SUPPLEMENTARY DATA

Splicing of Lon Non-Coding RNAs Primarily Depends on Polypyrimidine Tract and 5’ Splice-Site Sequences Due to Weak Interactions with SR Proteins

Zuzana Krchňáková, Prasoon Kumar Thakur, Michaela Krausová, Nicole Bieberstein, Nejc Haberman, Michaela Müller-McNicoll and David Staněk

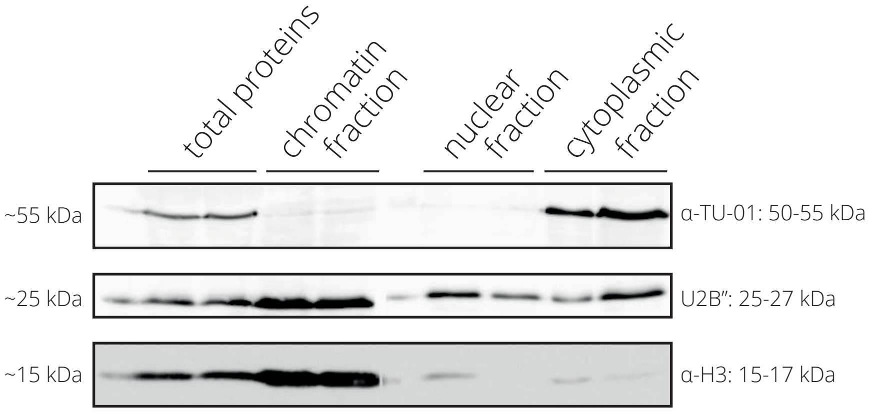


Figure S1. Western blot line of total proteins and proteins from cellular fractionations of HeLa cells (two biological replicates). Anti-a-tubulin, anti-U2B'', and anti-histone3 antibodies detect the cytoplasmic, nucleoplasmic, and chromatin fractions, respectively.

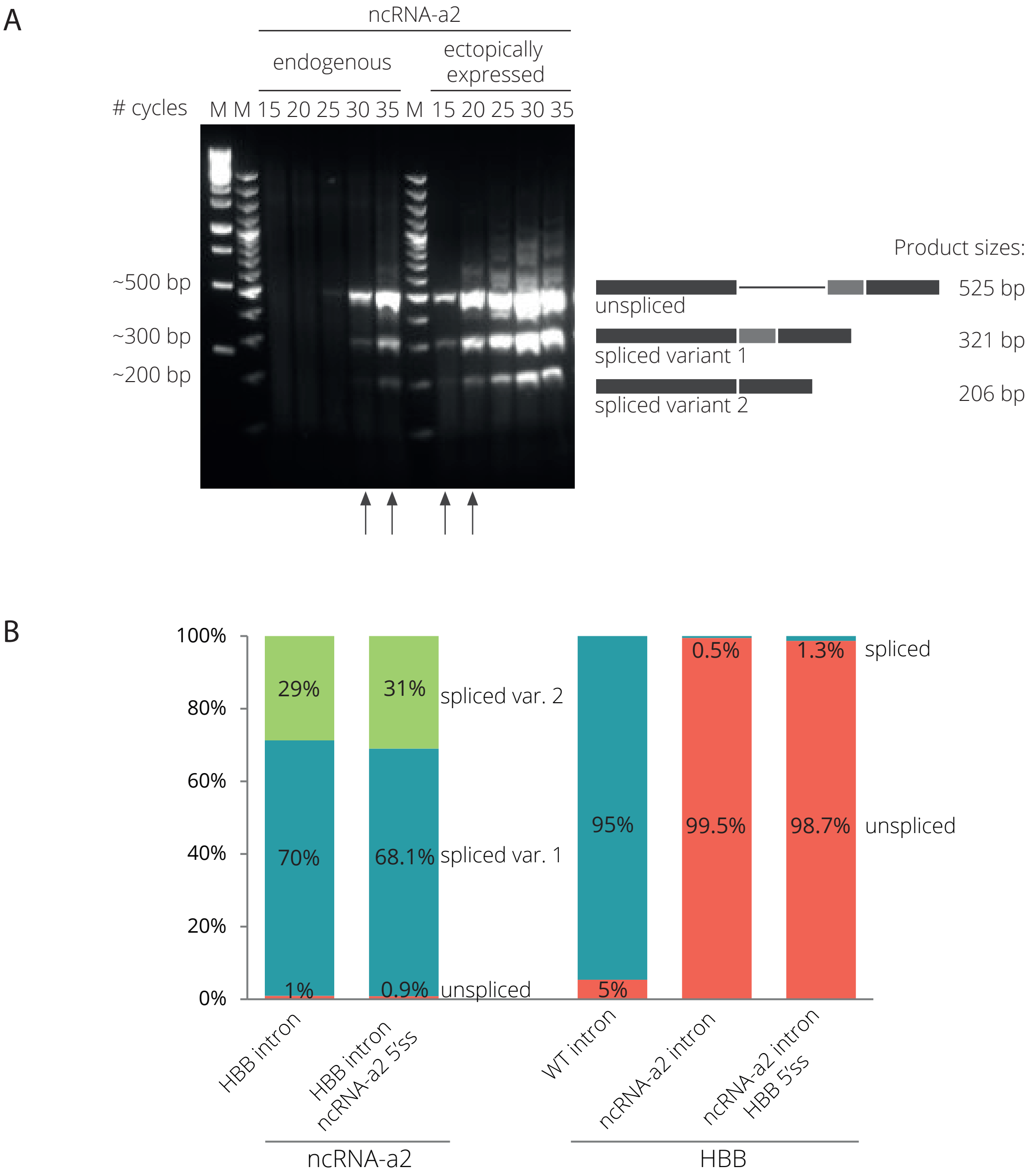


Figure S2. (**A**) Splicing efficiency of endogenous and ectopically expressed ncRNA-a2 measured by semiquantitative RT-PCR. Results are quantified from experiments indicated by arrows; M - marker. (**B**) Fractions of spliced and unspliced transcripts of ncRNA-a2 and HBB after intron swopping.

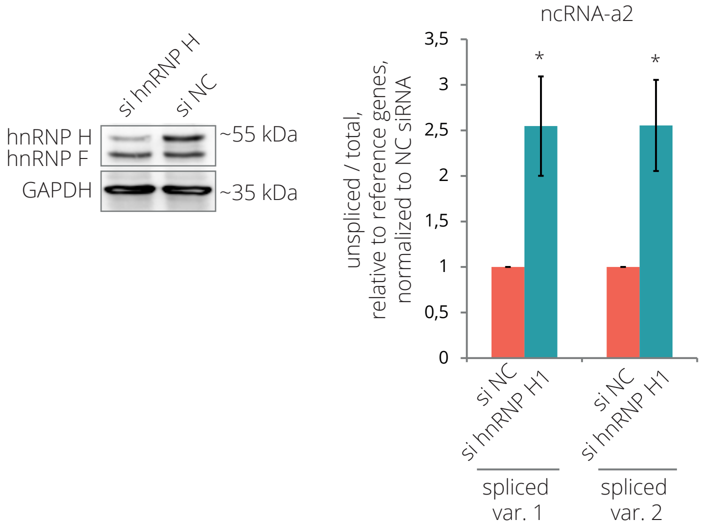


Figure S3. The fraction of unspliced transcripts of the ncRNA-a2 gene after hnRNP H siRNA knock-down. Barplots show relative RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the hnRNP H knock-down with negative control, $^*p<0.05$.

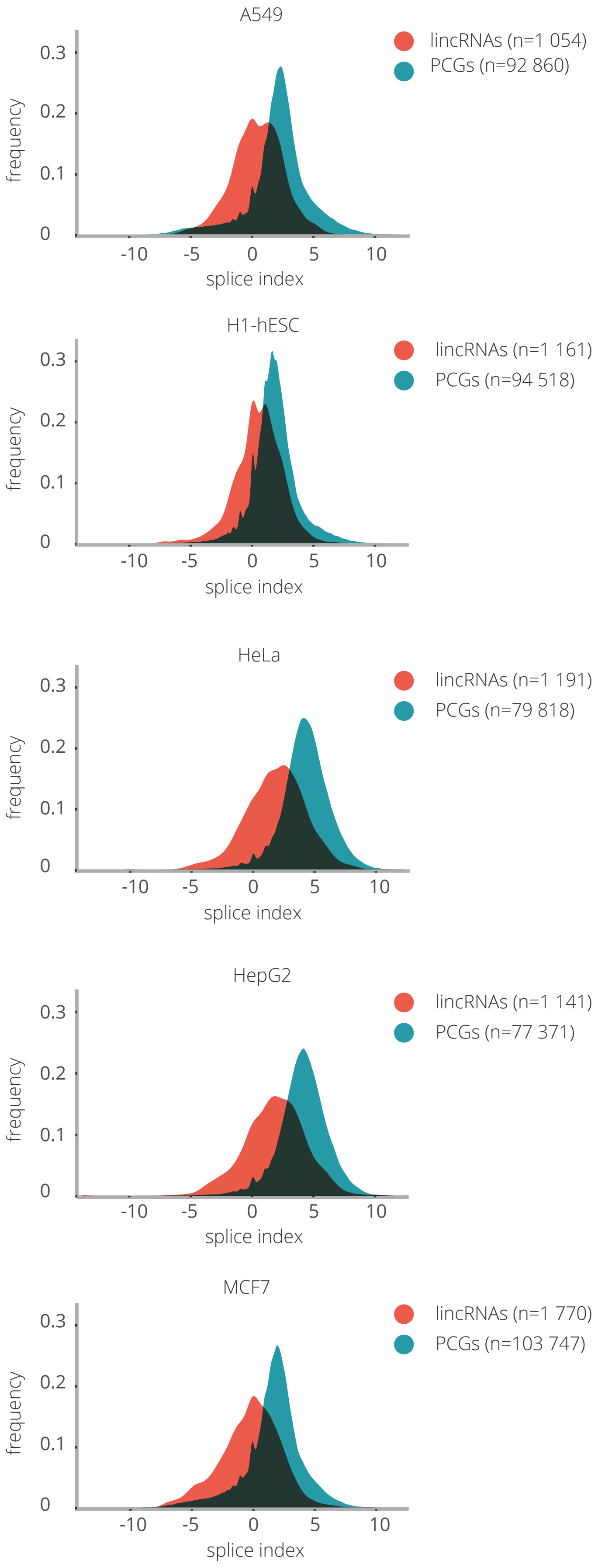
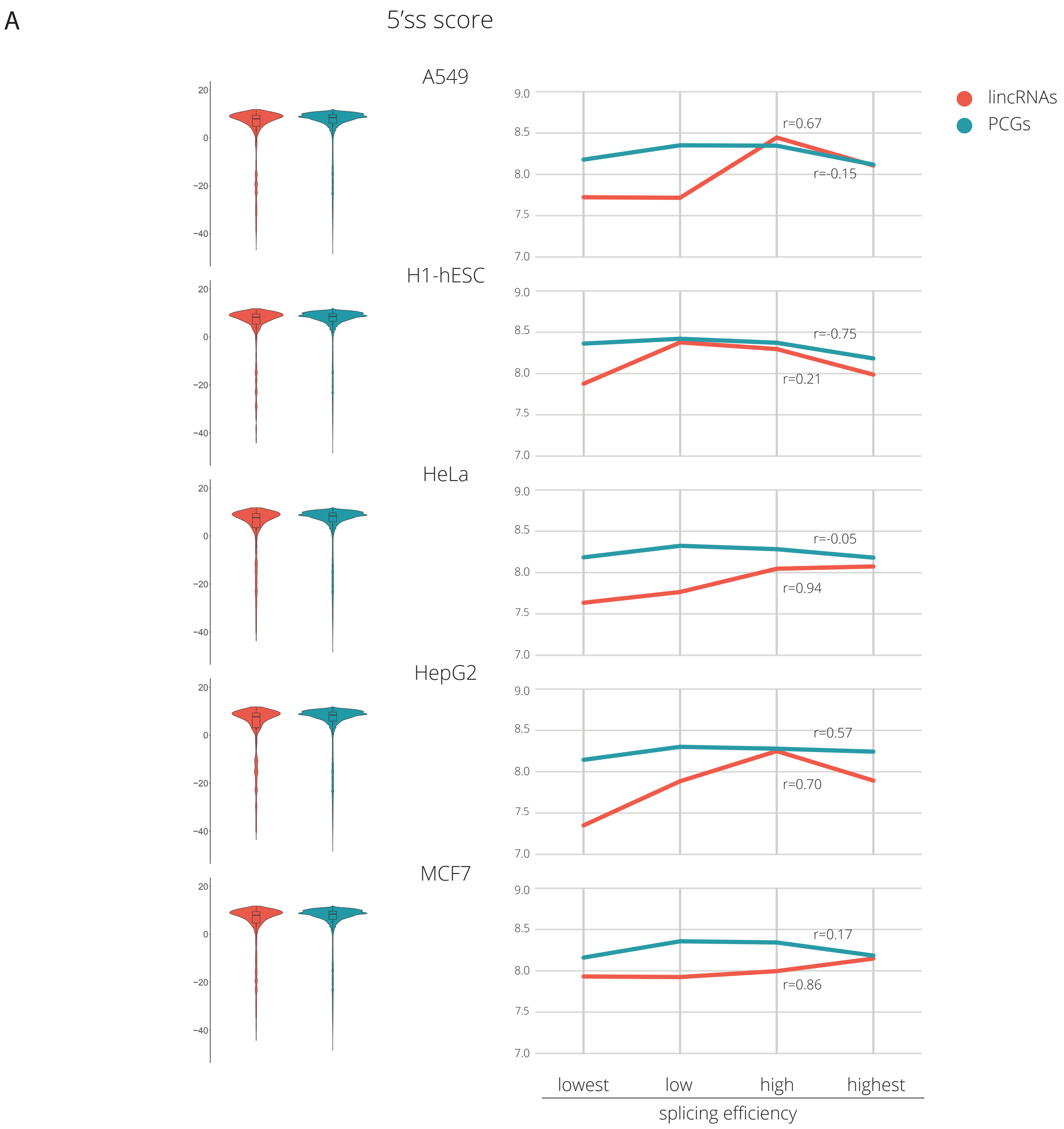


Figure S4. Splice index scores of introns in lincRNAs and PCGs in five different human cell lines; n = the number of introns.



B

Average splicing index (SI)

splicing efficiency	lincRNAs A549	lincRNAs H1-hESC	lincRNAs HeLa	lincRNAs HepG2	lincRNAs MCF7	PCGs A549	PCGs H1-hESC	PCGs HeLa	PCGs HepG2	PCGs MCF7
lowest	-2.17	-1.98	-1.25	-1.36	-3.44	-0.83	-0.29	1.90	1.67	-1.26
low	-0.18	0.04	1.24	1.17	-0.75	1.83	1.32	3.82	3.66	1.41
high	1.23	1.18	2.76	2.72	0.72	2.80	2.17	4.88	4.74	2.44
highest	2.98	2.79	4.82	4.91	2.72	4.97	4.06	6.67	6.50	4.37

Figure S5. (**A**) 5'ss score distribution in lincRNAs and PCGs in five different cell lines (left panel); 5'ss scores divided according to splicing efficiency of lincRNAs and PCGs (right panel); r - Pearson's correlation coefficient between 5'ss score and splicing index. (**B**) A table of mean splicing indexes (SIs) of lincRNAs and PCGs in four groups divided according to the splicing efficiencies.

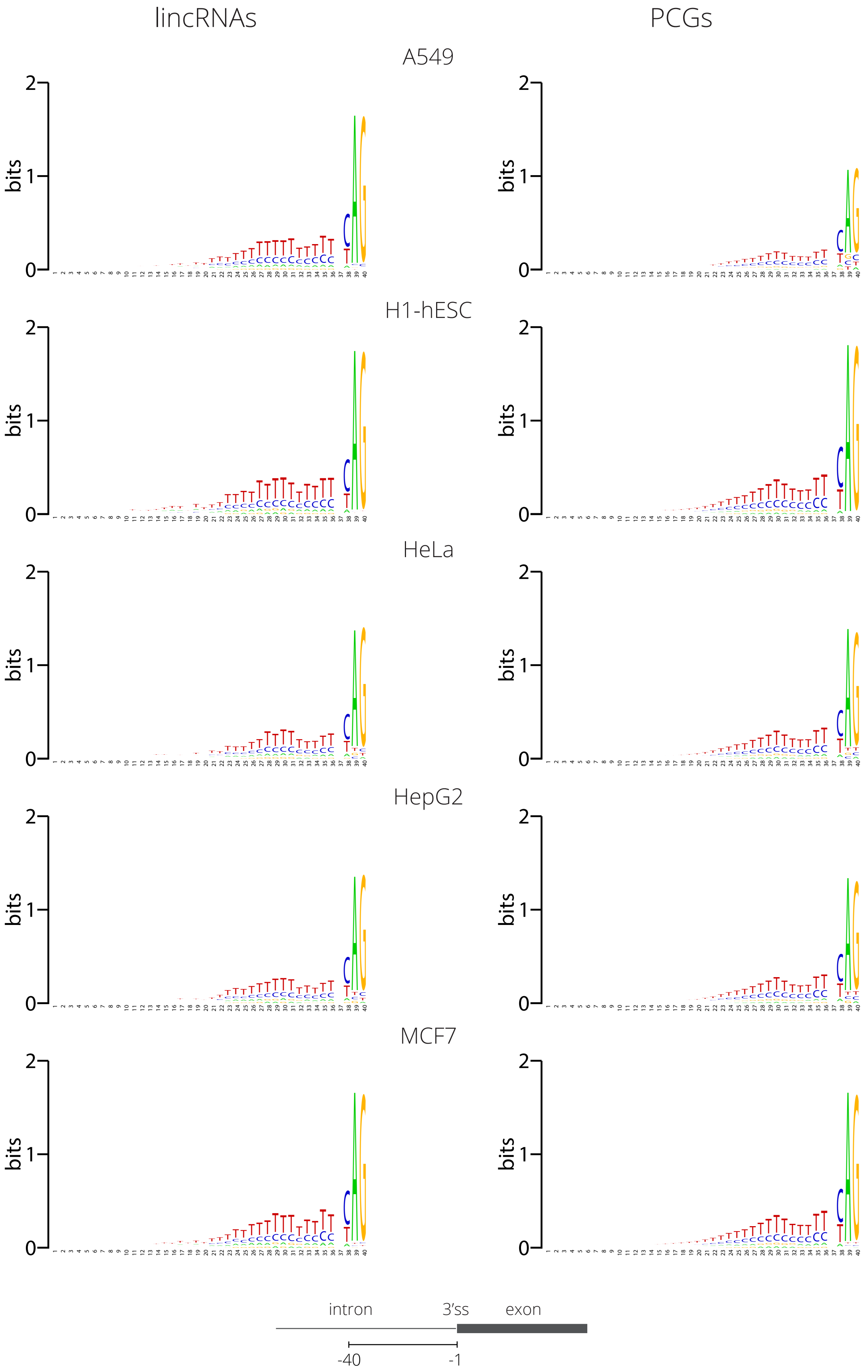


Figure S6. Sequence logos of 3'ss and PPT regions in lincRNAs and PCGs in different cell lines.

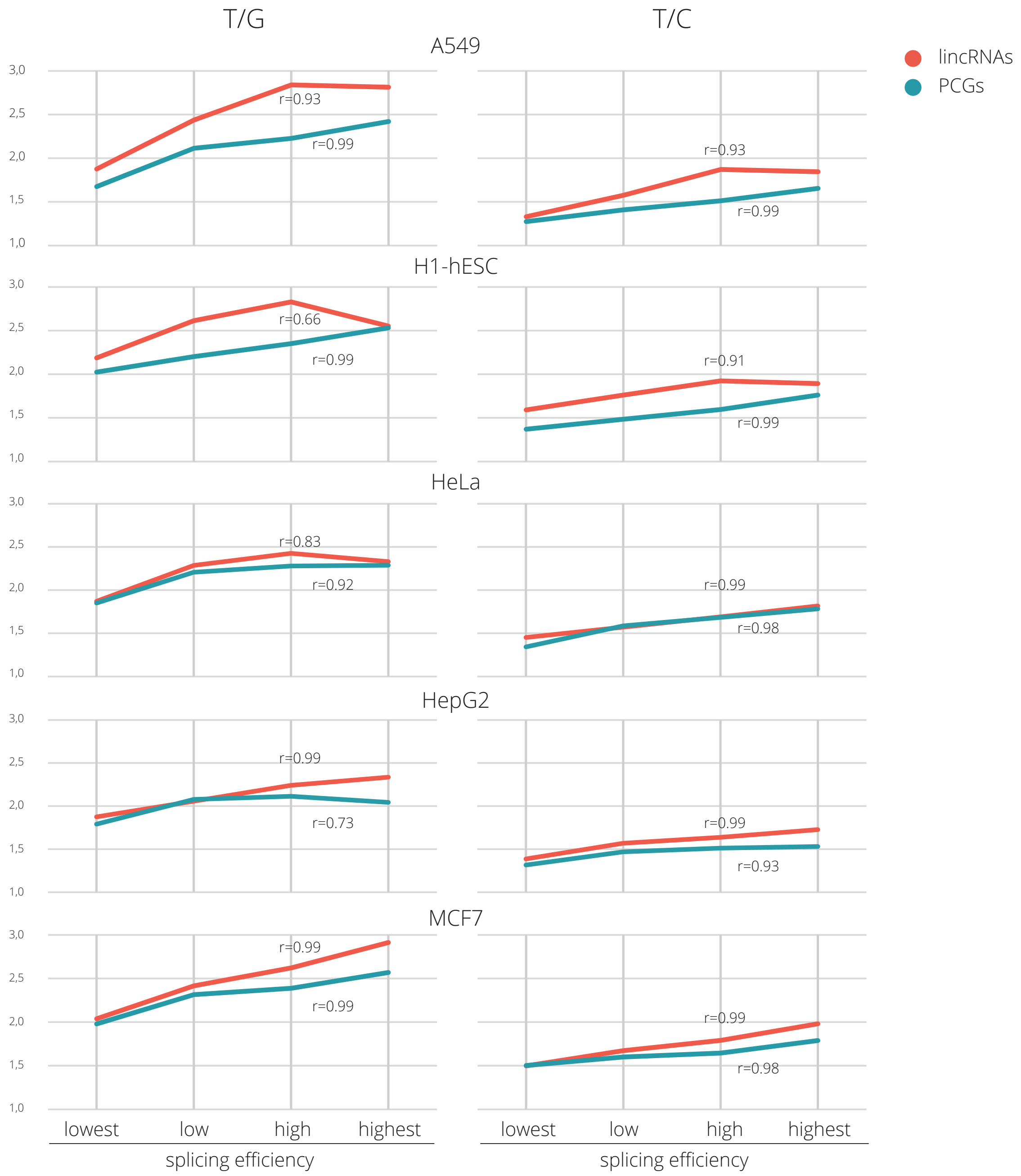


Figure S7. Thymidines over guanidines (T/G) or cytidines (T/C) ratio frequencies in the PPT region in different cell lines divided according to splicing efficiency in lincRNAs and PCGs; r - Pearson's correlation coefficient between T content in PPT and splicing index. Splicing groups are the same as at Figure S5.

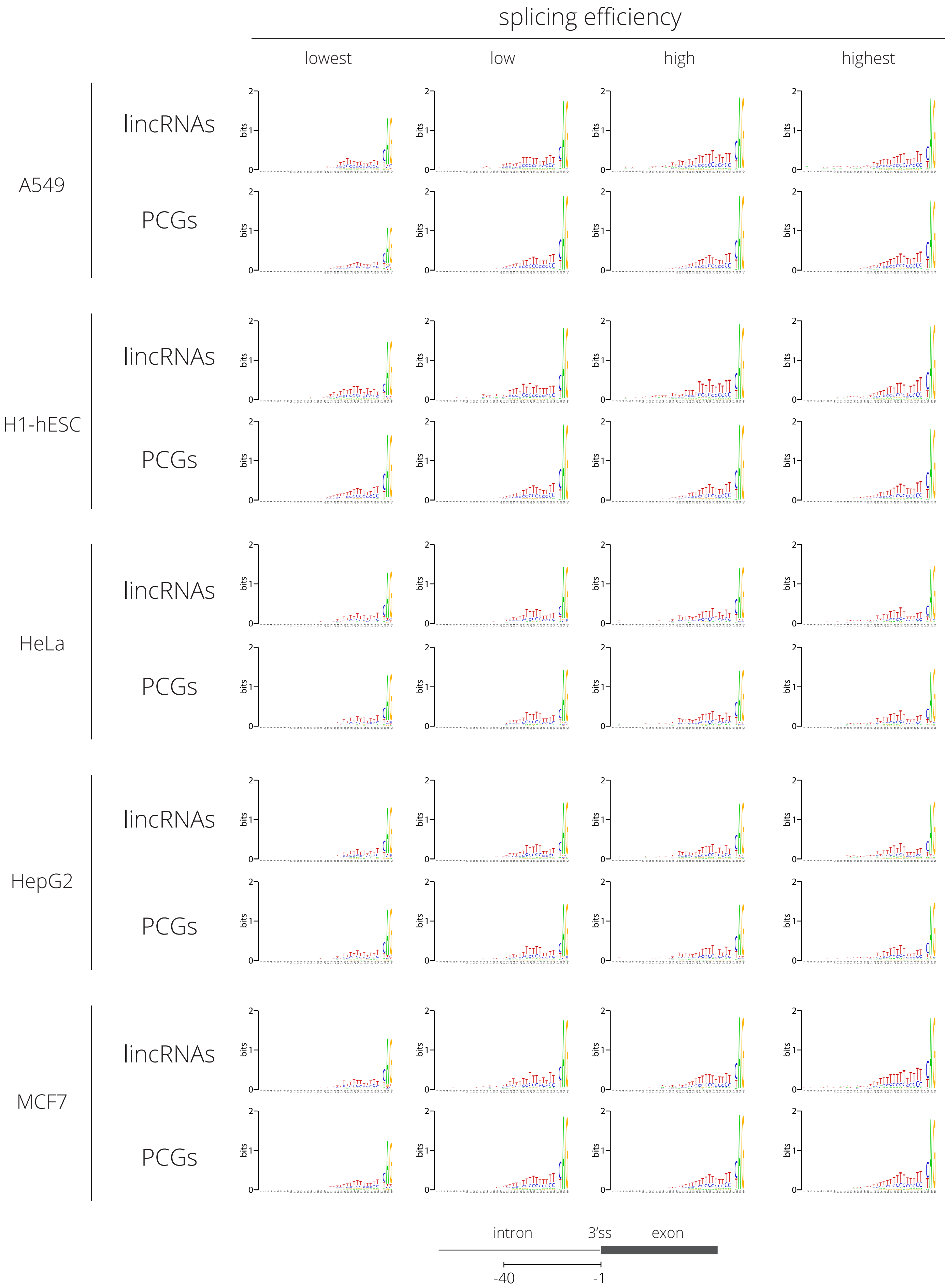


Figure S8. Sequence logos in 3’ss and PPT regions in lincRNAs and PCGs in different cell lines divided according to splicing efficiency.

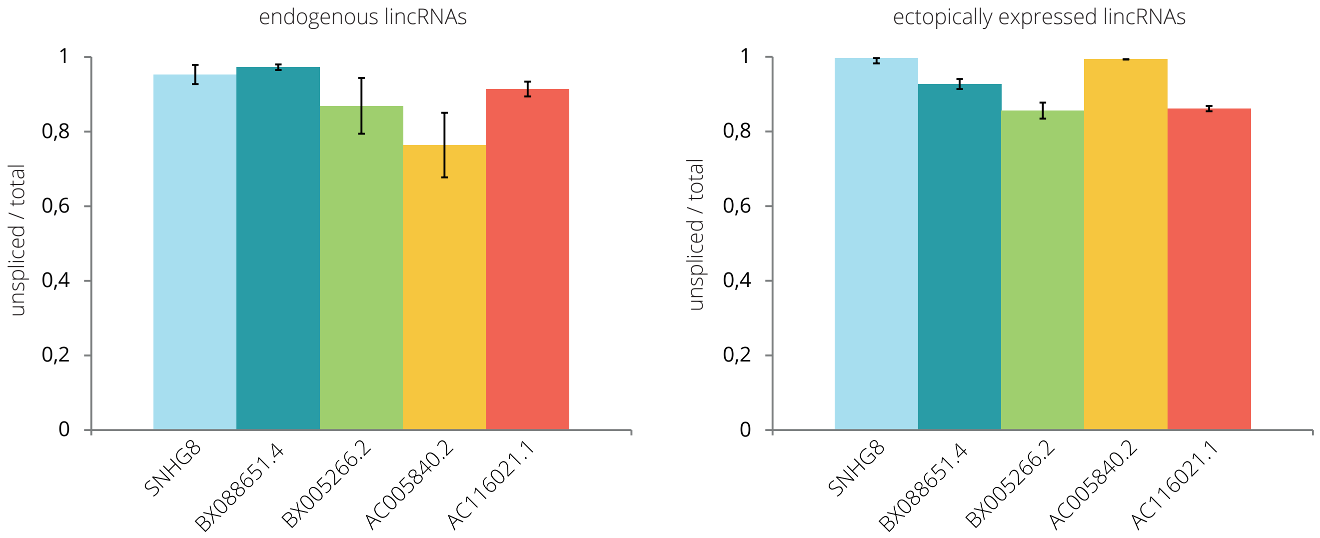
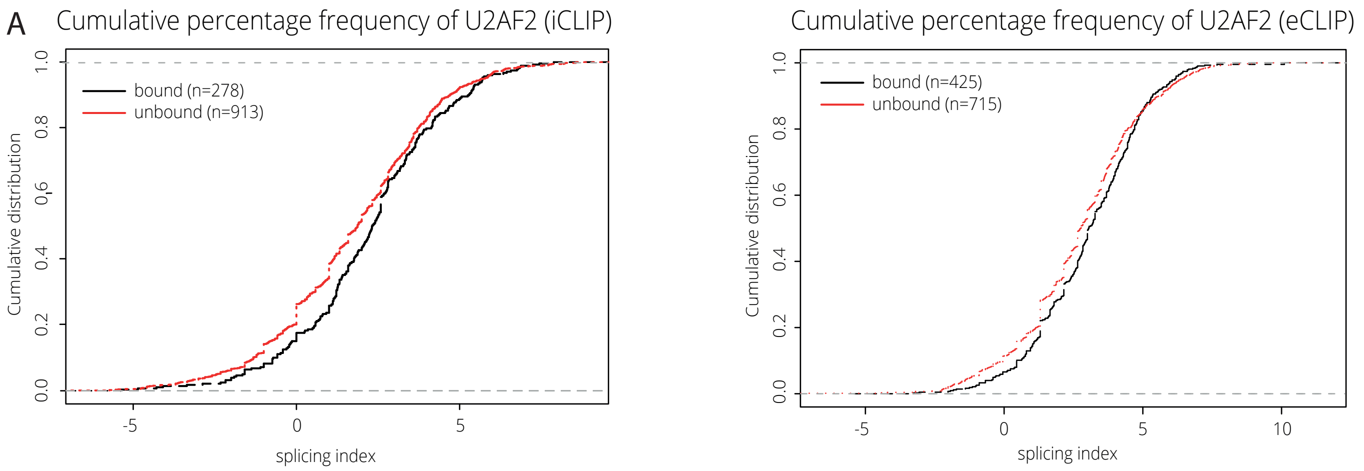


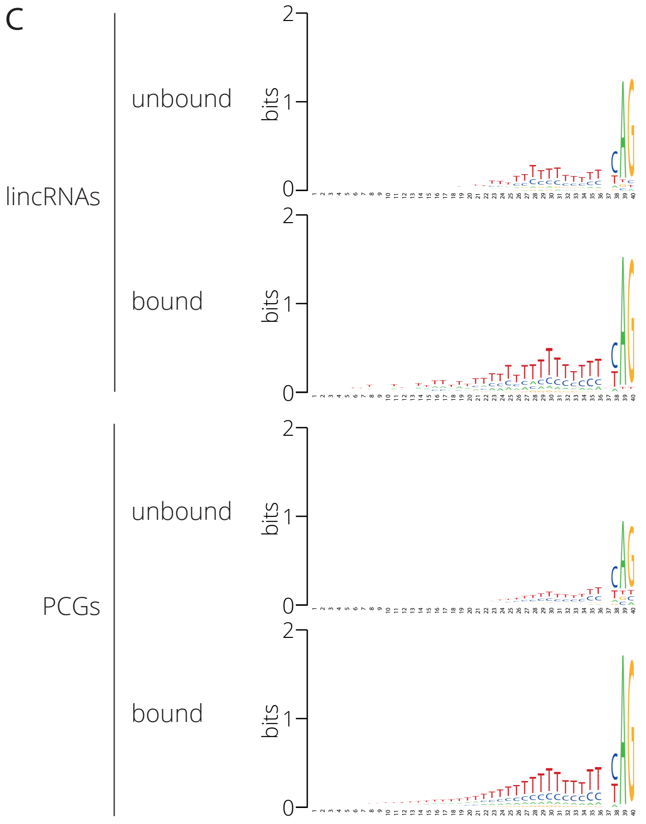
Figure S9. Inefficient splicing of selected lincRNAs. Splicing efficiencies of endogenous lincRNAs (left) and splicing efficiencies of transiently expressed lincRNAs (right). Splicing efficiencies are measured as a fraction of unspliced transcripts relative to total amount of transcripts. Bar-plots show RNA levels as determined by RT-qPCR. Average of at least three independent experiments is shown. Error bars indicate ± SEM.

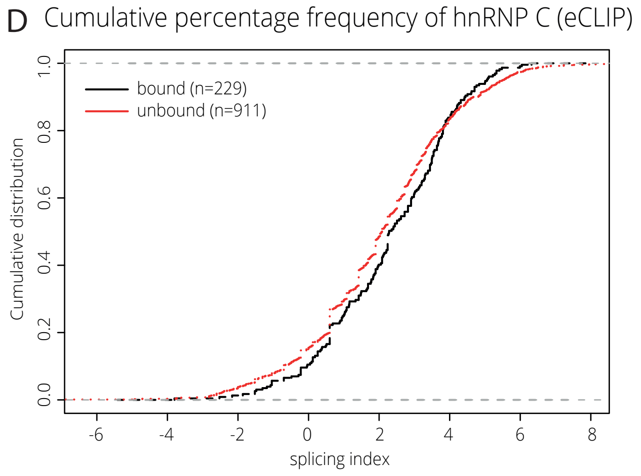


B

Sequence composition of introns bound/unbound by UA2F2 (iCLIP)

		A	C	G	T
lincRNAs	unbound	0.22	0.24	0.18	0.36
	bound	0.22	0.21	0.15	0.42
PCGs	unbound	0.20	0.27	0.21	0.32
	bound	0.22	0.21	0.15	0.42





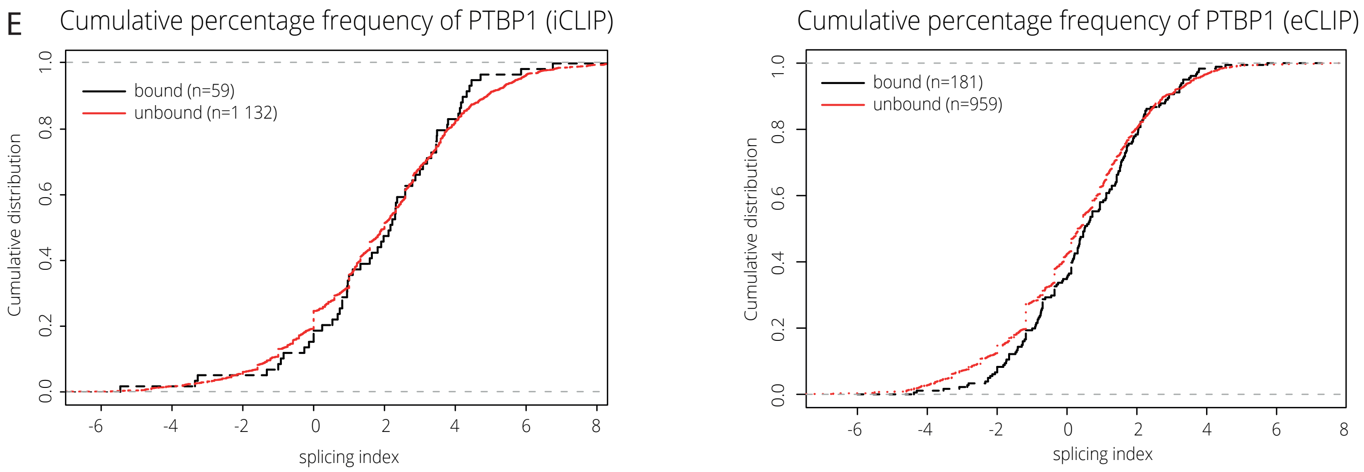


Figure S10. Cumulative distribution of splicing efficiency (**A**), sequence composition (**B**) and logos of 40 nt upstream of 3’ss (**C**) of lincRNA introns bound and unbound by U2AF2. Cumulative distribution of splicing efficiency comparing hnRNP C-bound and unbound lincRNAs (**D**), and PTBP1-bound and unbound lincRNAs (**E**); n=number of genes. PCGs serves as a control.

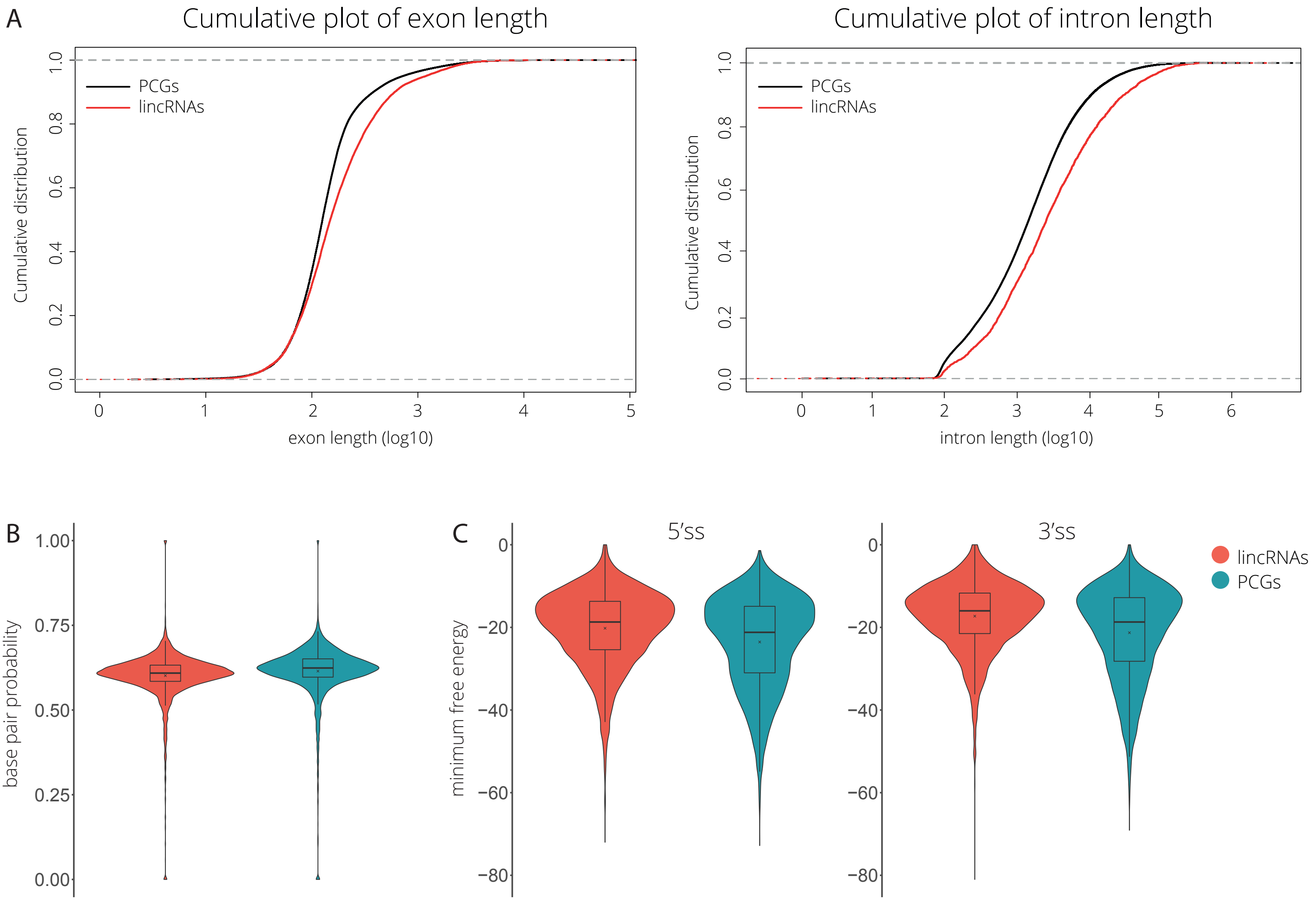


Figure S11. (**A**) Cumulative distributions of exon and intron lengths in lincRNAs and PCGs. (**B**) Base pair probability of 5000 randomly chosen lincRNAs and PCGs shorter than 2000 nts. (**C**) Minimum free energies of RNA sequences 100 nt upstream and 100 nt downstream of 5'ss (left) and 3'ss (right). 5000 randomly chosen introns from lincRNAs and PCGs were analyzed.

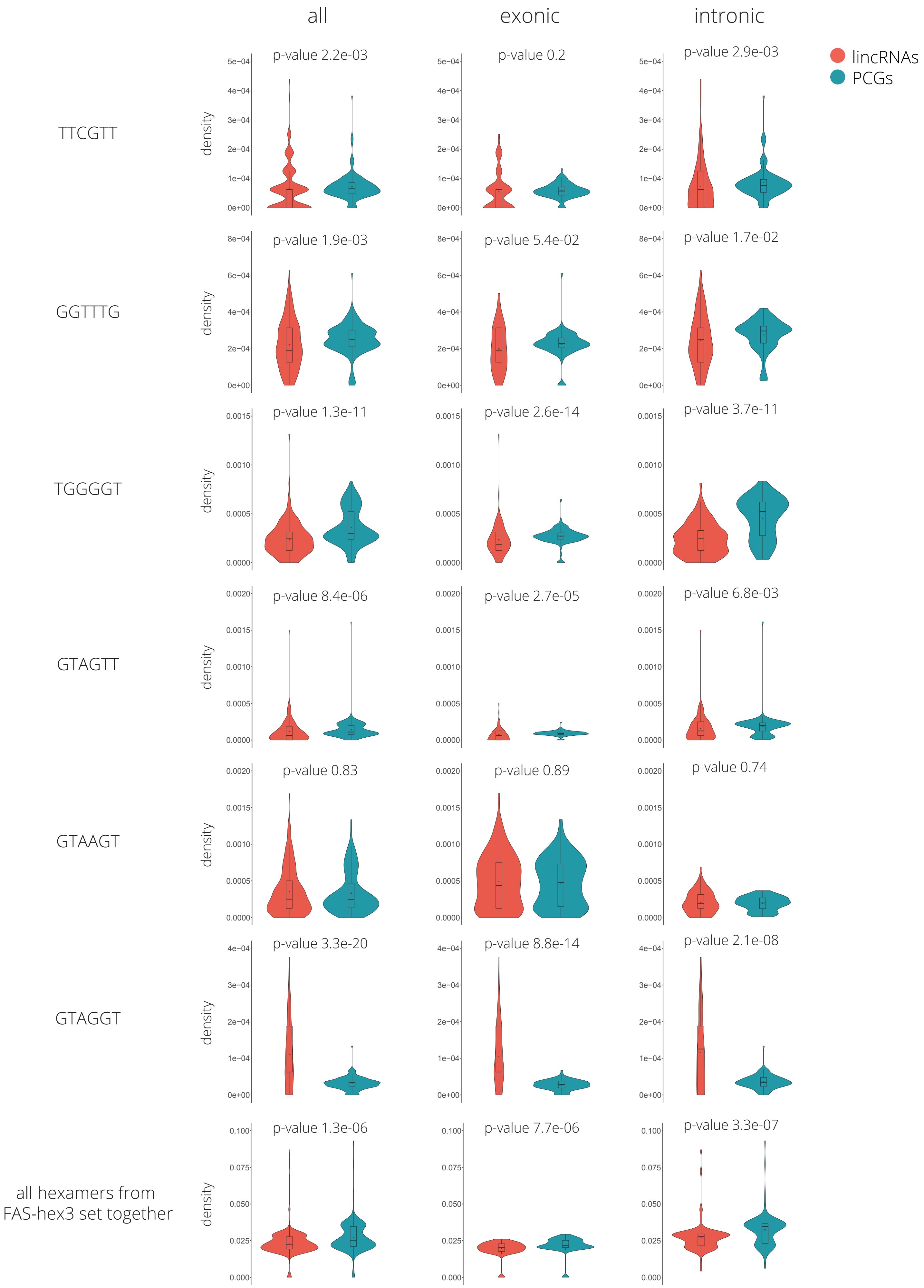


Figure S12. Splicing silencer motif distributions 100 nt upstream (intronic) and 100 nt downstream (exonic) of 3’ss in lincRNAs and PCGs. Motifs taken from Wang et al. (2004) and FAS-ESS web server (http://genes.mit.edu/fas-ess/). P-values are calculated by Wilcoxon rank sum test.

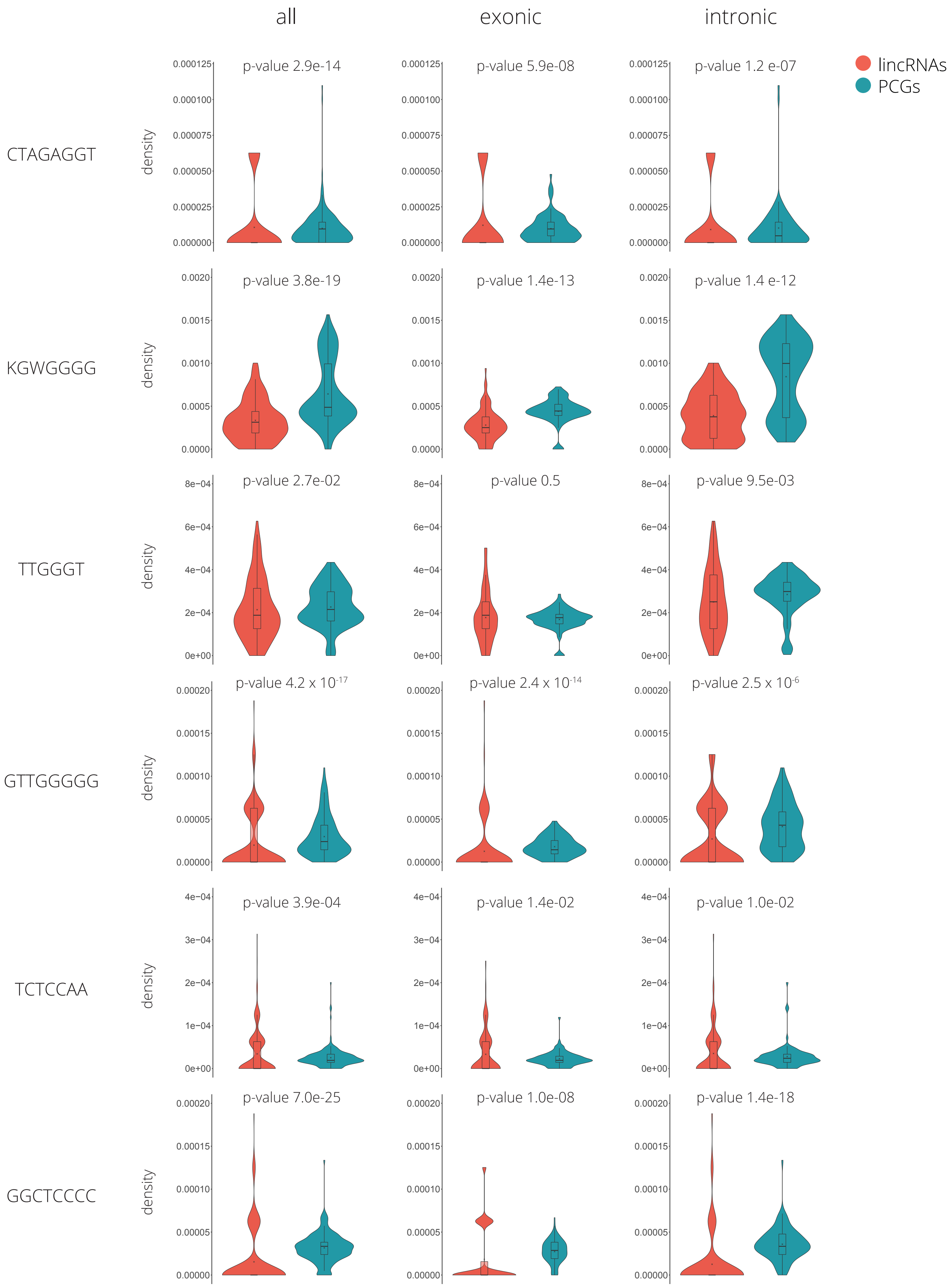


Figure S13. Splicing silencer motif distributions 100 nt upstream (intronic) and 100 nt downstream (exonic) of 3’ss in lincRNAs and PCGs. Motifs taken from Sironi et al. (2004). P-values are calculated by Wilcoxon rank sum test.

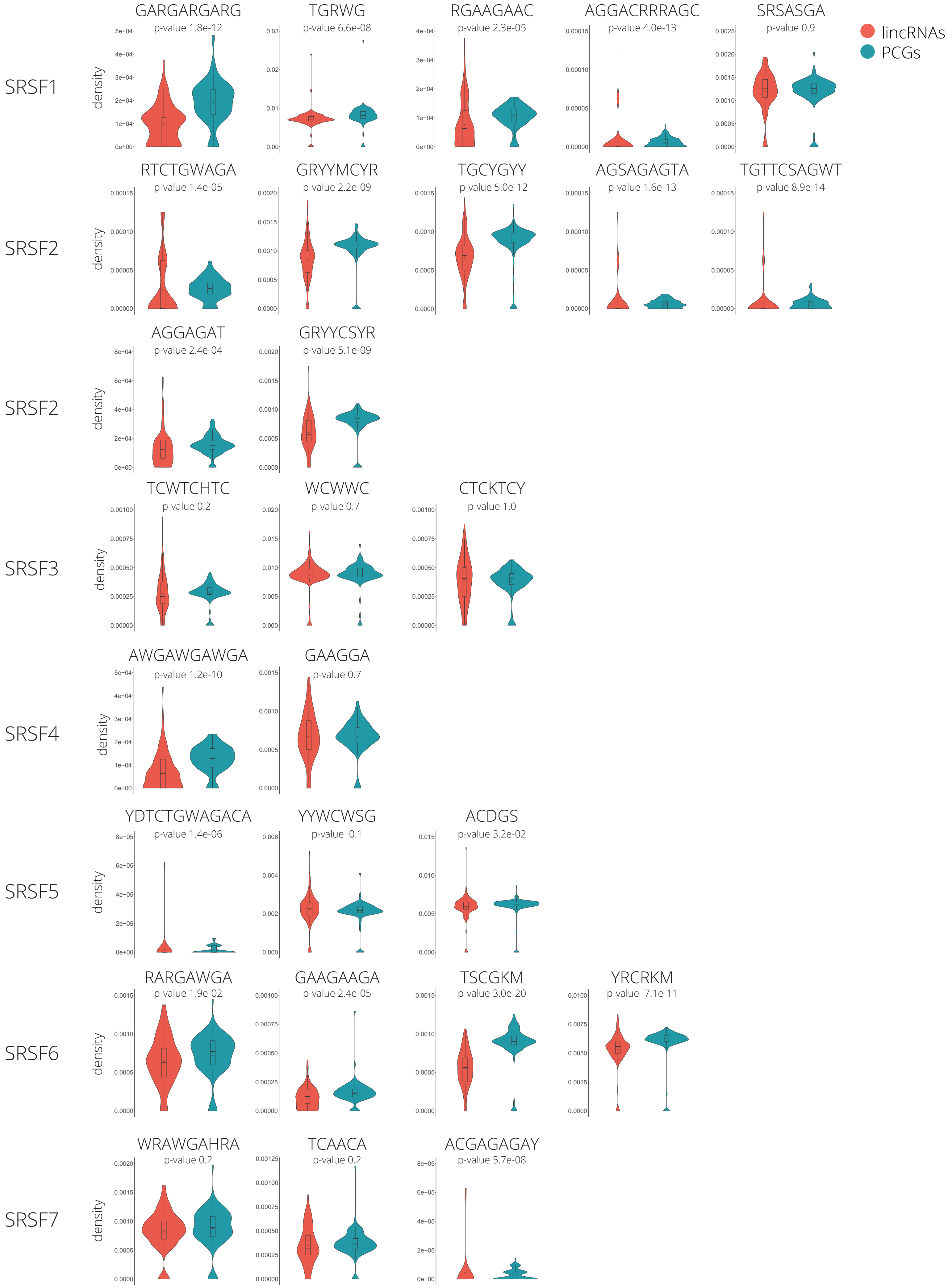


Figure S14. SRSF binding motif distributions in lincRNAs and PCGs 100 nt upstream of 5'ss. References for motifs are provided in Supplementary Data. P-values are calculated by Wilcoxon rank sum test. Two binding motifs of SRSF9 are not shown because no motifs were found in lincRNAs.

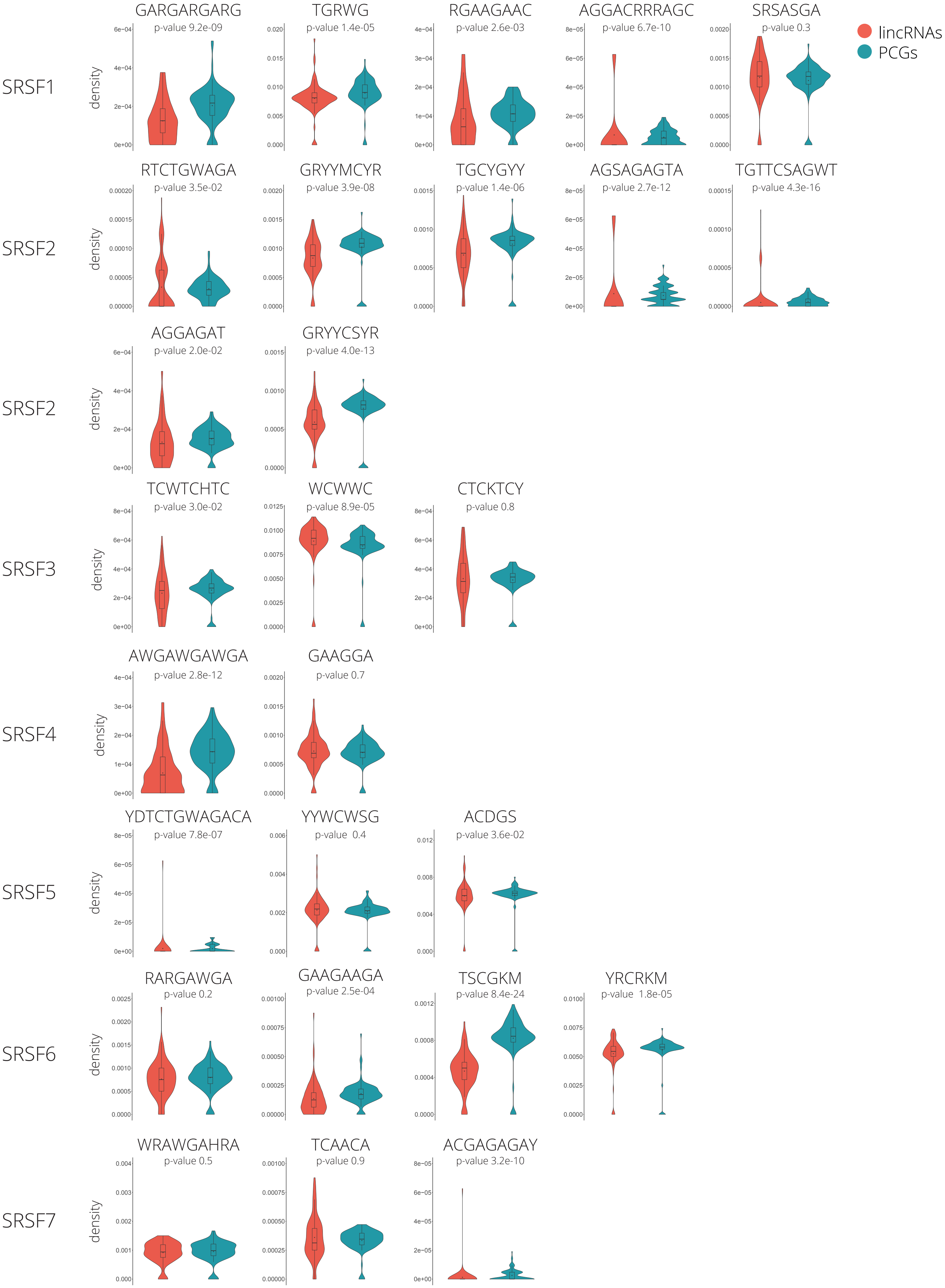


Figure S15. SRSF binding motif distributions in lincRNAs and PCGs 100 nt downstream of 3’ss. References for motifs are provided in Supplementary Data. P-values are calculated by Wilcoxon rank sum test. Two binding motifs of SRSF9 are not shown because no motifs were found in lincRNAs.

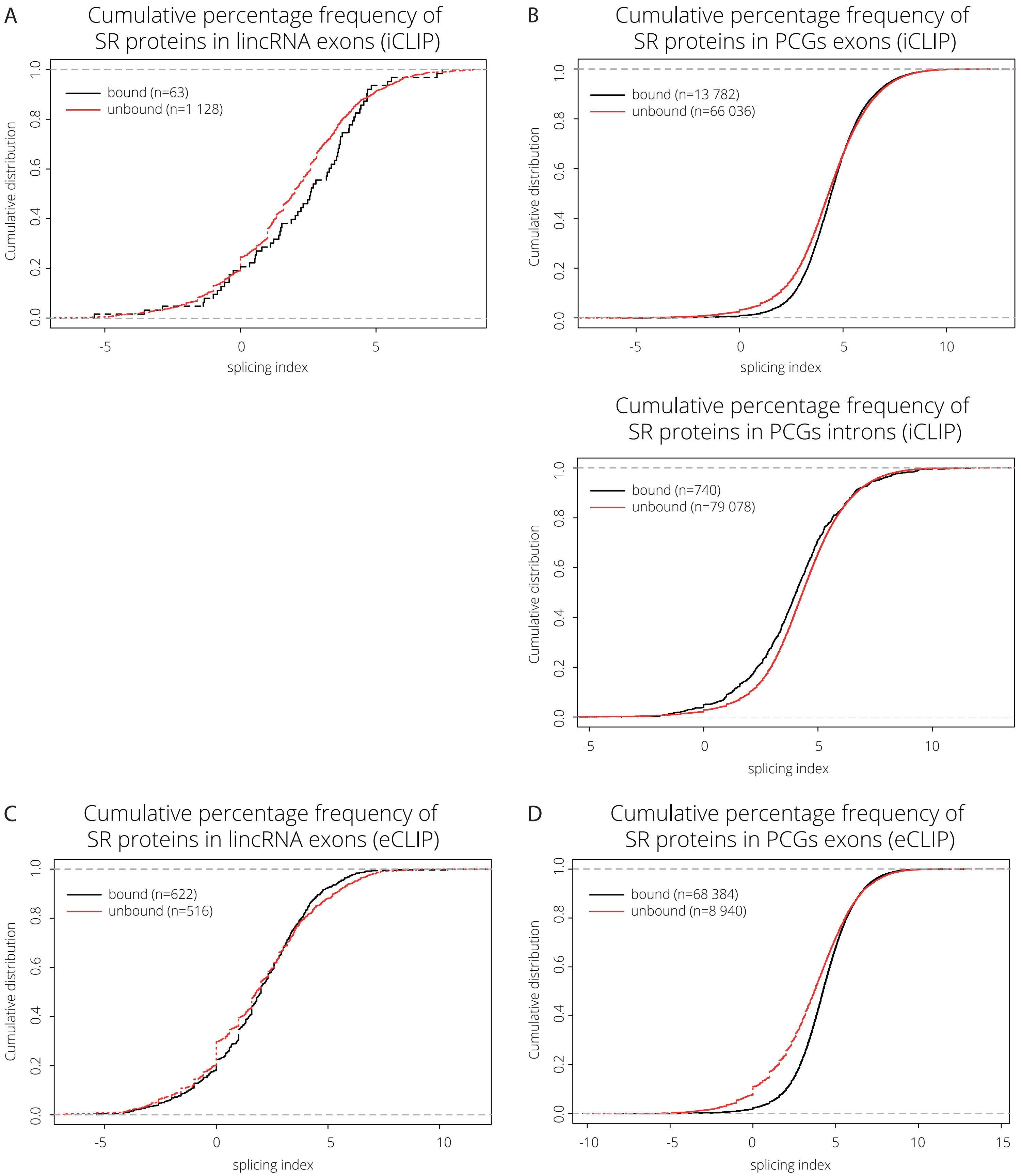


Figure S16. Cumulative distributions of splicing efficiencies comparing SRSF2, SRSF5 and SRSF6-bound and unbound lincRNAs (**A**) and PCGs (**B**) in exons (100 nt downstream of 3’ss) and introns (100 nt upstream of 3’ss). Cumulative distributions of splicing efficiencies comparing SRSF1, SRSF7 and SRSF9-bound and unbound lincRNAs (**C**) and PCGs (**D**) in exons (100 nt downstream of 3’ss); n = number of genes.

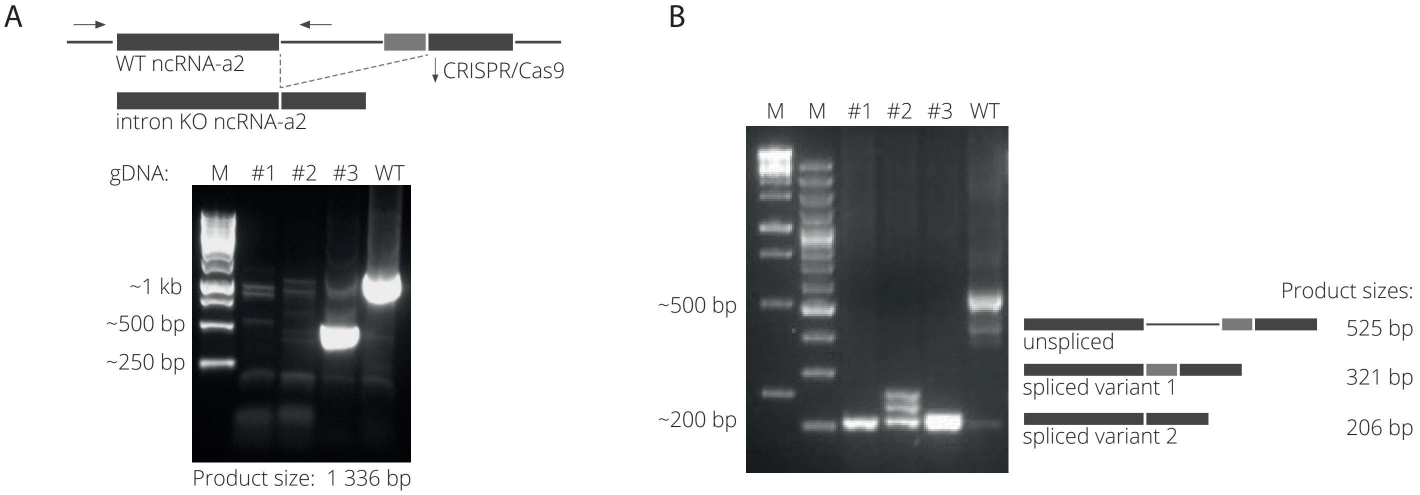


Figure S17. NcRNA-a2 intron KOs. (**A**) Intron retention assay of ncRNA-a2; horizontal arrows - primers, M - marker. Unspecific products caused by mispriming in lanes 2-4 were confirmed to be unspecific by cloning into plasmid vectors and sequencing. (**B**) RT-PCR analysis of ncRNA-a2 in three selected intron KO cell lines.

Supplementary List of Primers

Primers for Quantitative PCR – Splicing Efficiency in Fractions

Target	sequence
ncRNA-a2 spliced variant 1	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- ACCTGGAGCCCGGAAGGAAC - 3’
ncRNA-a2 spliced variant 2	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- GCACTGGAGACGGAAGGAAC - 3’
ncRNA-a2 unspliced	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- GGAGGGGAGGGAGGAAGT - 3’
ncRNA-a5 spliced	5’- CAGGTCAGGTCCTCTGGGTA - 3’
	5’- CATCCCTTTCCTGGGGTAGT - 3’
ncRNA-a5 unspliced	5’- GCTTGAGTCCTCCCAAGGTT - 3’
	5’- CATCCCTTTCCTGGGGTAGT - 3’
GAPDH spliced	5’- ACATCGCTCAGACACCATGG - 3’
	5’- GTTAAAAGCAGCCCTGGTGA - 3’
GAPDH unspliced	5’- CAGGGAAGCTCAAGGGAGAT - 3’
	5’- GTTAAAAGCAGCCCTGGTGA - 3’
LDHA spliced	5’- TGGCAGCCTTTTCCTTAGAA - 3’
	5’- CTTTCTCCCTCTTGCTGACG - 3’
LDHA unspliced	5’- TGGCAGCCTTTTCCTTAGAA - 3’
	5’- TGTGCAACTGCACTCTACCC - 3’

Primers for Semiquantitative PCR

Target	sequence
ncRNA-a2 semi	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- TGATGGCATTGAATTGGAGA - 3’

Primers for Quantitative PCR – Intron Swap

Target	sequence
MID3-4 RT	5’- CTACAGTGCTGAGTGCGTCT - 3’
ncRNA-a2 spliced variant 1	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- ACCTGGAGCCCGGAAGGAAC - 3’
ncRNA-a2 spliced variant 2	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- GCACTGGAGACGGAAGGAAC - 3’
ncRNA-a2 unspliced	5’- CTGCCGGGTTGCAAACTTG - 3’
	5’- TGGTAGAAGCACGAGCAAGG- 3’
ncRNA-a2 MID3/4	5’- GTGCCCTGGAATGTTTGCTG - 3’

	5’- AGTGCTGAGTGCGTCTCTGA - 3’
HBB spliced	5’- TTGGACCCAGAGGTTCTTTG - 3’
	5’- TGCCCAGGAGCCTGAAGTTC - 3’
HBB unspliced	5’- TAGCAGCTACAATCCAGCTACC - 3’
	5’- CACACAGACCAGCACGTTG - 3’
HBB total	5’- TTGGACCCAGAGGTTCTTTG - 3’
	5’- CCTGAAGTTCTCAGGATCCA - 3’

Primers for Quantitative PCR – Splicing Efficiency of All 2NcRNA-a2 Mutants

Target	sequence
MID3-4 RT	5’- CTACAGTGCTGAGTGCGTCT - 3’
ncRNA-a2 spliced variant 1	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- ACCTGGAGCCCGGAAGGAAC - 3’
ncRNA-a2 spliced variant 2	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- GCACTGGAGACGGAAGGAAC - 3’
ncRNA-a2 MID3/4	5’- GTGCCCTGGAATGTTTGCTG - 3’
	5’- AGTGCTGAGTGCGTCTCTGA - 3’

Primers for Quantitative PCR – Splicing Efficiency after hnRNP H siRNA KD

Target	sequence
ncRNA-a2 spliced variant 1	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- ACCTGGAGCCCGGAAGGAAC - 3’
ncRNA-a2 spliced variant 2	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- GCACTGGAGACGGAAGGAAC - 3’
ncRNA-a2 unspliced 2	5’- CTGCCGGGTTGCAAACTTG- 3’
	5’- TGGTAGAAGCACGAGCAAGG - 3’
GAPDH unspliced	5’- CAGGGAAGCTCAAGGGAGAT - 3’
	5’- GTTAAAAGCAGCCCTGGTGA - 3’
UBB	5’- GCTTTGTTGGGTGAGCTTGT - 3’
	5’- TCACGAAGATCTGCATTTTGA - 3’

Primers for Quantitative PCR – Splicing Efficiency of other LincRNAs

Target	sequence
MID5-6 RT	5’- CTCGCGATATCGTGTCTGAT- 3’
SNHG8 spliced	5’- TTAGGTGAAAGTCGCCGGGC - 3’
	5’- TCAAACTGACGGTTCTCGGG - 3’
SNHG8 unspliced	5’- GGCTTTGGAAACCCTTAAGT - 3’
	5’- TCAAACTGACGGTTCTCGGG - 3’
BX088651.4spliced	5’- GAAAGCTCAGACTCAGGGCC- 3’
	5’- AAGCACTAGACTGCGCACGC- 3’

BX088651.4unspliced	5’- GAAAGCTCAGACTCAGGGCC- 3’
	5’- AAAACCAGAGGCGGTGGAAG- 3’
BX005266.2spliced	5’- GAAAGCTCAGACTCAGGGCC- 3’
	5’- GGATCGAGATCTGCGCACGC- 3’
BX005266.2unspliced	5’- GAAAGCTCAGACTCAGGGCC- 3’
	5’- AAAACCAGAGGCGGTGGAAG- 3’
AC005840.2 spliced	5’- TGGCAAGTTTACCACCCTGA- 3’
	5’- ACAGTGCTTCCGTCCTCATG- 3’
AC005840.2 unspliced	5’- TGGCAAGTTTACCACCCTGA- 3’
	5’- GAAAGCAGCCATCCCCTTAC- 3’
AC116021.1spliced	5’- CGAGCTCTTGGAAACCCGGG- 3’
	5’- GGGTTGACATGAGGATGGCA- 3’
AC116021.1unspliced	5’- CTTTGCAGGCAACAACCCTC- 3’
	5’- GGGTTGACATGAGGATGGCA- 3’

Primers for Quantitative PCR – Splicing Efficiency of NcRNA-a2 with HBB PPT

Target	sequence
MID3-4 RT	5’- CTACAGTGCTGAGTGCGTCT - 3’
ncRNA-a2 spliced variant 1	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- ACCTGGAGCCCGGAAGGAAC - 3’
ncRNA-a2 spliced variant 2	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- GCACTGGAGACGGAAGGAAC - 3’
ncRNA-a2 MID3/4	5’- GTGCCCTGGAATGTTTGCTG - 3’
	5’- AGTGCTGAGTGCGTCTCTGA - 3’

Primers for Quantitative PCR – U2AF2 RIP

Target	sequence
MID3-4 RT	5’- CTACAGTGCTGAGTGCGTCT - 3’
ncRNA-a2 unspliced 2	5’- CTGCCGGGTTGCAAACTTG- 3’
	5’- TGGTAGAAGCACGAGCAAGG - 3’
ncRNA-a2 MID3/4	5’- GTGCCCTGGAATGTTTGCTG - 3’
	5’- AGTGCTGAGTGCGTCTCTGA - 3’

Primers for Quantitative PCR – siRNA KDs of LncRNA-as

Target	sequence
ncRNA-a2	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- TGGTAGAAGCACGAGCAAGG - 3’
KDM5B	5’- TCAGTGCAGAGAGCCAGAGA - 3’
	5’- GGATAGATCGGCCTCGTGTA - 3’

RABIF	5’- AGGGACCGCTCTCTTCTCTC - 3’
	5’- AGTGTTCCTGGAGGAGATCG - 3’
ADIPOR1	5’- ACATCTGGACCCATCTGCTT - 3’
	5’- CCCAAAAACCACCTTCTCCT - 3’
KLHL12	5’- TCAAGTGCGACGAAATTCAG - 3’
	5’- GCTCTTTCTTGGCATGCTTC - 3’
GAPDH	5’- ACATCGCTCAGACACCATGG - 3’
	5’- GTTAAAAGCAGCCCTGGTGA - 3’
LDHA	5’- TGGCAGCCTTTTCCTTAGAA - 3’
	5’- CTTTCTCCCTCTTGCTGACG - 3’
ACTB	5’- GGCATCCTCACCCTGAAGTA - 3’
	5’- AGGTGTGGTGCCAGATTTTC - 3’
UBB	5’- GCTTTGTTGGGTGAGCTTGT - 3’
	5’- TCACGAAGATCTGCATTTTGA - 3’

ncRNA-a5	5’- CAGGTCAGGTCCTCTGGGTA - 3’
	5’- CATCCCTTTCCTGGGGTAGT - 3’
PQLC3	5’- CCTCAGCCTTCCGAGTTTAC - 3’
	5’- GAGGATGGGGTACTCCAGGT - 3’
E2F6	5’- AGGAATGGGCTCCAGAGAGA - 3’
	5’- TCGGACTCCCAGTTTCGTTG - 3’
ROCK2	5’- TGAAGCCTGACAACATGCTC - 3’
	5’- AATCCGGTGTTCCAACTGCT - 3’
GAPDH	5’- ACATCGCTCAGACACCATGG - 3’
	5’- GTTAAAAGCAGCCCTGGTGA - 3’

Primers for Quantitative PCR – Ectopic Overexpression

Target	sequence
ncRNA-a2	5’- CCTTACTCTTGGACAACACTCC - 3’
	5’- TTGGATGGACCGAATGAGGATG - 3’
KLHL12	5’- ACAAGCCTGCTGTGAGTTCT - 3’
	5’- TCCACCTCTCCTTGACTCAGA - 3’
GAPDH	5’- ATTTGGTCGTATTGGGCGCC - 3’
	5’- TGAGGTCAATGAAGGGGTCA - 3’

Primers for Genotyping CRISPR Mutants

Target	sequence
ncRNA-a2 intron KO	5’- GGGGGTGGGTATAGAGGAATTTGATGC - 3’
	5’- GCGGACCCTACACGTGCGCTACGT - 3’
ncRNA-a2 semi	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- TGATGGCATTGAATTGGAGA - 3’

Primers for Quantitative PCR – NcRNA-a2 Intron KOs

Target	sequence
ncRNA-a2	5’- CCTTACTCTTGGACAACACTCC - 3’
	5’- TTGGATGGACCGAATGAGGATG - 3’
KDM5B	5’- TCAGTGCAGAGAGCCAGAGA - 3’
	5’- GGATAGATCGGCCTCGTGTA - 3’
RABIF	5’- AGGGACCGCTCTCTTCTCTC - 3’
	5’- AGTGTTCCTGGAGGAGATCG - 3’
ADIPOR1	5’- ACATCTGGACCCATCTGCTT - 3’
	5’- CCCAAAAACCACCTTCTCCT - 3’
KLHL12	5’- ACAAGCCTGCTGTGAGTTCT - 3’
	5’- TCCACCTCTCCTTGACTCAGA - 3’
GAPDH	5’- ATTTGGTCGTATTGGGCGCC - 3’
	5’- TGAGGTCAATGAAGGGGTCA - 3’

Supplementary List of Used ISE Motifs

Name of the mutant	sequence *(mutated nucleotides are underlined)*
1xISE	5’- TTTGGGC - 3’
2xISE	5’- TTTGGGCTATTGG - 3’
3xISE	5’- TTTGGGCTTTGGGCTATTGG - 3’
ISEctrl	5’- TTCGCGC - 3’

Supplementary List of Modified PPTs

Name of the mutant	sequence *(mutated nucleotides are underlined)*
ncRNAa-2 WT	5’- TCGCCGCCTCTGACAACTTTT - 3’
ncRNAa-2 T21	5’- TTTTTTTTTTTTTTTTTTTTT- 3’
ncRNAa-2 CtoT	5’- TTGTTGTTTTTGATAATTTTT- 3’
ncRNAa-2 GAtoT	5’- TCTCCTCCTCTTTCTTCTTTT- 3’
SNHG8 WT	5’- GCATGCGCGGACTTGAGTGCTCAT - 3’
SNHG8 T25	5’- TTTTTTTTTTTTTTTTTTTTTTTTTT- 3’
BX088651.4 WT	5’- CATCGCGTCCTCTTC - 3’
BX088651.4 T15	5’- TTTTTTTTTTTTTTT- 3’
BX005266.2 WT	5’- CATCGCGTCCTCTTCCAGTCTAGTGCTTTTTT - 3’
BX005266.2 T32	5’- TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT- 3’
AC005840.2 WT	5’- GAGACTCACTCTAGTCTTTTCCCG - 3’
AC005840.2 T24	5’- TTTTTTTTTTTTTTTTTTTTTTTT- 3’
AC116021.1 WT	5’- TCCCTATTTTG - 3’
AC116021.1 T11	5’- TTTTTTTTTTT- 3’

Supplementary List of MaxEnt Scores

Name of the mutant	5‘ss		3‘ss	
	sequence	MaxEnt score	sequence	MaxEnt score
ncRNA-a2	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC	6.13
			gatcctctcgtctcccccagTCT	9.31
ncRNA-a5 intron 2	GAGgtaagc	9.85	ttgagatccttctttttcagGTG	9.33
GAPDH intron 2	CGGgtgagt	9.89	accctcacgtattcccccagGTT	8.32
LDHA intron 3	AAGgttgat	4.19	attattccccttttctctagACT	8.08
ncRNA-a2 HBB intron	CCGgtgagt	10.90	ctcttatcttcctcccacagGGC	12.56
ncRNA-a2 HBB intron ncRNA-a2 5‘ss	CCGgtaacc	5.28	ctcttatcttcctcccacagGGC	12.56
HBB WT intron 2	AGGgtgagt	9.25	ctcttatcttcctcccacagCTC	11.43
HBB ncRNA-a2 intron	AGGgtaacc	5.40	cgcctctgacaacttttcagCTC	4.97
HBB ncRNA-a2 intron HBB 5‘ss	AGGgtgagt	9.25	cgcctctgacaacttttcagCTC	4.97
ncRNA-a2 FΔ1=Δ1	CCGgtaacg	7.34	cgcctctgacaacttttcagGGC	6.13
			gatcctctcgtctcccccagTCT	9.31
ncRNA-a2 FΔ2	CCGgtaacg	7.34	cgcctctgacaacttttcagGGC	6.13
			gatcctctcgtctcccccagTCT	9.31
ncRNA-a2 FΔ3	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC	6.13
			gatcctctcgtctcccccagTCT	9.31
ncRNA-a2 FΔ4	CCGgtaaca	5.92	cgcctctgacaacttttcagGGC	6.13
			gatcctctcgtctcccccagTCT	9.31
ncRNA-a2 FΔ5	CCGgtaact	6.39	cgcctctgacaacttttcagGGC	6.13
			gatcctctcgtctcccccagTCT	9.31
ncRNA-a2 FΔ6	CCGgtaaca	5.92	cgcctctgacaacttttcagGGC	6.13
			gatcctctcgtctcccccagTCT	9.31
ncRNA-a2 FΔ7	CCGgtaaca	5.92	cgcctctgacaacttttcagGGC	6.13
			gatcctctcgtctcccccagTCT	9.31
ncRNA-a2 FΔ8	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC	6.13
			gatcctctcgtctcccccagTCT	9.31
ncRNA-a2 RΔ1=Δ8	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC	6.13
			gatcctctcgtctcccccagTCT	9.31

ncRNA-a2 RΔ2	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 RΔ3	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 RΔ4	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 RΔ5	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 RΔ6	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 RΔ7	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 Δ2	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 Δ3	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 Δ4	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 Δ5	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 Δ6	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 Δ7	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 7.53 5‘ss	ACGgtaact	7.53	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 9.6 5’ss	CAGgtgagc	9.6	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 11.08 5‘ss	CAGgtaagg	11.08	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ1 7.66 5‘ss	CAGgtgacg	7.66	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ1 9.35 5’ss	CCGgtgagg	9.35	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ1 11.08 5‘ss	CAGgtaagg	11.08	cgcctctgacaacttttcagGGC	6.13

			gatcctctcgtctcccccagTCT	9.31
ncRNA-a2 T21 7.53 5‘ss	ACGgtaact	7.53	tttttttttttttttttcagGGC gatcctctcgtctcccccagTCT	13.12 9.31
ncRNA-a2 T21 9.6 5’ss	CAGgtgagc	9.6	tttttttttttttttttcagGGC gatcctctcgtctcccccagTCT	13.12 9.31
ncRNA-a2 T21 11.08 5‘ss	CAGgtaagg	11.08	tttttttttttttttttcagGGC gatcctctcgtctcccccagTCT	13.12 9.31
ncRNA-a2 T21	CCGgtaacc	5.28	tttttttttttttttttcagGGC	13.12
ncRNA-a2 CtoT	CCGgtaacc	5.28	tgtttttgataatttttcagGGC	8.62
ncRNA-a2 GAtoT	CCGgtaacc	5.28	ctcctctttcttcttttcagGGC	13.39
ncRNA-a2 ΔPPT	CCGgtaacc	5.28	tcgccgcctctgacaaccagGGC	4.76
ncRNA-a2 3xISE-T21	CCGgtaacc	5.28	tttttttttttttttttcagGGC	13.12
ncRNA-a2 with HBB PPT	CCGgtaacc	5.28	ctcttatcttcctcccacagGGC	12.56
SNHG8 WT	AAAgtacgt	7.99	cggacttgagtgctcattagGTC	-0.49
SNHG8 T25	AAAgtacgt	7.99	ttttttttttttttttttagGTC	13.77
BX088651.4 WT	CAGgtaaag	9.65	tccatcgcgtcctcttccagTCT	8.29
BX088651.4 T15	CAGgtaaag	9.65	tctttttttttttttttcagTCT	10.63
BX005266.2 WT	CAGgtaaag	9.65	agtctagtgctttttttccagATC	5.60
BX005266.2 T32	CAGgtaaag	9.65	ttttttttttttttttccagATC	12.67
AC005840.2 WT	ACGgtaagg	10.38	actctagtcttttcccgcagGAA	9.46
AC005840.2 T24	ACGgtaagg	10.38	tttttttttttttttttcagGAA	13.15
AC116021.1 WT	TTGgtaaaa	3.23	ttatcatccctattttgcagGAA	10.48
AC116021.1 T11	TTGgtaaaa	3.23	ttatcatttttttttttcagGAA	12.22

Supplementary List of Intron Sequences

Name		Sequence	Length (n)
ncRNA-a2 intron	5′-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggatgagggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3′	204
ncRNA-a2 HBB intron 2 ncRNA-a2 5‘ss	5′-	gtaaccgcatgggacgcttgatgttttctttccccttcttttctatggttaagttcatgtcataggaaggggataagtaacagggtacagtttagaatgggaaacagacgaatgattgcatcagtgtggaagtctcaggatcgttttagtttcttttatttgctgttcataacaattgttttcttttgtttaattcttgctttcttttttttttcttctccgcaatttttactattatacttaatgccttaacattgtgtataacaaaaggaaatatctctgagatacattaagtaacttaaaaaaaaactttacacagtctgcctagtacattactatttggaatatatgtgtgcttatttgcatattcataatctccctactttattttcttttatttttaattgatacataatcattatacatatttatgggttaaagtgtaatgttttaatatgtgtacacatattgaccaaatcagggtaattttgcatttgtaattttaaaaaatgctttcttcttttaatatacttttttgtttatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaatgtatcatgcctctttgcaccattctaaagaataacagtgataatttctgggttaaggcaatagcaatatctctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttcatattgctaatagcagctacaatccagctaccattctgcttttattttatggttgggataaggctggattattctgagtccaagctaggcccttttgctaatcatgttcatacctcttatcttcctcccacag - 3′	850
ncRNA-a2 intron with HBB PPT	5′-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggatgagggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctgttcatacctcttatcttcctcccacag - 3′	209
ncRNA-a2 HBB 5‘ss	5′-	gtgagtgcgttgcgaagaccacgctgccgggttgcaaacttgggggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggatgagggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3′	204
ncRNA-a2 T21 HBB 5‘ss	5′-	gtgagtgcgttgcgaagaccacgctgccgggttgcaaacttgggggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggatgagggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctttttttttttttttttttttcag - 3′	204
HBB intron 2	5′-	gtgagtctatgggacgcttgatgttttctttccccttcttttctatggttaagttcatgtcataggaaggggataagtaacagggtacagtttagaatgggaaacagacgaatgattgcatcagtgtggaagtctcaggatcgttttagtttcttttatttgctgttcataacaattgttttcttttgtttaattcttgc	850

		tttcttttttttttcttctccgcaatttttactattatacttaatgccttaacattgtgtataacaaaag gaaatatctctgagatacattaagtaacttaaaaaaaaactttacacagtctgcctagtacatt actatttggaatatatgtgtgcttatttgcatattcataatctccctactttattttcttttatttttaat tgatacataatcattatacatatttatgggttaaagtgtaatgttttaatatgtgtacacatattg accaaatcagggtaattttgcatttgtaattttaaaaaatgctttcttcttttaatatactttttttgtt tatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaatgtatcatgcct ctttgcaccattctaaagaataacagtgataatttctgggttaaggcaatagcaatatctctgc atataaatatttctgcatataaattgtaactgatgtaagaggtttcatattgctaatagcagcta caatccagctaccattctgcttttattttatggttgggataaggctggattattctgagtccaagc taggcccttttgctaatcatgttcatacctcttatcttcctcccacag - 3′	
HBB ncRNA-a2 intron HBB 5‘ss	5′-	gtgagtctgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctcccctc cccctgggcgccgtgcaactgccctgggaccgggttctgggatgagggggcagaccgggct ccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccctcgccgcc tctgacaacttttcag - 3′	204
HBB intron 2 with ncRNA-a2 PPT	5′-	gtgagtctatgggacgcttgatgttttctttccccttcttttctatggttaagttcatgtcataggaa ggggataagtaacagggtacagtttagaatgggaaacagacgaatgattgcatcagtgtgg aagtctcaggatcgttttagtttcttttatttgctgttcataacaattgttttcttttgtttaattcttgc tttcttttttttttcttctccgcaatttttactattatacttaatgccttaacattgtgtataacaaaag gaaatatctctgagatacattaagtaacttaaaaaaaaactttacacagtctgcctagtacatt actatttggaatatatgtgtgcttatttgcatattcataatctccctactttattttcttttatttttaat tgatacataatcattatacatatttatgggttaaagtgtaatgttttaatatgtgtacacatattg accaaatcagggtaattttgcatttgtaattttaaaaaatgctttcttcttttaatatactttttttgtt tatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaatgtatcatgcct ctttgcaccattctaaagaataacagtgataatttctgggttaaggcaatagcaatatctctgc atataaatatttctgcatataaattgtaactgatgtaagaggtttcatattgctaatagcagcta caatccagctaccattctgcttttattttatggttgggataaggctggattattctgagtccaagc taggcccttttgctaatcatcgccgcctctgacaacttttcag - 3′	845
WT	5′-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctcccct cccccctgggcgccgtgcaactgccctgggaccgggttctgggatgagggggcagaccggg ctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccctcgccg cctctgacaacttttcag - 3′	204
FΔ1= Δ1	5′-	gtaacgccgggttgcaaacttgggggggacttcctccctcccctcccccctgggcgccgtgcaact gccctgggaccgggttctgggatgagggggcagaccgggctccccagcggccggcgcagc acgtagcgcacgtgtagggtccgctccccacccctcgccgcctctgacaacttttcag - 3′	184
FΔ2	5′-	gtaacggacttcctccctcccctcccccctgggcgccgtgcaactgccctgggaccgggttctgg gatgagggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtaggg	164

		tccgctccccaccccctcgccgcctctgacaacttttcag - 3’	
FΔ3	5’-	gtaacccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	144
FΔ4	5’-	gtaacccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	124
FΔ5	5’-	gtaacatgaggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	104
FΔ6	5’-	gtaactccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	84
FΔ7	5’-	gtaacacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	64
FΔ8	5’-	gtaacccgctccccaccccctcgccgcctctgacaacttttcag - 3’	44
RΔ1= Δ8	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcagcccgctccccaccccctcgccgcctctgacaacttttcag - 3’	184
RΔ2	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggcccgctccccaccccctcgccgcctctgacaacttttcag - 3’	164
RΔ3	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggccgctccccaccccctcgccgcctctgacaacttttcag - 3’	144
RΔ4	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctccccctgggcgccgtgcaactgccgctccccaccccctcgccgcctctgacaacttttcag - 3’	124
RΔ5	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctccccgctccccaccccctcgccgcctctgacaacttttcag - 3’	104
RΔ6	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggccgctccccaccccctcgccgcctctgacaacttttcag - 3’	84

RΔ7	5’-	gtaaccgcgttgcgaagaccacgctccgctccccaccccctcgccgcctctgacaacttttcag - 3’	64
Δ2	5’-	gtaaccgcgttgcgaagaccacgctggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	184
Δ3	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	184
Δ4	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctccccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	184
Δ5	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctccccctgggcgccgtgcaactgatgaggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	184
Δ6	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggtccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	184
Δ7	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	184
1xISE	5’-	gtaaccgcgttgcgaagaccacgcttttgggcgccgggttgcaaacttggggggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	211
2xISE	5’-	gtaaccgcgttgcgaagaccacgcttttgggctattgggccgggttgcaaacttggggggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	217
3xISE	5’-	gtaaccgcgttgcgaagaccacgcttttgggctttgggctattgggccgggttgcaaacttggggggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3’	224

ISEctrl	5′-	gtaaccgcgttgcgaagaccacgctttcgcgcgccgggttgcaaacttggggggacttcctccctcccctcccccctgggcgccgtgcaactgccctgggaccgggttctgggatgagggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccctcgccgcctctgacaacttttcag - 3′	211
Δ60	5′-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctcctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3′	144
3xISE-Δ60	5′-	gtaaccgcgttgcgaagaccacgcttttgggctttgggctattgggccgggttgcaaacttgggggggacttcctccctcccctcctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3′	164
T21	5′-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctcccccctgggcgccgtgcaactgccctgggaccgggttctgggatgagggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccttttttttttttttttttttttcag - 3′	204
CtoT	5′-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctcccccctgggcgccgtgcaactgccctgggaccgggttctgggatgagggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccttgttgtttttgataatttttcag - 3′	204
GAtoT	5′-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctcccccctgggcgccgtgcaactgccctgggaccgggttctgggatgagggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctctcctcctctttcttcttttcag - 3′	204
ΔPPT	5′-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctcccccctgggcgccgtgcaactgccctgggaccgggttctgggatgagggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaaccag - 3′	200
Δ60-T21	5′-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggggacttcctccctcccctcctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccaccccccttttttttttttttttttttttcag - 3′	144

Supplementary List of SRSF Binding Motifs

Protein	Motif	Reference
SRSF1	GARGARGARG	(1)
	TGRWG	(2)
	RGAAGAAC	(2)
	AGGACRRAGC	(2)
	SRSASGA	(2)
SRSF2	RTCTGWAGA	(1)
	GRYYMCYR	(3)
	TGCYGYY	(3)
	AGSAGAGTA	(2)
	TGTTCSAGWT	(2)
	AGGAGAT	(2)
	GRYYCSYR	(2)
SRSF3	TCWTCHTC	(1)
	WCWWC	(2)
	CTCKTCY	(2)
SRSF4	AWGAWGAWGA	(1)
	GAAGGA	(2)
SRSF5	YDTCTGWAGACA	(1)
	YYWCWSG	(3)
	ACDGS	(2)
SRSF6	RARGAWGA	(1)
	GAAGAAGA	(1)
	TSCGKM	(2)
	YRCRKM	(3)
SRSF7	WRAWGAHRA	(1)
	TCAACA	(2)
	ACGAGAGAY	(2)
SRSF9	GACGAC	(2)
	CTGGATT	(2)

Supplementary List of sgRNAs and Guide Target Sequences

Guide	sequence
ncRNA-a2 intron KO #1	5’- TTCGCAACGCGGTTACCGGA - 3’
#2	5’- AGTCTCCAGTGCACCGGCTT - 3’

Guide target	sequence
ncRNA-a2 intron KO #1	5’- CCTTCCGGTAACCGCGTTGCGAA - 3’
#2	5’- CCCAGTCTCCAGTGCACCGGCTT - 3’

References

1. Müller-McNicoll, M., Botti, V., de Jesus Domingues, A.M., Brandl, H., Schwich, O.D., Steiner, M.C., Curk, T., Poser, I., Zarnack, K. and Neugebauer, K.M. (2016) SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev.*, **30**, 553-566.
2. Mueller, W.F. and Hertel, K.J. (2011) The role of SR and SR-related proteins in pre-mRNA splicing. In Lorkovic, Z. (ed.), *RNA Binding Proteins*. Landes Bioscience and Springer Science+Business Media, New York, Vol. I, pp. 1-21.
3. Paz, I., Akerman, M., Dror, I., Kosti, I. and Mandel-Gutfreund, Y. (2010) SFmap: a web server for motif analysis and prediction of splicing factor binding sites. *Nucleic Acids Res.*, **38**, W281-285.
4. Sironi, M., Menozzi, G., Riva, L., Cagliani, R., Comi, G.P., Bresolin, N., Giorda, R. and Pozzoli, U. (2004) Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res.*, **32**, 1783-1791.
5. Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M. and Burge, C.B. (2004) Systematic Identification and Analysis of Exonic Splicing Silencers. *Cell*, **119**, 831-845.
6. Yeo, G. and Burge, C.B. (2004) Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *J. Comput. Biol.*, **11**, 377-394.