Regulated intron retention and nuclear pre-mRNA decay contribute to PABPN1 autoregulation

Danny Bergeron†, Gheorghe Pal†, Yves B. Beaulieu, and Francois Bachand.

Supplemental Information Inventory

Supplemental Information includes eight figures, Supplemental Experimental Procedures, and Supplemental References.

SUPPLEMENTARY FIGURES:

SUPPLEMENTARY FIGURES

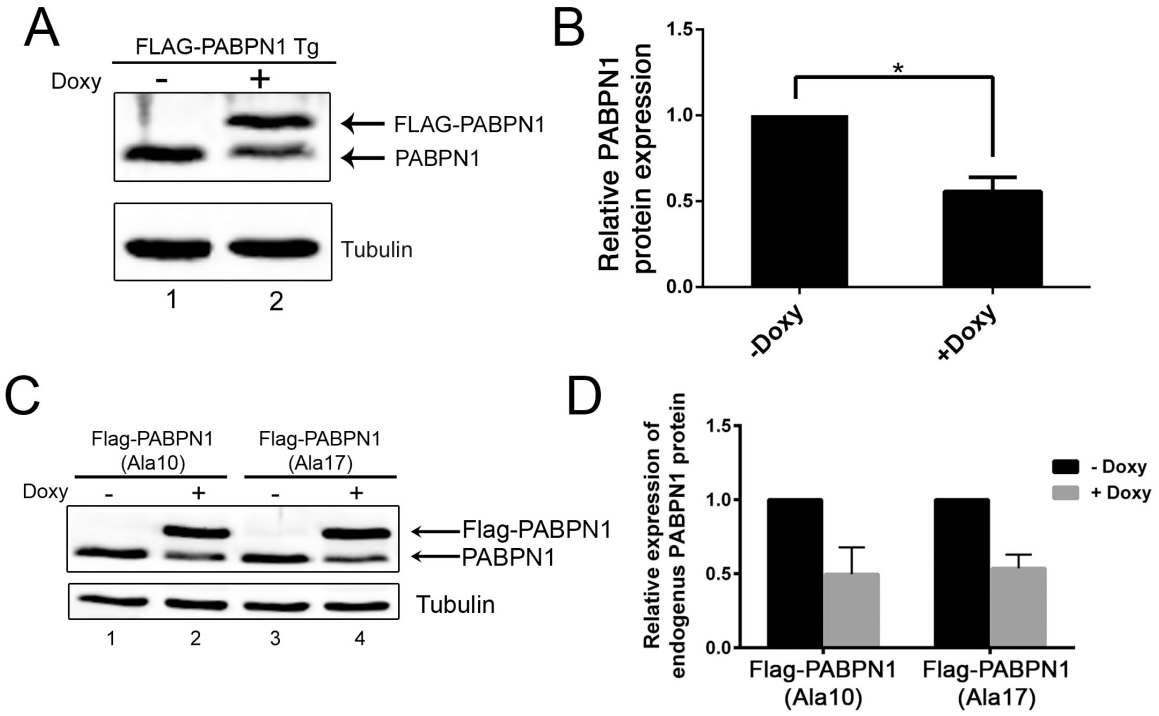

Figure S1, related to Figure 1: Down-regulation of endogenous PABPN1 protein in a transgenic cell line that induces FLAG-PABPN1 expression.

(A) Western blot analysis of the indicated proteins using extracts from an untreated (-) or doxycycline-treated (+) transgenic (Tg) cell line that express FLAG-PABPN1 (lanes 1-2). The PABPN1-specific antibody recognizes both endogenous and FLAG-tagged PABPN1.

(B) Quantification of endogenous PABPN1 protein in the presence and absence of doxycycline in the FLAG-PABPN1 transgenic (Tg) cell line. (✶) *P*-value <0.05, Student's *t*-test. Note that the reduced levels of autoregulation observed in the FLAG-PABPN1 Tg cell line is due to the lower expression levels of the *FLAG-PABPN1* transgene relative to the *GFP-PABPN1* transgene.

(C) Western blot analysis of the indicated proteins using extracts from HEK293T cells that conditionally express normal (Ala10, lanes 1-2) and alanine-expanded (Ala17, lanes 3-4) versions of PABPN1 in the presence of doxycycline (lanes 2 and 4). The PABPN1-specific antibody recognizes both endogenous and FLAG-tagged PABPN1.

(D) Quantification of endogenous PABPN1 protein in the presence and absence of doxycycline in the cell lines expressing normal (Ala10) and alanine-expanded (Ala17) PABPN1.

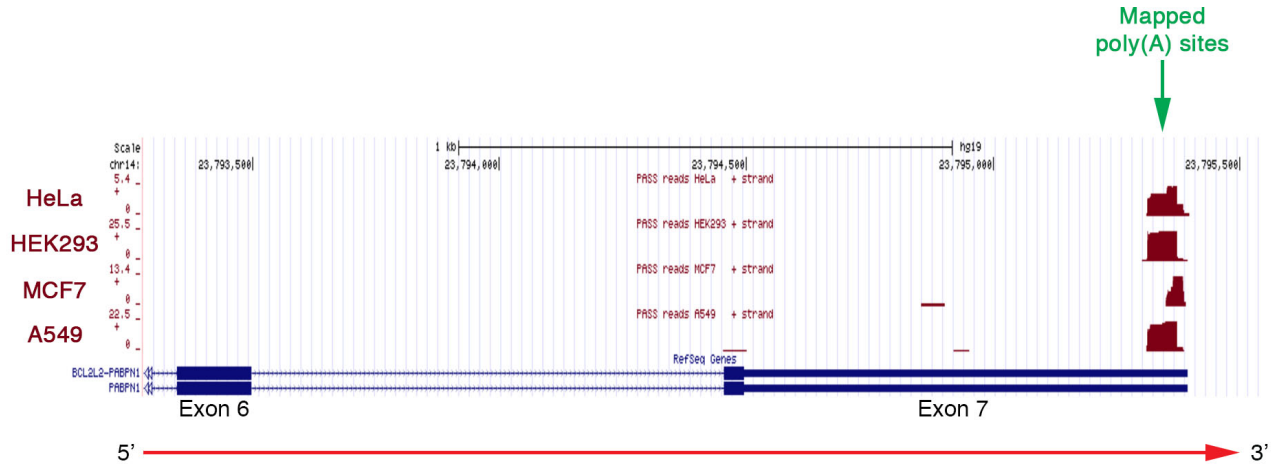

Figure S2, related to Figure 1: 3' READS analysis of *PABPN1* polyadenylation sites in different human cell lines.

RNA-seq coverage of poly(A) site-supporting (PASS) reads for *PABPN1* in HeLa, HEK293, MCF7, and A549 human cell lines, as determined by 3' READS (Hoque et al. 2013). In all four cell lines, PASS reads identified one major region of poly(A) site selection for *PABPN1* ~875-900-nt downstream of the stop codon.

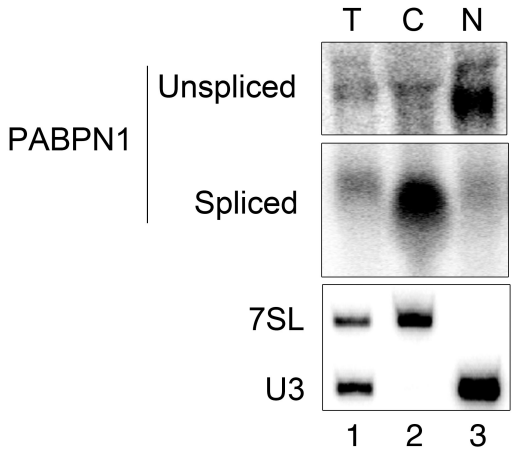

Figure S3, related to Figure 1: The unspliced *PABPN1* pre-mRNA that retains the 3'-terminal intron is enriched in the nucleus.

HEK293T were fractionated to isolate nuclear (N) and cytoplasmic (C) RNA, whereas the remaining one-third was lysed to analyze total (T) RNA. Total and fractionated RNA populations were analyzed by Northern blotting using probes complementary to *PABPN1* intron 6 (unspliced; upper panel) and exon 2 (spliced; middle panel). As a control for the RNA fractionation, the U3 snoRNA and signal recognition particle RNA (7SL) were enriched in the nucleus and cytoplasm, respectively.

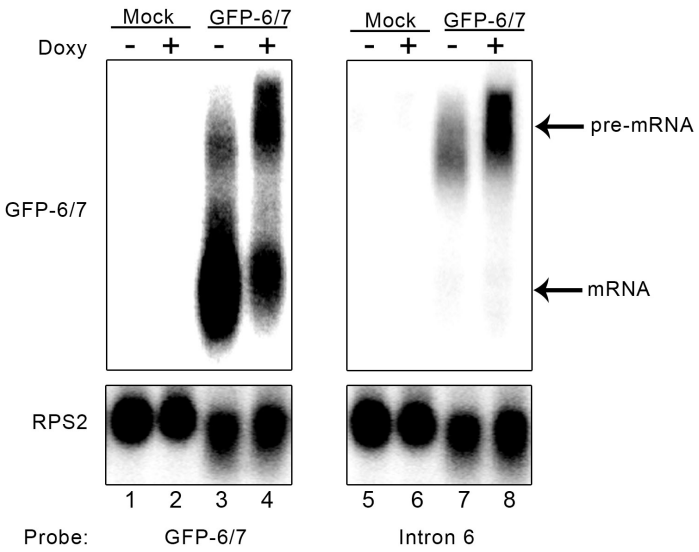

Figure S4, related to Figure 2: The long RNA isoform expressed from the GFP-6/7 minigene corresponds to unspliced transcript.

Northern blot analysis of total RNA prepared from untreated (-) and doxycycline-treated (+) GFP-PABPN1 transgenic cell line that were mock-transfected (lanes 1-2 and 5-6) or transfected with the GFP-6/7 minigene (lanes 3-4 and 7-8). The blots were analyzed using probes complementary to GFP-6/7-specific (lanes 1-4) and intron 6-specific (lanes 5-8) sequences. The *RPS2* mRNA was analyzed as a loading control.

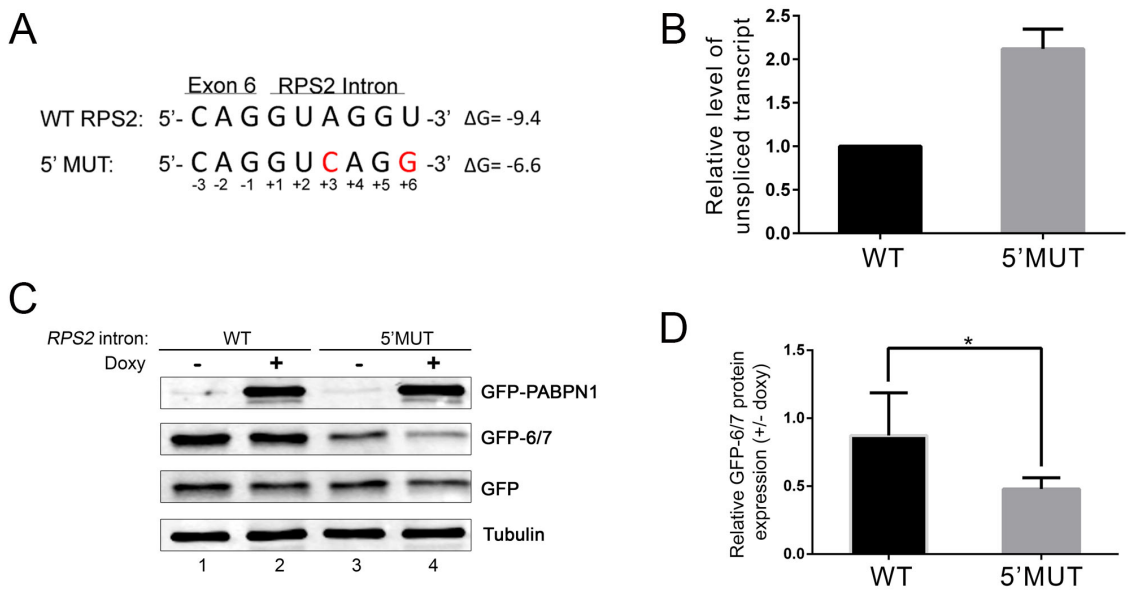

Figure S5, related to Figure 3: Mutations in the 5'ss of the *RPS2* 3'-terminal intron that reduce splicing efficiency render expression from the GFP-6/7 minigene sensitive to PABPN1-dependent regulation.

(A) 5' ss sequences of wild-type (WT) and mutated (5' MUT) *RPS2* 3'-terminal intron are shown. The nucleotide changes introduced in the 3'-terminal intron of *RPS2* to reduce splicing efficiency are shown in red. Nucleotide positions are numbered and indicated under each nucleotide. ΔG, predicted free energy of 5' ss.

(B) Increased level of unspliced transcript in cells that express GFP-6/7 minigene with mutations in the *RPS2* 5'ss that reduce base pairing with the U1 snRNA. RT-qPCR analysis of unspliced transcript using total RNA prepared from HEK293T cells that were previously transfected with GFP-6/7 constructs containing wild-type and 5' ss mutant *RPS2* intron. The unspliced transcript was analyzed using a forward primer specific to *RPS2* intronic sequences and a reverse primer specific to *PABPN1* 3' UTR sequences. Data were normalized to endogenous *RPS2* mRNA and expressed relative to the GFP-6/7 construct containing the wild-type *RPS2* intron.

(C) The reduced splicing efficiency of the *RPS2* intron with 5' ss mutations sensitize GFP-6/7 expression to PABPN1-dependent regulation. Western blot analysis of the indicated proteins using extracts from untreated (-) and doxycycline-treated (+) GFP-PABPN1 transgenic cells that were previously transfected with GFP-6/7 constructs containing wild-type (lanes 1-2) and 5' ss mutant (lanes 3-4) *RPS2* intron. Cells were also co-transfected with a vector expressing GFP to control for transfection efficiency.

(D) GFP-6/7 protein levels were normalized to tubulin and to GFP (to control for transfection efficiency) and expressed relative to uninduced control (-Doxy). (✶) *P*-value <0.05, Student's *t*-test.

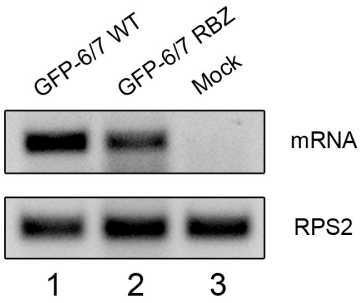

Figure S6, related to Figure 4: Spliced GFP-6/7 mRNA is produced from the ribozyme construct.

RT-PCR analysis of spliced GFP-6/7 mRNA using total RNA prepared from HEK293T cells that were either mock-transfected (lane 3) or transfected with the wild-type (lane 1) and ribozyme (RBZ) GFP-6/7 constructs. To specifically detect spliced GFP-6/7 mRNA, the PCR amplification was performed using a forward primer complementary to GFP-specific sequences and a reverse primer positioned across the exon 6-exon 7 junction of *PABPN1.*

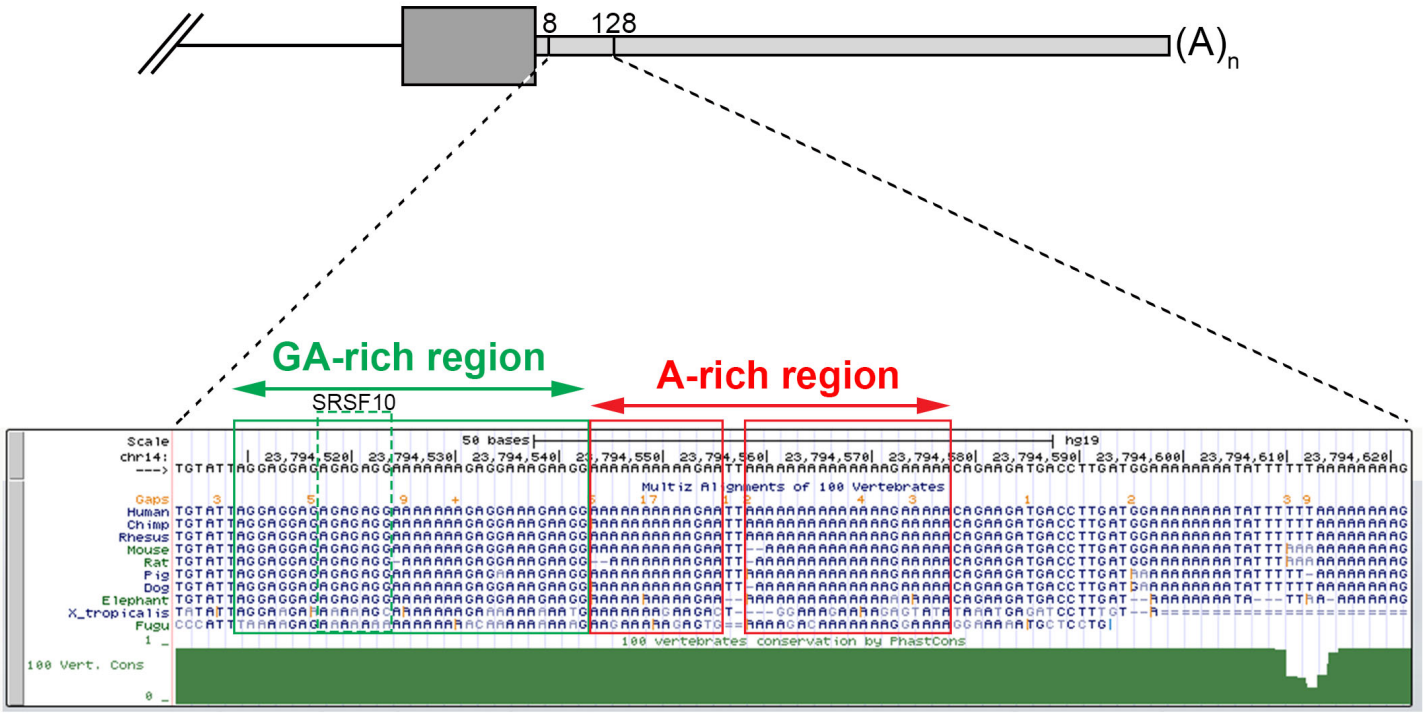

Figure S7, related to Figures 5 and 6: Conserved GA- and A-rich sequences in the *PABPN1* 3' UTR.

Diagram showing the 3' end of the *PABPN1* gene, including the intron 6-exon 7 junction as well as ~900-nt of 3' UTR sequences. The nucleotide sequence of a conserved 120-nt region (nt 8-128 starting from the stop codon) among ten species is shown at the bottom. The phastCons plot (green) shows the probability of sequence conservation across 100 vertebrate species. The green and red double-headed arrows indicate the conserved GA- and A-rich sequences, respectively. The consensus SRSF10-binding site, as determined by RNAcompete (Ray et al. 2013), is shown in the dotted box within the GA-rich region.

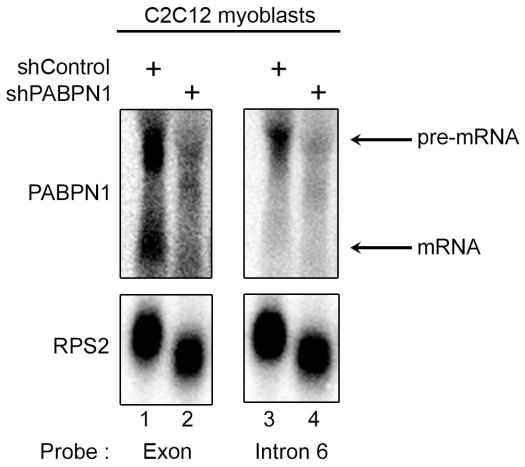

Figure S8, related to Discussion: Detection of intron-retained PABPN1 transcripts in mouse C2C12 myoblasts.

Northern blot analysis using total RNA from mouse C2C12 cells induced to express PABPN1-specific (lanes 2 and 4) and control (lanes 1 and 3) shRNAs. Blots were analyzed using RNA probes complementary to *PABPN1* exons 1-3 (lanes 1-2) and intron 6 (lanes 3-4) sequences, as well as mouse *RPS2* mRNA as a loading control. As the mouse *PABPN1* mRNA co-migrated with the 18S rRNA, detection of the *PABPN1* mRNA by northern was sub-optimal, making estimates of mRNA:pre-mRNA ratios unreliable.

SUPPLEMENTARY EXPERIMENTAL PROCEDURES

RNAi-mediated silencing. siRNAs were from Dharmacon/Thermo Scientific. Control siRNAs (ON-TARGETplus Non-targeting siRNA #4), siUPF1 (GGAACCACCUGCUGAACUATT); sihRRP40 (CACGCACAGUACUAGGUCAdTdT), siMTR4 (CAAUUAAGGCUCUGAGUAAUU), and siXRN2 (GAGUACAGAUGAUCAUGUUUU), and siPABPN1 (AGUCAACCGUGUUACCAUAUU) were previously described (Beaulieu et al. 2012). siRNAs to hnRNP A2/B1 (AAGCUUUGAAACCACAGAAGAdTdT), hnRNP H (GCACAGGUAUAUUGAAAUCdTdT), and hnRNP C (AGGCGCUUGUCUAAGAUCAdTdT) were also previously described. (Venables et al. 2008).

Expression Constructs. The wild-type GFP-6/7 construct (pFB864) was generated by PCR amplification of a ~2100-nt *PABPN1* genomic fragment that started with the seventh base pair of exon 6 up to approximately 1-Kb downstream of the stop codon in exon 7, using a 5' primer containing EcoRI sequence (5'-GGCGAATTCCAAACGAACCAACAGACCAGG-3') and a 3' primer containing BamHI sequence (5'-GATCCCCACAACACCAGTACTTCTACCACAAGG-3'), which was cloned in frame to GFP coding sequence into peEGFP-C1. The intronless GFP-6/7 construct (pFB908) was generated from pFB864 using the Gibson assembly cloning kit (New England Biolabs) via PCR amplification of *PABPN1* exon 6 and exon 7 that included overlapping sequences. The GFP-6/7 construct with mutations in the *PABPN1* 5' ss (pFB969) was generated by site-directed mutagenesis using pFB864 as template. GFP-6/7 constructs with *RPS2* (pFB952) and *HNRNPK* (pFB963) introns were generated by Gibson assembly cloning using PCR fragments containing *RPS2* and *HNRNPK* intronic sequences fused to sequences from exon 6 and exon 7 of *PABPN1*. The GFP-6/7 construct with mutations in the *RPS2* 5' ss (pFB996) was generated by site-directed mutagenesis using pFB952 as template. The GFP-6/7 Ribozyme construct (pFB1026) was generated by inserting a previously described fragment (Bird et al. 2005) containing a stabilizing stem loop from the histone *H2A* mRNA and a variant of the hepatitis delta ribozyme upstream of the PABPN1 polyadenylation signal. cDNAs to SRSF10 and hnRNP C were obtained from DNASU plasmid repository and cloned into pEGFP-C1. The GFP-6/7 A-to-C mutants A-MUT1 (pFB838) and A-MUT2 (pFB839) were generated by site-directed mutagenesis using pFB781 as template. All of the constructs used in this study were verified by automated sequencing.

Sequence of EMSA probes. The wild-type *PABPN1* 3' UTR corresponding to the A-rich region (CATCTTCTGTTTTTCTTTTTTTTTTTTTTTAATTCTTTTTTTTTTCCTTCTTTCCTCTTTCCTGTCTC) , or A-to-C mutants (A-MUT1; CATCTTCTGTTTTTCTTTTTTTTTTTTTTTAATTCGTGTGTGTGTCCTTCTTTCCTCTTTCCTGTCTC) and (A-MUT2; CATCTTCTGTTTTTCGTGTGTGTGTTTTTAATTCTTTTTTTTTTCCTTCTTTCCTCTTTCCTGTCTC) were generated by *in vitro* transcription using the mirVana miRNA Probe Construction kit (Ambion, AM1550) according to the manufacturer's instructions.

SUPPLEMENTARY REFERENCES

Beaulieu YB, Kleinman CL, Landry-Voyer AM, Majewski J, Bachand F. 2012. Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. *PLoS genetics* **8**: e1003078.

Bird G, Fong N, Gatlin JC, Farabaugh S, Bentley DL. 2005. Ribozyme cleavage reveals connections between mRNA release from the site of transcription and pre-mRNA processing. *Mol Cell* **20**: 747-758.

Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. 2013. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods* **10**: 133-139.

Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A et al. 2013. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**: 172-177.

Venables JP, Koh CS, Froehlich U, Lapointe E, Couture S, Inkel L, Bramard A, Paquet ER, Watier V, Durand M et al. 2008. Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol Cell Biol* **28**: 6033-6043.