nature structural & molecular biology

A complex network of factors with overlapping affinities represses splicing through intronic elements

Yang Wang[1], Xinshu Xiao[2,3], Jianming Zhang[2,4], Rajarshi Choudhury[1], Alex Robertson[2], Kai Li[1], Meng Ma[1,5], Christopher B Burge[2] & Zefeng Wang[1]

To better understand splicing regulation, we used a cell-based screen to identify ten diverse motifs that inhibit splicing from introns. Motifs were validated in another human cell type and gene context, and their presence correlated with *in vivo* splicing changes. All motifs exhibited exonic splicing enhancer or silencer activity, and grouping these motifs according to their distributions yielded clusters with distinct patterns of context-dependent activity. Candidate regulatory factors associated with each motif were identified, to recover 24 known and new splicing regulators. Specific domains in selected factors were sufficient to confer intronic-splicing-silencer activity. Many factors bound multiple distinct motifs with similar affinity, and all motifs were recognized by multiple factors, which revealed a complex overlapping network of protein-RNA interactions. This arrangement enables individual *cis* elements to function differently in distinct cellular contexts, depending on the spectrum of regulatory factors present.

The specificity of splicing is mainly defined by splice-site and branch-point sequences located near the 5′ and 3′ ends of introns. Beyond these core signals, multiple *cis*-acting splicing-regulatory elements (SREs) have essential roles in controlling splicing specificity. These SREs are conventionally classified as exonic splicing enhancers (ESEs) or silencers (ESSs) and intronic splicing enhancers (ISEs) or silencers (ISSs). SREs generally function by recruiting *trans*-acting splicing factors that interact favorably or unfavorably with the core splicing machinery such as the small nuclear ribonucleoprotein (snRNP) complexes U1 or U2 (refs. 1,2). In mammals, the splicing of each gene is controlled by multiple SREs and corresponding splicing factors, whose combinatorial actions determine the final splicing outcome[3].

To determine comprehensive rules of splicing regulation, systematic identification of SREs and the associated *trans*-acting factors is needed. Significant progress has been made recently toward comprehensive identification of exonic SREs[4–9]. However, fewer large-scale analyses of intronic SREs have been conducted[9,10], despite the expectation that intronic elements are as important in splicing regulation as their exonic counterparts. Intronic SREs are probably recognized by a greater diversity of proteins, including proteins of the hnRNP class as well as others[2], but knowledge of these factors remains incomplete.

The best-known ISSs include binding sites for the splicing repressors hnRNP I (or PTB) and hnRNP A1 (refs. 11,12), as well as CA-rich sequences bound by hnRNP L (ref. 13). PTB can bind ISSs and inhibit the transition from an exon-definition complex to an intron-spanning spliceosome[11], whereas hnRNP A1 may block recognition of splice sites by cooperative propagation along the pre-mRNA from the primary binding site[12] or by other mechanisms[14]. Overall, the sequence composition and mechanisms of action of known ISSs are quite diverse, which suggests that many more ISSs remain to be discovered.

To systematically identify ISSs, we screened a random library for short sequences that inhibit splicing in cultured human cells. All identified ISS motifs enhanced or inhibited exon inclusion when inserted into exons, and most altered the choice of alternative 5′ splice sites (5′ SSs). Analysis of genomic distributions of these motifs identified four motif classes, each having similar patterns of context-dependent splicing-regulatory activity. Identification of *trans*-acting splicing factors for each motif revealed a densely connected network, which suggests that individual exons are often controlled by multiple factors with overlapping RNA-binding specificities that function through the same element at different temporal or cellular contexts.

RESULTS

Identification of 102 ISSs by using a cell-based screen

To identify ISS sequences without bias, we developed a cell-based screen called fluorescence-activated screen of ISSs (FAS-ISS). We constructed a reporter gene (pZW11) whose exons 1 and 3 form a complete mRNA encoding GFP and whose exon 2 is a small constitutive exon (exon 6 of human *SIRT1*) that is normally included in ~100% of mRNAs (**Supplementary Table 1**). Insertion of known ISSs containing binding sites of hnRNP I ('Test 1') or hnRNP A1 ('Test 2') in the intron downstream of exon 2 caused partial skipping of this exon. After transient transfection with the ISS-containing reporters, we observed about 15–35% GFP-positive cells (**Fig. 1a**), whereas cells with the control construct had negligible fluorescence (<5% green cells).

[1]Department of Pharmacology, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [3]Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, California, USA. [4]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA. [5]Department of Computer Science and Technology, Anhui University, China. Correspondence should be addressed to Z.W. (zefeng@med.unc.edu) or C.B.B. (cburge@mit.edu).

Received 17 July; accepted 5 November; published online 16 December 2012; doi:10.1038/nsmb.2459

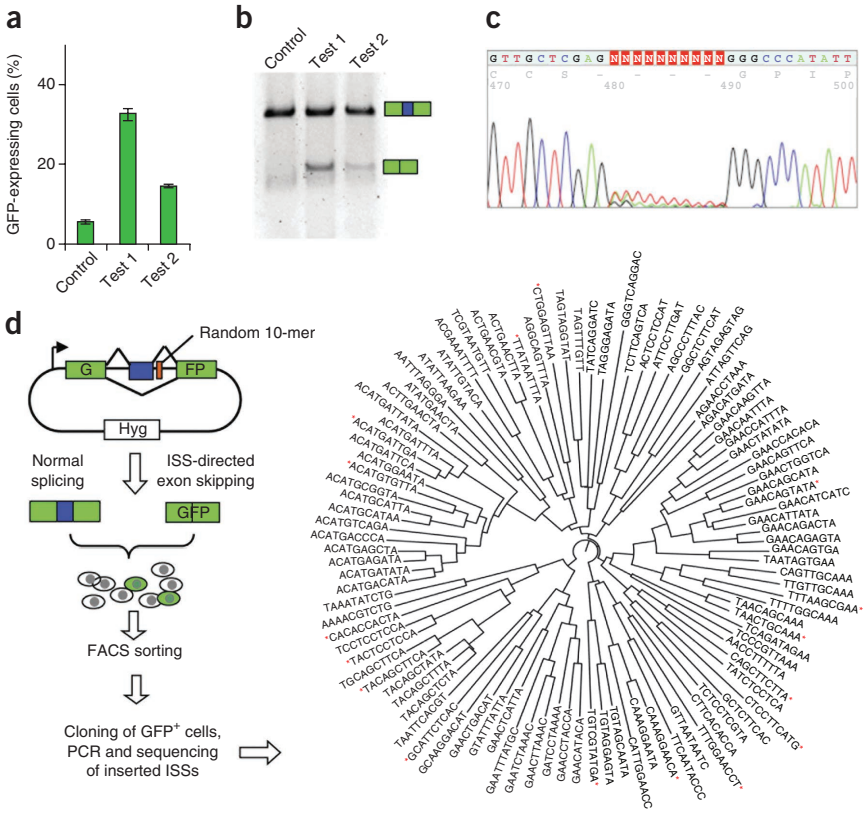

Figure 1 Fluorescence-activated screen for intronic splicing silencers (FAS-ISS) identifies 102 unique ISS 10-mers. (**a**) Test of the reporter system. Two known ISS sequences, Test 1 (PTB binding site TCCTCCTCCA) and Test 2 (hnRNP A1 binding site, TAGGTAGGTA), and an arbitrary control sequence (ACCTCAGGCG) were inserted into the GFP-based reporter (pZW11) and transiently transfected into HEK293T cells. The fraction of GFP-positive cells was measured with flow cytometry 24 h after transfection (mean and s.d. of three or more replicates shown). (**b**) Measurement of splicing with RT-PCR using the RNA templates purified from the transfected cells and primers targeting exons 1 and 3 of the reporter. (**c**) Chromatogram from sequencing of the random 10-mer region of the minigene. DNAs amplified by PCR from stably transfected Flp-In-293 cells shows the randomness of the initial library before screening. Sequences around the insertion region (positions 480–489) are shown. (**d**) Diagram of screening strategy and resulting ISS motifs. The 102 unique ISS 10-mers identified through FAS-ISS are clustered on the basis of sequence similarity by using ClustalW2. The 16 ISS 10-mers that were reinserted into the original construct and retested for ISS activity by transient transfection are indicated (red asterisks).

We confirmed that splicing patterns were consistent with the GFP expression, through reverse-transcription PCR (RT-PCR) (**Fig. 1b**).

We inserted a random pool of decanucleotides (10-mers) at 17 nucleotides (nt) downstream of the test exon, far enough into the intron to avoid interference with the 5′ SS. We transformed the resulting library into sufficient numbers of *Escherichia coli* to achieve approximately two-fold coverage of possible DNA 10-mers. To evaluate the library quality, we randomly picked 24 colonies for plasmid extraction and sequencing and found that all had ten-base inserts with no bias at any position. We further transfected the resulting library into Flp-In-293 cells and sequenced inserted fragments from a pool of stably transfected cells to ensure no sequence bias (**Fig. 1c**).

The reporter was inserted by site-specific recombination into a single target site of the Flp-In-293 cell line. Typically, about one in 1,000–5,000 cells was GFP positive (**Supplementary Fig. 1a**). These GFP-positive cells were sorted by fluorescence-activated cell sorting (FACS) and clonally expanded, and putative ISS sequences were identified by PCR and sequencing (**Fig. 1d**). We conducted 315 transfections in 21 batches, from which 114 ISS 10-mers were identified, 102 of which were unique. (In a few cases, 9-mers or 11-mers were identified; **Supplementary Table 2**.) We identified nine 10-mers at least twice in independent transfections, and 19 pairs differed by a single nucleotide, which suggested that the screen approached saturation. These ISS 10-mers could readily be clustered by sequence similarity (**Fig. 1d**). We inserted 16 of the recovered ISS 10-mers (asterisks in **Fig. 1d**) spanning all clusters back into the reporter pZW11 to assess their activity with a transient-transfection assay[7]. All 16 candidates produced 20–45% GFP-positive cells, compared to <3% for random controls (**Supplementary Fig. 1b**). The percentage of green cells was presumably limited by transfection efficiency, as stable lines containing the same 10-mers generally yielded a higher fraction of green cells.

Identification and validation of ten core ISS motifs

To extract core motifs with intrinsic ISS activity, we used statistical overrepresentation in the recovered 10-mer set as a criterion[7]. Forty-one 5-mers occurred at least five times in the recovered sequences, a frequency about three-fold higher than expected ($P < 10^{-4}$, $n = 10{,}000$ samplings of random 10-mers). Using similar criteria, we identified overrepresented 6-mers and 7-mers, and these k-mers ($k = 5$, 6 or 7) were then clustered by sequence similarity and multiply aligned by using CLUSTALW to identify candidate ISS motifs. At a cutoff dissimilarity score of 3.6, most ISS k-mers fell into one of nine clusters of at least six sequences each, which were designated FAS-ISS groups A to I (**Fig. 2a**). We also analyzed an isolated 7-mer designated group U (unclustered). Such clustering was inherently inexact, as the number of clusters may vary depending on the dissimilarity cutoff. However, this clustering was still informative, as it provided distinct consensus motifs for further analyses. The group A consensus (CTCCT) resembled the binding motif of PTB and of its neuronal homolog, nPTB, and the group I consensus (TAGG) matched the hnRNP A1 motif. Both factors were known to inhibit splicing from intronic binding sites[11,12,14], which confirmed the ability of our screen to identify authentic ISS motifs.

The ISSs identified had higher content of A (37%) or T (28%) and lower levels of C (19%) or G (16%) (**Supplementary Fig. 1c**). This composition was distinct from previously identified exonic SREs, where ESEs were largely A or G rich[5,6] and ESSs were G or T rich[6,7]. When considering individual motifs, we found that sequences identified by FAS-ISS more often resembled the A- or G-rich ESE motifs than ESSs. Consistently, there are examples where SR proteins promoted splicing from exonic sites but inhibited splicing from intronic sites[15–18]. At the dinucleotide level, AC, CA and GA were overrepresented in FAS-ISSs, and CG was under-represented. By comparison, AC and GA were overrepresented in ESEs and under-represented in ESSs[5,7], which

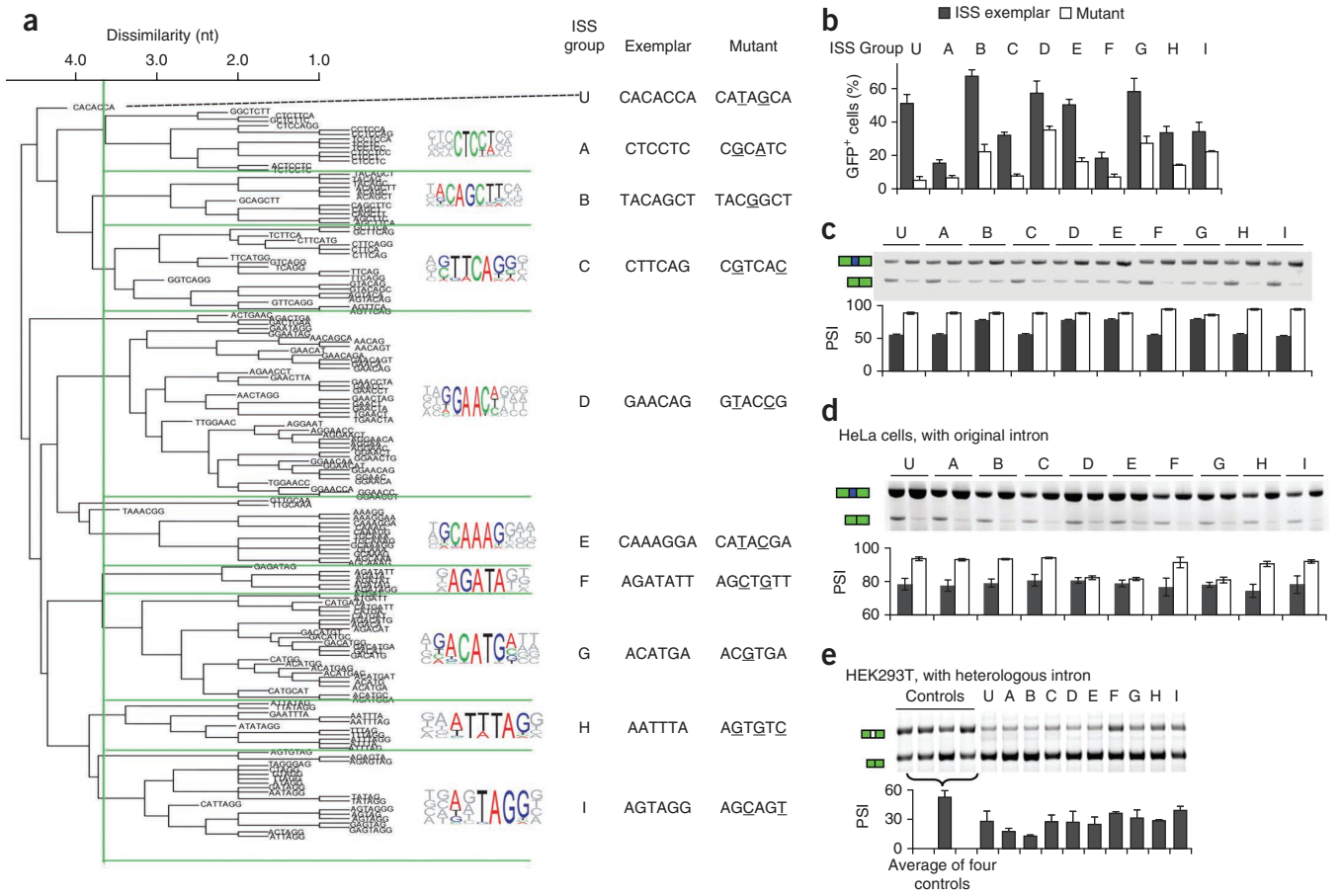

Figure 2 Identification of core ISS motifs and exemplars and validation of activity. (**a**) Overrepresented oligonucleotide *k*-mers, extracted from FAS-ISS 10-mers, are shown clustered on the basis of sequence similarity. The consensus motifs, the exemplars (representative 6-mers or 7-mers) and tested mutations for each group are listed at right. (**b**,**c**) Assessment of splicing by flow cytometry as the fraction of GFP positive cells (**b**) or by semiquantitative RT-PCR (sqRT-PCR) (**c**). Samples are the ISS exemplar and mutant from each group, inserted into the pZW11 plasmid and transiently transfected into HEK293T cells. The PSI values of three independent experiments are shown beneath a representative gel. Error bars, s.d.; $n = 3$. (**d**) Validation of ISS activity in a second cell type. Samples are from HeLa cells transfected with pZW11 containing ISS exemplars and their mutants, and inclusion of the test exon was determined by sqRT-PCR. Error bars, s.d.; $n = 3$. (**e**) Validation of ISS activity in a heterologous intron context. sqRT-PCR as above on HEK293T cells transfected with two copies of ISS exemplars (**Supplementary Table 3**) from each group, inserted downstream of a cassette exon in a different splicing reporter (pZW2C, described in Online Methods). Four neutral sequences that did not match known SREs (**Supplementary Table 4**) were used as negative controls. Error bars, s.d.; $n = 3$.

again suggested similarity between ISSs and ESEs (**Supplementary Fig. 1d**). An exception was the ISS group I (consensus TAGG) that matched motifs bound by hnRNP A1, which typically represses splicing from intronic and exonic locations[19–21].

There are additional ISSs undetectable by our screen, such as motifs whose corresponding *trans* factors were poorly expressed in HEK293T cells or motifs with weak activity to repress the test exon in our reporter. Our screen might also miss very long ISSs or those that function only when presented in multiple copies. A previous screen for intronic SREs using a reporter based on nonsense-mediated mRNA decay, called SPLICE[10], predominantly identified G-rich and G- or T-rich elements. The SPLICE elements exhibited limited overlap with the ISS motifs identified here but overlapped somewhat with motifs identified in a parallel screen for ISEs[22].

To test whether the overrepresented motifs possess intrinsic ISS activity, we analyzed a representative 6-mer or 7-mer (an 'exemplar') from each ISS group (**Fig. 2a**). We used a mutant of the exemplar differing by 1–3 nt as a control (**Supplementary Tables 3** and **4**). Using transient transfection, we observed ISS activity for each exemplar, as judged by generation of GFP-positive cells (**Fig. 2b**). Control sequences all yielded significantly decreased levels of green cells ($P < 0.05$, paired *t*-test), and the mutations of groups A, C, F and U essentially abolished ISS activity. We further confirmed these results by semiquantitative RT-PCR (**Fig. 2c**). These observations suggested that our approach successfully extracted core ISS motifs, and this motivated our use of these exemplars as representatives of the ISS groups in the remainder of this study.

To examine the functionality of FAS-ISSs at a genomic scale, we determined whether the presence of the ISSs affected alternative splicing patterns in the Human BodyMap 2.0 data set. We searched downstream of alternative exons for the ISS sequences and for a set of decoy sequences generated by permuting the ISSs. For each ISS group, we compared the percent spliced in (PSI) values of the exons with an ISS to those associated with a decoy. Eight out of ten ISS groups had decreased inclusion rates as compared to those of their decoy sets, with four groups (B, E, D and G) having significant changes, whereas only one ISS (I) showed a substantially

increased PSI (**Supplementary Table 5**) with a *P*-value cutoff of 0.01, which suggests that most FAS-ISSs can indeed interpret the splicing of alternative exons from RNA-sequencing (RNA-seq) data.

ISSs function in a heterologous intron and cell type

SREs generally function by recruiting specific *trans* factors whose activities may vary in different cell types and pre-mRNA contexts. Thus we tested whether the FAS-ISS motifs were specific to the cell type used in the original screen by transiently transfecting HeLa cells with reporters containing the same exemplars. Whereas no skipping of the test exon was observed in reporters containing randomly chosen sequences, exemplars of all ten ISS groups caused a marked degree of exon skipping, and most mutants showed decreased levels of exon skipping in HeLa cells (**Fig. 2d**).

To test whether these motifs had silencer activity in a second pre-mRNA context, we constructed a new splicing reporter by inserting exon 2 of the Chinese hamster dihydrofolate reductase gene and portions of its flanking introns between two GFP exons. We inserted exemplars of ten ISS groups at downstream of this exon (**Supplementary Tables 3** and **4**) and assayed exon inclusion. We observed that eight of the ten ISS motifs exhibited silencer activity in this heterologous intron compared to the controls (**Fig. 2e**). Two exceptions were groups F and I, whose exon inclusion levels were not markedly different from controls, even though sequences similar to the group I motif had been shown to function as ISSs[20]. Taken together, these data indicated that although the screen was conducted with a single reporter in a specific cell line, all the obtained ISS motifs had activity across different cell types, and most functioned in a heterologous intron context.

ISSs activate or repress splicing from exonic locations

The activities of SREs often depend on their locations relative to nearby splice sites, collectively described as the 'context dependence' of SREs[1]. To determine whether FAS-ISS motifs function in other pre-mRNA locations, we inserted the ISS exemplars into the exon of a modular splicing reporter (**Fig. 3a**). Notably, all ISSs functioned either as ESEs to promote exon inclusion (exemplars C, D, E, F, G and U) or as ESSs to inhibit exon inclusion (exemplars A, H, I). We observed similar functions for the ISSs in both HeLa and HEK293T cells (**Supplementary Fig. 2a**). The only exception was exemplar B, which functioned as an ESE in HEK293T cells but lacked detectable ESE activity in HeLa cells.

Using a reporter with two competing 5′ SSs, we previously found that all FAS-ESSs inhibited use of the proximal 5′ SS when inserted between the two sites, whereas most ESEs promoted use of this site[23]. When inserting each ISS exemplar between competing 5′ SSs of the same reporter, we observed more diverse activities (**Fig. 3b** and **Supplementary Fig. 2b**): three ISS motifs (C, D and E) showed 'ESE-like' activity by promoting use of the proximal site, whereas most ISSs (A, B, F, G, H and I) inhibited the proximal-site usage (that is, ESS like).

We scanned the Illumina Human BodyMap 2.0 data set for alternative exons containing either ISS sequences or their decoys and examined the correlation of ISSs with the change of exon inclusion rate. The ISSs with ESS activity (group A, H and I, shown in red) all showed decreased PSI values when compared to the decoy set (**Supplementary Table 5**), whereas most ISSs with ESE activity were associated with increased PSI value, although only groups D and G produced substantial increases. Similar analyses for exons with alternative 5′ SSs did not produce statistically significant results, probably because the sample size was small in that case.

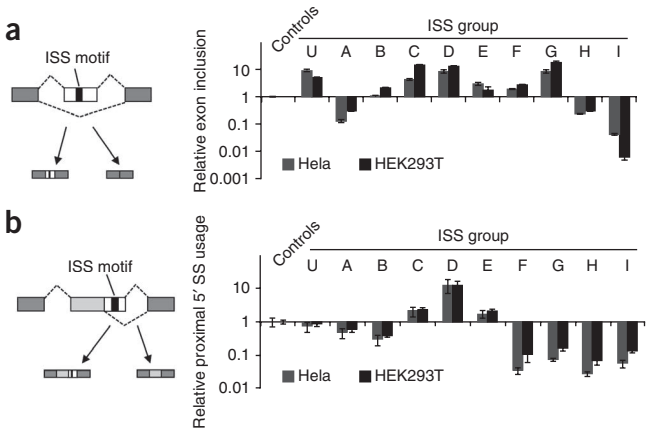

Figure 3 ISS motifs regulate splicing from an exonic context. (**a**) FAS-ISS motifs can function as ESEs or ESSs. Relative exon-inclusion (PSI values divided by the mean of three negative controls (described in **Supplementary Table 4**)) are shown. Samples are from HeLa and HEK293T cells transfected with two copies of ISS exemplars inserted into the test exon of a modular splicing reporter[55]. The mean and s.d. of three independent transfections are shown. (**b**) FAS-ISS motifs can regulate alternative 5′ SSs. Relative proximal 5′ SS usage (**Supplementary Table 4**), given by the observed PSI value (defined as the fractional use of the proximal (downstream) 5′ SS, measured by sqRT-PCR) divided by that observed in controls is shown. Samples are from HeLa and HEK293T cells transfected with a single copy of the exemplar motif from each group, inserted between two alternative 5′ SSs of a minigene reporter[23]. Error bars, s.d.; $n = 3$.

Genomic distributions define distinct ISS clusters

Specific classes of SREs often have characteristic distribution patterns in pre-mRNAs. For example, many ESEs are enriched in constitutive exons relative to introns and in exons with weak splice sites, consistent with their function in enhancing exon inclusion[5,6]. Conversely, ESSs are enriched in introns relative to exons and in skipped exons compared to constitutive exons, consistent with their function in repressing both exon inclusion and use of decoy splice sites[7,23]. Thus the distributional biases of SREs reflect their functions in different pre-mRNA locations and may be used in SRE classification[24]. Previously we observed that all FAS-ESSs exhibited similar distribution biases[7,23]. By contrast, the relative frequencies of different ISS groups varied substantially in constitutive exons, skipped exons and flanking introns—for example, some groups were enriched in constitutive exons, others in skipped exons or introns—revealing unexpected complexity (**Supplementary Fig. 3**).

To explore this complexity and uncover any new patterns, we measured the relative enrichment of FAS-ISSs in 13 pairs of pre-mRNA regions (**Supplementary Table 6**) and analyzed the vectors of enrichment values by using principal component analysis (PCA). The distribution of each ISS group in space of the first two principal components, which represented 86% of the variance, suggested the existence of four clusters of motifs (numbered by cluster size), including two tight clusters (Cluster 1, motifs C, D, E, G; Cluster 3, B, U), the looser Cluster 2 (A, H, I) and the isolated motif F as Cluster 4 (**Fig. 4a**). This classification corresponded closely with the context-dependent activities of the motifs observed above (**Fig. 4b**). Cluster 1 motifs were characterized by enrichment in constitutive exons relative to skipped exons and introns and by a convex distribution along exons (that is, enriched near both ends relative to the centers of exons). All ISSs in this cluster promoted inclusion from an exonic location and favored proximal 5′ SS usage when located between competing 5′ SSs (**Fig. 3**),

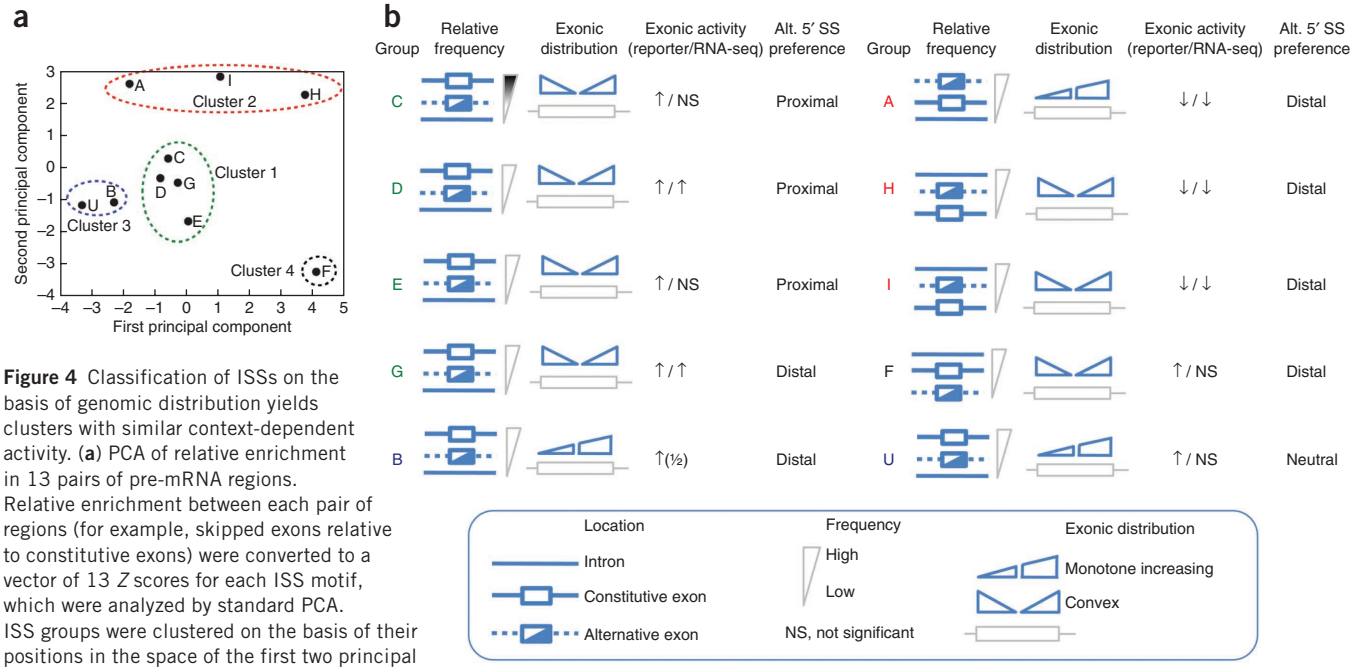

Figure 4 Classification of ISSs on the basis of genomic distribution yields clusters with similar context-dependent activity. (**a**) PCA of relative enrichment in 13 pairs of pre-mRNA regions. Relative enrichment between each pair of regions (for example, skipped exons relative to constitutive exons) were converted to a vector of 13 *Z* scores for each ISS motif, which were analyzed by standard PCA. ISS groups were clustered on the basis of their positions in the space of the first two principal components, which explain 86% of the variance. Both axes were labeled with *Z* scores (unitless). (**b**) Summary of the distributional biases and context-dependent activity of each ISS group. Diagrams representing constitutive exons, skipped exons and introns are shown adjacent to each motif in order from highest (above) to lowest motif abundance, followed by diagrams representing the distribution of the motif from 5′ to 3′ ends of exons. The exonic activities of each ISS motif in different pre-mRNA locations, based on data from **Figure 3** and **Supplementary Table 5**, are listed at right. (The 1/2 indicates that motif B had ESE activity in one of two cell lines tested.) Alternative 5′ SS preference is also shown. The ISS motif labels are color coded according to the PCA clusters shown in **a**.

a pattern typical of ESEs[23]. Cluster 2 motifs were enriched in introns and skipped exons relative to constitutive exons and had variable exonic distributions. Conversely they inhibited exon inclusion and proximal 5′ SS usage from an exonic location (**Figs. 3** and **4b**), similar to ESSs[23]. Motifs of Cluster 3 were exon enriched but had monotone increasing along exons. Unlike canonical ESEs, these ISSs promoted exon inclusion but inhibited (or did not affect) proximal 5′SS usage. Motif F distributed differently from all other motifs (**Fig. 4a**). It was intron enriched like Cluster 2 but promoted exon inclusion like Clusters 1 and 3. Taken together, all six exon-enriched ISS motifs promoted exon inclusion, and all but one of the intron-enriched motifs inhibited exon inclusion, consistent with the RNA-seq results in endogenous exons (**Fig. 4b**). The existence of four clusters of ISS motifs with distinct patterns of activity contrasts with the relatively homogeneous distributions and activities of FAS-ESSs, which suggests that the functions of intronic SREs are more diverse.

Identification of ISS-binding factors

SREs generally function through recruiting *trans* factors to affect splicing[1]. Alternatively, they affect splicing by forming inhibitory structures to block recognition of splice signals[25], by changing the long-range RNA structure[26] or by altering the spliceosome assembly dynamics[9]. Because we didn't detect particular changes of pre-mRNA structures in the original intron-containing ISSs versus control 10-mers by using Mfold (data not shown), and because most FAS-ISS motifs had silencer activity in a heterologous intron (**Fig. 3b**), these ISSs are most likely to function through recruiting inhibitory *trans* factors. To identify associated factors without bias, we incubated a short RNA fragment of ISS exemplars with whole-cell extract, isolated RNA–protein complexes with streptavidin beads and identified proteins specifically bound by ISSs by using MS[27] (**Fig. 5a**).

Our approach was illustrated by using the exemplar from group U (CACACCA). Four major bands were present in the exemplar RNA but were reduced or absent from controls (**Fig. 5b**). The strongest band (band 2) corresponded to hnRNP L, which is known to bind CA-rich elements[13,28]. Band 1 contained two similar proteins (IGF-II mRNA-binding proteins 1 and 3) that are involved in RNA trafficking and translation[29]. Band 3 was identified as Y-box binding protein 1 (YB-1), which can bind to both DNA and RNA. YB-1 was reported to influence alternative splicing of CD44 (ref. 30), but its RNA-binding specificity remained unclear. Band 4 contained an alternative isoform of hnRNP L and the protein SF2P32. SF2P32 was originally identified through co-purification with the splicing factor SRSF1 (ref. 31) but was later localized to the mitochondrial matrix[32], which suggested that it may be a co-purification artifact resulting from its ability to bind other RNA-binding proteins (RBPs). Thus, hnRNP L and YB-1 were identified as putative splicing repressors associated with motif U.

We conducted RNA interference (RNAi) and overexpression of identified proteins in cells transfected with splicing reporters containing cognate ISSs. Knockdown of either hnRNP L or YB-1 derepressed the inclusion of an exon flanked by the group U ISS, whereas RNAi knockdown of an unrelated RBP (GRSF-1) had no effect (**Fig. 5c** and **Supplementary Fig. 4a**). These splicing-inhibitory activities were specific to the CA-rich motif, as the knockdown had no effect on the reporter containing another ISS (group D). Conversely, the overexpression of YB-1, and to a lesser extent hnRNP L, enhanced exon skipping in the reporter containing the CA-rich ISS but not in a similar reporter containing other ISSs (**Fig. 5d** and **Supplementary Fig. 4b**). These data implicated hnRNP L and YB-1 as factors responsible for the ISS activity of motif U. Notably, simultaneous knockdown of hnRNP L and YB-1 had similar effects to knockdown of each

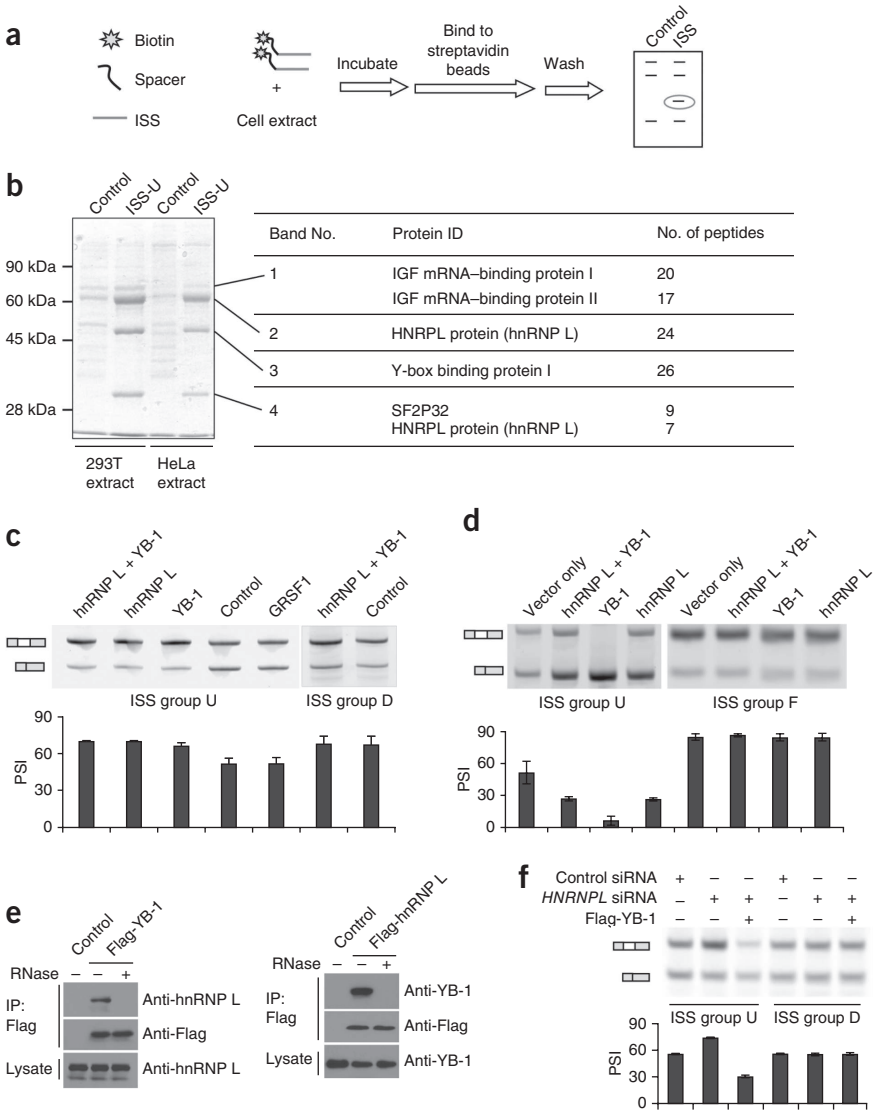

Figure 5 Identification and validation of ISS-associated splicing repressors. (**a**) Diagram of protein-identification procedure. (**b**) Affinity purification using motif U (CA rich) and its mutant as bait incubated with extracts from HEK293T and HeLa cells. Four specific bands were excised for MS, yielding peptides from the proteins listed at right. (**c**) RNAi analysis of the putative splicing-regulatory factors. Derepression of the target exon was observed following knockdown of hnRNP L and/or YB-1. Results for reporters containing another ISS (motif D) as a specificity control are shown at right. The mean and s.d. of PSI values measured by sqRT-PCR from three independent experiments are shown below a representative gel. Error bars, s.d.; $n = 3$. (**d**) RT-PCR experiments and representative gels showing that overexpression of *trans* factors specifically inhibits exon inclusion in reporters with the cognate ISS. Reporters containing ISSs of other groups were used as specificity controls. Error bars, s.d.; $n = 3$. (**e**) Co-immunoprecipitation (IP) in HEK293T cells, using anti-Flag antibody for precipitation and antibodies against endogenous proteins for detection. The cells were lysed in the absence or presence of RNase A. (**f**) RT-PCR experiments and representative gels showing that overexpression of YB-1 reversed the splicing derepression phenotype caused by RNAi knockdown of hnRNP L. The splicing reporters containing cognate ISSs (group U) or control ISSs (group D) were used as specificity controls. Error bars, s.d.; $n = 3$.

protein alone, and overexpression of both factors repressed splicing less than overexpression of YB-1, which suggests that ISS-dependent splicing repression by YB-1 and hnRNP L is not simply additive. The two proteins did not directly interact with each other, as judged by co-immunoprecipitation assay, although the indirect association through RNA bridges was observed (**Fig. 5e**). The overexpression of YB-1 can rescue the RNAi effect of hnRNP L in splicing reporters containing cognate ISSs but not control ISSs (**Fig. 5f**). Therefore the two proteins might compete for binding to the same ISS, with YB-1 repressing splicing more strongly than hnRNP L.

Identification of putative *trans* factors for all ISS motifs

We expanded the affinity purifications to identify *trans* factors for other ISS motifs (groups B to I). Group A was excluded because it matched known binding motifs of PTB and nPTB that inhibit splicing from intronic positions[33–36]. We purified the RNA–protein complexes associated with these ISSs and two additional negative controls (mutant exemplars from groups C and F) and resolved total proteins by SDS-PAGE (**Supplementary Fig. 5**).

Each ISS motif exhibited a distinct protein-binding profile (**Supplementary Fig. 5**), which implied sequence-specific binding of RNAs by associated factors. The protein bands present only in ISS samples were identified by MS. Most of the identified factors were known RBPs or contained putative RNA-binding domains, whereas a few were annotated as DNA-binding or cytoskeletal proteins (**Supplementary Table 7**). Some proteins may have been identified through indirect interaction with other specific RBPs. For example, multiple proteins in U1 snRNP (for example, U1-70k, U1A, SmD1 and SmD2) co-purified with ISS motif E, a purine-rich element recognized by SFRS1 (**Table 1**). Because SRSF1 was detected as a component of the U1 snRNP complex[37], these U1 snRNP core components probably bound to group E ISSs indirectly through interaction with SFRS1.

We also identified some abundant proteins that probably bound RNA in a non–sequence-specific manner (for example, La protein was co-purified with three motifs, and nucleolin was co-purified with motif C). Of the 24 identified putative sequence-specific splicing factors (**Table 1**), 12 belonged to the hnRNP class, and many were previously shown to regulate splicing, including hnRNPs A1, A2, H1, I and L[35,36,38–40]. Several of the 12 non-hnRNP proteins identified have not been previously reported to regulate splicing.

A complex interaction map between ISSs and *trans* factors

Most of the identified RBPs contained multiple RNA-binding domains, and many appeared to have flexible specificity capable of binding multiple distinct motifs. Conversely, individual ISS motifs were recognized by multiple proteins, which suggests that the relationship between splicing factors and ISS motifs may be most compactly represented by an RNA-protein connectivity map (**Fig. 6a**) rather than a table of one-to-one interactions. The ISS motifs from

Table 1 Putative *trans* factors of FAS-ISSs identified through affinity purification.

ISS group	Representative ISS *k*-mer	Associated proteins: Putative splicing factors (protein motifs)[a]	Associated proteins: Other RNA-binding proteins
A	CTCCTC	hnRNP I/PTB[b] (4 RRM + alanine-rich)	
B	TACAGCT	Nono (2 RRM) YB-1 (CSD) hnRNP I/PTB (4 RRM + alanine-rich)	La protein
C	CTTCAG	SFPQ (2 RRM + Gln-Glu-Pro–rich) hnRNP UL1 (glycine-rich + proline-rich) hnRNP H1 (3 RRM + glycine-rich) hnRNP A1 (2 RRM + glycine-rich) hnRNP A2/B1 (2 RRM + glycine-rich)	Nucleolin
D	GAACAG	GRSF-1 (3 RRM + alanine-rich) hnRNP F (3 RRM) SFRS1 (2 RRM + glycine-rich) RBM45/CELF-3 homolog (3 RRM)	CSDE1
E	CAAAGGA	SFRS1 (2 RRM + glycine-rich) hnRNP Q (3 RRM) hnRNP A1 (2 RRM + glycine-rich) hnRNP A2/B1 (2 RRM + glycine-rich) hnRNP H1 (3 RRM + glycine-rich) hnRNP F (3 RRM) FUS	Several U1 snRNP core proteins (for example, U1-70k, U1 snRNP protein A and C, SmD1, SmD2)
F	AGATATT	KHSRP (4 KH + glycine-rich) DAZAP1 (2 RRM + proline-rich) hnRNP D0 (2 RRM + glycine-rich) hnRNP A0 (2 RRM + glycine-rich) hnRNP A1 (2 RRM + glycine-rich) hnRNP A2/B1 (2 RRM + glycine-rich) hnRNP A3 (2 RRM + glycine-rich) CIRBP (1 RRM + glycine-rich) G3BP-1 (1 RRM + glycine-rich) G3BP-2 (1 RRM + glycine-rich)	La protein
G	ACATGA	G3BP1 (1 RRM + glycine-rich) hnRNP L (3 RRM + glycine-rich + proline-rich)	GTF2I
H	AATTTA	DAZAP1 (2 RRM + proline-rich) hnRNP A0 (2 RRM + glycine-rich) hnRNP D0 (2 RRM + glycine-rich) hnRNP DL (2 RRM + glycine-rich) hnRNP A1 (2 RRM + glycine-rich) hnRNP A2/B1 (2 RRM + glycine-rich) hnRNP A3 (2 RRM + glycine-rich)	La protein
I	AGTAGG	DAZAP1 (2 RRM + proline-rich) hnRNP A1 (2 RRM + glycine-rich) hnRNP A2/B1 (2 RRM + glycine-rich) hnRNP A3 (2 RRM + glycine-rich) hnRNP H1 (3 RRM + glycine-rich) hnRNP D0 (2 RRM + glycine-rich) hnRNP DL (2 RRM + glycine-rich) FUS	DHX36
U	CACACCA	hnRNP L (3 RRM + glycine-rich + proline-rich) YB-1 (CSD)	IGF mRNA binding proteins I, II

[a]Characteristic RNA-binding or splicing-regulatory protein motifs are indicated (RRM: RNA-recognition motif; CSD: cold-shock domain; Glycine-rich: glycine-rich motif; etc.).
[b]Based on the literature rather than affinity purification and MS.

the same cluster (**Fig. 4a**) tended to bind similar proteins, whereas a few specific proteins (for example, members of the hnRNP A, D and H families) recognized motifs belonging to two different clusters.

The complexity of the RNA-protein connectivity map raised questions about the extent of direct binding. An extreme case was presented by ISS motifs F, H and I, which bound a highly overlapping set of factors despite considerable sequence diversity. These factors included hnRNPs A0, A1, A2, A3, D0, DL and DAZAP1, with most containing two N-terminal RRM domains and a C-terminal glycine-rich motif. We used surface plasmon resonance (SPR) to determine direct interactions between the ISS exemplar and single purified proteins. (hnRNPs A0, A1, A2, D0 and DL were tested; hnRNP A3 failed to express.) Notably, the ISS motifs F, H and I could all directly bind multiple proteins in the hnRNP A family (A0, A1, A2) and bind hnRNPs D0 and DL (**Fig. 6b** and **Supplementary Fig. 6**), although each protein exhibited somewhat different affinities for individual exemplars. For example, hnRNPs A0 and D0 preferentially bound the A- and U-rich motif H, whereas hnRNP DL showed preferential binding to motif I. Nevertheless, the binding affinities to distinct motifs were fairly similar, with apparent K_d values in the range of 10–100 nM for each pair (R.C. and Z.W., unpublished data), which suggested that the RNA-protein connections detected in our affinity purification typically reflected direct interactions.

Glycine-rich and RS domains inhibit splicing from introns

The hnRNP proteins represented roughly half of the putative *trans* factors identified above, with many containing regions of repetitive sequence such as glycine-rich domains (for example, the hnRNP A1 family). In addition, we found one SR protein (SRSF1) binding to ISSs of groups D and E. It is well established that SRSF1 activates splicing by binding ESEs, and that hnRNP A1 inhibits splicing by recognizing ESSs (reviewed in refs. 1,41). Both of these prototypical splicing factors have modular structures, recognizing RNA targets through their N-terminal RRM domains and affecting splicing through C-terminal functional domains (RS or glycine rich)[19,42]. However, other mechanisms may exist, as the RS domains have been shown to recognize RNA near branchpoints to promote splicing[43]. When tethered to an exon, most RS domains can activate exon splicing[42,44,45] whereas the glycine-rich domain of hnRNP A1 can suppress splicing[19,45]. Because both hnRNP A1 and SRSF1 were identified as putative ISS-associated splicing repressors, we speculated that their RS and glycine-rich domains could inhibit splicing from an intronic location.

We tested this hypothesis by fusing an RS domain (from SRSF1) or glycine-rich domain (from hnRNP A1) with modified pumilio (PUF) domains that recognized specific 8-nt RNA sequences[46]. Two PUF domains, PUF(3-2), specifically recognizing UGUAUGUA, and PUF(6-2), recognizing UUGAUAUA[45,46], were combined with the RS or glycine-rich domain (from SRSF1 or hnRNP A1, respectively) to yield four fusion proteins that were co-transfected with the splicing reporter–containing cognate target sites (**Fig. 6c**). As predicted, both the RS and glycine-rich domains inhibited exon inclusion from downstream of the 5′ SS, whereas the binding of the PUF domain itself had no detectable effect (**Fig. 6c**, left). Splicing inhibition was detected only in the cognate PUF-RNA pairs but not between noncognate pairs.

To examine whether the splicing inhibition at intronic locations is a general activity of the RS and glycine-rich domains, additional proteins were generated by fusing PUF(6-2) with the RS domains of SFRS2, SFRS5 and SFRS7, the glycine-rich domains of hnRNPs A1, A2 and A3, a short peptide of $(RS)_6$ or a 19–amino acid glycine-rich sequence (**Fig. 6c**, right). Notably, compared to the controls (lanes 13 and 14), all of these fusions strongly inhibited inclusion of the exon

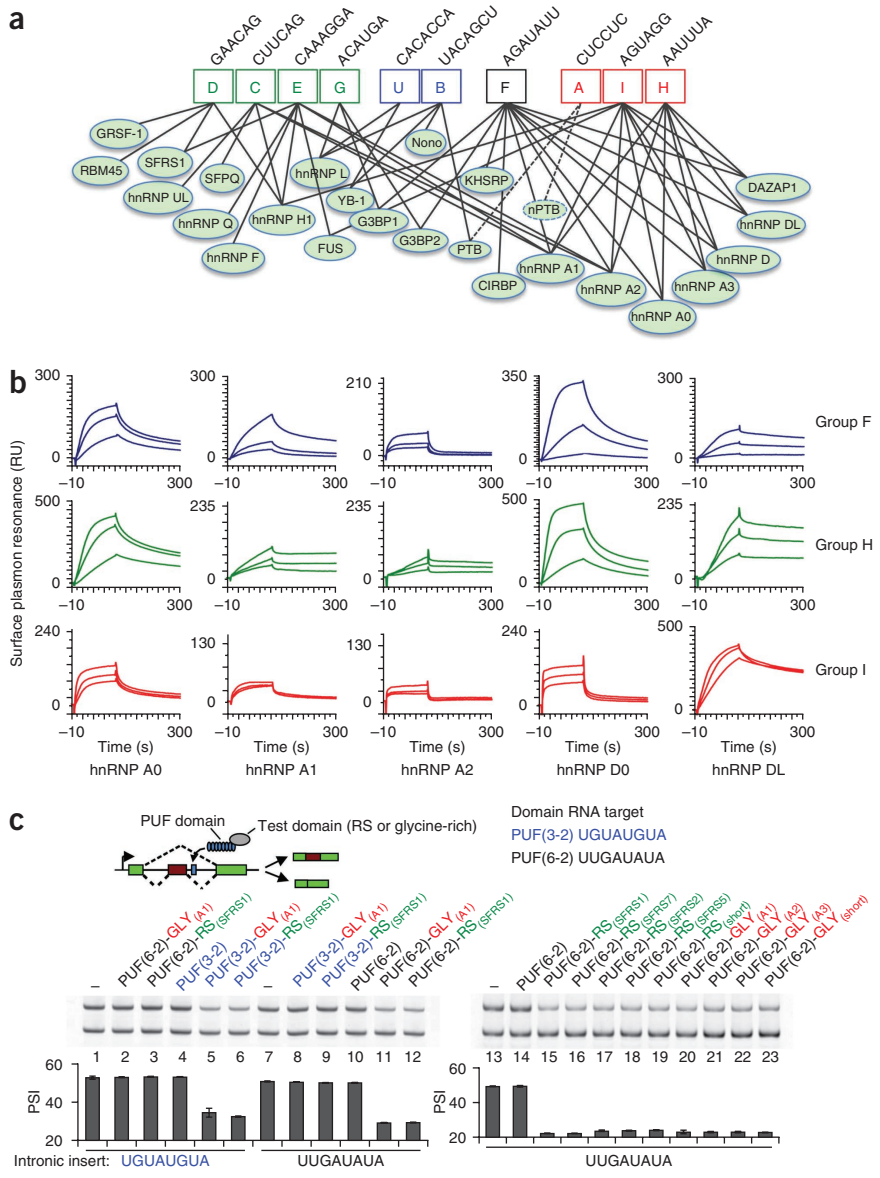

Figure 6 ISSs are recognized by a complex overlapping network of factors. (**a**) Interactions between putative splicing factors and all ISS groups are represented by solid lines (dashed lines represent interactions inferred from the literature), yielding a many-to-many connectivity map encompassing all ten ISSs and their cognate factors. (**b**) SPR presenting direct RNA-protein binding between ISS exemplars and proteins of the hnRNP A and D families. Three different protein concentrations (50, 100 and 200 nM, from bottom to top in each graph) were injected at a flow rate of 50 µl min^{-1} for 120 s. For hnRNP A1, 20, 40 and 80 nM were used to compensate for the strong binding signal. (**c**) RT-PCR results showing that tethering RS and glycine-rich domains to a downstream intronic location can repress splicing. Fusion proteins consisting of a PUF domain fused to an RS or glycine-rich domain were assayed for their effects on splicing. Two target 8-mers, UGUAUGUA and UUGAUAUA, recognized respectively by PUF(3-2) and PUF(6-2), were inserted into the splicing reporter pZW2C and co-transfected with the indicated fusion proteins into HEK293T cells. All combinations of fusion proteins and targets were tested, with matching PUF-target pairs indicated by the same colors (blue or black). The PUF domains alone and empty vector served as negative controls. Right panel, fusions between PUF(6-2) and a variety of RS and glycine-rich domains were tested. Error bars, s.d.; $n = 3$.

containing the cognate target at downstream introns, which suggests that a polypeptide containing RS repeats or a short glycine-rich fragment is sufficient to inhibit splicing when recruited downstream of an exon.

DISCUSSION

We adapted a cell-based screening approach developed previously for comprehensive identification of ESSs[7] to identify ten distinct ISS motifs. The similarities in the reporters and other aspects of the two screens make comparison of their results worthwhile for understanding intrinsic differences between exonic and intronic silencers. All the newly identified ISS motifs had activity in a second cell type, and most functioned in a heterologous intron, which supported their functional modularity and portability as observed for ESSs. However, in some respects, ISSs behaved very differently from ESSs, exhibiting far greater diversity and complexity of function.

The first clue to these differences was the greater sequence complexity of the ISSs. Whereas the ESSs were G or T rich (74% G + T), the sequence bias in the ISSs was less pronounced (65% A + T), and a larger number of ISS motifs were identified (ten ISSs versus seven ESSs). All ESS motifs had similar patterns of distributional bias[7,23] and were able to repress intron-proximal 5′ SSs and 3′ SSs, which demonstrated remarkable consistency of activity[23]. By contrast, the ISS motifs had diverse patterns of distributional bias, with different motifs being most enriched in either constitutive exons, skipped exons or introns and having either convex or increasing frequency along exons. Placement of the ISSs into exonic locations revealed a diversity of functions, including activation or repression of exon inclusion and inhibition or enhancement of proximal 5′ SS usage. Such context-dependent activities matched the genomic distribution patterns, which emphasizes the need for subclassification of ISSs. We used a PCA approach to define four clusters of functionally different ISS motifs on the basis of distinct genomic distribution patterns, thus the motifs within the same cluster may still have quite different consensus sequences. The associated *trans* factors can act through multiple motifs and sometimes even across multiple PCA classes, which suggests that this classification is likely to be improved by taking *trans* factors into account.

It has been noted that the sequence specificity of most RBPs is relatively low, causing high noise in affinity purifications of proteins associated with ISSs. Our method had a number of improvements to minimize such noise, including use of a short RNA (21–23 nt) with three copies of ISS exemplars to increase affinity, use of a long spacer separating biotin from RNA to increase the RNA accessibility and incubation in a large volume (that is, low protein concentration) for an extended time (>4 h) to help reach binding equilibrium. Many of the identified proteins probably regulate splicing, as suggested in this study and by other labs, but some may also function in other

RNA-processing pathways. For example, G3BP-1 was reported to mediate RNA translation and degradation[47], and GRSF-1 was shown to affect translation[48]. Even some canonical splicing factors like SRSF1 or hnRNP A1 were reported to control mRNA translation or degradation[49,50].

Most factors bound multiple ISS motifs, and each motif was recognized by multiple factors, which supports a complex network of protein-RNA interactions responsible for ISS activity. Such overlapping specificities may enable a variety of regulatory relationships. For example, multiple factors with similar regulatory activities may bind the same motif to confer functional redundancy, whereas antagonism would be expected in cases where one factor can displace another and confer opposite splicing-regulatory function[22]. Other relationships could provide subtler tuning of splicing levels, for example, displacement of a repressive factor by another repressive factor with stronger or weaker activity or recruitment of distinct splicing-regulatory complexes to a given site in the pre-mRNA. The relationship between PTB and nPTB may illustrate some of these regulatory relationships. For example, in HeLa cells nPTB can compensate for PTB depletion[34], whereas in neural development replacement of PTB by nPTB may alter the regulation of a substantial alternative-splicing program[51].

A variety of regulatory relationships are possible in the examples identified here. For example, both hnRNP L and YB-1 bound ISS motif U. Although overexpression or depletion of either factor was sufficient to repress or derepress splicing, respectively, their relationship is not one of simple redundancy. Instead they may compete for the same site, with hnRNP L repressing splicing less strongly than YB-1 (**Fig. 5c**). As another example, ISS motifs D and E were both recognized by SFRS1 and hnRNP H1, which typically have the opposite pattern of activity[38,52–54]. Our data suggested that these factors might antagonistically regulate introns containing motifs D or E—in HEK293T cells, binding of motifs D and E by SRSF1 or other intron-repressive factors likely predominates, as these motifs have ISS activity. The less-characterized factors RBM45 and GRSF-1 also bind motif D, potentially adding additional regulatory complexity through this motif.

Our analyses using fusions of candidate splicing-regulatory domains indicated that the requirements for splicing repression were unexpectedly minimal. Tethering of an RS or glycine-rich domain from any of various SR or hnRNP proteins (or even a short RS repeat or glycine-rich peptide) downstream of an exon was sufficient to repress splicing (**Fig. 6c**). This observation raises the possibility that many other proteins with similar domains also control splicing when binding to pre-mRNA. An analysis of the human proteome identified a total of 24 human proteins that contain both a short glycine-rich segment and at least one RRM domain, representing an eight-fold enrichment above the frequency expected from occurrence of these domains individually (A.R., unpublished data).

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Hui (Shanghai Institute of Biological Science, Shanghai, China) and A. Willis (University of Nottingham, UK) for providing expression constructs of *trans* factors and B. Graveley (University of Connecticut Health Center, Farmington, Connecticut, USA) for constructs containing RS domains. We thank T. Nilsen and A. Berglund for critical reading of manuscripts and Z. Dominski and B. Marzluff for helping in RNA affinity purifications. This work was supported by an American Heart Association grant (0865329E) and US National Institutes of Health grant (R01CA158283) to Z.W. and (2-R01-GM085319) to C.B.B.

AUTHOR CONTRIBUTIONS

Y.W., C.B.B. and Z.W. designed the research. Y.W., Z.W., J.Z., K.L. and R.C. performed the experiments. M.M., X.X. and A.R. developed computational methods to analyze the data. Y.W., C.B.B. and Z.W. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nsmb.2459.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Wang, Z. & Burge, C.B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802–813 (2008).
2. Matlin, A.J., Clark, F. & Smith, C.W. Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6**, 386–398 (2005).
3. Blencowe, B.J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37–47 (2006).
4. Liu, H.X., Zhang, M. & Krainer, A.R. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* **12**, 1998–2012 (1998).
5. Fairbrother, W.G., Yeh, R.F., Sharp, P.A. & Burge, C.B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).
6. Zhang, X.H. & Chasin, L.A. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.* **18**, 1241–1250 (2004).
7. Wang, Z. *et al.* Systematic identification and analysis of exonic splicing silencers. *Cell* **119**, 831–845 (2004).
8. Goren, A. *et al.* Comparative analysis identifies exonic splicing regulatory sequences–the complex definition of enhancers and silencers. *Mol. Cell* **22**, 769–781 (2006).
9. Yu, Y. *et al.* Dynamic regulation of alternative splicing by silencers that modulate 5′ splice site competition. *Cell* **135**, 1224–1236 (2008).
10. Culler, S.J., Hoff, K.G., Voelker, R.B., Berglund, J.A. & Smolke, C.D. Functional selection and systematic analysis of intronic splicing elements identifies active sequence motifs and associated splicing factors. *Nucleic Acids Res.* **38**, 5152–5165 (2010).
11. Sharma, S., Kohlstaedt, L.A., Damianov, A., Rio, D.C. & Black, D.L. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat. Struct. Mol. Biol.* **15**, 183–191 (2008).
12. Kashima, T., Rao, N. & Manley, J.L. An intronic element contributes to splicing repression in spinal muscular atrophy. *Proc. Natl. Acad. Sci. USA* **104**, 3426–3431 (2007).
13. Hui, J. *et al.* Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J.* **24**, 1988–1998 (2005).
14. Blanchette, M. & Chabot, B. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* **18**, 1939–1952 (1999).
15. Kanopka, A., Muhlemann, O. & Akusjarvi, G. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* **381**, 535–538 (1996).
16. Ibrahim, E.C., Schaal, T.D., Hertel, K.J., Reed, R. & Maniatis, T. Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proc. Natl. Acad. Sci. USA* **102**, 5002–5007 (2005).
17. Shen, M. & Mattox, W. Activation and repression functions of an SR splicing regulator depend on exonic versus intronic-binding position. *Nucleic Acids Res.* **40**, 428–437 (2012).
18. McNally, L.M. & McNally, M.T. SR protein splicing factors interact with the Rous sarcoma virus negative regulator of splicing element. *J. Virol.* **70**, 1163–1172 (1996).
19. Del Gatto-Konczak, F., Olive, M., Gesnel, M.C. & Breathnach, R. hnRNP A1 recruited to an exon *in vivo* can function as an exon splicing silencer. *Mol. Cell Biol.* **19**, 251–260 (1999).
20. Hua, Y., Vickers, T.A., Okunola, H.L., Bennett, C.F. & Krainer, A.R. Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *Am. J. Hum. Genet.* **82**, 834–848 (2008).
21. Tange, T.O., Damgaard, C.K., Guth, S., Valcarcel, J. & Kjems, J. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J.* **20**, 5748–5758 (2001).
22. Wang, Y., Ma, M., Xiao, X. & Wang, Z. Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. *Nat. Struct. Mol. Biol.* **19**, 1044–1052 (2012).
23. 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).
24. Lim, K.H., Ferraris, L., Filloux, M.E., Raphael, B.J. & Fairbrother, W.G. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc. Natl. Acad. Sci. USA* **108**, 11093–11098 (2011).
25. Huang, C. *et al.* A structured RNA in HBV PRE represses alternative splicing in a sequence-independent and position-dependent manner. *FEBS J.* **278**, 1533–1546 (2011).
26. Pervouchine, D.D. *et al.* Evidence for widespread association of mammalian splicing and conserved long-range RNA structures. *RNA* **18**, 1–15 (2012).

27. Dominski, Z., Yang, X.C., Kaygun, H., Dadlez, M. & Marzluff, W.F. A 3′ exonuclease that specifically interacts with the 3′ end of histone mRNA. *Mol. Cell* **12**, 295–305 (2003).
28. Rothrock, C.R., House, A.E. & Lynch, K.W. HnRNP L represses exon splicing via a regulated exonic splicing silencer. *EMBO J.* **24**, 2792–2802 (2005).
29. Nielsen, F.C., Nielsen, J. & Christiansen, J. A family of IGF-II mRNA binding proteins (IMP) involved in RNA trafficking. *Scand. J. Clin. Lab. Invest. Suppl.* **234**, 93–99 (2001).
30. Allemand, E., Hastings, M.L., Murray, M.V., Myers, M.P. & Krainer, A.R. Alternative splicing regulation by interaction of phosphatase PP2Cgamma with nucleic acid-binding protein YB-1. *Nat. Struct. Mol. Biol.* **14**, 630–638 (2007).
31. Krainer, A.R., Conway, G.C. & Kozak, D. Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev.* **4**, 1158–1171 (1990).
32. Muta, T., Kang, D., Kitajima, S., Fujiwara, T. & Hamasaki, N. p32 protein, a splicing factor 2-associated protein, is localized in mitochondrial matrix and is functionally important in maintaining oxidative phosphorylation. *J. Biol. Chem.* **272**, 24363–24370 (1997).
33. Oberstrass, F.C. *et al.* Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**, 2054–2057 (2005).
34. Spellman, R., Llorian, M. & Smith, C.W. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol. Cell* **27**, 420–434 (2007).
35. Jin, W., Bruno, I.G., Xie, T.X., Sanger, L.J. & Cote, G.J. Polypyrimidine tract-binding protein down-regulates fibroblast growth factor receptor 1 α-exon inclusion. *Cancer Res.* **63**, 6154–6157 (2003).
36. Côté, J., Dupuis, S. & Wu, J.Y. Polypyrimidine track-binding protein binding downstream of caspase-2 alternative exon 9 represses its inclusion. *J. Biol. Chem.* **276**, 8535–8543 (2001).
37. Das, R. *et al.* SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* **26**, 867–881 (2007).
38. Chou, M.Y., Rooke, N., Turck, C.W. & Black, D.L. hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol. Cell Biol.* **19**, 69–77 (1999).
39. Kashima, T., Rao, N., David, C.J. & Manley, J.L. hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. *Hum. Mol. Genet.* **16**, 3149–3159 (2007).
40. Hui, J., Stangl, K., Lane, W.S. & Bindereif, A. HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat. Struct. Biol.* **10**, 33–37 (2003).
41. Chen, M. & Manley, J.L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* **10**, 741–754 (2009).
42. Graveley, B.R. & Maniatis, T. Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol. Cell* **1**, 765–771 (1998).
43. Shen, H., Kan, J.L. & Green, M.R. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol. Cell* **13**, 367–376 (2004).
44. Graveley, B.R., Hertel, K.J. & Maniatis, T. A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *EMBO J.* **17**, 6747–6756 (1998).
45. Wang, Y., Cheong, C.G., Hall, T.M. & Wang, Z. Engineering splicing factors with designed specificities. *Nat. Methods* **6**, 825–830 (2009).
46. Cheong, C.G. & Hall, T.M. Engineering RNA sequence specificity of Pumilio repeats. *Proc. Natl. Acad. Sci. USA* **103**, 13635–13639 (2006).
47. Tourrière, H. *et al.* The RasGAP-associated endoribonuclease G3BP assembles stress granules. *J. Cell Biol.* **160**, 823–831 (2003).
48. Ufer, C. *et al.* Translational regulation of glutathione peroxidase 4 expression through guanine-rich sequence-binding factor 1 is essential for embryonic brain development. *Genes Dev.* **22**, 1838–1850 (2008).
49. Michlewski, G., Guil, S., Semple, C.A. & Caceres, J.F. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol. Cell* **32**, 383–393 (2008).
50. Guil, S., Long, J.C. & Caceres, J.F. hnRNP A1 relocalization to the stress granules reflects a role in the stress response. *Mol. Cell Biol.* **26**, 5744–5758 (2006).
51. Boutz, P.L. *et al.* A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* **21**, 1636–1652 (2007).
52. Caputi, M. & Zahler, A.M. Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H′/F/2H9 family. *J. Biol. Chem.* **276**, 43850–43859 (2001).
53. Schaub, M.C., Lopez, S.R. & Caputi, M. Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. *J. Biol. Chem.* **282**, 13617–13626 (2007).
54. Chen, C.D., Kobayashi, R. & Helfman, D.M. Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev.* **13**, 593–606 (1999).
55. Xiao, X., Wang, Z., Jang, M. & Burge, C.B. Coevolutionary networks of splicing cis-regulatory elements. *Proc. Natl. Acad. Sci. USA* **104**, 18583–18588 (2007).

ONLINE METHODS

Splicing reporters. All splicing reporters were constructed from a backbone vector, pZW1, which contains a multiple cloning site between two GFP exons[7]. The random-sequence library was generated from a fold-back primer extended with Klenow and digested and ligated into pZW11 that contains a multiple cloning site downstream of a constitutive exon. To test ISSs in a heterologous context, a new reporter (pZW2C) was constructed by inserting exon 2 of the *DHFR* gene and part of its flanking introns between the two GFP exons. The reporter with the competing 5′ SS and the modular splicing reporter was described previously[23,55,56]. Additional details are described in **Supplementary Note**.

Cell culture and transfections. HEK293 cells and HeLa cells were cultured with DMEM supplemented with 10% FBS. Transfections were carried out with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions.

For stable transfection, Flp-In-293 (Invitrogen) cells were cotransfected with a 10-fold excess of pOG44 plasmid. To select stable transfectants, the cells were expanded by a 1:4 dilution one day after transfection, grown for one more day, and then hygromycin was added to a final concentration of 100 μg/ml. For each transfection in the FAS screen, pZW11 vector (1.6 μg) containing random library was transfected into Flp-In-293 cells in a 15-cm tissue culture dish. After selection for 10 d, the positive clones were trypsin digested, pooled and sorted by using a Cytomation MoFlo high-speed sorter into 96-well plates at one cell per well. The survival colonies were visible after 10 d and were checked for green fluorescence. Total DNA from GFP-positive cells was used for PCR and sequencing. The ISS sequences were further analyzed as previously described[7].

RNA purification and semiquantitative RT-PCR. Total RNA was isolated from transfected cells with TRIzol reagent (Invitrogen) followed by DNase I (Invitrogen) treatment. Total RNA (2 μg) was then reverse transcribed with SuperScript III (Invitrogen) with poly(T) primer or gene-specific primer (for splicing reporter), and one-tenth of the RT product was used as the template for PCR amplification (25 cycles of amplification, with trace amount of Cy5-dCTP in addition to nonfluorescent dNTPs). The resulting gels were scanned with a Typhoon 8600 Imager (GE Healthcare) and analyzed with Image Quant 5.2 software (Molecular Dynamics/GE Healthcare). All experiments were repeated at least three times. The primers used to amplify GFP-based minigene reporters were 5′-AGTGCTTCAGCCGCTACCC-3′ for GFP exon 1 and 5′-GTTGTACTCCAGCTTGTGCC-3′ for exon 3.

RNA affinity purification and identification of *trans* factors. The RNA affinity-purification method was modified from a previously described protocol[27]. A short (~21-nt) RNA fragment containing three copies of ISS exemplar was synthesized with 5′ biotin followed by two 18-carbon spacers (Ambion/Invitrogen). For each RNA sample, approximately 2.5×10^8 HeLa cells (NCCC, Minneapolis) were harvested at ~95% confluence and resuspended with 2.5-ml ice-cold buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl). The cells were mixed with 2.5ml 2× lysis buffer (50mM Tris-HCl, pH 8.0, 150mM NaCl, 15 mM NaN_3, 1% (v/v) NP-40, 2 mM DTT, 2 mM PMSF, 2× protease-inhibitor mix) and lysed for 5 min, then centrifuged at 12,000*g* for 20 min at 4 °C. Subsequently, 0.75 nmol biotinylated RNA with two 18-atom spacers (Dharmacon) were added and incubated for 2 h at 4 °C. We then added 50 μl streptavidin-agarose beads (Sigma) into the mixture and incubated for 2 h at 4 °C with slow rotation. The beads were rinsed three times with 4 ml lysis buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 15 mM NaN_3, 0.5% NP-40, 1 mM DTT, 1 mM PMSF, 1× protease-inhibitor mix), resuspended in 40 μl final volume and mixed with 10 μl 5× SDS loading buffer. The proteins were then separated with a 10% SDS-PAGE gel and stained with Coomassie blue. The gels were kept in 3% acetic acid for further MS analysis. Bands specific to the ISS RNA but not to controls were excised, and the gel slices containing the candidate proteins were digested with trypsin, following standard protocols. Peptides eluted from the gel were analyzed by ESI-MS/MS on a Q-Tof (Micromass) mass spectrometer. The in-gel digestion and MS were carried out by the UNC Proteomics Center.

Analysis of correlation between ISSs and tissue-specific splicing data. The human-tissue sequence data (Body Map 2.0) were obtained from the ENA archive with accession number ERP000546 (available at http://www.ebi.ac.uk/ena/data/view/ERP000546). The 50-bp paired-end reads were mapped by using Tophat, and PSI values of alternative exons were calculated by using MISO[57]. Decoy ISS-motif sets were generated as random permutations of each of the ISS cluster members, and the known SREs were avoided. ISS sequences and decoys were identified in the skipped-exon body and in the 200 bases following the skipped exon. The significance of the overall ISS effect was calculated as the median of comparisons between PSI values for events with an ISS and those with a decoy. Skipped-exon PSI values were taken for each event from the tissue where their gene is most highly expressed. Significance of tissue specificity was assessed by a bootstrapping comparison.

Knockdown and overexpression of putative *trans* factors. The siRNAs were purchased from Dharmacon (On-target SMARTpool, with scrambled dsRNA controls), and transfections were conducted with Lipofectamine 2000 (Invitrogen) in cells grown in 24-well plates. After 48 h of siRNA transfection, we transfected with 0.2 μg splicing reporter and harvested the cells 24 h after the second transfection for further analyses. In overexpression experiments, we cotransfected 0.8 μg expression vector and 0.2 μg splicing reporters. The cells were harvested after 72 h for further analyses. For the rescue experiment, HEK293T cells were transfected with target siRNA or control siRNA. After 48 h, we co-transfected Flag-tagged YB1 and ISS splicing reporters and collected the cells 24 h after the second transfection for further analysis.

Coimmunoprecipitation of hnRNP L and YB1. Flag-tagged splicing factors were transfected into HEK293T cells as described above. After 48 h, the transfected cells were lysed in buffer containing 50 mM HEPES, 150 mN NaCl, 1mM EDTA, 1% (w/v) CHAPS and protease-inhibitor cocktail (Sigma; with or without 50 μg/ml of RNaseA) at room temperature for 10 min. The M2 anti-Flag–agarose resin (40 μl) was prepared as described by manufacturer instructions and incubated with 1,000 μl of cell-lysate supernatant with gentle agitation at 4 °C overnight. The IP samples were spun down at 8,000 r.p.m. for 30 s and washed with buffer (50 mM HEPES, 150 mM NaCl, 0.1% Triton X-100, 10% glycerol, pH 7.5) three times. Then the proteins were eluted with Flag peptide (200ng/μl) and transferred into a new tube for further analysis.

Exon and intron data sets. The exon and intron data sets were generated by using similar filters as described earlier[23]. The A3Es and A5Es were further filtered by requiring that the longer isoform differ from the shorter isoform by at least 6 bases, and skipped exons were required to be at least 6 bases in length. The following numbers of orthologous human-mouse exon pairs were obtained by using our criteria: 1,232 A5Es, 1,408 A3Es, 2,964 skipped exons and 44,368 constitutive exons. **Supplementary Note** provides details in computational analyses.

Positional frequency of ISSs in different pre-mRNA locations. The ISS *k*-mers in each group were analyzed separately for the positional frequency along the pre-mRNAs of different exon-intron data sets. The first and last 60 nt of exons, together with the 200 nt flanking the introns were used in the analyses. The positional frequency was defined as the number of transcripts that contain a certain set of ISS *k*-mers divided by the total number of transcripts in the position under consideration. To smooth the data, the average frequency in a 10-nt window was plotted.

Biacore SPR analysis. The hnRNPs A0, A1, A2, D0 and DL were cloned into the pT7HtB expression vector with His_6 tag and expressed in the *E. coli* BL21(DE3) strain. The proteins were purified to >90% purity with the His GraviTrap Kit (GE Health Care), following the manufacturer's instructions. The purified protein was buffer exchanged to remove excess imidazole and stored at 4 °C in buffer (20mM HEPES, pH 7.4, 0.5M NaCl, 10% glycerol and 2mM DTT).

The 5′-biotinylated RNAs were synthesized by Dharmacon and immobilized at 150 fmol per flow cell on a Biacore streptavidin-coated chip (Sensor Chip SA, GE Health Care). A control flow cell was loaded with a 21-nt CA-rich control sequence, and all the SLR responses were normalized to the controls. All experiments were done on the Biacore 2000 platform at least twice for every concentration, and the data were fitted with BIAevaluation software. The protein was diluted to its final concentration by HBS-EP buffer (GE Health Care, Cat No. BR-1001-88) before experiments to avoid bulk shift. Data acquisition was performed by using Kinject mode at a flow rate of 50 μl/min in the same running buffer. The surface was regenerated by using 100-μl injection of 2M NaCl followed by 200 μl of running buffer.

56. Xiao, X. *et al.* Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat. Struct. Mol. Biol.* **16**, 1094–1100 (2009).
57. Katz, Y., Wang, E.T., Airoldi, E.M. & Burge, C.B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).