Supplemental Data

Materials and Methods

Constructs. All the splicing reporters used were constructed from a backbone vector, pZW1, which contains a multicloning site between two GFP exons (*1*). To make the competing 5' splice site (5'ss) reporter, 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. The first PCR amplified 327 bp of the upstream intron 5, the exon and 11 bp of the downstream intron 6 using primers 1 and 2 (all primer sequences are listed in Table S1). The PCR fragment was digested with *Sal*I/*Hind*III, and then inserted into pZW1 digested with *Xho*I/*Hind*III to produce the plasmid pZW1a. The second PCR targeted position 12 to position 266 of the downstream intron (counting from the 5' end of the intron) using primers 3 and 4, and this PCR fragment was digested with *Eco*RI/*Sac*II, then inserted into pZW1 digested with same enzymes. The resulting construct, pZW9, contains exon 6 of SIRT1 and flanking introns with restriction sites 11 bp downstream of the 5'ss. To generate the second 5'ss, primers 5 and 6 were annealed and digested with *Hind*III/*Eco*RI, and inserted into pZW9. The sequences between the two 5'ss are also shown in Fig 1C.

To make the competing 3'ss reporter, we used pZW8, a plasmid that has exon 6 of the human *SIRT1* gene (with restriction sites inside the exon at 22 bp downstream of the 3'ss) together with portions of its flanking introns inserted into pZW1 (*1*). The second 3'ss derived from exon 2 of the human gamma globin gene (Ensembl ID: ENSG00000197458) which was

amplified (together with its branch point) by PCR using primers 7 and 8, and inserted into pZW8.

To make the intron retention reporter diagrammed in Fig. 4, we amplified exons 3 and 4 of the human *NKIRAS2* gene (Ensembl ID: ENSG00000168256) together with their intervening and flanking introns by two PCR reactions targeted to the 5' and 3' halves of the middle intron and corresponding flanking sequences (using primers 9 and 10 for the first PCR, and primers 11 and 12 for the second PCR). The two fragments were digested and inserted sequentially into the multi-cloning site of pZW1 as described for the competing 5'ss reporter, leaving a multi-cloning site inside the middle intron 19 bp downstream of the 3'ss.

To insert candidate ESS sequences and controls into the competing 5'ss reporter, we used a forward primer CACCTCGAG(N_{6-10})GAATTCCAC and reverse primer GTGGAATTC(N_{6-10})CTCGAGGTG containing the candidate sequences (N_{6-10}) flanked by *Xho*I and *Eco*RI sites. The two primers were annealed, digested, and ligated into the vectors. To insert candidate ESS sequences and controls into the competing 3'ss and intron retention reporters, we used a forward primer CACCTCGAG(N_{6-10})GGGCCCCAC and reverse primer GTGGGGCCC(N_{6-10})CTCGAGGTG containing the candidate sequences (N_{6-10}) flanked by *Xho*I and *Apa*I sites. The two primers were annealed, digested, and ligated into the vectors. The same strategy was used to insert the MS2 hairpin sequences into the competing splice site reporters. The sequence for the MS2 hairpin (aka MS hairpin) is CGTACACCATCAGGGTACG, and the sequence for MSΔ is CGTACCCATCAGGGTACG (folded structures are shown in Fig. 3).

The expression vectors for the MS2 coat protein fused to splicing factors SF2/ASF, hnRNP A1 and the Gly-rich domain of hnRNP A1 were a generous gift from Dr. Richard Breathnach and have been described previously (*2*).

To make the constructs for the natural A5E or A3E exons, 4 primers were designed for each gene. Primers P1 and P4 targeted the exon and flanking introns, whereas primers P2 and P3 are overlapping primers covering the region between the two splice sites to introduce mutations. For the wild type, a PCR product was generated using P1 and P4, and inserted into pZW1 using the *Xho*I/*Sac*II site. For the mutant sequences, we first generated two PCR products, one using the P1/P2 pair, and the other using the P3/P4 pair. In the third PCR reaction, the P1/P4 primer pair and a mixture of the first two PCR products was used as a template, producing the mutant form of the exon. This mutant product was inserted into pZW1 using the *Xho*I/*Sac*II site.

Cell culture and transfection 293 cells were cultured with D-MEM medium supplemented with 10% fetal bovine serum (FBS). Transfections were carried out with lipofectamine 2000 (Invitrogen) in 12-well culture plates according to the manufacturer's instructions. For co-transfection experiments, 0.5 μg of the competing 5'ss or 3'ss reporter construct was mixed with appropriate amounts of fusion protein expression plasmid (or empty vector control), as described in the text and Fig. 3, and the DNA mixture was transfected with Lipofectamine 2000.

RNA purification and RT-PCR Total RNA was purified from transfected cells using PURESCRIPT RNA isolation kit (Gentra Systems), followed by DNase I treatment. The reverse transcription reaction was carried out using 2 μg total RNA with SuperScript III (Invitrogen) according to the manufacturer's instructions. One tenth of the product from RT reaction was used for PCR (20 cycles of amplification, with trace amounts of α-^{32}P-dCTP in addition to non-radioactive dNTPs). The primers used were AGTGCTTCAGCCGCTACCC for the upstream GFP exon and GTTGTACTCCAGCTTGTGCC for the downstream GFP exon.

Quantitation of spliced isoforms RT-PCR products were separated by 5% PAGE gels, exposed to phosphorimager, and scanned with a Typhoon 9400 scanner from Amersham Biosciences. The amount of each splice form was measured with ImageQuant 5.2 (Molecular Dynamics). All experiments were repeated at least twice.

For the data in Fig. 1, the proportion of the upper band from 14 transfections (1 test and 6 controls, in duplicate) were ranked and the rank-sum test was applied to the ranks of the two test samples. The probability that these two samples ranked lowest and second lowest of the 14 under the null hypothesis is $2/(14 \times 13) = 1/91$ (listed as $P < 0.02$ in the figure legend). To quantify the four splice forms in the intron retention reporter (Fig. 4), the intensity of each splice form was measured with ImageQuant 5.2, and the percentage of each isoform was calculated and plotted. By convention, we have left the results in terms of phosphorimager intensity values rather than molar ratios. Because body-labeling was used, these values are different, and of course the intensity value tends to under-estimate the true molar ratios of shorter products relative to longer products.

Exon and Intron Datasets Alignments of human and mouse cDNA and EST sequences to the human genome (hg17) and the mouse genome (mm6) were obtained from the UCSC Genome Browser, http://genome.ucsc.edu (*3, 4*). Constitutive exons (CEs), pairs of alternative 3'ss exons (A3Es), pairs of alternative 5'ss exons (A5Es), skipped exons (SEs) and retained introns (RIs) were defined as in (*5*). 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 differ 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 5631 A5Es, 7155 A3Es, 34795 SEs, 96338 CEs and 1259 RIs. 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 (*4*). For SEs and CEs, we required that the first and last nucleotide positions of the exon be aligned in human/mouse orthologous exons. 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. Applying these criteria yielded the following numbers of orthologous human/mouse exon pairs: 281 A5Es, 301 A3Es, 1649 SEs and 26340 CEs.

To generate the datasets of exons with and without upstream or downstream decoy splice sites for Fig. 1, intronic regions of maximum 200 bases flanking constitutive internal exons were extracted from introns of at least 50 bases in length. Each authentic splice site was scored using a maximum entropy model which accounts for statistical dependencies between adjacent and non-adjacent splice site positions (*1*). The intronic segments flanking each exon were searched for potential splice sites using the same model. For an authentic 5'ss with score C bits, if a potential splice site with score $\geq$ C bits was found within 100 bases downstream of the 5'ss, this exon was classified as an exon with a downstream decoy 5'ss and the region between the 5'ss and the potential splice site was extracted for the analysis shown in Fig. 1A. If no potential splice site with score $\geq$ C-3 bits was found within 200 bases downstream of the 5'ss, this exon was classified as an exon without a downstream decoy and the intronic region adjacent to the authentic 5'ss was extracted for the analysis shown in Fig. 1A. The cutoff C-3 bits was used in

this instance to exclude ambiguous cases where a potential decoy of score close to the authentic splice site was present. A similar protocol was applied to classify exons with and without upstream decoy 3'ss.

The natural A5E and A3E exons used in the mutational analyses shown in Fig. 2 and Fig. S3 were chosen from the subset of A5E and A3E events that had high EST coverage of both alternative exon forms in a range of tissues. This criterion was used in order to identify candidates for which both isoforms were likely to be expressed in 293 cells.

Sets of orthologous alternative 3'ss and alternative 5'ss exons. To obtain datasets of high quality for analysis of conservation of splicing regulatory elements in A3Es and A5Es, we devised a procedure to obtain putative human/dog/mouse/rat orthologous A3Es and A5Es based on human sequences and multi-genome alignments.

(i) For the 5631 human A5Es and 7155 human A3Es, we considered the alignment of the first two intronic positions upstream and downstream of both the long and short isoforms in human, dog, mouse and rat genomes using the UCSC multi8way alignment. Requiring these positions to be present and aligned in all 4 species yielded 2053 A5Es and 2593 A3Es.

(ii) We required that the lengths of regions aligned to the long and short human isoforms of each A3E or A5E event differed by 0, 3, 6 or 9 bases in the human, dog, mouse and rat genomes. This criterion derived from study of transcript-based human/mouse orthologous A3Es and A5Es. Of the 281 human/mouse A5E ortholog pairs, 99% of the short isoforms and 96% of the long isoforms differed by a multiple of 3 bases. Of the 301 human/mouse A3E orthologs, 98% of the short isoforms and 94% of the long isoforms differed by a multiple of 3 bases. In

addition, more than 99% of the orthologous isoforms differed by no more than 9 bases. Using this filter, we retained 1531 A5Es and 1832 A3Es.

(iii) For both the short and long isoforms of A5Es and A3Es from all 4 genomes, we required presence of at least one open reading frame without stop codons. In this step, we retained 1480 A5Es and 1759 A3Es.

(iv) We required that the scores of the splice sites of both short and long isoforms of A5Es and A3Es differed by no more than 4 bits among all 4 genomes. This cutoff was set by considering human/mouse orthologous A5E and A3E pairs for which both isoforms were supported by transcripts in both organisms, among which 99% had splice sites differing by no more than 4 bits.

Overall, 1074 A5Es and 1318 A3Es passed the above filtering procedure and are categorized as potential orthologous A5Es and A3Es in human, dog, mouse and rat genomes. These data sets were used for the conservation analysis shown in Fig. 2B and Fig. S2.

Conservation of ESS and ESE Hexamers in A3E and A5E exons. To analyze the conservation of ESS and ESE hexamers in the A3E and A5E extension regions, we defined a measure called Conserved Occurrence Rate (COR):

$$COR = \frac{COR_H + COR_M}{2}$$

where COR_H is the COR measure for human and COR_M is the COR measure for mouse. COR_H is calculated as follows:

$$COR_H = 1 - \frac{\sum_{i,j}(ESX_j^{Hexon_i} - ESX_j^{Mexon_i})}{\sum_{i,j} ESX_j^{Hexon_i}} \ ,$$

where the upper sum is taken over all i, j pairs such that $(ESX_j^{Hexon_i} - ESX_j^{Mexon_i}) > 0$

Similarly, COR_M is calculated as:

$$COR_M = 1 - \frac{\sum\limits_{i,j}(ESX_j^{Mexon_i} - ESX_j^{Hexon_i})}{\sum\limits_{i,j}ESX_j^{Mexon_i}},$$

where the upper sum is taken over all i, j pairs such that $(ESX_j^{Mexon_i} - ESX_j^{Hexon_i}) > 0$.

Here, ESX_j is the count of the jth ESS (or ESE) hexamer in the region between alternative splice sites in an exon, with the superscripts $Hexon_i$ and $Mexon_i$ representing the ith orthologous human and mouse exons, respectively. The difference in the occurrence of each ESS/ESE hexamer bewteen human/mouse is summed together for all ESS/ESE hexamers and all exons. This measure of conservation does not require positional alignment of the specific hexamer in the human and mouse orthologs since there was no evidence that the precise location between splice sites is critical for their function in our experimental tests.

To evaluate the conservation of occurrence of ESS and ESE hexamers, we constructed control sets of hexamers in the following way. For each ESS or ESE hexamer (ESX_j), we randomly picked a control hexamer that was not in the FAS-hex3 (*1*) or the RESCUE-ESE set (*6*), but had the same number of occurrences in the A3E or A5E extension regions as ESX_j. As in the case of microRNA target site analysis (*7*), it is important to control for the abundance of the hexamer under study when analyzing conservation because both the number of conserved occurrences and the mutation rate are correlated to hexamer abundance. To generate a broad random control set, we further required the existence of at least 5 eligible control hexamers for each ESS or ESE. This constraint removed 3 ESS and 2 ESE hexamers from the A3E analysis and 1 ESS and 13 ESEs from the A5E analysis. The constraint that the control hexamer have

exactly the same number of occurrences as the hexamer under study was enforced throughout, but could be relaxed somewhat without affecting any of the conclusions.

Density of ESS in retained introns. The majority of retained introns (RI), a total of 885, alter the open reading frame of the transcripts when retained. These "intron-like" RIs were analyzed in Fig. 5, comparing the ESS density in these RIs with the ESS density in constitutive introns (CIs) with similar length and sequence composition as the RIs. The control sets of CIs were constructed in the following way. For each RI, we randomly picked a CI whose length differed from the length of the RI by less than 5% and whose total expected count of ESS hexamers differed from that of the RI by less than 5%. The expected count of ESS hexamers in each intron was calculated using a first-order Markov model if the length of the intron was longer than 100 bases. For example, for a hexamer $x_1x_2x_3x_4x_5x_6$, the expected count in an intron of length L is:

$$E[N(x_1x_2x_3x_4x_5x_6)] = (L\text{-}5)\cdot P(x_1) \cdot P(x_2|x_1) \cdot P(x_3|x_2) \cdot P(x_4|x_3) \cdot P(x_5|x_4) \cdot P(x_6|x_5)$$

where $P(\cdot)$ is the probability (or conditional probability) of the nucleotide in the intron which was estimated via the maximum likelihood method from the empirical frequency (or conditional frequency) of the nucleotide in question. When the length of the intron was no more than 100 bases, a zero-order Markov model was used, i.e. assuming independence between consecutive nucleotides so that the conditional probabilities are replaced by unconditional probabilities above. Three RIs had no control constitutive introns satisfying both the length and sequence composition criteria and were excluded from the analysis.

Supplemental Text

Comments on Figures 1 and 4

The 3 sequences which failed to significantly inhibit the intron-proximal 3'ss in Fig. 1D were: the unclustered decanucleotide TTTCCTGATG (sample 6) that uniquely failed to affect 5'ss selection (Fig. 1C), the hexamer TTCGTT, and the tandem overlap decamer TTCGTTCGTT (the latter two belonging to FAS-ESS group A). The third group A sequence shown in Fig. 1C had modest but significant effects on 3'ss usage. Thus, group A ESSs, which had robust effects on 5'ss usage, appear to have weaker or negligible effects on 3'ss usage.

Consistent with the effects on 3'ss usage, the same three sequences also failed to significantly reduce levels of the intron retention isoform in Fig. 4B. Thus, group A sequences deviated from the pattern of activity observed for other ESS groups in two different assays. Consistently, the unclustered decanucleotide TTTCCTGATG failed to affect splicing in our assays, suggesting that it may represent an exception to the patterns observed for other ESSs, or that this sequence has ESS activity only in very particular flanking sequence contexts.

The results shown in Figures 1-4 raise a number of interesting mechanistic questions about exactly how the factors that recognize ESSs mediate their effects on splice site choice, which were beyond the scope of the present study. One such question is how ESS elements mediate directional inhibition of both downstream 5'ss and upstream 3'ss. Based on previous biochemical studies, two general classes of mechanistic models for the effects of hnRNPs on splice site usage can be proposed. In the first model, drawing from studies of hnRNP A1 (*8*), hnRNPs influence 3'ss selection by directly occluding recognition of an upstream 3'ss by U2AF and/or U2 snRNP, perhaps by directional (3'-to-5') oligomerization along the transcript from a

nearby ESS. This model could be extended to include occlusion of recognition of downstream 5'ss by U1 snRNP, e.g., if 5'-to-3' oligomerization occurs in the vicinity of competing 5'ss (Model 1, Fig. S6). An alternative model which can more simply explain the effects at both 5'ss and 3'ss draws on recent studies of hnRNP I/PTB (*9, 10*). In this model, hnRNPs function by disrupting exon definition interactions that span across the location of a bound ESS, e.g., perhaps by disrupting the important exon-spanning interactions that occur between U1 and U2 snRNPs (Model 2, Fig. S6). Variations of these models can also be envisaged, e.g., with hnRNP-mediated looping out of splice sites either blocking access to snRNPs (Model 1) or preventing their participation in exon definition interactions (Model 2) (*11*).

These two models make identical predictions about the effects on splicing of insertion of an ESS into the middle of an exon or between competing 5'ss or 3'ss (as in Figs. 1-3). However, they make distinct predictions about the effects of ESSs inserted into a retained intron which is recognized by exon defintion as an exon-intron-exon unit (12). If we assume that *NKIRAS2* exon-3-intron-3-exon-4 is recognized by exon definition in steps leading to the retained intron isoform, then the hnRNPs that act by Model 1 and those that act by Model 2 have predictable and distinct effects on the levels of the dual-skipped versus fully spliced isoforms (Fig. 4). In primary transcripts which would otherwise be processed to produce the retained intron isoform, recruitment of hnRNPs that act via Model 1 to an ESS located in the retained intron might inhibit recognition of the upstream 3'ss (belonging to *NKIRAS2* exon 3) by U2 snRNP and/or inhibit recognition of the downstream 5'ss (belonging to *NKIRAS2* exon 4) by U1 snRNP, which might be expected to lead to increased levels of the dual-skipped isoform. On the other hand, recruitment of hnRNPs that act via Model 2 to an ESS located in the retained intron might inhibit the exon spanning interactions between the 5'ss of exon 4 and the 3'ss of exon 3 that would lead

11

to the retained isoform but would not block recognition of these splice sites by snRNPs. This activity would presumably result in an increase in the fully spliced isoform, again at the expense of the retained intron isoform. Thus, one explanation for the differences between Class 1 and Class 2 ESSs is that Class 1 ESSs act by recruiting hnRNPs that act via Model 1 and that Class 2 ESSs act by recruiting hnRNPs that act via Model 2. Variations of these models can also be envisaged, e.g., with hnRNP-mediated looping out of splice sites either blocking access to snRNPs (Model 1) or preventing their participation in exon definition interactions (Model 2) (*11*).

Those ESSs which resembled the binding motifs for hnRNP A1 (FAS-ESS 10-mers F and G in Fig. 4B) led to increases in levels of the dual-skipped isoform, consistent with function through Model 1, which was based on previous studies of A1 function (*8*). However, many other ESSs increased the level of the fully spliced isoform, consistent with activity via Model 2 in which ESSs act to inhibit ESS-spanning exon definition interactions. For those ESSs that resemble 5'ss motifs (e.g., the hexamer GTAAGT), inhibition of exon definition might occur through recruitment of U1 snRNP which could directly compete with that bound to the authentic 5'ss for interactions with U2 snRNP bound to the upstream branch/3'ss. Other means of inhibiting exon definition can also be imagined.

Patterns of conservation in A5E and A3E exons

A high level of sequence conservation was observed in the orthologous human and mouse exons, with the variable 'extension' regions (yellow) of both A5Es and A3Es slightly more conserved than the constant regions (blue) (Fig. S2A), consistent with the presence of a high frequency of conserved splicing regulatory elements. The level of sequence conservation in

the introns flanking these A5Es and A3Es was higher than for introns flanking 'included-conserved exons' (ICEs, i.e. conserved constitutive exons), consistent with previous analyses (*12*), but was lower than that seen in introns flanking alternative-consesrved exons (ACEs, i.e. exons subject to conserved skipping/inclusion). These observations suggest that intronic splicing regulatory elements play some role in alternative splice site selection, but play a more common or more important role in control of exon skipping. Focusing on ESS elements specifically, we observed a substantially higher density of conserved ESSs (according to UCSC multi8way alignment of human, mouse, rat and dog genomes) in the extension regions than in the constant (core) regions for both A5Es and A3Es (Fig. S2B, P = 2.7e-3 and 1.4e-7, respectively, using a discrete version of the Kolmogorov-Smirnov test). A higher density of conserved RESCUE-ESE hexamers was observed in the extension regions than in the core regions for both A5Es and A3Es (Fig. S2C, P = 0.008 and 0.04 respectively), consistent with a role for ESEs in regulation of alternative splice site usage. The COR values for the RESCUE-ESE hexamers are illustrated in Fig. S2D. Occurrence of ESEs was found to be conserved in the A5E and A3E extension regions (P= 2e-3 and 4e-14, respectively).

Supplementary Figure Legends

Figure S1. Effects of ESE sequences on splice site usage. (A) Body-labeled RT-PCR of RNAs from cells transfected with competing 5'ss reporters inserted with control sequences (listed in table S5). **(B)** RT-PCR results with competing 5'ss reporters inserted with RESCUE-ESE sequences (*6*) (listed in Table S4). Five out of 10 ESEs significantly enhanced usage of the proximal splice site (indicated by asterisks, p < 0.02 by rank-sum test). **(C)** RT-PCR of total RNA from cells transfected with competing 3'ss reporters inserted with control sequences (listed in Table S5). **(D)** RT-PCR results with competing 3'ss reporters inserted with RESCUE-ESE sequences. All transfections were repeated at least twice, and a representative is shown.

Figure S2. Patterns of sequence and regulatory element conservation in A3E and A5E exons. (A) Sequence conservation in the vicinity of orthologous human and mouse A3E and A5E exons. **(B)** Conservation of ESS hexamers in A3E and A5E core and extension regions. The cumulative density function (CDF) of the number of conserved ESS hexamers in the core (blue curve) and extension (red curve) regions of orthologous A5Es (left panel, P = 0.003, KS test) and A3Es (right panel, P = 1.4e-7) is shown. **(C)** The CDF of the number of conserved ESE hexamers in the core (blue curve) and extension (green curve) regions of A5Es (left panel, P = 0.008) and A3Es (right panel, P = 0.04) is shown. **(D)** Conserved occurrence rate (COR) of RESCUE-ESE hexamers is shown in green dot. The histograms show COR values for 5000 control sets of hexamers (Supporting Methods). P-values were calculated based on a normal approximation (purple curves) of the distribution of the controls.

Figure S3. Effects of ESS-disrupting mutation on 3'ss usage in IL17RE exon 14.
Alternative 3'ss usage assessed by RT-PCR for a minigene based on exon 14 of human IL17RE gene (as in Fig. 2). A portion of the wild-type sequence between the alternative 3'ss is shown above with FAS-hex3 hexamers in red; the mutant sequence is shown below with mutations highlighted in blue. Duplicate transfections are shown.

Figure S4. Effects of MS2-SF2/ASF fusion protein on splice site usage. (A) The competing 5'ss reporter was inserted with either MS hairpin or MSΔ (a single nucleotide deletion) between two 5'ss, and co-transfected with different amounts of expression construct for the MS2-SF2/ASF fusion protein (0.01 μg to 0.125 μg of expression construct was used, as indicated above each lane). The effect on splicing was assayed by body-labeled RT-PCR with primers to flanking exons of total RNAs from co-transfected cells. **(B)** The competing 3'ss reporter was inserted with either MS hairpin or MSΔ between the two 3'ss, and co-transfected with different amount of expression construct of SF2/ASF fused to MS2 coat proteins (0.025 μg to 2.5 μg of expression construct was used, as shown above each lane). RT-PCR was conducted as panel A. **(C)** The competing 5'ss reporter was inserted with either MS hairpin or MSΔ (a single nucleotide deletion) between two 5'ss, and co-transfected with different amounts of expression construct for an MS2 coat-hnRNP A1 Glycine rich domain fusion protein (0.01 μg to 0.5 μg of expression construct was used, as indicated above each lane). The increased expression of fusion protein at higher plasmid concentrations was confirmed by western blotting (not shown). RT-PCR was conducted as panel A.

Figure S5. Replication of ESS mechanism experiment. (A) Diagram of intron retention

reporter. As in Fig. 4B, the two alternative exons are shown in purple and yellow, and the

retained intron is shown as a dashed box. **(B)** Radioactive RT-PCR results using RNAs purified

from cells transfected with intron retention reporters. This is a duplicate of the experiment

shown in Fig. 4C.

Figure S6. Model and prediction of regulatory roles of ESS in the intron retention reporter

minigene. (A) Two possible models for how ESSs influence splice site selection (see supporting

text). **(B)** Intron retention reporter minigene used to distinguish between models shown in panel

A. Alternative exons are shown as purple and yellow boxes, with the retained intron shown as a

dashed box. Predicted effects on levels of specific spliced isoforms are shown by filled black

arrows.

Table S1: Selected primers used in construction of splicing reporters. Abbreviations of gene names: SirT1: Sirtuin 1; NKIRAS: NFKB inhibitor interacting Ras-like 2; AGER: Advanced glycosylation end product-specific receptor; H2RSP: HAI-2 related small protein; IL17RE: Interleukin 17 receptor E isoform 5.

Primer ID	Sequence	Description/Comments
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	CACGAATTCACTGCAGGTGAGTTCAGGGCCCCAC	Primer pair for additional 5' splice site
6	GTGGGGCCCTGAACTCACCTGCAGTGAATTCGTG	
7	CACAAGCTTCTCGAGGAATTCATGTGGGCCCTGGCACCTTCTGACTGTCA	Forward primer of gamma-globin 3'ss
8	CACGGTACCGGAGCCTGTGAGATTGACAAG	Reverse primer of gamma-globin 3'ss
9	CACGTCGACAGGAATAGGACTTGATCACAAC	These are the primer pair for NKIRAS2 exon3 and flanking introns
10	CACAAGCTTCTCGAGTGTCATAGAAACGCACCTGCTC	
11	CACGAATTCGGGCCCGGGGGCTCCGAGATGGG	These are the primer pair for NKIRAS2 exon4 and flanking introns
12	CACCCGCGGGCATAGGCCTGTGAATGTAC	
13	CACCTCGAGAGAATTGCTTGAACCCAGGAG	P1 of AGER
14	GACTATCCTCGCAGTCACATGTGTTGGGGG	P2 of AGER
15	GTGACTGCGAGGATAGTCAACAAG	P3 of AGER
16	CACCCGCGGGCCTGGAAGCCCTAGGTCTG	P4 of AGER
17	CACCTCGAGCGCCTCATCACCGGCTGCGG	P1 of H2RSP
18	CGTGACTCAGCCGCCTGAAGTTGGAG	P2 of H2RSP
19	CGGCTGAGTCACGGGCGCCGCCTC	P3 of H2RSP
20	CACCCGCGGTGGGAGAGAAGAGATGGTTC	P4 of H2RSP
21	CACCTCGAGTTGAAAAATCAGAAAAACCTAG	P1 of IL17RE exon 14
22	GAGCCGGCCACCGCTATGAGCC	P2 of IL17RE
23	ATAGCGGTGGCCGGCTCTGTG	P3 of IL17RE
24	CACCCGCGGCCAAGCCCAGTTCCTCAGT	P4 of IL17RE

Table S2. FAS-ESS decamer groups. Adapted from (*1*).

FAS-ESS Group	Consensus motif	Putative *trans*-factor
A		?
B		hnRNP A1 and ?
C		hnRNP F/H
D		?
E		?
F		U1 snRNP
G		hnRNP A1 / U1 snRNP
u (unclustered)	not applicable	?

Table S3. FAS-ESS sequences tested.

Sample no.	Sequence tested	FAS-ESS group
1	GTAGGTAGGT	F
2	GGTCCACTAG	u
3	TCGTTCCTTA	A
4	GGGATGGGGT	B
5	GGGGTTGGGA	C
6	TTTCCTGATG	u
7	TGTTTAGTTA	E
8	GTTAGGTATA	G
9	TAATAGTTTA	D
10	TTCGTT	A
11	GTAAGT	F
12	TGGGGT	C
13	GTAGTT	D/E
14	GGTTTG	B
15	GTAGGT	G
16	TTCGTTCGTT	A
17	GTAAGTAAGT	F
18	TGGGGGGAGG	C
19	TTAGTGTAGT	D/E
20	GGTTTGGGTT	B
21	TAGGTAGGTA	G

Table S4. The RESCUE-ESE-based ESE sequences tested. Predicted ESE decamers derived

from overlapping RESCUE-ESE hexamers (*8*).

Sample no.	Sequence	RESCUE-ESE group
1	TTGGATTGGA	5A/3G
2	GAATCATCAG	5B/3A
3	GAAGAAGAAG	5C/3D
4	ATCTTCTTCA	5D
5	ACTACACTAC	5E
6	TCAGATCAGA	3B
7	GACAAGACAA	3C
8	TGAAGTGAAG	3E
9	AAAACAAAAC	3F
10	ACTTCACTTC	3H

Table S5. Control sequences used. These sequences were arbitrarily picked from a random

library of decanucleotides and lacked detectable ESS activity.

Sample no.	Sequence
Ctl 1	CGTGCAATTT
Ctl 2	CGATTGGAAC
Ctl 3	ACACGCGGGT
Ctl 4	AATCAATTCC
Ctl 5	GAATTCATGT
Ctl 6	ATTTGATACC

Supporting References

1. 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.

2. Del Gatto-Konczak, F., Olive, M., Gesnel, M. C., and Breathnach, R. (1999). hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. Mol Cell Biol *19,* 251-260.

3. Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y. T., Roskin, K. M., Schwartz, M., Sugnet, C. W., Thomas, D. J., et al. (2003). The UCSC Genome Browser Database. Nucleic Acids Res *31,* 51-54.

4. Kent W.J., Sugnet C.W., Furey T.S., Roskin KM, Pringle TH, Zahler AM, Haussler D. (2002). The human genome browser at UCSC. Genome Res. *12,* 996-1006.

5. Yeo, G., Holste, D., Kreiman, G., and Burge, C. B. (2004). Variation in alternative splicing across human tissues. Genome Biol *5,* R74.

6. Lewis BP, Burge CB, Bartel DP (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120,* 15-20.

7. Sugnet, C. W., Kent, W. J., Ares, M., Jr., and Haussler, D. (2004). Transcriptome and genome conservation of alternative splicing events in humans and mice. Pac Symp Biocomput, 66-77.

8. Fairbrother, W. G., Yeh, R. F., Sharp, P. A., and Burge, C. B. (2002). Predictive identification of exonic splicing enhancers in human genes. Science *297,* 1007-1013.

9. Zhu J, Mayeda A, Krainer AR. (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol Cell *8,* 1351-61.

10. Izquierdo JM, Majos N, Bonnal S, Martinez C, Castelo R, Guigo R, Bilbao D, Valcarcel J. (2005). Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. Mol Cell *19,* 475-84.

11. Sharma S, Falick AM, Black DL. (2005). Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. Mol Cell *19,* 485-96.

12. Robberson BL, Cote GJ, Berget SM. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol Cell Biol. *10,* 84-94.

13. Nasim FU, Hutchison S, Cordeau M, Chabot B. (2002). High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. RNA *8,* 1078-89.
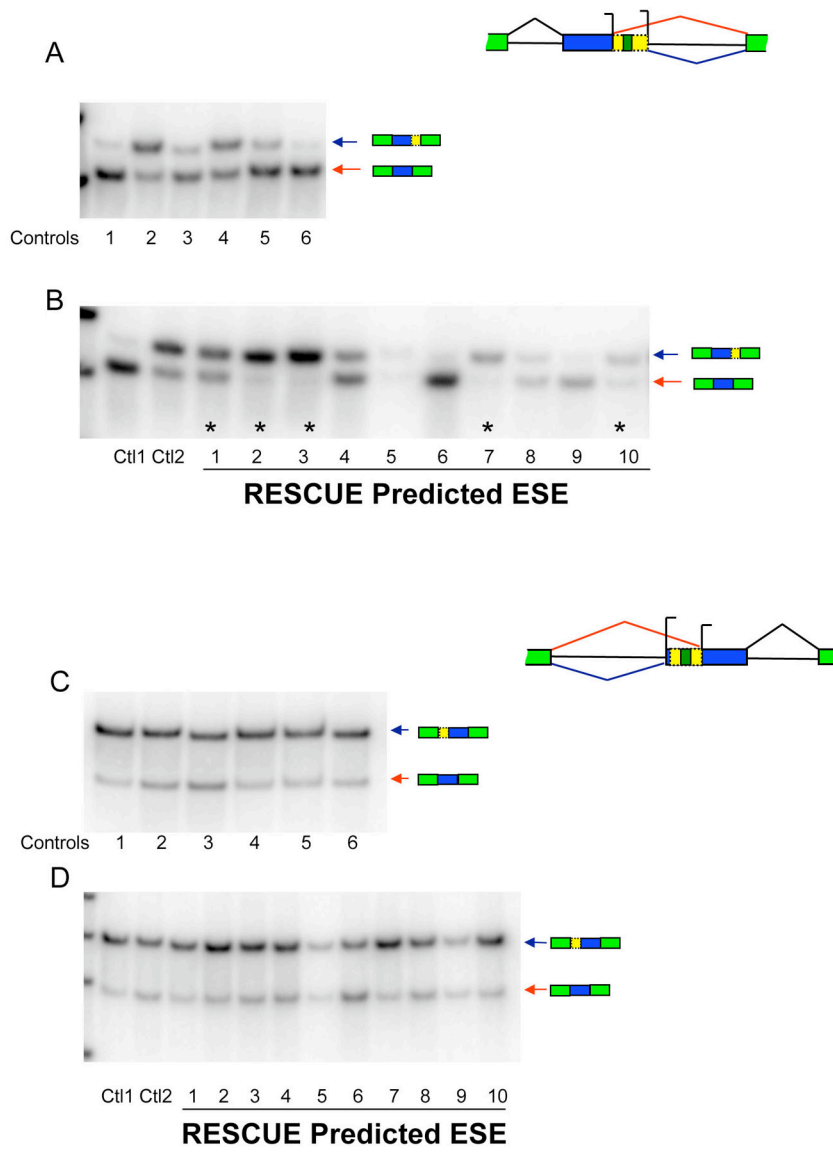
Figure S1

A

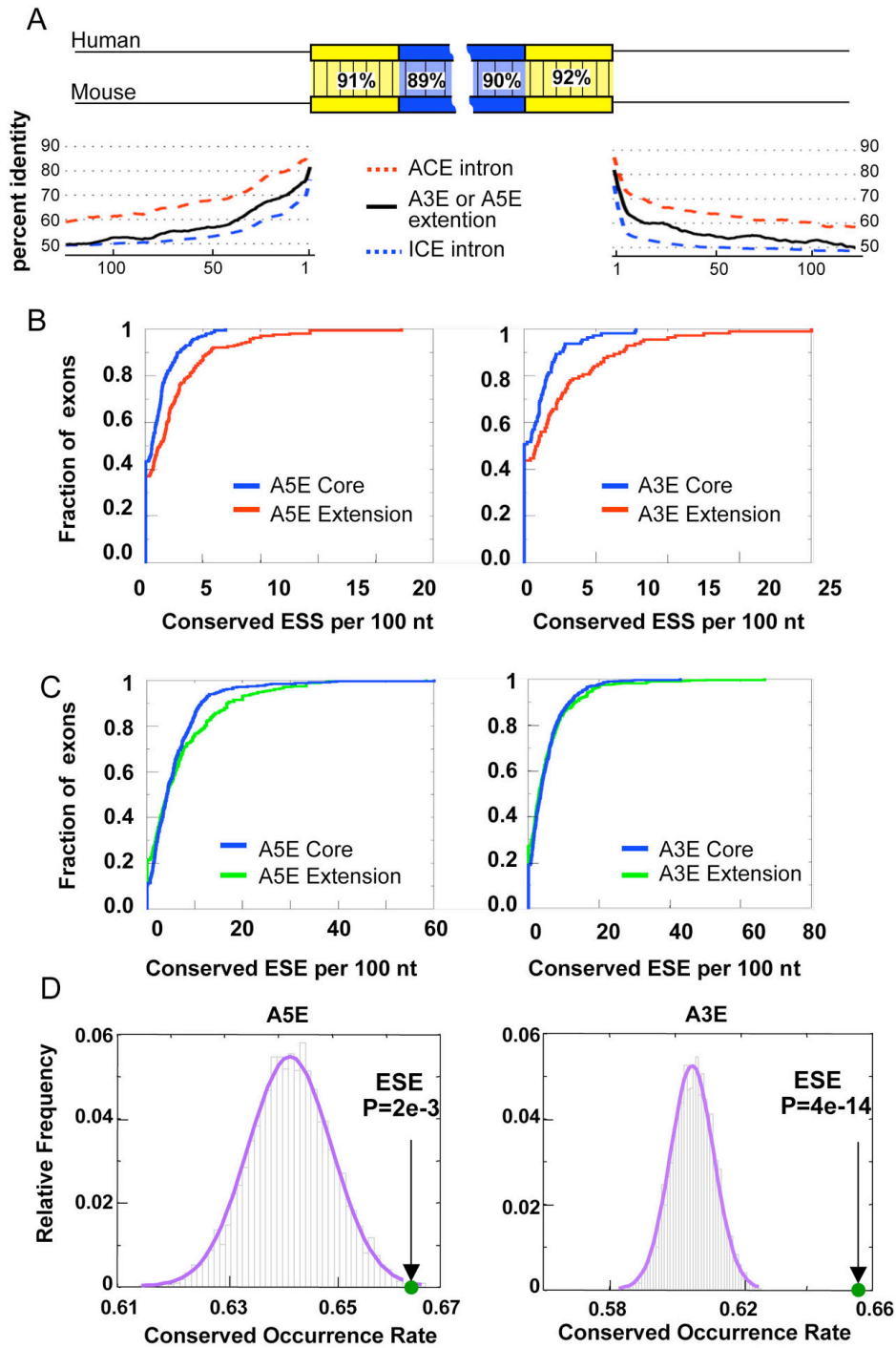Controls 1 2 3 4 5 6

B

*   *   *   *   *
Ctl1 Ctl2 1 2 3 4 5 6 7 8 9 10
RESCUE Predicted ESE

C

Controls 1 2 3 4 5 6

D

Ctl1 Ctl2 1 2 3 4 5 6 7 8 9 10
RESCUE Predicted ESE

23

Figure S2

...ATA**GGGGTGGGGGGC**TC...

↓

...ATAG**c**GGTGG**cc**GGCTC...

Wildtype (+ESS)   **Mutant (-ESS)**
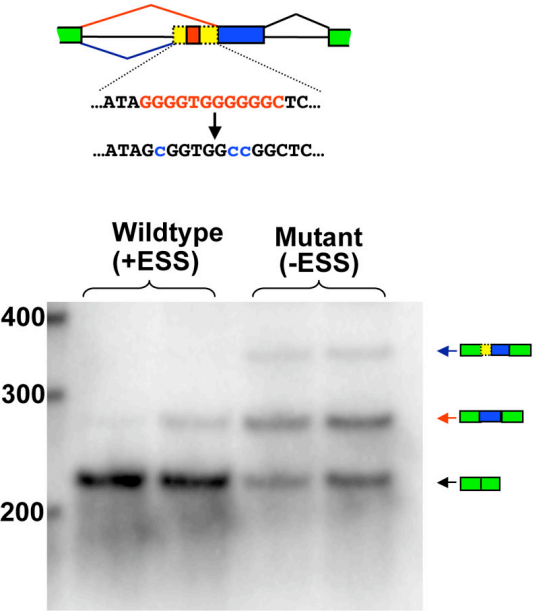
**Exon 14 of IL17RE
(interleukin 17 receptor E isoform 5)**

Figure S4

A

MS2 coat-SF2/ASF

MS

MSΔ

SF2/ASF	0		0.01 μg		0.025 μg		0.05μg		0.125 μg	
Insert	MS	MSΔ	MS	MSΔ	MS	MSΔ	MS	MSΔ	MS	MSΔ

B

MS2 coat-SF2/ASF

SF2/ASF	0		0.025μg		0.1 μg		0.5 μg		2.5 μg	
Insert	MS	MSΔ	MS	MSΔ	MS	MSΔ	MS	MSΔ	MS	MSΔ

C

MS2 coat-Gly domain of A1

Gly domain	0		0.01 μg		0.025 μg		0.05μg		0.125 μg		0.5 μg	
Insert	MS	MSΔ	MS	MSΔ	MS	MSΔ	MS	MSΔ	MS	MSΔ	MS	MSΔ

26

Figure S5

A

B

Figure S6