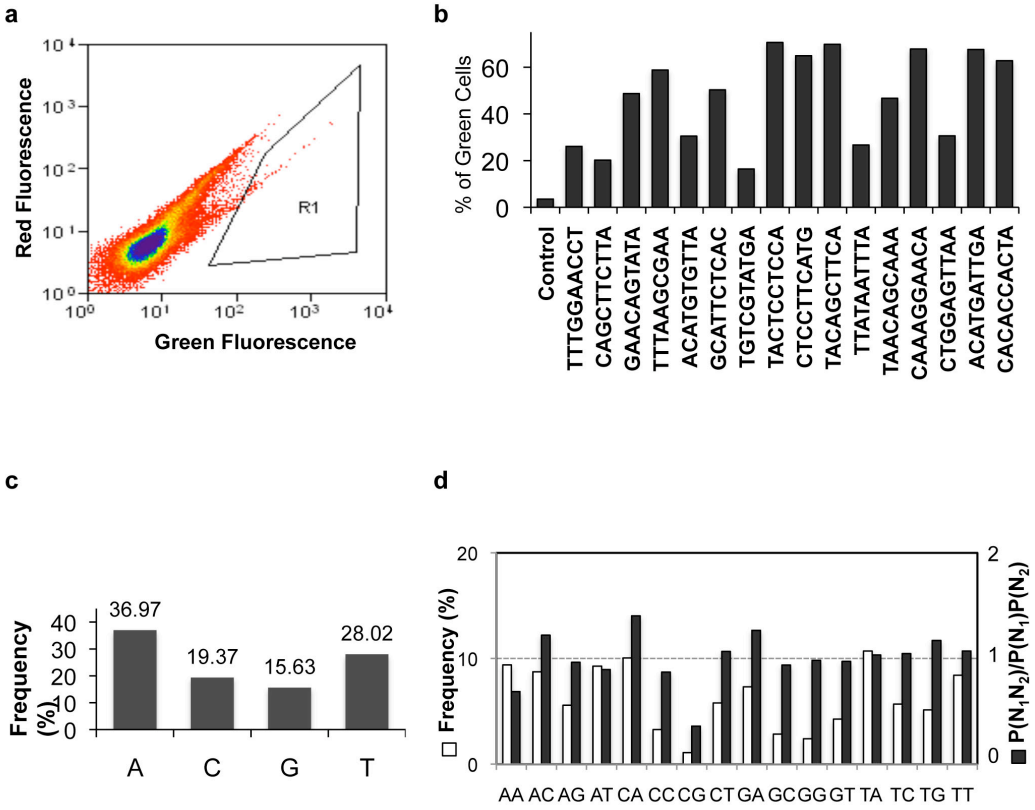
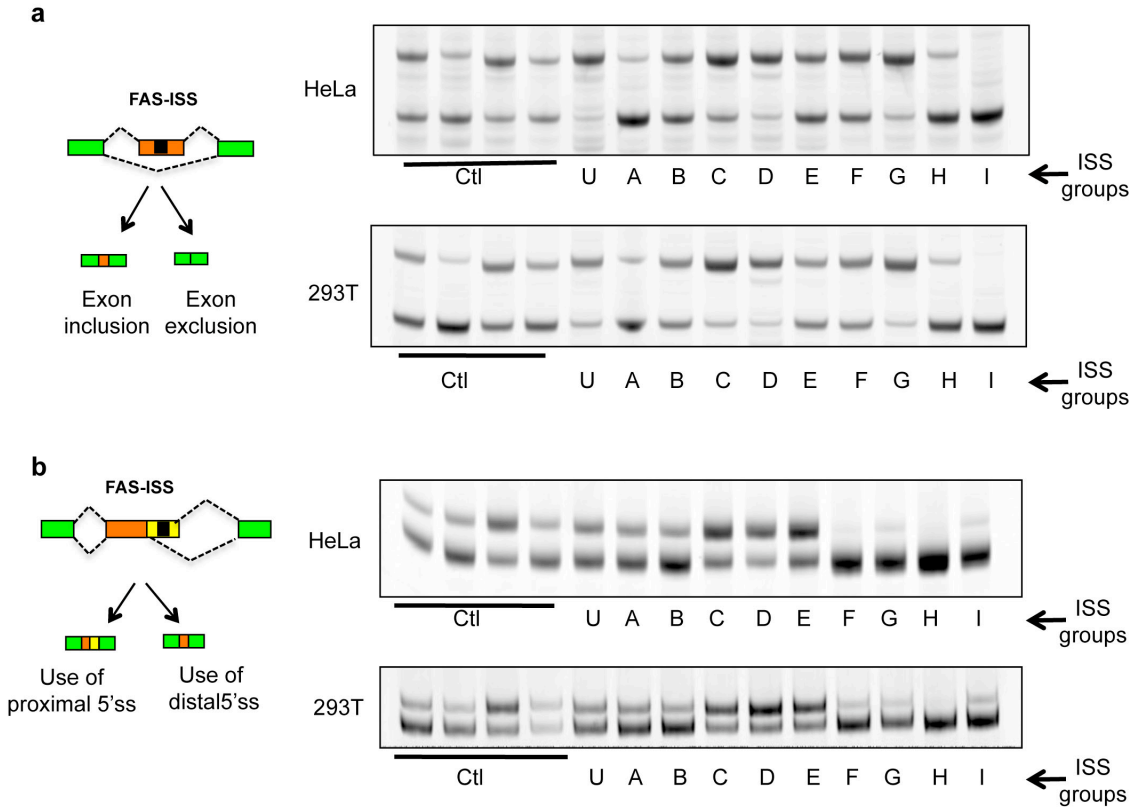
Wang et al., Supplementary Fig. 1



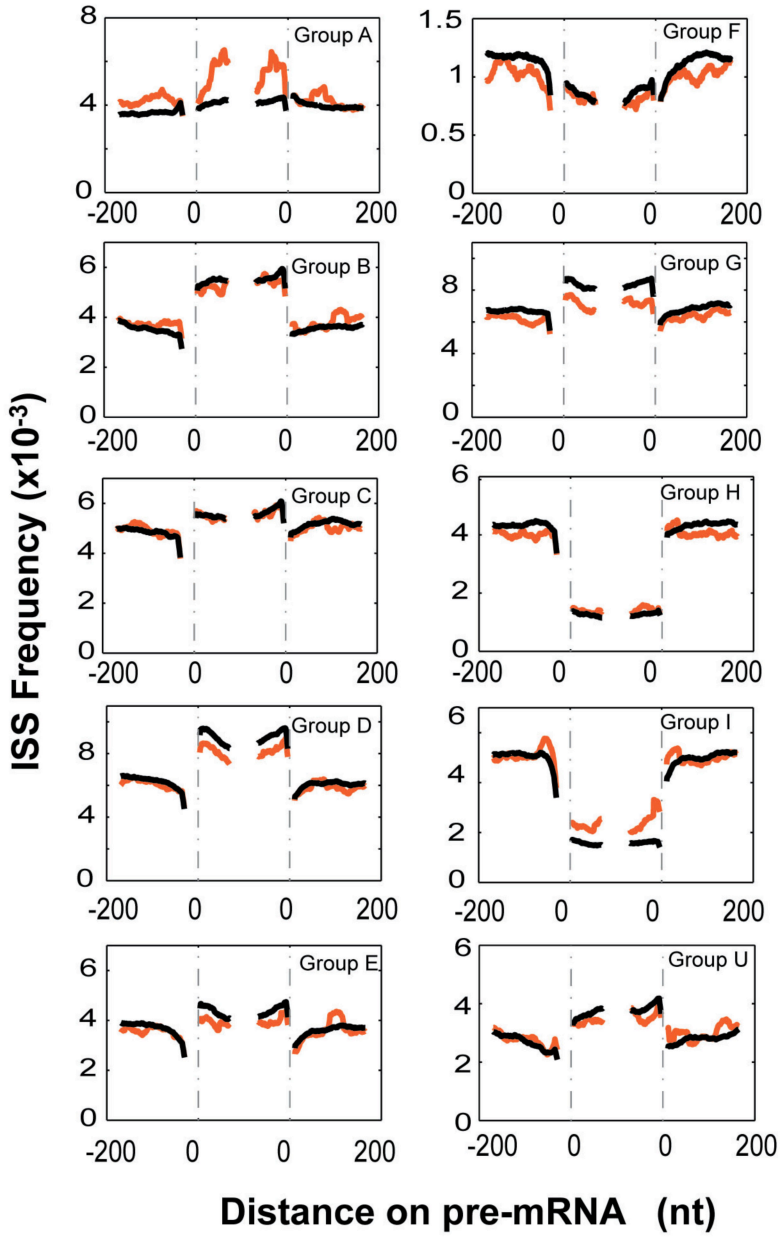
Supplementary Fig. 1. Identification of FAS-ISS decamers. (a) Flow cytometry profile of 293 FlipIn cells transfected with the pZW11 reporter inserted with random library. After selection for stable integration, all hygromycin resistant cells were pooled and analyzed by flowcytometry. Both red and green fluorescence signals were measured to correct for self-fluorescence background. The GFP-positive cells (R1 region) were sorted using a Cytomation MoFlo high-speed sorter into 96 well plates to recover all of the ISS sequences. (b) Validation of FAS-ISS decamer activity. To validate the silencer activities of the newly identified ISS decamers, 293T cells were transiently transfected with pZW11 containing 16 arbitrarily selected ISS decamers and control sequence. The GFP-positive cells were examined by flow cytometry at 24h after transfection. (c) The frequencies of mononucleotide in the screened ISS decamer set. (d) The frequencies and odds ratios of dinucleotides in the screened ISS decamers set.

Wang et al., Supplementary Fig. 2



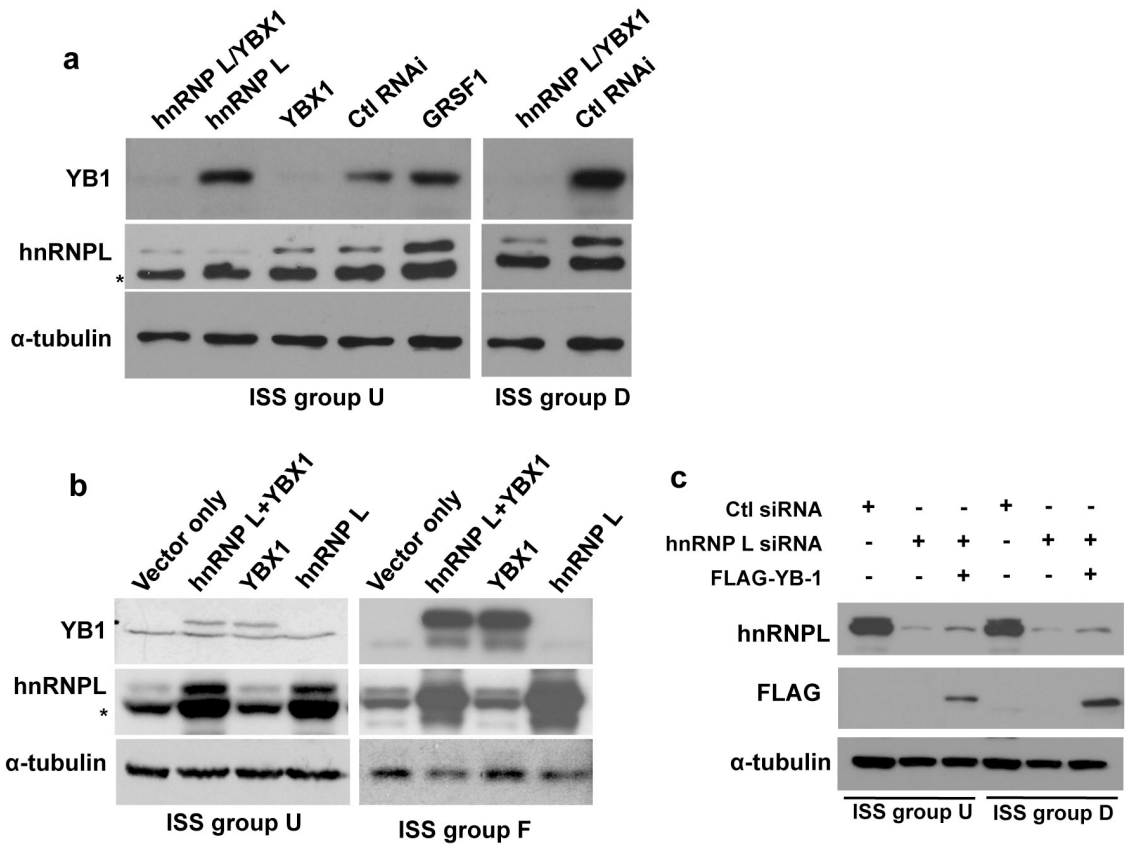
Supplementary Fig. 2. Diverse functions of ISSs when inserted in the exons or between two 5′ SS. (a) The ISS exemplars of each group were inserted into the exon of a splicing reporter and transiently transfected into both HeLa and 293T cells. After 48 hours, RNAs were isolated from the transfected cells to determine the functions of ISSs by RT-PCR. (b) The representative ISSs of each group were inserted between two 5′ SS in the exonic extension region of the mini-gene reporter. The resulting reporters were transiently transfected into HeLa and 293T cells. After 48 hours, RNAs were extracted from the transfected cells and the functions of ISSs were examined by RT-PCR and shown in the gel figures.

Wang et al., Supplementary Fig.3



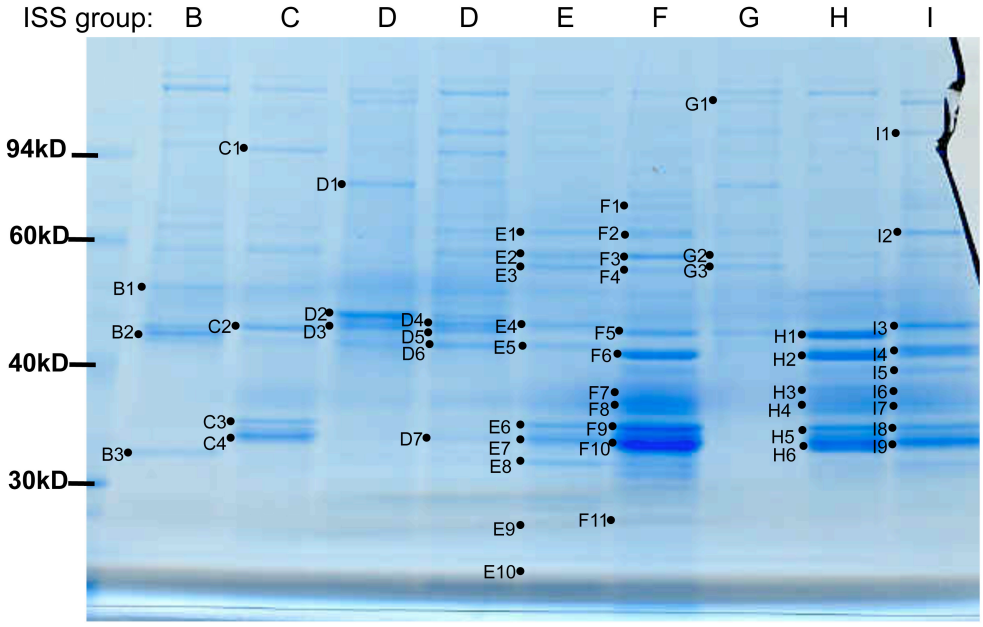
Supplementary Fig. 3. The positional frequency of FAS-ISS k-mers in different pre-mRNA locations. Distribution of ISS k-mers near the constitutive exon (black) and skipped exon (red). The number of transcripts containing ISS k-mers divided by total number of transcripts at each position is plotted as ISS frequency. The first and last 50 bases of exons and the first and last ~200 bases of introns are shown, excluding the last 20 bases of the upstream intron and the first 10 bases of the downstream intron to avoid overlaps with the splice site motifs.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.2459



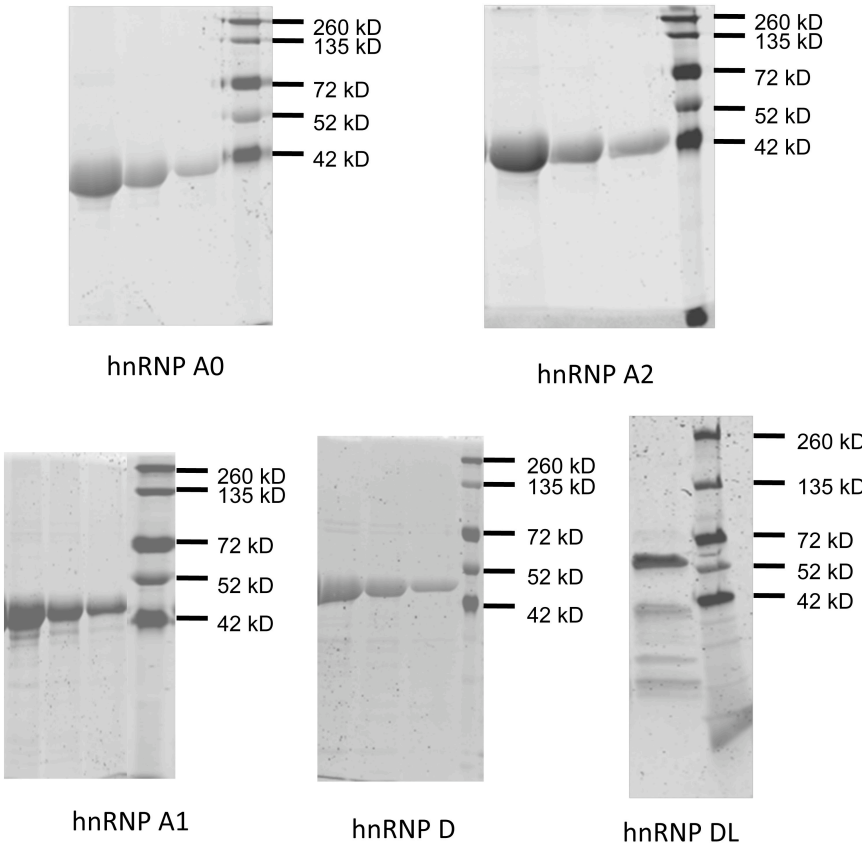
Supplementary Fig. 4. Depletion or overexpression of hnRNP L and YB1 proteins. (a) 293T cells were transfected with the siRNA of *HNRNP L* and *YB1*. After 48 hours of siRNA transfection, we transfected 0.2 μg splicing mini-gene reporters containing ISS group U or group D into the cells respectively, and harvested the cells 24 hours after the second transfection to examine the protein levels of HNRNP L and YB1 through Western blot, and the tubulin level was detected as loading control. *GRSF1* siRNA was used as the specificity control. The asterisk indicated a non-specific protein that cross-react with HNRNP L antibody. (b) 293T cells were transiently transfected with expression vectors of HNRNP L and YB1. After 72 hours, proteins were extracted from the transfected cells to determine the expression levels of HNRNP L and YB1 with western blot. The asterisk indicated a non-specific protein that cross-react with HNRNP L antibody. The tubulin level was examined as loading control. (c) 293T cells were transfected with siRNA of scramble control or *HNRNP L*. After 48 hours of siRNA transfection, the cells were co-transfected with 0.2 μg splicing reporter containing ISS group U or group D and FLAG-YB1 expression vector (lane 3 and 6). The cells were harvested after another 24 hours for the protein analysis. The tubulin level was measured as loading control.

Wang et al., Supplementary Fig. 5



Supplementary Fig. 5. **The SDS-PAGE gel of putative protein factors that bind to different ISS groups.** Biotinylated RNA oligos of each ISS group were incubated with HeLa whole cell extract, bound to streptavidin beads and washed, the RNA-protein complex were eluted and separated on a SDS-PAGE gel. The specific bands (marked with a dot and labeled according to each group) were cut and identified by mass spectrometry. Two batches of the affinity purification experiments were carried out for group D ISS, and we separated two samples in the same gel for protein identification.

Wang et al., Supplementary Fig. 6



Supplementary Fig. 6. **Purification of recombinant proteins for measurement of direct RNA-protein binding.** The putative *trans*-factors binding to groups F, H, I were cloned into bacterial expression system (pT7HtB) and purified with His GraviTrap Kit. The final protein products were assayed with SDS-PAGE to check purity, and were later used in Biacore assay to measure the direct RNA-protein binding. Three fractions eluded from Ni column were shown for hnRNP A0, A2, A1 and D, and the elution fractions were combined for hnRNP DL.

Supplementary Table 1: Selected primers used in construction of splicing reporters.

Primer ID:	Sequence	Note
1	CACGTCGACCTGCAGGATTTTAGCCCTG	Forward primer of SirT1 intron 5
2	CACAAGCTTCTCGAGCAACAAATTACCTGATTAAAAAT	Reverse primer of SirT1 exon 6 +intron 6
3	CACGAATTCATGTGGGCCCATATTTTAGGAATTGTTC	Forward primer of SirT1 intron 6
4	CACCCGCGGACAACTTGCTTATGATCCTGAC	Reverse primer of SirT1 intron 6
5	TCGACTACGTACATGCGGCC	Primer pairs used in filling in pZW2 and destroy the exonic restriction sites
6	GCATGTACGTAG	
7	TCTCGAGACTGGGGCCCTAAGATGAGGATTCTAGGGG	Primer pairs used to introduce XhoI/ApaI site at downstream of DHFR exon 2
8	TAGGGCCCCAGTCTCGAGACCCCAAATTACCTTC	
9	CACGCTAGCGCTACCGGTCGCCAC	Forward primer for GFP exon 1
10	AAGGGCTGCAGAAAGGCTGGAAC	Reverse primer for DHFR exon 2 and downstream intron

Supplementary Table 2. The sequences of the ISS decamers obtained from screen. The first 16 decamers are those that were reconfirmed by transient transfection. The ISS sequences obtained for multiple times in independent transfections are marked with a superscript number to indicate how many times it was obtained.

ISS Sequences	
TTTGGAACCT	ISS #1
CAGCTTCTTA	ISS #2
GAACTGACAT	ISS #3
TTTAAGCGAA	ISS #4
ACATGTGTTA	ISS #5
GCATTCTCAC	ISS #6
TGTCGTATGA	ISS #7
TACTCCTCCA [2]	ISS #8
CTCCTTCATG	ISS #9
TACAGCTTCA [2]	ISS #10
TTATAATTTA	ISS #11
TAACAGCAAA	ESS #12
CAAAGGAACA	ESS #13
CTGGAGTTAA [2]	ESS #14
ACATGATTGA	ESS #15
CACACCACTA	ISS #16
TACAGCTTTA [2]	
ACATGATTTA[2]	
CATTGGAACC [2]	
TTGTTGCAAA [4]	
GAACAGTATA [3]	
AGACATGATA	
ATTCCTTGAT	
AACCTTTTTA	
TCAGATAGAA	
TCCCGTTAAA	
GAATCTAAAC	
TCGTAATGTT	
ACATGCGGTA	
GAACAGTGA	
ATATGAACTA	
ACATGTCAGA	

GAACAATTTA	
GAATTTATGC	
GCTCTTCAC	
GAACATACA	
ACATGATTCA	
TTTTGGCAAA	
TACAGCTCTA	
GAACAGAGTA	
ACATGATATA	
GAACTATATA	
GCAAGGACAT	
AGAACCTAAA	
AAAACGTCTG	
GAACTCATTA	
ACATGACCCA	
GATCCTAAAA	
ACATGGAATA	
ACTTGAACTA	
AGCCCTTTAC	
CTTCACACCA	
GAACTGGTCA	
TACAGCTATA	
ACATGATTATA	
TAGTTTGTT	
ATATTGTACA	
TAAATATCTG	
GTTAATAATC	
TAACTGCAAA	
TATCAGGATC	
GAACTTAAAC	
ACGAAATTTT	
TGTAGGAGTA	
ACTCCTCCAT	
GGCTCTTCAT	
ACTGAACGTA	
ACATGACATA	
GAACAGTTCA	
GAACAGCATA	
GAACATTATA	
TAGTAGGTAT	
TATCTCCTCA	

ACATGAGCTA	
ACTGAACTTA	
GAACCATTTA	
TCTCCTCGTA	
ATATTAAGAA	
ACATGAGATA	
AGTAGAGTAG	
TCCTCCTCCA	
GAACAAGTTA	
GAACCACACA	
CAGTTGCAAA	
TCTTCAGTCA	
AATTTAGGGA	
ACATGCATTA	
TAATTCACGT	
CTTCACACCA	
GAACCTACCA	
TGTAGCAATA	
CAAAGGAATA	
GGGTCAGGAC	
TTCAATACCC	
TAATAGTGAA	
GAACAGACTA	
GTATTTATTA	
GAACATCATC	
TGCAGCTTCA	
AGGCAGTTTA	
ATTAGTTCAG	
ACATGCATAA	
TAGGGAGATA	

Supplementary Table 3. The representative ISS sequences in each group that were tested for their activities in splicing regulation. The control sequences were also indicated.

Sequence Name	Sequences	Note
ISS-k0	CACACCA	The representative k-mer and its mutation in group U
ISS-k0-mut	CATAGCA	
ISS-k1	CTCCTC	The representative k-mer and its mutation in group A
ISS-k1-mut	CGCATC	
ISS-k2	TACAGCT	The representative k-mer and its mutation in group B
ISS-k2-mut	TACGGCT	
ISS-k3	CTTCAG	The representative k-mer and its mutation in group C
ISS-k3-mut	CGTCAC	
ISS-k4	GAACAG	The representative k-mer and its mutation in group D
ISS-k4-mut	GTACCG	
ISS-k5	CAAAGGA	The representative k-mer and its mutation in group E
ISS-k5-mut	CATACGA	
ISS-k6	AGATATT	The representative k-mer and its mutation in group F
ISS-k6-mut	AGCTGTT	
ISS-k7	ACATGA	The representative k-mer and its mutation in group G
ISS-k7-mut	ACGTGA	
ISS-k8	AATTTA	The representative k-mer and its mutation in group H
ISS-k8-mut	AGTGTC	
ISS-k9	AGTAGG	The representative k-mer and its mutation in group I
ISS-k9-mut	AGCAGT	
ISS-kk0 (group U)	CACACCAGCACACCA	To increase the sensitivity, two copies of representative ISS k-mers were used in Fig 2e and 5 to test the function of ISS in different contexts.
ISS-kk1 (group A)	CTCCTCCTCCTCCT	
ISS-kk2 (group B)	TACAGCTTACCAGCT	
ISS-kk3 (group C)	ACTTCAGGCTTCAGA	
ISS-kk4 (group D)	GAACAGGAACAG	
ISS-kk5 (group E)	CAAAGGAGCAAAGGA	
ISS-kk6 (group F)	AGATATTGAGATATT	
ISS-kk7 (group G)	ACATGAGACATGAT	
ISS-kk8 (group H)	CAATTTAGAATTTA	
ISS-kk9 (group I)	CAGTAGGTAGTAGG	
ISS-kkk0 (group U)	CACACCAGCACACCAUCACACCA	Three copies of the representative ISS k-mers were used to do the affinity
ISS-kkk2 (group B)	UACAGCUACAGCUUACAGCUU	

ISS-kkk3 (group C)	GUUCAGGCUUCAGACUUCAGG	purification assay.
ISS-kkk4 (group D)	GGAACAGAGGAACUGGAACAG	
ISS-kkk5 (group E)	CAAAGGAGCAAAGGACAAAGG	
ISS-kkk6 (group F)	GAGAUAUUAGAUAGAGAUAUG	
ISS-kkk7 (group G)	GACAUGAUACAUGAUACAUGA	
ISS-kkk8 (group H)	GAAUUUAGAAUUUAGAAUUUA	
ISS-kkk9 (group I)	CAGUAGGUAGUAGGGAUAGGG	

Supplementary Table 4. Control sequences used. The control 1 was the sequence in the empty pZW2C vector. The control 2 and 5 were arbitrarily picked hexamer and octamers that lack detectable ISS activity. All other control sequences were arbitrarily picked from a random library of decanucleotides and lacked known SRE activities.

Sequence ID	Sequence	Note
Control 1	ACTG	These are the control sequences used when testing ISS activities in other introns or exons. Control 1-4 were used in Fig. 2e and controls 4-7 were used in Fig. 3a.
Control 2	AGCAGT	
Control 3	GAATTGTTT	
Control 4	TCGAGTTTAG	
Control 5	TTGATATA	
Control 6	TTACTGTACT	
Control 7	GTCCGTCAGT	
Control 8	CGATTGGAAC	These are the control sequences inserted between two competing 5′ splice sites (Fig. 3b). The same controls were used previously [1].
Control 9	ACACGCGGGT	
Control 10	AATCAATTCC	
Control 11	GAATTCATGT	
Control 12	CAUAGCAGAUUGCAUCAUACAU	The control sequence used for the affinity purification assay
Control 13	ACGCGAUACGCGAUACGCGAU	

Supplementary Table 5: Correlation of ISSs to the splicing changes in the entire transcriptome. The tissue specific splicings of the skipped exons were obtained from the RNA-seq results of Illumina's Body Map 2.0 dataset. The alternative exons with ISS at donwnstrean or within the exons were extracted, and their PSI values were compared with exon sets containing decoy ISSs. The overall p-value were calculated by bootstrapping comparison, and we set its threshold at 0.01 to ensure the false discovery rate <1 for each group. The significant changes of PSI were highlighted in light green. Within the exonic context, the ISSs that promoted splicing in a reporter (with ESE activity) were marked with green color, whereas the ISSs with ESS activity in a reporter were marked with red.

Location	ISS Groups	Overall PSI changes	Overall p-value	Individual tissues with significant changes (p-value)	# of total tissues tested
downstream introns	A	Increase	0.0785	lymph/adrenal (0.006),muscle (0.022)	11
	B	Decrease	0.00204	adrenal (0.022)	10
	C	Decrease	0.206	none	14
	D	Decrease	0.00232	none	15
	E	Decrease	0	none	14
	F	Decrease	0.120	none	12
	G	Decrease	0	none	14
	H	Decrease	0.225	none	12
	I	Increase	0.001	colon (0.035)	15
Within exon	A	Decrease	0	heart (0.024),adrenal (0.004)	8
	B	Increase	0.0664	none	7
	C	Increase	0.0126	prostate (0.002), kidney (0.047)	12
	D	Increase	0.00545	adipose (0.006),prostate (0.019)	14
	E	Decrease	0.0334	thyroid (0.005), breast (0.015), wbc (0.032)	14
	F	Increase	0.0589	none	6
	G	Increase	0	none	13
	H	Decrease	0	ovary (0.024)	5
	I	Decrease	0	heart (0.022)	7

Supplementary Table 6. Raw data from principal component analyses.

ISS group	A	B	C	D	E	F	G	H	I	U
CEexon-CEintron	2.405	9.7078	2.6973	5.7305	5.6043	-4.4737	4.0952	-14.3941	-6.8802	8.0452
CE5ssWeak-strong_Exon	-0.2521	0.4468	-0.1373	0.3872	-0.6009	-0.8141	0.0656	-1.55	-0.8004	0.7057
CE5ssWeak-Strong_Intron	0.1441	-0.6207	-0.3262	-0.2191	-0.9326	-0.7235	-0.2775	0.1319	-0.1128	-0.309
CE3ssWeak-Strong_Exon	-0.7826	-0.02	0.1384	0.9053	0.8982	2.3456	0.4212	-0.5442	-0.1415	-0.3075
CE3ssWeak-Strong_Intron	-1.8419	-0.9312	-0.3705	-0.2323	-0.0484	2.3032	0.0333	2.4127	0.1823	-1.4715
SE-CEexon	1.6307	-0.2319	-0.0247	-0.4782	-0.7312	-0.724	-0.7973	0.546	1.445	-0.2795
SE-CEintron	0.8214	0.6768	-0.0636	-0.1643	-0.1423	-2.0106	-0.5369	-0.8253	0.1308	0.7956
CEregion1-2	-0.1218	0.0577	0.1429	0.1468	0.3997	1.337	0.319	0.7561	0.3216	-0.2713
CEregion3-4	0.4802	1.151	0.6595	0.5866	0.8808	1.8461	0.6133	0.9012	0.1557	1.0113
A5Eext-core	0.2187	-0.58	0.1518	-0.2107	0.3291	1.8494	-0.0443	1.9011	1.0155	0.2992
A5Eext-intron	0.3655	1.0453	0.4058	0.7926	1.4546	1.1002	0.9997	-1.0203	-0.5882	3.0072
A3Eext-core	0.5743	-0.9077	0.0148	-0.5708	-0.3481	1.5398	-0.2222	2.2625	1.1569	-0.3551
A3Eext-intron	0.9223	0.8867	0.2812	0.4961	0.9932	-0.2447	0.9482	-1.5887	-0.7504	2.51

CEexon-CEintron: Z-score for the difference of frequency between the exon of CE versus that in the introns of CEs, others defined similarly.

Exon convexity:

CE region 1: +1 - +25 in exon

CE region 2: +26 +50 in exon

CE region 3: -25 - -1 in exon

CE region 4: -25 - -1 in exon

Supplementary Table 7. The proteins identified through RNA affinity purification. Each protein band in Supplementary Fig. 5 was cut from SDS-PAGE gel and trypsin digested for mass-spectrometry analyses to obtain protein IDs. Only the putative RNA binding proteins were shown.

Group ID (representative k-mer)	Protein Band #	Protein ID	Number of peptides identified
Group B (TACAGCT)	B1	NONO (non-POU domain containing octamer-bindin)	26
		hnRNP I (PTB)	6
	B2	La protein	19
		YB-1 (Y box-binding protein 1)	12
	B3	SF2P32	10
Group C (CTTCAG)	C1	SFPQ (PTB-associated splicing factor)	21
		Nucleolin	19
		hnRNP U like1	19
	C2	La	26
		eEF1A1	12
		hnRNP H1	13
	C3	hnRNP A2/B1	25
	C4	hnRNP A1	15
Group D (GAACAG)	D1	Cold shock domain-containing protein E1 (CSDE1)	13
	D2, D4	GRSF-1	36
	D3, D5	GRSF-1	33
		RBM45 (CELF-3 homolog)	7
	D6	La	11
		hnRNP H	37
	D7	ASF/SF2	10
Group E (CAAAGGA)	E1	Fusion (Involved in t(12;16) in malignant liposarcoma)	10
	E2	RNA helicase DDX5	22
		hnRNP Q	11
		U1-70K (SNRNP70)	20
			13
		Fusion (Involved in t(12;16) in malignant	

		liposarcoma).	
	E3	U1-70K (SNRNP70)	34
	E4	HnRNP H1	19
	E5	hnRNP F	16
	E6	hnRNP A2/B1	26
	E7	hnRNP A1	18
		ASF/SF2	12
	E8	ASF/SF2	16
		U1 snRNP protein A	12
	E9	U1 snRNP protein C	7
		ASF/SF2	10
	E10	Small nuclear ribonucleoprotein Sm D1	4
		small nuclear ribonucleoprotein Sm D2	13
Group F (AGATATT)	F1	KH type-splicing regulatory protein (KHSRP)	22
	F2	Keratin, type I cytoskeletal 9	26
		Keratin, type II cytoskeletal 1	13
	F3	Ras GTPase-activating protein-binding protein 1 (G3BP1)	17
	F4	Ras GTPase-activating protein-binding protein 1 (G3BP1)	9
		Ras GTPase-activating protein-binding protein 2 (G3BP2)	16
	F5	La protein	31
		hnRNP H1	13
	F6	DAZAP1 (DAZ-associated protein 1)	13
	F7	hnRNP D0 (AU-rich element RNA-binding protein 1)	17
		hnRNP A3	8
	F8	hnRNP D-like (AU-rich element RNA-binding protein)	11
		hnRNP A3	10
	F9	hnRNP A1	15
		hnRNP A2/B1	25
	F10	hnRNP A2/B1	13
		hnRNP A1	20
	F11	hnRNP A1	12
		CIRBP (Cold inducible RNA binding protein)	7

Group G (ACATGA)	G1	GTFII-I (General transcription factor II-I)	20
	G2	g3BP1 (Ras GTPase-activating protein-binding protein 1)	11
	G3	hnRNP L	20
Group H (AATTTA)	H1	La protein	32
	H2	DAZAP1 (DAZ-associated protein 1)	16
	H3	hnRNP D0 (AU-rich element RNA-binding protein 1)	17
		hnRNP A3	8
	H4	hnRNP D0 (AU-rich element RNA-binding protein 1)	8
		hnRNP D-like (AU-rich element RNA-binding protein)	11
		hnRNP A3	10
	H5	hnRNP A2/B1	25
		hnRNP A0	6
		hnRNP A1	16
	H6	hnRNP A2/B1	13
		hnRNP A0	9
		hnRNP A1	20
Group I (AGTAGG)	I1	DHX36 (DEAH box protein 36)	35
	I2	Fusion(Involved in t(12;16) in malignant liposarcoma)	13
	I3	hnRNP H1	21
	I4	DAZAP1 (DAZ-associated protein 1)	12
		hnRNP F	17
	I5	hnRNP D0 (AU-rich element RNA-binding protein 1)	15
		hnRNP F	12
	I6	hnRNP D0 (AU-rich element RNA-binding protein 1)	17
		hnRNP A3	8
	I7	hnRNP D0 (AU-rich element RNA-binding protein 1)	8
		hnRNP D-like (AU-rich element RNA-binding protein)	11
		hnRNP A3	10
	I8	hnRNP A2/B1	25

		hnRNP A0	6
		hnRNP A1	16
	I9	hnRNP A2/B1	13
		hnRNP A1	20

Supplementary Note

Vectors and constructs:

All the splicing reporters were constructed from a backbone vector, pZW1, which contains a multicloning site between two GFP exons [2]. To construct the reporter for FAS-ISS screen, a constitutive exon - exon 6 of the human *SIRT1* gene (Ensembl ID: ENSG00000096717) - was amplified together with portions of its flanking introns in two PCR reactions [1]. The first PCR reaction amplified 327 bp of the upstream intron 5, the exon 6 and 11 bp of the downstream intron 6 using primers 1 and 2 (Supplementary Table 1), and the second PCR targeted position 12 to position 266 of the downstream intron using primers 3 and 4 (Supplementary Table 1). The two PCR fragments were cloned into pZW1 to generate the resulting construct, pZW9, which contains a three-exon minigene with exon 1 and 3 forming an intact GFP gene and a multicloning site at 11 bp downstream of the 5′ss of the test exon 2 (SirT1 exon 6) [1]. The three-exon minigene of pZW9 was transferred into the site-specific integration plasmid pcDNA5/FRT by *Nhe*I/*Bam*HI digestion and ligation, generating the vector pZW11 that was stably transfected into 293 FlpIn cell line. To confirm the ISS activities in different cell types, both pZW9 and pZW11 were used in the transient transfection experiments as they contain the same minigene driven by CMV promotor.

The candidate ISS sequences and controls were inserted into reporter pZW11 or other minigenes using XhoI/ApaI sites. To this end, we used a forward primer CACCTCGAG(N_{6-10})GGGCCCCAC and reverse primer GTGGGGCCC(N_{6-10})CTCGAGGTG which contained the candidate sequences (designated N_{6-10}) flanked by

*Xho*I and *Apa*I sites. The two primers were annealed, digested, and ligated into the reporter vectors.

To make the random sequence library, we extended the foldback primer CACCTCGAG(N_{10})GGGCCCACACGTTTTTTTCGTGTGGGCCC with Klenow and dNTPs for 20 min at room temperature, then heat-inactivated the polymerase, cut the DNA with *Xho*I and *Apa*I and ligated into pZW11. No purification steps were needed since the reaction buffer was diluted at each step. The final amount of primer in the ligation reaction was 2.4 pmol [2]. The ligation product was used to transform ElectroMax DH-5α, and we transform sufficient numbers of *E. coli* cells to obtain >$2x10^6$ colonies, achieving ~2-fold coverage of the 4^{10} = ~10^6 possible DNA decamers.

To test ISSs in a heterologous exon context, a reporter vector, pZW2C, was constructed by inserting the exon 2 of Chinese hamster dihydrofolate reductase (DHFR) gene and part of its flanking introns between the two GFP exons. This reporter was modified from pZW2 that was originally used in the FAS-ESS screen and contains an XhoI/ApaI restriction site inside the test exon 2 [2]. The pZW2 was digested with Xho1/ApaI and filled in with an oligonucleotide (obtained by annealing primers 5 and 6, Supplementary Table 1) to destroy the exonic restriction sites. We then introduced a new XhoI/ApaI restriction site at 18 nt downstream of the exon 2 by three consecutive PCR reactions: the first PCR used primers 8 and 9 to amplify the pZW2 vector whose exonic restriction sites was destroyed, the second PCR reaction used primers 7 and 10 to amplify the pZW2, and the third PCR reaction use the products from PCR 1 and 2 as template and use primers 9 and 10 to amplify a fragment that contain intronic restriction sites [2]. The resulting product of PCR3 was inserted into pZW2 digested with NheI/PstI to obtain the

reporter pZW2B. To increase the ISS detection sensitivity, the pZW2C was further generated by weakening the 3′ ss of exon 2 in pZW2B with site-directed mutagenesis so that the exon 2 was included in ~50% of mRNA in the absence of ISS.

The reporter with competing 5′ ss was described previously [1]. We also used a modular splicing reporter to test how the FAS-ISS sequences affect splicing when inserted into an exon. This modular reporter is constructed by inserting a test exon (Exon 12 of the human *IGF2BP1*, Ensembl ID ENSG00000159217), together with portion of its flanking introns, into the backbone vector pZW1 [3,4]. Inside and adjacent to the test exon, several restriction enzyme cloning sites were strategically engineered, allowing to change different splicing regulatory sequences and the splicing sites in combination with others. The ISSs were inserted into this vector using XhoI/ApaI sites.

Knock down and overexpression of *trans*-acting splicing factors

The siRNAs used to knock down candidate *trans*-factors were purchased from Dharmacon (On-target SMARTpool). One day before the transfection, 1.25×10^5 293T cells were seeded onto 24-well plates. For each well, 60 pmol siRNA were transfected with Lipofectamine 2000 (Invitrogen) following manufacturer's instructions. After 48 hours of siRNA transfection, we transfected 0.2 μg splicing mini-gene reporter constructs in each sample, and harvested the cells 24 hours after the second transfection. The cells were split into 2 portions, one for total RNAs purification and RT-PCR, and the other for protein analyses by Western blot. YB-1 antibody was purchased from Cell Signaling (Cat#2749, 1:1000 dilution). hnRNP L antibody was purchased from Sigma

(Clone#4D11, 1:10000 dilution). Flag antibody was purchased from Sigma (Clone #M2, 1:1000 dilution).

To over express *trans*-acting factors, cells were seeded onto 24-well plates 1 day prior to transfection. For each transfection, 0.2 μg of splicing mini-gene reporters were mixed with 0.8 μg of protein expression vector and 2 μl of Lipofectamine 2000 (Invitrogen). After 20 min incubation at room temperature, the mixtures were added to each well, and the cells were harvested after 72 h for further RNA or protein analysis. The hnRNP L expression vector is a gift from Dr. Jingyi Hui in Shanghai Institute for Biological Science, and the YB1 expression vector is a gift from Dr. Anne Willis in University of Nottingham.

The expression constructs for the fusion proteins of a PUF domain and different functional domains were generated as previously described [6]. The RS domains used in the fusion proteins are: residues 123-238 of SFRS7 (NP001026854), residues 180-272 of SFRS5 (NP008856), residues 117-221 of SFRS2 (NP003007) and an $(RS)_6$ sequence. The Gly-rich domains are residues 195-320 of hnRNP A1 (NP_002127), residues 203-353 of hnRNP A2 (NP112533), residues 211-378 of hnRNP A3 (NP919223), and the short peptide of YGGGGPGYGNQGGGYGGG. For each transfection, 0.8 μg of expression vectors for fusion protein and 0.2 μg of splicing reporters were mixed and transfected into 293T cells with Lipofectamine 2000 (Invitrogen), and the splicing outcomes were analyzed with semi-quantitative RT-PCR.

Exon and Intron Datasets

The exon and intron datasets were generated using similar filters as described earlier [1]. Alignments of human and mouse cDNA and EST sequences to the human genome and the mouse genome were obtained from the UCSC Genome Browser, http://genome.ucsc.edu [7,8]. Constitutive exons (CEs), skipped exons (SEs) and pairs of alternative 5'ss exons (A5Es) were defined as in [9]. All splice site pairs were required to conform to the GT-AG or GC-AG consensus (or be supported by multiple ESTs). To avoid potential EST alignment artifacts, the A3Es and A5Es were further filtered by requiring that the longer isoform differs from the shorter isoform by at least 6 bases, and SEs were required to be at least 6 bases in length. For the human genome, we obtained 6,736 A5Es, 8,142 A3Es, 15,256 SEs and 91,045 CEs. For the mouse genome, we obtained 2967 A5Es, 3922 A3Es, 15050 SEs and 95424 CEs.

The above human and mouse exons were identified independently by using transcript data specific to each organism. Human/mouse orthologous A3Es, A5Es, SEs and CEs were identified based on the human-centric multiz (multiz8way) alignment obtained from the UCSC Genome Browser [10]. For SEs and CEs, we required that the first and last nucleotide positions of the exon be aligned in human/mouse orthologous exons. This procedure yielded 2,964 orthologous human/mouse SEs and 44,368 CEs. For A3Es and A5Es, we required that the first and last nucleotide positions of both the short isoform and the long isoform be aligned in orthologous exons. To expand the set of orthologous A5Es and A3Es, we devised a procedure to obtain putative human/dog/mouse/rat orthologous A3Es and A5Es based on human sequences and multi-genome alignments [1]. Overall, 1,232 A5Es and 1,408 A3Es are categorized as potential

orthologous A5Es and A3Es in human, dog, mouse and rat genomes. These data sets were used for the conservation analysis.

Reference:

1. Wang, Z., Xiao, X., Van Nostrand, E. & Burge, C.B. General and specific functions of exonic splicing silencers in splicing control. *Mol Cell* **23**, 61-70 (2006).
2. Wang, Z. et al. Systematic identification and analysis of exonic splicing silencers. *Cell* **119**, 831-45 (2004).
3. Xiao, X., Wang, Z., Jang, M. & Burge, C.B. Coevolutionary networks of splicing cis-regulatory elements. *Proc Natl Acad Sci U S A* **104**, 18583-8 (2007).
4. Xiao, X. et al. Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat Struct Mol Biol* **16**, 1094-100 (2009).
5. O'Gorman, S., Fox, D.T. & Wahl, G.M. Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251**, 1351-5 (1991).
6. Wang, Y., Cheong, C.G., Hall, T.M. & Wang, Z. Engineering splicing factors with designed specificities. *Nat Methods* **6**, 825-30 (2009).
7. Karolchik, D. et al. The UCSC Genome Browser Database. *Nucleic Acids Res* **31**, 51-4 (2003).
8. Chung, S. et al. Crooked neck is a component of the human spliceosome and implicated in the splicing process. *Biochim Biophys Acta* **1576**, 287-97 (2002).
9. Yeo, G., Holste, D., Kreiman, G. & Burge, C.B. Variation in alternative splicing across human tissues. *Genome Biol* **5**, R74 (2004).
10. Kent, W.J. et al. The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).