Nat Struct Mol Biol. Author manuscript; available in PMC 2011 March 01.

Published in final edited form as:
Nat Struct Mol Biol. 2010 September ; 17(9): 1114–1123. doi:10.1038/nsmb.1881.

Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB

Miriam Llorian[1,6], **Schraga Schwartz**[2,6], **Tyson A. Clark**[3], **Dror Hollander**[2], **Lit-Yeen Tan**[1], **Rachel Spellman**[1,4], **Adele Gordon**[1], **Anthony C. Schweitzer**[3], **Pierre de la Grange**[4], **Gil Ast**[2], and **Christopher WJ Smith**[1]

[1]Department of Biochemistry, University of Cambridge 80, Tennis Court Road, Cambridge CB2 1GA, UK

[2]Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel-Aviv University, Ramat Aviv 69978, Israel

[3]Affymetrix, Inc., 3420 Central Expressway, Santa Clara, CA 95051

[4]GenoSplice technology, Centre Hayem, Hôpital Saint-Louis 1, avenue Claude Vellefaux, 75010, Paris, France

Abstract

To gain global insights into the role of the well-known repressive splicing regulator PTB we analyzed the consequences of PTB knockdown in HeLa cells using high-density oligonucleotide splice-sensitive microarrays. The major class of identified PTB-regulated splicing event was PTB-repressed cassette exons, but there was also a substantial number of PTB-activated splicing events. PTB repressed and activated exons showed a distinct arrangement of motifs with pyrimidine-rich motif enrichment within and upstream of repressed exons, but downstream of activated exons. The N-terminal half of PTB was sufficient to activate splicing when recruited downstream of a PTB-activated exon. Moreover, insertion of an upstream pyrimidine tract was sufficient to convert a PTB-activated to a PTB-repressed exon. Our results demonstrate that PTB, an archetypal splicing repressor, has variable splicing activity that predictably depends upon its binding location with respect to target exons.

Introduction

Alternative splicing (AS) is the major mechanism that accounts for the much higher complexity of mammalian proteomes compared with the number of protein-coding genes in their genomes. While much has been learned from the experimental dissection of model systems of alternative splicing, it is often difficult to disentangle the truly general properties that typify a particular program of regulated splicing from those that are peculiar to the particular model system being investigated. One of the major current goals of research upon alternative splicing is to try to decipher the underlying logic of coregulated programs of alternative splicing[1-5]. This type of approach has been enabled by the development of

cwjs1@cam.ac.uk.
Author contributions
ML and RS prepared materials for array analyses. TAC and ACS carried out the array analyses. ML, L-YT, AG validated array predictions. SS, DH, PG and GA carried out bioinformatic analyses. ML carried out all other wet experimental work. ML, SS and CWJS wrote the manuscript with input from other authors. CWJS conceived the project and coordinated the collaborations with important input from GA, ML, SS and TAC.
[6]The two first authors contributed equally.
[5]Current address: Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, Fulham Road, London, SW3 6JB

Europe PMC Funders Author Manuscripts

technologies such as splice-sensitive microarrays that allow profiling of large numbers of AS events (ASEs) in parallel. By comparing the transcriptomes from large numbers of tissues it is possible to define sets of coregulated exons and then to analyse them for the enrichment of sequence motifs in particular locations relative to the regulated exons[6-8]. A complementary approach is to analyze the contribution of individual splicing regulators to regulated splicing programs by profiling changes in AS after experimental perturbation of splicing regulator levels. We have used this approach to define a large set of AS events that are regulated by human Polypyrimidine Tract Binding protein (PTB) in HeLa cells.

PTB is an hnRNP protein that commonly acts as a repressive splicing regulator[9,10] as well as influencing other post-transcriptional steps in gene expression[11]. Structurally, PTB consists of four RNA Recognition Motif (RRM) domains, each of which recognise pyrimidine motifs[12]. The optimal binding site consists of pyrimidine motifs typified by UCUU[13,14]. Splicing silencer elements containing such motifs have been characterized in introns flanking PTB-repressed exons e.g.[15-18]. Proposed mechanisms of repression range from simple competition with U2AF65 binding at the polypyrimidine tract[19,20] to interference with cross-intron or cross-exon splicing complex assembly[21-23]. In addition various lines of evidence suggest that PTB-mediated looping of RNA between binding sites may play a role in its repressive activity[12,24,25]. On the other hand, just the second RRM domain and the following interdomain linker can induce exon skipping when artificially tethered adjacent to an exon[26]. Until recently, there were only sporadic reports of PTB activating some splicing events[27,28], possibly by antagonizing other repressors[27,29,30]. However, a CLIP-Seq global survey of PTB-binding RNAs in HeLa cells indicated that PTB-repressed and activated exons could be distinguished by the arrangement of PTB binding sites. PTB binding around regulated exons was associated with repression, while PTB binding in the region of the flanking constitutive exons was associated with activation of cassette exons[31].

Here we report the use of a high-density Affymetrix microarray, featuring probe-sets corresponding to all well annotated exons and exon-exon junctions in ~24,000 human genes, to identify ASEs affected by knockdown of PTB in HeLa cells. We identified 196 PTB-repressed events and 67 PTB-activated events. PTB-repressed and activated exons showed distinct arrangements of enriched sequence motifs. PTB-repressed exons showed enrichment of PTB-binding motifs both within the exons themselves as well as in the flanking upstream intron, but not the downstream intron. In stark contrast PTB-activated exons were associated with enrichment of PTB-binding motifs only within the downstream intron and were associated with weaker 5′ splice sites. Strikingly, we found a similar pattern of PTB CLIP-tag enrichment upstream of repressed exons but downstream of activated exons. Overall, our data indicate that PTB shows activity that is dependent on the location at which it binds relative to a target exon, similar to the Nova[32], Fox[33-35] and Mbnl[36] proteins,.

Results

PTB promotes both exon skipping and inclusion

PTB and nPTB were knocked down in HeLa cells[37], with levels of knockdown routinely over 80-90% (Fig. 1). Total RNA from three experimental and three control samples was used to generate target for hybridization to Affymetrix arrays featuring probesets corresponding to all exons and exon-exon junctions of approximately 24,000 human genes[38] (Fig. 1b). Alterations in AS between control and knockdown were discovered using the Splicing Index approach[38,39]. Candidate PTB-regulated ASEs were identified by association with one or more significant probesets. The resulting 226 ASEs were loosely grouped by confidence level taking into account the number of significant probesets, whether or not reciprocal changes were observed, and the level of cDNA sequence support for the event.

The HIGHEST confidence category included 26 events for which every associated probeset showed significant changes, while the HIGH confidence category contained 88 events for which reciprocal changes were observed. At the other extreme, the LOW category consisted of 10 events with only one supporting probeset and minimal cDNA sequence data supporting AS in that genomic region. In addition, a further 37 events were identified using target generated from all six knockdown and six control samples hybridized to the Affymetrix Human Exon 1.0 ST array, which features probesets to exons but not exon-junctions. Candidate ASEs were identified using the Splicing Index approach[39] and categorized into confidence groups using criteria similar to those used for the junction array data.

The full list of identified PTB-regulated events is in Supplementary data 1. Validation of array predictions was carried out by reverse transcription followed by quantitative real time PCR (qRT-PCR). For mutually exclusive splicing events where the two exons were identically sized, RT-PCR was followed by isoform specific restriction digestion (Fig. 2). Events were considered validated if a significant change (T-test, $p < 0.05$) was observed in the same direction as the array prediction. The validation rate for junction array predictions was 67/72 (93%). Even within the LOW category, 8 out of the 10 predictions were validated, while 9 of 10 MED LOW cases were validated. Thus the junction array analysis identified a large number of PTB-regulated AS events with a very low false-positive rate. While some genuine PTB-regulated ASEs will not have been identified (false-negatives), the set of PTB-regulated exons was suitable for further analysis.

The PTB-regulated events encompassed most categories of ASE except retained introns (Fig. 1c). They included 133 simple cassette exons, 27 exons that were parts of multiple cassettes, 18 mutually exclusive (ME) exons from a total of 13 ME pairs (reciprocal changes were observed in both members of 5 ME pairs), 32 alternative 3′ terminal exons, 5 alternative 5′ splice sites and 4 alternative 3′ splice sites. The majority of events (196 of 263) were upregulated upon PTB knockdown, consistent with PTB's conventional role as a splicing repressor[9,10]. The remaining 67 were down-regulated upon knockdown, implying that they are activated by PTB, as suggested previously[27,28]. Although some apparent PTB-activated exons may be indirect targets, subsequent motif-enrichment and functional analyses indicated that a substantial proportion are likely to be authentic direct PTB targets (see below).

PTB regulated mutually exclusive exons

ME splicing events are interesting from the perspective that some special mechanism is required to prevent the two exons from splicing together[40]. All the array predictions of PTB-regulated ME splicing were validated by RT-PCR (Fig.2 ref.[37] and data not shown). Of the 13 ME pairs, in 10 cases the exon upregulated upon PTB/nPTB knockdown was the downstream of the pair, while in 3 cases it was the upstream exon. In 2 cases (*C14ORF118* and one pair of *ACTN4* exons) the predicted branch point of the downstream exon was sufficiently close to the upstream exon to prevent the two exons splicing together. In 9 cases, splicing together of the two exons was predicted to lead to NMD, which can lead to the appearance of strict ME splicing. We tested whether PTB acted to enforce ME behavior of each of these exon pairs by carrying out knockdown of PTB and nPTB in combination with UPF1[41] (Fig. 2, and Supplementary Fig. 1). In 6 cases PTB/nPTB knockdown led to an increase in the double-included product. However, in no case did double inclusion become the major spliced product. For example, with *TPM1* exons 8 and 9, double inclusion increased from 0.8 to 5.8% while single inclusion of exon 9 increased from 3.5 to 47.3% (Fig. 2). Thus, while PTB acts to repress *TPM1* exon 9, it only partially restricts double splicing of exons 8 and 9, indicating that other mechanisms enforce ME splicing.

Tissue specificity of PTB-regulated exons

To enable bioinformatic analyses of the PTB-regulated exons we discarded all cases from the two lowest confidence groups which we had not validated experimentally, any complex events which could not be clearly classified, and ME exons that were down-regulated upon PTB knockdown. The resulting ASEs included 145 PTB repressed cassettes, 32 PTB activated cassettes, 21 PTB repressed alternative 3′ terminal events, and 10 PTB activated alternative 3′ terminal events. As a control dataset we used 4,892 cassette exons which showed no overlap with the PTB-dataset.

To investigate tissue-specificity of PTB regulated exons we extracted tissue-specific profiles based on data for 24,426 tissue-regulated ASEs across 48 human cell lines and tissues[6]. In this manner we obtained sets of tissue measurements for 1,520 control cassette exons, 69 PTB-repressed cassettes and 20 PTB-activated cassettes. We also extracted *PTB* and *nPTB* transcript levels across the tissues/cell lines. Unsupervised hierarchical clustering of tissues/cell lines according to the splicing profiles of control, PTB-activated and PTB-repressed cassettes yielded two distinct clusters: a cluster of brain and striated muscle (heart and skeletal), and one of all remaining tissues (Fig. 3). PTB-repressed cassettes were included to highest levels in brain and striated muscle. Conversely, PTB-activated cassettes were included at lower levels in these tissues compared to others (Fig. 3a). Moreover, there was a significant inverse correlation between mean inclusion levels of PTB-repressed cassettes and transcript levels of PTB ($R = 0.84$, $P = 2.5 \times 10^{-14}$) (Fig. 3b) across the 48 tissue/cell samples. Conversely, mean inclusion levels of PTB-activated cassettes showed a positive correlation with *PTB* levels ($R = 0.55$, $P = 3.8 \times 10^{-5}$). *nPTB* expression correlates inversely with *PTB*, consequently *nPTB* levels showed direct correlation with the inclusion levels of PTB-repressed cassettes and inverse correlation with PTB-activated cassettes. We repeated these analyses using Affymetrix ExonArray data for 11 human tissues[39], and obtained similar results (data not shown).

Consistent with the observed tissue specificity, gene ontology (GO) analysis[42] indicated that the most significantly enriched terms for molecular function and biological process were associated with the cytoskeleton, G-protein modulators and membrane traffic (Supplementary Table 1), similar to the enrichment terms noted for PTB-regulated events in N2A cells[27,43]. Among the PTB-regulated ASEs were cassette events within the genes for both PTB paralogs nPTB and *ROD1*[27,37,43], as well as in the genes of other splicing regulators such as *MBNL2* and *SRPK2* (Supplementary data file 1). Indeed, when GO analysis was repeated using a 3-fold larger set of predictions produced using ASPIRE[44], nucleic acids binding protein and pre-mRNA processing were significantly enriched terms, supporting previous observations that splicing regulators form complex regulatory networks[34]. A number of the PTB-regulated exons have also previously been identified to be misregulated in cancer, including mutually exclusive *PKM2*[45], *TPM1* and *ACTN1* exons[46], which may be related to elevated PTB levels in some cancers[47-49].

Features of PTB-regulated exons

The 208 PTB-regulated ASEs were analyzed for various splice site sequence features (Table 1). PTB-repressed cassettes and 3′ end exons differed from control and PTB-activated sets in having longer, higher scoring polypyrimidine tracts. They also had significantly longer AG dinucleotide exclusion zones upstream of the 3′ splice site, which is indicative of distant branch point location[50]. In contrast, PTB-activated cassettes had significantly weaker 5′ splice sites by two out of three separate measures of 5′ splice site strength.

We were next interested in deciphering the regulatory code of PTB. Though it is known that PTB binds sequences typified by ‘TCTT’, we applied a naïve approach and looked for

enrichment of k-mers, 4-7 nt RNA words, within the PTB-regulated exons and in the adjacent 200 nt of each flanking intron. We observed significant enrichment of UC-rich k-mers within PTB-repressed exons and in the 200 nt upstream intronic flank (Fig. 4a-b, left and middle panels). Clustering of enriched k-mers revealed classical optimal CU-rich PTB-binding motifs in the exons and upstream intronic flank (Fig. 4c, left and middle columns). In contrast, only 8 motifs were enriched within the downstream intronic flank of PTB-repressed exons. The two most significantly enriched motifs (UGCU and UGCATG) correspond to optimal binding sites for Mbnl[51], and Fox proteins[52], respectively (Fig 4b right panel). Two other enriched motifs also contained UGCU, and three others contained part of the FOX hexamer. Clustering of the 8 motifs produced a hybrid motif resembling both Fox and Mbnl sites (Fig. 4b,c right panels). To pursue this further we extracted 3,547 Fox2 CLIP tags[34], and looked for overlap with PTB-regulated and control exons. While 8.9% of PTB-repressed exons overlapped with Fox2 tags, this was true of only 1.6% of control exons, a 5.6 fold enrichment (exact Fisher test, $P=1.33\times10^{-6}$). Thus, PTB-repressed exons are significantly associated with observed Fox2 binding.

We next analyzed the density of PTB-binding sites at positions along the 200 nt of flanking intron and within the exons themselves. The analysis included flanking constitutive exons and their adjacent 200 nt intron flank but we found no major enrichments in these areas (Supplementary Fig. 2). We defined PTB binding sites either as any stretch consisting of YTCY or YCTY (where Y stands for pyrimidine) or by matches to $YCTN_{1\text{-}6}CTN_{3\text{-}8}YCT$ (N represents any nucleotide)[12]. Essentially identical results were obtained using the two methods. A peak of PTB sites was observed within the first 20 nt upstream of control exons. In contrast, the PTB repressed exons showed a higher density of PTB sites extending further upstream, approaching background levels at ~−120 nt (Fig. 4d, left panel). The density of PTB sites was higher within PTB-repressed exons than control exons with no obvious peaks (Fig. 4d, centre panel). On the downstream side of repressed exons, the density of PTB sites was the same as for control cassettes. Similar to the cassette exons, 3′-end PTB repressed exons (but not 3′-end activated exons) also showed an elevated density of PTB sites in both the upstream intron flank and the exons themselves (data not shown). The highest density of downstream PTB motifs was associated with PTB-activated cassettes (Table 1), although this apparent enrichment did not pass a test of significance ($p = 0.063$). When we restricted analysis to 100 nt adjacent to the exon the enrichment of downstream PTB motifs was significant ($p = 0.02$). To pursue this further we looked for k-mer enrichment within the PTB-activated exons and their flanking regions (Fig. 5). Ten k-mers were significantly enriched on the downstream side of the PTB-activated exons, seven of which were pyrimidine-rich. The most significantly enriched motif (1.8 fold, $p = 2.96 \times 10^{-6}$) was the optimal PTB-binding motif TCTT[18], and clustering of enriched motifs revealed a similar motif (Fig. 5c). This suggests that exons that are downregulated upon PTB/nPTB knockdown are authentic targets of activation by PTB. The enrichment of PTB sites at this location could be connected with weaker 5′ splice sites of PTB-activated exons (Table 1). No motif enrichment was observed within the exons, while on the upstream side the enriched motifs resembled a CELF protein binding site[53].

PTB represses via exon splicing silencers

The observed enrichment of PTB-binding motifs within, but not downstream of, PTB-repressed exons was unexpected. We therefore analyzed individual cases with predicted PTB-dependent Exon Splicing Silencers (ESS). *ANXA7* exon 6 has extended pyrimidine tracts with a number of potential PTB binding motifs (Fig. 6a). The exon along with 200 nt of flanking introns was cloned into a GFP-based exon-trapping vector. Constructs were transfected into HeLa cells and splicing analyzed by RT-PCR under normal conditions and in response to PTB/nPTB knockdown. Inclusion of *ANXA7* exon 6 was strongly PTB-

dependent, increasing from 3.9 to 67% inclusion upon PTB/nPTB knockdown (Fig. 6b). Three point mutations in the exon were designed to impair PTB-binding, while avoiding the creation of splicing enhancers[54]. UV crosslinking of exon probes in HeLa nuclear extract showed that these mutations abolished PTB crosslinking (Fig. 6c, immunoprecipitation of PTB not shown). Electrophoretic mobility shift assay (EMSA) and supershift with PTB antibodies showed that the major gel-shifted complex contained PTB (Fig. 6d,e). Moreover, this complex formed on the mutant RNA only with higher levels of nuclear extract, indicating lower affinity for PTB (Fig. 6d). Consistent with the exonic PTB binding site acting as an ESS, the mutations caused an increase of exon inclusion from 3.9 to 46% (Fig. 6b). The mutant construct remained responsive to PTB/nPTB knockdown, presumably reflecting the presence of additional PTB binding sites within and upstream of the exon. Similar observations were made with an exon in the *MTDH* gene (data not shown). These data support the prediction from motif enrichment that PTB commonly acts as a repressor by binding to ESS's.

PTB activation via downstream ISE

A key insight from the bioinformatic analyses was the enrichment of PTB-binding motifs downstream of PTB activated exons. To investigate the possible role of downstream PTB-binding ISEs we selected an exon (chr14:55,209,643-55,209,726) from the *KTN1* gene (Fig. 7a), which showed decreased inclusion upon PTB knockdown (Fig. 7b). This effect was reproduced in the exon-trapping vector, with a 94 to 75% decrease (Fig. 7c, WT construct). We identified two pyrimidine-rich blocks, Py1 and Py2, downstream of the exon (Fig. 7a). Deletion of Py1 (ΔPy1) had very little effect upon exon inclusion under control conditions. In contrast, mutant ΔPy2 had lower levels of exon inclusion (~75%) and was unaffected by PTB knockdown. This indicates that the Py2 element has PTB-dependent ISE activity. The combined deletion ΔPy1,2 further reduced exon inclusion to 37%, showing that in the absence of the Py2 ISE, Py1 can also activate exon inclusion.

A number of well studied PTB-repressed exons have PTB-binding elements on both sides of the exon[15,17,55], while the *KTN1* exon has only a downstream site. We next asked if we could convert a PTB-activated to a PTB-repressed exon by inserting an upstream PTB-binding element. To this end we created the PyUP series of mutants in which a pyrimidine rich tract (from *ANXA7*, Fig. 6) or a control sequence was inserted 68 nt upstream of the *KTN1* exon. These insertions in the upstream intron were combined with either wild type, ΔPy1, ΔPy2 or ΔPy1, 2 versions of the downstream intron. Insertion of the PyUP, but not the control sequence, led to 70% exon skipping, consistent with insertion of an upstream ISS (Fig. 7d, lanes 1-3). The same effect was seen when the upstream insertions were combined with the ΔPy1 downstream intron (Fig. 7d lanes 4-6). The effect of the PyUP insertion was less dramatic in the context of the ΔPy2 downstream intron (Fig. 7d, lanes 7-9), suggesting that the full activity of the PyUP ISS is partially dependent upon Py2. In the context of the ΔPy1,2 downstream intron, insertion of PyUP had very little effect (Fig 7d, lanes 10-12), consistent with previous reports that PTB-mediated repression frequently requires more than one PTB-binding element[15,24,25]. Moreover, while the Py2 element has ISE activity in the context of the wild type or nrs insertion in the upstream intron (Fig 7d compare lanes 1 with 7, and 3 with 9), in the context of the PyUP insertion Py2 has silencer activity, albeit modest (Fig 7d, compare lanes 2 and 8).

We tested the dependency of the PyUP series of constructs on PTB by RNAi of PTB/nPTB (Fig. 7e). While the WT and nrs UP constructs showed decreased exon inclusion upon PTB knockdown (Fig. 7e lanes 1,2,5,6) the constructs with PyUP were, as expected, repressed by PTB (lanes 3,4,9,10). Constructs without upstream or downstream PTB sites were unresponsive to PTB knockdown (lanes 7,8,11,12). Paradoxically, construct PyUP_ΔPy1,2, responded to PTB knockdown with increased exon inclusion (Fig 7e, lanes 9,10), even

though ΔPy1,2 itself did not respond to PTB knockdown (lanes 7,8) and insertion of PyUP into ΔPy1,2 (lanes 7,9) did not lead to increased exon skipping. Formally, comparing ΔPy1,2 with PyUP_ΔPy1,2 the PyUP element has no activity in control conditions (lanes 7,9) but acts as an ISE under PTB knockdown conditions (lanes 8, 10). A possible explanation for the behavior of PyUP_ΔPy1,2, is that PTB binds to the PyUP element but does not repress, consistent with the requirement for more than one PTB binding element for repression[15]. However, in the absence of PTB the element may be able to bind one or more other proteins that cause activation of splicing. Nevertheless, taken together, these results suggest that the activity of an upstream PTB binding site has a repressive activity that is dominant over downstream PTB-dependent ISEs.

A minimal PTB splicing activator domain

We next tested whether downstream activation by PTB could be restored by artificial tethering using MS2 coat protein fusions (Fig. 8). Replacement of Py2 with an MS2 coat protein binding hairpin ΔPy2xMS2 reduced exon inclusion from 95 to 63%, a slightly larger effect than simple deletion of Py2 (Fig. 8b, lanes 1-3). Cotransfection with PTB4-MS2 increased exon inclusion from 63 to 83% (lane 4) while MS2 coat protein alone had no effect (Fig. 8b, lane 9). A fusion of MS2 to the splicing repressor hnRNPA1[56] led to enhanced exon skipping (lane 10), demonstrating the specificity of activation by PTB recruitment. We next tested four deletion mutants of PTB fused to MS2 (Fig. 8a), all of which were expressed to similar levels (Fig. 8b, bottom). The mutations comprised deletions of two or more of the RRM domains[12], as well as the interdomain linker between RRMs 2 and 3, which contains the 26 amino acid insert characteristic of the PTB4 isoform. Strikingly, fusion proteins RRM12L and RRM2L both increased exon inclusion similar to full length PTB (Fig. 8b lanes 5,7). In contrast, RRM12 was without effect (lane 6), demonstrating the importance of the linker between RRMs 2 and 3. The C-terminal RRM34 had an activity opposite to that of full length PTB, producing slightly lower levels of exon inclusion (lane 8). Thus the minimal domain of PTB that is sufficient to activate the *KTN1* exon is RRM2 combined with the following interdomain linker. Strikingly, this is identical to the minimal domain that acted as a splicing repressor domain when tethered adjacent to *Tpm1* exon 3[26].

Discussion

The most striking insight provided by our analyses is that, like Nova[32], Fox[33-35] and Mbnl[36,51] proteins, PTB shows position-dependent splicing activity (Fig. 9). When binding upstream or within a cassette exon it acts as a repressor, but when bound only on the downstream side it activates inclusion. The position-dependent activity correlates with splice site features. PTB-activated exons tend to have weaker 5′ splice sites than control cassette exons, while PTB-repressed exons have longer and stronger polypyrimidine tracts (Table 1).

Our results contrast with a recent UV Cross-Linking and Immunoprecipitation (CLIP) analysis of PTB targets in HeLa cells. Xue *et al.*[31] also found that the position of PTB binding differed markedly between PTB-repressed and activated cassette exons, but the resultant activity map and model for PTB-activation differ from ours. For PTB-repressed exons they observed enrichment of PTB binding in both flanking introns, but not within the exons themselves. For PTB-activated exons, binding was associated with the flanking constitutive exons. This suggested a model in which activation by PTB results from its familiar repressor activity, with partial kinetic inhibition of the flanking constitutive sites promoting cassette exon inclusion[31]. This contrasts with our model of direct activation by binding to adjacent downstream ISEs. The reasons for the differences between the two sets of results are not entirely clear. Our analyses are based upon a larger number of PTB-regulated exons (177 compared to 55). The differences may also relate to the methods used

to identify PTB-regulated events. The CLIP based pipeline produced a set of 22 validated PTB-regulated ASEs from the 10,372 genes containing PTB CLIP tags. Of these, nearly half (10/22) were activated by PTB, which contrasts with the generally higher prevalence of PTB repression, and suggests that the CLIP pipeline may have an unintentional inbuilt bias that leads to identification of a subset of PTB-regulated ASEs. Consistent with this suggestion, only 14 PTB-repressed exons, and not a single PTB-activated exon, were shared between the two data-sets.

To address the differences between the two experimental approaches and datasets, we combined the CLIP-seq data[31] (Gene Expression Omnibus accession GSE19323) with our datasets of regulated ASEs. Mean number of reads per nucleotide were calculated across our three datasets of PTB repressed, PTB activated and control exons (Fig. 9a, upper panel). The map of CLIP tag density closely resembled the motif enrichment that we had already observed (Supplementary Fig 2, Fig 9a). To ensure that these enrichments did not arise from a small number of highly expressed transcripts we calculated the percentage of exons in each class with at least one CLIP tag in the corresponding transcript region. The most significant enrichments were again observed immediately downstream of activated exons, upstream of repressed exons and within repressed exons (Fig. 9a, lower panel). In addition, many other regions of the PTB-regulated exons showed increased incidence of PTB-binding compared to control cassette exons. Finally, we determined the percentage of Xue et al's[31] PTB-regulated exons with at least one CLIP tag within or adjacent to the exon (Supplementary Fig. 4). Again, the highest incidence of CLIP tags occurred immediately downstream of PTB-activated exons. Additional support for our model of direct activation by PTB is also provided by the *KTN1* minigene constructs in which all the sequence elements necessary for PTB activation were shown to lie within 200 nt of the exon (Figs 7, 8). Nevertheless, it appears that this archetypal splicing repressor is able to activate cassette exons in at least two ways; by kinetically slowing splicing at flanking constitutive splice sites[31], as well as by direct activation.

The fact that PTB-activated exons have weaker 5′ splice sites (Table 1) suggests that PTB may promote U1 snRNP binding. Indeed, activation of U1 snRNP binding has been suggested as a general mechanism for downstream ISEs[57]. In the case of activation by TIA1 a direct interaction with U1C protein assists U1 snRNP binding[58]. A striking finding which should provide clues to the mechanism of activation was the identification of a minimal activation domain of PTB consisting of the second RRM domain and the following interdomain linker (Fig. 8). This is identical to the minimal repressor domain when tethered downstream of *Tpm1* exon 3[26]. RRM2 interacts not only with RNA[12] but also with [S/G][I/L]LGxxP peptide motifs in the corepressor protein Raver1[59]. It will be of interest to determine whether the minimal domain uses the same molecular surfaces and interacting partners for activation and repression. An alternative possibility is that RRM2 may interact with other proteins containing [S/G][I/L]LGxxP motifs. By characterizing minimal effector domains and their functional targets we should be able to address the more general issue of why regulators such as PTB, Fox and Nova activate splicing when bound downstream while repressing when bound upstream.

We also gained insights into the intensively investigated repressor activity of PTB. Analysis of numerous PTB-regulated exons has identified PTB-binding silencer elements most frequently in upstream introns, but also in downstream introns and within the exons themselves[10]. Analysis of large data-sets (>20,000) of tissue-regulated exons has shown enrichment of PTB-like motifs up to 150 nt upstream of exons that are activated in striated muscle and brain, but not within or downstream of the exons[6,8]. The enrichment of PTB-motifs within PTB-repressed exons (Figure 4) is consistent with the general depletion of PTB-binding from exons revealed by genomic SELEX[14]. In contrast to our findings, several

well-studied model PTB-regulated exons have PTB-binding sites flanking the exons[15,17,55], and models for PTB-mediated repression consequently often involve looping out of entire exons between PTB bound at flanking sites[10,12]. Our data suggests that such looping might more commonly occur between the exon and upstream intron rather than between the flanking introns.

PTB has been well characterized as a repressor of neuronal specific exons[27,43]. By combining our set of PTB-regulated exons with published exon-level expression data[6] we found that exons that are regulated by PTB show not only neuronal but also striated muscle specificity (Fig.3). This adds direct evidence to the association that was previously noted between muscle and brain specific exons and enrichment of PTB binding sites upstream of such exons[5-8]. Regulation by PTB contrasts with the Fox proteins, which are also associated with neuron and muscle specific splicing, but which are expressed at high levels in those tissues. In contrast, it is lower levels of PTB that contribute to muscle and brain specific exon inclusion. The association of PTB-repressed exons with downstream Fox and Mbnl motifs (Figure 5), a position associated with activation of splicing[6,8,33-36,51] indicates that a significant subset of muscle and neuron specific exons are co-regulated by withdrawal of PTB-mediated repression combined with activation by Fox or Mbnl proteins.

In identifying PTB targets in HeLa cells we needed to knock-down not just PTB but also nPTB, which is upregulated upon PTB knockdown and in proteomic analyses compensates for the loss of PTB[37]. However, PTB and nPTB are not fully redundant. The replacement of PTB by nPTB is responsible for a substantial fraction of AS changes that occur during neurogenic differentiation[27,43]. Boutz et al[27] were able to classify exons in mouse N2A cells that were regulated by either PTB or nPTB or both proteins. It will be of future interest to determine whether these different classes of exons show distinct activity maps for PTB and nPTB binding.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a Wellcome Trust programme grant to CWJS (077877), EC grant EURASNET-LSHG-CT-2005-518238 (CWJS and GA). GA is supported by grants from the Israel Science Foundation (ISF 61/09), Joint Germany-Israeli Research Program (ca-139), Deutsche-Israel Project (DIP MI-1317), the Israel Cancer Association, and the Israel Cancer Research Foundation (ICRF). PG was supported by AFM (“Association Française contre les Myopathies”) and EC grant EURASNET-LSHG-CT-2005-518238. S.S. and D.H. are fellows of the Edmond J. Safra bioinformatics program at Tel Aviv University. This work was performed in partial fulfillment of the requirements for a PhD degree of S.S. and D.H., Sackler Faculty of Medicine, Tel Aviv University, Israel. We thank Martina Hallegger, Nick McGlincy and Jernej Ule for comments on the manuscript.

Appendix

Methods

Tissue culture, cell extracts and analysis

HeLa S3 cells were cultured and transfected by standard procedures[37,60]. Protein lysates were obtained by direct lysis in SDS loading buffer and equivalent amounts separated on 15% acrylamide gels. RNA was harvested with TRI reagent (SIGMA) following manufacturer’s instructions. Reverse Transcriptase reactions were carried out using 1μg of total RNA, oligodT or a minigene specific RT-primer CGRT and Superscript II (Invitrogen) following manufacturer’s instructions. Transiently expessed transcripts were detected with 10 pmol of CG5′ primer, 4 pmol ^{32}P-labeled GFPN1 primer, under the following

conditions: 94°C 30s, 56°C 30s, 72°C 30 s, 35 cycles. Samples were separated on 8M urea gels, and quantitated by Molecular Dynamics Phosphorimager. Data shown is mean ± s.d., n=3. Microarray validation was carried out by qPCR with SensiMix *NoRef* DNA Kit (Quantace) in a Corbett Rotor Gene 6000 series detector[61]. Data acquisitions were performed with Rotor Gene 6000 Series software (version 1.7). Relative quantification analysis was performed using a dynamic amplification efficiency determination for each amplification run as provided in the Comparative Quantification method. Melting curve analysis and gel electrophoresis verified that a single product had been formed. A full list of primers is in Supplementary file 2. Gel mobility shift and UV crosslinking assays were carried out as described previously[17,61]. Anti-PTB rabbit polyclonal was obtained from[62], anti-PTB mouse monoclonal (BB7) from DL Black, anti-nPTB rabbit polyclonal from A. Willis, anti-ERK from Santa Cruz, and anti-Actin A2066 from SIGMA. Anti-MS2 was raised against recombinant full length MS2 coat protein.

DNA Constructs

ANXA7 exon 6; *KTN1* exon (chr14:55,209,643-55,209,726) and approximately 200 nt of its corresponding native flanking introns were cloned into the Kpn I/Sph I or Not I/Sph I sites of pCAGGPNS vector[61,63]. Mutations were created by PCR mediated mutagenesis using Stratagene native Pfu polymerase and verified by DNA sequencing. For the ESS mutations in *ANXA7* exon 6 we used the ESR prediction tool (http://esrsearch.tau.ac.il/)[54] to avoid creating potential ESEs. MS2 hairpins were introduced in EcoRV sites 80 nt downstream of the 5′ splice site of *KTN1* exon. Pyrimidine rich sequences from the *ANXA7* exon or an unrelated sequence were inserted 68 nt upstream of the 3′ splice site of *KTN1* exon. PTB-MS2 fusion constructs were from [26].

Microarrays

Labeled target was prepared from 1 μg of Total RNA using Affymetrix WT Sense Target Labeling Kit and hybridized overnight (16 hours) to either the Human Exon 1.0 ST Array or the non-commercial human exon-junction array (HJAY)[38]. Array data was sketch-quantile normalized and PLIER was used for probeset summarization. The Splicing Index algorithm was used to identify probesets with statistically significant differences in exon-inclusion rate between control and knockdown samples using cut-offs of $p > 0.001$ (T-Test) and a log2 magnitude of $> |0.5|$[38,39]. Probesets were mapped to the UCSC human genome browser to observe genomic context. ASEs were loosely classified into six confidence categories (HIGHEST, HIGH, MED HIGH, MED, MED LOW, LOW) based on the number of probesets with significant changes, whether or not reciprocal changes were observed, and the extent of cDNA/EST sequence data supporting alternative splicing at that genomic locus.

Bioinformatic analyses

Conservation measures were assessed by phastCons scores for 18 placental organisms, downloaded from the UCSC browser. Gene Ontology analysis was carried out using PANTHER[42]. For analyses of PTB-regulated exons we discarded all cases from the two lowest confidence groups which we had not validated experimentally, and also complex events which could not be clearly classified. This yielded 145 PTB repressed cassettes, 32 PTB activated cassettes, 21 PTB repressed alternative 3′ terminal events, and 10 PTB activated alternative 3′ terminal events. As controls we obtained a dataset of 4,965 cassette exons from EST data[64]. After filtering exons also present in the PTB dataset, 4,892 control cassettes remained.

Scoring of splicing signals

5′SS, defined as the 3 terminal exonic and the first 6 intronic nucleotides, were scored by the Shapiro and Senapathy method[65], free energy (ΔG°) of binding between U1 snRNA and the 5′SS[66] and the maximum entropy method[67]. The 3′SS, defined as the last 20 intronic and the first two exonic nucleotides, was scored using position specific scoring matrix (PSSM) based log-odd scores, which were normalized to the range 0 to 100[64]. (C) PPTs were identified within the last 50 intronic nucleotides and scored as in[68]. AG exclusion zones were scored as the distance between the terminal AG dinucleotide and the most adjacent AG upstream. PTB binding sites were defined either as any YTCY or YCTY motif, or as $YCTN_{1\text{-}6}CTN_{3\text{-}8}YCT$[12]. FOX2 binding sites were defined as TGCATG.

Identification and clustering of enriched motifs

Significantly enriched motifs within each of the four PTB regulated event types across three segments (200 upstream intronic nucleotides, exon, and 200 downstream intronic nucleotides) were identified by counting occurrences of each 4-6 nt k-mer in the PTB regulated group and in the corresponding control group. For each k-mer length, we also counted the total number of k-mers of that length within each of the groups. The ratio between the prevalences was used to calculate the fold change between the two groups. We used the exact Fisher test to test the hypothesis that the prevalence of a k-mer was not different between the two groups. We applied a FDR correction for multiple testing and demanded that FDR < 0.05.

Clustering of k-mers relies on a distance metric between two k-mers, which relies on finding the optimum alignment. To optimally align two k-mers, we exhaustively went through all combinations of aligning one k-mer against the other without introducing gaps within the sequence, but allowing gaps at the termini. Each alignment was assigned a score, calculated as the sum of the number of mismatches between all non-gap positions plus the minimum of the number of gaps at the termini. The final distance between two k-mers was then defined as the minimum of the possible alignment scores. To cluster a set of k-mers, we began by generating an all-against-all distance matrix, based on the distance metric. We next iteratively clustered kmers, using the following steps as long as there are more than 5 kmers in the matrix: (1) We calculate the sum of all columns, and select the column with the minimal sum.

This indicates that the kmer in this column is closest to all other remaining kmers, and this kmer is therefore defined as a seed. (2) We extract all kmers whose distance (according to our scoring) is within k from the seed, and together these form a cluster. We used $k=1$, although we found that for our data the clustering was not very sensitive to values of k and similar results were produced using other values. (3) We align all the kmers in the cluster against the seed, using the above defined alignment procedure. (4) We then coverted this alignment into a position specific scoring matrix (PSSM), indicating the frequency of each nucleotide at each position. This PSSM was then visualized using the SeqLogo package[69] in R[70]. (5) All kmers within this cluster are then removed from the initial matrix.

Tissue profiling of PTB regulated exons

We used published microarray data for 24,426 ASEs across 48 tissues and cell lines[6]. For each event, this dataset comprises a set of 48 values, indicative of the change in relative abundance of the splice forms (as change in percent of splice form composition) across the tissues and cell lines. From this dataset we extracted a subset of 7,795 single cassette exons. We merged our dataset of PTB repressed and activated cassettes, as well our dataset of control cassettes, with the tissue specific profiles. We thereby obtained tissue measurements for 1,520 control cassette exons, 69 PTB repressed cassettes and 20 PTB activated cassettes.

References

1. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol. 2009; 10:741–54. [PubMed: 19773805]
2. Matlin AJ, Clark F, Smith CW. Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol. 2005; 6:386–98. [PubMed: 15956978]
3. Blencowe BJ. Alternative splicing: new insights from global analyses. Cell. 2006; 126:37–47. [PubMed: 16839875]
4. Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA. 2008; 14:802–13. [PubMed: 18369186]
5. Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ. Deciphering the splicing code. Nature. 2010; 465:53–9. [PubMed: 20445623]
6. Castle JC, Zhang C, Shah JK, Kulkarni AV, Kalsotra A, Cooper TA, Johnson JM. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. Nat Genet. 2008; 40:1416–25. [PubMed: 18978788]
7. Kalsotra A, Xiao X, Ward AJ, Castle JC, Johnson JM, Burge CB, Cooper TA. A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. Proc Natl Acad Sci U S A. 2008; 105:20333–8. [PubMed: 19075228]
8. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008; 456:470–6. [PubMed: 18978772]
9. Wagner EJ, Garcia-Blanco MA. Polypyrimidine tract binding protein antagonizes exon definition. Mol Cell Biol. 2001; 21:3281–8. [PubMed: 11313454]
10. Spellman R, Smith CW. Novel modes of splicing repression by PTB. Trends Biochem Sci. 2006; 31:73–6. [PubMed: 16403634]
11. Sawicka K, Bushell M, Spriggs KA, Willis AE. Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. Biochem Soc Trans. 2008; 36:641–7. [PubMed: 18631133]
12. Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, Reymond L, Amir-Ahmady B, Pitsch S, Black DL, Allain FH. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. Science. 2005; 309:2054–7. [PubMed: 16179478]
13. Perez I, McAfee JG, Patton JG. Multiple RRMs contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. Biochemistry. 1997; 36:11881–90. [PubMed: 9305981]
14. Reid DC, Chang BL, Gunderson SI, Alpert L, Thompson WA, Fairbrother WG. Next-generation SELEX identifies sequence and structural determinants of splicing factor binding in human pre-mRNA sequence. RNA. 2009; 15:2385–97. [PubMed: 19861426]
15. Amir-Ahmady B, Boutz PL, Markovtsov V, Phillips ML, Black DL. Exon repression by polypyrimidine tract binding protein. RNA. 2005; 11:699–716. [PubMed: 15840818]
16. Ashiya M, Grabowski PJ. A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. RNA. 1997; 3:996–1015. [PubMed: 9292499]
17. Gooding C, Roberts GC, Smith CW. Role of an inhibitory pyrimidine element and polypyrimidine tract binding protein in repression of a regulated alpha-tropomyosin exon. Rna. 1998; 4:85–100. [PubMed: 9436911]
18. Perez I, Lin CH, McAfee JG, Patton JG. Mutation of PTB binding sites causes misregulation of alternative 3′ splice site selection in vivo. RNA. 1997; 3:764–78. [PubMed: 9214659]
19. Lin CH, Patton JG. Regulation of alternative 3′ splice site selection by constitutive splicing factors. RNA. 1995; 1:234–45. [PubMed: 7489496]
20. Singh R, Valcarcel J, Green MR. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. Science. 1995; 268:1173–6. [PubMed: 7761834]
21. Izquierdo JM, Majos N, Bonnal S, Martinez C, Castelo R, Guigo R, Bilbao D, Valcarcel J. Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. Mol Cell. 2005; 19:475–84. [PubMed: 16109372]

22. Sharma S, Falick AM, Black DL. Polypyrimidine tract binding protein blocks the 5′ splice site-dependent assembly of U2AF and the prespliceosomal E complex. Mol Cell. 2005; 19:485–96. [PubMed: 16109373]

23. Sharma S, Kohlstaedt LA, Damianov A, Rio DC, Black DL. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. Nat Struct Mol Biol. 2008; 15:183–91. [PubMed: 18193060]

24. Chou MY, Underwood JG, Nikolic J, Luu MH, Black DL. Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. Mol Cell. 2000; 5:949–57. [PubMed: 10911989]

25. Matlin AJ, Southby J, Gooding C, Smith CW. Repression of alpha-actinin SM exon splicing by assisted binding of PTB to the polypyrimidine tract. RNA. 2007; 13:1214–23. [PubMed: 17592047]

26. Robinson F, Smith CW. A splicing repressor domain in polypyrimidine tract-binding protein. J Biol Chem. 2006; 281:800–6. [PubMed: 16282332]

27. Boutz PL, Stoilov P, Li Q, Lin CH, Chawla G, Ostrow K, Shiue L, Ares M Jr. Black DL. A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev. 2007; 21:1636–52. [PubMed: 17606642]

28. Xing Y, Stoilov P, Kapur K, Han A, Jiang H, Shen S, Black DL, Wong WH. MADS: a new and improved method for analysis of differential alternative splicing by exon-tiling microarrays. RNA. 2008; 14:1470–9. [PubMed: 18566192]

29. Lou H, Helfman DM, Gagel RF, Berget SM. Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3′-terminal exon. Mol Cell Biol. 1999; 19:78–85. [PubMed: 9858533]

30. Paradis C, Cloutier P, Shkreta L, Toutant J, Klarskov K, Chabot B. hnRNP I/PTB can antagonize the splicing repressor activity of SRp30c. RNA. 2007; 13:1287–300. [PubMed: 17548433]

31. Xue Y, Zhou Y, Wu T, Zhu T, Ji X, Kwon Y-S, Zhang C, Yeo G, Black DL, Sun H, Fu X-D, Zhang Y. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. Mol Cell. 2009; 36:996–1006. [PubMed: 20064465]

32. Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB. An RNA map predicting Nova-dependent splicing regulation. Nature. 2006; 444:580–6. [PubMed: 17065982]

33. Venables JP, Klinck R, Koh C, Gervais-Bird J, Bramard A, Inkel L, Durand M, Couture S, Froehlich U, Lapointe E, Lucier JF, Thibault P, Rancourt C, Tremblay K, Prinos P, Chabot B, Elela SA. Cancer-associated regulation of alternative splicing. Nat Struct Mol Biol. 2009

34. Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat Struct Mol Biol. 2009; 16:130–7. [PubMed: 19136955]

35. Zhang C, Zhang Z, Castle J, Sun S, Johnson J, Krainer AR, Zhang MQ. Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. Genes Dev. 2008; 22:2550–63. [PubMed: 18794351]

36. Du H, Cline MS, Osborne RJ, Tuttle DL, Clark TA, Donohue JP, Hall MP, Shiue L, Swanson MS, Thornton CA, Ares M Jr. Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. Nat Struct Mol Biol. 2010; 17:187–93. [PubMed: 20098426]

37. Spellman R, Llorian M, Smith CW. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. Mol Cell. 2007; 27:420–34. [PubMed: 17679092]

38. Yamamoto ML, Clark TA, Gee SL, Kang JA, Schweitzer AC, Wickrema A, Conboy JG. Alternative pre-mRNA splicing switches modulate gene expression in late erythropoiesis. Blood. 2009; 113:3363–70. [PubMed: 19196664]

39. Clark TA, Schweitzer AC, Chen TX, Staples MK, Lu G, Wang H, Williams A, Blume JE. Discovery of tissue-specific exons using comprehensive human exon microarrays. Genome Biol. 2007; 8:R64. [PubMed: 17456239]

40. Smith CW. Alternative splicing--when two's a crowd. Cell. 2005; 123:1–3. [PubMed: 16213205]
41. Mendell JT, ap Rhys CM, Dietz HC. Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts. Science. 2002; 298:419–22. [PubMed: 12228722]
42. Mi H, Guo N, Kejariwal A, Thomas PD. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. Nucleic Acids Res. 2007; 35:D247–52. [PubMed: 17130144]
43. Makeyev EV, Zhang J, Carrasco MA, Maniatis T. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. Mol Cell. 2007; 27:435–48. [PubMed: 17679093]
44. Ule J, Ule A, Spencer J, Williams A, Hu JS, Cline M, Wang H, Clark T, Fraser C, Ruggiu M, Zeeberg BR, Kane D, Weinstein JN, Blume J, Darnell RB. Nova regulates brain-specific splicing to shape the synapse. Nat Genet. 2005; 37:844–52. [PubMed: 16041372]
45. Christofk HR, Vander Heiden MG, Harris MH, Ramanathan A, Gerszten RE, Wei R, Fleming MD, Schreiber SL, Cantley LC. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature. 2008; 452:230–3. [PubMed: 18337823]
46. Thorsen K, Sorensen KD, Brems-Eskildsen AS, Modin C, Gaustadnes M, Hein AM, Kruhoffer M, Laurberg S, Borre M, Wang K, Brunak S, Krainer AR, Torring N, Dyrskjot L, Andersen CL, Orntoft TF. Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. Mol Cell Proteomics. 2008; 7:1214–24. [PubMed: 18353764]
47. He X, Pool M, Darcy KM, Lim SB, Auersperg N, Coon JS, Beck WT. Knockdown of polypyrimidine tract-binding protein suppresses ovarian tumor cell growth and invasiveness in vitro. Oncogene. 2007; 26:4961–8. [PubMed: 17310993]
48. Wang C, Norton JT, Ghosh S, Kim J, Fushimi K, Wu JY, Stack MS, Huang S. Polypyrimidine tract-binding protein (PTB) differentially affects malignancy in a cell line-dependent manner. J Biol Chem. 2008; 283:20277–87. [PubMed: 18499661]
49. David CJ, Chen M, Assanah M, Cannoll P, Manley JL. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. Nature. 2009
50. Gooding C, Clark F, Wollerton MC, Grellscheid SN, Groom H, Smith CW. A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. Genome Biol. 2006; 7:R1. [PubMed: 16507133]
51. Goers ES, Purcell J, Voelker RB, Gates DP, Berglund JA. MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. Nucleic Acids Res. 2010
52. Jin Y, Suzuki H, Maegawa S, Endo H, Sugano S, Hashimoto K, Yasuda K, Inoue K. A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. EMBO J. 2003; 22:905–12. [PubMed: 12574126]
53. Faustino NA, Cooper TA. Identification of putative new splicing targets for ETR-3 using sequences identified by systematic evolution of ligands by exponential enrichment. Mol Cell Biol. 2005; 25:879–87. [PubMed: 15657417]
54. Goren A, Ram O, Amit M, Keren H, Lev-Maor G, Vig I, Pupko T, Ast G. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. Mol Cell. 2006; 22:769–81. [PubMed: 16793546]
55. Wagner EJ, Garcia-Blanco MA. RNAi-mediated PTB depletion leads to enhanced exon definition. Mol Cell. 2002; 10:943–9. [PubMed: 12419237]
56. Del Gatto-Konczak F, Olive M, Gesnel MC, Breathnach R. hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. Mol Cell Biol. 1999; 19:251–60. [PubMed: 9858549]
57. Venables JP. Downstream intronic splicing enhancers. FEBS Lett. 2007; 581:4127–31. [PubMed: 17716673]
58. Forch P, Puig O, Martinez C, Seraphin B, Valcarcel J. The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5′ splice sites. EMBO J. 2002; 21:6882–92. [PubMed: 12486009]
59. Rideau AP, Gooding C, Simpson PJ, Monie TP, Lorenz M, Huttelmaier S, Singer RH, Matthews S, Curry S, Smith CW. A peptide motif in Raver1 mediates splicing repression by interaction with the PTB RRM2 domain. Nat Struct Mol Biol. 2006; 13:839–48. [PubMed: 16936729]

60. Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CW. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol Cell. 2004; 13:91–100. [PubMed: 14731397]
61. Coles JL, Hallegger M, Smith CW. A nonsense exon in the Tpm1 gene is silenced by hnRNP H and F. RNA. 2009; 15:33–43. [PubMed: 19037011]
62. Hunt SL, Jackson RJ. Polypyrimidine-tract binding protein (PTB) is necessary, but not sufficient, for efficient internal initiation of translation of human rhinovirus-2 RNA. RNA. 1999; 5:344–59. [PubMed: 10094304]
63. Ellis PD, Smith CW, Kemp P. Regulated tissue-specific alternative splicing of enhanced green fluorescent protein transgenes conferred by alpha-tropomyosin regulatory elements in transgenic mice. J Biol Chem. 2004; 279:36660–9. [PubMed: 15194683]
64. Schwartz S, Gal-Mark N, Kfir N, Oren R, Kim E, Ast G. Alu exonization events reveal features required for precise recognition of exons by the splicing machinery. PLoS Comput Biol. 2009; 5:e1000300. [PubMed: 19266014]
65. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. 1987; 15:7155–74. [PubMed: 3658675]
66. Carmel I, Tal S, Vig I, Ast G. Comparative analysis detects dependencies among the 5′ splice-site positions. RNA. 2004; 10:828–40. [PubMed: 15100438]
67. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004; 11:377–94. [PubMed: 15285897]
68. Schwartz SH, Silva J, Burstein D, Pupko T, Eyras E, Ast G. Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. Genome Res. 2008; 18:88–103. [PubMed: 18032728]
69. Bembom, O. seqLogo: An R package for plotting DNA sequence logos. 2007.
70. R Development Core Team. R. A Language and Environment for Statistical Computing. 2009.

Europe PMC Funders Author Manuscripts

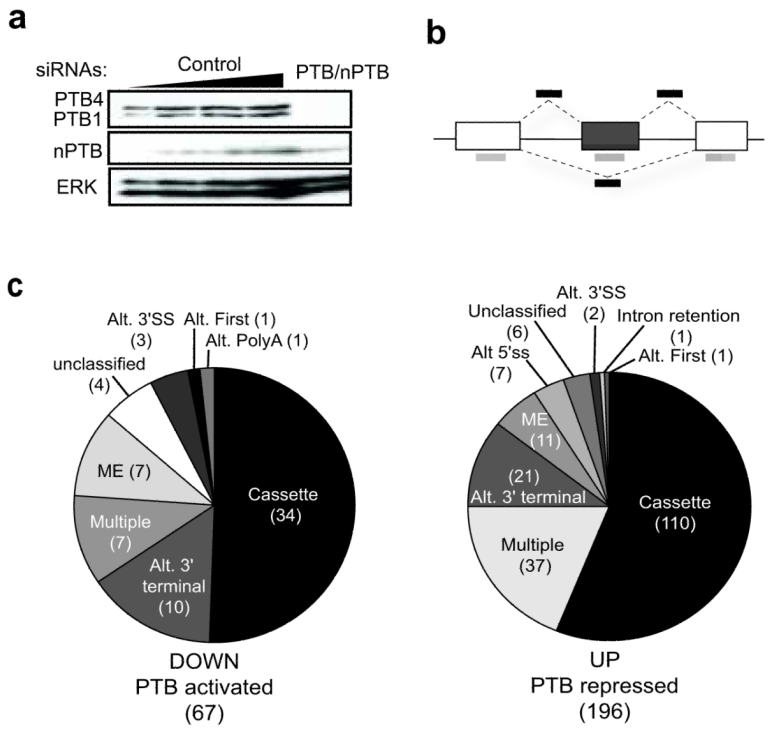

Figure 1. Microarray analysis of PTB/nPTB targets in Hela cells

(a) Western blot of HeLa cells transfected with control or PTB/nPTB siRNAs using the indicated antibodies. Equivalent amounts of control and knockdown samples were loaded together with 1:2, 1:4 and 1:8 dilutions of the control lysate to allow estimation of knockdown efficiency. (b) Schematic representation of the Affymetrix Human Exon-Junction Array. White boxes represent constitutive exons, dark gray box represent alternative spliced exon, dotted lines indicate splicing outcomes. Light and dark bars represent exon-body and exon-junction probesets respectively. (c) Classification of PTB regulated events. Pie charts showing distribution of PTB activated or repressed alternative spliced events, number of events on each situation given in brackets. ME, mutually exclusive exon; Alt. 5′SS, alternative 5′ splice site; Alt. 3′SS, alternative 3′ splice site; Multiple, multiple cassette exon; Alt. start, alternative first exon; Alt. PolyA, alternative polyadenlylation site; unclassified, unassigned/ambiguous splicing events.

Europe PMC Funders Author Manuscripts

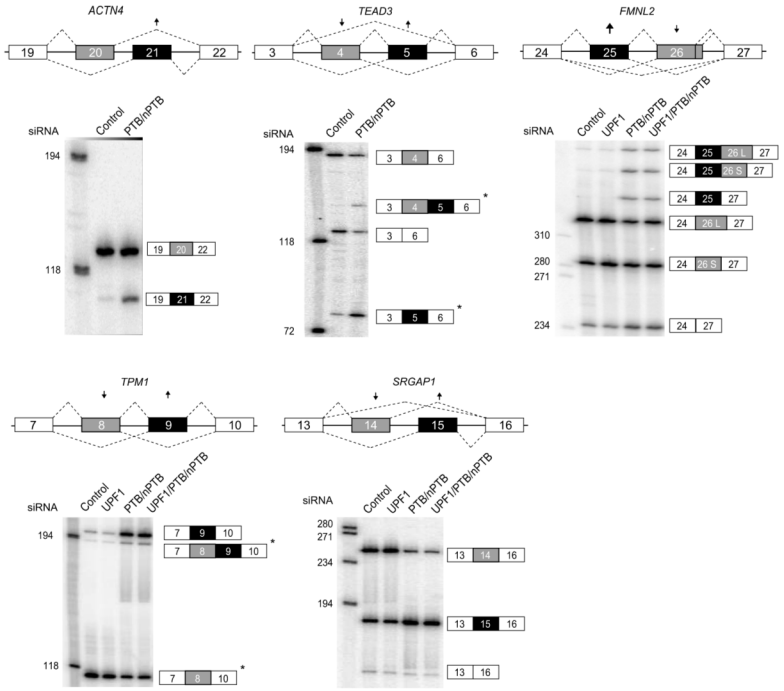

Figure 2. PTB regulation of mutually exclusive exons

Diagrams show possible transcripts arising from splicing of mutually exclusive exons in *ACTN4, TEAD3, FMNL2, TPM1* and *SRGAP1*. Flanking constitutive exons are coloured white, PTB repressed exons are black and PTB activated exons are gray. Direction of change predicted by array is indicated by arrows above exons (largest arrow – HIGHEST, smaller arrow – HIGH confidence level). Asterisks indicate *TPM1* RT-PCR products digested with PstI and *TEAD3* RT-PCR products digested with HinfI. Molecular markers are shown to the left of gels and the identity of bands indicated to the right. Quantitation is shown in Supplementary Fig. 1.

Europe PMC Funders Author Manuscripts

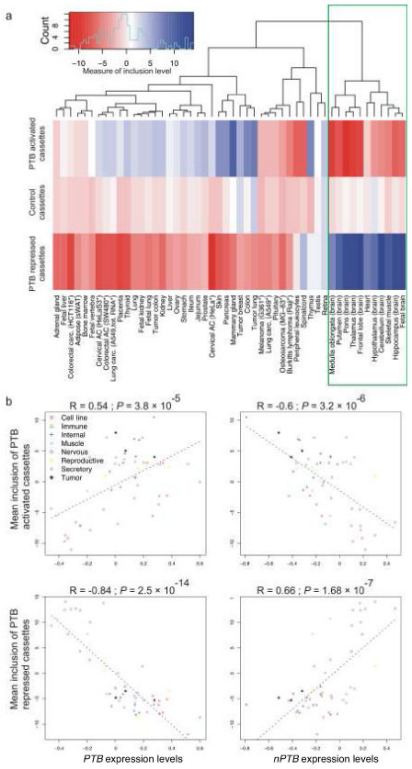

Figure 3. Tissue specificity of PTB-regulated exons

(a) Heat-map of mean tissue-specific exon inclusion profiles for PTB-activated (top), control (middle) and PTB-repressed (bottom) cassette exons across 50 human tissues and cell lines. Note that the major separation is between brain and striated muscles and all remaining tissues. (b) Correlations between relative inclusion levels of PTB-activated (top row) and PTB-repressed (bottom row) cassette exons and transcript levels of *PTB* (left column) and *nPTB* (right column). Each point represents the relative inclusion level of PTB-regulated exons in a particular tissue sample. Tissues/cell lines were grouped into 8 coherent groups of tissues, each represented by a different symbol.

Europe PMC Funders Author Manuscripts

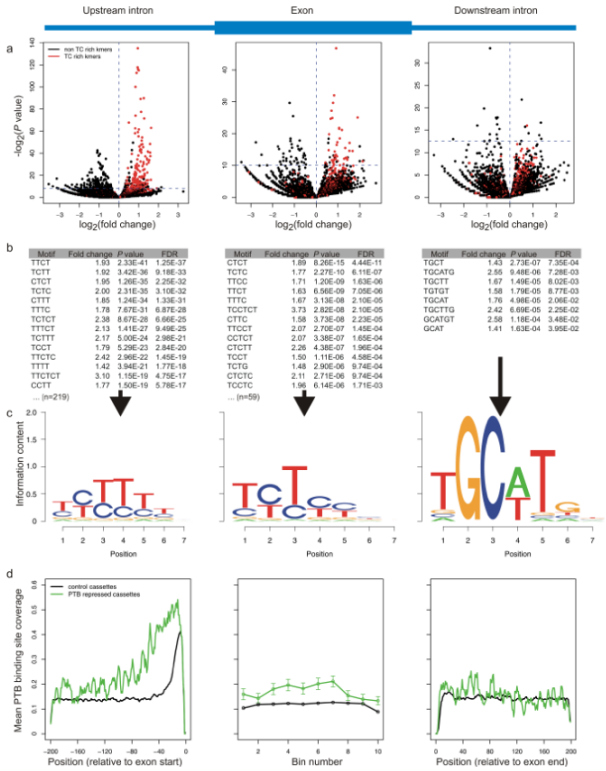

Figure 4. Motifs associated with PTB-repressed exons

(a) Volcano plots representing log transformed p-values as a function of fold-change for k-mers of 4-7 nt in length. *P* values and fold changes were derived based on a comparison of PTB-repressed compared to control cassette exons. Plots are shown for the 200 nt upstream flank (left), the exon (center) and 200 nt downstream intron flank (right). K-mers with a pyrimidine content greater than 75% are colored red. The remainder are colored black. Enriched motifs lie in the upper right quadrant. (b) Enriched motifs ranked by *P* value. (c) Representation of clustered significant motifs using sequence logos, in which the height of a nucleotide at a given position reflects its frequency. (d) Density of known PTB sites, defined as YTCY or YCTY, within the 200 nt upstream intron flank (left), the exon (center) and the 200 nt downstream intron flank (right). In each case the black line represents control cassette exons and the green line PTB-repressed cassette exons.

Europe PMC Funders Author Manuscripts

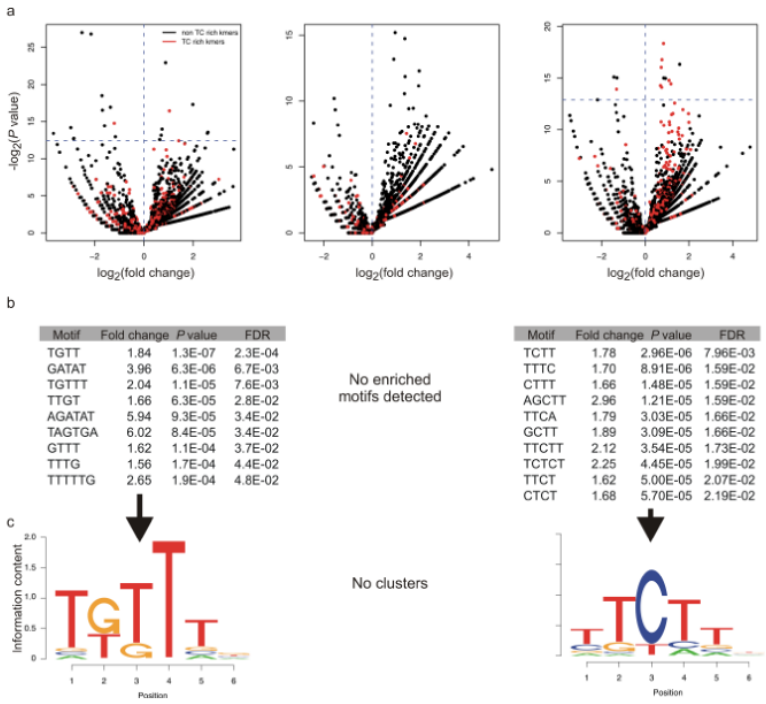

Figure 5. Motifs associated with PTB-activated exons

a) Volcano plots as in Fig. 4a for PTB-activated compared to control cassette exons. (b) Enriched motifs ranked by *P* value as in Fig 4b. No significantly enriched motifs were found within the exon sequences. (c) Clustered significant motifs as in Fig. 4c.

Europe PMC Funders Author Manuscripts

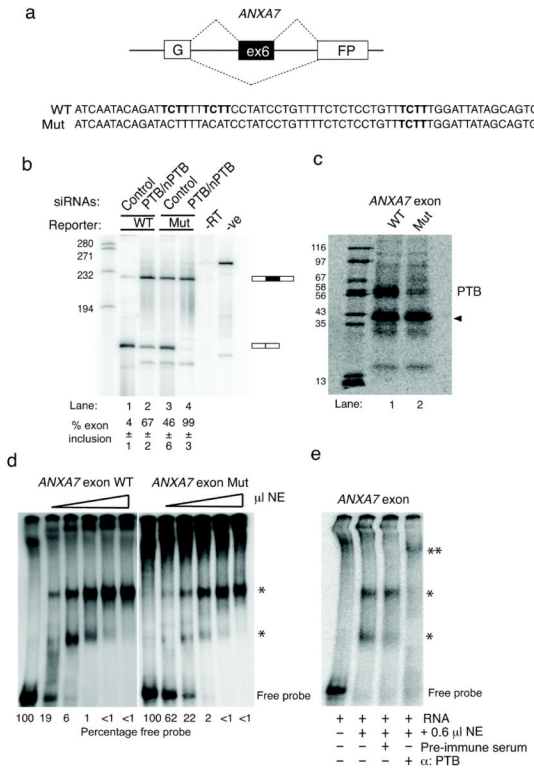

Figure 6. PTB acts via a ESS

(a) Schematic representation of *ANXA7* exon 6 minigene reporter (top). *ANXA7* exon 6 sequence (black rectangle) and 200 nucleotides of the flanking introns were cloned into the GFP reporter vector. WT and mutant sequences of the *ANXA7* exon are shown at the bottom. UCUU type motifs are shown in bold (b) RT-PCR of *ANXA7* minigene wild type (Wt) or mutated version (Mut) cotransfected with siRNAs against PTB and nPTB or control siRNAs. Molecular markers (nt) are shown to the left. (c) UV-crosslinking of *ANXA7* exon 6 WT and mutant (Mut) RNAs in HeLa nuclear extracts. Molecular marker sizes (kDa) are shown to the left. The position of the PTB crosslinked doublet is shown to the right. (d) Electrophoretic mobility shift assay of WT and mutant RNAs with HeLa nuclear extracts. Each sample contains 10 fmol of RNA and 0, 0.2, 0.4, 0.8, 1.4 and 2 µl of nuclear extract. Asterisks mark the position of the shifted complexes. Quantification of the extent of the shift is given as percentage of free RNA. (e) Supershift of WT *ANXA7* exon 6 RNA with 0.6 µl of HeLa NE and with 2 µl pre-immune serum or PTB antibody as indicated. The position of shifted complexes is marked with an asterisk and a double asterisk shows the supershifted complex.

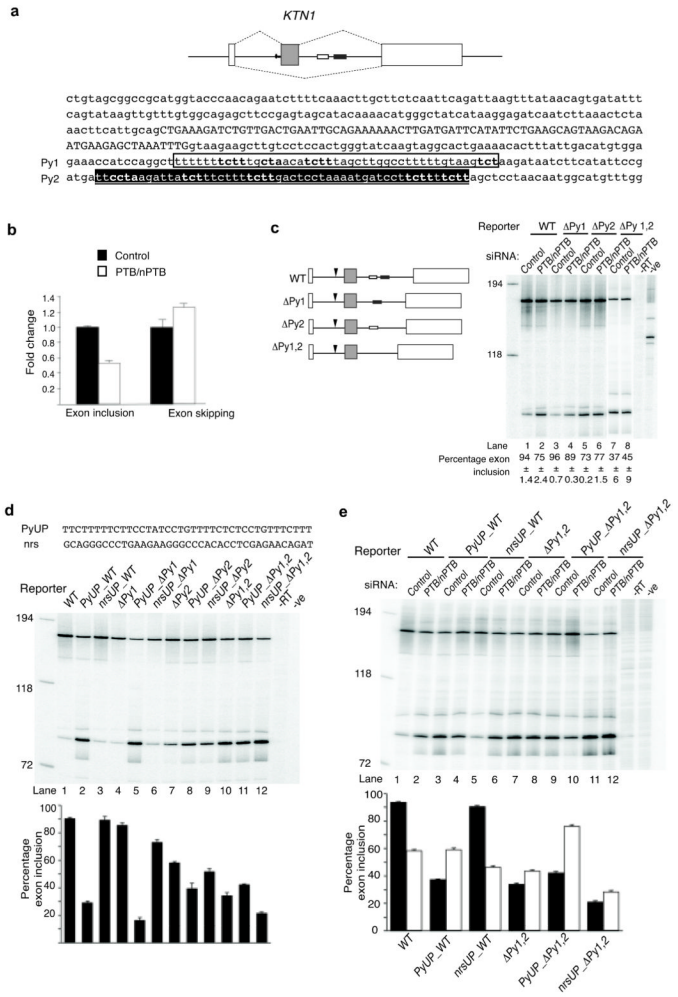

Figure 7. PTB promotes exon inclusion via an ISE

(a) Schematic representation of KTN1 minigene (top). Sequence of *KTN1* exon (chr14:55,209,643-55,209,726), 152 nucleotides of upstream and 200 nucleotides of downstream flanking introns were cloned into the GFP reporter (bottom). Intron sequences shown in lower case, exonic sequences in upper case, boxed sequence correspond to identified extended pyrimidine rich sequences Py1 (white box) and Py2 (black box). (b) Changes in inclusion levels of the endogenous *KTN1* exon in response to PTB/nPTB knockdown quantified by qPCR in control (black) and PTB/nPTB knockdown (white) conditions. Values shown are mean ± s.d., n=3 (c) The minigene reporter and intronic deletions are shown schematically to the left. ΔPy1 is deletion of the first pyrimidine tract (white box on panel a); ΔPy2 is deletion of the second pyrimidine tract (black box on panel a) or ΔPy1,2 deletion of both sequences. RT-PCR of *KTN1* minigene reporter (wt) and deletions cotransfected with control or PTB/nPTB siRNAs. Size markers are shown to the left of the gel and % of exon inclusion is shown at the bottom of the gel. Values shown are mean ± s.d., n=3, except ΔPy1,2 that is n=15. (d) Sequence of pyrimidine rich tract (PyUP) or a non related sequence (nrs) inserted 68 nucleotides upstream of the *KTN1* exon (top). RT-PCR of the indicated minigene reporters transfected in HeLa cells (middle). Values shown are mean percentage exon inclusion ± s.d., n=3 (Bottom). (e) RT-PCR of the indicated minigene reporters cotransfected with control or PTB/nPTB siRNAs. The histogram aligned below shows mean percentage exon inclusion ± s.d., n=3.

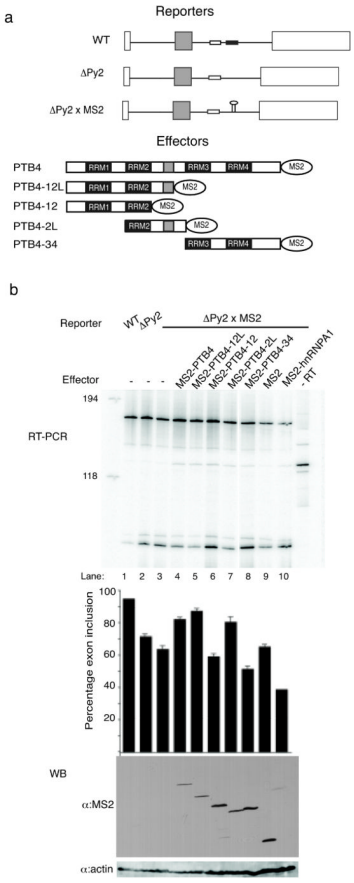

Figure 8. MS2 recruitment to downstream intron reveals PTB minimal activation domain

(a) Schematic representation of minigene reporters and effectors utilized in the experiment. (b) RT-PCR (top), quantification of percentage exon inclusion (middle) and western blot (bottom) of the same transfection experiments probed with anti MS2 and anti actin as loading control. Histogram bars and western blots are aligned with the lanes of the preceding RT-PCR gel. Values shown are mean ± s.d., n=3.

Europe PMC Funders Author Manuscripts

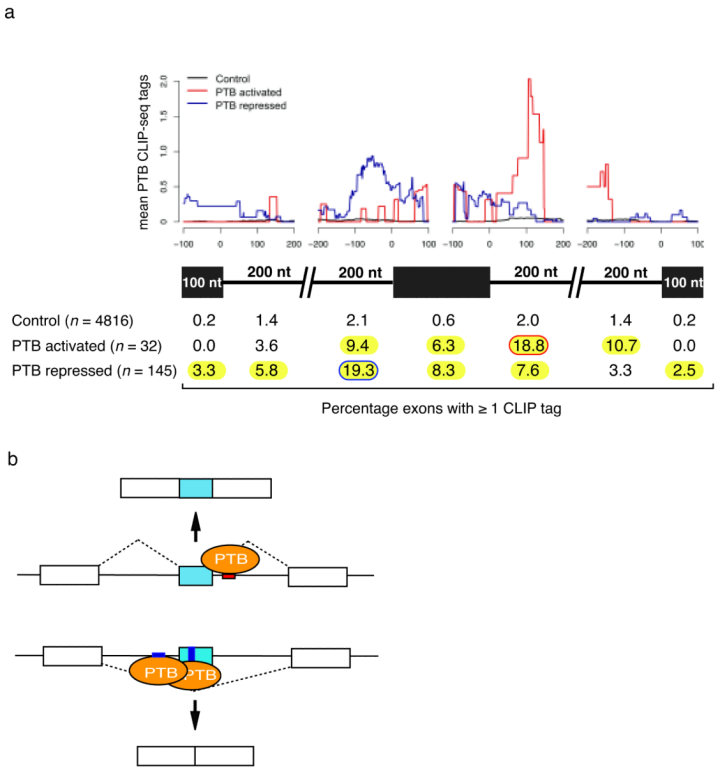

Figure 9. Position dependent PTB mode of action

(a) Density of CLIP-seq tags{Xue, 2009 #235} (top) and percentage of exons and intronic segments overlapped by ≥1 CLIP tag (bottom) on control, PTB-activated and PTB-repressed exons (this study) within 7 regions around regulated exons, including 200nt of intronic regions, the regulated exons as well as 100nt in the flanking constitutive exons. Black line represents control exons, red line PTB- activated exons and purple line PTB-repressed exons. Significant differences between PTB-regulated and control exons were determined by Fisher exact tests and are highlighted in yellow ($P < 0.05$). The most significant enrichments were observed immediately downstream of activated exons ($P = 4.9 \times 10^{-5}$), immediately upstream of repressed exons ($P = 1.2 \times 10^{-17}$) and within repressed exons ($P = 1.7 \times 10^{-9}$). (b) The cartoon represents an alternative cassette exon (blue box) and flanking constitutive exons (white boxes). Intronic sequences are shown as thin lines, exonic/intronic splicing silencers are depicted as green rectangles and splicing enhancers are depicted as red rectangles. PTB (orange oval shape) binding to downstream sequences promotes exon inclusion and binding to upstream intronic or exonic sequences promotes exon skipping.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Nat Struct Mol Biol. Author manuscript; available in PMC 2011 March 01.

Table 1

Properties of PTB-regulated cassette and 3′ terminal exons compared to control cassette exons

	n	PPT start	PPT length	PPT score	3′ splice site score	AGEZ	5′ splice site score			Exon length	PTB site density		
							pssm	ΔG	Max ent		Up intron	Exon	Down intron
control cassettes	4892	21.40 ± 0.02	13.70 ± 0.14	9.92 ± 0.09	71.59 ± 0.12	39.43 ± 0.36	81.15 ± 0.11	−5.82 ± 0.03	7.85 ± 0.04	156.91 ± 4.1	0.16 ± 1.1e-03	0.12 ± 1.2e-03	0.14 ± 1.1e-03
PTB activated 3′ terminal	10	20.60 ± 4.3	11.00 ± 2.7	8.85 ± 1.7	68.49 ± 2.2	40.80 ± 10	N/A	N/A	N/A	841.90 ± 320	0.16 ± 2.5e-02	0.13 ± 2.5e-02	0.13 ± 1.6e-02
PTB repressed 3′ terminal	21	28.71 ± 3.1	21.52 ± 2.7	16.18 ± 2.1	75.27 ± 1.8	65.76 ± 10	N/A	N/A	N/A	712.19 ± 120	0.25 ± 2.7e-02	0.16 ± 1.1e-02	0.15 ± 1.3e-02
PTB activated cassettes	32	22.47 ± 2.8	12.03 ± 1.7	8.30 ± 0.99	68.55 ± 1.6	49.34 ± 4.8	77.98 ± 1.4	−4.82 ± 0.36	7.32 ± 0.66	111.69 ± 17	0.15 ± 1.5e-02	0.11 ± 1.7e-02	0.17 ± 1.8e-02
PTB repressed cassettes	146	34.99 ± 1.2	25.03 ± 1.0	16.92 ± 0.70	70.80 ± 0.77	67.85 ± 3.3	80.40 ± 0.65	−5.70 ± 0.16	7.71 ± 0.25	132.94 ± 24	0.26 ± 8.0e-03	0.17 ± 1.0e-02	0.15 ± 6.8e-03

Light yellow: $p < 0.05$

dark yellow: $p < 0.001$ (T-test)