Supplemental Fig S2

A. Color code of the relative frequency (%) of hexanucleotides in SRSF-activated and -repressed exons. After recovering the 10 most enriched hexanucleotides in SRSF-activated exons, their relative frequency was calculated by comparing it to the average frequency calculated in 10 000 sets of control exons. The relative frequency of these hexanucleotides was also calculated in SRSF-repressed exons. Red and green colors indicate when the frequency is higher and lower, respectively, in the set of regulated exons, when compared to the sets of control exons. The sets of SRSF-regulated exons originated from publicly available RNAseq datasets generated from the K562 (1), HepG2 (2), GM19238 (3), HeLa (4), K562 (5), Huh7 (6), HepG2 (7), GM19238 (8), and MDA-MB-231 (9) cell lines. Purine, SSNG motifs, Cs, and As are underlined in the enriched hexanucleotides identified in SRSF1-, SRSF2-, SRSF3-, and TRA2-activated exons, respectively. (*) Randomization test FDR < 0.05 (for the 10 most enriched hexanucleotides).

B. Relative frequency (%), when compared to sets of control exons of G nucleotides in SRSF1-regulated exons identified in different cell lines as described in Fig S2A (K562 (1), HepG2 (2), GM19238 (3), HeLa (4)). (*) Randomization test FDR < 0.003.

C. Relative frequency (%), when compared to sets of control exons of the G nucleotide in SRSF3-activated and -repressed exons identified in HepG2 (1), GM19238 (2). (*) Randomization test FDR < $1x10^{-4}$.

D. Color code of the relative frequency (%) of dinucleotides, compared to sets of control exons, in SRSF-activated and SRSF-repressed exons across different cell lines. The sets of SRSF-regulated exons originated from publicly available RNAseq datasets generated from the K562 (1), HepG2 (2), GM19238 (3), HeLa (4), K562 (5), Huh7 (6), HepG2 (7), GM19238 (8), and MDA-MB-231 (9) cell lines. The frequency of each dinucleotide was calculated in SRSF-activated and SRSF-repressed exons and expressed as the % of the average frequency calculated in sets of control exons. Red and green colors indicate when the dinucleotide frequency is higher and lower, respectively, in the sets of regulated exons when compared to sets of control exons. Only the two most enriched dinucleotides in SRSF-activated compared to SRSF-repressed exons are represented. (*) Randomization test FDR < 0.05.

E. Relative frequency (%), when compared to sets of control exons, of glycine (Gly corresponding to GGN codons), alanine (Ala corresponding to GCN codons), proline (Pro corresponding to CCN codons), and lysine (Lys corresponding to AAR codons) encoded by SRSF1-activated and -repressed exons identified across different cells lines. The sets of SRSF1-regulated exons originated from publicly available RNAseq datasets generated from the K562 (1), HepG2 (2), GM19238 (3), and HeLa (4), cell lines. (**) Randomization test FDR < 0.05.

F. Relative frequency (%), when compared to sets of control exons, of G and C nucleotides in HNRNPK-repressed exons identified in K562 (1) and HepG2 (2) cells. (**) Randomization test FDR < 0.002.

G. Relative frequency (%), when compared to sets of control exons, of C (C1-2) nucleotides at the first and second codon position from HNRNPK-repressed exons identified in K562 (1) and HepG2 (2) cells. (**) Randomization test FDR < $1x10^{-4}$.

H. Relative frequency (%), when compared to sets of control exons, of glycine (Gly corresponding to GGN codons) and proline (Pro corresponding to CCN codons) encoded by HNRNPK-repressed exons identified in K562 (1) and HepG2 (2) cells. (**) Randomization test FDR < $1x10^{-4}$.

I. Relative frequency (%), when compared to sets of control exons, of polar uncharged, charged or neutral (in terms of hydropathy) amino acids encoded by HNRNPK-repressed exons identified in K562 (1) and HepG2 (2) cells. (**) Randomization test FDR < 0.005.