Ch 4 – Lesson 2.1

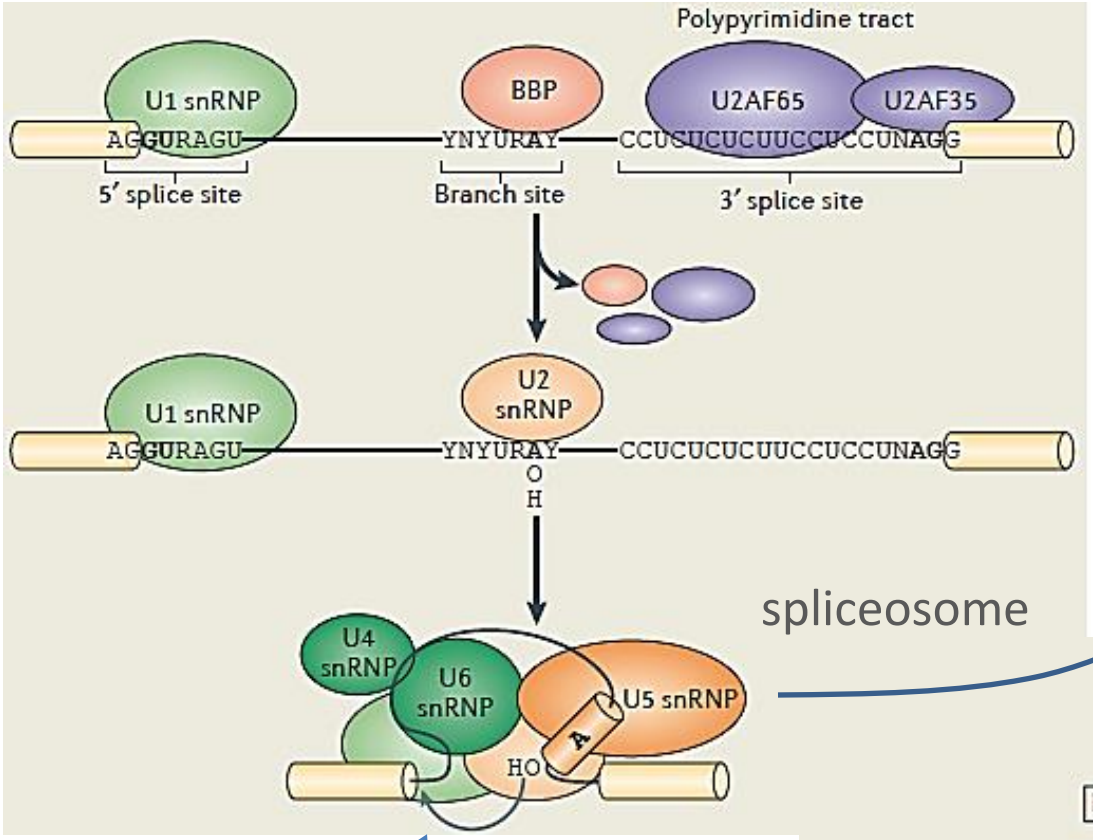
Alternative splicing regulation

Textbook

Alternative splicing: a pivotal step between eukaryotic transcription and translation

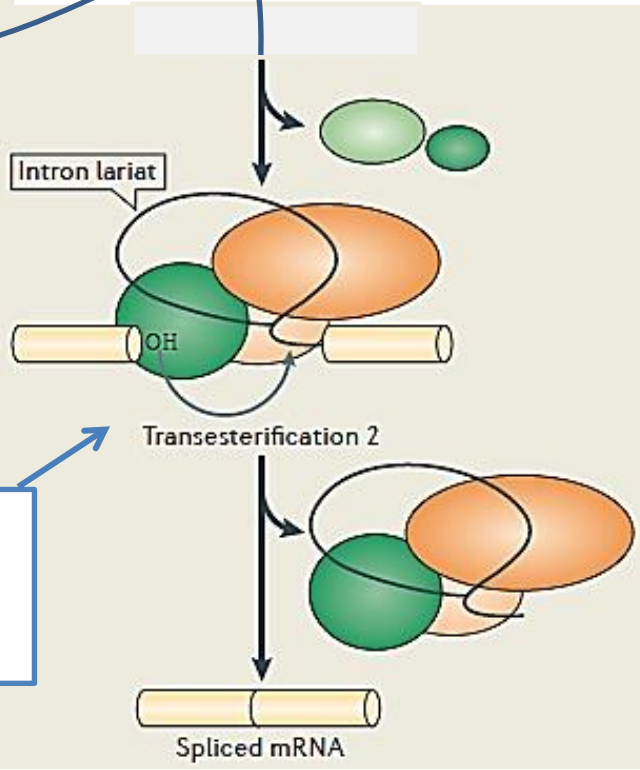
Alberto R. Kornblihtt, Ignacio E. Schor, Mariano Alló, Gwendal Dujardin, Ezequiel Petrillo and Manuel J. Muñoz

Abstract | Alternative splicing was discovered simultaneously with splicing over three decades ago. Since then, an enormous body of evidence has demonstrated the prevalence of alternative splicing in multicellular eukaryotes, its key roles in determining tissue- and species-specific differentiation patterns, the multiple post- and co-transcriptional regulatory mechanisms that control it, and its causal role in hereditary disease and cancer. The emerging evidence places alternative splicing in a central position in the flow of eukaryotic genetic information, between transcription and translation, in that it can respond not only to various signalling pathways that target the splicing machinery but also to transcription factors and chromatin structure.



exon splicing

spliceosome



1st transesterification: the 2’-OH of the branching point «A» attacks the phospho-diester bond betwee exon and intron

2nd transesterification: the 3’-OH of the exon attacks the phosphodiester bond betwee intron and exon

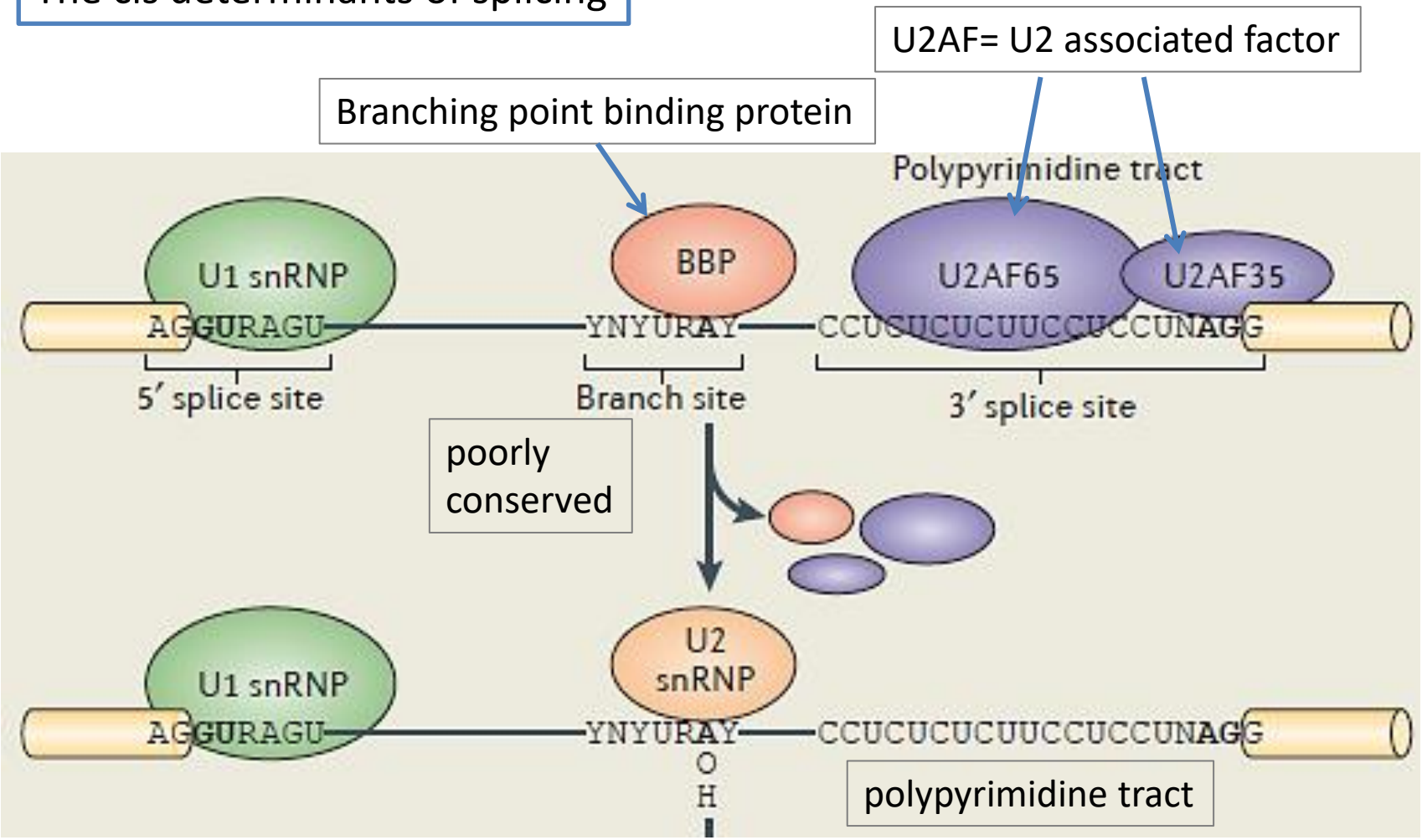
Spliceosome: snRNPs and associated proteins

Table 1. Core Subunits of Human U snRNPs

U snRNP	subunit gene	common subunit name(s)[a]	molecular mass (kDa)[b]	% U snRNP	recognizable domain/ functional site[c]
U1 (248.1 kDa)	RNU1	U1 snRNA	53.5	21.6	
	SNRPB, -B2, -D1, -D2, -D3, -E, -F, -G	seven Sm proteins	94.3	38.0	Sm
	SNRNPA	U1-A	31.3	12.6	RRM
	SNRNP70	U1-70k	51.6	20.8	RRM; SR repeat
	SNRNPC	U1-C	17.4	7.0	Znf
U2 (987.4 kDa)	RNU2	U2 snRNA	61.2	6.2	
	SNRPB, -B2, -D1, -D2, -D3, -E, -F, -G	seven Sm proteins	94.3	9.6	Sm
	SNRPA1	U2A′	28.4	2.9	LRR
	SNRPB2	U2B″	25.4	2.6	RRM
	SF3A1	SF3a120	88.9	9.0	SWAP; UBQ domain
	SF3A2	SF3a66	49.3	5.0	Znf
	SF3A3	SF3a60	58.6	5.9	Znf; SAP
	SF3B1	SF3b155	145.8	14.8	HEAT repeat
	SF3B2	SF3b145	100.2	10.1	SAP
	SF3B3	SF3b130	135.5	13.7	DExH/D
	SF3B4	SF3b49	44.4	4.5	RRM
	SF3B5	SF3b10	10.1	1.0	
	SF3B14	SF3b14a; p14	14.6	1.5	RRM
	PHF5A	SF3b14b; Rds3	12.4	1.3	PHD-like
	DDX46	DDX46; hPrpSp	117.4	11.9	DExH/D; SR repeat
	SMNDC1	SPF30/SMNrp	26.7	2.7	Tudor domain
U5 (1055.7 kDa)	RNU5	U5 snRNA	37.6	3.6	
	SNRPB, -B2, -D1, -D2, -D3, -E, -F, -G	seven Sm proteins	94.3	8.9	Sm
	TXNL4A	U5-15K	16.9	1.6	TRX
	SNRNP40	U5-40K	39.3	3.7	WD40
	CD2BP2	U5-52K	37.6	3.6	GYF
	DDX23	U5-100K; hPrp28	95.6	9.1	DExH/D; SR repeat
	PRPF6	U5-102K; hPrp6	106.9	10.1	HAT/TPR repeats
	EFTUD2	U5-116K; hSnu114	109.4	10.4	EF2-like fold; GTPase
	SNRNP20	U5-200K; hBrr2	244.5	23.2	DExH/D
	PRPF8	U5-220k; hPrp8	273.6	25.9	RNase H-fold; RRM; Jab1/MPN
U4/U6 (589.1 kDa)	RNU4	U4 snRNA	46.9	8.0	
	RNU6	U6 snRNA	34.6	5.9	
	SNRPB, -B2, -D1, -D2, -D3, -E, -F, -G	seven Sm proteins	94.3	16.0	Sm
	LSM2, -3, -4, -5, -6, -7, -8	seven LSm proteins	78.9	13.4	Sm
	NHP2L1	15.5K	14.2	2.4	
	PPIH	U4/U6-20K; SnuCyp-20	19.2	3.3	cyclophilin-like
	PRPF31	U4/U6-61K; hPrp31	55.5	9.4	Nop
	PRPF4	U4/U6-60K; hPrp4	58.4	9.9	WD40
	PRPF3	U4/U6-90K; hPrp3	77.5	13.1	PWI
	SART3	p110; SART3; hPrp24	109.6	18.6	HAT repeats; RRM

From: Van der Feltz et al., 2012

The cis determinants of splicing



This is GT-AG introns (by far the most frequent)
A secondary type exist (xx-xx), requiring U11 and U12 snRNP

Code	Represents	Complement
A	Adenine	T
G	Guanine	C
C	Cytosine	G
T	Thymine	A
Y	Pyrimidine (C or T)	R
R	Purine (A or G)	Y
W	weak (A or T)	W
S	strong (G or C)	S
K	keto (T or G)	M
M	amino (C or A)	K
D	A, G, T (not C)	H
V	A, C, G (not T)	B
H	A, C, T (not G)	D
B	C, G, T (not A)	V
X/N	any base	X/N
-	Gap	-

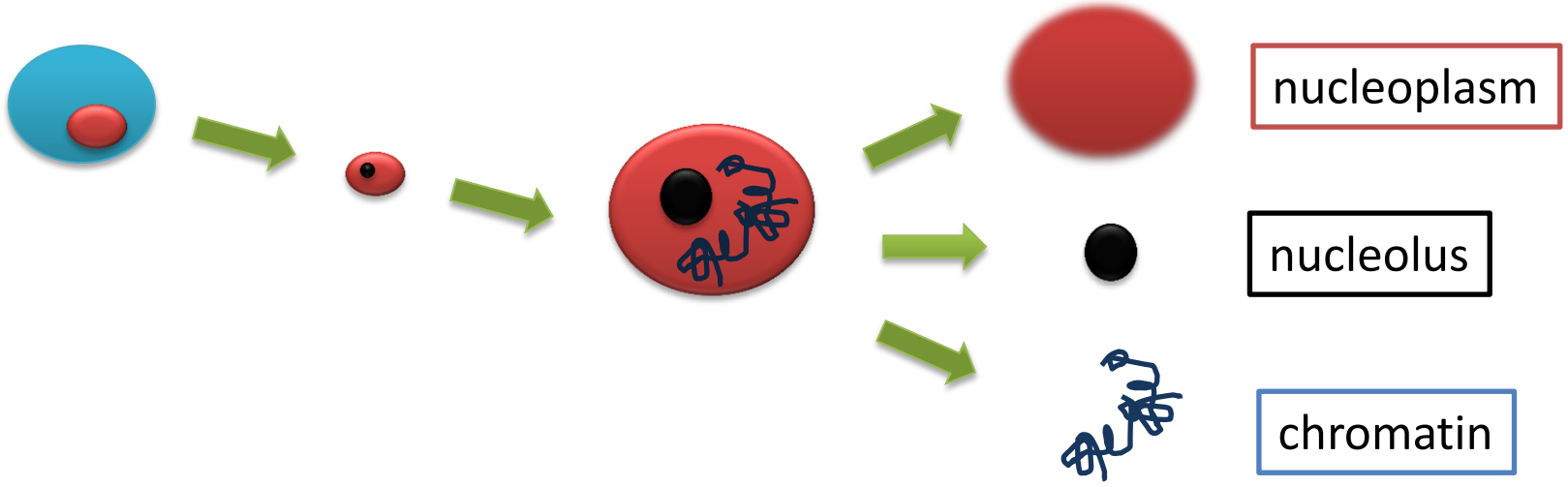
DNA base code

splicing news

When, where ?

Co-transcriptional or post-transcriptional ?

let’s see how ENCODE has approached it



The transcriptome of nuclear subcompartments

(reading Djebali et al, 2012 ENCODE series)

.....

*For the K562 cell line, we also analysed RNA isolated from three subnuclear compartments (**chromatin, nucleolus and nucleoplasm**).*

Almost half (18,330) of the GENCODE (v7) annotated genes detected for all 15 cell lines (35,494) were identified in the analysis of just these three nuclear subcompartments. In addition, there were as many novel unannotated genes found in K562 subcompartments as there were in all other data sets combined.

.....

enormous variety of transcripts

The interrogation of different subcellular RNA fractions provides snapshots of the status of the RNA population along the RNA processing pathway.

Djebali et al., 2012

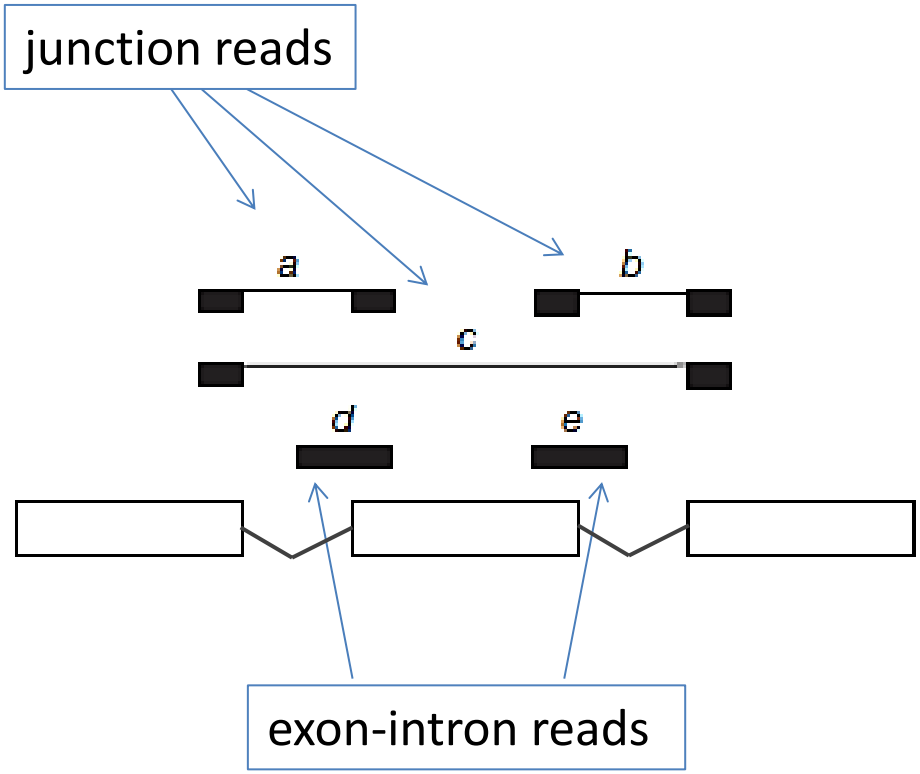
...we confirm that splicing predominantly occurs during transcription.

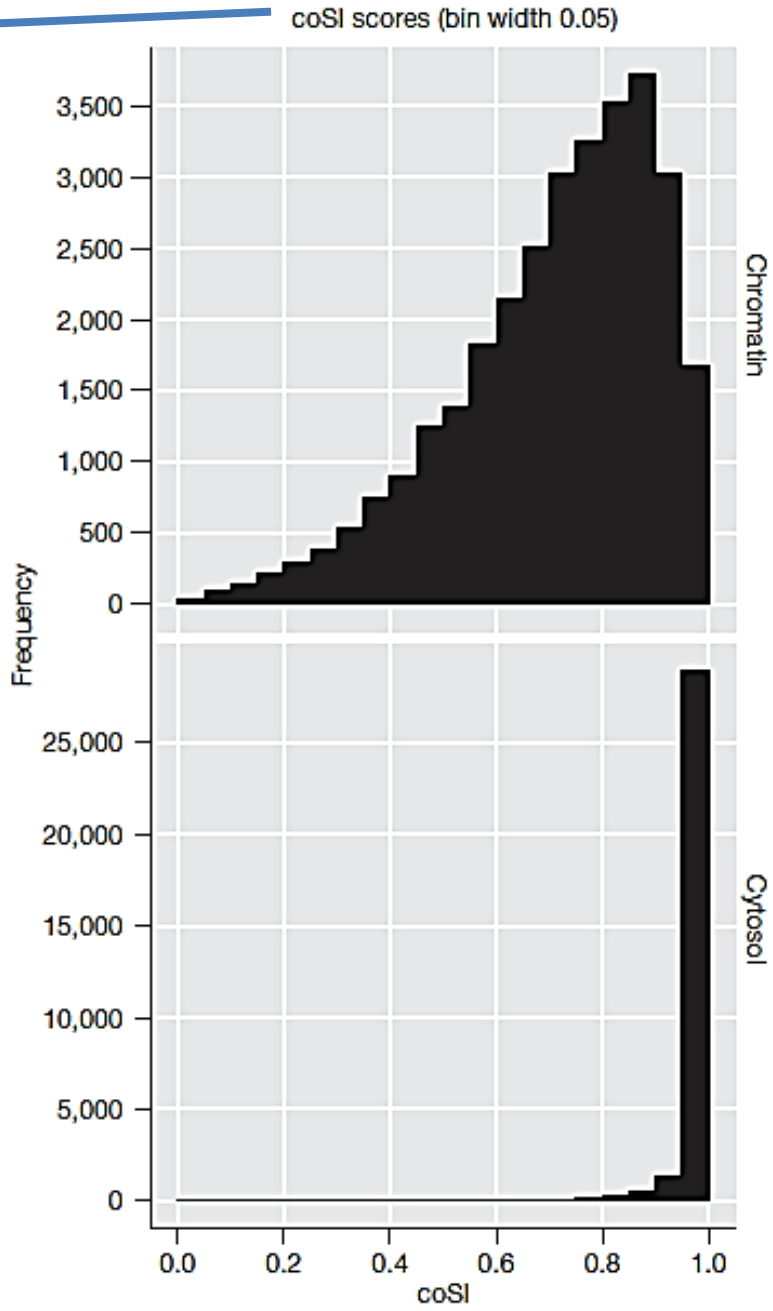
...introns are already being spliced in chromatin-associated RNA—the fraction that includes RNAs in the process of being transcribed

....we found strong enrichment specifically of spliceosomal small nuclear RNAs in this RNA fraction

..... we have observed that ***exons in the process of being spliced are enriched in a number of chromatin marks***

coSI = the ratio between junction reads and exon-intron reads





Djebali et al., 2012

Alternative splicing

Sequences at the borders of exon-intron and within the intron are similar but can vary.

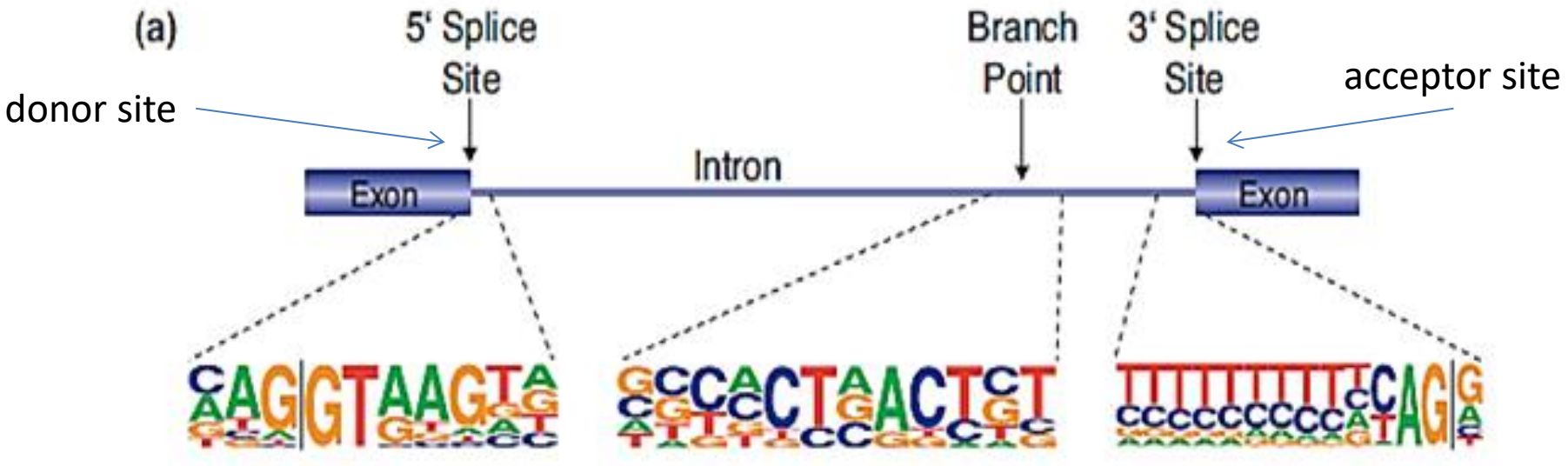
Splice sites can be **strong** or **weak** depending on how far their sequences diverge from the consensus sequence.

This determines their affinities for cognate splicing factors

In general, strong splice sites lead to constitutive splicing and full usage of the site

Pay attention to this concept, it will ground the discussion on PolII elongation rate as determinant of AS

The first chance to obtain regulation derives from how exons are recognized



Variations of these sequences can give «stronger» and «weaker» splicing sites

genomic view

From: McManus & Graveley, COGD, 2011

Indeed, **in addition** to the sequences directly regulating binding of spliceosome components, in both Exons and Introns sequence motifs exist that regulate the use of splice sites.

Named after their location and effect:

ESE: exonic splicing enhancers
ESS: exonic splicing silencers
ISE: intronic splicing enhancers
ISS: intronic splicing silencers

They are therefore ***cis*-elements** for splicing regulation

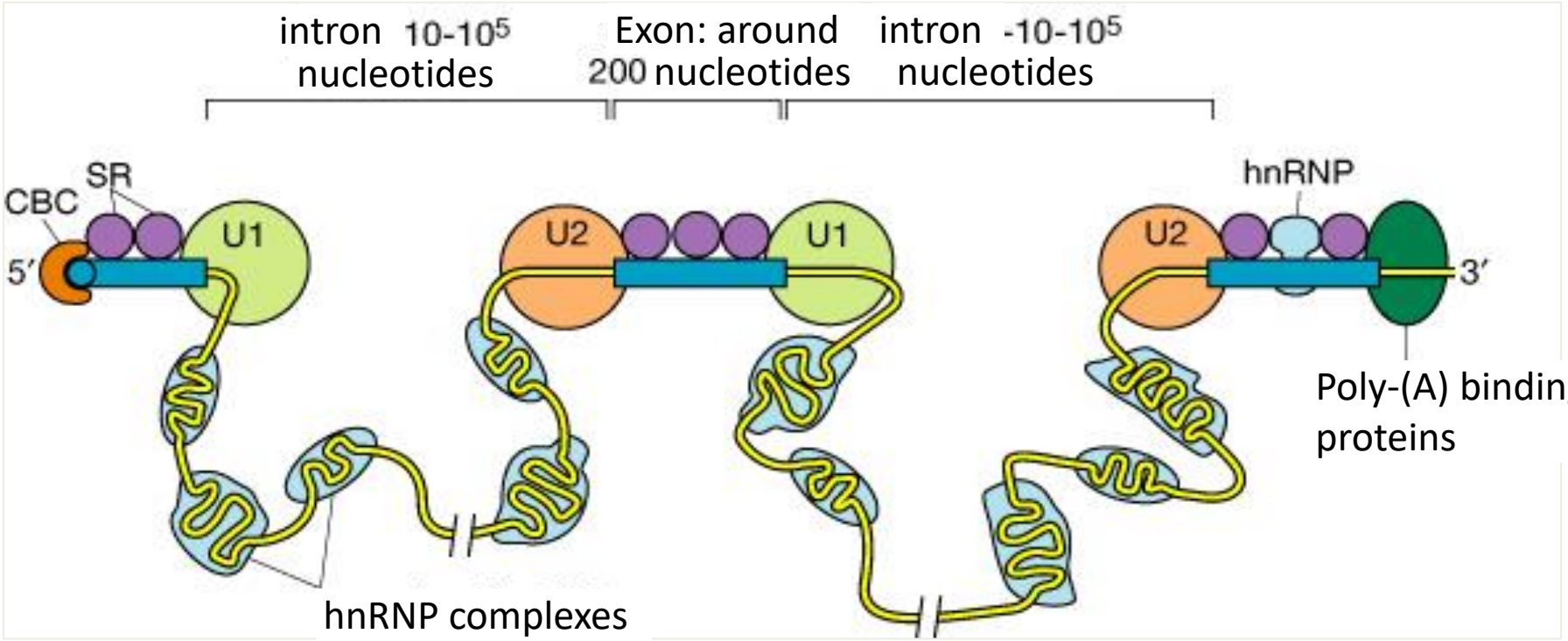
Trans-acting factors of splicing are proteins binding to these elements.
They belong to the general class of RNA Binding Proteins (**RBP**)

Three categories:

1. SR proteins and SR-like

2. hnRNP

3. tissue-specific and context-specific factors


intron 10-10^5 nucleotides
Exon: around 200 nucleotides
intron -10-10^5 nucleotides
CBC
SR
5′
U1
U2
U1
U2
hnRNP
3′
Poly-(A) binding proteins
hnRNP complexes

SR proteins = splicing regulators

Domains:
The most typical domain is an alternating Arginine-Serine-rich domain, called “RS domain”: it is a protein-protein interaction domain.

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

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

Binding sites:
Mostly at Exons, sometimes also to ISE (intronic splicing enhancers)

Activity:
Mostly activatory toward the most proximal exon.
Exon definition.

Canonical SR proteins

SR-like
(other protein containing an RS domain

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*

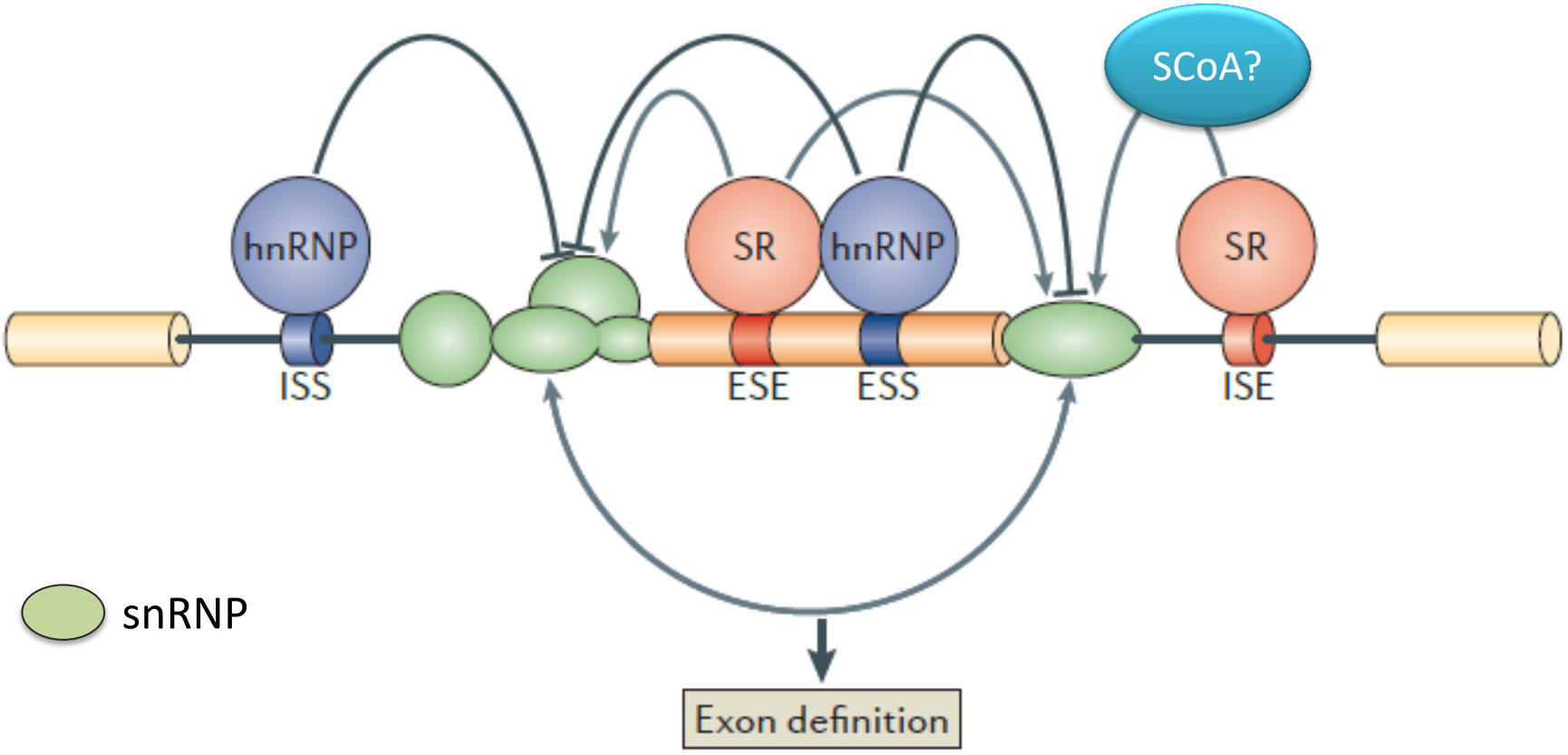
hnRNP proteins (heterogeneous nuclear Ribo Nucleic Protein)

- Many different families
- Usually bind intronic sites
- Intron definition
- Several other roles have been ascribed to individual members, e.g. cytoplasmic localization.

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*

The effect of SR and hnRNP binding is either to stabilize or destabilize the interaction of basal splicing factors (snRNPs) with the splicing sites
This action can be direct protein-protein contact, or mediated by splicing co-activators.



SRs and (some) hnRNPs can be regulated by signalling pathways (*e.g. by phosphorylation*);

most SRs and most hnRNPs have ubiquitous expression

tissue-specific AS (very frequent) requires additional tissue-specific factors

Hence, additional cis-elements sould be present in regulated pre-mRNAs.

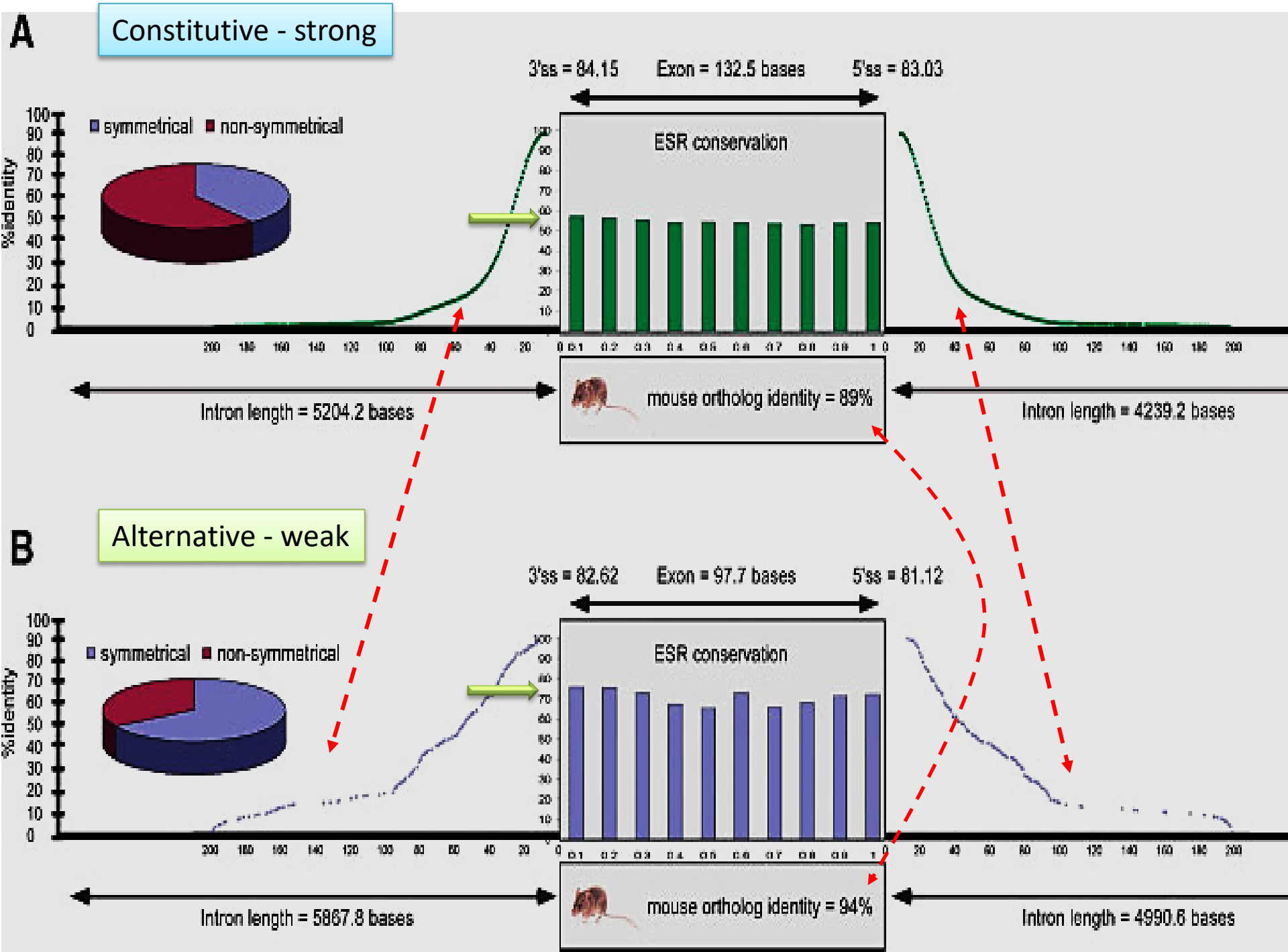
One way to answer this question is to compare sequences around **alternative** exons to those of **constitutive** exons.

comparative genomics

Human and mouse transcriptomes are quite well characterized, making it possible to classify exons as **constitutive** or **alternative** based on real expression data (microarrays, RNA-seq).

Exons were normalized in lenght and flanking introns explored within 200 bp.

From Kim et al., BioEssays 30:38–47


A
Constitutive - strong
3'ss = 84.15
Exon = 132.5 bases
5'ss = 83.03
%identity
symmetrical
non-symmetrical
ESR conservation
Intron length = 5204.2 bases
mouse ortholog identity = 89%
Intron length = 4239.2 bases
B
Alternative - weak
3'ss = 82.62
Exon = 97.7 bases
5'ss = 81.12
%identity
symmetrical
non-symmetrical
ESR conservation
Intron length = 5867.8 bases
mouse ortholog identity = 94%
Intron length = 4990.6 bases

2.1 Features of **Alternative** *versus* **Constitutive** Exons

More conserved

A-Exons in both human and mouse are **more conserved** than C-Exons

Conservation higher toward exon edges and extends farther in introns

Weak splice sites

A-Exons have **weak splice sites** (especially *Cassette Exons*)

Shorter

A-Exons (esp. *cassette exons*) are **shorter** and flanked by longer introns.

Symmetric

The percentage of **symmetrical** exon is definitely higher in A-Exons

Cassette exon: a symmetric exon that may undergo exon skipping

From Kim et al., 2007, BioEssays 30:38–47