



HHS Public Access

Author manuscript

Nat Struct Mol Biol. Author manuscript; available in PMC 2013 July 01.

Published in final edited form as:

Nat Struct Mol Biol. 2013 January ; 20(1): 36–45. doi:10.1038/nsmb.2459.

A Complex Network of Factors with Overlapping Affinities Repress Splicing through Intronic Elements

Yang Wang¹, Xinshu Xiao^{2,3}, Jianming Zhang^{2,4}, Rajarshi Choudhury¹, Alex Robertson², Kai Li¹, Meng Ma^{1,5}, Christopher B. Burge^{2,6}, and Zefeng Wang^{1,6}

¹Department of Pharmacology, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599.

²Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

³Department of Integrative Biology and Physiology, University of California, Los Angeles, CA 90095.

⁴Department of Biological Chemistry & Molecular Pharmacology, Harvard Medical School, Boston, MA 02115.

⁵Department of Computer Science and Technology, Anhui University, China 23061.

Abstract

To better understand splicing regulation, we used a cell-based screen to identify ten diverse motifs that inhibit splicing from intron. Each motif was 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 based on their distributions yielded clusters with distinct patterns of context-dependent activity. Candidate regulatory factors associated with each motif were identified, recovering 24 known and novel splicing regulators. Specific domains in selected factors were sufficient to confer ISS activity. Many factors bound multiple distinct motifs with similar affinity, and all motifs were recognized by multiple factors, revealing a complex, overlapping network of protein:RNA interactions. This arrangement enables individual *cis*-element to function differently in distinct cellular contexts depending on the spectrum of regulatory factors present.

Keywords

splicing regulation; splicing factors; intronic splicing silencers; RNA binding protein; context dependent activity

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

⁶To whom correspondence should be addressed: Zefeng Wang, Phone: 919-966-0131; zefeng@med.unc.edu, Christopher B. Burge: cburge@mit.edu.

Author contribution: 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.

Introduction

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) play 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^{1,2}. In mammals, splicing of each gene is controlled by multiple SREs and corresponding splicing factors, whose combinatorial actions determine the final splicing outcome³.

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 SREs⁴⁻⁹. 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 likely recognized by a greater diversity of proteins, including proteins of the hnRNP class as well as others², 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^{11,12}, as well as CA-rich sequences bound by hnRNP L¹³. PTB can bind to ISSs and inhibit the transition from an exon-definition complex to an intron-spanning spliceosome¹¹, whereas hnRNP A1 may block recognition of splice sites by cooperative propagation along the pre-mRNA from primary binding sites introns¹² or by other mechanisms¹⁴. Overall, the sequence composition and mechanisms of action of known ISSs are quite diverse, suggesting 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'SS). Analysis of genomic distributions of these motifs identified four classes of motifs, each having similar patterns of context-dependent splicing regulatory activity. Identification of *trans*-acting splicing factors for each motif revealed a densely connected network, suggesting 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 using a cell-based screen

To unbiasedly identify ISS sequences, we developed a cell-based screen called FAS-ISS (fluorescence activated screen of ISS). We constructed a reporter gene (pZW11) whose exons 1 and 3 form a complete mRNA encoding the GFP, and 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 ISS containing binding sites of hnRNP I (Test 1) or hnRNP A1 (Test 2) in the downstream intron of exon 2 caused partial skipping of this exon. Following 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). We confirmed that splicing patterns were consistent with the GFP expression via RT-PCR (Fig. 1b).

We inserted a random pool of decanucleotides (10mers) at 17 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 *E. coli* to achieve ~2-fold coverage of possible DNA 10mers. To evaluate the library quality, we randomly picked 24 colonies for plasmid extraction and sequencing, and found that all had 10 base inserts with no bias at any position. We further transfected the resulting library into 293-FlpIn 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 293-FlpIn cell line. Typically about one in 1000–5000 cells was GFP positive (Supplementary Fig. 1a). These GFP positive cells were sorted by FACS, 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 10mers were identified, 102 of which were unique (In a few cases, 9mers or 11mers were identified, see Supplementary Table 2). We identified nine 10mers at least twice in independent transfections and 19 pairs differed by a single nucleotide, suggesting that the screen approached saturation. These ISS 10mers could be readily clustered by sequence similarity (Fig. 1d). We inserted 16 of the recovered ISS 10mers (asterisks in Fig. 1d) spanning all clusters back into the reporter pZW11 to assess their activity with a transient transfection assay⁷. 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 10mers 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 over-representation in the recovered 10mer set as a criterion⁷. Forty-one 5mers occurred at least five times in the recovered sequences, about 3-fold higher than expected ($p < 10^{-4}$, based on 10,000 samplings of random 10mers). Using similar criteria, we identified over-represented 6mers and 7mers, and these k -mers ($k = 5, 6, 7$) were then clustered by sequence similarity and multiply aligned 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}, confirming 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}. Considering individual motifs, FAS-ISSs 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 over-represented in FAS-ISSs, and CG was under-represented. By comparison, AC and GA were over-represented in ESEs and under-represented in ESSs ^{5,7}, again suggesting 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 293T 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 ISE ²².

To test whether the over-represented motifs possess intrinsic ISS activity, we analyzed a representative 6mer or 7mer (an “exemplar”) from each ISS group (Fig. 2a). We used a mutant of the exemplar differing by 1–3 nt as a control (Supplementary Table 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 semi-quantitative RT-PCR (Fig. 2c). These observations suggested that our approach successfully extracted core ISS motifs, and 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 in genomic scale, we determined whether the presence of these ISSs affected alternative splicing patterns in Human BodyMap 2.0 dataset. We searched the 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 PSI (percent-spliced-in) values of the exons with an ISS to those associated with a decoy. Eight out of ten ISS groups had decreased inclusion rate than their decoy sets, with four groups (B, E, D and G) having significant changes, whereas only one ISS (I) showed substantially increased PSI (Supplementary Table 5), suggesting that most FAS-ISSs can indeed interpret the splicing of alternative exons from RNA-seq data.

ISSs function in a heterologous intron and cell type

SREs generally function through 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 original screen by transiently transfecting HeLa cells with reporters containing the same exemplars. While no skipping of the test exon was observed in reporters containing randomly chosen sequences, exemplars of all 10 ISS groups caused a significant 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 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, 4) and assayed exon inclusion. We observed that 8 of the 10 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 ISSs²⁰. 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 SREs¹. 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). Strikingly, 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 293T cells (Supplementary Fig. 2a). The only exception was exemplar B, which functioned as an ESE in 293T cells but lacked detectable ESE activity in HeLa cells.

Using a reporter with two competing 5'SS, 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 site²³. When inserting each ISS exemplar between competing 5'SS 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 (i.e., ESS-like).

We scanned the Illumina's Human BodyMap 2.0 dataset for alternative exons containing either ISS sequences or their decoys, and examined the correlation of ISS with the change of exon inclusion rate. The ISSs with ESS activity (group A, H, and I, shown in red) all showed remarkably decreased PSI values compared to decoy set (Supplementary Table 5), whereas most ISSs with ESE activity increased the PSI value although only groups D and G produced substantial increases. Similar analyses for exons with alternative 5'SS did not

produce statistically significant results, probably because the sample size was small in this case.

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 (CE) 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 (SE) compared to CE, consistent with their function in repressing 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 classification²⁴. 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 CE, SE and flanking introns – e.g., some groups were enriched in CEs, others in SEs 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 thirteen pairs of pre-mRNA regions (Supplementary Table 6), and analyzed the vectors of enrichment values 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 4 clusters of motifs (numbered by cluster size) including two tight clusters (Cluster 1: motifs C, D, E, G, and 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 CEs relative to SEs and introns, and by a convex distribution along exons (i.e. 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'SS (Fig. 3), a pattern typical of ESEs²³. Cluster 2 motifs were enriched in introns and SEs relative to CEs and had variable exonic distributions. Conversely they inhibited exon inclusion and proximal 5'SS usage from an exonic location (Fig. 3 and 4b), similar to ESSs²³. 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 intron-enriched motifs inhibited exon inclusion, consistent with the RNA-seq results in endogenous exons (Fig. 4b). The existence of 4 clusters of ISS motifs with distinct patterns of activity contrasts with the relatively homogeneous distributions and activities of FAS-ESSs, suggesting that the functions of intronic SREs are more diverse.

Identification of ISS binding factors

SREs generally function through recruiting *trans*-factors to affect splicing¹. Alternatively, they affect splicing by forming inhibitory structures to block recognition of splice signals²⁵ or changing the long range RNA structure²⁶, or by altering the spliceosome assembly

dynamics⁹. Since we didn't detect particular changes of pre-mRNA structures in the original intron containing ISSs versus control 10mers using Mfold (not shown) and most FAS-ISS motifs had silencer activity in a heterologous intron (Fig. 3b), they are most likely to function through recruiting inhibitory *trans*-factors. To unbiasedly identify associated factors, we incubated a short RNA fragment of ISS exemplars with whole cell extract, and isolated RNA-protein complexes with streptavidin beads and identified proteins specifically bound by ISSs with mass spectrometry²⁷. (Fig. 5a).

Our approach was illustrated 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 (no. 2) corresponded to hnRNP L that 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 translation²⁹. Band 3 was identified as Y box binding protein 1 (YB-1) that can bind to both DNA and RNA. YB-1 was reported to influence alternative splicing of CD44³⁰, but its RNA binding specificity was 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 SRSF1³¹ but was later localized to the mitochondrial matrix³², suggesting 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 RNAi and over-expression of identified proteins in cells transfected with splicing reporters containing cognate ISS. Knockdown of either hnRNP L or YB-1 derepressed the inclusion of an exon containing the group U ISS, whereas RNAi of an unrelated RBP (GRSF-1) had no effect (Fig. 5c, 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 over-expression 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, supplementary Fig. 4b). These data implicated hnRNP L and YB-1 as factors responsible for the ISS activity of motif U. Interestingly, simultaneous knockdown of hnRNP L and YB-1 had similar effects as RNAi of each protein alone, and over-expression of both factors repressed splicing less than over-expression of YB-1, suggesting 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 over-expression of YB-1 can rescue the RNAi effect of hnRNP L in splicing reporters containing cognate ISS but not control ISS (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 positions³³⁻³⁶. 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), implying sequence-specific binding of RNAs by associated factors. The protein bands present only in ISS samples were identified by mass spectrometry. Most of identified factors were known RBPs or contained putative RNA binding domains, while 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 (e.g. U1-70k, U1A, SmD1 and SmD2) co-purified with ISS motif E, a purine-rich element recognized by SFRS1 (Table 1). Since SRSF1 was detected as a component of U1 snRNP complex³⁷, these U1 snRNP core components probably bound to group E ISS indirectly through interacting with SFRS1.

We also identified some abundant proteins that probably bound RNA in a non-sequence-specific manner (e.g., La protein was co-purified with 3 motifs and nucleolin co-purified with motif C). Of the 24 identified putative sequence-specific splicing factors (Table 1), 12 belonged to the hnRNP class, many of which were previously shown to regulate splicing, including hnRNPs A1, A2, H1, I and L^{35,36,38–40}. Several non-hnRNP proteins have not previously been 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, suggesting 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 the same cluster (Fig. 4a) tended to bind similar proteins, while a few specific proteins (e.g., members of 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 that bound to a highly overlapping set of factors despite significant sequence diversity. These factors included hnRNP A0, A1, A2, A3, D0, DL and DAZAP1, most containing two N-terminal RRM domains and a C-terminal Gly-rich motif. We used surface plasmon resonance (SPR) to determine direct interactions between the ISS exemplar and single purified protein (hnRNP A0, A1, A2, D0 and DL were tested, hnRNP A3 failed to express). Notably, the ISS motifs F, H and I can all directly bind to multiple proteins in the hnRNP A family (A0, A1, A2) and to hnRNP D0 and hnRNP DL (Fig. 6b, supplementary Fig. 6), although each protein exhibited somewhat different affinities for individual exemplar. For example, hnRNP A0 and D0 preferentially bound the A or U-rich motif H, while hnRNP DL showed preferential binding to the motif I. Nevertheless, the binding affinities to distinct motifs were fairly similar, with apparent K_d values in the range 10–100 nM for each pair (unpublished), suggesting that the RNA-protein connections detected in our affinity purification typically reflected direct interactions.

Gly-rich and RS domains inhibit splicing from introns

The hnRNP proteins comprised roughly half of the putative *trans*-factors identified above, many containing regions of repetitive sequence like Gly-rich domains (e.g. 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 to ESEs and hnRNP A1 inhibits splicing by recognizing ESSs (reviewed in ^{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 Gly-rich) ^{19,42}. However other mechanism may exist, as the RS domains were shown to recognize RNA near branch-point to promote splicing ⁴³. When tethered to an exon, most RS domains can activate exon splicing ^{42,44,45} whereas the Gly-rich domain of A1 can suppress splicing ^{19,45}. Since both hnRNP A1 and SRSF1 were identified as putative ISS-associated splicing repressors, we speculated that their RS and Gly-rich domains can inhibit splicing from an intronic location.

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

To examine whether the splicing inhibition at intronic locations is a general activity of the RS and Gly-rich domains, additional proteins were generated by fusing PUF(6-2) with the RS domains of SFRS2, SFRS5 and SFRS7, the Gly-rich domains of hnRNP A1, A2 and A3, a short peptide of (RS)₆ or a 19-aa Gly-rich sequence (Fig. 6c right panel). Remarkably, compared to the controls (lanes 13 and 14), all of these fusions strongly inhibited inclusion of the exon containing the cognate target at downstream introns, suggesting that a polypeptide containing RS repeats or a short Gly-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 to understand 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, supporting 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. While 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 (10 ISSs versus 7 ESSs). All ESS motifs had similar patterns of distributional bias^{7,23} and were able to repress intron-proximal 5'SS and 3'SS, demonstrating remarkable consistency of activity²³. By contrast, the ISS motifs had diverse patterns of distributional bias, with different motifs being most enriched in either CE, SE 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 distributional patterns, emphasizing the need for subclassification of ISSs. We used a PCA approach to define four clusters of functionally different ISS motifs based on distinct genomic distribution patterns, thus the motifs within same cluster may still have quite different consensus sequence. The associated *trans*-factors can act through multiple motifs and sometimes even across multiple PCA classes, suggesting this classification is likely to be improved by taking into account of *trans*-factors.

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, which include using a short RNA (21–23 nt) with 3 copies of ISS exemplars to increase affinity, using a long spacer separating biotin from RNA to increase the RNA accessibility, and incubating in a large volume (i.e. 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 degradation⁴⁷ and GRSF-1 was shown to affect translation⁴⁸. Even the 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, supporting 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 function²². Other relationships could provide subtler tuning of splicing levels, e.g., 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, e.g., in HeLa cells nPTB can compensate for PTB depletion³⁴, while in neural development replacement of PTB by nPTB may alter the regulation of a substantial alternative splicing program⁵¹.

A variety of regulatory relationships are possible in the examples identified here. For example, both hnRNP L and YB-1 bound ISS motif U. While over-expression or depletion of either factors was sufficient to repress or derepress splicing respectively, their relationship is not 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 that typically have the opposite pattern of activity^{38,52–54}. Our data suggested these factors might antagonistically regulate introns containing motifs D or E - in 293T cells, binding of motifs D and E by SFRS1 or other intron-repressive factors likely predominates, since these motifs have ISS activity. The less-characterized factors RBM45 and GRSF-1 also bind motif D, potentially adding additional regulatory complexity via this motif.

Our analyses using fusions of candidate splicing regulatory domains indicated that the requirements for splicing repression were surprisingly minimal. Tethering of an RS or Gly-rich domain from any of various SR or hnRNP proteins (or even a short RS repeat or Gly-rich peptide) at 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 human proteome identified a total of 24 human proteins that contain both a short Gly-rich segment and at least one RRM domain, representing an 8-fold enrichment above the frequency expected from occurrence of these domains individually (unpublished).

Online Methods

Splicing reporters

All splicing reporters were constructed from a backbone vector, pZW1, which contains a multi-cloning site between two GFP exons⁷. The random sequence library was generated from a foldback primer extended with Klenow and digested and ligated into pZW11 that contain a multi-cloning site at downstream of a constitutive exon. To test ISSs in a heterologous context, a new reporter (pZW2C) was constructed by inserting the exon 2 of DHFR gene and part of its flanking introns between the two GFP exons. The reporter with competing 5'SS and the modular splicing reporter were described previously^{23,55,56}. Additional details were described in supplemental note.

Cell culture and transfections

293 cells and HeLa cells were cultured with D-MEM medium supplemented with 10% fetal bovine serum (FBS). Transfections were carried out with lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions.

For stable transfection, 293 FlpIn cells were co-transfected with 10 fold excess of pOG44 plasmid. To select stable transfectants, the cells were expanded by a 1 to 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, 1.6 µg pZW11 vector containing random library were transfected into 293 FlpIn cells in a 15 cm tissue culture dish. After selection for 10 days, the positive clones were trypsin digested, pooled and sorted using a Cytomation MoFlo high-speed sorter into 96 well plates at one cell per well. The survival colonies will be visible after 10 days, and were checked for green fluorescence. The total DNAs from GFP positive cells were used for PCR and sequencing. The ISS sequences were further analyzed with a computational pipeline as previously described.

RNA purification and Semi-quantitative 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 non-fluorescent 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 AGTGCTTCAGCCGCTACCC for GFP exon 1 and GTTGTACTCCAGCTTGTGCC for exon 3.

RNA affinity purification and identification of *trans*-factors

The RNA affinity purification method was modified from a protocol of Marzluff's lab²⁷. 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% confluent and resuspended with ice-cold 2.5 ml resuspension buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl). The cells were mixed with 2.5ml 2x lysis buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 15 mM NaN₃, 1%(V/V) NP-40, 2 mM DTT, 2 mM PMSF, 2x protease inhibitor mix) and lysed for 5 min, and then centrifuge at 12000 g for 20 min at 4°C. Subsequently 0.75 nmol biotinylated RNA with two 18 atom spacers (purchased from Dharmacon) were added and incubated for 2 hrs at 4 °C. We then added 50 µl Streptavidin-agarose beads (purchased from Sigma) into the mixture and incubated for 2 hrs at 4 °C with slow rotation. The beads were rinsed 3 times using 4 ml lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 15 mM NaN₃, 0.5% NP-40, 1 mM DTT, 1 mM PMSF, 1x protease inhibitor mix), resuspended in 40 µl final volume, and mixed with 10 µl 5x SDS loading buffer. The proteins were then separated with a 10% SDS-PAGE gel and stained with coomassie blue. The gels was kept in 3% acetic acid for the further mass spectrometry analysis. Bands specific to the ISS RNA but not to controls were excised and the gel slices containing the candidate proteins was 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 was carried out by the UNC Proteomics Center.

Analyze the correlation between ISSs and the 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 50bp paired end reads were mapped using Tophat and PSI values of alternative exons were calculated using MISO⁵⁷. 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 identified in the skipped exon body and in the 200 bases following the skipped exon. Significance of overall ISS effect was calculated as the median of comparisons between PSI values for events with an ISS and those with a decoy. SE 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.

Knock down and over-expression of putative *trans*-factors

The siRNAs were purchased from Dharmacon (On-target SMARTpool, with scramble dsRNA controls), and transfections were conducted with Lipofectamine 2000 (Invitrogen) in cells grown in 24-well plates. After 48 hours of siRNA transfection, we transfected with 0.2 µg splicing reporter, and harvested the cells 24 hours after the second transfection for further analyses. In over-expression experiments, we co-transfected 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, 293T cells were transfected with siRNA or control siRNA. After 48 hours, we co-transfected FLAG-tagged YB1 and ISS splicing reporters, and collected the cells 24 hours after the second transfection for further analysis.

Co-immunoprecipitation of hnRNP L and YB1

FLAG-tagged splicing factors were transfected into 293T cell as described above. After 48 hours, the transfected cells were lysed in lysis buffer containing 50 mM HEPES, 150 mM NaCl (4.38g), 1mM EDTA, 1% (w/v) CHAPS and Sigma protease inhibitor cocktail (with or without 50 µg/ml of RnaseA) at room temperature for 10 minutes. The M2 FLAG agarose resin (40 µl) was prepared as described by manufacturer instruction, and incubated with 1000 µl of cell lysate supernatant with gentle agitation at 4 °C for overnight. The IP samples were span down at 8000rpm for 30 seconds and washed with wash buffer (50 mM HEPES, 150 mM NaCl, 0.1% Triton X-100, 10% glycerol, pH to 7.5) for 3 times. Then the proteins were eluted with FLAG peptide (200ng/µl) and transferred into a new tube for further analysis.

Exon and intron datasets

The exon and intron datasets were generated using similar filters as described earlier²³. 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. The following numbers of orthologous human/mouse exon pairs were obtained using our criteria: 1232 A5Es, 1408 A3Es, 2964 SEs and 44368 CEs. See supplemental method for details in computational analyses.

Positional frequency of ISS in different pre-mRNA locations

The ISS-kmers in each group were analyzed separately for the positional frequency along the pre-mRNA of different exon/intron dataset. The first and last 60-nt of exons, together with the 200 nt flanking introns were used in the analyses. The positional frequency was defined as number of transcripts that contain 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 analysis

The hnRNP A0, A1, A2, D0 and DL were cloned into pT7HtB expression vector with 6xHis tag and expressed in *E. coli* BL21(DE3) strain. The proteins were purified into >90% purity with His GraviTrap Kit (GE Health Care) following manufacturer's instruction. The purified protein was buffer exchanged to remove excess imidazole and stored at 4 °C in storage buffer (20mM HEPES pH 7.4 0.5M NaCl, 10% glycerol and 2mM DTT).

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

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the J. Hui from Shanghai Institute of Biological Science, Shanghai, China and A. Willis from University of Nottingham, UK providing expression constructs of *trans*-factors, and Brenton Graveley from University of Connecticut Health Center, Farmington, CT for constructs containing RS domains. We thank T. Nilsen and A. Berglund for critical reading of manuscripts. We thank Z. Dominski and B. Marzluff for helping in RNA affinity purifications. This work is supported by an AHA grant (0865329E) and NIH grant (R01CA158283) to ZW, and NIH grant to CB.

Reference

1. Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*. 2008; 14:802–813. [PubMed: 18369186]
2. Matlin AJ, Clark F, Smith CW. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol*. 2005; 6:386–398. [PubMed: 15956978]
3. Blencowe BJ. Alternative splicing: new insights from global analyses. *Cell*. 2006; 126:37–47. [PubMed: 16839875]
4. Liu HX, Zhang M, Krainer AR. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev*. 1998; 12:1998–2012. [PubMed: 9649504]
5. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. Predictive identification of exonic splicing enhancers in human genes. *Science*. 2002; 297:1007–1013. [PubMed: 12114529]
6. Zhang XH, Chasin LA. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev*. 2004; 18:1241–1250. [PubMed: 15145827]
7. Wang Z, et al. Systematic identification and analysis of exonic splicing silencers. *Cell*. 2004; 119:831–845. [PubMed: 15607979]
8. Goren A, et al. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol Cell*. 2006; 22:769–781. [PubMed: 16793546]
9. Yu Y, et al. Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. *Cell*. 2008; 135:1224–1236. [PubMed: 19109894]

10. Culler SJ, Hoff KG, Voelker RB, Berglund JA, Smolke CD. Functional selection and systematic analysis of intronic splicing elements identifies active sequence motifs and associated splicing factors. *Nucleic Acids Res.* 2010
11. Sharma S, Kohlstaedt LA, Damianov A, Rio DC, Black DL. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat Struct Mol Biol.* 2008; 15:183–191. [PubMed: 18193060]
12. Kashima T, Rao N, Manley JL. An intronic element contributes to splicing repression in spinal muscular atrophy. *Proc Natl Acad Sci U S A.* 2007; 104:3426–3431. [PubMed: 17307868]
13. Hui J, et al. Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J.* 2005; 24:1988–1998. [PubMed: 15889141]
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.* 1999; 18:1939–1952. [PubMed: 10202157]
15. Kanopka A, Muhlemann O, Akusjarvi G. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature.* 1996; 381:535–538. [PubMed: 8632829]
16. Ibrahim el C, Schaal TD, Hertel KJ, Reed R, Maniatis T. Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proc Natl Acad Sci U S A.* 2005; 102:5002–5007. [PubMed: 15753297]
17. Shen M, Mattox W. Activation and repression functions of an SR splicing regulator depend on exonic versus intronic-binding position. *Nucleic Acids Res.* 2012; 40:428–437. [PubMed: 21914724]
18. McNally LM, McNally MT. SR protein splicing factors interact with the Rous sarcoma virus negative regulator of splicing element. *J Virol.* 1996; 70:1163–1172. [PubMed: 8551577]
19. Del Gatto-Konczak F, Olive M, Gesnel MC, Breathnach R. hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. *Mol Cell Biol.* 1999; 19:251–260. [PubMed: 9858549]
20. Hua Y, Vickers TA, Okunola HL, Bennett CF, Krainer AR. Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *Am J Hum Genet.* 2008; 82:834–848. [PubMed: 18371932]
21. Tange TO, Damgaard CK, Guth S, Valcarcel J, Kjems J. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *Embo J.* 2001; 20:5748–5758. [PubMed: 11598017]
22. Wang Y, Ma M, Xiao X, Wang Z. Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. *Nat Struct Mol Biol.* 2012
23. Wang Z, Xiao X, Van Nostrand E, Burge CB. General and specific functions of exonic splicing silencers in splicing control. *Mol Cell.* 2006; 23:61–70. [PubMed: 16797197]
24. Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc Natl Acad Sci U S A.* 2011; 108:11093–11098. [PubMed: 21685335]
25. Huang C, et al. A structured RNA in HBV PRE represses alternative splicing in a sequence-independent and position-dependent manner. *FEBS J.* 2011
26. Pervouchine DD, et al. Evidence for widespread association of mammalian splicing and conserved long-range RNA structures. *RNA.* 2012; 18:1–15. [PubMed: 22128342]
27. Dominski Z, Yang XC, Kaygun H, Dadlez M, Marzluff WF. A 3' exonuclease that specifically interacts with the 3' end of histone mRNA. *Mol Cell.* 2003; 12:295–305. [PubMed: 14536070]
28. Rothrock CR, House AE, Lynch KW. HnRNP L represses exon splicing via a regulated exonic splicing silencer. *Embo J.* 2005; 24:2792–2802. [PubMed: 16001081]
29. Nielsen FC, Nielsen J, Christiansen J. A family of IGF-II mRNA binding proteins (IMP) involved in RNA trafficking. *Scand J Clin Lab Invest Suppl.* 2001; 234:93–99. [PubMed: 11713986]
30. Allemand E, Hastings ML, Murray MV, Myers MP, Krainer AR. Alternative splicing regulation by interaction of phosphatase PP2Cgamma with nucleic acid-binding protein YB-1. *Nat Struct Mol Biol.* 2007; 14:630–638. [PubMed: 17572683]
31. Krainer AR, Conway GC, Kozak D. Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev.* 1990; 4:1158–1171. [PubMed: 2145194]

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.* 1997; 272:24363–24370. [PubMed: 9305894]
33. Oberstrass FC, et al. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science.* 2005; 309:2054–2057. [PubMed: 16179478]
34. Spellman R, Llorian M, Smith CW. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol Cell.* 2007; 27:420–434. [PubMed: 17679092]
35. Jin W, Bruno IG, Xie TX, Sanger LJ, Cote GJ. Polypyrimidine tract-binding protein down-regulates fibroblast growth factor receptor 1 alpha-exon inclusion. *Cancer Res.* 2003; 63:6154–6157. [PubMed: 14559795]
36. Cote J, Dupuis S, Wu JY. Polypyrimidine track-binding protein binding downstream of caspase-2 alternative exon 9 represses its inclusion. *J Biol Chem.* 2001; 276:8535–8543. [PubMed: 11116151]
37. Das R, et al. SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell.* 2007; 26:867–881. [PubMed: 17588520]
38. Chou MY, Rooke N, Turck CW, Black DL. hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol Cell Biol.* 1999; 19:69–77. [PubMed: 9858532]
39. Kashima T, Rao N, David CJ, Manley JL. hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. *Hum Mol Genet.* 2007; 16:3149–3159. [PubMed: 17884807]
40. Hui J, Stangl K, Lane WS, Bindereif A. HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat Struct Biol.* 2003; 10:33–37. [PubMed: 12447348]
41. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol.* 2009; 10:741–754. [PubMed: 19773805]
42. Graveley BR, Maniatis T. Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol Cell.* 1998; 1:765–771. [PubMed: 9660960]
43. Shen H, Kan JL, Green MR. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol Cell.* 2004; 13:367–376. [PubMed: 14967144]
44. Graveley BR, Hertel KJ, Maniatis T. A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *Embo J.* 1998; 17:6747–6756. [PubMed: 9822617]
45. Wang Y, Cheong CG, Hall TM, Wang Z. Engineering splicing factors with designed specificities. *Nat Methods.* 2009; 6:825–830. [PubMed: 19801992]
46. Cheong CG, Hall TM. Engineering RNA sequence specificity of Pumilio repeats. *Proc Natl Acad Sci U S A.* 2006; 103:13635–13639. [PubMed: 16954190]
47. Tourriere H, et al. The RasGAP-associated endoribonuclease G3BP assembles stress granules. *J Cell Biol.* 2003; 160:823–831. [PubMed: 12642610]
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.* 2008; 22:1838–1850. [PubMed: 18593884]
49. Michlewski G, Guil S, Semple CA, Caceres JF. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol Cell.* 2008; 32:383–393. [PubMed: 18995836]
50. Guil S, Long JC, Caceres JF. hnRNP A1 relocalization to the stress granules reflects a role in the stress response. *Mol Cell Biol.* 2006; 26:5744–5758. [PubMed: 16847328]
51. Boutz PL, et al. A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* 2007; 21:1636–1652. [PubMed: 17606642]
52. Caputi M, Zahler AM. Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J Biol Chem.* 2001; 276:43850–43859. [PubMed: 11571276]
53. Schaub MC, Lopez SR, 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.* 2007; 282:13617–13626. [PubMed: 17337441]

54. Chen CD, Kobayashi R, Helfman DM. 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.* 1999; 13:593–606. [PubMed: 10072387]
55. Xiao X, Wang Z, Jang M, Burge CB. Coevolutionary networks of splicing cis-regulatory elements. *Proc Natl Acad Sci U S A.* 2007; 104:18583–18588. [PubMed: 17998536]
56. Xiao X, et al. Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat Struct Mol Biol.* 2009; 16:1094–1100. [PubMed: 19749754]
57. Katz Y, Wang ET, Airoidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods.* 2010; 7:1009–1015. [PubMed: 21057496]

prior to screening. Sequences around the insertion region (positions 480–489) are shown. (d) Diagram of screening strategy and resulting ISS motifs. The 102 unique ISS 10mers identified through FAS-ISS were clustered based on the sequence similarity using ClustalW2. The 16 ISS 10mers that were reinserted into the original construct and retested for ISS activity by transient transfection are indicated (red asterisks).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

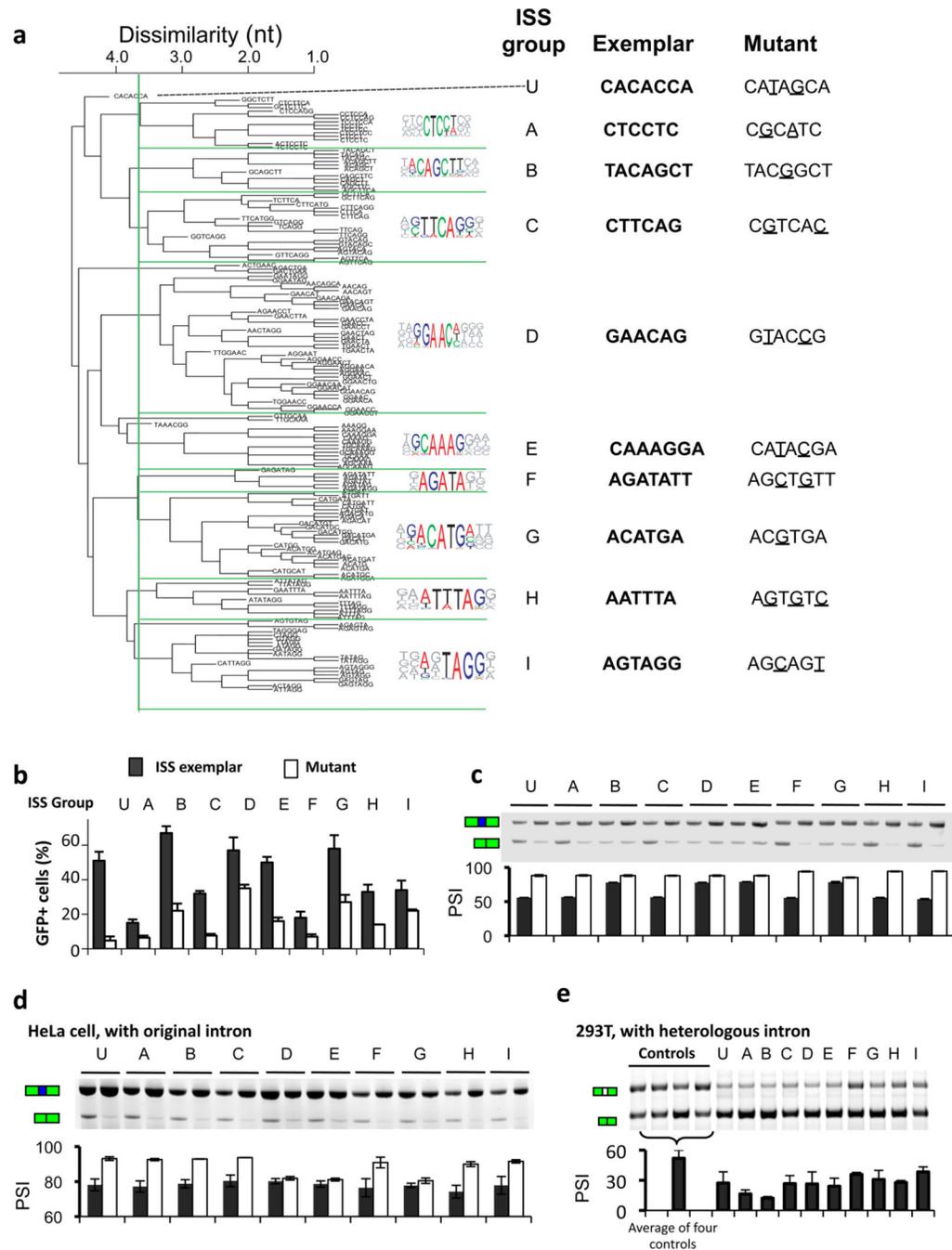


Figure 2. Identification of core ISS motifs and exemplars and validation of activity

(a) Over-represented oligonucleotide k -mers were extracted from FAS-ISS 10mers and clustered based on sequence similarity. The consensus motifs, the exemplars (representative 6- or 7-mers) and tested mutations for each group are listed at right. (b) The ISS exemplar and mutant from each group were inserted into the pZW11 plasmid, transiently transfected into 293T cells, and splicing was assessed by flow cytometry as the fraction of GFP positive cells, or (c) by semi-quantitative RT-PCR (sqRT-PCR). The percent-spliced-in (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. HeLa cells were transfected with pZW11 containing ISS exemplars and their mutants, and the inclusion of the test exon was determined by sqRT-PCR. Error bars, s.d.; n=3. (d) Validation of ISS activity in a heterologous intron context. Two copies of ISS exemplars (Supplementary Table 3) from each group were inserted downstream of a cassette exon in a different splicing reporter (pZW2C, see Methods) and transfected into 293T cells, followed by sqRT-PCR as above. Four neutral sequences that did not match known SREs (Supplementary Table 4) were used as negative controls. Error bars, s.d.; n=3.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

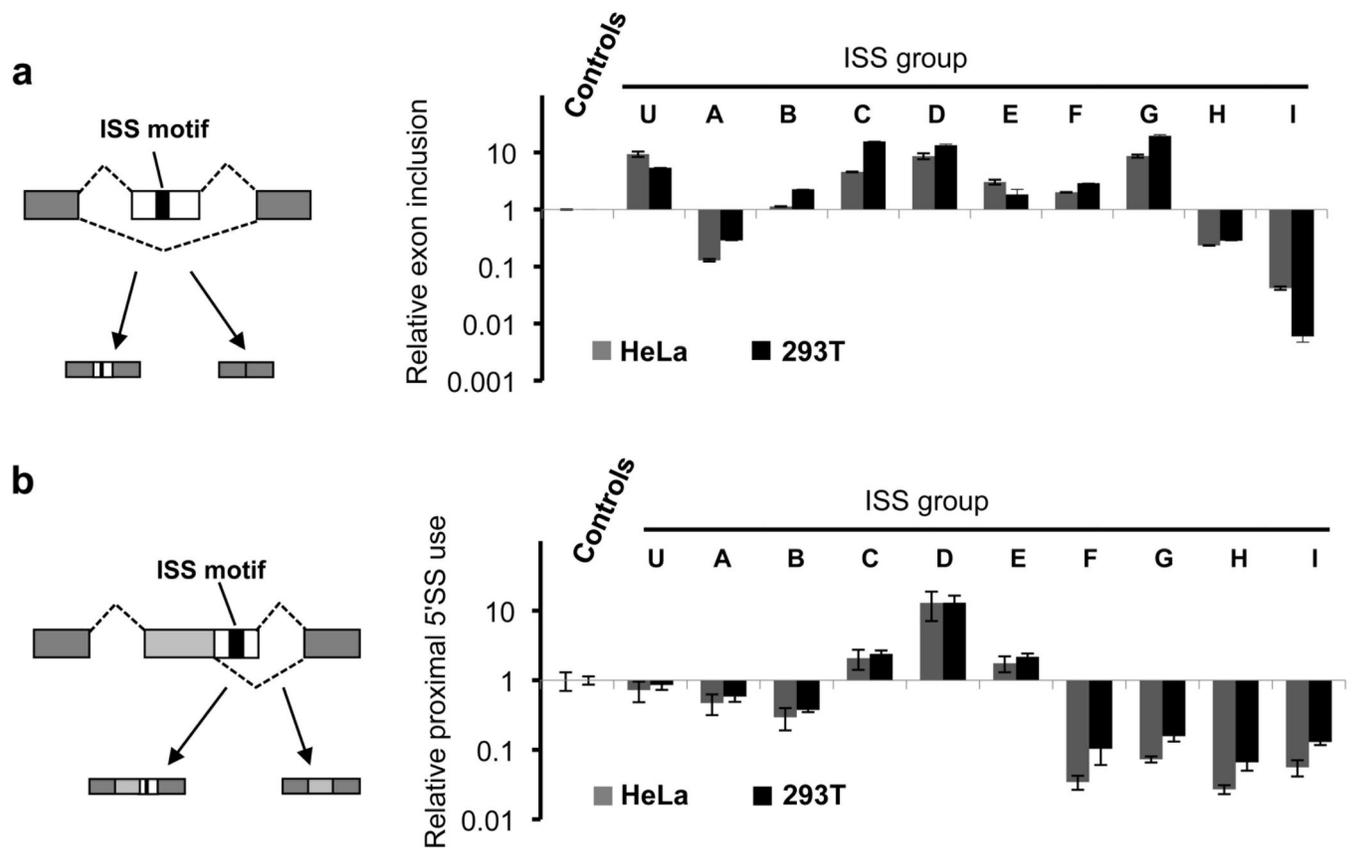


Figure 3. ISS motifs regulate splicing from an exonic context

(a) FAS-ISS motifs can function as ESEs or ESSs. Two copies of ISS exemplars were inserted into the test exon of a modular splicing reporter⁵⁵, and transfected into HeLa and 293T cells. The resulting PSI values divided by the mean of 3 negative controls are shown as “relative exon inclusion” levels (Supplementary Table 4). The mean and s.d. of three independent transfections are shown. (b) FAS-ISS motifs can regulate alternative 5'SS. A single copy of the exemplar motif from each group was inserted between two alternative 5'SS of a minigene reporter²³ and transfected into HeLa and 293T cells. The observed PSI value – defined as the fractional use of the proximal (downstream) 5'SS – was measured by sqRT-PCR and divided by that observed in controls to give relative proximal 5'SS usage (Supplementary Table 4). Error bars, s.d.; n=3.

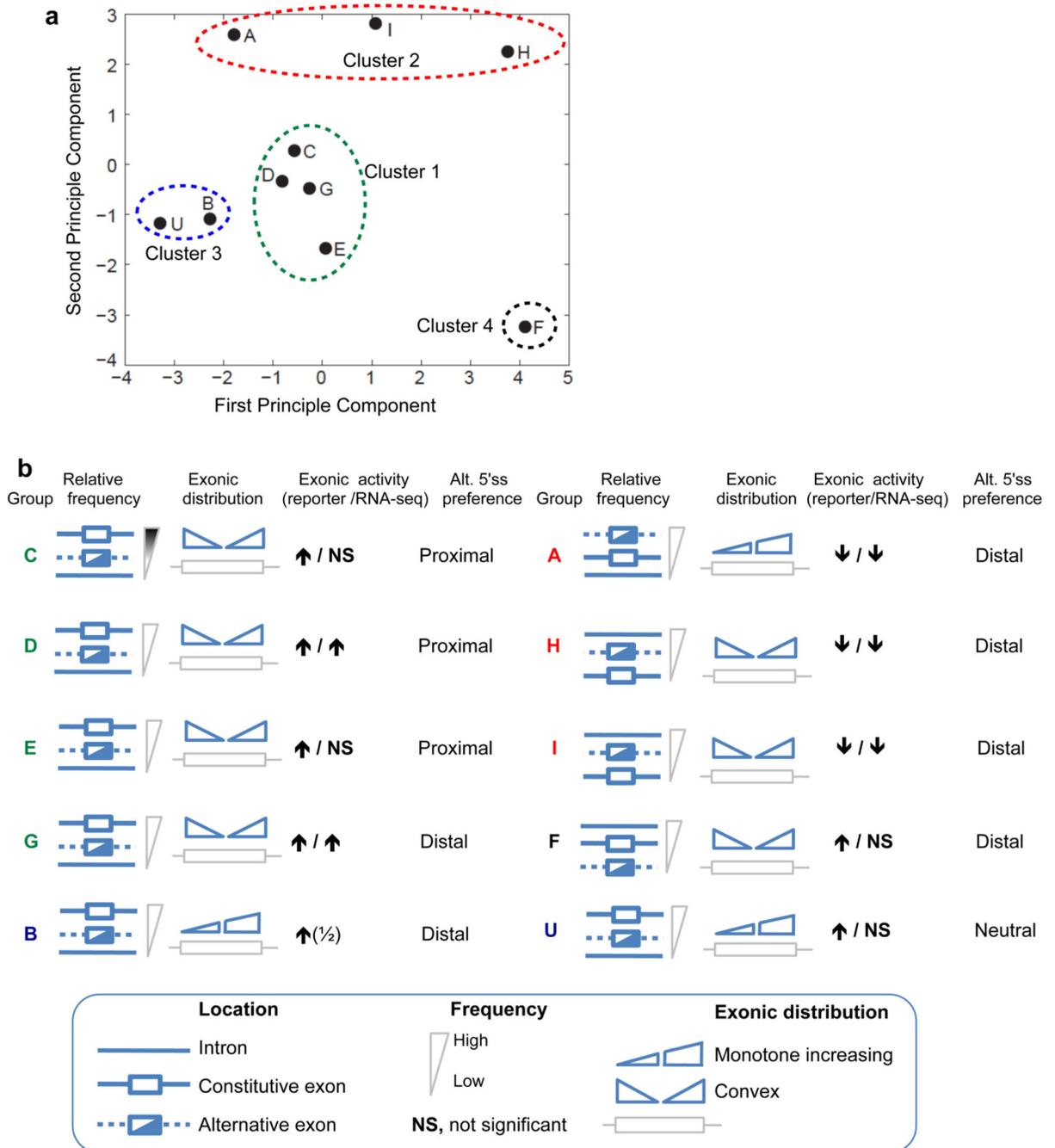


Figure 4. Classification of ISSs based on genomic distribution yields clusters with similar context-dependent activity

(a) Principal component analysis (PCA) of relative enrichment in 13 pairs of pre-mRNA regions. Relative enrichment between each pair of regions (e.g., SE relative to CE) were converted to a vector of 13 z-scores for each ISS motif, which were analyzed by standard PCA. ISS groups were clustered based on 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 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 $\frac{1}{2}$ indicates that motif B had ESE activity in 1 of 2 cell lines tested). The ISS motif labels are color-coded according to the PCA clusters shown in panel A.

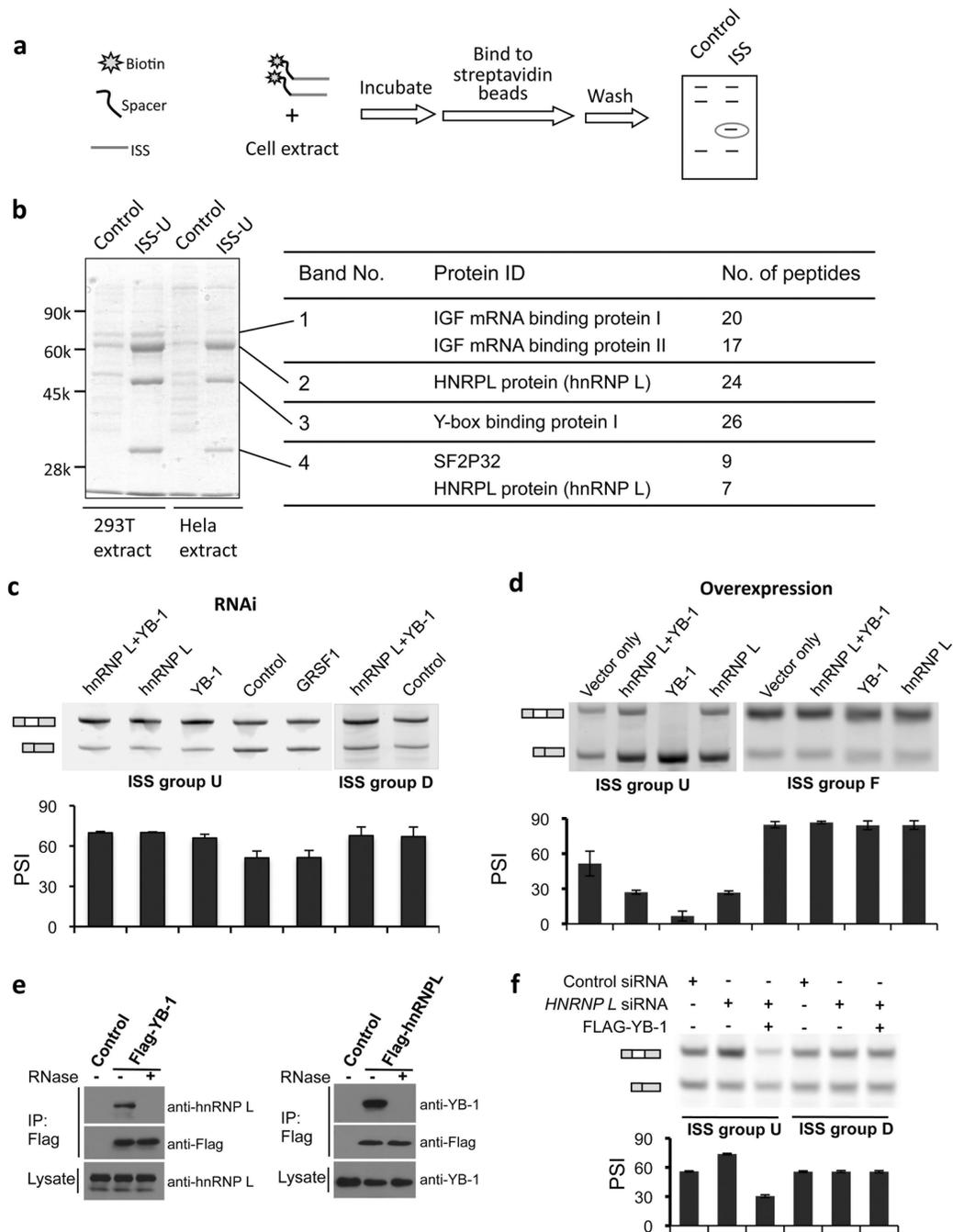


Figure 5. Identification and validation of ISS-associated splicing repressors

(a) Diagram of protein identification procedure. (b) Affinity purification using motif U (CA-rich) and its mutant as bait incubated with extracts from 293T and HeLa cells. Four specific bands were excised for mass spectrometry, yielding peptides from the proteins listed at right. (c) RNAi analysis of the putative splicing regulatory factors. Derepression of the target exon is 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 semi-quantitative RT-PCR from three independent experiments are

shown below a representative gel in all panels c, d and f (error bars, s.d.; n=3). (d) RT-PCR experiments showing the over-expression of *trans*-factors specifically inhibits exon inclusion in reporters with the cognate ISS. Reporters contain ISSs of other group were used as a specificity control. (e) Co-immunoprecipitation in 293T 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 showing that over-expression of YB-1 reversed the splicing derepression phenotype caused by RNAi of hnRNP L. The splicing reporters containing cognate ISS (group U) or control ISS (group D) were used as a specificity control.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

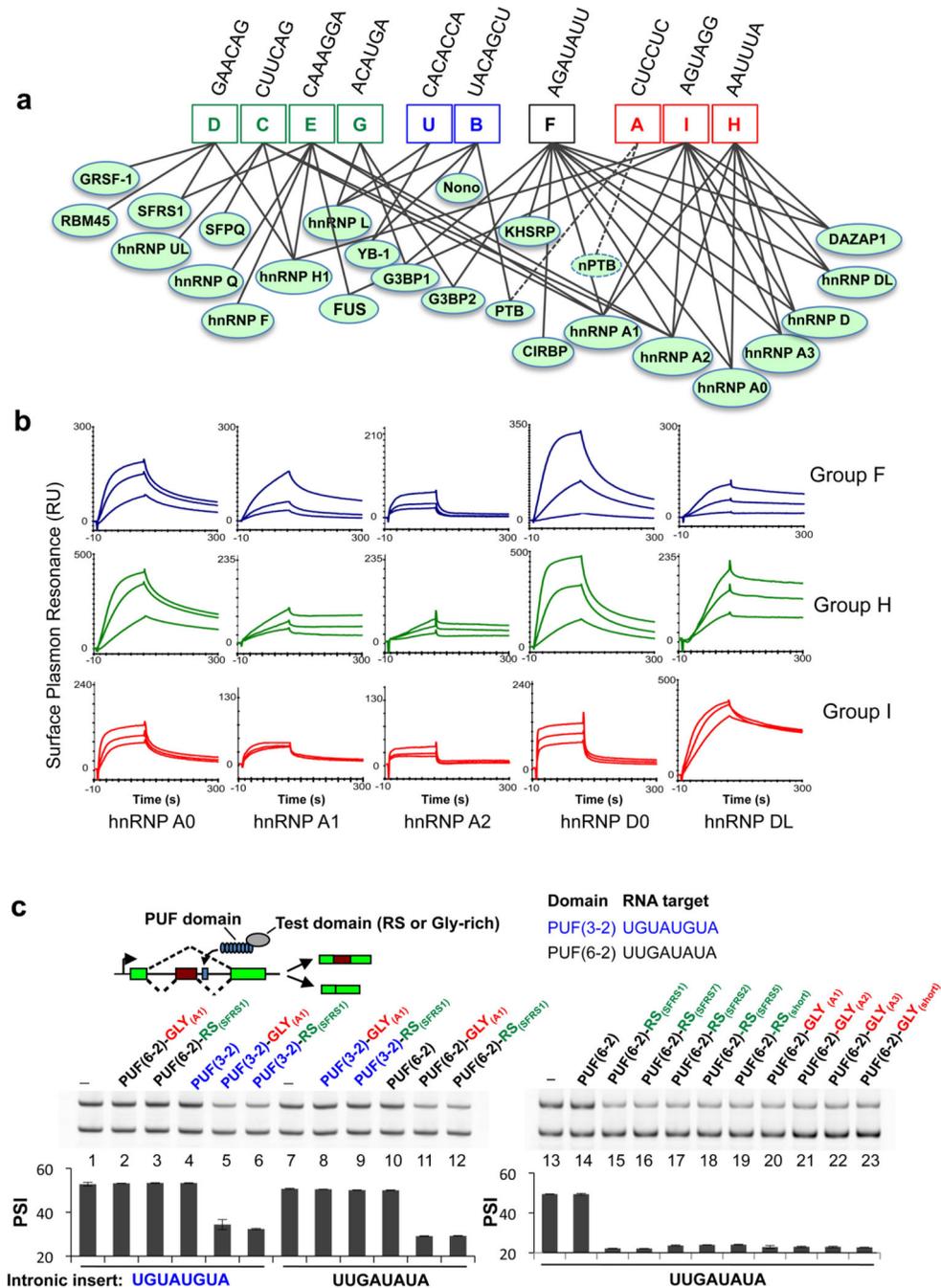


Figure 6. ISSs are recognized by a complex, overlapping network of factors

(a) Interactions between putative splicing factors and all ISS groups were represented by solid lines (dashed lines represent interactions inferred from the literature), yielding a many-to-many connectivity map encompassing all 10 ISSs and their cognate factors. (b) Surface plasmon resonance 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 panel) were injected at a flow rate of 50 μ l/min 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 Gly-rich to a downstream intronic location can repress splicing. Fusion proteins consisting of a PUF domain fused to an RS or Gly-rich domain were assayed for their effects on splicing. Two target 8mers, UGUAUGUA and UUGAUUAU 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 293T cells. All combinations of fusion proteins and targets were tested, with matching PUF-target pairs indicated by 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 Gly-rich domains were tested. Error bars, s.d.; $n=3$.

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

ISS Group	Representative ISS k-mer	Associated Proteins	
		Putative splicing factors (protein motifs) ¹	Other RNA binding proteins
A	CTCCTC	hnRNP I/PTB ² (4 RRM + Ala-rich)	
B	TACAGCT	Nono (2 RRM) YB-1 (CSD) hnRNP I/PTB (4 RRM + Ala-rich)	La protein
C	CTTCAG	SFPQ (2 RRM + Gln/Glu/Pro-rich) hnRNP UL1 (Gly-rich + Pro-rich) hnRNP H1 (3 RRM + Gly-rich) hnRNP A1 (2 RRM + Gly-rich) hnRNP A2/B1 (2 RRM + Gly-rich)	Nucleolin
D	GAACAG	GRSF-1 (3 RRM + Ala-rich) hnRNP F (3 RRM) SFRS1 (2 RRM + Gly-rich) RBM45/CELF-3 homolog (3 RRM)	CSDE1
E	CAAAGGA	SFRS1 (2 RRM + Gly-rich) hnRNP Q (3 RRM) hnRNP A1 (2 RRM + Gly-rich) hnRNP A2/B1 (2 RRM + Gly-rich) hnRNP H1 (3 RRM + Gly-rich) hnRNP F (3 RRM) FUS	several U1 snRNP core proteins (e.g., U1-70k, U1 snRNP protein A and C, SmD1, SmD2)
F	AGATATT	KHSRP (4 KH + Gly-rich) DAZAP1 (2 RRM + Pro-rich) hnRNP D0 (2 RRM + Gly-rich) hnRNP A0 (2 RRM + Gly-rich) hnRNP A1 (2 RRM + Gly-rich) hnRNP A2/B1 (2 RRM + Gly-rich) hnRNP A3 (2 RRM + Gly-rich) CIRBP (1 RRM + Gly-rich) G3BP-1 (1 RRM + Gly-rich) G3BP-2 (1 RRM + Gly-rich)	La protein
G	ACATGA	G3BP1 (1 RRM + Gly-rich) hnRNP L (3 RRM + Gly-rich + Pro-rich)	GTF2I
H	AATTTA	DAZAP1 (2 RRM + Pro-rich) hnRNP A0 (2 RRM + Gly-rich)	La protein

ISS Group	Representative ISS k-mer	Associated Proteins	
		Putative splicing factors (protein motifs) ¹	Other RNA binding proteins
		hnRNP D0 (2 RRM + Gly-rich) hnRNP DL (2 RRM + Gly-rich) hnRNP A1 (2 RRM + Gly-rich) hnRNP A2/B1 (2 RRM + Gly-rich) hnRNP A3 (2 RRM + Gly-rich)	
I	AGTAGG	DAZAP1 (2 RRM + Pro-rich) hnRNP A1 (2 RRM + Gly-rich) hnRNP A2/B1 (2 RRM + Gly-rich) hnRNP A3 (2 RRM + Gly-rich) hnRNP H1 (3 RRM + Gly-rich) hnRNP D0 (2 RRM + Gly-rich) hnRNP DL (2 RRM + Gly-rich) FUS	DHX36
U	CACACCA	hnRNP L (3 RRM + Gly-rich + Pro-rich) YB-1 (CSD)	IGF mRNA binding proteins I, II

¹Characteristic RNA binding or splicing regulatory protein motifs are indicated (RRM: RNA Recognition Motif; CSD: Cold shock domain; Gly-rich: Glycine rich motif; etc.).

²Based on the literature rather than affinity purification/mass spec.