Regolazione dello splicing alternativo (AS)

Nelle lezioni della parte 2.1 abbiamo imparato:

-risultati di analisi con deep-sequencing → forse fino a 98% dei geni con AS (hu)

-AS diviene sempre più importante nell'evoluzione verso i Vertebrati

-meccanismo biochimico dello splicing/spliceosoma e proteine associate

-sequenze che definiscono confine esone/introne e sequenze introniche

-modelli di AS (exon skipping, alternative 5'/ 3', mutually excl., etc.)

-Alternative TSS, poly(A) signals, frequenze di uso alternativo di esoni.

-Esempi funzionali di AS (compreso il gene Dscam di D. melanogaster)

-Studi genome-wide con microarrays esonici

-Studi genome-wide con microarrays exon-junction

nature

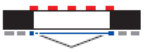
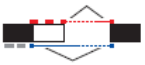
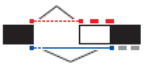
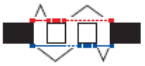
Vol 456|27 November 2008|doi:10.1038/nature07509

ARTICLES

Alternative isoform regulation in human tissue transcriptomes

Eric T. Wang[1,2]*, Rickard Sandberg[1,3]*, Shujun Luo[4], Irina Khrebtukova[4], Lu Zhang[4], Christine Mayr[5], Stephen F. Kingsmore[6], Gary P. Schroth[4] & Christopher B. Burge[1]

Through alternative processing of pre-messenger RNAs, individual mammalian genes often produce multiple mRNA and protein isoforms that may have related, distinct or even opposing functions. Here we report an in-depth analysis of 15 diverse human tissue and cell line transcriptomes on the basis of deep sequencing of complementary DNA fragments, yielding a digital inventory of gene and mRNA isoform expression. Analyses in which sequence reads are mapped to exon–exon junctions indicated that 92–94% of human genes undergo alternative splicing, ~86% with a minor isoform frequency of 15% or more. Differences in isoform-specific read densities indicated that most alternative splicing and alternative cleavage and polyadenylation events vary between tissues, whereas variation between individuals was approximately twofold to threefold less common. Extreme or 'switch-like' regulation of splicing between tissues was associated with increased sequence conservation in regulatory regions and with generation of full-length open reading frames. Patterns of alternative splicing and alternative cleavage and polyadenylation were strongly correlated across tissues, suggesting coordinated regulation of these processes, and sequence conservation of a subset of known regulatory motifs in both alternative introns and 3′ untranslated regions suggested common involvement of specific factors in tissue-level regulation of both splicing and polyadenylation.

Alternative transcript events		Total events ($\times 10^3$)	Number detected ($\times 10^3$)	Both isoforms detected	Number tissue-regulated	% Tissue-regulated (observed)	% Tissue-regulated (estimated)
Skipped exon		37	35	10,436	6,822	65	72
Retained intron		1	1	167	96	57	71
Alternative 5′ splice site (A5SS)		15	15	2,168	1,386	64	72
Alternative 3′ splice site (A3SS)		17	16	4,181	2,655	64	74
Mutually exclusive exon (MXE)		4	4	167	95	57	66
Alternative first exon (AFE)		14	13	10,281	5,311	52	63
Alternative last exon (ALE)		9	8	5,246	2,491	47	52
Tandem 3′ UTRs		7	7	5,136	3,801	74	80
Total		105	100	37,782	22,657	60	68

Constitutive exon or region — Body read — Junction read — pA Polyadenylation site
Alternative exon or extension — Inclusive/extended isoform — Exclusive isoform — Both isoforms

UNDERSTANDING ALTERNATIVE SPLICING: TOWARDS A CELLULAR CODE

Arianne J. Matlin‡, Francis Clark and Christopher W. J. Smith‡*

Abstract | In violation of the 'one gene, one polypeptide' rule, alternative splicing allows individual genes to produce multiple protein isoforms — thereby playing a central part in generating complex proteomes. Alternative splicing also has a largely hidden function in quantitative gene control, by targeting RNAs for nonsense-mediated decay. Traditional gene-by-gene investigations of alternative splicing mechanisms are now being complemented by global approaches. These promise to reveal details of the nature and operation of cellular codes that are constituted by combinations of regulatory elements in pre-mRNA substrates and by cellular complements of splicing regulators, which together determine regulated splicing pathways.

Nature Rev Mol Cell Biol (2005) 6:386.

Review (27/11/09)

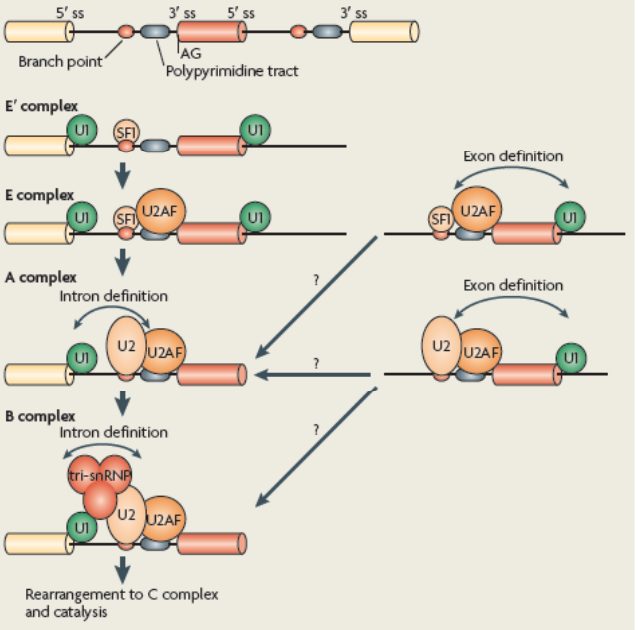
Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches

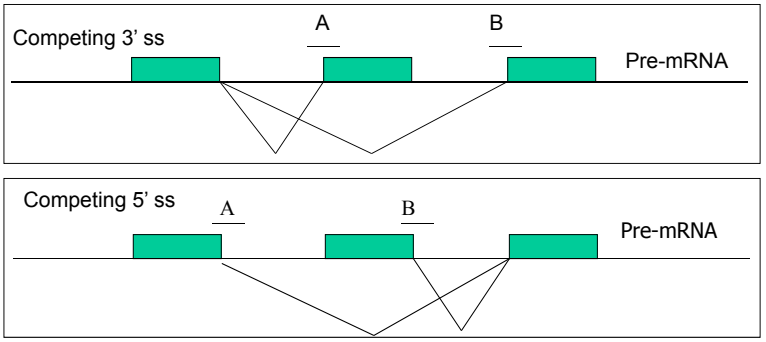
Mo Chen and James L. Manley

Abstract | Alternative splicing of mRNA precursors provides an important means of genetic control and is a crucial step in the expression of most genes. Alternative splicing markedly affects human development, and its misregulation underlies many human diseases. Although the mechanisms of alternative splicing have been studied extensively, until the past few years we had not begun to realize fully the diversity and complexity of alternative splicing regulation by an intricate protein–RNA network. Great progress has been made by studying individual transcripts and through genome-wide approaches, which together provide a better picture of the mechanistic regulation of alternative pre-mRNA splicing.

review

Box 1 | **Splicing and spliceosome assembly**





First point: What does mean "strong" or "weak" splice site?

a) Of course the degree of complementarity to the RNA U comes first

b) Are there additional sequences that contribute strength to the machinery ? Are there specific proteins?

1) strength of RNA-RNA interaction

In **yeast** there is a clear correlation of 5'-ss sequence with usage:

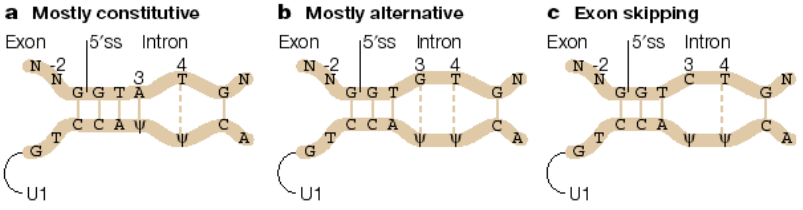


Figure 4 | **Base paring between different types of 5′ss and U1 snRNA.** Positions 3 and 4 of the 5′ss are different between the panels; stacking energy is illustrated by distances between nucleotides that base pair with U1 snRNA. Solid and dashed lines indicate Watson–Crick and non-Watson–Crick pairing, respectively. 'N' indicates unspecified nucleotides, Ψ indicates pseudo-uridine. **a** | Mostly constitutive. **b** | Mostly alternative. **c** | Exon skipping.

From: Ast G. (2004)
Nature Rev Genetics 5: 773.

....but vertebrates do not:

5'-ss has major differences.

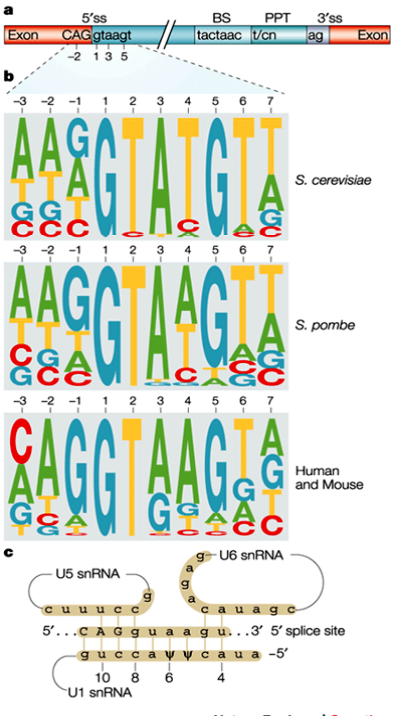
S.cerevisiae introns have 6 nt well conserved: 5'-GTATGT-3'

the -1 "G" increases in frequency from S.cerevisiae (37%) to humans (80%), positions +1 to +6 degenerate a little.

BS (branch site) in S. cerevisiae is very conserved (5'-UACUAAC-3') in S. pombe is much more variable, as in mammals (5'-CURAY-3')

Polypyrimidine tract is also variable, as well as the distance between BS and 3'-ss. (distance very short in S. pombe).

From: Ast G. (2004)
"How did alternative splicing evolve?"
Nature Rev Gen 5: 773-782.



Functional recognition of the 3′ splice site AG by the splicing factor U2AF^{35}

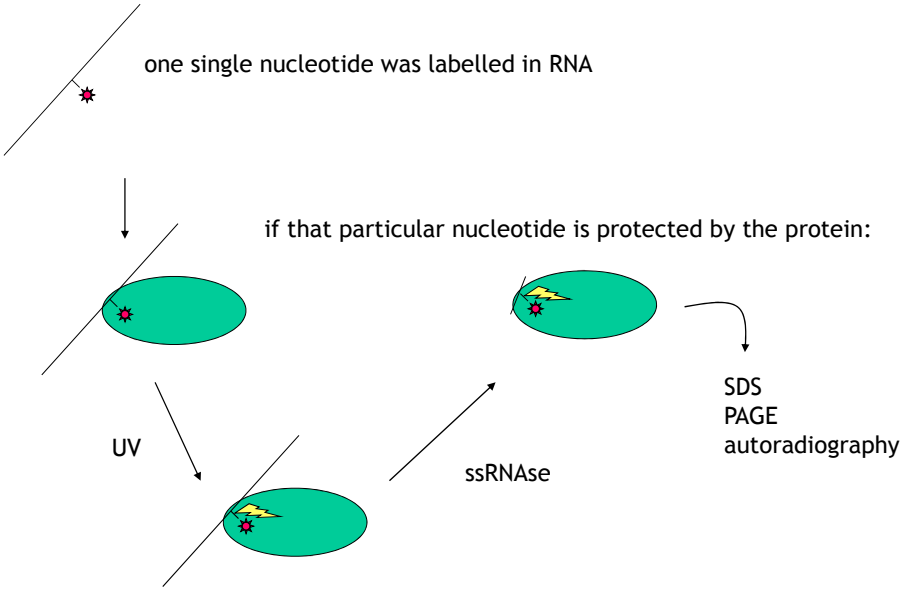
Shaoping Wu*, Charles M. Romfo†, Timothy W. Nilsen† & Michael R. Green*

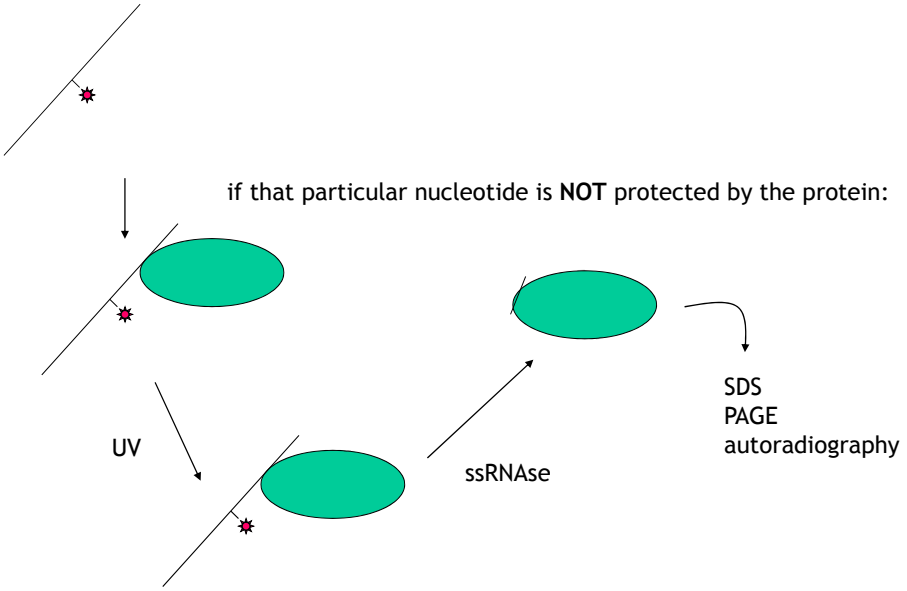
** Howard Hughes Medical Institute, Program in Molecular Medicine, University of Massachusetts Medical Center, 373 Plantation Street, Worcester, Massachusetts 01605, USA*
† Center for RNA Molecular Biology, Department of Molecular Biology and Microbiology, Case Western Reserve University, School of Medicine, 10,900 Euclid Avenue, Cleveland, Ohio 44106, USA

In metazoans, spliceosome assembly is initiated through recognition of the 5′ splice site by U1 snRNP and the polypyrimidine tract by the U2 small nuclear ribonucleoprotein particle (snRNP) auxiliary factor, U2AF (refs 1, 2). U2AF is a heterodimer comprising a large subunit, U2AF^{65}, and a small subunit, U2AF^{35} (ref. 3). U2AF^{65} directly contacts the polypyrimidine tract and is required for splicing *in vitro*[4]. In comparison, the role of U2AF^{35} has been puzzling: U2AF^{35} is highly conserved[5–7] and is required for viability[6,7], but can be dispensed with for splicing *in vitro*[4,8,9]. Here we use site-specific crosslinking to show that very early during spliceosome assembly U2AF^{35} directly contacts the 3′ splice site. Mutational analysis and *in vitro* genetic selection indicate that U2AF^{35} has a sequence-specific RNA-binding activity that recognizes the 3′-splice-site consensus, AG/G. We show that for introns with weak polypyrimidine tracts, the U2AF^{35}–3′-splice-site interaction is critical for U2AF binding and splicing. Our results demonstrate a new biochemical activity of U2AF^{35}, identify the factor that initially recognizes the 3′ splice site, and explain why the AG dinucleotide is required for the first step of splicing for some but not all introns.

In 3'-SS a clear consensus sequence in addition to the AG(G) rule is not present, as well: in this papers authors tried to see whether U2AF35 has any additional sequence preference in addition to strict 3'-splice site

1) they determined the footprint of U2AF35 over the splice site
2) they run RNA- SELEX to see sequence preference by U2AF35


one single nucleotide was labelled in RNA
if that particular nucleotide is protected by the protein:
UV
ssRNAse
SDS
PAGE
autoradiography


if that particular nucleotide is **NOT** protected by the protein:
UV
ssRNAse
SDS
PAGE
autoradiography

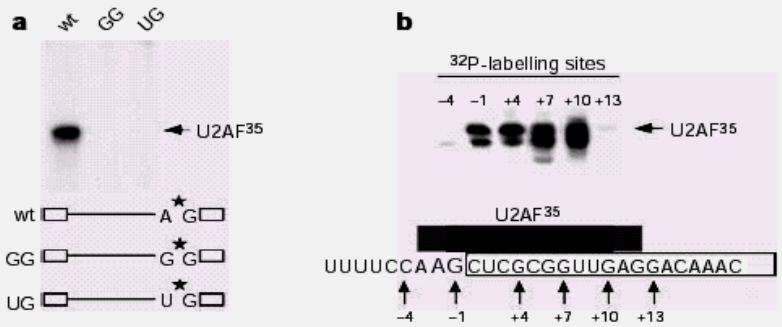
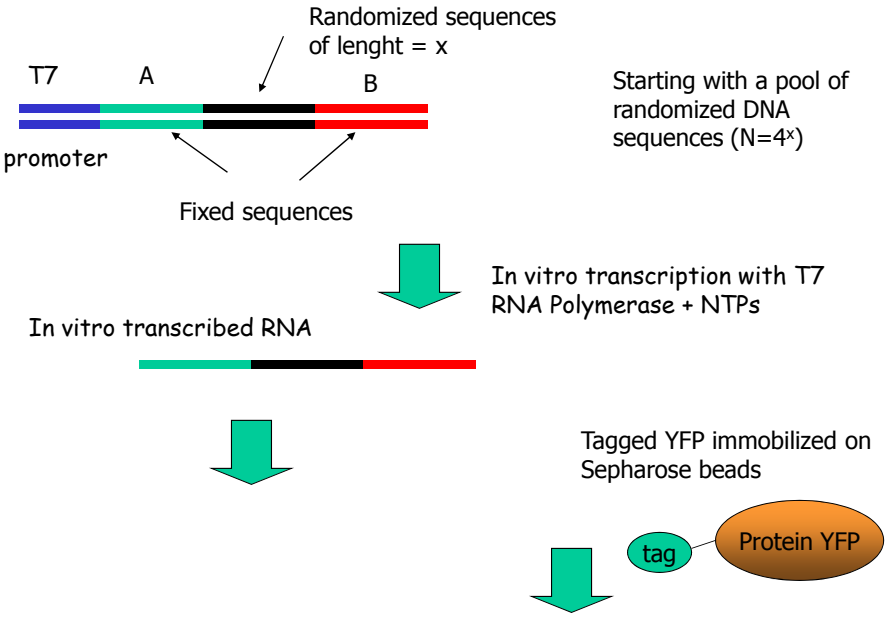


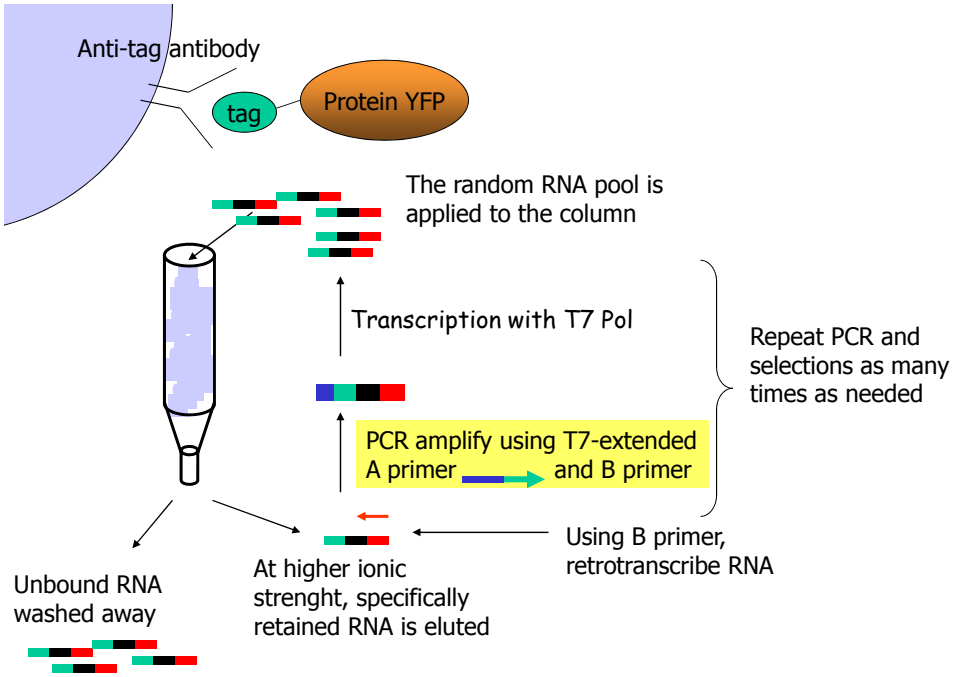
Figure 3 Boundaries and AG-dinucleotide requirement for the U2AF^{35}–3′-splice-site interaction. **a**, AG-dinucleotide requirement for U2AF^{35}–3′-splice-site interaction. UV crosslinking as in Fig. 1c with U2AF from HeLa and the indicated pre-mRNAs. wt, wild type. Diagram shows the mutations and the position of the site-specific ^{32}P label (asterisk). **b**, Boundaries of U2AF^{35} interaction. UV crosslinking as in Fig. 1a, except that a series of site-specifically labelled RNAs containing the ^{32}P label at the positions indicated by arrows were used. The AG dinucleotide at the 3′ splice site is bold and the exon sequences are in boxes. The minimal (black) and maximal (grey) boundaries of the U2AF^{35} interaction site are indicated.

from Wu et al. (1999), Nature 402: 832

In vitro evolution of molecules (RNA)

RNA - SELEX





After 6 cycles of SELEX, the eluted RNAs are RT, amplified, ligated, **cloned** (published in 1999) and sequenced in series, in order to read the sequences selected and their relative fequencies.

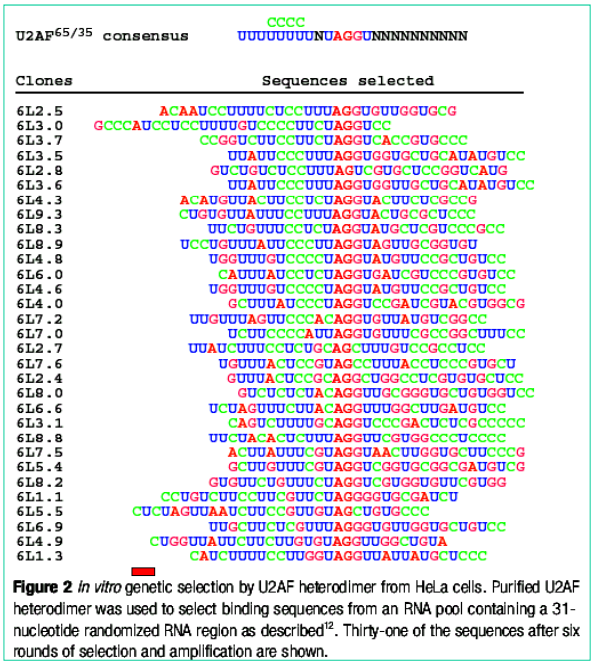


Figure 2 *in vitro* genetic selection by U2AF heterodimer from HeLa cells. Purified U2AF heterodimer was used to select binding sequences from an RNA pool containing a 31-nucleotide randomized RNA region as described[12]. Thirty-one of the sequences after six rounds of selection and amplification are shown.

Implication of additional exonic or intronic sequences

-in some cases, experimentally detected using mutation and in vitro splicing;

-conservation between organisms when alternative exons are conserved is observed within and around alternative exon;

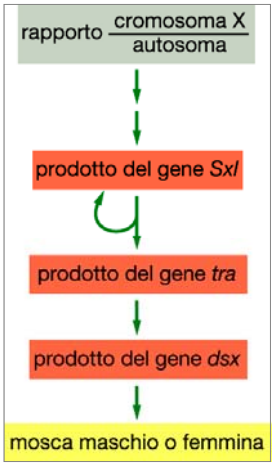
-difficult to find out by bioinformatics (very variable in terms of nucleotide and position)

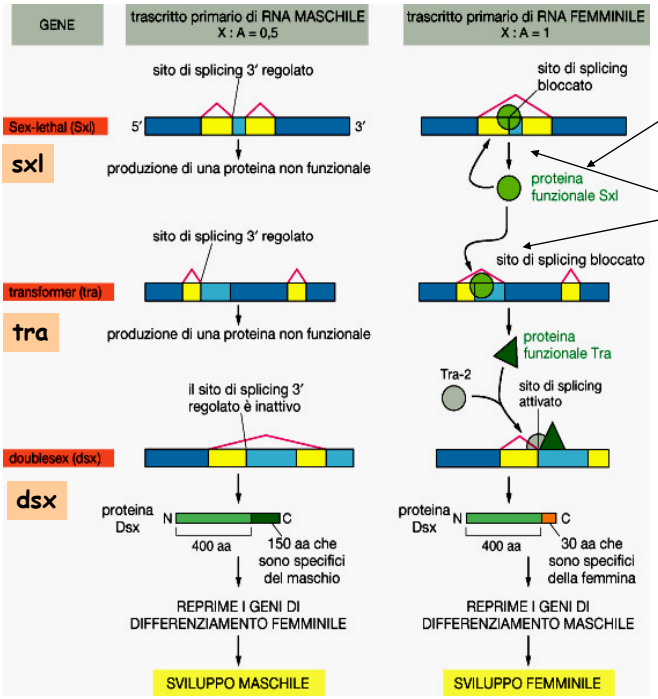
-specificity of interacting proteins often very difficult to identify

Exonic regulatory elements?

Determination of sex in Drosophila gave the **first example** of exonic sequences enhancing a "poor" 3'-ss utilization.

The first and most known model of regulated alternative splicing is the determination of sex in Drosophila. The primary determinat is the X:A chromosome ratio.
This determines a cascade of splicing regulatory signals, resulting in the production of two alternative splicing isoforms of the *dsx* transcription factor, repressing either female-specific or male-specific genes.

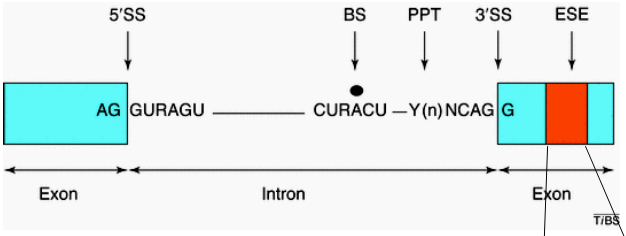




X:A=1 produces transient activation of an alternative promoter in the *sxl* gene, giving rise to a functional *sxl* protein

The *sxl* protein competes with U2AF for binding to the poly-pyrimidine tract

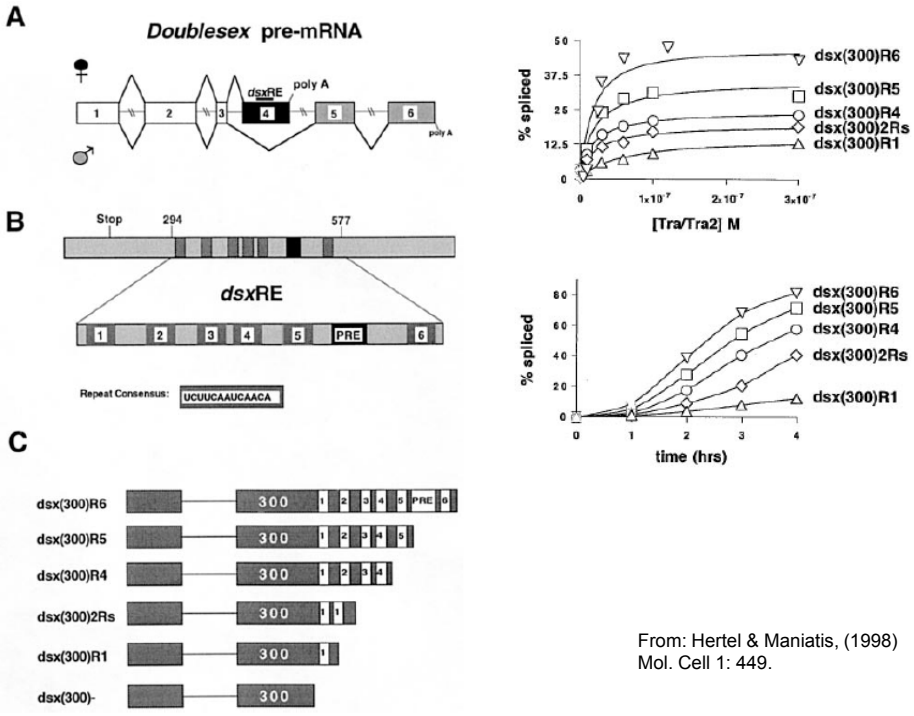
In the drosophila doublesex gene, the ESE element is present in exon 4 in addition to common intronic elements



Le ripetizioni dell'esone 4 di doublesex in Drosophila sono stati i primi Exonic Splicing Enhancer conosciuti (ESE)

ESE sono poi state riconosciute in numerosissimi esoni di varie specie.

$(rep)_6$



From: Hertel & Maniatis, (1998) Mol. Cell 1: 449.

Con che cosa interagiscono queste ESE ?

SR proteins = splicing regulators

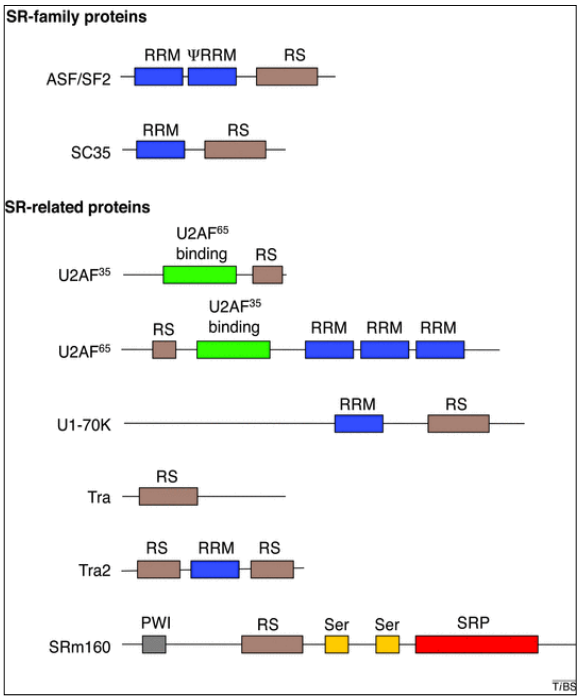
The most typical domain is an alternating Arginine-Serine domain, called "RS domain": it is a protein-protein interaction domain.

SR are phosphorylated at Ser by several kinases → regulates interaction with each other and with other proteins.

SR proteins are "proximalizing factors", whenever there is a "ss" choice, i.e. they have function in constitutive splicing, promoting the formation of complexes with pre-mRNA, snRNP U1 and U2.

SR proteins also interact with the CAP-binding protein and with poly-A binding proteins.

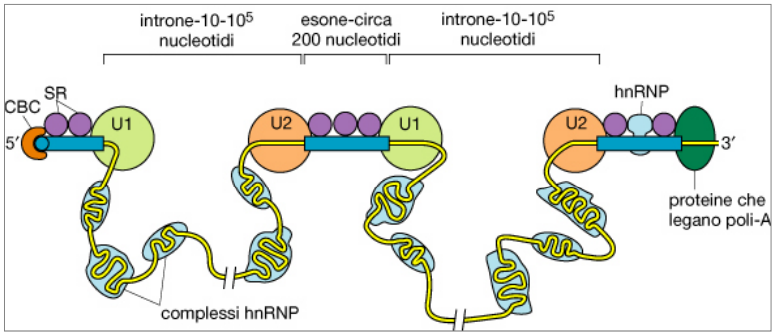
Useful characteristics: they precipitate with 10 mM Mg++, so that it is easy to deplete nuclear extracts of SR factors (e.g. the S100 extract)



SR proteins display an RS motif accompanied by one or more RRN domain.

Other related proteins possess RS domains.

Name*	Domains	Binding sequence	Target genes
Canonical SR proteins			
SRp20 (SFRS3)	RRM and RS	GCUCCUCUUC	*SRP20*, *CALCA* and *INSR*
SC35 (SFRS2)	RRM and RS	UGCUGUU	*ACHE* and *GRIA1–GRIA4*
ASF/SF2 (SFRS1)	RRM, RRMH and RS	RGAAGAAC	*HIPK3*, *CAMK2D*, HIV RNAs and *GRIA1–GRIA4*
SRp40 (SFRS5)	RRM, RRMH and RS	AGGAGAAGGGA	*HIPK3*, *PRKCB* and *FN1*
SRp55 (SFRS6)	RRM, RRMH and RS	GGCAGCACCUG	*TNNT2* and *CD44*
SRp75 (SFRS4)	RRM, RRMH and RS	GAAGGA	*FN1*, *E1A* and *CD45*
9G8 (SFRS7)	RRM, zinc finger and RS	(GAC)n	*TAU*, *GNRH* and *SFRS7*
SRp30c (SFRS9)	RRM, RRMH and RS	CUGGAUU	*BCL2L1*, *TAU* and *HNRNPA1*
SRp38 (FUSIP1)	RRM and RS	AAAGACAAA	*GRIA2* and *TRD*
Other SR proteins			
SRp54	RRM and RS	ND	*TAU*
SRp46 (SFRS2B)	RRM and RS	ND	NA
RNPS1	RRM and Ser-rich	ND	*TRA2B*
SRrp35	RRM and RS	ND	NA
SRrp86 (SRrp508 and SFRS12)	RRM and RS	ND	NA
TRA2α	RRM and two Arg-rich	GAAARGARR	dsx
TRA2β	RRM and two RS	(GAA)n	*SMN1*, *CD44* and *TAU*
RBM5	RRM and RS	ND	*CD95*
CAPER (RBM39)	RRM and RS	ND	*VEGF*



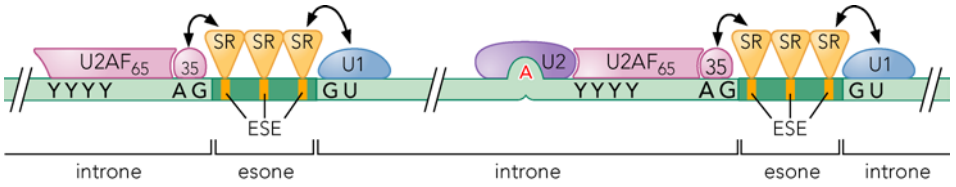


Table 1 | **Ribonucleoproteins that are involved in pre-mRNA splicing**

Name	Other names	Domains*	Binding sequences	Target genes
hnRNP A1	NA	RRM, RGG and G	UAGGGA/U	*SMN2* and *RAS*
hnRNP A2 hnRNP B1	NA	RRM, RGG and G	(UUAGGG)n	HIV *tat* and *IKBKAP*
hnRNP C1 hnRNP C2	AUF1	RRM	U rich	*APP*
hnRNP F	NA	RRM, RGG and GY	GGGA and G rich	*PLP*, *SRC* and *BCL2L2*
hnRNP G	NA	RRM and SRGY	CC(A/C) and AAGU	*SMN2* and *TMP1*
hnRNP H hnRNP H′	DSEF1	RRM, RGG, GYR and GY	GGGA and G rich	*PLP*, HIV *tat* and *BCL2L1*
hnRNP I	PTB	RRM	UCUU and CUCUCU	*PTB*, *nPTB*, *SRC*, *CD95*, *TNTT2*, *CALCA* and *GRIN3B*
hnRNP L	NA	RRM	C and A rich	*NOS* and *CD45*
hnRNP LL	SRRF	RRM	C and A rich	*CD45*
hnRNP M	NA	RRM and GY	ND	*FGFR2*
hnRNP Q	NA	RRM and RGG	ND	*SMN2*

hnRNP = heterogeneous nuclear ribonucleoproteins

Everything has to be tought as a dynamic process since it is now clear that exon and intron definition occurs co-transcriptionally

review

"Cotranscriptionality": The Transcription Elongation Complex as a Nexus for Nuclear Transactions

Roberto Perales[1] and David Bentley[1,*]
[1]Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, UCHSC, MS8101, P.O. Box 6511, Aurora CO, 80045, USA
*Correspondence: david.bentley@uchsc.edu
DOI 10.1016/j.molcel.2009.09.018

Much of the complex process of RNP biogenesis takes place at the gene cotranscriptionally. The target for RNA binding and processing factors is, therefore, not a solitary RNA molecule but, rather, a transcription elongation complex (TEC) comprising the growing nascent RNA and RNA polymerase traversing a chromatin template with associated passenger proteins. RNA maturation factors are not the only nuclear machines whose work is organized cotranscriptionally around the TEC scaffold. Additionally, DNA repair, covalent chromatin modification, "gene gating" at the nuclear pore, Ig gene hypermutation, and sister chromosome cohesion have all been demonstrated or suggested to involve a cotranscriptional component. From this perspective, TECs can be viewed as potent "community organizers" within the nucleus.

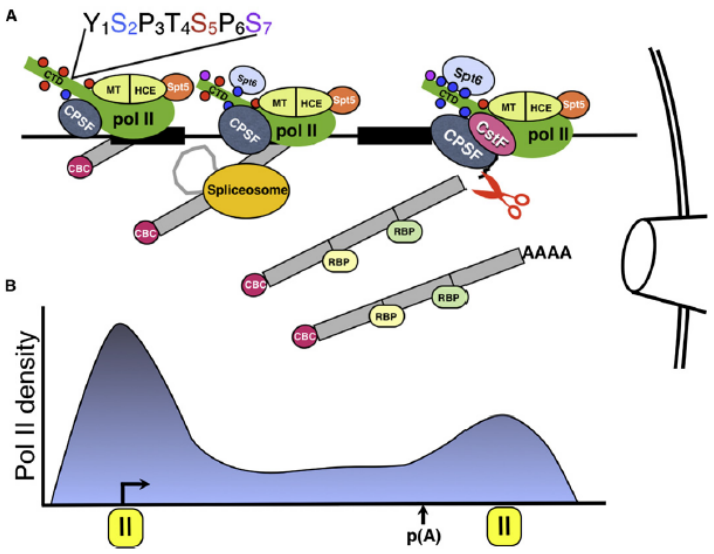


Figure 1. Processing and Elongation Factors Associated with the Human Pol II TEC

Extensive research on several model systems has led to definition of several sequences that regulale alternative splicing:

ESE – Exonic splicing enhancer

ESS – Exonic splicing silencer

ISE – Intronic splicing enhancer

ISS – Intronic splicing silencer

As a general rule, enhancers interact – directly or indirectly – with SR proteins or related, while silencers generally work through **hnRNPs.**

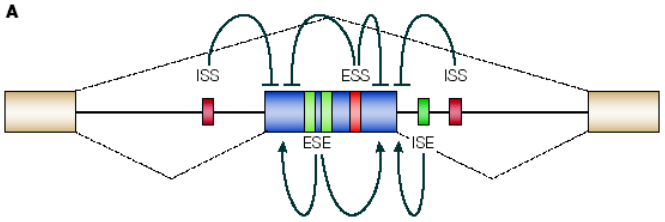
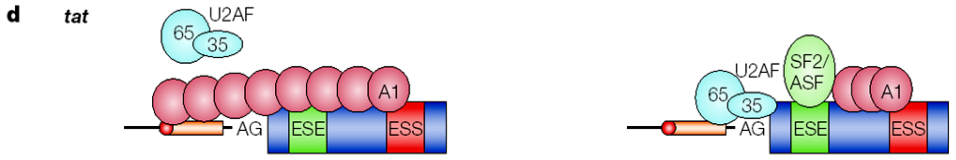


Figure 1 | **Elementary alternative splicing events and regulatory elements. A** | In addition to the splice-site consensus sequences, a number of auxiliary elements can influence alternative splicing. These are categorized by their location and activity as exon splicing enhancers and silencers (ESEs and ESSs) and intron splicing enhancers and silencers (ISEs and ISSs). Enhancers can activate adjacent splice sites or antagonize silencers, whereas silencers can repress splice sites or enhancers. Exon inclusion or skipping is determined by the balance of these competing influences, which in turn might be determined by relative concentrations of the cognate RNA-binding activator and repressor proteins.

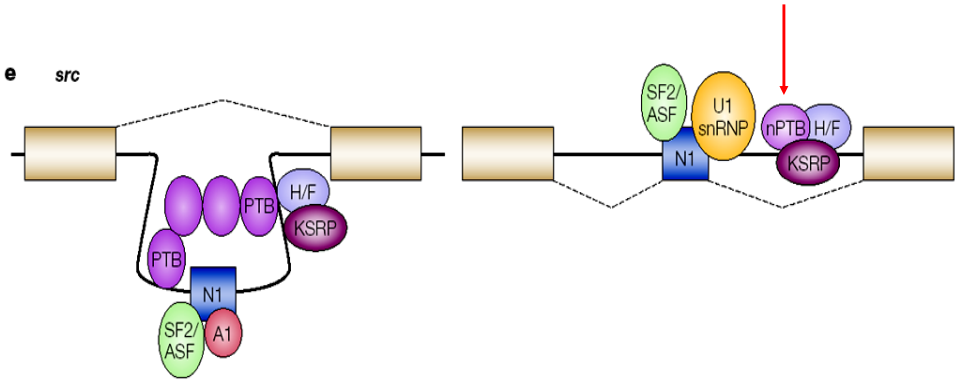
From: Matlin et al. (2005), Nature Rev Mol Cell Biol, 6: 386.

b | A weak 5' splice site in the *FAS* transcript is enhanced by TIA1 binding to a downstream intron splicing enhancer (ISE). TIA1 cooperatively promotes the interaction of U1 small nuclear ribonucleoprotein particles (snRNPs) with the pre-mRNA.

c | Repression of the non-sex-specific *tra* 3' splice site involves the interaction of SXL with an intron splicing silencer (ISS) embedded in the polypyrimidine tract and the prevention of U2AF binding. This leads to selection of the downstream female-specific 3' splice site.



d | Inclusion of exon 3 of HIV1 *tat* pre-mRNA is determined by the nuclear ratio of specific heterogeneous nuclear ribonucleoprotein (hnRNP) and SR proteins. Propagative multimerization of hnRNPA1 from a high-affinity exon splicing silencer (ESS) is sterically blocked by the interaction of SF2/ASF with the upstream ESE. In this case, ESE function requires the RRM domains but not the RS domain of SF2/ASF.



e | The regulation of N1 exon splicing in the *src* transcript provides an example of combinatorial control by cooperation and antagonism between numerous positively and negatively acting factors. In non-neuronal cells (left), N1 is excluded, whereas in neurons (right), it is included in the mature mRNA. Constitutive exons are shown as beige boxes, whereas alternative exons are shown as blue boxes. KSRP, KH-type splicing regulatory protein; nPTB, neural polypyrimidine tract binding protein.

Splicing regulatory elements ESE, ISE, ESS, ISS

how do they look like ?

Very difficult to define, poor conservation, superposition with other sequence algorithms, possibly combinatorial interaction with many different RNA-binding proteins

We will go through one of the first paper addressing this question, starting from the known ESE in the dsx gene, then we will move to more “modern” approaches that put together bioinformatics and functional assays to explore the whole genome in search of sequence elements or “motifs” regulating AS.

MOLECULAR AND CELLULAR BIOLOGY, Mar. 1999, p. 1705–1719
0270-7306/99/$04.00+0

Vol. 19, No. 3

D- melanogaster

Selection and Characterization of Pre-mRNA Splicing Enhancers: Identification of Novel SR Protein-Specific Enhancer Sequences

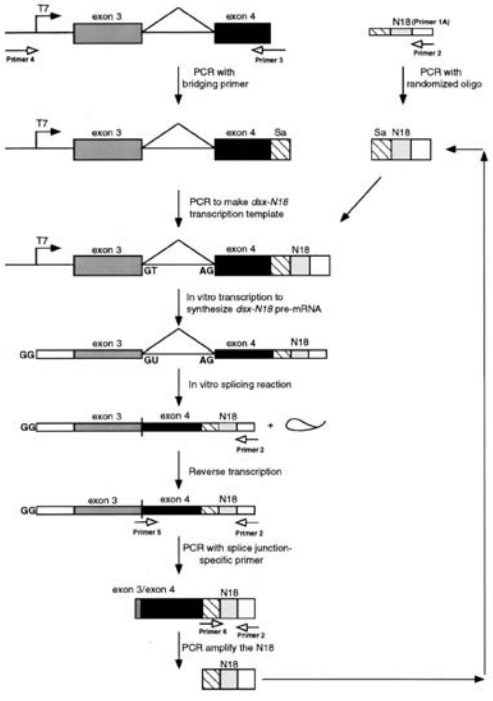
THOMAS D. SCHAAL AND TOM MANIATIS*

Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138

Received 14 September 1998/Returned for modification 28 October 1998/Accepted 23 November 1998

Splicing enhancers are RNA sequences required for accurate splice site recognition and the control of alternative splicing. In this study, we used an in vitro selection procedure to identify and characterize novel RNA sequences capable of functioning as pre-mRNA splicing enhancers. Randomized 18-nucleotide RNA sequences were inserted downstream from a *Drosophila doublesex* pre-mRNA enhancer-dependent splicing substrate. Functional splicing enhancers were then selected by multiple rounds of in vitro splicing in nuclear extracts, reverse transcription, and selective PCR amplification of the spliced products. Characterization of the selected splicing enhancers revealed a highly heterogeneous population of sequences, but we identified six classes of recurring degenerate sequence motifs five to seven nucleotides in length including novel splicing enhancer sequence motifs. Analysis of selected splicing enhancer elements and other enhancers in S100 complementation assays led to the identification of individual enhancers capable of being activated by specific serine/arginine (SR)-rich splicing factors (SC35, 9G8, and SF2/ASF). In addition, a potent splicing enhancer sequence isolated in the selection specifically binds a 20-kDa SR protein. This enhancer sequence has a high level of sequence homology with a recently identified RNA-protein adduct that can be immunoprecipitated with an SRp20-specific antibody. We conclude that distinct classes of selected enhancers are activated by specific SR proteins, but there is considerable sequence degeneracy within each class. The results presented here, in conjunction with previous studies, reveal a remarkably broad spectrum of RNA sequences capable of binding specific SR proteins and/or functioning as SR-specific splicing enhancers.

SELEX procedure for selecting 18-mers with exon splicing enhancer properties



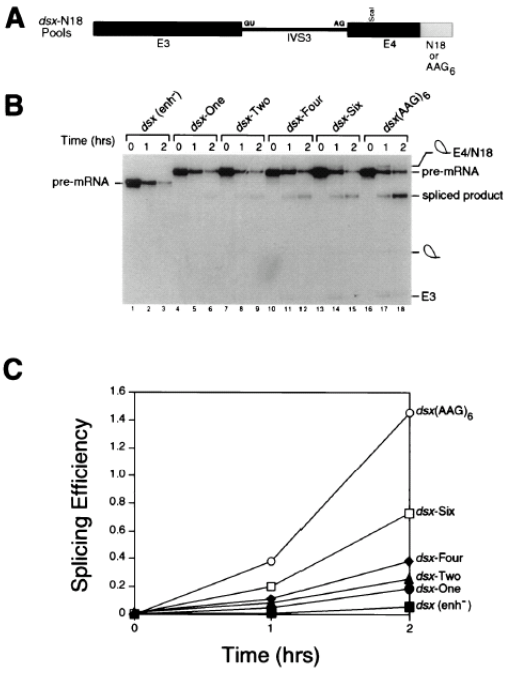


FIG. 2. Evolution of the *dsx-N18* pool. (A) The *dsx-N18* and *dsx-(AAG)6* constructs are shown schematically. Exon 3, intron 3, exon 4, and the enhancer(s) are indicated by E3, IVS3, E4, and N18 or AAG6, respectively. The 59 and 39 splice sites are indicated by GU and AG, respectively. (B) Kinetic analysis showing in vitro splicing assays performed with HeLa cell nuclear extracts and uniformly labeled pre-mRNA splicing substrates comprising the total pool of *dsx-N18* pre-mRNAs after various rounds of the selection (rounds 1, 2, 4, and 6 are shown in lanes 4 to 6, 7 to 9, 10 to 12, and 13 to 15, respectively). The negative control pre-mRNA (lanes 1 to 3) is an *dsx* pre-mRNA lacking an enhancer [*dsx*(enh2)]. The positive control pre-mRNA (lanes 16 to 18) is a *dsx* pre-mRNA activated by six consecutive copies of a multimerized AAG trinucleotide splicing enhancer (modeled after a synthetic polypurine splicing enhancer in reference 66) that is otherwise isogenic to the *dsx-N18* construct. In the kinetic analysis shown, the reaction mixtures were incubated for the number of hours indicated at the top, and positions of the precursors, intermediates, and products of the splicing reaction are indicated to the left and right. The RNAs were analyzed on a 10% denaturing gel in order to resolve the lariat-exon 4 intermediate from the spliced product. (C) Quantitation of the in vitro splicing reactions in panel B. The splicing efficiency (ratio of spliced product to precursor) is calculated from quantitation of individual bands after subtraction of background using a BAS2000 phosphorimager.

TABLE 1. Purine- and pyrimidine-rich sequences in the selected splicing enhancers

Clone no.	Splicing enhancer sequence[a]	Splicing efficiency[b] (%)
Class I, purine-rich enhancers (≥65% purine content)		
Motif A, GGGGA		
3-7	GCAAC**GGGGA**CGCGGC	40
3-1	AGCGGUCGCGGUU**GGGGGag**	32
6-43	GC**GGAGGAGG**CCCGU**GGGag**	50
Motif B, GGAGGA		
6-43	GC**GGAGGAGG**CCCGU**GGGag**	50
6-19	GCCAGC**GGAGGA**UGCGG	53
Motif C, GGAGA		
3-35	CUGGAAUAC**GGAGA**CCGG	36
6-40	GGUGAGC**GGAGA**UGCUGC	31
Others		
3-36	GGACCU**AGAGG**UGGCGAC	40
6-29	GACCGUCGGAC**AGGAG**C	36
Class II, pyrimidine-rich enhancers (≥67% pyrimidine content)		
Motif D, UCUCC		
6-13	**au**CUCCACGUCGCCUGCUGC	38
6-16	**au**CUCCACGUCGCCUGCUGC	37
6-24	UUUGCGG**UCUCC**GGC**CUCC**	56
Motif E, UCUUC		
6-5	UGCCACCCGCGG**UCUUCC**	26
6-12	UCGUCG**UCUUC**GCGGCCC	49
3-32	CCUGCUGCG**UCUUU**GUCC	27
Motif F, UCCUC		
6-7	CCUG**UCCUC**GGUGUUGC	36
6-22	CG**UCCUC**GUGUCACCGCC	37
6-6	GGU**UCCUG**UCGCCGCCCC	41
Controls (reference)		
dsx, enhancerless		≤1
hβ-globin (51)		81
dsx-ASLV (60)		58

TABLE 2. Recurring motifs in selected splicing enhancers and other strong enhancers

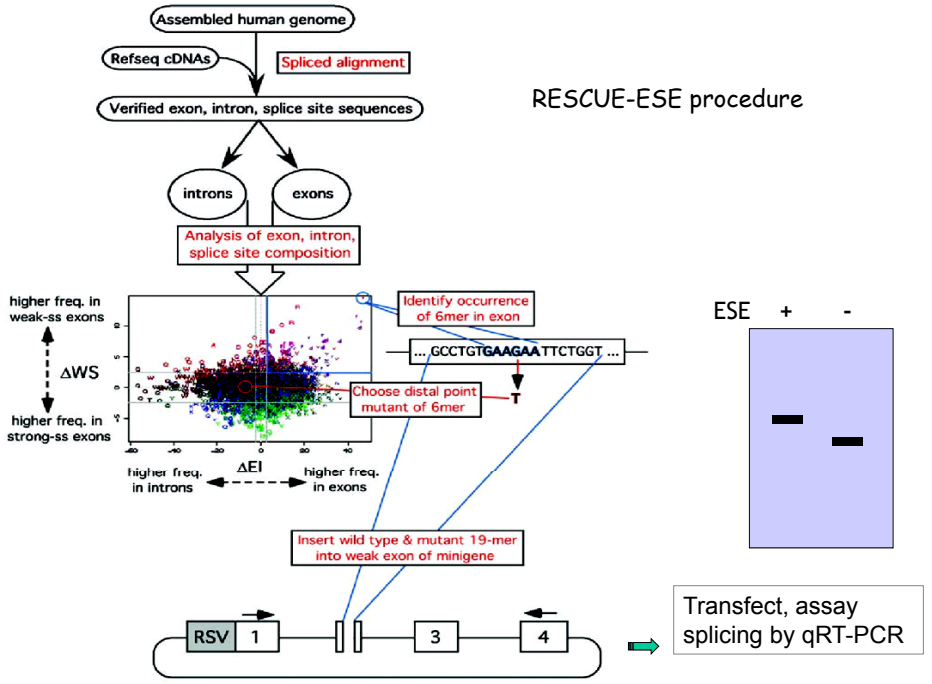
Clone	Selected enhancer sequence[a]	Splicing efficiency[b] (%)
Class III, enhancers containing permutations of the sequence (U)GGACCNG		
6-14	GCCGCCGCUUCGU**GGACCag**	53
6-25	CACGCUCCUCGCU**GGACCag**	53
6-38	GCCGCCGUGGU**GGACCGG**ag	50
6-26	CCGAGCUACA**GGACCGG**ag	35
6-29	GACCGUC**GGACAGG**AGC	36
3-35	CUGGAAUACGG**AGACCGGag**	36
3-36	u**GGACCUA**GAGGUGGCGAC	40
6-9	u**GGACCGC**CCUGCCAUACC	34
3-3	CAGGCG**GGACCGC**GACG	17
Class IV, enhancers containing the sequence (C)CACC(C)		
6-28	CCGGAG**CCACCC**GGUACC	29
6-5	UG**CCACCC**GCGGUCUUCC	26
6-2	CGUCG**CACCC**UGUCUGCC	29
6-22	CGUCCUCGUGU**CACC**GCC	37
6-35	UCCUGGCGU**CACC**GUAC	27
Class V, enhancers containing the sequence YGCCGCC		
6-14	**uGCCGCC**GCUUCGUGGACC	53
6-38	**uGCCGCC**GUGGUGGACCGG	50
6-45	**uGCCGCC**GCGAGUUGGGGC	32
6-8	GCCAGUAGU**UGCCGCC**GC	24
6-6	GGUUCCUGU**CGCCGCC**CC	41
6-1	GGACACCUGUG**CGCCGCC**ag	43
Class VI, enhancers containing the sequence RGAACYU		
3-25	CCACGU**GGAACCU**CGUCC	35
6-44	ACGGCGC**GGAACCU**UUCC	47
6-23	GCCCG**AGAACUU**CUUGCC	40
Class VII, other strong enhancers		
6-18	CCGACGCCAUGGACGACGag	55
6-3	GGCUGCCAGUCGGAAUUGG	52
6-47	CCGUGACAGCAUCGGCGG	50
3-23	CGUCGGCAGGUGGUCCCG	47
6-39	UCUGGAUCCUGCGGAUGG	44

Predictive Identification of Exonic Splicing Enhancers in Human Genes

William G. Fairbrother,[1,2]* Ru-Fang Yeh,[1]* Phillip A. Sharp,[1,2] Christopher B. Burge[1]†

Specific short oligonucleotide sequences that enhance pre-mRNA splicing when present in exons, termed exonic splicing enhancers (ESEs), play important roles in constitutive and alternative splicing. A computational method, RESCUE-ESE, was developed that predicts which sequences have ESE activity by statistical analysis of exon-intron and splice site composition. When large data sets of human gene sequences were used, this method identified 10 predicted ESE motifs. Representatives of all 10 motifs were found to display enhancer activity in vivo, whereas point mutants of these sequences exhibited sharply reduced activity. The motifs identified enable prediction of the splicing phenotypes of exonic mutations in human genes.

Science (2002) 297: 1007-1013.



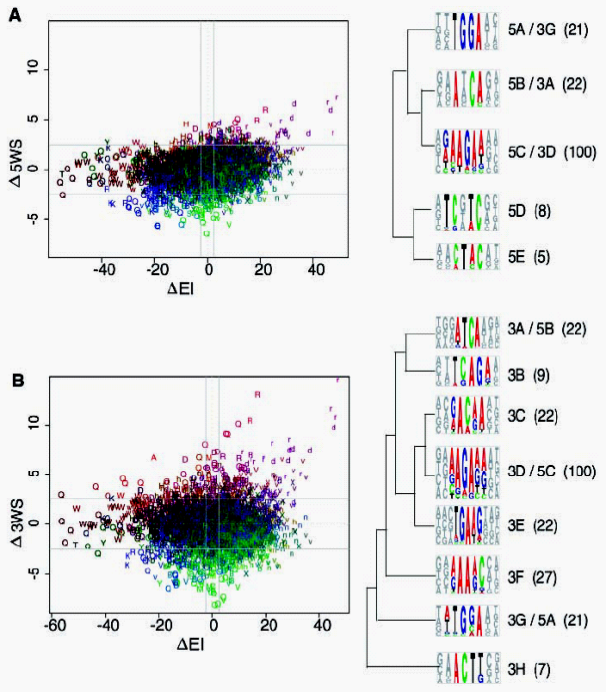
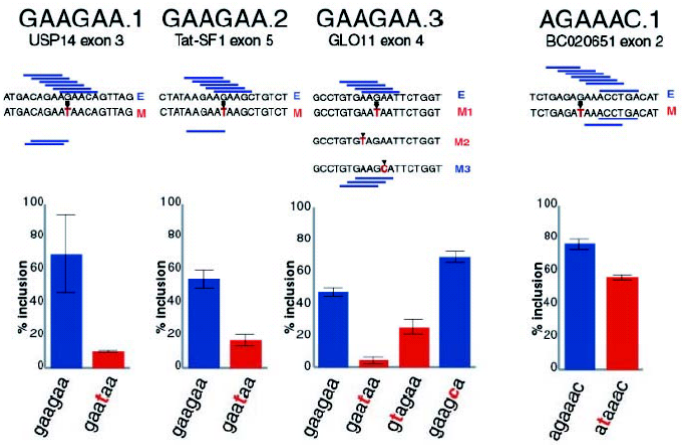


Fig. 2. RESCUE-ESE prediction of 5′ and 3′ ESEs in human genes.

(**A**) Scatterplot for prediction of 5ESE activity. Hexamers are represented by colored letters as described in Fig. 1. Simplified dendrogram shows clustering of 5′ESE hexamers (total of 103 hexamers with ΔEI>2.5 and Δ5WS>2.5) into five clusters of four or more hexamers.

(**B**) Scatterplot for prediction of 3′ESE activity. Simplified dendrogram shows clustering of 3′ESE hexamers (total of 198 hexamers with ΔEI>2.5 and Δ3WS>2.5) into eight clusters of four or more hexamers. Complete dendrograms of all hexamers are shown in fig. S3. The aligned sequences in each cluster are represented as Pictograms (http://genes.mit.edu/pictogram.html).

Cluster labels (e.g., 3B, 5A/3G) are listed to the right of each Pictogram, with the total number of hexamers in the cluster indicated in parentheses.



Examers + 19-b upstream and 6-b downstream for each “exemplar”

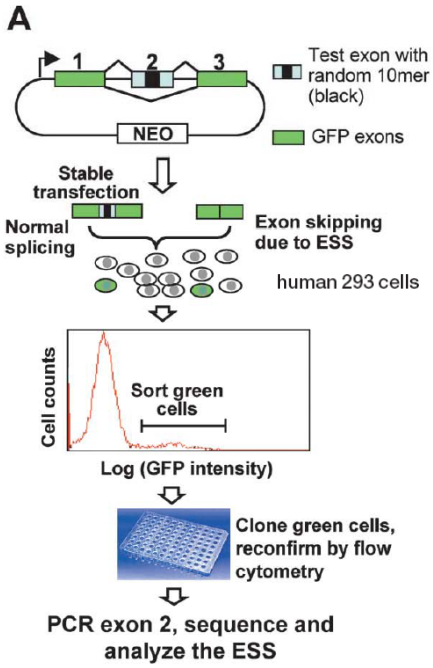
Cell, Vol. 119, 831–845, December 17, 2004, Copyright ©2004 by Cell Press

Systematic Identification and Analysis of Exonic Splicing Silencers

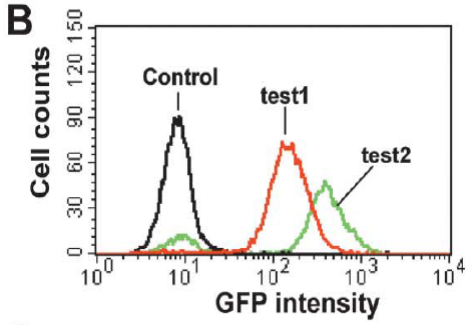
Zefeng Wang,[1] Michael E. Rolish,[1,2] Gene Yeo,[1,3] Vivian Tung,[1] Matthew Mawson,[1] and Christopher B. Burge[1,*]
[1]Department of Biology
[2]Department of Electrical Engineering and Computer Science
[3]Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

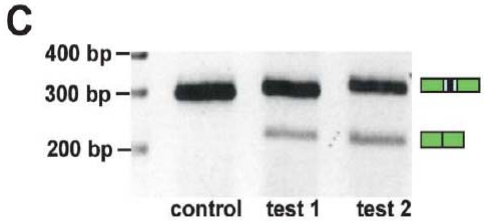
Summary

Exonic splicing silencers (ESSs) are *cis*-regulatory elements that inhibit the use of adjacent splice sites, often contributing to alternative splicing (AS). To systematically identify ESSs, an in vivo splicing reporter system was developed to screen a library of random decanucleotides. The screen yielded 141 ESS decamers, 133 of which were unique. The silencer activity of over a dozen of these sequences was also confirmed in a heterologous exon/intron context and in a second cell type. Of the unique ESS decamers, most could be clustered into groups to yield seven putative ESS motifs, some resembling known motifs bound by hnRNPs H and A1. Potential roles of ESSs in constitutive splicing were explored using an algorithm, ExonScan, which simulates splicing based on known or putative splicing-related motifs. ExonScan and related bioinformatic analyses suggest that these ESS motifs play important roles in suppression of pseudoexons, in splice site definition, and in AS.



exon 2 of the Chinese hamster dihydrofolate reductase (*DHFR*) gene, was used as the test exon.

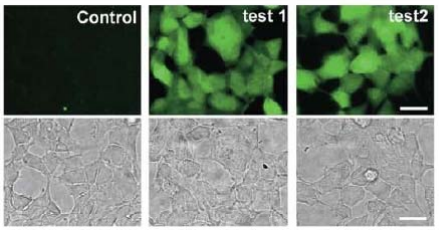




(B) Test of the reporter system with two known ESS sequences. Test 1 (hnRNP A1 binding site, ATGATAG GGACTTAGGGT [Burd and Dreyfuss, 1994]) and test 2 (U2AF65 binding site, TTTTTTTTTCCTTTTT TTTTCCTTTT [Singh et al., 1995]) were inserted into the pZW4 reporter construct and transfected into 293 Flp-In cells, and positive transfectants were pooled for flow cytometry. The "Control" was a randomly chosen 10-mer sequence (ACCTCAGGCG) inserted into the same vector.

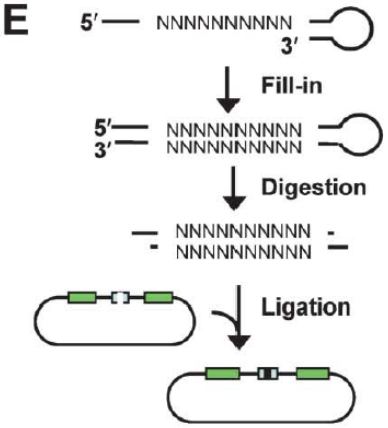
(C) RT-PCR results using RNA purified from the transfected cells as template, with primers targeted to exons 1 and 3 of pZW4.

D

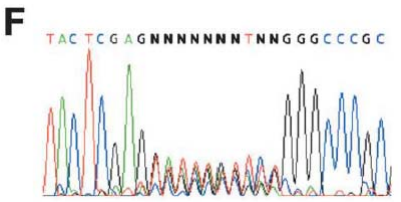


D) Microscopic images of transfected cells. Upper panel, GFP fluorescence. Lower panel, phase images. Scale bar, 50 μM.

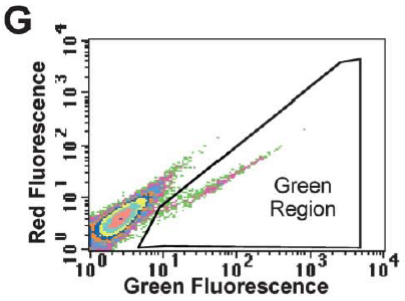
E



(E) Construction of random decamer library. The foldback primer was synthesized with a random sequence of 10 bp, then extended with Klenow fragment, digested, ligated into pZW4, and transformed into *E. coli*.

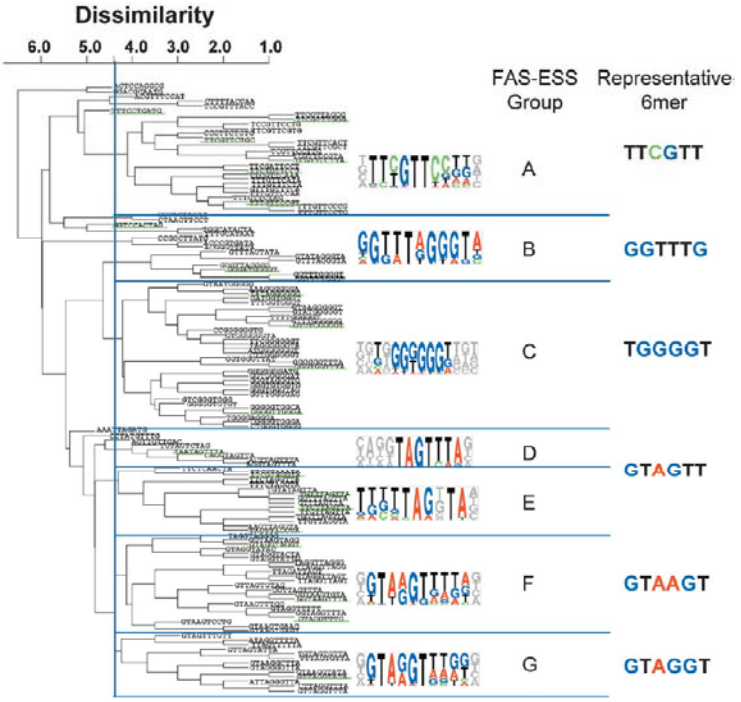


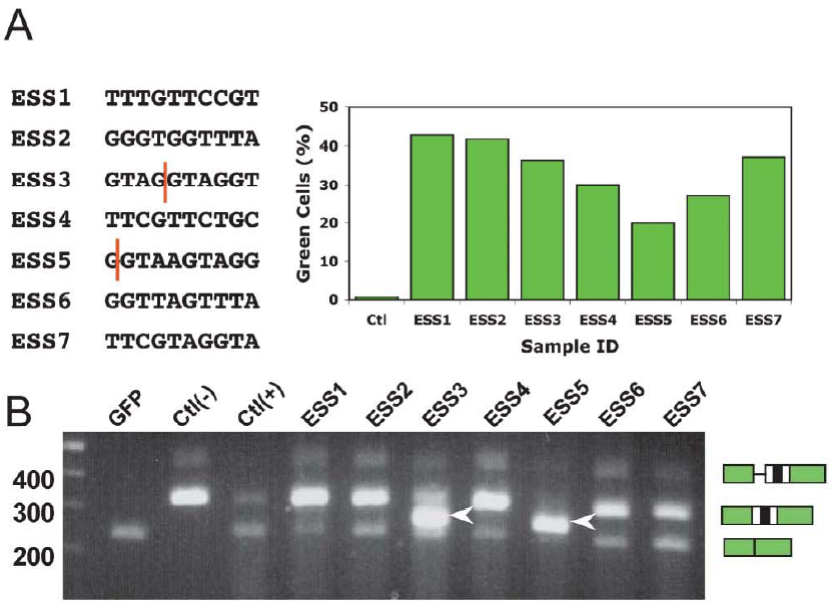
(F) Sequencing of the random decamer region. 293 cells stably transfected with the pZW4 library were pooled to purify total DNAs, from which minigene fragments were amplified by PCR and sequenced. Sequences around the insertion region are shown.



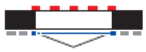
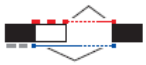
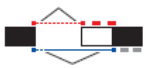
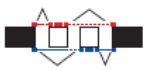
(G) Flow cytometry profile of single transfection using pZW4 random decamer library.

Sequences in the green cells cloned and sequenced





In a large number of cases: tissue-specific regulation

Alternative transcript events		Total events ($\times 10^3$)	Number detected ($\times 10^3$)	Both isoforms detected	Number tissue-regulated	% Tissue-regulated (observed)	% Tissue-regulated (estimated)
Skipped exon		37	35	10,436	6,822	65	72
Retained intron		1	1	167	96	57	71
Alternative 5′ splice site (A5SS)		15	15	2,168	1,386	64	72
Alternative 3′ splice site (A3SS)		17	16	4,181	2,655	64	74
Mutually exclusive exon (MXE)		4	4	167	95	57	66
Alternative first exon (AFE)		14	13	10,281	5,311	52	63
Alternative last exon (ALE)		9	8	5,246	2,491	47	52
Tandem 3′ UTRs		7	7	5,136	3,801	74	80
Total		105	100	37,782	22,657	60	68

Constitutive exon or region — Body read — Junction read — pA Polyadenylation site
Alternative exon or extension — Inclusive/extended isoform — Exclusive isoform — Both isoforms

nature Vol 444 | 30 November 2006 | doi:10.1038/nature05304

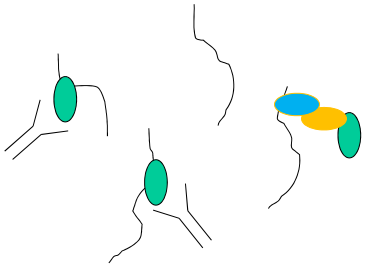
ARTICLES

An RNA map predicting Nova-dependent splicing regulation

Jernej Ule[1,2]*†, Giovanni Stefani[1,2]*†, Aldo Mele[1,2], Matteo Ruggiu[1,2], Xuning Wang[3], Bahar Taneri[4]†, Terry Gaasterland[4]†, Benjamin J. Blencowe[5] & Robert B. Darnell[1,2]

Nova proteins are neuron-specific alternative splicing factors. We have combined bioinformatics, biochemistry and genetics to derive an RNA map describing the rules by which Nova proteins regulate alternative splicing. This map revealed that the position of Nova binding sites (YCAY clusters) in a pre-messenger RNA determines the outcome of splicing. The map correctly predicted Nova's effect to inhibit or enhance exon inclusion, which led us to examine the relationship between the map and Nova's mechanism of action. Nova binding to an exonic YCAY cluster changed the protein complexes assembled on pre-mRNA, blocking U1 snRNP (small nuclear ribonucleoprotein) binding and exon inclusion, whereas Nova binding to an intronic YCAY cluster enhanced spliceosome assembly and exon inclusion. Assays of splicing intermediates of Nova-regulated transcripts in mouse brain revealed that Nova preferentially regulates removal of introns harbouring (or closest to) YCAY clusters. These results define a genome-wide map relating the position of a *cis*-acting element to its regulation by an RNA binding protein, namely that Nova binding to YCAY clusters results in a local and asymmetric action to regulate spliceosome assembly and alternative splicing in neurons.

Cross-link (U.V.)



Nova: the first vertebrate tissue-specific splicing factors (neurons)
(Nova1 – Nova2)

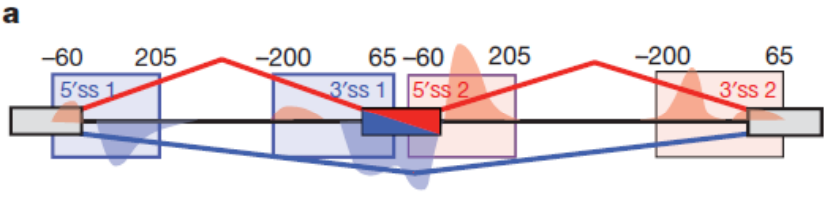
CLIP (cross-link immunoprecipitation)
(Assignment BMG11)

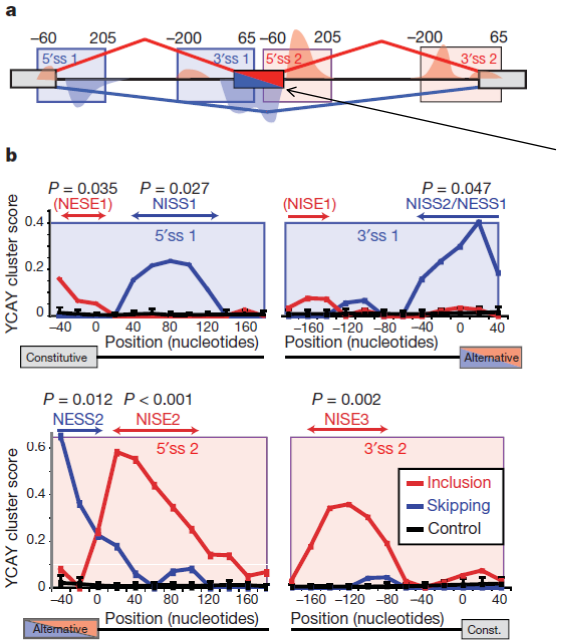
IMPT

RNA identified on microarrays
or sequenced

48 targets identified in previous studies

Clustering of “YCAY” Nova recognition sequences in 48 Nova-regulated exons





Scoring YACY clusters (median 28-base long) in 200-bp boxes at 5'- and 3'- SS
Control: genome-wide positions Nova-independent

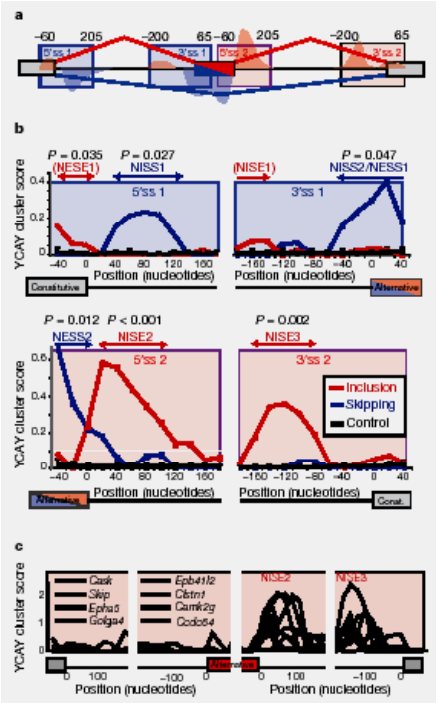
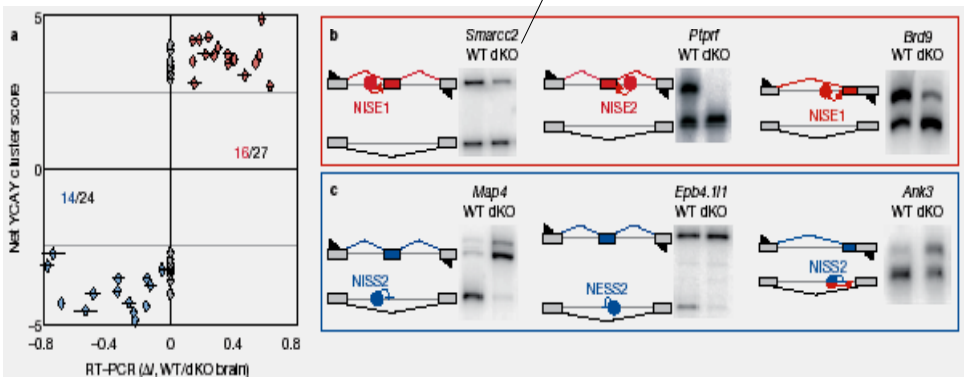


Figure 1 | Definition of the Nova-RNA binding map. **a**, A generic premRNA showing the four regions that define the Nova-RNA binding map (the start and end of each region is labelled by a nucleotide distance to the splice site). Peaks demonstrate the positions of Nova-dependent **splicing enhancers** (red) or **silencers** (blue).
b, A conserved YCAY cluster score (y axis) was calculated as described in Supplementary Methods. The x axis shows the nucleotide position of the centre of the sequence window. At each significant peak of YCAY cluster enrichment in Nova-regulated versus control pre-mRNAs, the P-value (two-tailed t-test, unequal variance) of YCAY clusters is shown. The error value of the control pre-mRNAs represents standard deviation of the mean values of 100 random groups of 20 control pre-mRNAs.
c, YCAY cluster distribution in eight Nova-regulated pre-mRNAs that contain both NISE2 and NISE3 elements.

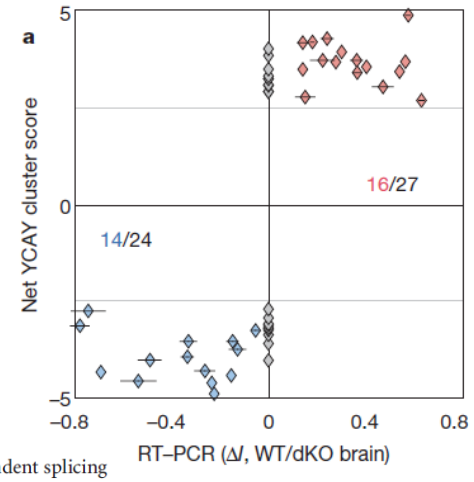
In K.O. mice brain

Red are intronic sites



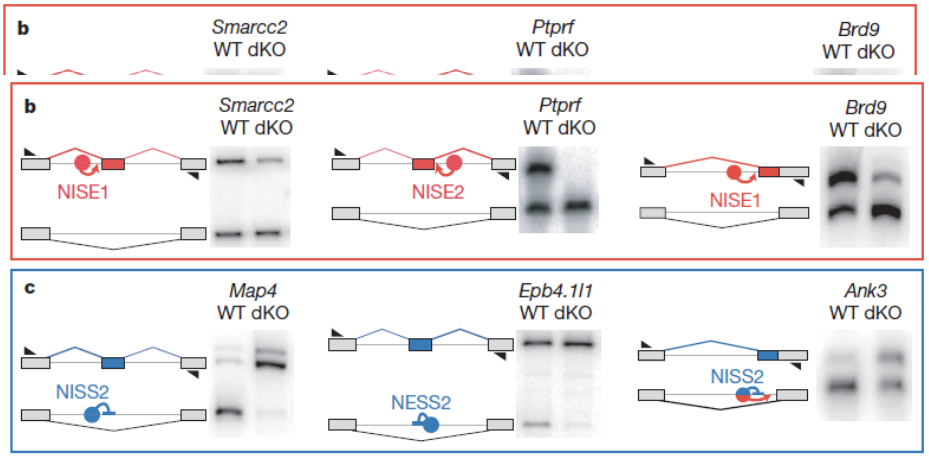
Blue are exonic sites or adjacent to ss

Genome-wide scanning with the Nova-score definition finds out 51 new candidate exons



To test whether the RNA map can predict Nova-dependent splicing regulation *de novo*, we calculated a net YCAY cluster score by subtracting Nova silencer from enhancer cluster scores (Supplementary Fig. 4). Using a stringent scoring method (|net YCAY cluster score| > 2.7), we identified 51 candidate Nova-regulated alternative exons in a genomic database of bioinformatically predicted alternative exons (B.T. and T.G., personal communication). Ten previously validated Nova-regulated exons were among these top predictions.

Examples of predicted Nova-regulated exons: analysis in brain tissues from Nova1$^{-/-}$ / Nova2 $^{-/-}$ double K.O. mice.



Mechanism: by studying the different steps in vitro on selected alternative exons, the authors demonstrated that Nova regulates assembly of the early spliceosomal complex