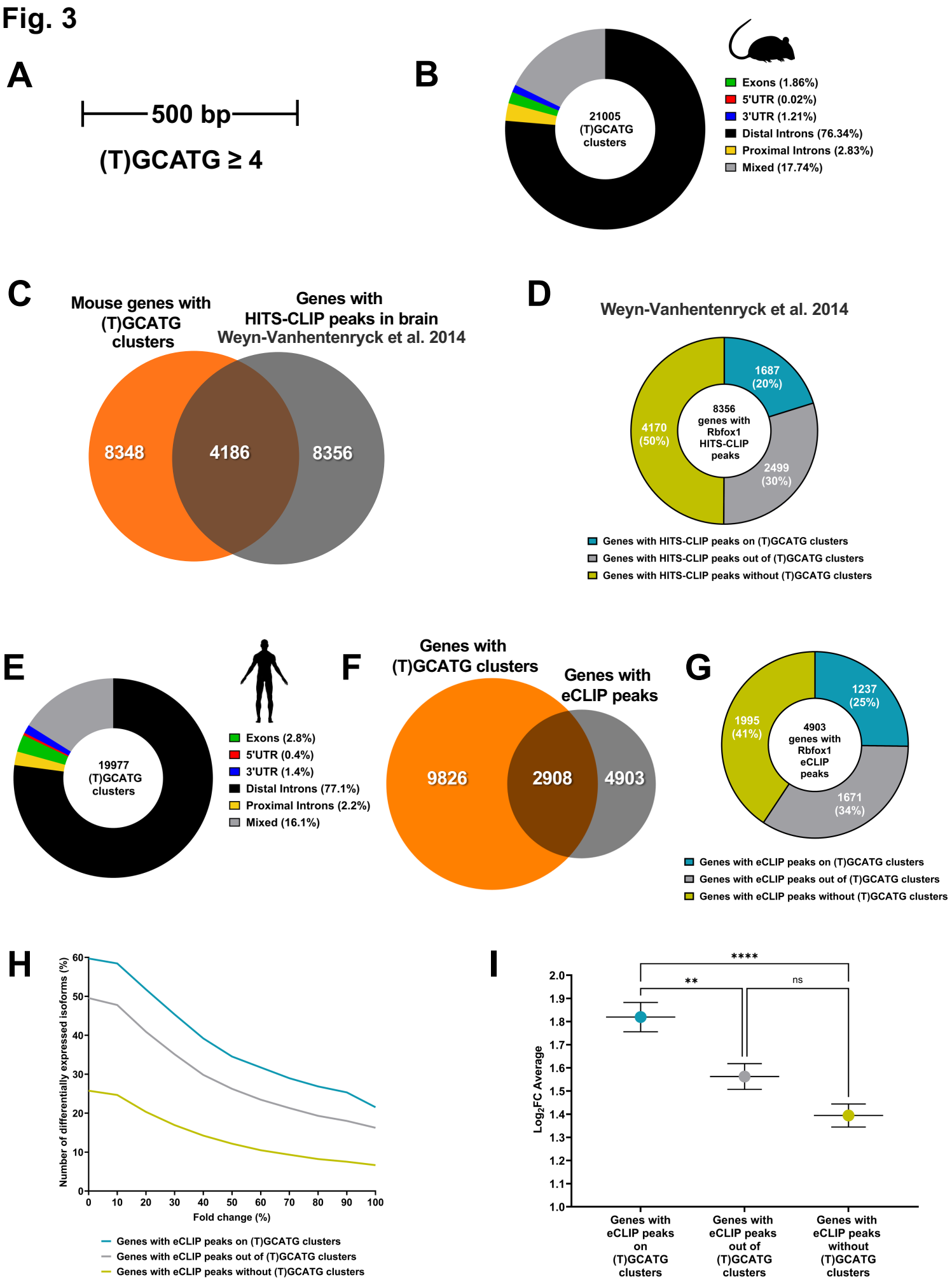
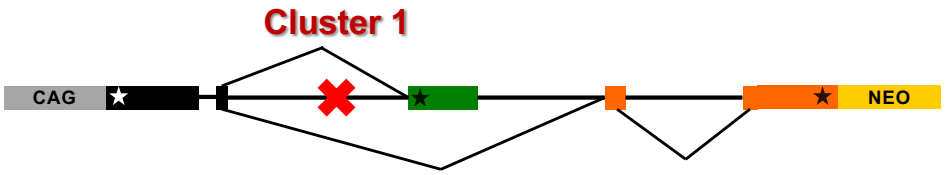
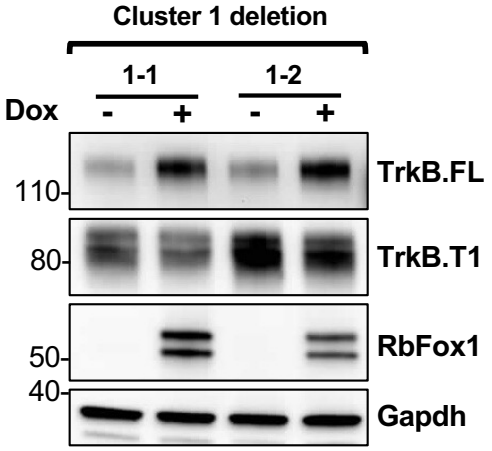

A
500 bp
(T)GCATG ≥ 4
B
21005 (T)GCATG clusters
Exons (1.86%)
5'UTR (0.02%)
3'UTR (1.21%)
Distal Introns (76.34%)
Proximal Introns (2.83%)
Mixed (17.74%)
C
Mouse genes with (T)GCATG clusters
Genes with HITS-CLIP peaks in brain Weyn-Vanhentenryck et al. 2014
8348
4186
8356
D
Weyn-Vanhentenryck et al. 2014
8356 genes with Rbfox1 HITS-CLIP peaks
1687 (20%)
2499 (30%)
4170 (50%)
Genes with HITS-CLIP peaks on (T)GCATG clusters
Genes with HITS-CLIP peaks out of (T)GCATG clusters
Genes with HITS-CLIP peaks without (T)GCATG clusters
E
19977 (T)GCATG clusters
Exons (2.8%)
5'UTR (0.4%)
3'UTR (1.4%)
Distal Introns (77.1%)
Proximal Introns (2.2%)
Mixed (16.1%)
F
Genes with (T)GCATG clusters
Genes with eCLIP peaks
9826
2908
4903
G
4903 genes with Rbfox1 eCLIP peaks
1237 (25%)
1671 (34%)
1995 (41%)
Genes with eCLIP peaks on (T)GCATG clusters
Genes with eCLIP peaks out of (T)GCATG clusters
Genes with eCLIP peaks without (T)GCATG clusters
H
Number of differentially expressed isoforms (%)
Fold change (%)
Genes with eCLIP peaks on (T)GCATG clusters
Genes with eCLIP peaks out of (T)GCATG clusters
Genes with eCLIP peaks without (T)GCATG clusters
I
Log2FC Average
****
**
ns
Genes with eCLIP peaks on (T)GCATG clusters
Genes with eCLIP peaks out of (T)GCATG clusters
Genes with eCLIP peaks without (T)GCATG clusters

Fig. 4

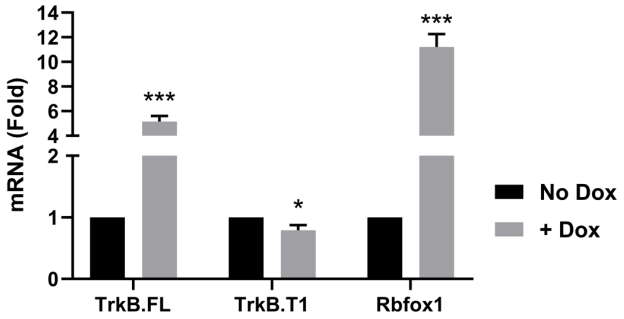
A



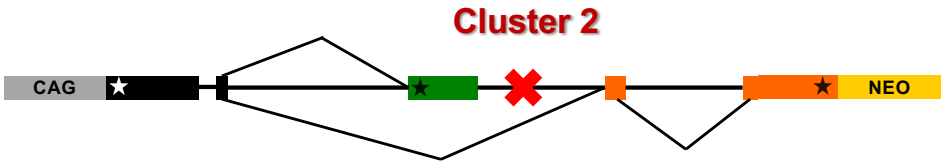
B



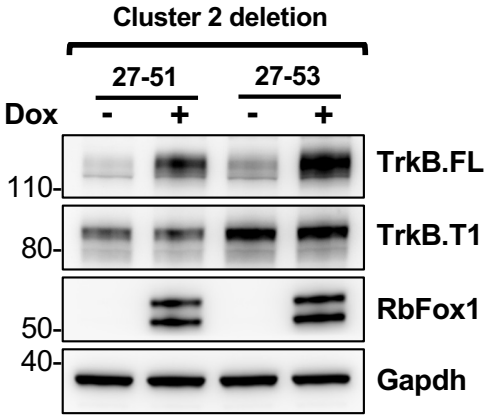
C



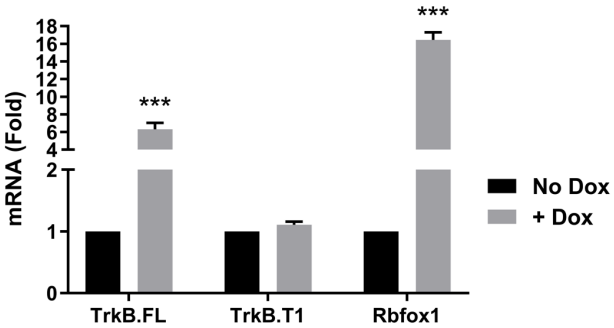
D



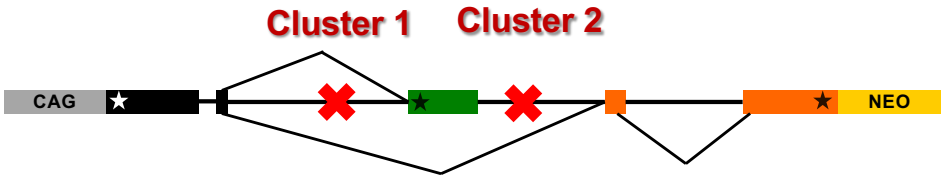
E



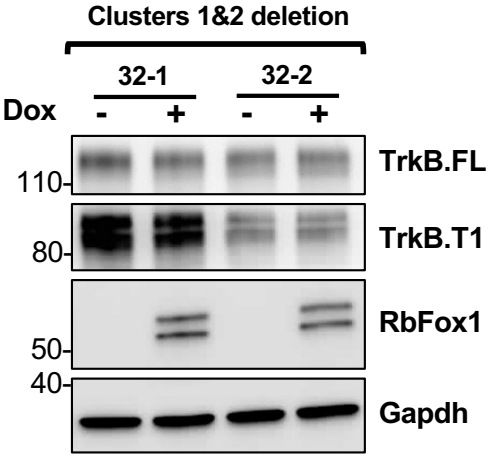
F



G



H



I

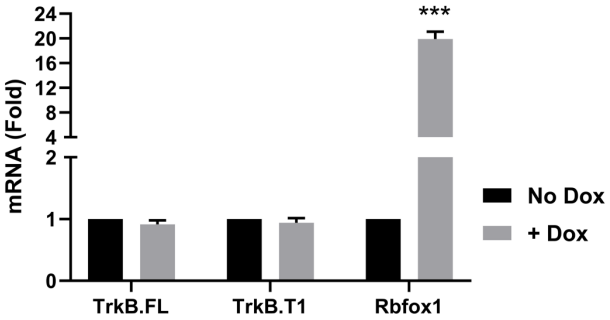


Fig. 5

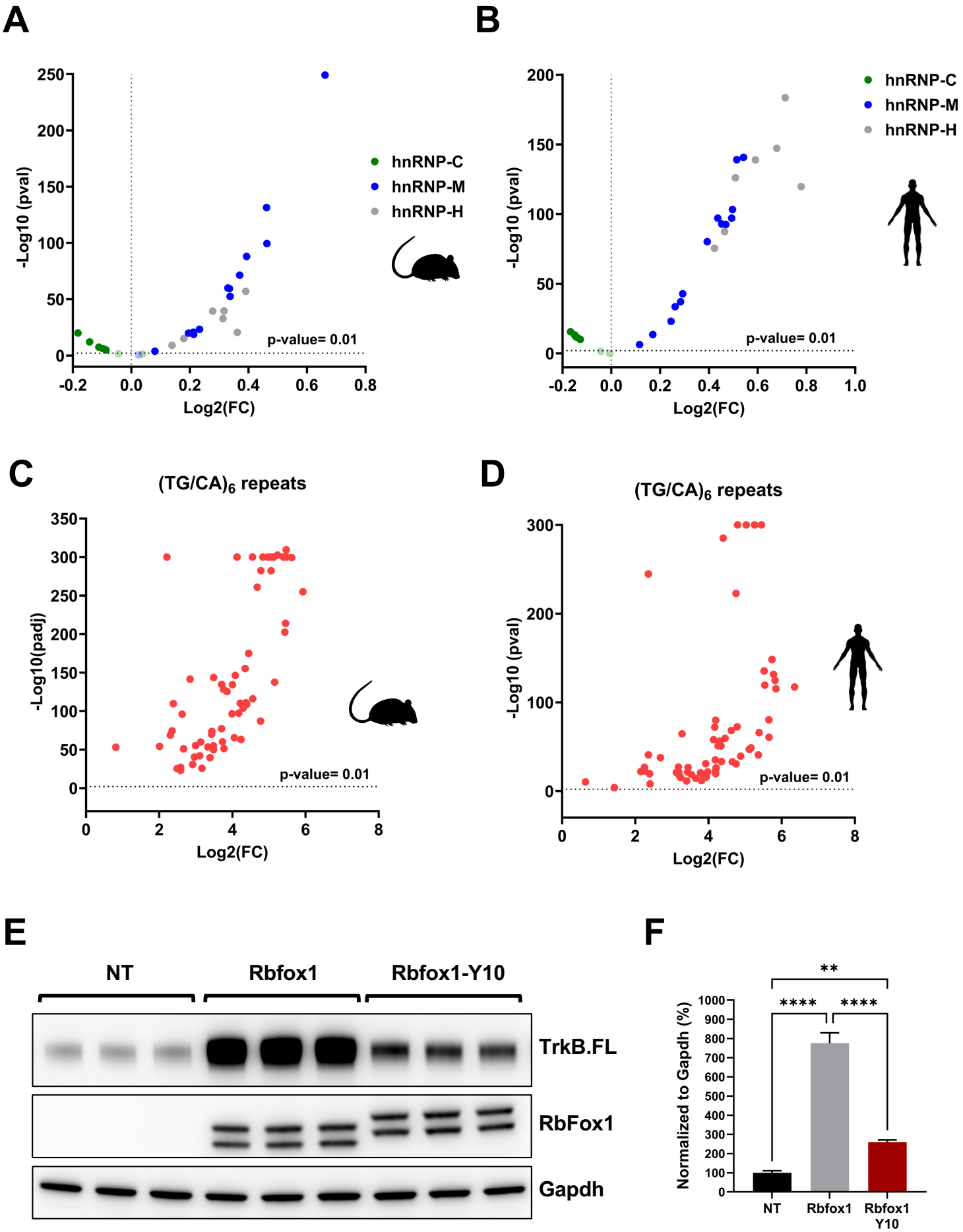
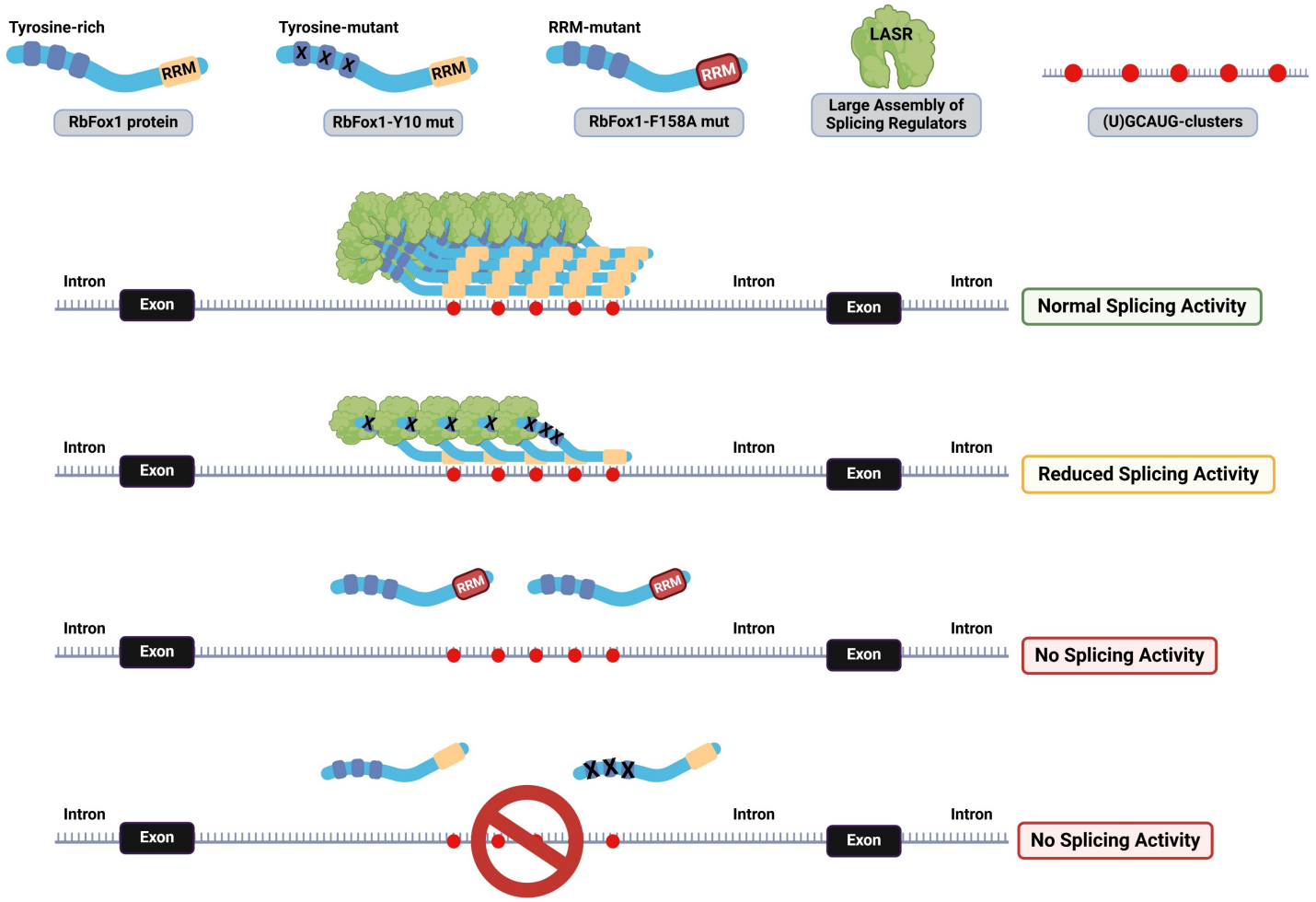

A
-Log10 (pval)
Log2(FC)
hnRNP-C
hnRNP-M
hnRNP-H
p-value= 0.01
B
-Log10 (pval)
Log2(FC)
hnRNP-C
hnRNP-M
hnRNP-H
p-value= 0.01
C
(TG/CA)6 repeats
-Log10(padj)
Log2(FC)
p-value= 0.01
D
(TG/CA)6 repeats
-Log10 (pval)
Log2(FC)
p-value= 0.01
E
NT
Rbfox1
Rbfox1-Y10
TrkB.FL
RbFox1
Gapdh
F
Normalized to Gapdh (%)
**
****
****
NT
Rbfox1
Rbfox1 Y10

Fig. 6


Tyrosine-rich
RRM
RbFox1 protein
Tyrosine-mutant
RRM
RbFox1-Y10 mut
RRM-mutant
RRM
RbFox1-F158A mut
LASR
Large Assembly of Splicing Regulators
(U)GCAUG-clusters
Intron
Exon
Intron
Exon
Intron
Normal Splicing Activity
Intron
Exon
Intron
Exon
Intron
Reduced Splicing Activity
RRM
RRM
Intron
Exon
Intron
Exon
Intron
No Splicing Activity
Intron
Exon
Intron
Exon
Intron
No Splicing Activity