

**Supplementary Table S1:** GEO number of publicly available datasets analyzed in this work.

**Supplementary Tables S2:** List of exons activated by each analyzed splicing factors and lists of the GC-exons and AT-exons.

**Supplementary Tables S3:** Annotation of isochores, LADs, and TADs.

#### **Figure S1**

Violin plots representing the relative 3' ss score (upper panel) and the relative 5' ss score (lower panel) for each set of splicing-factor activated exons, when compared to control exons. CCE=control coding exons; (\*) Wilcoxon's test FDR < 0.05.

#### **Figure S2**

Violin plots representing the relative adenine, cytosine, guanine, and thymine frequencies for each set of splicing-factor activated exons, when compared to control exons. CCE=control coding exons; (\*) Student's test FDR < 0.05.

#### **Figure S3**

**A.** Violin plots representing the relative GC frequency in each set of splicing factor-activated exons, when compared to control exons . CCE=control coding exons; (\*) Student's test FDR < 0.05.

**B.** Violin plots representing the relative GC frequencies in introns that are upstream and downstream, respectively, of splicing factor-activated exons. CCE=control coding exons; (\*) Wilcoxon's test FDR < 0.05.

**C.** Heatmaps representing the relative frequency of GC and AT nucleotides in the upstream (n-1) and downstream introns of splicing factor-activated exons. (\*) Wilcoxon's test, FDR < 0.05.

#### **Figure S4**

**A.** Violin plots representing the relative upstream intron size (upper panel) and the relative downstream intron size (lower panel) for each set of splicing-factor activated exons, when compared to control exons. CCE=control coding exons

**B** Heatmap representing the median size of the smallest intron flanking splicing factor-activated exons, when compared to the median size of human introns. The sets of splicing factor-activated exons are represented in the same order as in fig 1a.

**C.** Correlation between the relative median size of introns of genes hosting splicing factor-activated exons (compared to all human introns) and the relative gene GC-content (compared to all human genes);  $r$  = Pearson correlation coefficient.

### Figure S5

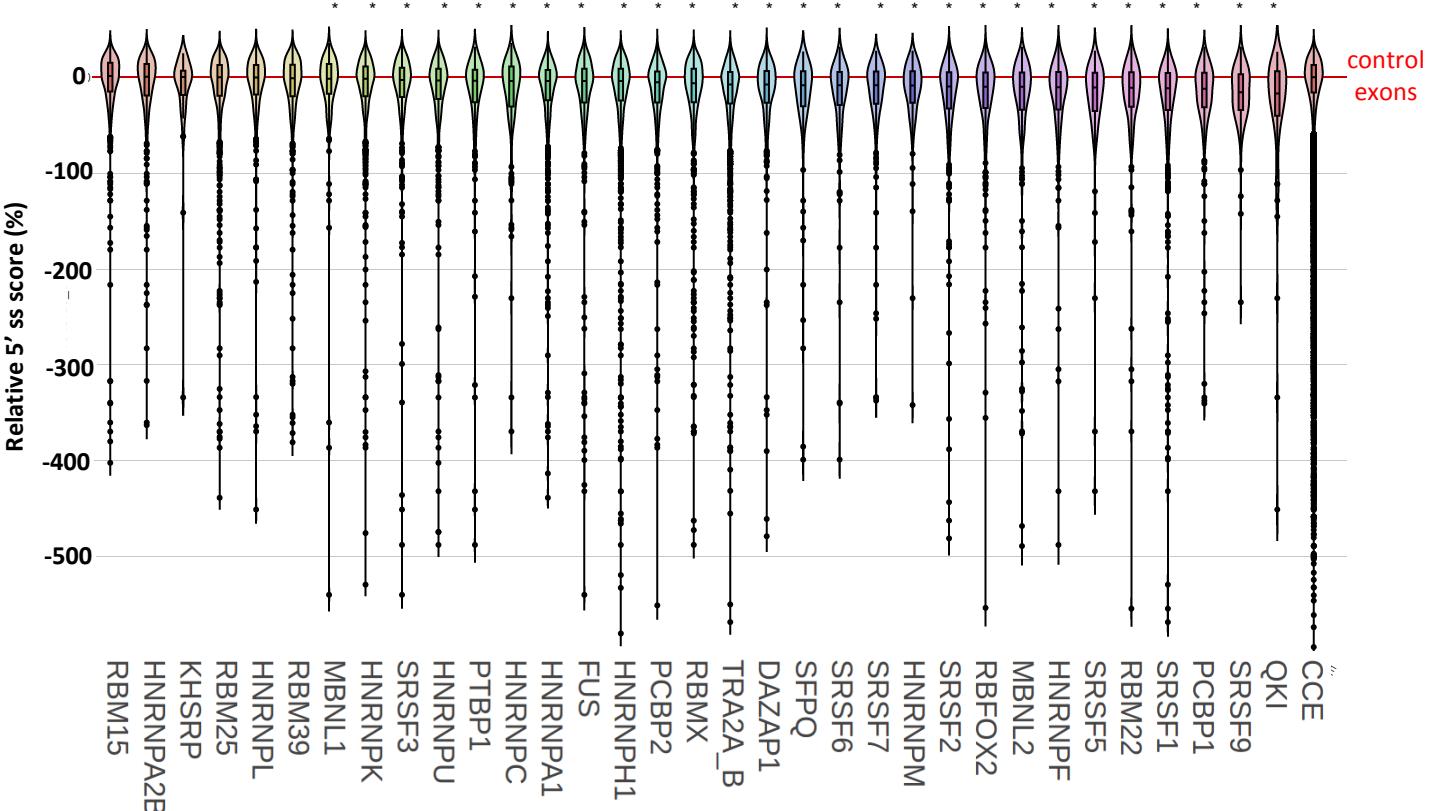
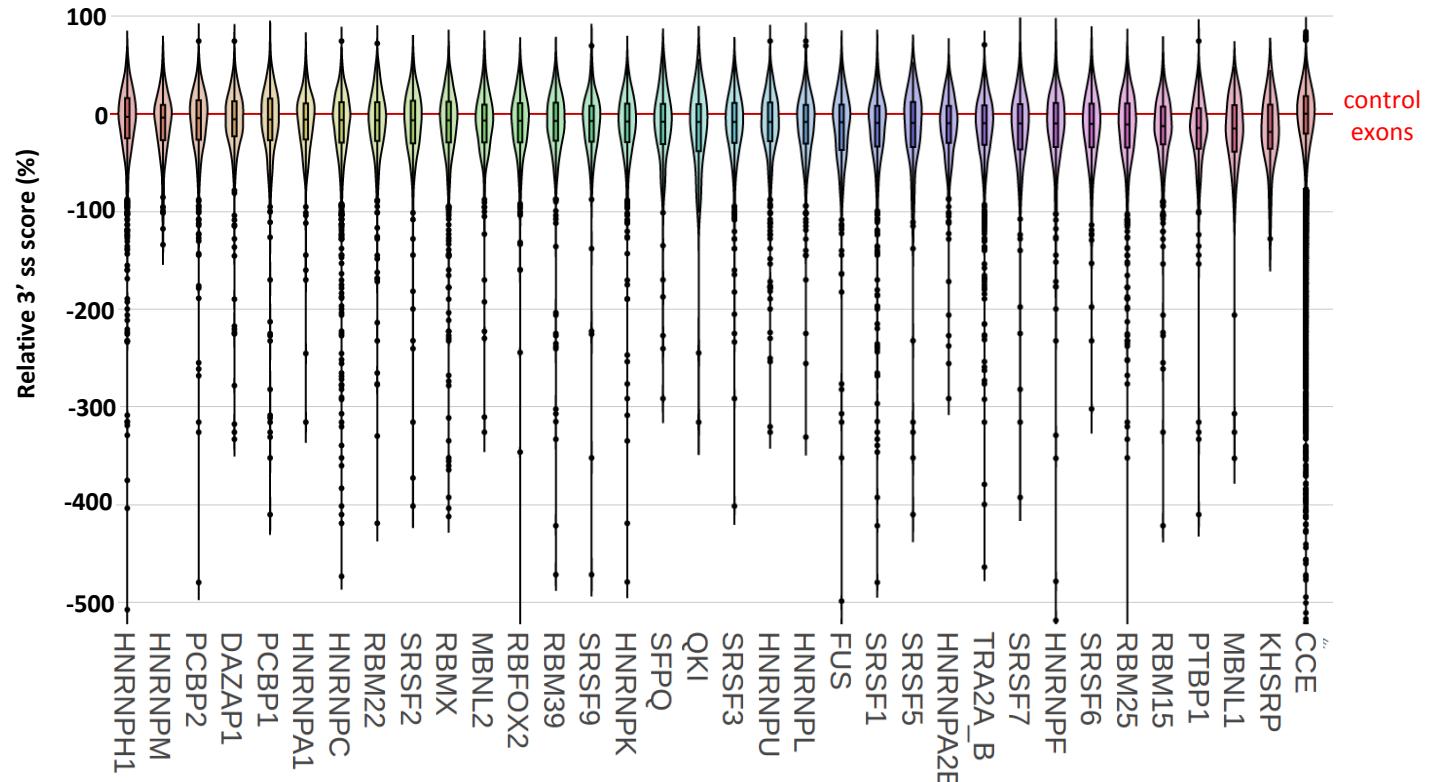
- A.** Density of reads obtained after immunoprecipitation of RNAPII in HEK293 cell line and mapping to different parts of the genes hosting GC-exons or AT-exons.
- B.** Density of reads obtained after DNA treatment with MNase (left panel) or after immunoprecipitation of the histone H3 (right panel) in Hela cells and mapping to GC-exons or AT-exons and their flanking introns.
- C.** Density of reads obtained after DNA treatment with MNase or after immunoprecipitation of the histone H3 in Hela cells and mapping to different parts of the genes hosting GC-exon or AT-exons.
- D.** Density of reads obtained from the HEK293 and Hela cell lines after immunoprecipitation of DNA using antibodies against H3K4me3 and H3K9ac and mapping different parts of genes hosting GC-exons or AT-exons.

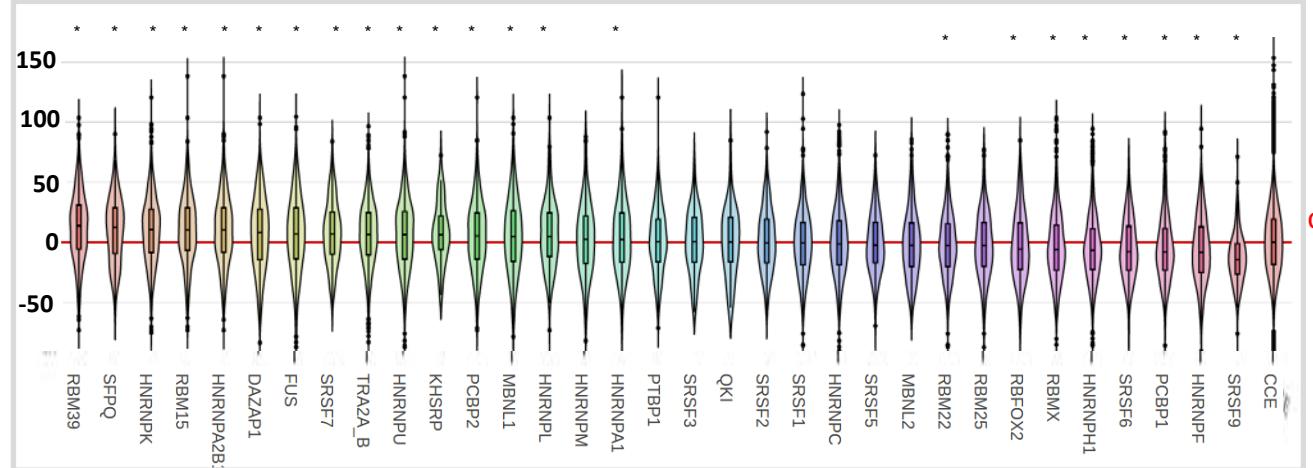
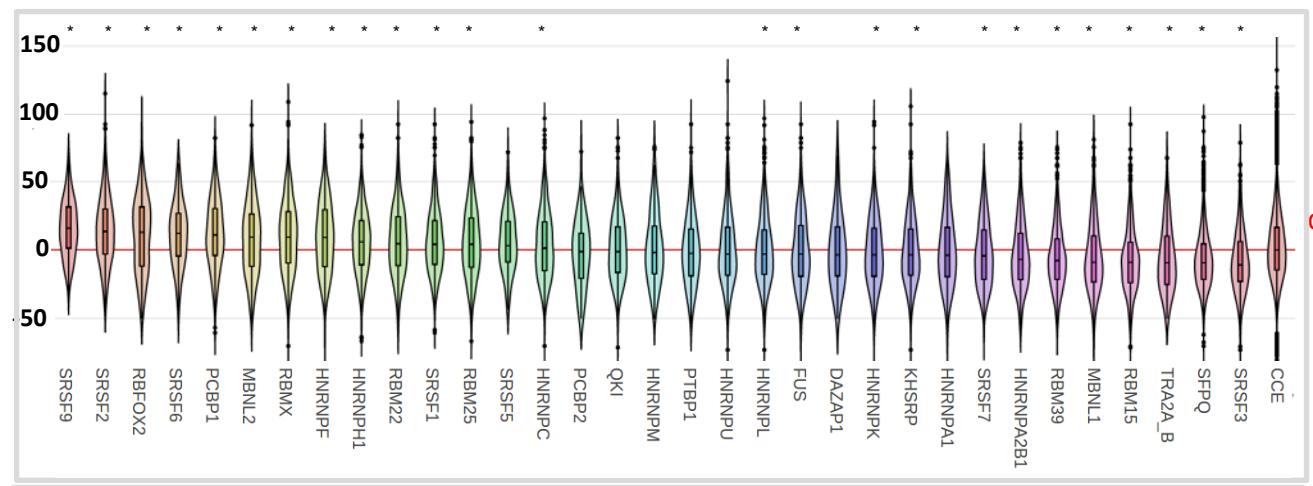
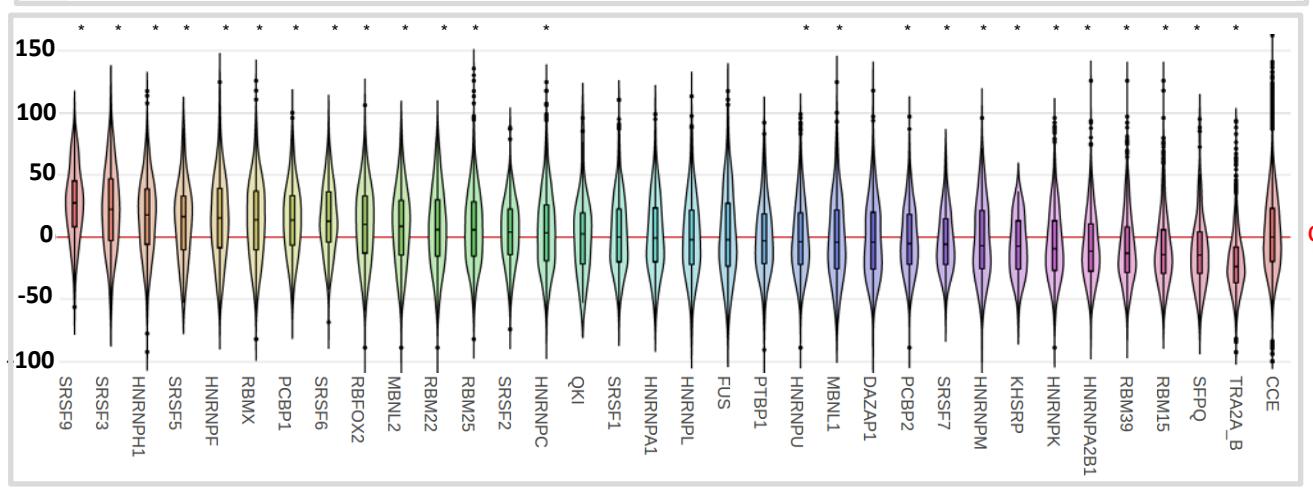
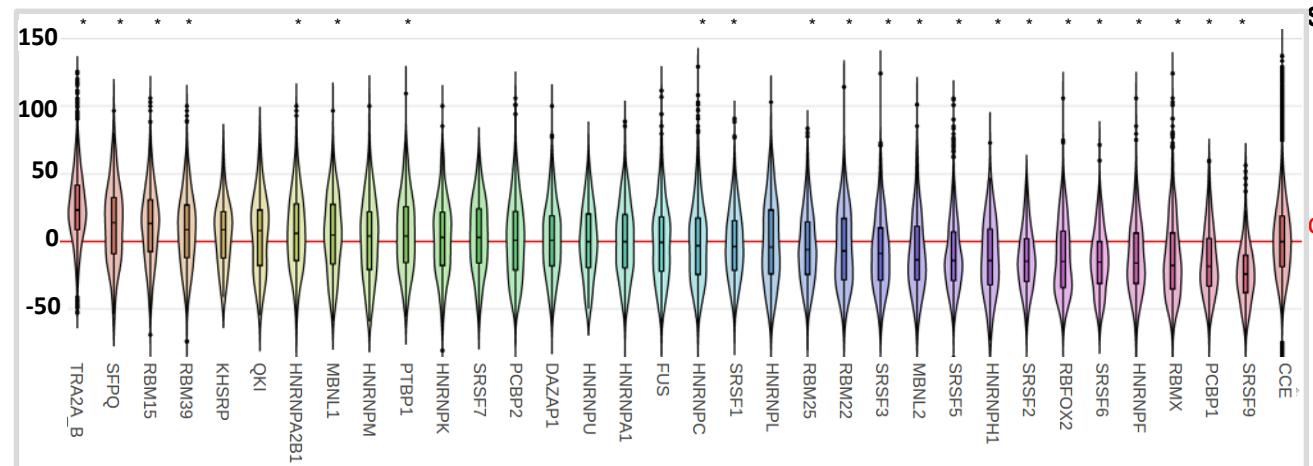
### Supplementary Figure S6

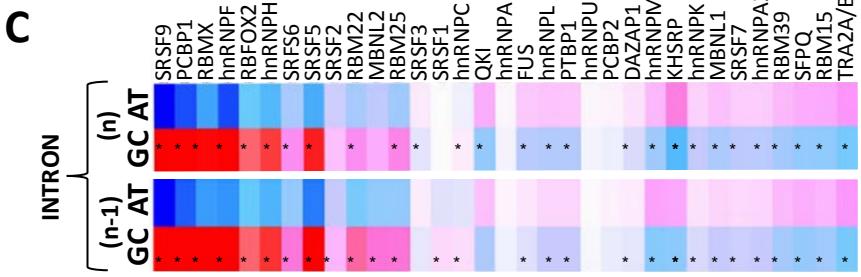
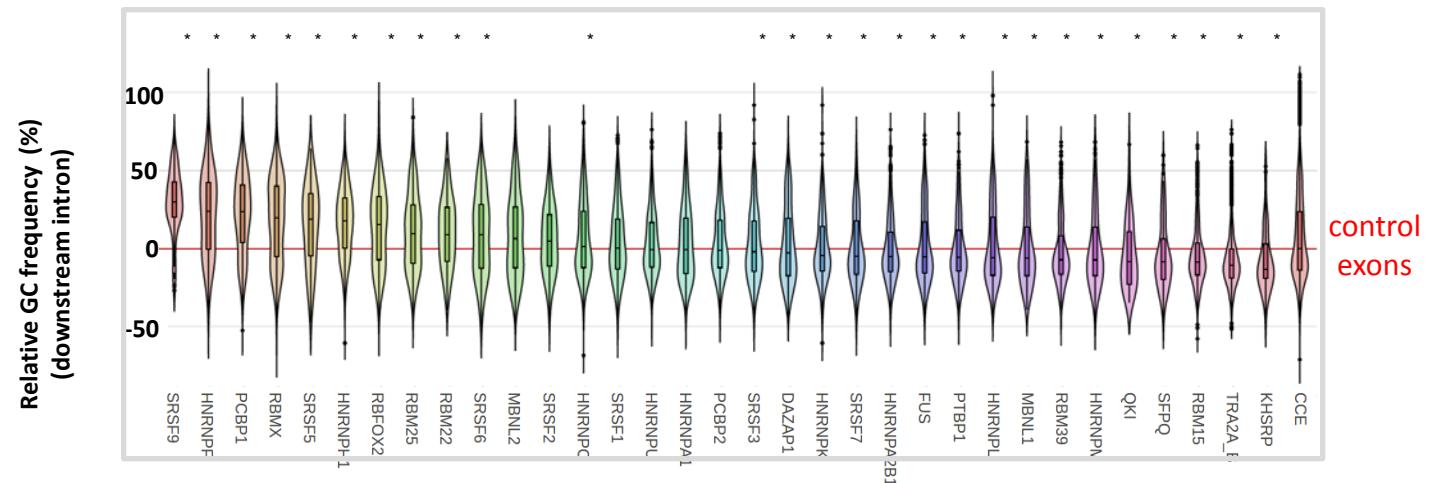
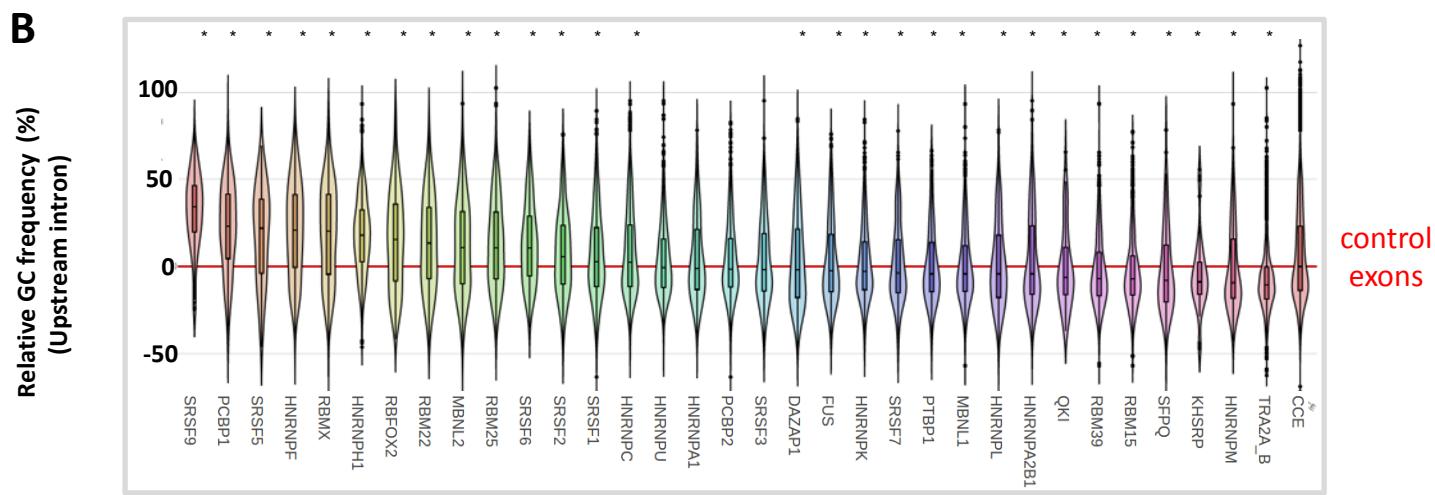
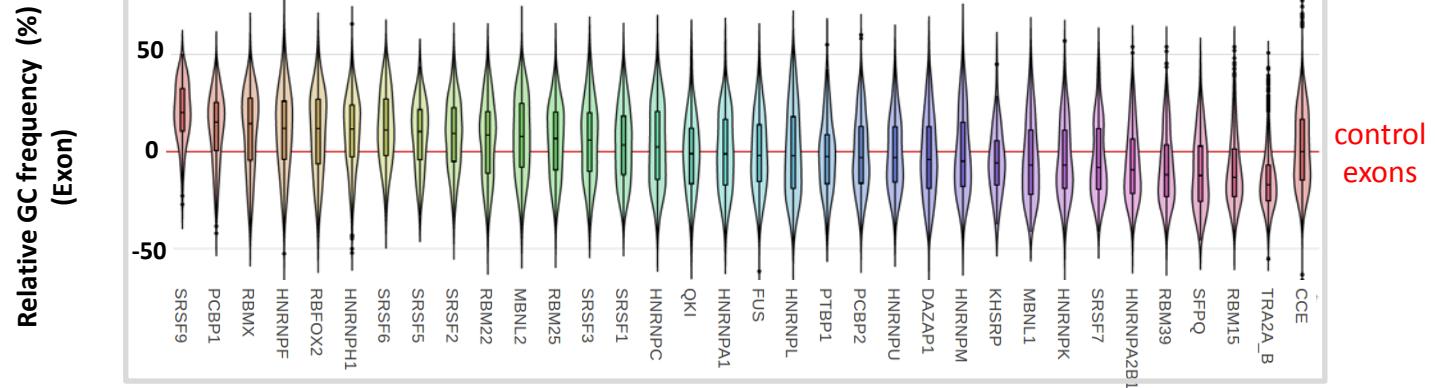
- A.** Correlation between the GC content of GC-exons and AT-exons and the GC content of their hosting isochores (left panel) defined by *Constantini et al* (*Constantini, M., Clay, O., Auletta, F. & Bernardi, G. An isochore map of human chromosomes. 2006; Genome Res 16, 536-541*). Proportion of AT-exons, GC-exons, and all human exons distributed across different isochore families defined by *Constantini et al.* (middle panel). Number of AT-exons and GC-exons present in individual isochores defined by *Constantini et al* (right panel). The left and right panels, represent, isochores containing preferentially GC-exons or AT-exons, respectively.
- B.** Same as in Supplementary Fig 6a but using isochores defined by Isosegmenter (<https://github.com/bunop/isoSegmenter>; Cozzi, P., Milanesi, L. & Bernardi, G. Segmenting the Human Genome into Isochores. 2015; Evolutionary bioinformatics online 11, 253-261).
- C.** Correlation between the GC content of GC-exons and AT-exons and the GC-content of their hosting TADs defined in the IMR90 cell line (left panel). Number of AT- and GC-exons present in individual TADs annotated from the IMR90 cell line (right panel).
- D.** Same as in Supplementary fig 6c but using TADs defined in the MCF7 cell line.

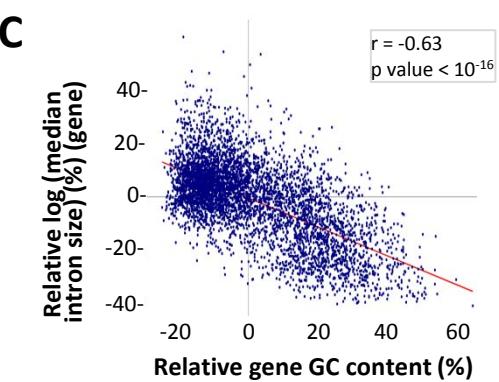
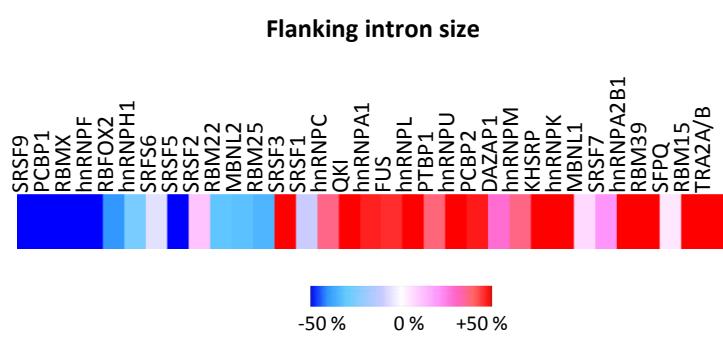
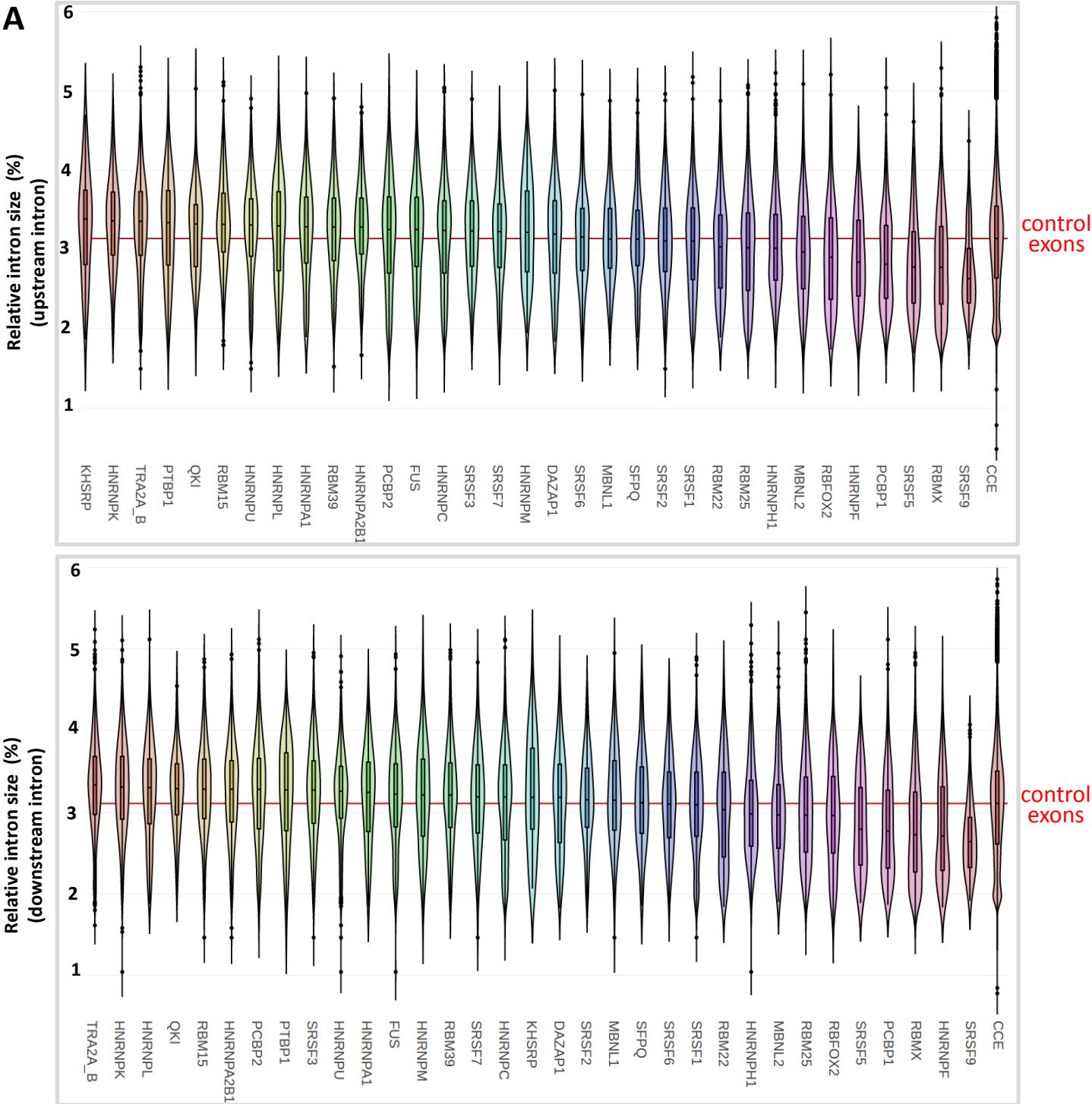
### Supplementary Figure S7

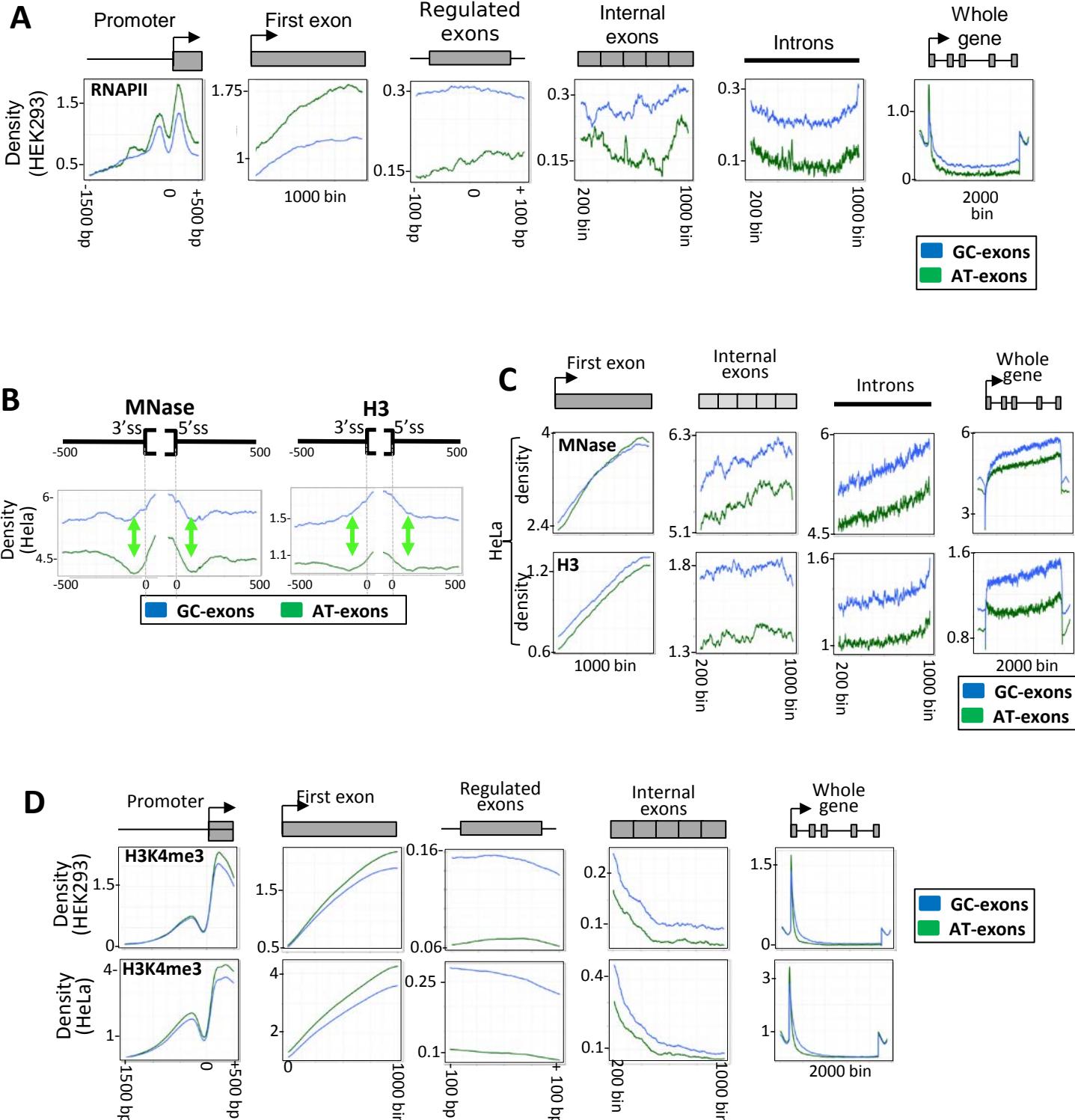
Splicing factor binding motifs retrieved from different resources. Splicing factors in blue color activate GC-exons, while splicing factors in green color activate AT-exons.

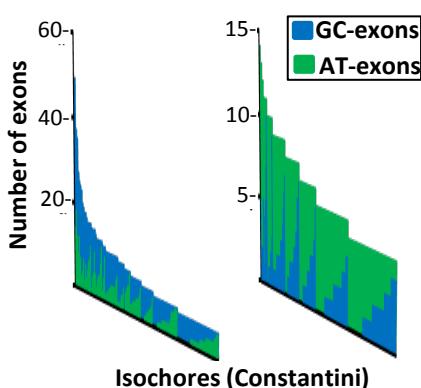
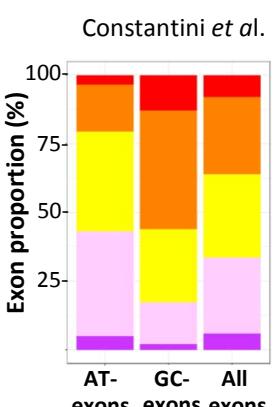
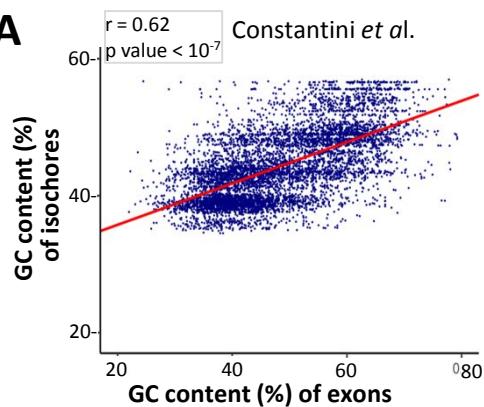
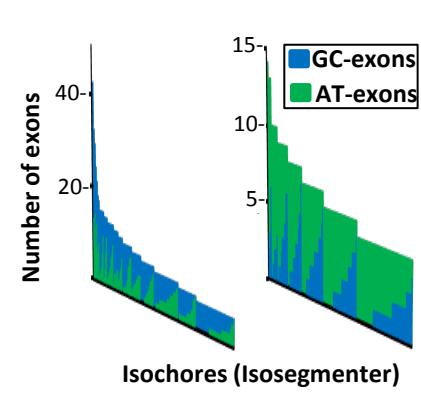
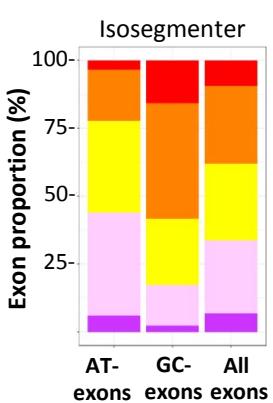
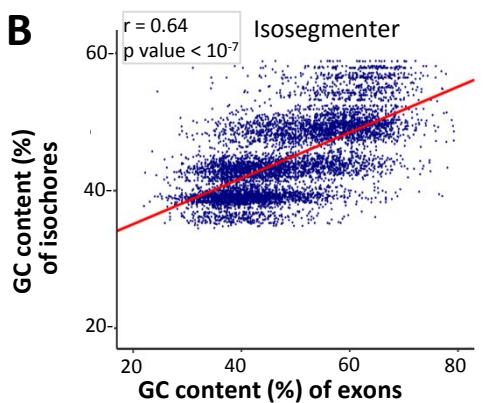
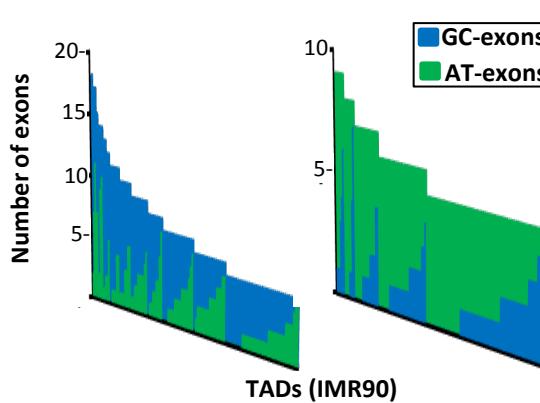
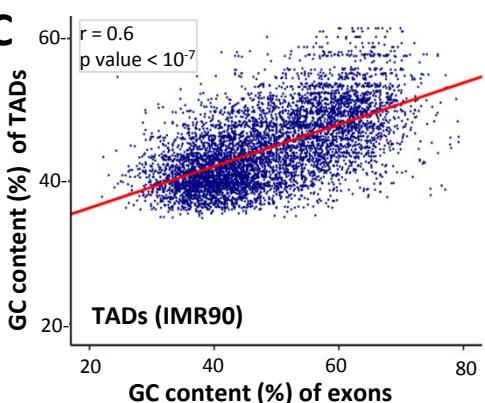
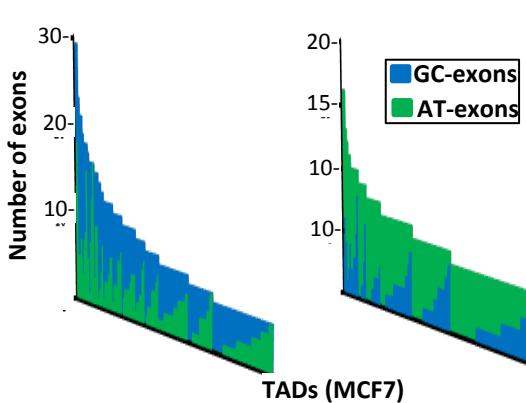
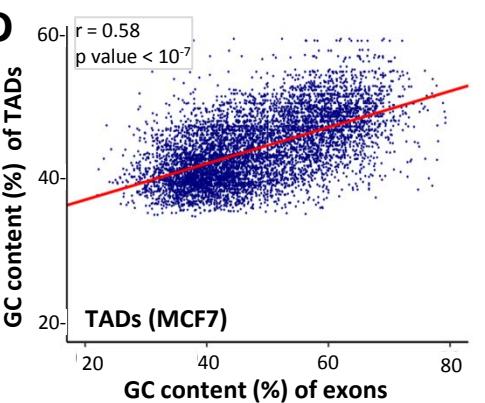










**A****B****C****D**

	Dominguez <sup>1</sup>	cisbp-rna <sup>2</sup>	ATTRACT <sup>3</sup>
SRSF1		GgA <b>G</b> G	A <b>C</b> GC <b>G</b> C
SRSF5	GCAGC		A C G <b>G</b> C
SRSF6		GGA <b>G</b> G	C <b>G</b> u <u>u</u>
SRSF9	GGA	A <b>G</b> A C <b>A</b>	AGGA
hnRNPF	GGGG		UGGGGU
hnRNPH			UGU <u>GGG</u>
PCBP1	CCC	CC <u>u</u> CC	
RBFOX2	GCA <u>G</u>		UC <u>CA</u> U <u>G</u>
RBM22	ACC <b>G</b> G		
RBM25	GGGG		
RBMX		CC <u>A</u>	
SFPQ	UAA	U <u>A</u> G <u>U</u> U	UGGAGAGAAC
DAZAP1	U <u>A</u> U <u>A</u>	U <u>A</u> G <u>U</u> U <u>A</u>	
KHSRP	U <u>GU</u> AU		UAGUAU
PTBP1		U <u>uu</u> u <u>cu</u>	UCUU
MBNL1	GC <u>u</u> ..	CC <u>U</u> U <u>cc</u>	u <u>GC</u> u
QKI		AC <u>U</u> AA <u>Ac</u>	A <u>UU</u> AA <u>AC</u>
TRA2A	GAA		AAGAAGAAGAA
hnRNPL	AC <u>A</u> CA	AC <u>A</u> ACA	
hnRNPA1		UAGG <u>G</u> A	UAGGG
SRSF7		AGAC.AC <u>G</u> A..	GGACGAC <u>GA</u>
hnRNPK	GCCC <b>A</b>	C <u>c</u> A <u>A</u> C <u>c</u>	CCC
hnRNPA2B1	GGGG	u <u>AGGG</u>	
FUS	GG <u>GG</u>	CGCG <u>c</u>	GGUG

Activation of GC-exons

Activation of AT-exons

<sup>1</sup> Dominguez et al. PMID 29883606<sup>2</sup> <http://attract.cnic.es><sup>3</sup> <http://cisbp-rna.ccbr.utoronto.ca/index.php>