



HHS Public Access

Author manuscript

Nat Struct Mol Biol. Author manuscript; available in PMC 2013 August 27.

Published in final edited form as:

Nat Struct Mol Biol. 2012 October ; 19(10): 1044–1052. doi:10.1038/nsmb.2377.

Intronic Splicing Enhancers, Cognate Splicing Factors and Context Dependent Regulation Rules

Yang Wang¹, Meng Ma^{1,2}, Xinshu Xiao³, and Zefeng Wang^{1,4}

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

²Department of Computer Science and Technology, Anhui University, China 23061

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

Summary

Most human genes produce multiple splicing isoforms with distinct functions. To systematically understand splicing regulation, we conducted an unbiased screen and identified >100 intronic splicing enhancers (ISEs) that were clustered by sequence similarity into six groups. All ISEs functioned in another cell type and heterologous introns, and their distribution and conservation patterns in different pre-mRNA regions are similar to exonic splicing silencers. Consistently all ISEs inhibited use of splice sites from exonic locations. The putative *trans*-factors of each ISE group were identified and validated. Five distinct ISE motifs were recognized by hnRNP H and F whose C-terminal domains were sufficient to render context-dependent activities of ISEs. The sixth group was controlled by factors that either activate or suppress splicing. This work provided a comprehensive picture of general ISE activities and provided new models of how a single element can function oppositely depending on its locations and binding factors.

Keywords

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

Introduction

Most human genes produce multiple isoforms through alternative splicing, which is tightly controlled in different tissues and developmental stages¹⁻³. The splicing specificity is mainly determined by 5' splice site (5'SS), 3' splice site (3'SS) and branch point sequences, as well as by multiple *cis*-acting splicing regulatory elements (SRE) that are conveniently classified as exonic splicing enhancers (ESEs) or silencers (ESSs), and intronic splicing enhancers (ISEs) or silencers (ISSs). These SREs generally function by recruiting *trans*-

Users may view, print, copy, and download 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

⁴To whom correspondence should be addressed: Zefeng Wang, Phone: 919-966-0131; Fax: 919-966-5640; zefeng@med.unc.edu.

Author contribution: Y W and Z W designed the research and performed the experiments. M M and X X developed computational methods to analyze the data. Y W and Z W wrote the paper.

factors to control splicing through diverse mechanisms⁴⁻⁷. Activities of SREs are often location-dependent⁶, however their underlying mechanisms are largely unclear. An important research goal is to study these SREs and their cognate factors on a global scale to derive a set of regulation rules for splicing (i.e. “splicing code”)^{6,8}.

Significant progress has been made to systematically identify exonic SREs (ESS and ESE) using experimental and computational approaches⁹⁻¹⁷. These reports provided a global picture that ESEs are more enriched and conserved in exons to promote the exon definition^{12,18-21}, whereas the ESSs are more enriched in introns to suppress pseudoexons and help define alternative splice sites^{19,22}. In comparison, the intronic SREs are less well understood. Several computational approaches were developed to predict general intronic SREs in human based on intronic sequence conservation or distribution biases, most predicted elements resemble the RNA motifs recognized by tissue-specific splicing factors such as Fox1, Nova and nPTB²³⁻²⁵.

Several ISEs were identified by analyzing sequences near alternative exons. For instance, the G-rich sequences containing at least one G-triplet were found to enhance splicing through recruiting hnRNP H and F to introns²⁶⁻²⁹. When located in the downstream intron, the activity of G-runs is dependent on the strength of nearby 5'SS³⁰. There are also tissue specific ISEs, such as YCAY motifs (Y = C or U) that is recognized by the neuron-specific protein Nova to control many brain-specific splicing events³¹ and the UGCAUG motif that is recognized by the brain- and muscle-specific factors Fox-1 and Fox-2^{32,33}.

To systematically study general ISEs, we developed a cell-based system to screen a random decamer library for sequences that promote splicing from an intronic location. We obtained 109 unique ISE 10mers, whose core motifs were clustered into six groups. We observed a systematic overlap between ISEs and ESSs and established a general rule that all ISE motifs consistently inhibited splicing in exons. The putative *trans*-factors for each ISE motif were further identified and analyzed. Altogether these data provided comprehensive rules of how ISEs regulate different alternative splicing events, and suggested two models of how the same SRE can either promote or suppress splicing at different pre-mRNA locations.

Results

Identification of 109 ISEs that promote intron splicing

To unbiasedly identify ISEs, we developed a cell-based system called fluorescence-activated screen (FAS) that used a splicing reporter (pZW15C) with two exons and a weak intron (Fig. 1a and Supplementary Table 1). When spliced together, the exons formed an mRNA encoding the enhanced GFP. The intron is normally retained during splicing, and insertion of an ISE will promote splicing to generate a functional eGFP. We tested this reporter by inserting a G-rich ISE or a control sequence and transfecting them into 293T cells, and found that this ISE indeed promoted splicing to produce ~70% GFP-positive cells whereas the control reporter generated ~2% green cells (Fig. 1b). Using semi-quantitative RT-PCR, we confirmed that the green cells were indeed resulted from correct splicing (Fig. 1c).

Since the cores of ISEs are thought to be relatively short, we inserted a 10-nt random pool of sequences at 23 nt into the intron, far enough to avoid interference with the 5'SS (Supplementary Fig. 1a). We transformed enough *E. coli* to obtain $\sim 2 \times 10^6$ colonies that provided at least 2-fold coverage of all possible DNA decamers. The quality of this library was examined to ensure that we started with an unbiased library. The resulting library was transfected into cultured 293-FlpIn cells that contain a single site-specific recombination site for stable integration. Similar to our previous screen for ESSs¹³, this system ensured high sensitivity and unbiased sequence recovery.

We carried out 208 transfections to obtain enough ($>10^6$) stable clones that gave roughly one fold coverage of the entire decamer library. In total 117 FAS-ISE decamers were identified through the screen, 109 of which were unique (Supplementary Table 2). We identified seven decamers twice and one decamers three times in independent transfections, suggesting that the screen is self-converged (Supplementary Table 2). Based on the sequence similarity, the resulting ISE decamers can be clustered using CLUSTALW, indicating that they contained common core motifs (Fig. 1d). Although the initial library was essentially random with $\sim 25\%$ of each base, the identified FAS-ISE decamers had higher contents of G (40%) and T (35%) compared to A (18%) and C (7%) (Supplementary Fig. 1b). We also identified overrepresented (e.g., GG, TA) and underrepresented (e.g., GA, CG, GC) dinucleotides in these ISEs (Supplementary Fig. 1c). Such composition biases are similar to FAS-ESS but are different from RESCUE-ESE^{11,13}. Consistently, the ESS hexamers (FAS-hex2)¹³ were overrepresented in the FAS-ISE decamers relative to random control sets (93 versus 35, $P < 10^{-4}$ based on random shuffles of ESS hexamers), whereas the RESCUE-ESE hexamers¹¹ were underrepresented compared to random hexamer sets (9 vs 45, $P < 10^{-4}$ based on random shuffles).

To validate our results, we arbitrarily selected 17 ISE decamers (marked by asterisks in Fig. 1d and Supplementary Table 2) to examine their activity in the original screen reporter. After transiently transfecting the 293T cells, all tested ISE decamers promoted splicing of the retained intron to generate 20-50% of green cells, whereas the control cells were essentially non-fluorescent (Supplementary Fig. 1d), suggesting our screen had a very low false positive rate.

Validate ISEs in a heterologous context and cell type

This screen was conducted using a constant intron from a particular gene, consequently certain ISEs identified may require sequence context specific to this intron for function. In addition, most alternative splicing events involve cassette exons and it was of interest to determine if these ISEs can function in other contexts. To directly address this, we generated a new splicing reporter (pZW2C) containing a cassette exon with its flanking introns, and inserted nine FAS-ISE sequences in the downstream intron (18 nt downstream of 5'SS). We found all the tested ISE decamers increased cassette exon inclusion in 293T cells as compared to the neutral sequence (Fig. 1e upper panel), suggesting that these ISEs generally function in a heterologous intron.

Most SREs are thought to function through recruiting specific *trans*-factors, whose expression levels or activities may vary in different cell types^{5,6,34}. Thus the splicing

pattern of the same gene can change in different tissues. We further determined whether the recovered FAS-ISEs are active in another cell type. By transfecting HeLa cells with pZW2C reporters containing the same set of ISEs, we found that all tested ISEs led to a marked increase of exon inclusion (Fig. 1e lower panel). These results suggest that the systematically identified FAS-ISEs have general enhancer activity in different introns and in another cell type.

Identification and validation of core ISE motifs

To extract the core motifs with intrinsic ISE activities, we identified hexamers that are statistically overrepresented in the recovered FAS-ISE decamers. Each decamer was extended by appending 2 nt of the vector sequence and subsequently broken into overlapping hexamers¹³. The number of hexamers occurring at least three times in the extended ISE set was more than 4-fold higher than what expected from random decamer sets ($P < 10^{-4}$, based on 10,000 samplings of 109 random decamers). This hexamer set, named as ISE-hex3 (Supplementary Table 3), was highlighted by the AGGTAT and GGGTGG that occurred 15 and 14 times respectively. Furthermore, 93 out of the 109 ISE decamers were covered by at least one hexamer, suggesting the ISE-hex3 represented a common pattern in ISE decamers. Based on sequence similarity, these ISE hexamers were clustered and multiply aligned to identify candidate motifs^{11,13}. At a dissimilarity cutoff of 2.45, most hexamers fell into one of six main clusters (Fig. 2a, groups A to F). We identified the consensus motif of each group by aligning all hexamers of that group, and found that group D resembled known ISEs bound by hnRNP H while others appeared to be novel.

We next used three strategies to test the intrinsic ISE activity of these significantly overrepresented hexamers. First, we selected six hexamers resembling the consensus of each group (i.e. exemplars) and inserted them into pZW15C. The exemplars were selected to represent the most common pattern from the consensus motifs, except for group D where the second most common pattern was selected to avoid synthesizing a string of six G's. For controls we used mutants of the exemplars and neutral sequences not resembling known SREs. Upon transfection into 293T cells, all exemplars promoted intron splicing, whereas the splicing was barely detectable for the controls (Fig. 2b). In the second experiment, we examined whether each ISE group could promote splicing of a cassette exon from a heterologous intron. To increase the sensitivity, we inserted two tandem copies of exemplars from each group (Supplementary Table 4) at downstream of a cassette exon, and found all ISEs substantially increased the exon inclusion (Fig. 2c). Finally, we tested whether the ISE hexamers can promote splicing from upstream of a 3'SS despite their original identification near 5'SS, and found that all the six ISE groups considerably increased exon inclusion from upstream introns (Fig. 2d). We also found that these exemplars functioned in another cell type (HeLa), consistent with the finding that FAS-ISEs were active in different cell types (Supplementary Fig. 2).

Distribution and conservation patterns of FAS-ISEs

The FAS-ISEs had a base composition remarkably similar to the FAS-ESSs identified from an independent screen¹³. The core motifs of both ISEs and ESSs have high G/U content and low content of C (Fig. 3a), which is very different from the purine-rich RESCUE-ESEs¹¹

and the AU rich FAS-ISSs (Wang et. al, data not shown). We next examined the positional distribution of ISE-hex3 in human exons and associated introns (Fig. 3b). All ISE groups are substantially enriched in introns vs. exons and most groups peaked at upstream of the 3'SS or downstream of the 5'SS, resembling the characteristic distribution pattern of ESSs^{12,13}. Even excluding the G-rich elements that are common for both ISEs and ESSs, they still had similar distribution patterns (Supplementary Fig. 3a). When calculating the frequency near exons with alternative 5'SS or 3'SS, we found that most ISEs were enriched in the exonic extension region compared to the core regions, again resembling that of FAS-ESSs²² (Supplementary Fig. 3b).

We further analyzed the relative conservation of ISE-hex3 in different pre-mRNA regions using a previously developed scoring system²². The exons conserved in human and mouse genomes were extracted and classified into skipped exons (SEs), constitutive exons (CEs) and alternative 5'SS or 3'SS exons (A5Es or A3Es). For each hexamer set we computed their *P* value of the conservation rate, thus a smaller *P* value representing more conserved hexamers in particular region (Fig. 3c). Remarkably, the ISE set had a similar conservation pattern to ESSs: both were more conserved in exons than in introns and were significantly conserved in exonic extension regions of the A5Es or A3Es (Fig. 3c). Different ISE groups had distinct patterns: four groups (A, D, E and F) were conserved in the extension region of A5Es and at downstream of CEs (Fig. 3c), whereas the others (B and C) were more conserved around A3Es than A5Es. In addition, most groups were conserved inside SEs and some groups, to a lesser extent, were conserved inside CEs.

Systematic overlap between ISEs and ESSs

Consistent with the similar distribution and conservation patterns between ISEs and ESSs, 46 hexamers were common to the 84 ISE-hex3 and the 176 FAS-ESS hexamers¹³, 12 times more than expected by chance. To analyze the systematic overlap between the two sets, we used the sequential pattern mining to score the probability of a short element belonging to a common set (supplementary notes)³⁵. This method considers both the nucleotide frequency and the dependency of different positions, thus served as a better classifier of SREs than position weight matrix. For all possible hexamers, we computed the ISE and ESS scores based on the sequential feature of FAS-ISE and FAS-ESS hexamer sets (Fig 3d). The hexamers with large ISE (or ESS) scores indicated that they are more likely to function as ISEs (or ESSs). We found a strong positive correlation between the ISE scores and ESS scores for all hexamers, suggesting that these two classes of SREs identified through unrelated screens of random sequences overlap systematically.

All ISEs consistently inhibit splicing from exonic contexts

Previous study suggested that ESSs play critical roles in regulating splicing by suppressing pseudoexons and inhibiting intron-proximal 5'SS or 3'SS²². Based on the similarity between ISEs and ESSs, we predicted that the new FAS-ISEs might inhibit splicing in exons. To test this prediction, we inserted ISE exemplars as tandem copies into a cassette exon of a modular splicing reporter and found that, as expected, all ISE groups consistently inhibited inclusion of the cassette exon (Fig. 4a, Supplementary Table 4 and 5). When inserted between competing 5'SS, all ISE groups significantly inhibited the use of the

proximal site compared to controls (Fig. 4b). The 3' splice site (3'SS) choice was also controlled by ISEs in a similar fashion, with the same panel of representative ISEs causing a consistent inhibition of the proximal 3'SS usage (Fig. 4c). These results support a general rule of SRE activity: that sequences capable of enhancing exon inclusion from introns location usually have the ability to inhibit splicing from various exonic contexts. However, the opposite is not true as some sequences can inhibit splicing when inserted into either exons or introns (Wang et al. unpublished data). The context dependent activities for splicing regulation were previously observed for selected elements⁶, our analyses established a general rule for an entire class of SREs.

Identify putative *trans*-factors for each ISE group

Because all ISE groups functioned in heterologous contexts, they probably act by specifically recruiting *trans*-factors. Thus we sought to unbiasedly identify such factors using an RNA affinity purification method³⁶. A 5'-biotin labeled short RNA "bait" (20 nt) containing three copies of ISE exemplars was synthesized and incubated with the extract of HeLa cells where these ISEs were consistently active (Fig. 1e). The RNA-protein complexes were purified and proteins specifically bound to the RNA "bait" were identified by mass spectrometry (Supplementary Fig. 4a). From all groups, we excised 30 bands and identified 17 known or predicted RNA binding proteins (Supplementary Fig. 4b and Table 6). Most identified proteins have RNA recognition motifs (RRM) that specifically bind to single-stranded RNA, and several proteins contain Zinc fingers or DEAH box that are known to bind nucleic acids (Fig. 5a).

Remarkably, five groups (A to E) with distinct motifs were all bound by three major proteins around 52 kDa (Supplementary Fig. 4b), which were identified as hnRNP H1, F, and to a lesser extent GRSF1 (Fig. 5a). The binding of hnRNP H1 to the group D (G-runs) was expected based on previous studies^{26,27,29}, however it is surprising that these proteins were major factors to recognize diverse sequences (Fig. 2a). A recent CLIP-seq study also suggested that the binding specificity of hnRNP H/F may be rather promiscuous³⁷, consistent with our *in vitro* data using a SELEX-like methods (Dong et al, unpublished data). Alternatively, these factors can bind to different ISE motifs indirectly through other proteins. The finding that ISE groups A to E were strongly bound by common splicing factors hnRNP H1 and F suggested these factors may play a predominant role in promoting splicing from introns. The consensus motif of group F is very different from the other groups (Fig. 2a). Consistently, this group had a different protein interaction profile (Supplementary Fig. 4b) and was recognized by RNA-binding proteins including DAZAP1, RBM4 and hnRNP D0, some being demonstrated to regulate splicing^{38,39}.

HnRNP H and F enhance splicing by recognizing various ISEs

With a comprehensive list of putative *trans*-factors, we next examined if they were indeed responsible for the ISE activity. Since the ISEs in groups A to E were recognized by hnRNP H1 and F, we chose to test their function using splicing reporters containing cognate ISEs. We selected group B because its consensus motif was not previously shown to be recognized by hnRNP H or F⁴⁰. Over-expression of both hnRNP H1 and F promoted splicing of the intron inserted with the group B ISE, and co-expression of both proteins had

synergistic effect (Fig. 5b left panel and Supplementary Fig. 5a). Consistently, the RNAi of hnRNP H1 or F decreased the intron splicing (Fig. 5b and Supplementary Fig. 5b). As controls, expressions or RNAi of hnRNP H1 or F had no effect on splicing of the same intron containing a non-cognate ISE (group F), suggesting they can promote splicing by specifically binding to the cognate targets.

HnRNP H1 control splicing through its C-terminal domain

hnRNP H1 represents a homologous protein family including hnRNP H2, F, 2H9 and GRSF1. Members of this family have similar domain organization (Fig. 5c), including three qRRM domains to specifically recognize G-runs⁴¹ and two Gly/Tyr rich domains, with the exception of GRSF1 that lacks the Gly/Tyr rich domain. We next examined the functional domains of hnRNP H by fusing different fragments to a programmable RNA binding domain (PUF domain) and co-expressing with splicing reporters containing cognate targets (Fig. 5d and 5e). We selected three hnRNP H1 fragments covering the non-qRRM region (Fig. 5c) and tethered them at downstream of a cassette exon. All fragments enhanced exon splicing compared to the PUF domain alone (Fig. 5d), suggesting that the Gly/Tyr rich domains were the functional modules of hnRNP H1. As controls, the Gly-rich domain of hnRNP A1 and the RS domain of SRSF1 inhibited splicing, consistent with the findings that both proteins control splicing by recognizing ISSs (Wang et.al, unpublished data). In addition, the control fusion proteins with a non-cognate PUF domain had no effect.

Remarkably, the same hnRNP H1 functional domains can inhibit exon splicing when recruited to the exon (Fig. 5e), supporting the finding that its cognate targets can function as either ISEs or ESSs in different contexts (Fig. 2 and 4). As controls, the Gly-rich domain of hnRNP A1 inhibited splicing and the RS domain of SRSF1 enhanced splicing, and fusion proteins with non-cognate PUF domain had no effect (Fig. 5e). Taken together, we proved that the recruitment of Gly/Tyr rich domains of hnRNP H1 (even the last 85 aa) was sufficient to cause the context dependent activity of cognate ISEs, suggesting that hnRNP H1 functions in a modular fashion similar to the SR proteins or hnRNP A1⁴²⁻⁴⁴. The activities of these C-terminal domain are exactly opposite to that of RS domains⁴²⁻⁴⁴, which may represent a new mechanism to control splicing.

Multiple antagonistic factors recognize ISEs in group F

The ISEs in group F were recognized by at least three putative splicing factors, DAZAP1, RBM4 and hnRNP D0 (Fig. 5a). We over-expressed each factor with the splicing reporter containing the group F exemplar, and found that DAZAP1 strongly promoted splicing to increase the PSI (percent spliced in) from 20% to 68%, whereas RBM4 and hnRNP D0 had an opposite effect to inhibit exon inclusion (Fig. 6a and Supplementary Fig. 6a). The activities of these factors were dependent on the presence of cognate ISE, as they did not affect the same reporter containing control ISE (group B) (Fig. 6a). These results suggested that DAZAP1 is responsible for the splicing enhancer activity of ISEs in group F, whereas the other factors may antagonize DAZAP1 activity.

To examine whether the antagonistic factors can compete for the same element and switch its activity between ISE and ISS, we co-expressed DAZAP1 and RBM4 in different ratios

with the splicing reporters containing the cognate ISE (Fig. 6b). The RBM4:DAZAP1 ratio indeed determined the splicing outcome: higher ratio inhibited splicing whereas a lower ratio promoted inclusion of the cassette exon (Fig. 6b). As controls, RBM4:DAZAP1 ratio had no effect on the reporter containing a non-cognate ISE. We noticed that the more abundant protein seems to “overshadow” the other protein in co-expression experiments (e.g., Fig. 6b lane 3 vs. lane 5, likely due to an over-expression artifact), however this does not change the main conclusion. The binding of the same SRE by antagonistic factors enables a delicate switch where the splicing outcome can be very sensitive to the subtle change of splicing factor levels.

Previous studies showed that DAZAP1 inhibited splicing through interaction with hnRNP A1/A2⁴⁵ and RBM4 could either activate or inhibit splicing in different genes⁴⁶⁻⁴⁸. We searched the entire human genome for cassette exons with the group F ISEs in their adjacent introns and tested how DAZAP1 and RBM4 over-expression affect their splicing. To achieve a more consistent expression level, we generated stable cell lines using Flp-In™-293 T-Rex system where the expression of DAZAP1 or RBM4 can be induced by tetracycline (Supplementary Fig. 6b). We found that in three endogenous genes (COR06, SF1 and ANKS3) with the ISE sequence at down stream of a cassette exon (Supplementary Table 7), DAZAP1 expression indeed promoted the exon inclusion, whereas RBM4 had an opposite effect to inhibit splicing (Fig. 6c).

Antagonistic factors compete for same ISE in exonic context

Since the group F ISE inhibited splicing in exons (Fig. 4), we next examined how its cognate factors contribute to this activity. We co-expressed DAZAP1 or RBM4 with the splicing reporter that contains group F exemplar inside a cassette exon, and found that DAZAP1 promoted splicing whereas RBM4 had opposite activity to inhibit splicing (Fig. 6d, lanes 1 to 3). Over-expression of these factors had no effect on the reporter containing non-cognate ISE, suggesting such activities were due to specific recognition by DAZAP1 or RBM4. Co-expression of two proteins in different ratios changed the splicing outcome: higher RBM4:DAZAP1 ratio caused splicing inhibition whereas a decreased ratio promoted exon inclusion (Fig. 6d, lanes 4 to 6). Such effect was again likely due to the direct competition of these factors, because the RBM4:DAZAP1 ratio did not change splicing of the control reporter.

To determine if the same regulatory rule also applies to endogenous splicing, we searched endogenous human cassette exons containing the group F ISEs and tested how DAZAP1 and RBM4 affect their splicing. For three genes (ZBT17, LAMP1 and NOL8) with the ISE inside a cassette exon (Supplementary Table 7), induction of DAZAP1 and RBM4 had opposite effects with DAZAP1 promoting exon inclusion and RBM4 inhibiting splicing (Fig. 6e). When examining two non-small cells lung cancer lines (H157 and A549) with different RBM4 levels, we found that higher RBM4 level in A459 cells correlated with a reduced exon inclusion in most endogenous genes containing group F ISEs (Supplementary Fig. 6d and 6e), suggesting that the RBM4:DAZAP1 ratio can also explain some endogenous splicing variation in different cell lines.

The binding of the same SRE by antagonistic factors provided another model of the context dependent activity for some ISEs: DAZAP1 was the dominant factor when binding to the group F element in intron, leading to the ISE activity; however, the RBM4 outcompeted DAZAP1 when the same element was inserted into exons, resulting in the ESS activity. This model represents an exception to many *cis*-elements that have positional dependent activities by recruiting same *trans*-factor to different locations, and can probably be better described as a “factor dependent activity”.

Discussion

This study was initiated from an unbiased screen for novel ISEs and generated several key conclusions. First, diverse sequences can function as general ISEs in different cell types and contexts. Second, we observed systematic overlaps between ISEs and ESSs identified through independent screens, suggesting a common rule for the location-dependent activity. Third, we identified multiple factors that specifically bind ISE motifs to promote splicing from introns. Fourth, hnRNP H1 and F play predominant roles in enhancing splicing from introns, the recruitment of hnRNP H1 Gly/Tyr rich domains to different locations explained the location-dependent activity of ISEs. Fifth, a novel class of ISEs (group F) can be recognized by antagonistic factors. Competition of these factors in introns *vs.* exons produced different splicing outcomes and caused the context-dependent activity of this group. Finally, single ISE element can be bound by multiple factors with distinct activities and the same factor can recognize multiple ISEs, suggesting that a complicated web of RNA-protein interactions control splicing to achieve certain degree of regulatory plasticity. Taken together, this study provided an integrated model in which the general ISEs can be considered as “intron-define elements” that support the splicing pathway to exclude the pre-mRNA region where they are located. These ISEs usually function through recruiting cognate splicing factors, however we cannot rule out the possibilities that some new ISEs may affect splicing by other mechanism such as by slowing down transcription rate.

Compared to the tissue-specific ISEs predicted by conservation^{23,24} and the ISREs identified from experimental screens⁴⁹, motifs identified here have consistent activities among heterologous gene contexts and cell types. Since this screen was conducted in 293 cells, some known tissue specific ISEs (e.g. binding sites for Nova or Fox-1) were not recovered in our screen. The motifs whose cognate factors are expressed at low levels in 293T cells may also be missed (e.g. Fox-2). Similar strategies can be applied in the future to identify and study intronic SREs in other cell types. The ISEs identified in this study can promote splicing when located in either upstream or downstream intron, presenting a new trend that is distinct from some tissue specific SREs (e.g. Nova or Fox-1/2 binding site) that activate splicing from the downstream intron but repress splicing from upstream introns. This trend may reflect how general splicing factors interact with the core splicing machinery, which appears to be distinct from known tissue specific factors. Therefore motifs identified here have distinct features compared to previously reported motifs and significantly expanded the SRE repertoire.

We observed that multiple *trans*-factors can specifically recognize the same ISE and the distinct ISEs can be bound by the same protein (Supplementary Table 6). We further

confirmed the functional relevance of such interaction network (Fig. 5 and 6). Different ISEs could be recognized either by splicing factors with similar activity to generate a synergistic regulation (Fig. 5) or by antagonistic factors to produce a sensitive regulatory switch (Fig. 6). Consistent with the RNA-protein interaction network, the RRM-containing proteins often interact with RNA motifs with very short consensus and the binding motifs of splicing factors were very flexible as judged by different studies⁵⁰. The interaction network between *cis*-acting SREs and *trans*-factors provides a possible mechanism for regulatory plasticity. To understand the splicing code, new models considering the complex interaction network have to be implemented to simulate such plasticity. Such model will require the integration of information obtained from this study and other transcriptome-wide studies.

Online methods

Splicing reporter constructs

All of the splicing mini-gene reporters were modified from the same backbone construct, pZW1, which included multi-cloning sites between two GFP exons¹³. To construct the reporter for FAS-ISE screen, a retained intron - intron 4 of C7orf26 (RefSeq: NM_024067)- was amplified with a PCR reaction using primers containing *XhoI/ApaI* restriction sites and inserted into the downstream of the 5' splice site (5'SS) of the first GFP exon. The resulting reporter construct, pZW15, contained two exons and a retained intron () with a multicloning site at 21 nt downstream of exon 1. To increase the sensitivity of FAS-screen, we introduced additional mutations in the 5'SS to make the site stronger. The resulting minigene was subcloned into the site-specific integration plasmid, pcDNA5/FRT, by *NheI/BamHI* sites, generating the vector pZW15C that was stably transfected with pOG44 (in 1:9 ratio) into 293 FlpIn cell line.

To insert candidate ISE sequences and controls into pZW15C, we used a forward primer CACCTCGAG(N₆₋₁₂)GGGCCCCAC and reverse primer GTGGGGCCC(N₆₋₁₂)CTCGAGGTG, which contained the candidate sequences (designated N₆₋₁₂) flanked by *XhoI* and *ApaI* sites. The two primers were annealed, digested by *XhoI/ApaI*, and ligated into pZW15C pre-digested with *XhoI/ApaI*.

To make the random sequence library, we extended the foldback primer CACCTCGAG(N₁₀)GGGCCCCACACGTTTTTTTCGTGTGGGCC with Klenow, cut the resulting DNA with *XhoI* and *ApaI* and ligated into pZW15C¹³. The ligation product was used to transform ElectroMax DH-5 α , and we transformed sufficient numbers of *E. coli* cells to obtain ~2-fold coverage of the possible DNA decamers. The resulting library was transfected into 293FlpIn cells in 15 batches to obtain enough stably transfected clones (>10⁶ clones) that cover the entire decamer space.

To test ISE in a heterologous context, we constructed the reporter pZW2C by inserting exon 2 of Chinese hamster DHFR gene and portions of its flanking introns between the two GFP exons. This reporter was modified from pZW2 that was originally used in the FAS-ESS screen and contained an *XhoI/ApaI* restriction site inside the test exon 2¹³. The pZW2 was digested with *XhoI/ApaI* and filled in with an oligonucleotide (obtained by annealing primers 1 and 2, Supplementary Table 1) to destroy the exonic restriction sites. We then

introduced a new *XhoI/ApaI* restriction site at 18 nt downstream of the exon 2 by three consecutive PCR reactions. The resulting product was inserted into pZW2 digested with *NheI/PstI* to obtain the reporter pZW2B. To increase the ISE detection sensitivity, the pZW2C was further generated by weakening the 3'SS of exon 2 in pZW2B with site-directed mutagenesis so that the exon 2 was included in ~50% of mRNA in the absence of ISE.

To test ISE activity near the 3'SS of an alternative exon, we used a modular reporter pGZ3¹⁹ and inserted ISEs at 33-nt upstream of the 3'SS for the test exon (Exon 12 of the human IGF-II mRNA-binding protein 1, IGF2BP1). We amplified the first GFP exon together with the following IGF2BP1 intron using the pGZ3 as template and the primers containing different ISE sequences (forward primer 5 and reverse primers 7-13, Supplementary Table 1). The resulting fragments were digested with *HindIII/SacI* and inserted into the pGZ3 vectors digested with same restriction enzymes.

The reporters with competing 5'SS and 3'SS were described previously, and we inserted ISEs and control sequences by annealing primers containing target sequence and cognate restriction sites²². To test if ISEs can affect splicing when inserted into a skipped exon, we used the same modular splicing reporter pGZ3. The ISEs were inserted into this vector using *XhoI/ApaI* sites located inside the test exon¹⁹.

To determine the functional modules of hnRNP H, we employed the pCI-new vector (Promega) to express fusion protein as described before⁴⁴. Briefly, we started with an expression construct that encodes from N- to C-terminals, Flag epitope, Gly-rich domain of hnRNP A1 (residues 195-320 of NP_002127), and the MS2 coat protein (gift of Dr. R. Breathnach from Institut de Biologie-CHR). The fragment encoding the MS2 coat protein fragment was removed using *BamHI/SalI* digestion and replaced with a fragment encoding a NLS (PPKKKRKV) and the PUF domain of human Pumilio1, resulting the PUF-Gly(hnRNP A1). To make an expression construct for PUF-RS(SRSF1), we replaced the fragment encoding the Flag/Gly-rich domain with a fragment that encodes the RS domain of SRSF1 protein with an N-terminal Flag epitope. To make expression constructs for PUF fused hnRNP H1 fragments, the RS domain was removed with *XhoI/BamHI* digestion and replaced with different fragments of hnRNP H1 (A, 188-449aa; B, 188-289aa; C, 364-449aa). To generate splicing reporters containing target sequences of PUF domains, we synthesized and annealed oligonucleotides containing UGUAUGUA sequence flanked by *XhoI* and *ApaI* sites, digested with *XhoI/ApaI*, and inserted into the exon of splicing reporter pGZ3 or the intron of splicing reporter pZW2C.

Cell culture and transfection

293T cells and HeLa cells were cultured with DMEM medium containing 10% of FBS. The ISE library stably transfected 293 FlpIn cells were maintained in DMEM medium with hygromycin at a final concentration of 100 µg/ml. For stable transfection, the random library was co-transfected with pOG44, which encodes the recombinase Flp, into 293 FlpIn cells. The stable transfectants were selected as previously described¹³.

To generate stable cell line expressing DAZAP1 upon tetracycline induction, we used pCDNA5 FRT/TO vector and 293 FlpIn/T-Rex cells (Invitrogen). The FLAG-tagged full length DAZAP1 was cloned into the vector, and transfected with pOG44 in 1:9 ratio. The stably integrated cells were selected with 100 µg/ml hygromycin. One day before the induction, the cells were transferred to hygromycin-free medium. The inductions were carried out by adding tetracycline to a final concentration of 2 µg/ml. The induced cells were collected 48 hours after induction to extract RNA and protein for further analysis.

To overexpress *trans*-factors, cells were plated into 24-well plates 1 day before transfection. To determine the effect of over-expression of *trans*-factors on splicing changes, 0.2 µg of mini-gene reporters were co-transfected with 0.4 µg of protein *trans*-factors using lipofectamine 2000 according to the manual. After 48 hours, cells were collected for further RNA and protein analysis. To knock down *trans*-factors, cells were transfected with 50 nM siRNA (siGENOME SMART pools, Dharmacon) according to the manual. At 48 hours post-transfection, cells were transfected with the splicing reporters containing ISE exemplars or control. The cells were harvested after another 24 hours for further analysis.

Assay of splicing reporters with semi-quantitative RT-PCR

The splicing reporters (pZW15C, pZW2C, pGZ3, pGZ3-intron, pEZ1B and pEZ2F) inserted with different ISEs in intronic or exonic locations were transfected into cultured cells (293T or HeLa) and samples were collected 24 hours after transfection. The total RNA were isolated from transfected cells with TRIzol reagent (Invitrogen) according to the manufacturer's instructions, followed by 1 h DNase I (Invitrogen) treatment at 37 °C and then heat inactivation of DNase I. Total RNA (2 µg) was then reverse-transcribed with SuperScript III (Invitrogen) with poly T primer or gene specific primer (for GFP based 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). RT-PCR products were separated on 10% PAGE gels, and scanned with a Typhoon 9400 scanner (Amersham Biosciences). The amount of each splicing isoform was measured with ImageQuant 5.2. The primers used to amplify GFP based minigene reporters were AGTGCTTCAGCCGCTACCC for GFP exon 1 and GTTGTACTCCAGCTTGTGCC for exon 3.

Identification of *trans*-factors with RNA affinity purification approach

The RNA affinity purification method was adopted from the previously described protocol³⁶. For each biotin labeled ISE RNA sample, about 2.5×10^8 HeLa cells were collected and resuspended with 2.5 ml ice cold resuspension buffer (50mM Tris-HCl pH 8.0, 150 mM NaCl). 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 centrifuged at 12000 g for 20 min at 4°C. Then 0.75 nmol biotinylated RNA with two 18 atom spacers (Dharmacon) were added to the supernatants and incubated for 2 hrs at 4 °C. Next, 50 µl Streptavidin-agarose beads (Sigma) were added into the mixture and incubated for 2 hrs at 4 °C with slow rotation. The beads were washed 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 μ l final volume, and mixed with 10 μ l 5x 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 the further mass spectrometry analysis.

The interested bands that contained candidate protein *trans*-factors were cut and analyzed by ESI-MS/MS on a Q-ToF (Micromass) mass spectrometer. This analysis was conducted by the UNC Proteomic Center.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Drs. Jingyi Hui from Shanghai Institute of Biological Science and Woan-Yuh Tarn from Institute of Biomedical Sciences in Academia Sinica for providing the expression constructs, and Drs. Bill Marzluff and Chris Burge for critical reading of manuscripts. We thank Dr. Zbigniew Dominski for helping in RNA affinity purifications. This work is supported by an AHA grant (0865329E) and NIH grant (R01CA158283) to ZW.

References

1. Wang ET, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008; 456:470–6. [PubMed: 18978772]
2. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*. 2008; 40:1413–5. [PubMed: 18978789]
3. Castle JC, et al. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat Genet*. 2008; 40:1416–25. [PubMed: 18978788]
4. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*. 2003; 72:291–336. [PubMed: 12626338]
5. Matlin AJ, Clark F, Smith CW. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol*. 2005; 6:386–98. [PubMed: 15956978]
6. Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*. 2008; 14:802–13. [PubMed: 18369186]
7. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature*. 2010; 463:457–63. [PubMed: 20110989]
8. Barash Y, et al. Deciphering the splicing code. *Nature*. 2010; 465:53–9. [PubMed: 20445623]
9. Tian H, Kole R. Selection of novel exon recognition elements from a pool of random sequences. *Mol Cell Biol*. 1995; 15:6291–8. [PubMed: 7565782]
10. 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]
11. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. Predictive identification of exonic splicing enhancers in human genes. *Science*. 2002; 297:1007–13. [PubMed: 12114529]
12. Zhang XH, Chasin LA. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev*. 2004; 18:1241–50. [PubMed: 15145827]
13. Wang Z, et al. Systematic identification and analysis of exonic splicing silencers. *Cell*. 2004; 119:831–45. [PubMed: 15607979]
14. Stadler MB, et al. Inference of splicing regulatory activities by sequence neighborhood analysis. *PLoS Genet*. 2006; 2:e191. [PubMed: 17121466]
15. Goren A, et al. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol Cell*. 2006; 22:769–81. [PubMed: 16793546]

16. Yu Y, et al. Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. *Cell*. 2008; 135:1224–36. [PubMed: 19109894]
17. Ke S, et al. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res*. 2011; 21:1360–74. [PubMed: 21659425]
18. Fairbrother WG, Holste D, Burge CB, Sharp PA. Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol*. 2004; 2:E268. [PubMed: 15340491]
19. Xiao X, Wang Z, Jang M, Burge CB. Coevolutionary networks of splicing cis-regulatory elements. *Proc Natl Acad Sci U S A*. 2007; 104:18583–8. [PubMed: 17998536]
20. Lam BJ, Hertel KJ. A general role for splicing enhancers in exon definition. *RNA*. 2002; 8:1233–41. [PubMed: 12403462]
21. Ke S, Zhang XH, Chasin LA. Positive selection acting on splicing motifs reflects compensatory evolution. *Genome Res*. 2008; 18:533–43. [PubMed: 18204002]
22. 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]
23. Yeo GW, Nostrand EL, Liang TY. Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genet*. 2007; 3:e85. [PubMed: 17530930]
24. Voelker RB, Berglund JA. A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing. *Genome Res*. 2007; 17:1023–33. [PubMed: 17525134]
25. Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc Natl Acad Sci U S A*. 2005; 102:2850–5. [PubMed: 15708978]
26. 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]
27. McCullough AJ, Berget SM. G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. *Mol Cell Biol*. 1997; 17:4562–71. [PubMed: 9234714]
28. Wang E, Mueller WF, Hertel KJ, Cambi F. G Run-mediated recognition of proteolipid protein and DM20 5' splice sites by U1 small nuclear RNA is regulated by context and proximity to the splice site. *J Biol Chem*. 2011; 286:4059–71. [PubMed: 21127064]
29. Wang E, Dimova N, Cambi F. PLP/DM20 ratio is regulated by hnRNPH and F and a novel G-rich enhancer in oligodendrocytes. *Nucleic Acids Res*. 2007; 35:4164–78. [PubMed: 17567613]
30. Xiao X, et al. Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat Struct Mol Biol*. 2009; 16:1094–100. [PubMed: 19749754]
31. Ule J, et al. CLIP identifies Nova-regulated RNA networks in the brain. *Science*. 2003; 302:1212–5. [PubMed: 14615540]
32. Brudno M, et al. Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. *Nucleic Acids Res*. 2001; 29:2338–48. [PubMed: 11376152]
33. Underwood JG, Boutz PL, Dougherty JD, Stoilov P, Black DL. Homologues of the *Caenorhabditis elegans* Fox-1 protein are neuronal splicing regulators in mammals. *Mol Cell Biol*. 2005; 25:10005–16. [PubMed: 16260614]
34. Blencowe BJ. Alternative splicing: new insights from global analyses. *Cell*. 2006; 126:37–47. [PubMed: 16839875]
35. Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM-SIGMOD international conference on management of data (SIGMOD'93)*. 1993:207–216.
36. 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]
37. Huelga SC, et al. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep*. 2012; 1:167–178. [PubMed: 22574288]

38. Pastor T, Pagani F. Interaction of hnRNPA1/A2 and DAZAP1 with an Alu-derived intronic splicing enhancer regulates ATM aberrant splicing. *PLoS One*. 2011; 6:e23349. [PubMed: 21858080]
39. Lai MC, Kuo HW, Chang WC, Tarn WY. A novel splicing regulator shares a nuclear import pathway with SR proteins. *EMBO J*. 2003; 22:1359–69. [PubMed: 12628928]
40. 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]
41. Dominguez C, Fiset JF, Chabot B, Allain FH. Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nat Struct Mol Biol*. 2010; 17:853–61. [PubMed: 20526337]
42. Graveley BR, Maniatis T. Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol Cell*. 1998; 1:765–71. [PubMed: 9660960]
43. 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; 19:251–60. [PubMed: 9858549]
44. Wang Y, Cheong CG, Hall TM, Wang Z. Engineering splicing factors with designed specificities. *Nat Methods*. 2009; 6:825–30. [PubMed: 19801992]
45. Goina E, Skoko N, Pagani F. Binding of DAZAP1 and hnRNPA1/A2 to an exonic splicing silencer in a natural BRCA1 exon 18 mutant. *Mol Cell Biol*. 2008; 28:3850–60. [PubMed: 18391021]
46. Kar A, Havlioglu N, Tarn WY, Wu JY. RBM4 interacts with an intronic element and stimulates tau exon 10 inclusion. *J Biol Chem*. 2006; 281:24479–88. [PubMed: 16777844]
47. Markus MA, et al. WT1 interacts with the splicing protein RBM4 and regulates its ability to modulate alternative splicing in vivo. *Exp Cell Res*. 2006; 312:3379–88. [PubMed: 16934801]
48. Lin JC, Tarn WY. Exon selection in alpha-tropomyosin mRNA is regulated by the antagonistic action of RBM4 and PTB. *Mol Cell Biol*. 2005; 25:10111–21. [PubMed: 16260624]
49. 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
50. Cook KB, Kazan H, Zuberi K, Morris Q, Hughes TR. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res*. 2011; 39:D301–8. [PubMed: 21036867]

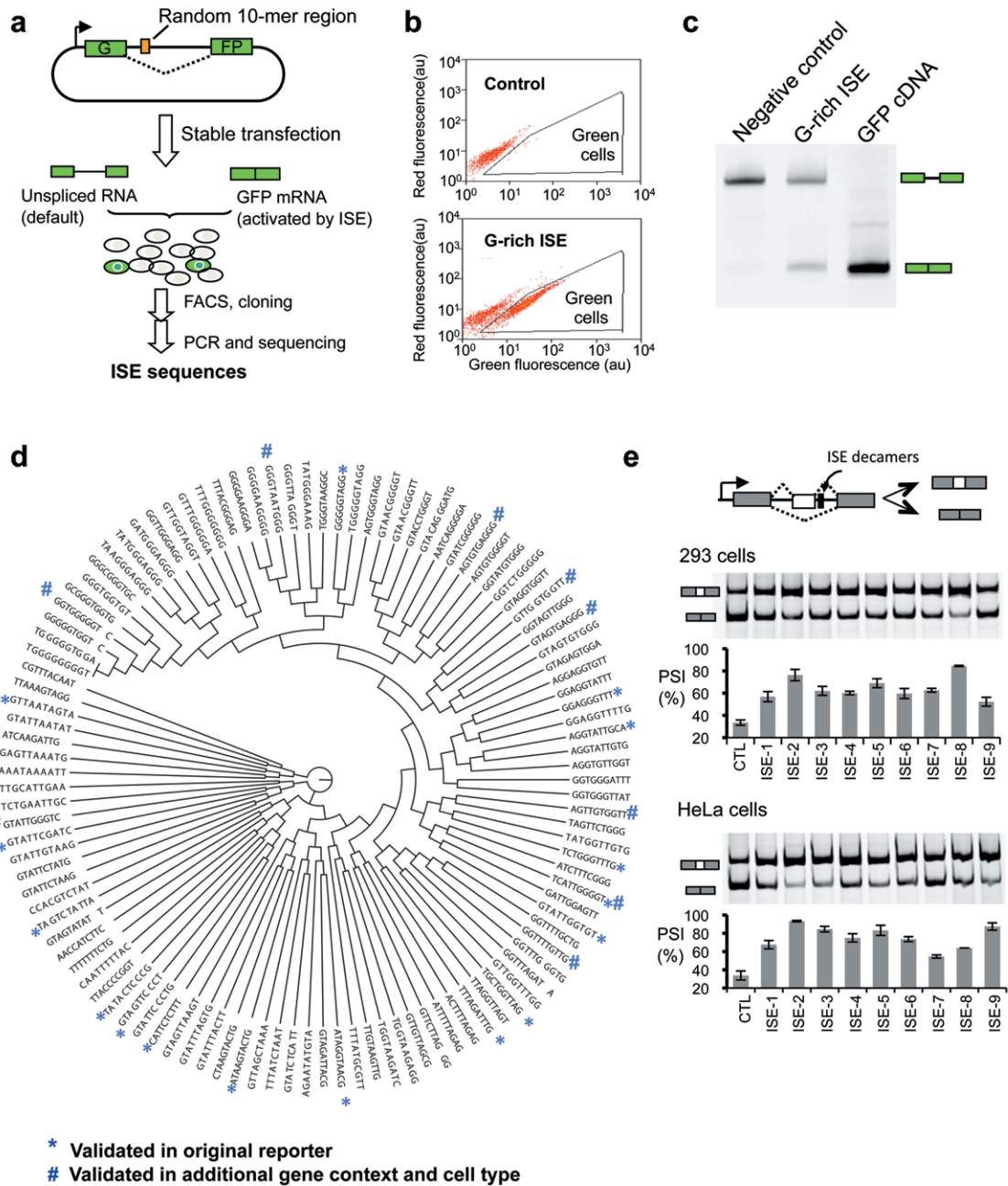


Figure 1. Identification and validation of ISEs

(a) Schematic diagram of screen strategy. (b) Test of ISE reporter system. A G-rich ISE sequence (TGGGGGGAGG) and a control sequence (CTCAAGCTTC) were inserted into the pZW15C reporter and transfected into 293T cells. Splicing of the retained intron was assayed with flow cytometry as percentage of green cells. Both red and green fluorescence intensity (in arbitrary units) were measured to correct for self-fluorescence background. (c) Validation of splicing reporter with semi-quantitative RT-PCR. The experiments were carried out as panel B, and the GFP expression vector was used as a positive control. (d) 109 unique ISE decamers were identified through FAS-ISE screen, which were clustered with

ClustalW2. The activities of selected ISE decamers were validated in the original screening reporter (asterisks) or in a heterologous gene context (indicated by #). (e) Testing FAS-ISEs in a heterologous intron and a different cell type. Nine FAS-ISE decamers (Supplementary Table 2) were randomly selected and inserted at downstream of a cassette exon. The reporters (pZW2C) were transiently transfected into 293T or HeLa cells to analyze splicing by semi-quantitative RT-PCR. At least two independent experiments were carried out for each sample, and the means and ranges of PSI (percent-spliced-in) values were plotted below representative gels.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

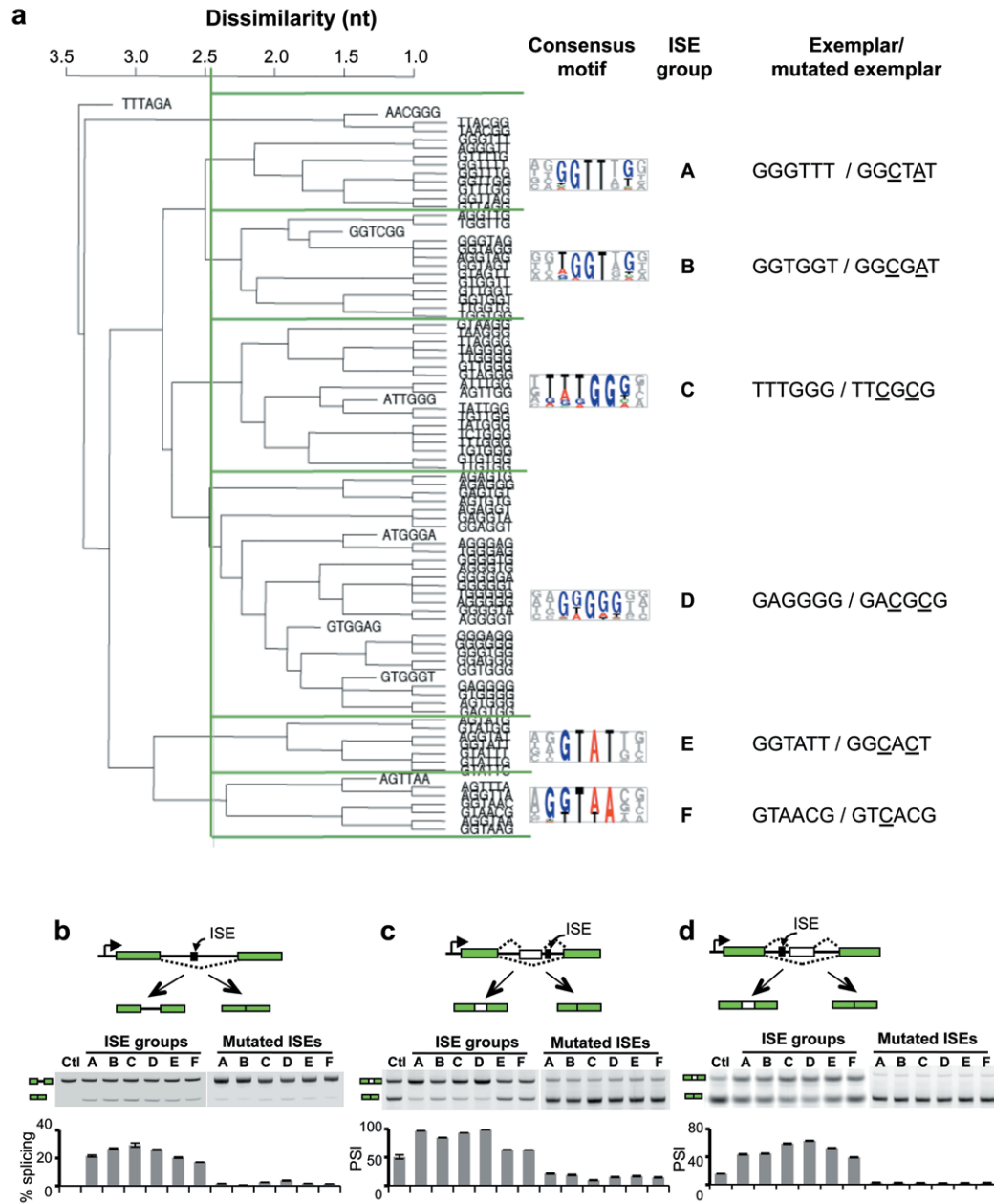


Figure 2. Identification and validation of the core motifs of FAS-ISEs

(a) Over-represented hexamers clustered by sequence similarity. We used an arbitrary cut-off of 2.45 in dissimilarity to generate six ISE groups that were labeled as group A to F. The hexamers in each group were aligned to obtain consensus motifs, and a representative hexamer (exemplar) and its mutant were selected in each group for further validations. (b) The representative exemplars and their mutants were inserted into the intron of pZW15C and transfected into 293T cells to examine the splicing efficiencies. In all panels, the experiments were repeated at least twice with means and ranges of splicing efficiency plotted under the representative gel. (c) All ISE groups enhanced splicing from downstream of a cassette exon. Two copies of the exemplars and their mutants were inserted in the

pZW2C reporter at the downstream of a cassette exon to examine their effects on splicing. **(d)** All ISE groups enhanced splicing from the upstream region of a cassette exon. Single copy of representative exemplars and mutants were inserted into the upstream of a cassette exon in a modular reporter ¹⁹.

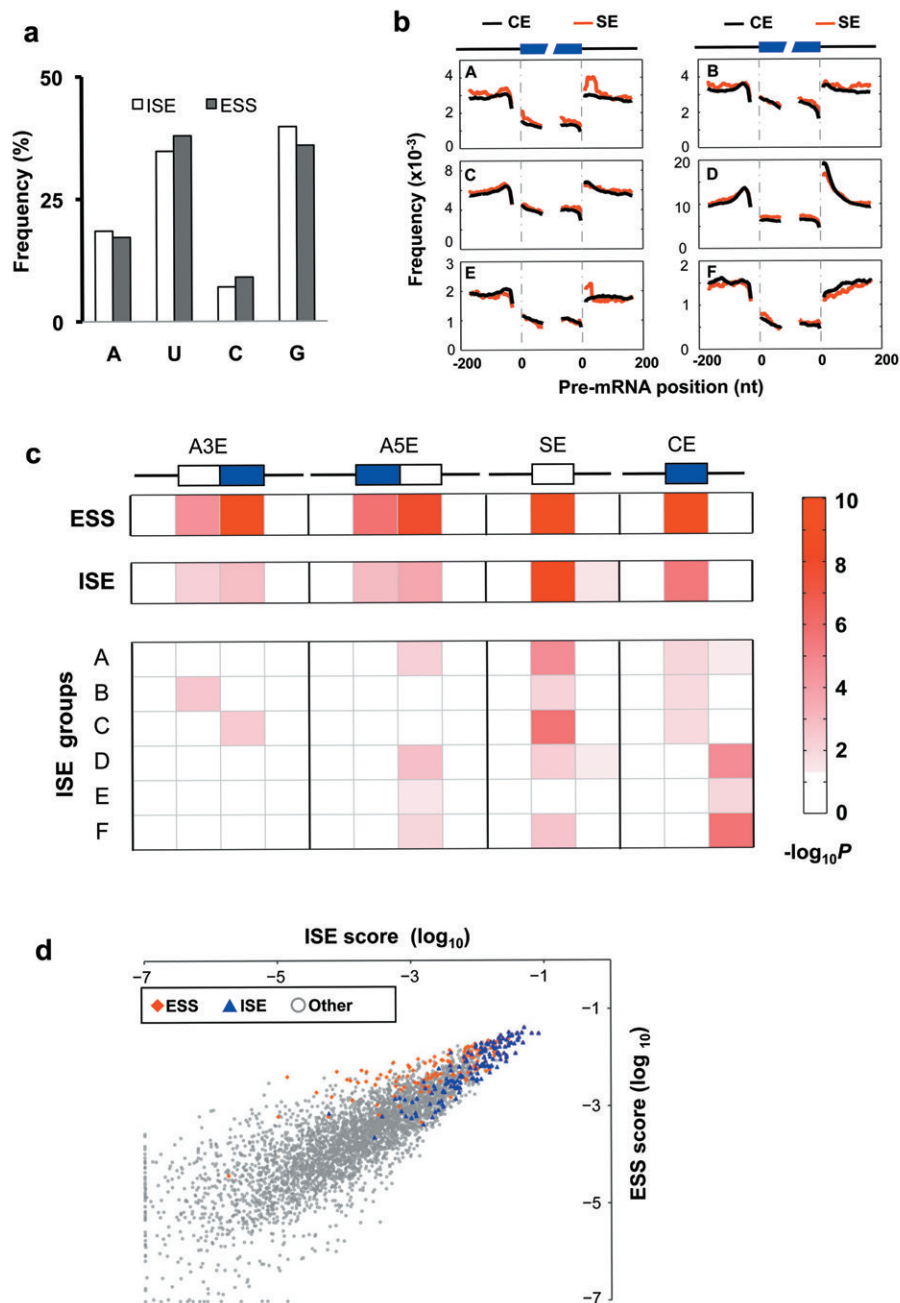


Figure 3. Similarity between FAS-ISE and FAS-ESSs

(a) Comparison of single nucleotide compositions between FAS-ISE and FAS-ESS hexamers. (b) The positional frequency of FAS-ISE. The FAS-ISE hexamers in each group were counted near the constitutive exons (CEs) and skipped exons (SEs), and their frequency was plotted as the number of ISE hexamers divided by total number of hexamers at each position. The first and last 100 bases of exons and the first and last 200 bases of introns were shown. (c) Conservation patterns of ISE and ESS hexamers near SEs, CEs, alternative 3'SS exons (A3Es) and alternative 5'SS exons (A5Es). The degrees of

conservation were calculated as the P values for the COR statistic²² and displayed in a heat map. **(d)** Comparison between ISE and ESS using a score system derived from sequential pattern mining. 199 ISE-hex2 hexamers (blue triangles) and 176 FAS-hex2 ESS hexamers (brown diamonds) were used as training sets to calculate the sequential feature scores, and the ISE and ESS scores of all 4096 hexamers were plotted in a scatter chart.

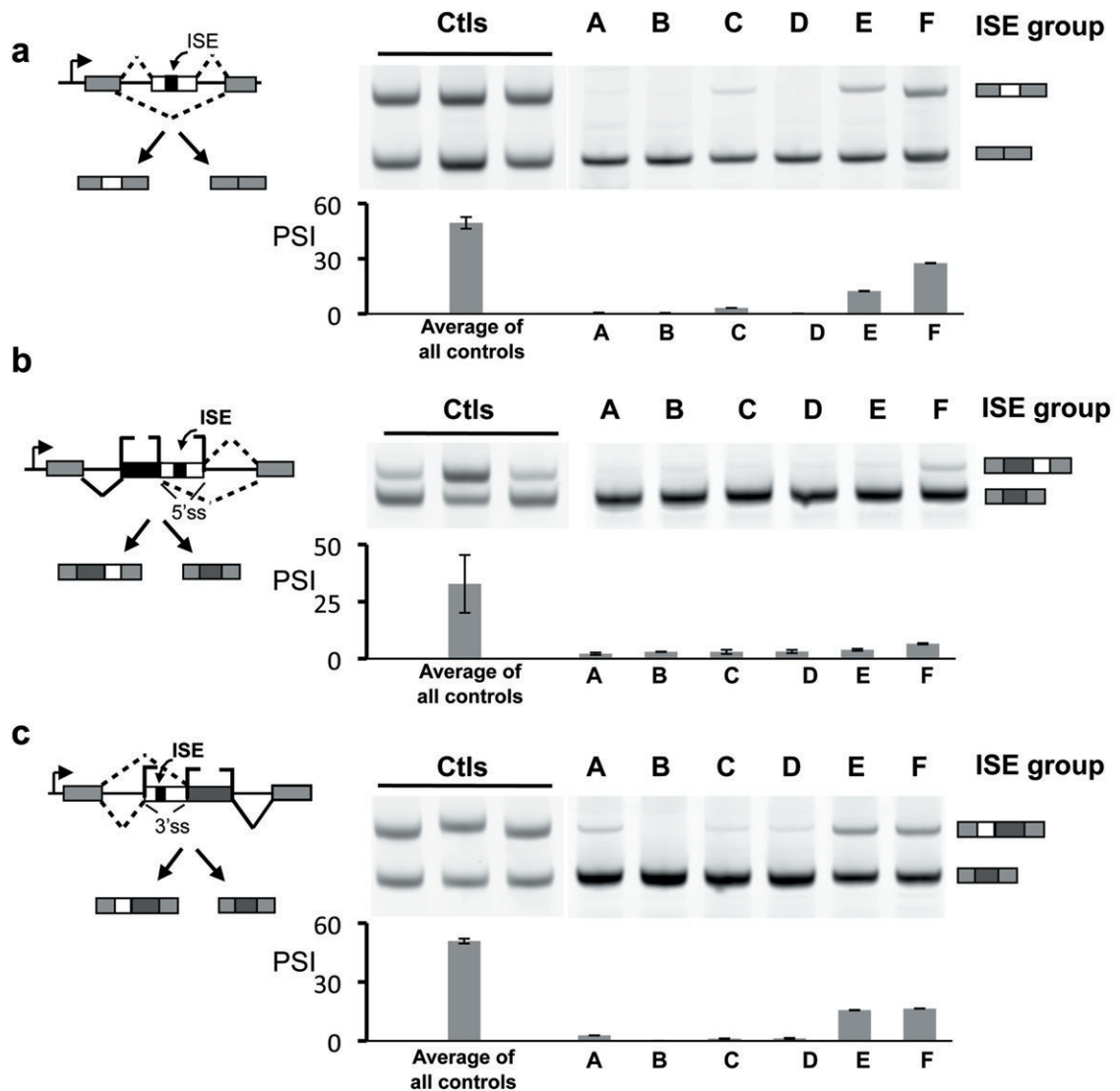


Figure 4. FAS-ISEs function as splicing silencers in exonic contexts

(a) FAS-ISEs inhibited cassette exon inclusion. Two copies of the exemplars in each group were inserted into the cassette exon of a modular reporter¹⁹, and transfected into 293T cells to assay for splicing. Inclusions of the cassette exon were determined using semi-quantitative RT-PCR. In all panels, two independent experiments were carried out and the means and ranges of PSI were measured and plotted below a representative gel. The standard deviation was used as the error bar in controls. (b) FAS-ISEs inhibited the use of downstream 5'SS when inserted between two competing sites in an A5E. The splicing reporters were transfected into 293T cells, and the usage of alternative sites was assayed by RT-PCR at 24 hours after transfection. (c) FAS-ISEs inhibited the usage of upstream 3'SS when inserted between two competing sites in an A3E. The splicing reporters were transiently transfected into 293T cells, and the use of alternative 3'SS was assayed by RT-PCR.

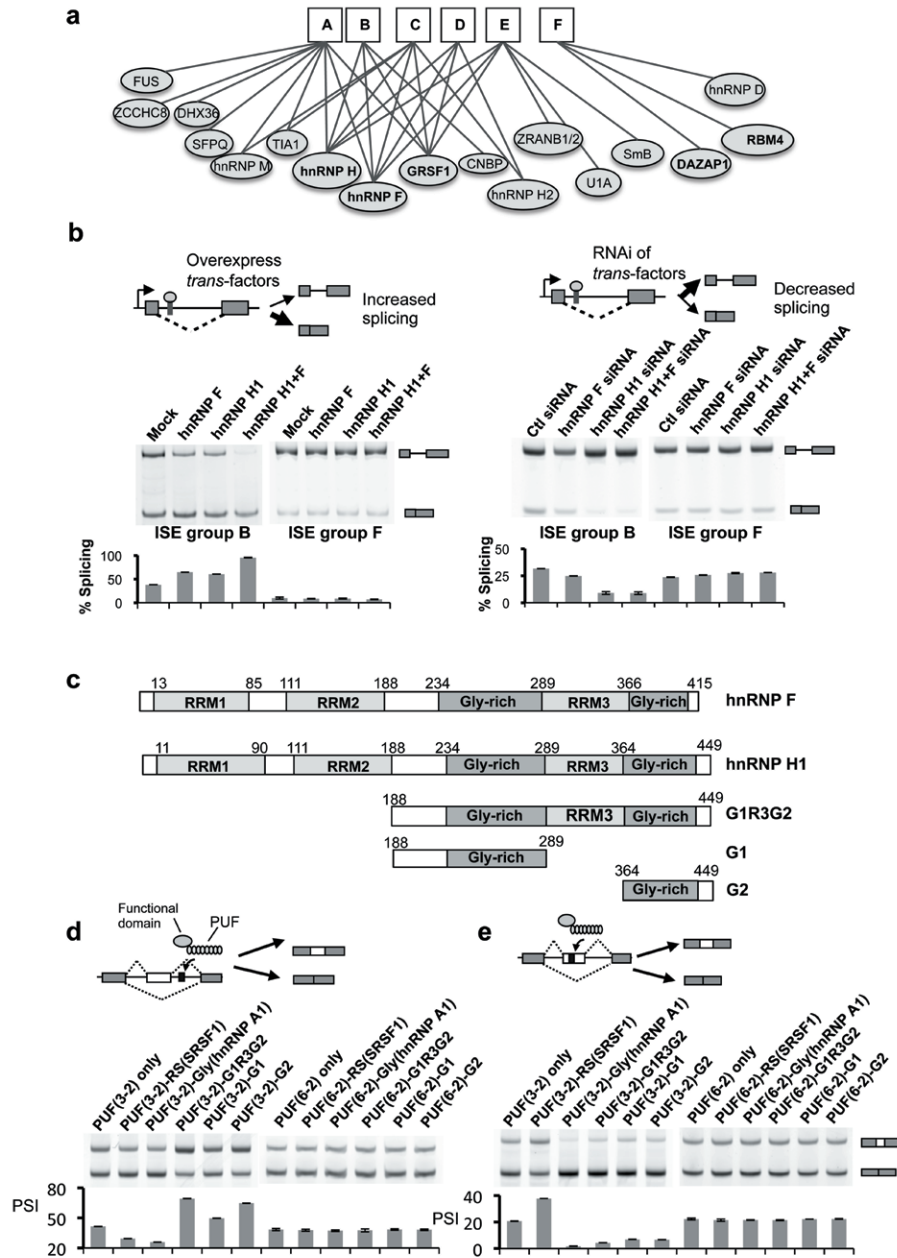


Figure 5. HnRNP H and F recognize distinct ISE groups to control splicing through C-terminal domains

(a) The connectivity map representing interactions of putative splicing factors to all ISE groups. (b) Over-expression and RNAi of hnRNP H1 and F specifically affected splicing of an intron containing the cognate ISE (group B). Left, the splicing reporters were co-expressed with *trans*-factors. Right, the splicing reporters were transfected 48 hours after RNAi of *trans*-factors. The splicing patterns were measured as in left panel. In both experiments, a reporter containing group F ISE was used as a control. (c) Schematics of the domains in hnRNP F and H1. (d) Different fragments of hnRNP H1 and the control domains were fused to a PUF domain, PUF(3-2) or PUF(6-2), to generate fusion proteins that

specifically bind to the target RNAs. The fusion proteins were co-transfected with a splicing reporter containing PUF(3-2) binding site at downstream of a cassette exon. We used the PUF(6-2) fusion proteins as specificity control. **(e)** The same set of fusion proteins were co-expressed with a reporter containing cognate targets within a cassette exon. In all panels, at least two independent experiments were conducted, and means and ranges of splicing rate were plotted below representative gels.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

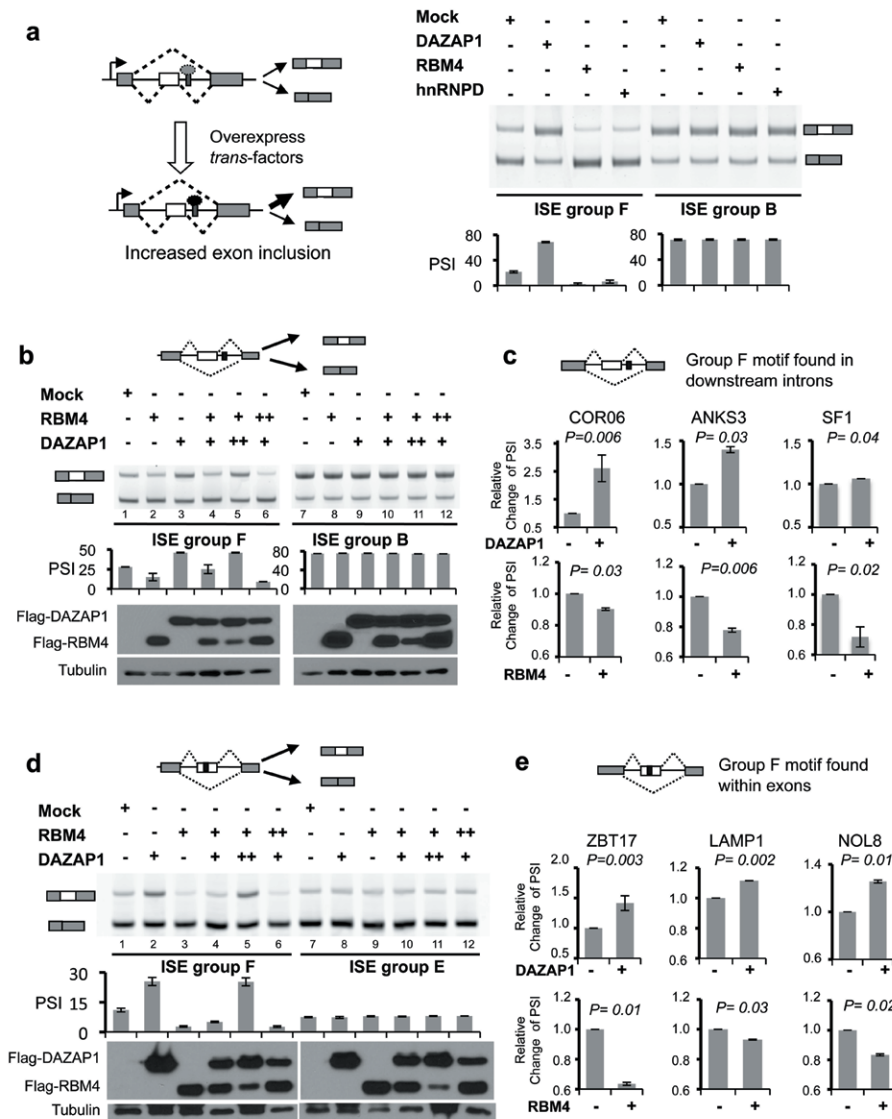


Figure 6. Multiple factors with antagonistic activities bind ISE group F to control splicing
(a) Over-expression of DAZAP1, RBM4 and hnRNP D0 affect splicing of a cassette exon containing the group F ISE at its downstream. The group B was used as a specificity control. The means and ranges of PSI from two independent experiments were plotted. **(b)** DAZAP1 and RBM4 antagonistically regulate splicing when binding to same intronic target. 293T cells were co-transfected with different amounts of RBM4 and DAZAP1 expression constructs (+, 0.2 μ g; ++, 1 μ g) and a reporter containing ISE exemplars at the downstream of a cassette exon. **(c)** DAZAP1 and RBM4 antagonistically control splicing of endogenous exons containing cognate ISEs in adjacent introns. Stable cell lines were induced with tetracycline to express DAZAP1 or RBM4. P values were calculated with the paired T test ($n=3$). **(d)** DAZAP1 and RBM4 antagonistically regulate splicing when binding to an exonic target. Different amounts of RBM4 and DAZAP1 constructs (+, 0.2 μ g; ++, 1 μ g) were co-transfected with a reporter containing the ISE exemplars inside a cassette exon.

Experimental conditions are same as panel b. (e) DAZAP1 and RBM4 antagonistically control splicing of endogenous cassette exons containing the group F ISE. Experiments and analyses were similar to the panel c.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript