Chapter 5

Post-transcriptional regulation

- **Co-transcriptional**
- **Post-transcriptional**

Co-transcriptional

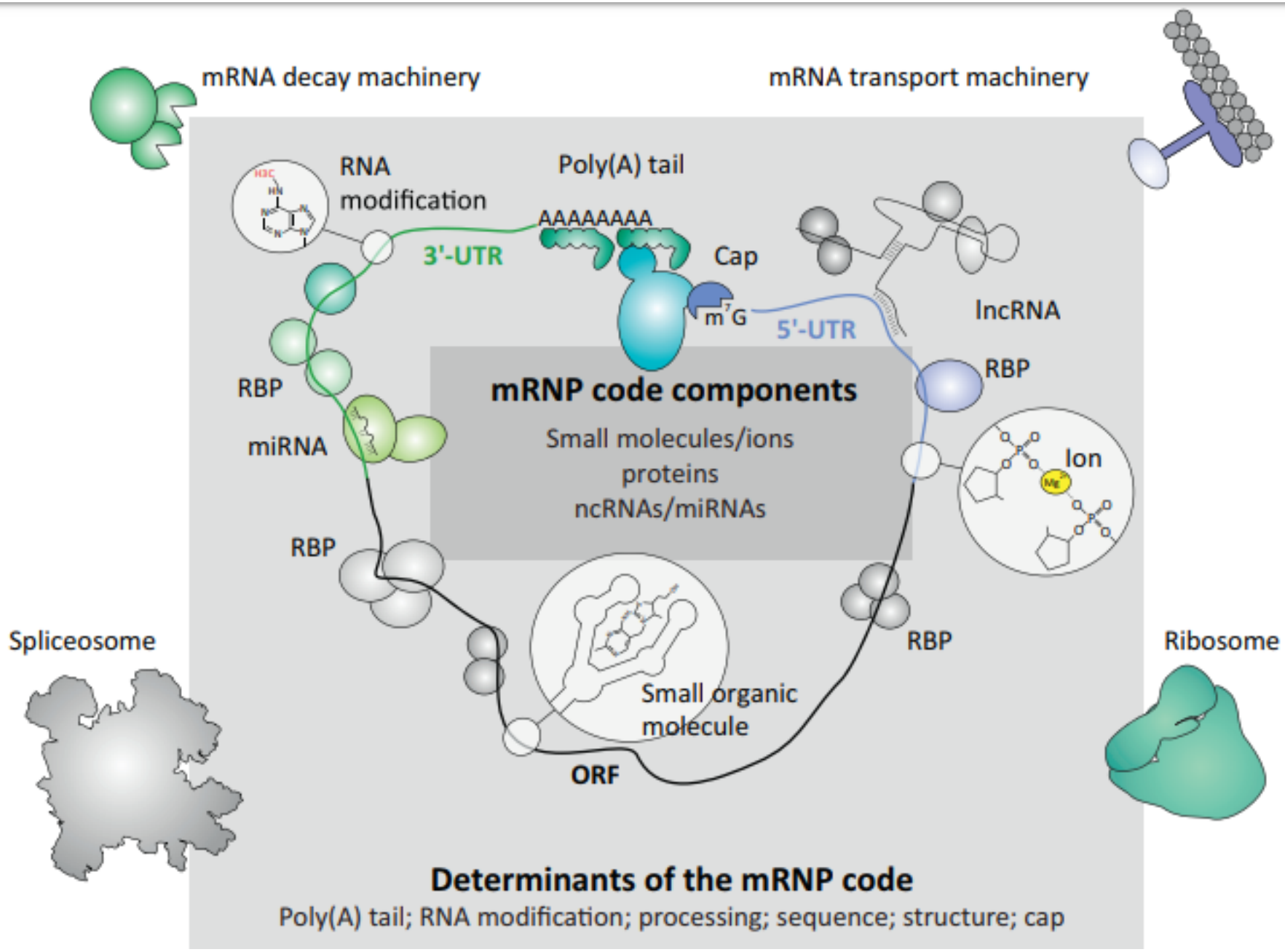
- 5’-capping
- Exon splicing
- 3’-polyadenylation
- RBP binding

mRNP

Post-transcriptional

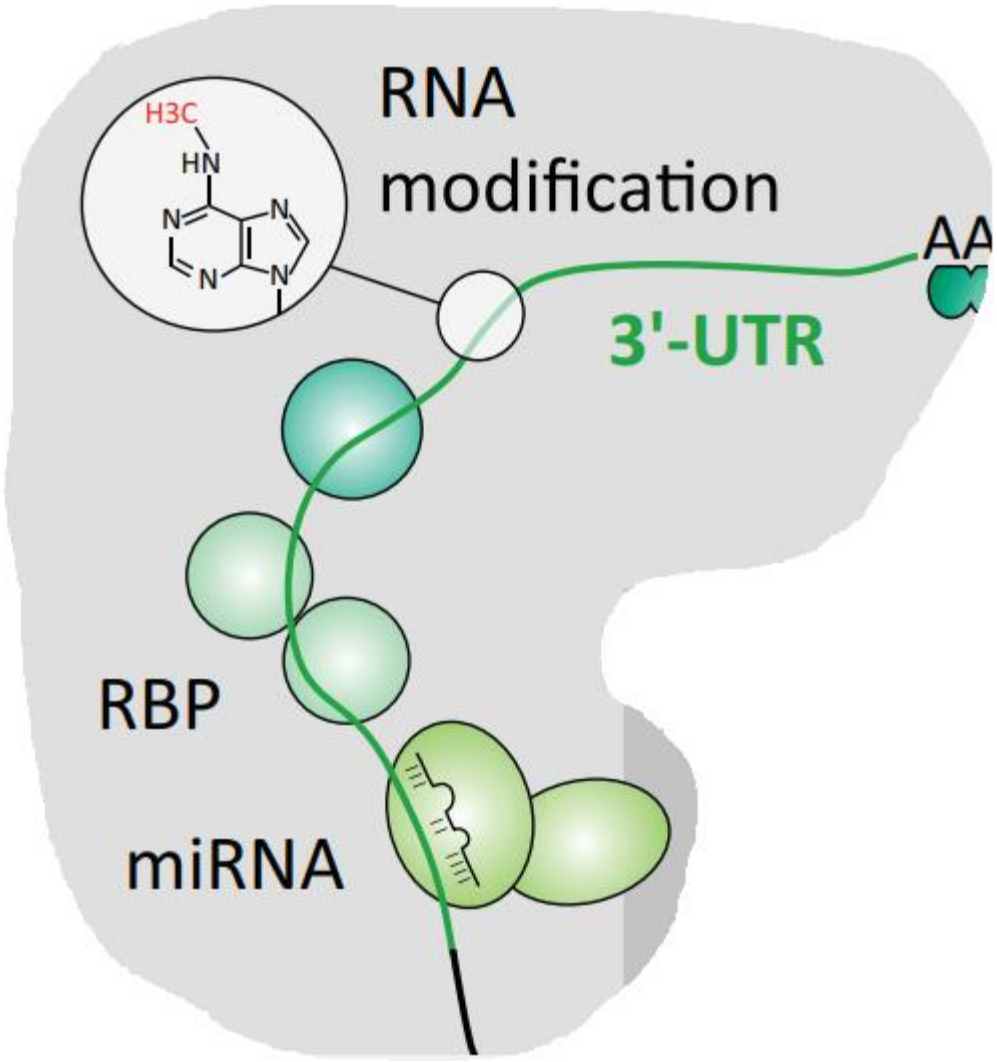
- Exon splicing
- Methylation of N6-Adenine
- Editing

Quality control
NRE (nuclear retention elimination)



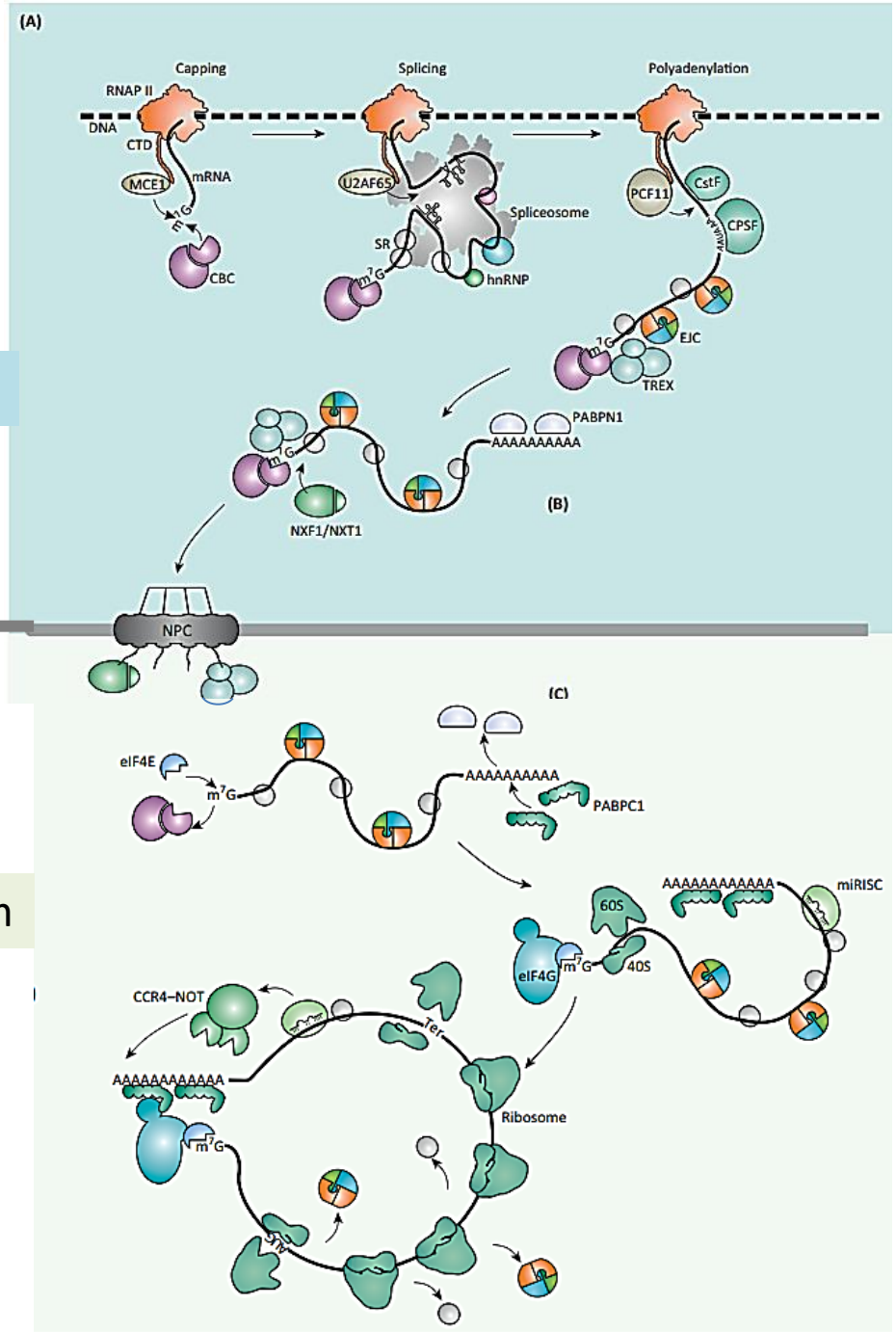
Gehring et al., Trends in Biochem Sci. 2017

Trends in Biochemical Sciences



Gehring et al., Trends in Biochem Sci., in press 2017

How the mRNP is assembled



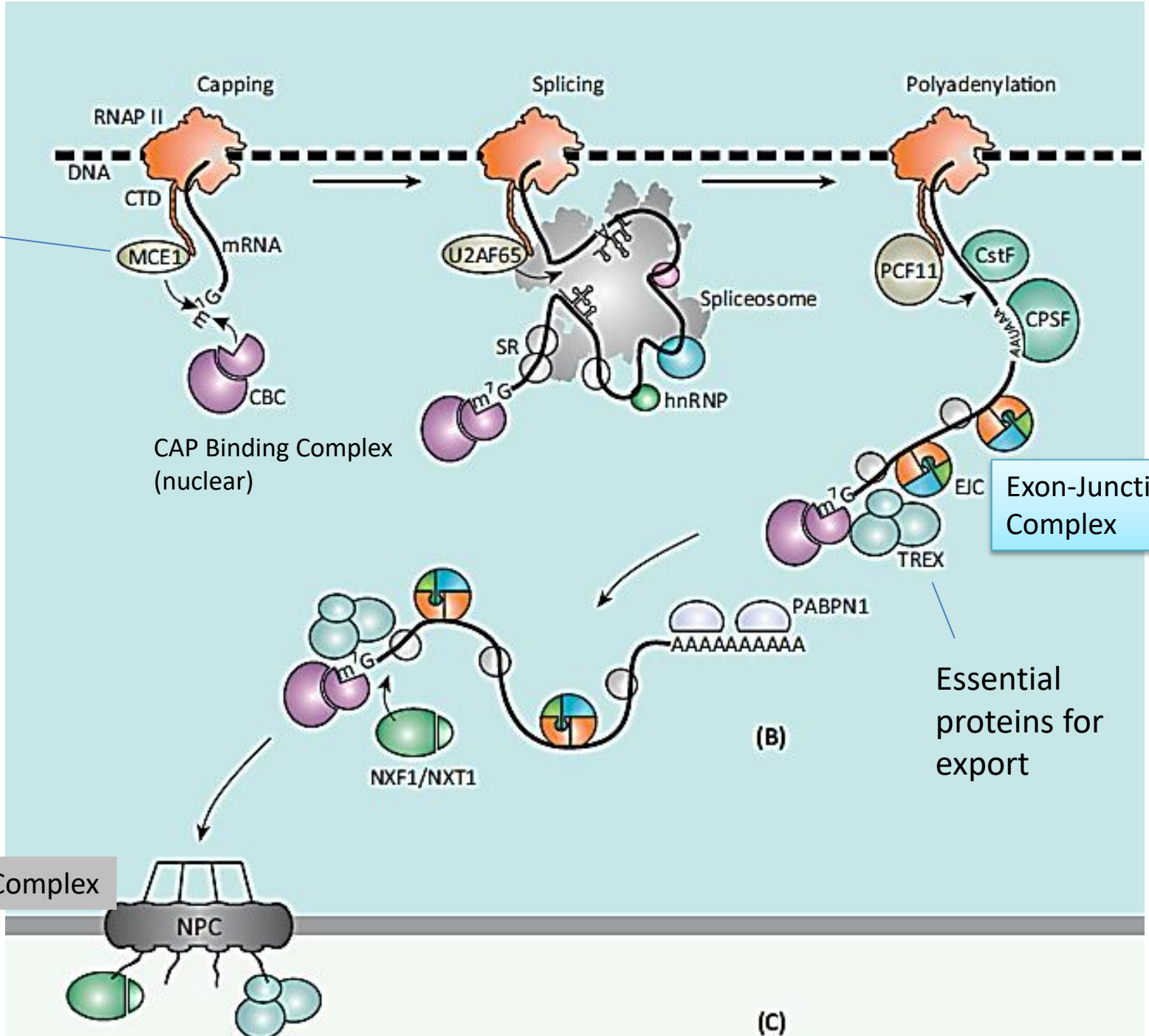
nucleus

Nuclear membrane

cytoplasm

Gehring et al., Trends in Biochem Sci., in press 2017

mRNA Capping Enzyme:
RNA Guanylyl-transferase and 5'-Phosphatase

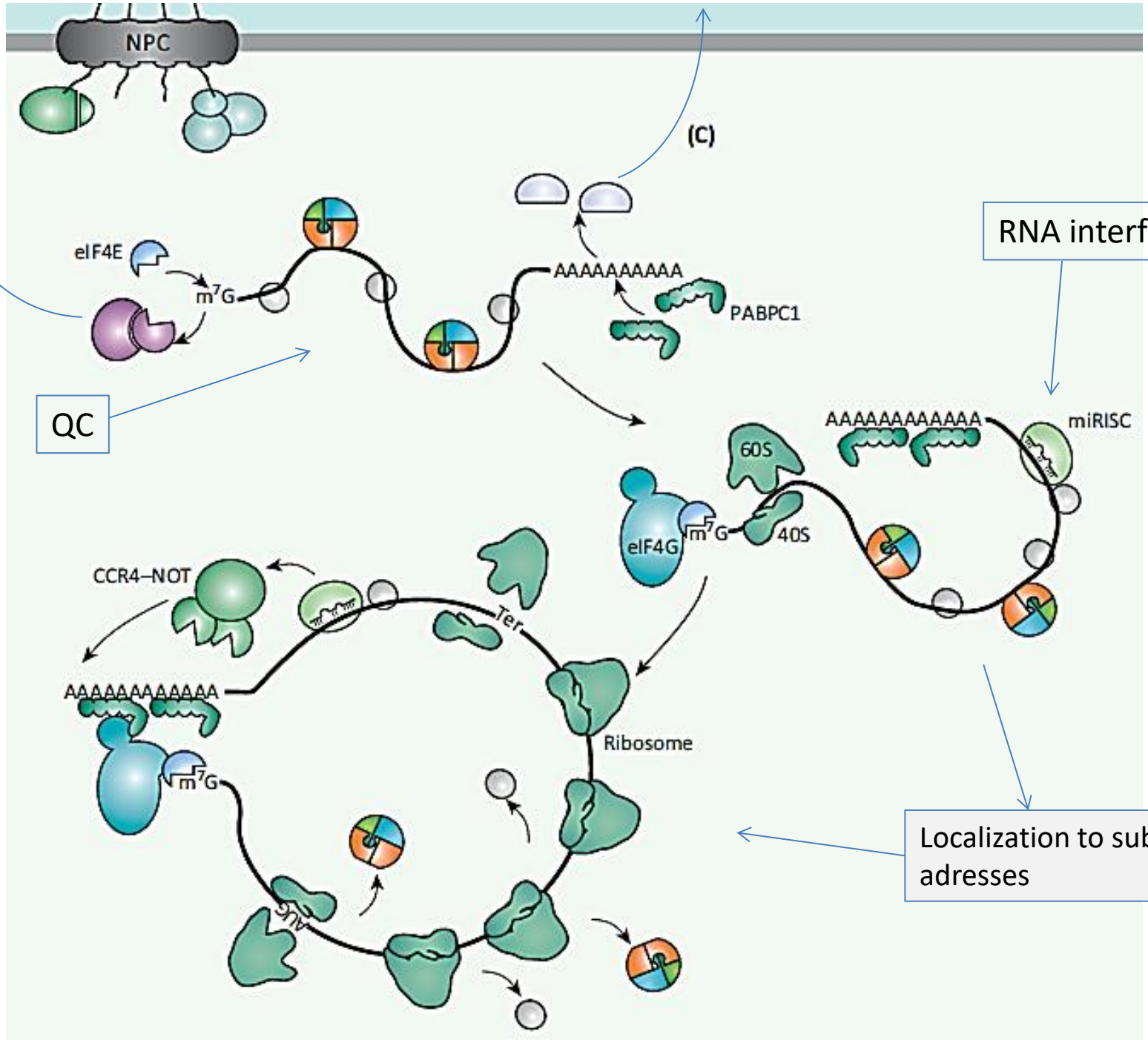


CAP Binding Complex (nuclear)

Exon-Junction Complex

Essential proteins for export

Nuclear Pore Complex


NPC
(C)
eIF4E
m7G
AAAAAAAAAA
PABPC1
QC
RNA interference
AAAAAAAAAAAA
miRISC
60S
eIF4G
m7G
40S
CCR4–NOT
Ter
AAAAAAAAAAAA
Ribosome
m7G
AUG
Localization to subcellular adresses

Alternative splicing: a pivotal step between eukaryotic transcription and translation

Textbook

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.

NATURE REVIEWS | MOLECULAR CELL BIOLOGY VOLUME 14 | MARCH 2013 | 153

- **RNA processing**
- **Splicing mechanisms**
- **Proteins and small RNA involved in splicing**

RNA processing
Capping – Splicing – polyadenylation - editing

- **RNA processing**
- **Splicing mechanisms**
- **Proteins and small RNA involved in splicing**

Socrative

Room:**3CKURJ4KY**

Quiz: RNA processing

https://b.socrative.com/teacher/#dashboard

5’ Capping

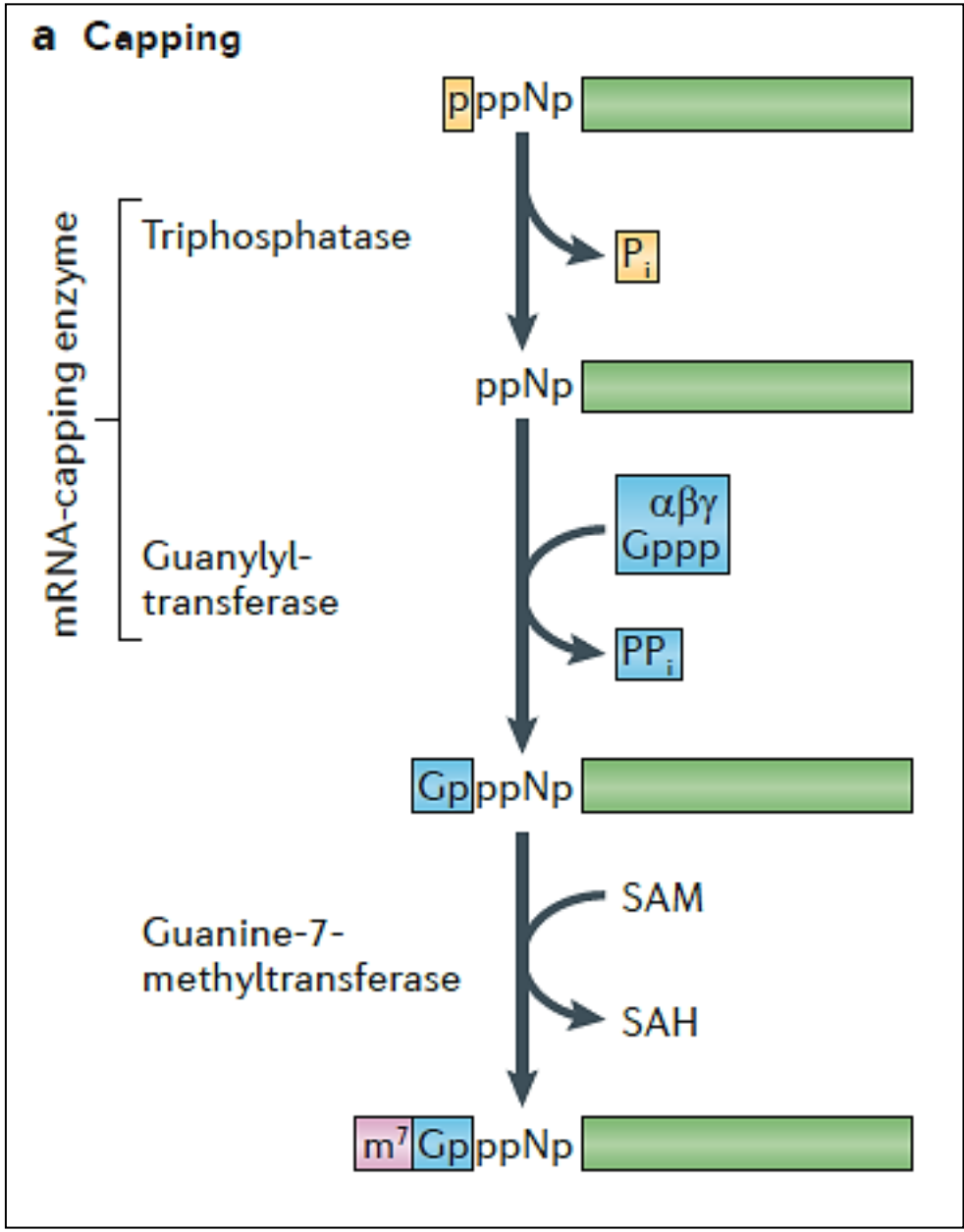


Figure 1 | **The major co-transcriptional mRNA processing steps.** Human protein names are given throughout.

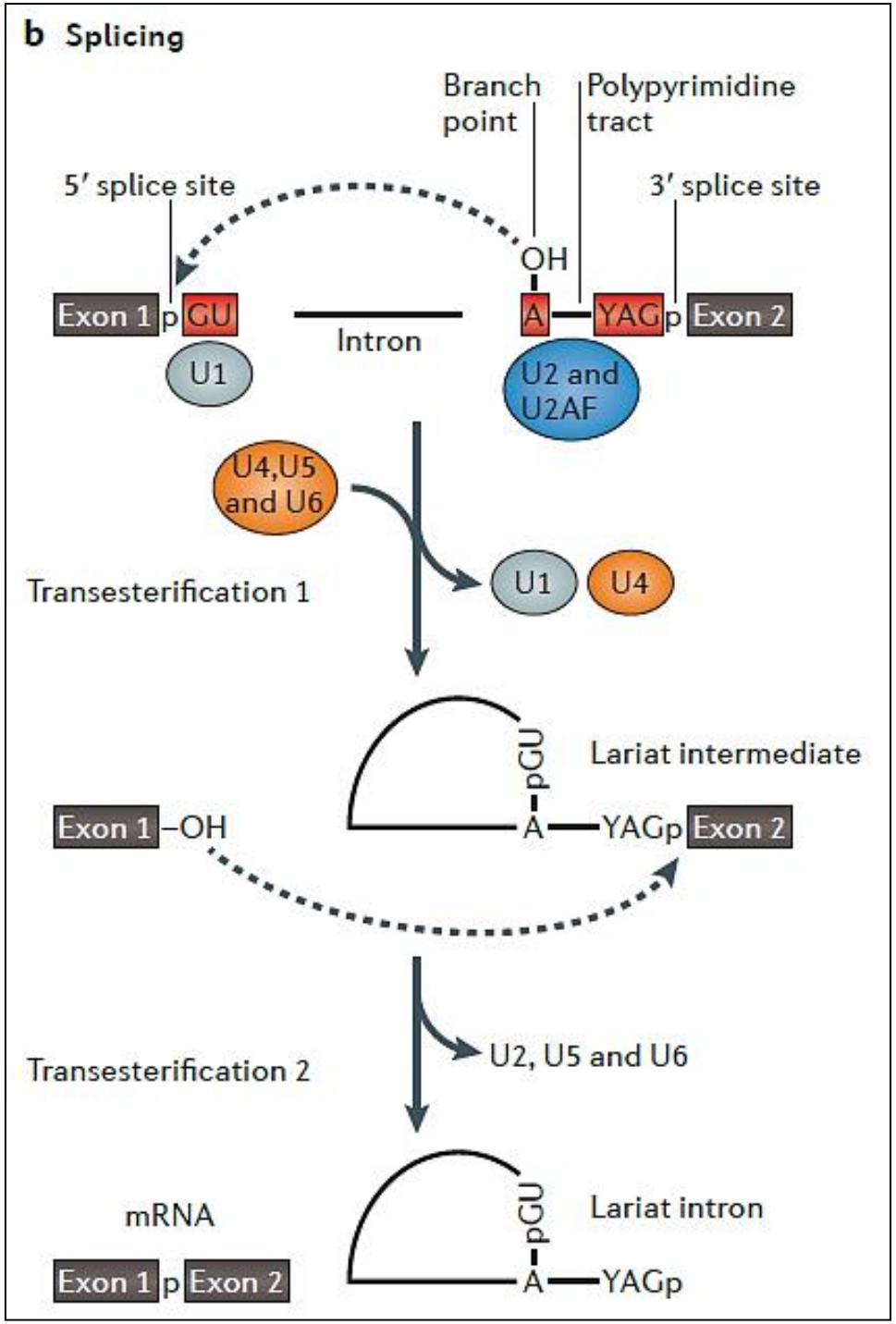
a | The RNA is shown in green; both GTP and the added guanosine cap (Gp) are shown in blue. The mRNA-capping enzyme in metazoans is bifunctional and has both triphosphatase and guanylyl-transferase activities that remove the γ-phosphate of the nascent transcript and transfer GMP from the GTP donor, respectively. The methyl donor *S*-adenosyl-l-methionine (SAM) is converted to S-adenosyl-l-homocysteine (SAH), which results in the 7-methylguanosine cap (shown in pink).

Bentley, 2014, Nat Rev Genetics 15:163-75.

Exon splicing

Figure 1 | **The major co-transcriptional mRNA processing steps.** Human protein names are given throughout.

b | Splicing clips out an intron or intervening sequences as a lariat and ligates the flanking exons together through two transesterification reactions. Conserved intronic splicing elements are indicated in red. Spliceosomal U1, U2, U4, U5 and U6 small nuclear ribonucleoprotein particles (snRNPs) and U2 auxiliary factor (U2AF) are shown, but numerous spliceosomal proteins are omitted.



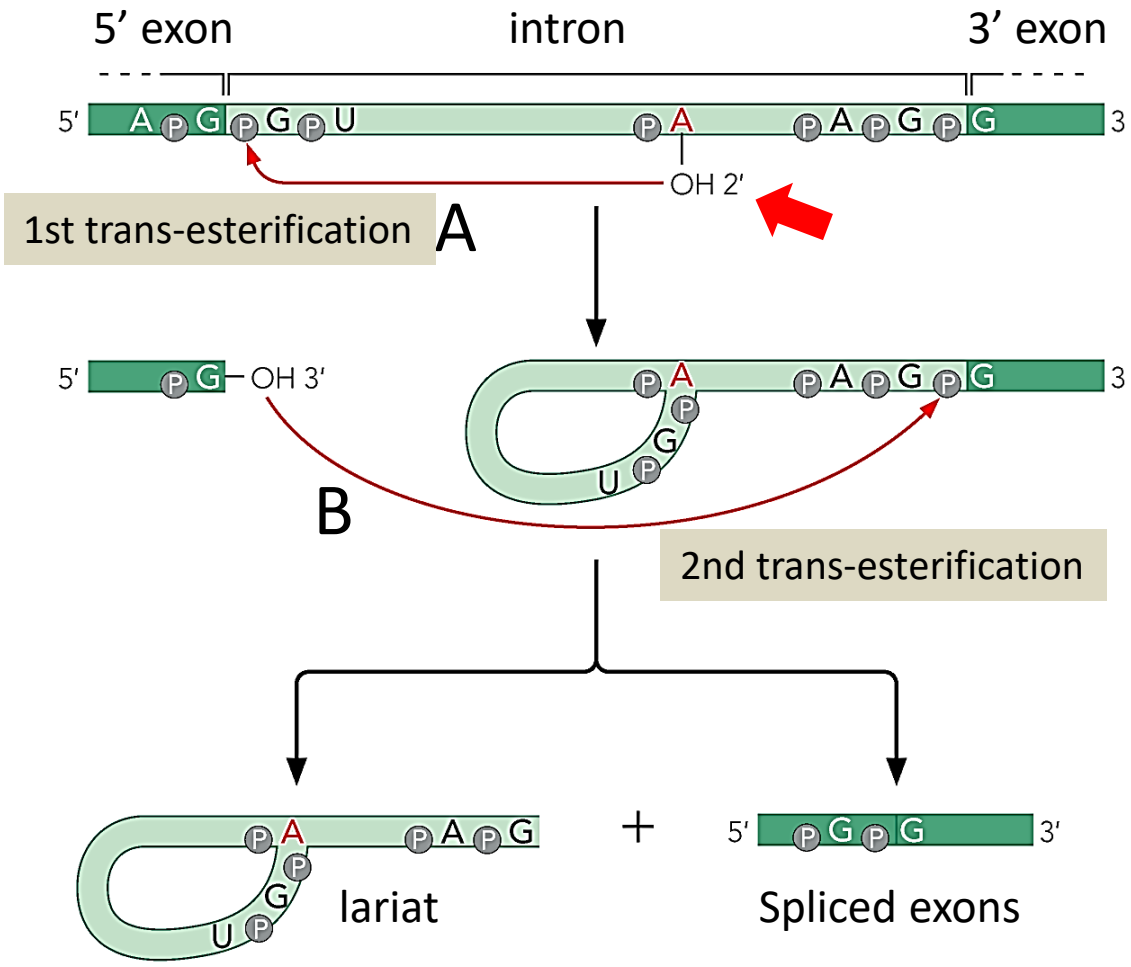
Bentley, 2014, Nat Rev Genetics 15:163-75.

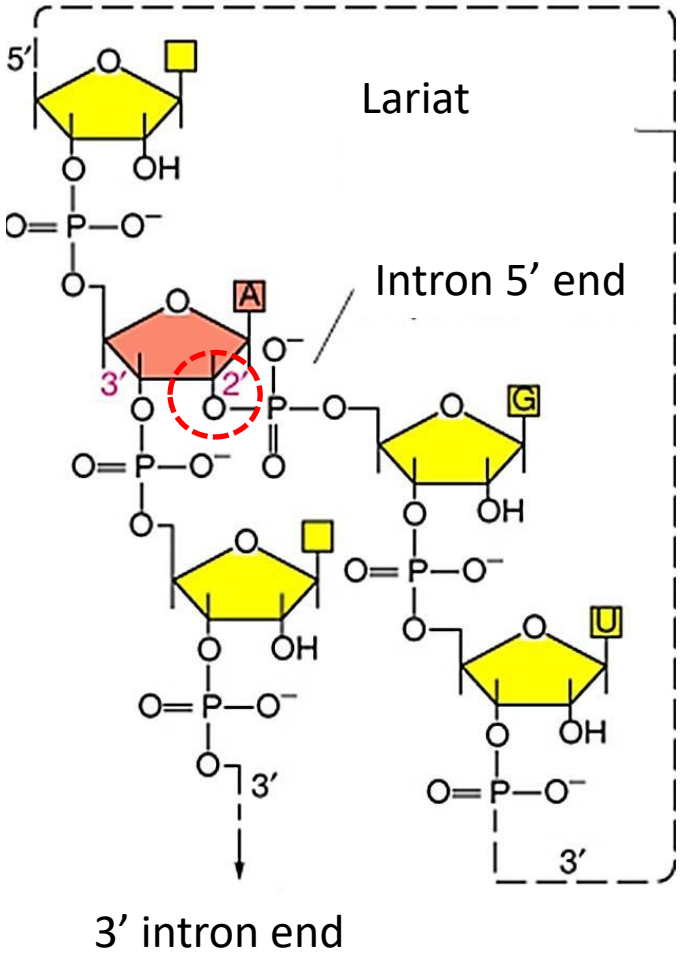
Pre-mRNA splicing occurs in two ATP-Independent transesterification reactions

A) first transesterification

B) second transesterification

Biochemistry of RNA processing





Is the spliceosome a simple machine ?

Not at all....!!

It has more or less the complexity of a ribosome

Indeed, the spliceosome is quite large complex:

- First, since it is formed by some three-hundred proteins and five U-RNAs
- Second, since it goes through dynamic assembly and disassembly

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Current Topic

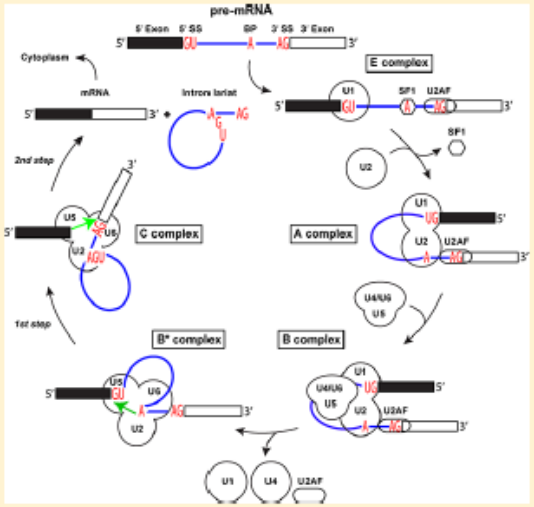
pubs.acs.org/biochemistry

Architecture of the Spliceosome

Clarisse van der Feltz, Kelsey Anthony, Axel Brilot, and Daniel A. Pomeranz Krummel*

Department of Biochemistry, Brandeis University, 415 South Street, Waltham, Massachusetts 02454, United States

ABSTRACT: Precursor-mRNA splicing is catalyzed by an extraordinarily large and highly dynamic macromolecular assemblage termed the spliceosome. Detailed biochemical and structural study of the spliceosome presents a formidable challenge, but there has recently been significant progress made on this front highlighted by the crystal structure of a 10-subunit human U1 snRNP. This review provides an overview of our current understanding of the architecture of the spliceosome and the RNA−protein complexes integral to its function, the U snRNPs.

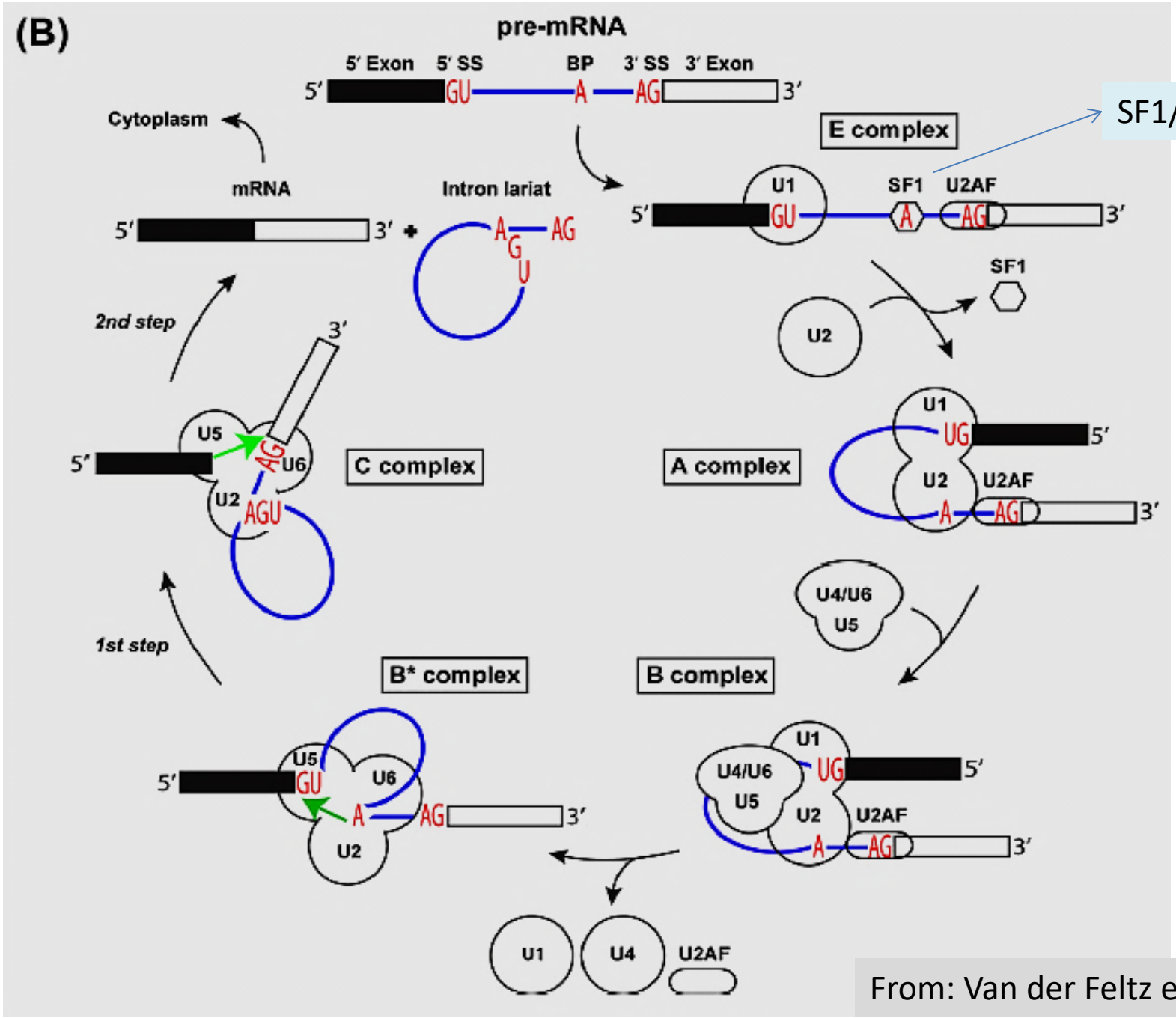


dx.doi.org/10.1021/bi201215r | Biochemistry 2012, 51, 3321−3333

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; hPrp5p	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



From: Van der Feltz et al., 2012

Do all introns start with GT and end with AG ?

Not all of them, there is a small number of introns following an alterative rule: they are called AT-AC introns, or also U12-type introns.

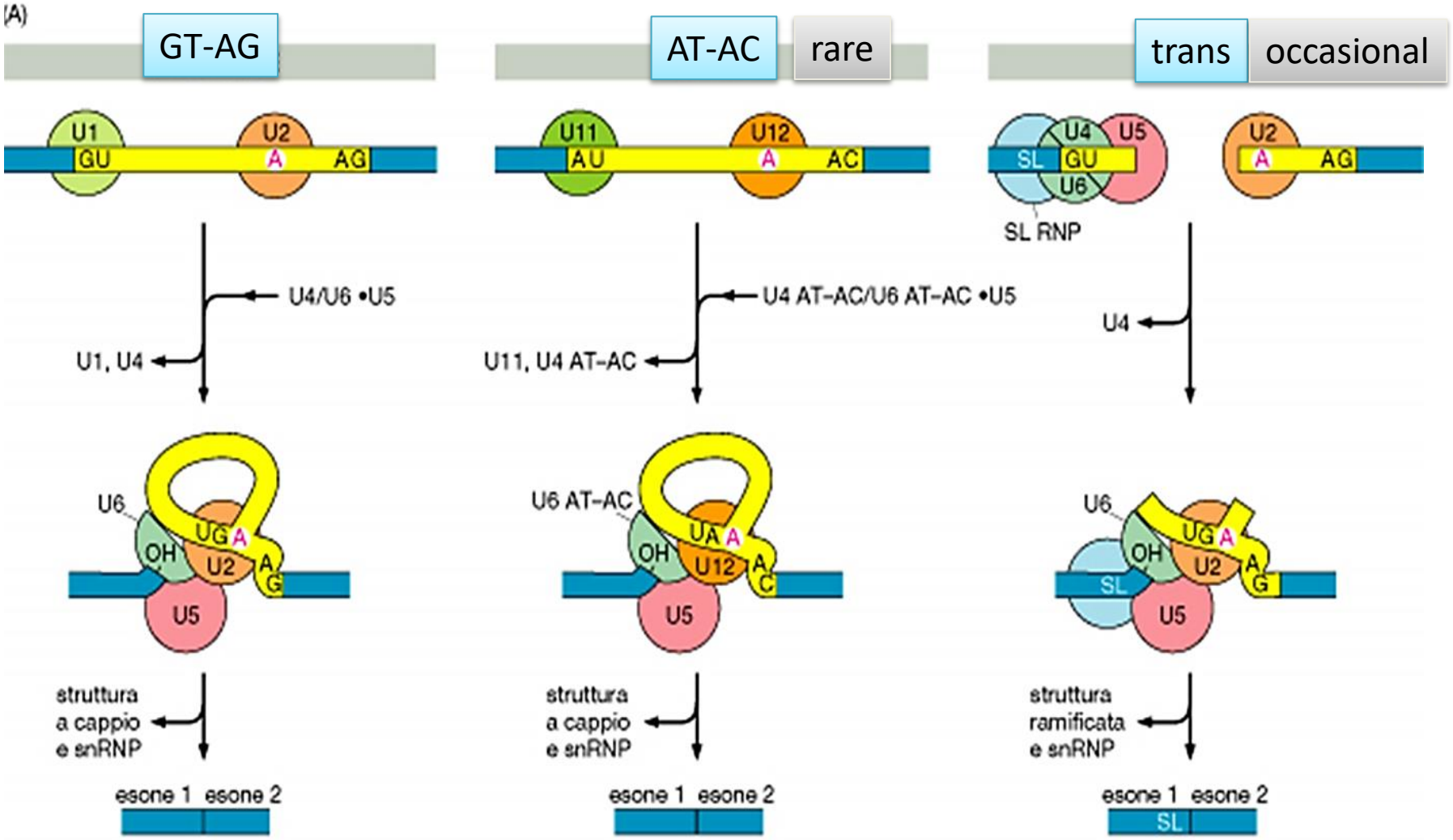
How many ?

Less than 0.5%

Is the same spliceosome taking care of them?

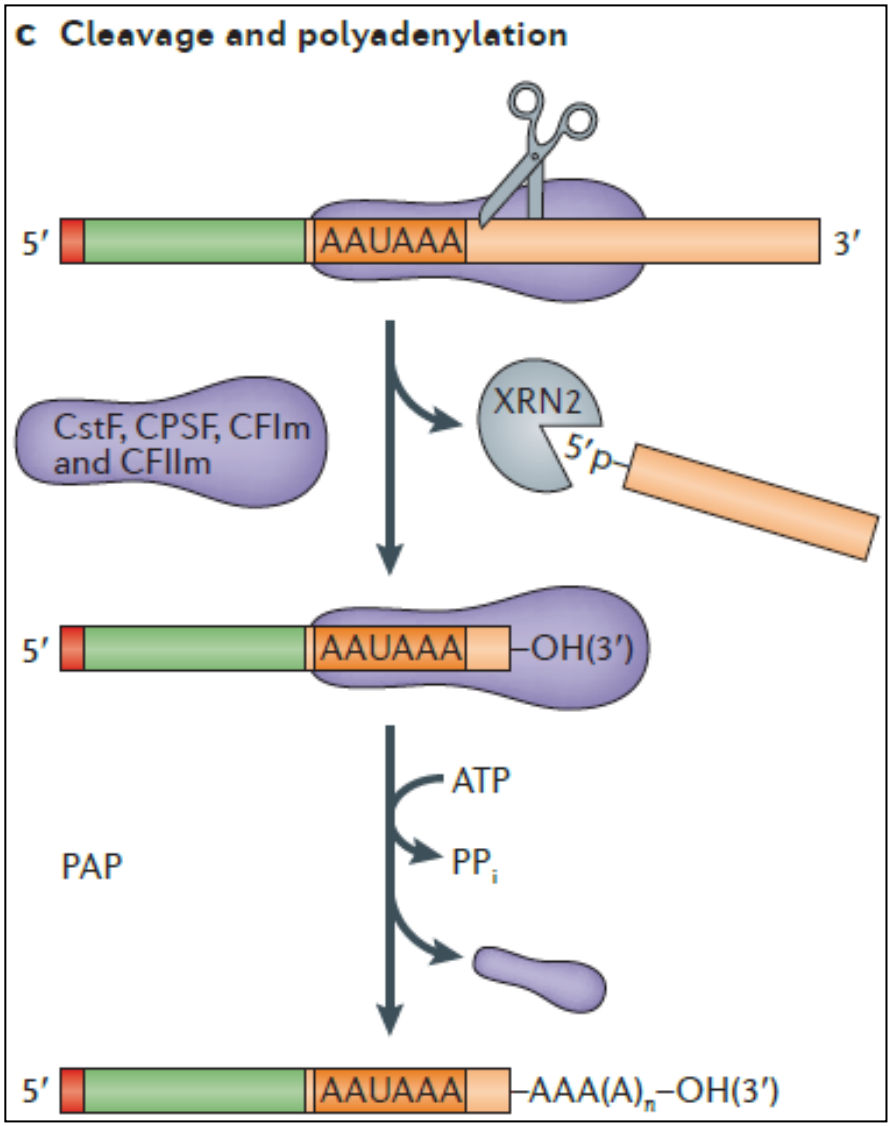
No: it is formed by U11, U12, U4atac, U6atac
While U5 snRNP is in common

Splicing types



3’-end cleavage and Polyadenylation

Figure 1 | **The major co-transcriptional mRNA processing steps.** Human protein names are given throughout.
c | 3‘ ends of mRNAs are formed by coupled cleavage and polyadenylation. Cleavage of mammalian pre-mRNAs occurs ~25 bases downstream of a consensus sequence (AAUAAA) and is carried out by the multi-subunit complex (shown in purple), which comprises cleavage stimulation factor (CstF), cleavage and polyadenylation specificity factor (CPSF) that bears the endonuclease, and cleavage factors I and II (CFIm and CFIIm). Poly(A) polymerase (PAP) adds the poly(A) tail. 3‘ ends of non-polyadenylated histone mRNAs (not shown) are also made co-transcriptionally by a cleavage complex that has many subunits in common with CstF and CPSF. The 5‘-to-3‘ RNA exonuclease 2 (XRN2) degrades RNA downstream of the cleavage site and facilitates transcription termination.



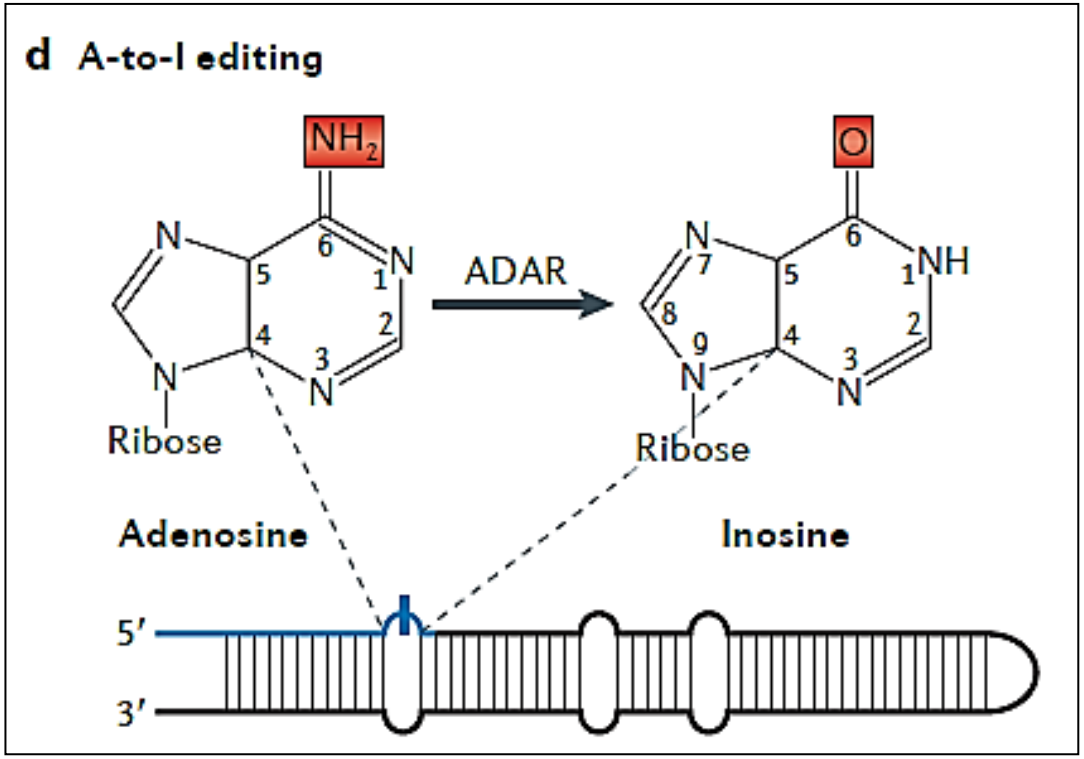
Bentley, 2014

RNA Editing

Figure 1 | **The major co-transcriptional mRNA processing steps.** Human protein names are given throughout.

d | Adenosine-to-inosine (A-to-I) editing is carried out by adenosine deaminases acting on RNA (**ADAR**), which deaminate adenosines into inosines. The folded *GLUR-2* pre-mRNA substrate is shown with the exon in blue and intron in grey.

Inosine behaves as a «**G**» in translation and secondary structures



Bentley, 2014

RNA Editing

Very difficult to detect ! Special sequencing techs and algorithms needed

Classical example is GluA2 glutamate receptor

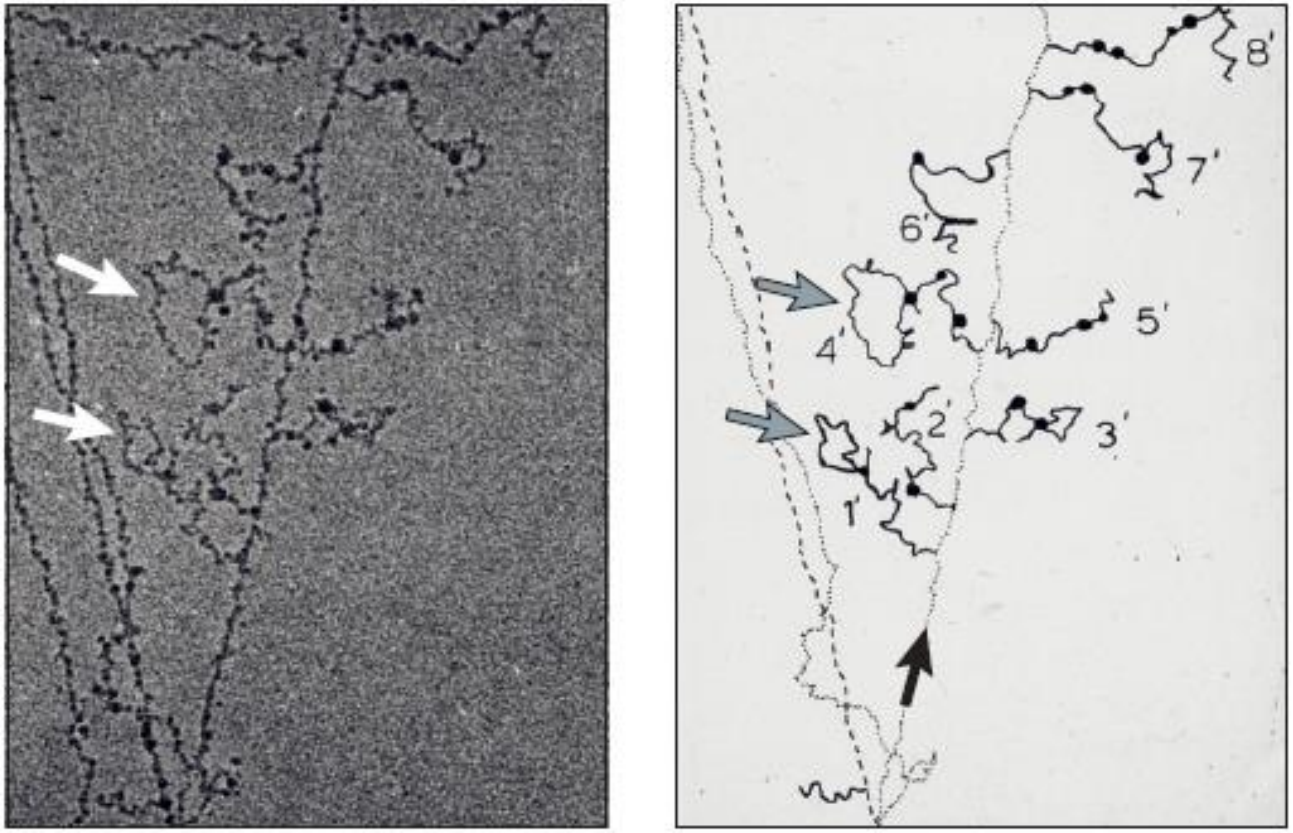
A to I editing at specific positions leads to critical amino acid substitutions in the receptor subunits, which induce profound alterations of the channel properties.

RNA-seq sudies evidenced many other cases, most to be confirmed experimentally.

Co-transcriptional RNA processing ?

Capping, polyadenylation and splicing are mainly co-transcriptional processes.

Remind: Research Paper -3 (Djebali et al., 2012 – ENCODE)



*This picture is taken from Beyer, A. L. & Osheim, Y. N. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. **2**, 754–765 (1988).*

Figure 2 | **The co-transcriptional nature of pre-mRNA processing. a** | Transcription of a gene and co-transcriptional processing of nascent transcripts are shown. 'Miller spread' electron micrograph (left) and its interpretation (right) are shown for a *Drosophila melanogaster* embryonic gene. The electron micrograph shows the DNA template with several engaged RNA polymerase II (Pol II) molecules and their associated nascent RNA transcripts with bound proteins (seen as dark blobs) that extend on either side of the DNA. Grey and white arrows mark introns that are spliced out co-transcriptionally. The black arrow indicates the direction of transcription along the DNA template.

Bentley, 2014

In all eukaryotes, RNA Polymerase II has a typical C-terminal domain (CTD) that is composed of repetitive peptide sequence. Human: (YSPTSPS) (52 repeats)

- CTD is not present in RNA Pol I and RNA Pol III (RNA produced by these undergo a totally different processing)
- Deletion, shortening or mutation of CTD **impairs processing**
- CTD co-immunoprecipitates a number of proteins needed for capping, splicing, poly-adenylation
- Factors needed for splicing and polyadenylation are found all along the transcribed region.

These observation led to the hypothesis that the RNA POL II CTD acts as a **platform** to carry processing factors and enzymes along nascent RNA

PolII CTD carries a number of processing factors & enzymes

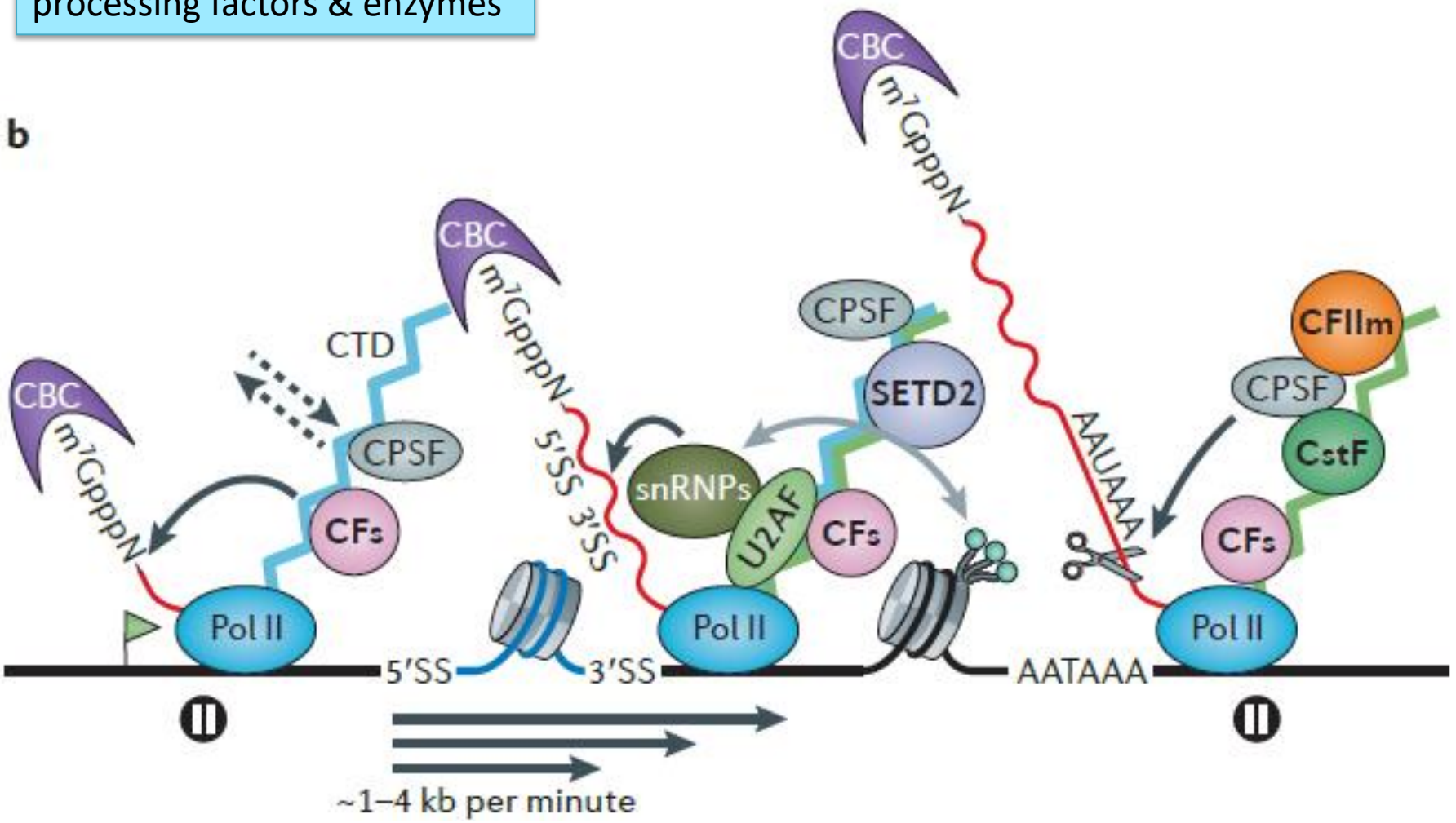


Figure 2 | **The co-transcriptional nature of pre-mRNA processing**

Bentley, 2014

Alternative splicing

H. sapiens	Estimated number of protein-coding genes: ca 20,000 Estimated number of transcripts: > 80,000 Estimated number of proteins: 250,000 – 1 million

S. cerevisiae:	253 genes contain introns only **3 genes** shown to have alternative splicing

H. sapiens:	>99% predicted to have exon-intron structure >95% predicted to undergo alternative splicing

Genecode statistics: https://www.gencodegenes.org/

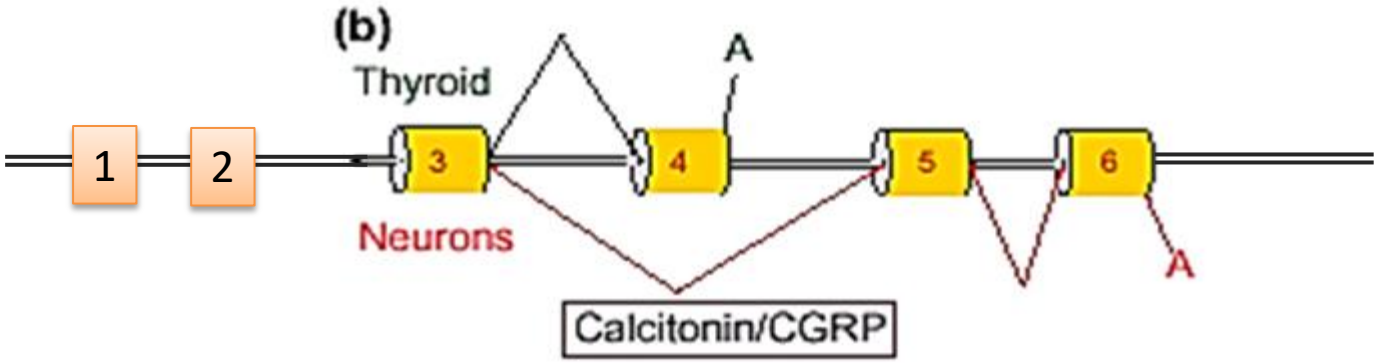
Does AS really increase the coding capacity of genomes ?

The majority of genes show many transcripts:
Some are protein-coding, other are noncoding RNAs

Sometimes alternative trascripts change UTRs (→ post-transcriptional regulation)

Sometimes the coding region is changed (alternative coding exons)

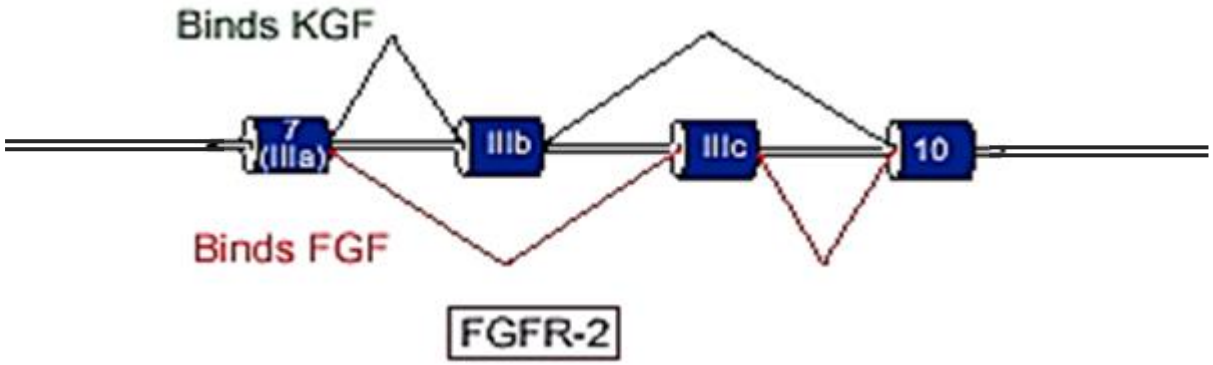
among the first discovered …



Calcitonin is a 32-aminoacid peptide hormone that is produced by the parafollicular tyroid cells in Humans. The first function of calcitonin is homeostatic: it lowers calcium concentration in blood.

CGRP is produced in both peripheral and central neurons. It is a potent peptide vasodilator and can function in the transmission of pain. In the spinal cord, the function and expression of CGRP may differ depending on the location of synthesis.

among the first discovered …



Mutual exclusion of either exon IIIb or IIIc in the fibroblast growth factor 2 RNA will change the specificity of binding of the encoded proteins showing high affinityfor Keratinocyte growth factor when expressing IIIb, or for FGF when expressing IIIc..

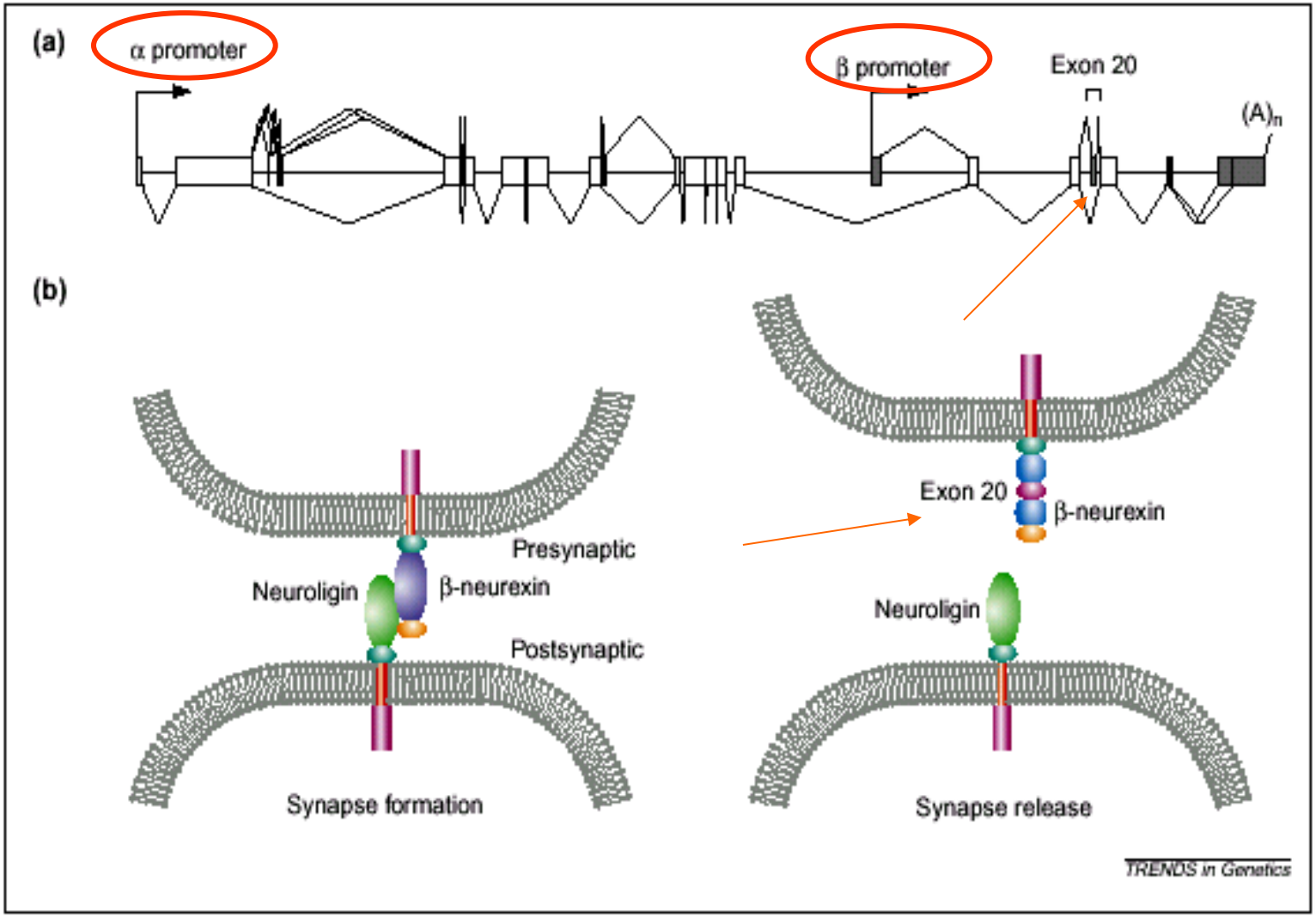


Fig. 2. Alternative splicing of the neurexin genes. (a) Organization of the human gene encoding neurexin I. The exon–intron structure of the human neurexin I gene is depicted (L. Rowen and B. Graveley, unpublished). The constitutive splicing events are indicated below the gene and alternative splicing events are depicted above the gene. The constitutive exons are white and the alternative exons are shaded. Exon 20 is indicated. Human neurexins II and III have a very similar exon–intron organization (L. Rowen and B. Graveley, unpublished). (b) Model for the function of the alternative splicing of exon 20 in b-neurexin I. b-neurexin I (present in the presynaptic cell) lacking sequences encoded by exon 20 can interact with neuroligin present in the postsynaptic cell, and thus function to initiate synaptogenesis. In contrast, b-neurexin I containing exon 20 encoded sequences can not interact with neuroligins. This form of b-neurexin I might indirectly function in releasing synapses.

Neurexins

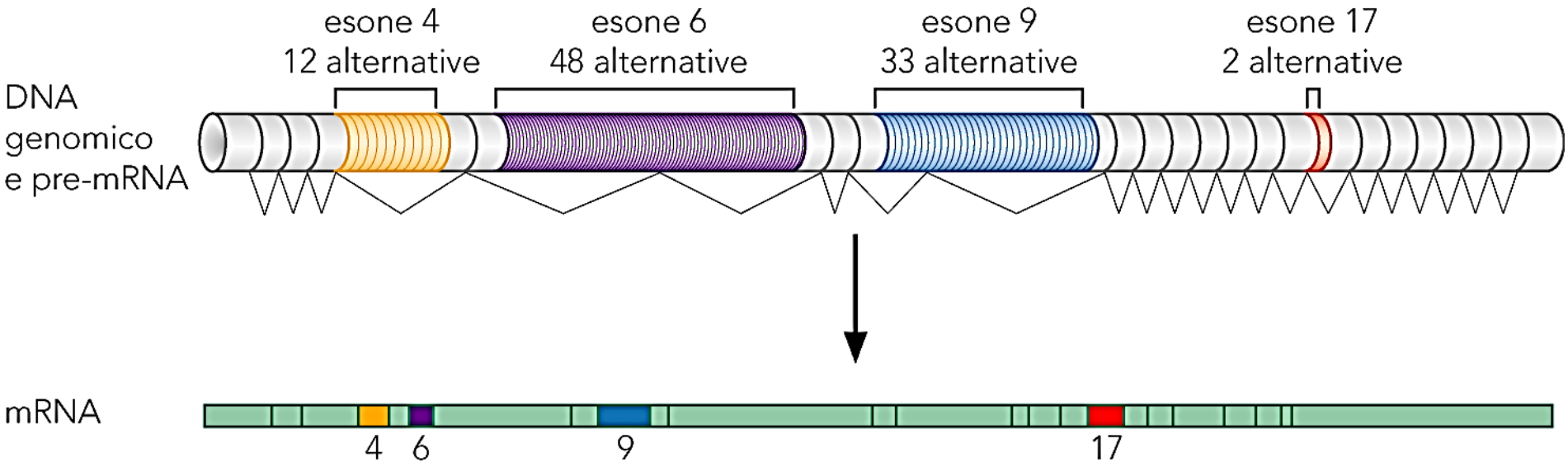
From: Graveley BR (2001) Trends Genet., 17:100-106.

Drosophila Dscam gene provides probably the **extreme** example of alternative splicing.

Perhaps the most complex event that takes place during development is the migration and connection of neurons. Even in a 'simple' organism such as *Drosophila melanogaster*, which contains only ~250 000 neurons, accurately wiring neurons together would appear to be a daunting task.

In flies, the gene encoding the Down syndrome cell adhesion molecule (*Dscam*) appears to fulfill at least part of this role. *Dscam* encodes an axon guidance receptor with an extracellular domain that contains ten immunoglobulin (Ig) repeats. The most striking feature of the *Dscam* gene is that it's pre-mRNA can be alternatively spliced into over 38,000 different mRNA isoforms (Fig. 3a). This is 2–3 times the number of predicted genes in the entire organism !

Each mRNA encodes a distinct receptor with the potential ability to interact with different molecular guidance cues, directing the growing axon to its proper location.



Potentially 38,000 splicing variants

«Pure» alternative splicing

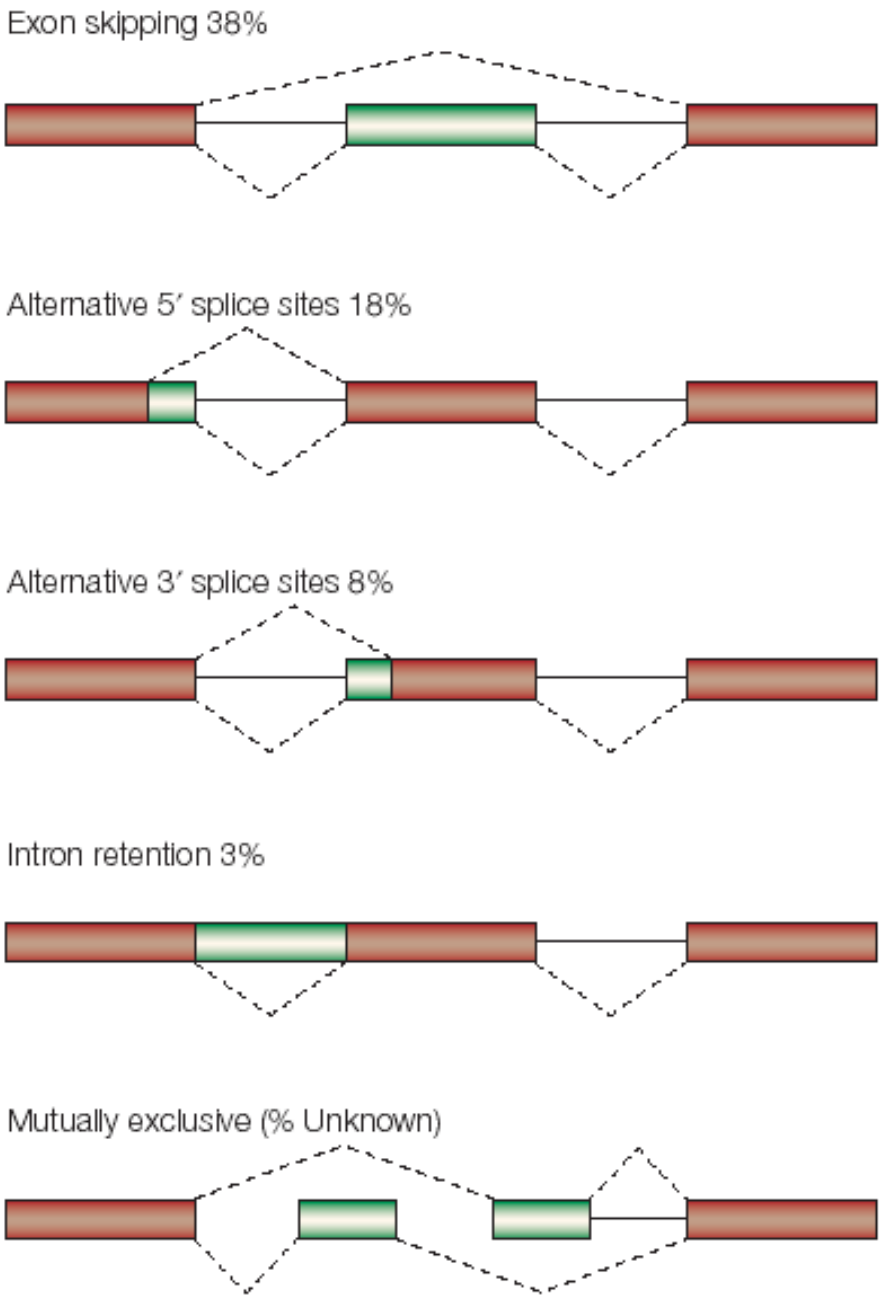


Figure 3

Types of alternative splicing.

In all five examples of alternative splicing, constitutive exons are shown in red and alternatively spliced regions in green, introns are represented by solid lines, and dashed lines indicate splicing activities. Relative abundance of alternative splicing events that are conserved between human and mouse transcriptomes are shown above each example (in % of total alternative splicing events).

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

Note that the indicated percentages derive from older studies and are slightly different from those demonstrated by recent, RNA-Seq based evaluations

Exons that are always present in processed transcripts are said «**constitutive**» or «strong» exons

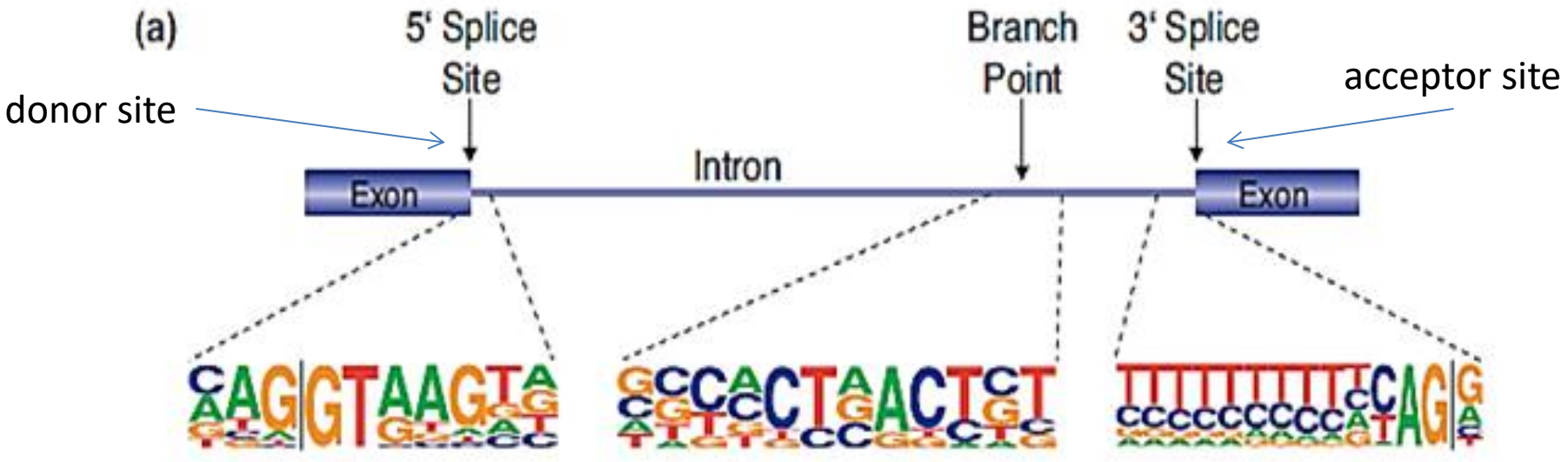
Exons that are present only in some processed transcripts are called «**alternative**» or «weak» exons

Alternative exon inclusion/exclusion : regulated process

- Exon-intron border sequence (splice-sites)
- Exon definition (intron definition in lower eukaryotes)
- SR and hnRNPs factors
- *cis*-elements and *trans*-regulatory factors
- Tissue-specific Splicing Regulators
- Nucleosome positioning over exons
- Histone PTMs and readers-splicing factors association
- Transcriptional speed and pausing

How is Alternative Splicing regulated ?

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



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

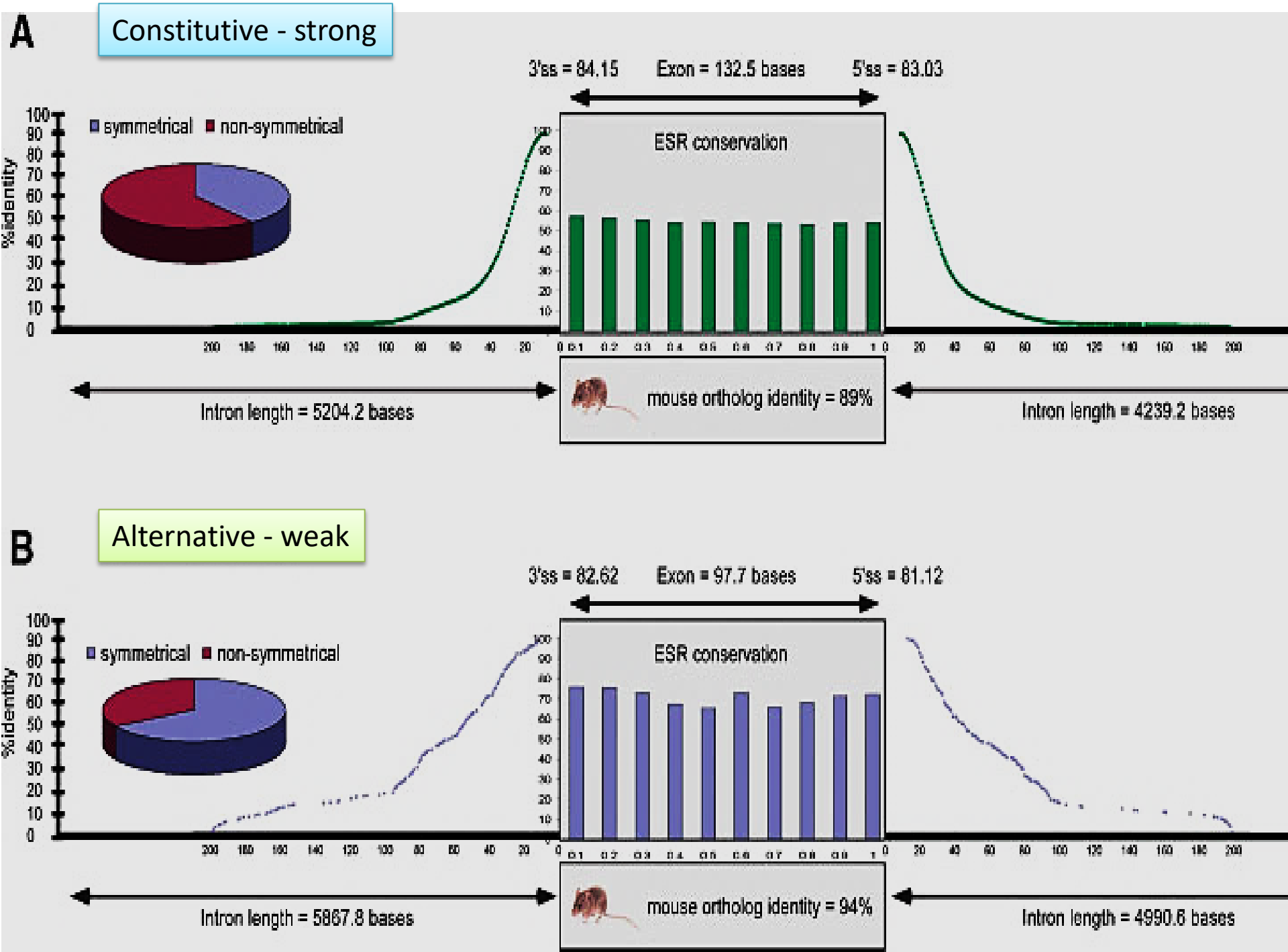
From: McManus & Graveley, COGD, 2011

Can **alternative exons** be distingushed from **constitutive exons** structurally?

Some Authors used comparative genomics to obtain insight. 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).

Sequences were then compared. Exons were normalized in lenght and flanking introns were explored within 200 pb.

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


A
Constitutive - strong
3'ss = 84.15
Exon = 132.5 bases
5'ss = 83.03
symmetrical
non-symmetrical
%identity
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
symmetrical
non-symmetrical
%identity
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

Exons that are alternatively spliced in both human and mouse are **more conserved** than constitutive exons

Conservation is higher toward exon edges and extends farther in introns: regulatory sequences ?

Weak splice sites

Cassette exons[1] have **weak splice sites**, compared to the strong ones in constitutive exons

Shorter

Alternative cassette exons are also **shorter** and are flanked by longer introns than constitutively spliced ones.

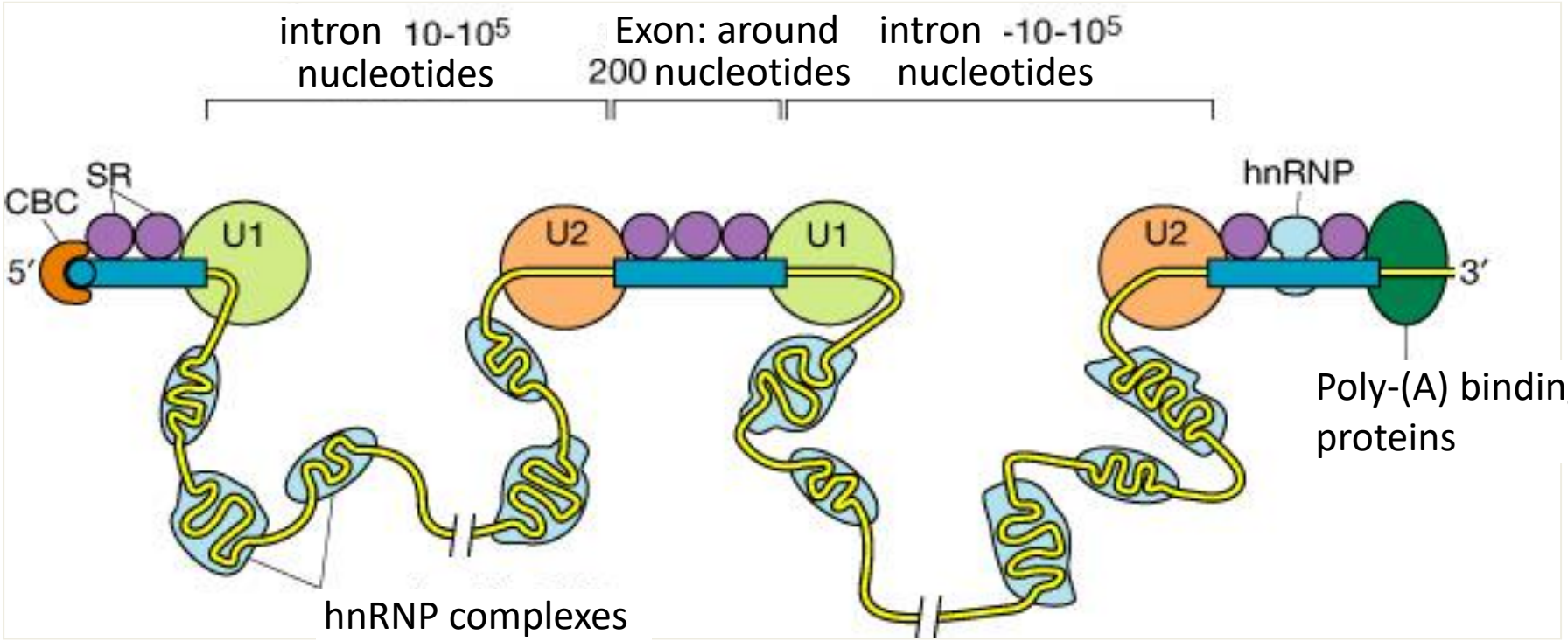
Symmetric

The percentage of **symmetrical** exon is definitely higher in alternative exons (symmetrical means “divisible-by-three” number of base pairs)

[1]*Cassette exons: refers to exon skipping*

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

Second point: how Exons (and Introns) are defined



SR proteins = splicing regulators

Domains:
The most typical domain is an alternating Arginine-Serine 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

RRM=RNA Recognition Motif

SR-like
(proteins 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**

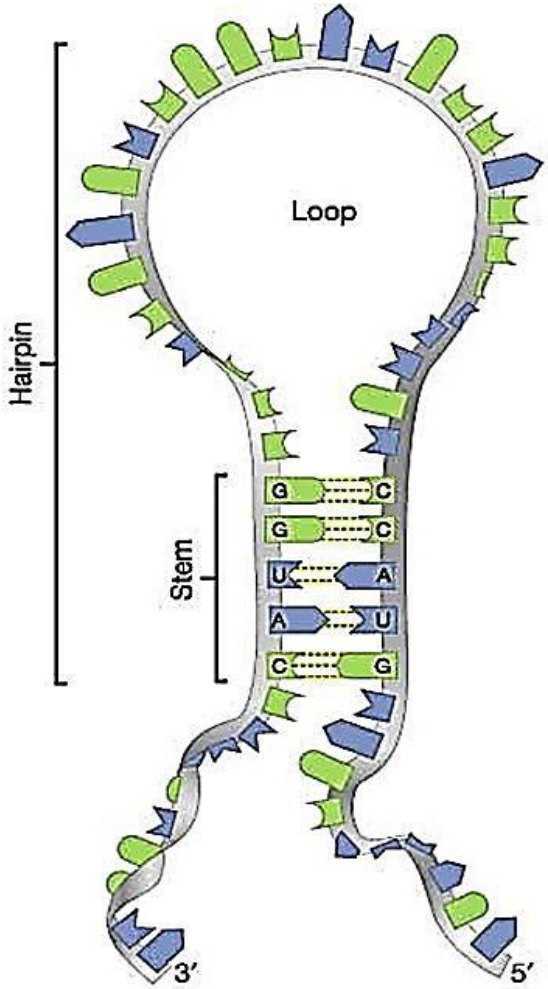
Name	Other names	Domains*	Binding sequences	Target genes
hnRNP A1	NA	RRM, RGG and G	UAGGGA/U	*SMN2* and *RAS*
hnRNP A2 hnRNP B1	NA	RRM, RGG and G	(UUAGGG)n	HIV *tat* and *IKBKAP*
hnRNP C1 hnRNP C2	AUF1	RRM	U rich	*APP*
hnRNP F	NA	RRM, RGG and GY	GGGA and G rich	*PLP*, *SRC* and *BCL2L2*
hnRNP G	NA	RRM 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*

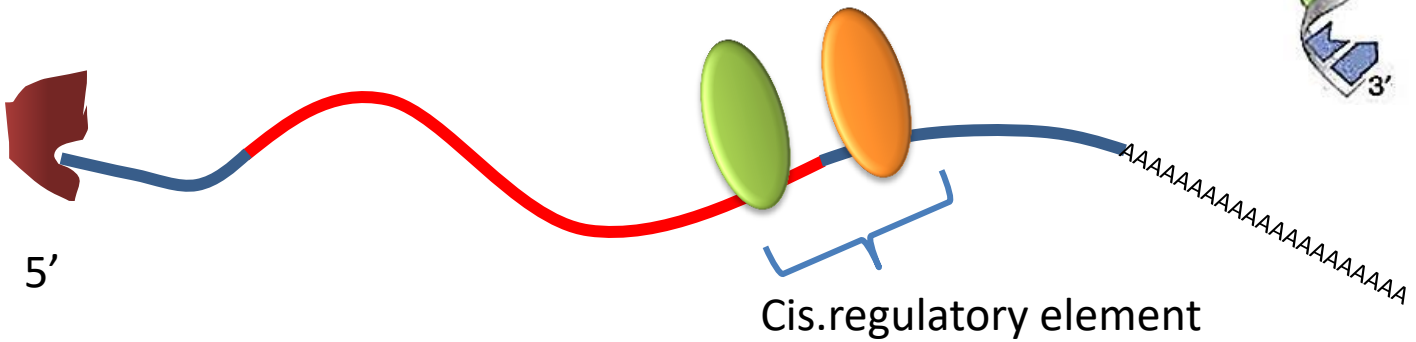
Identification of RNA binding proteins motifs

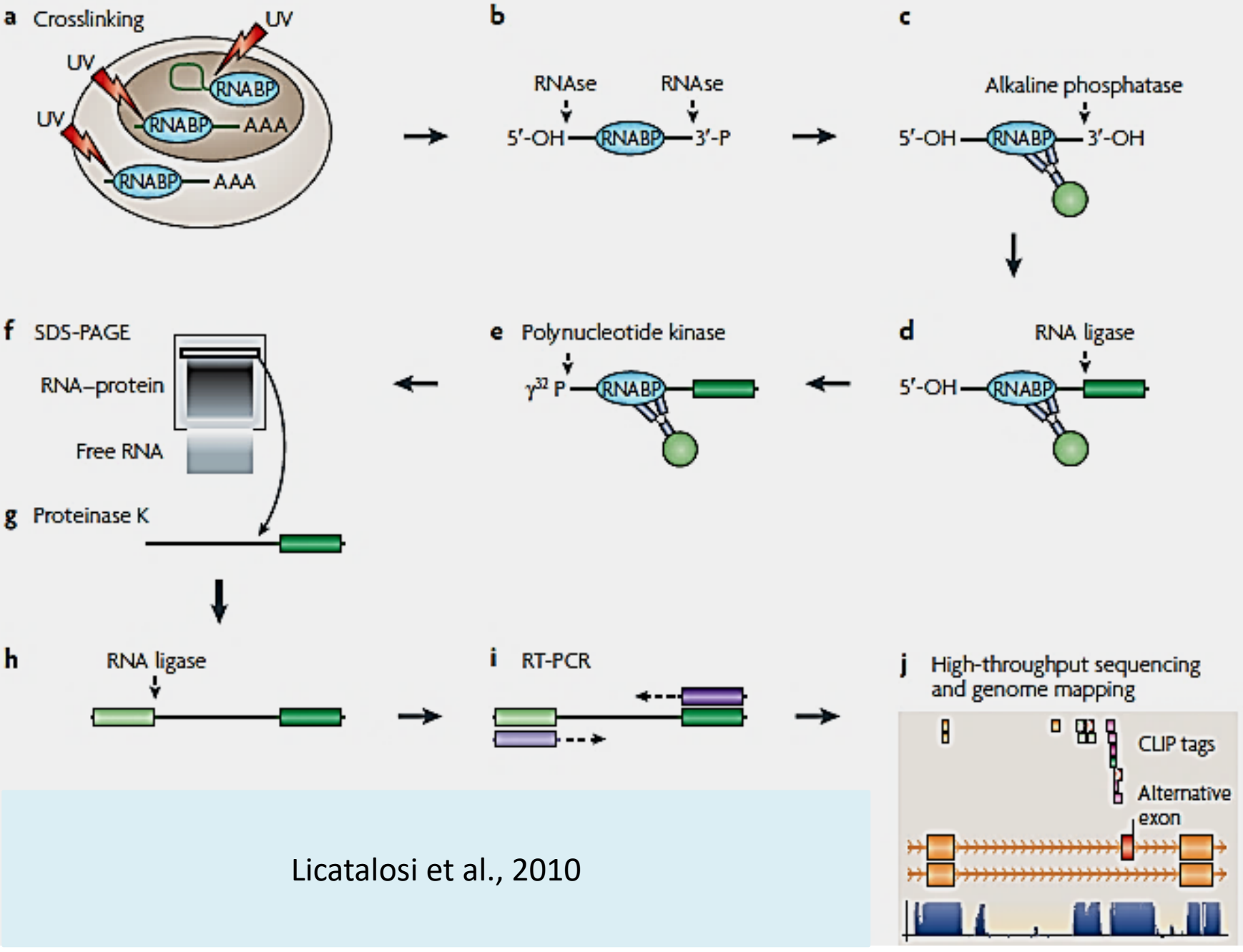
Specificity of RNA binding: both «sequence» and «structure» elements

Problems in predicting regulatory motifs:

- Localization (intron length)
- Sometimes dispersed elements
- Sometimes the structural component prevails upon pure sequence



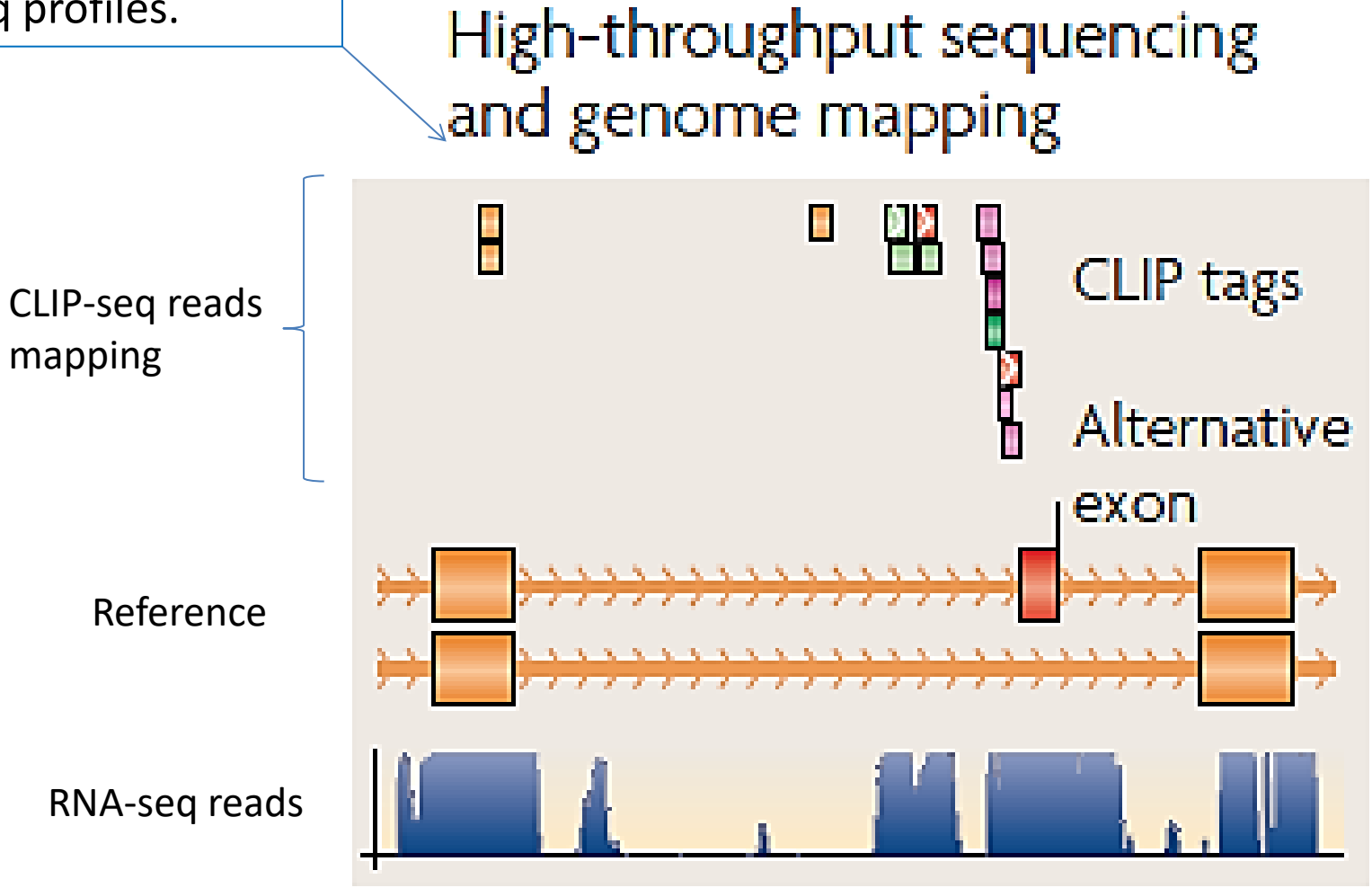




Licatalosi et al., 2010

Example:

Individual Splicing Factors HITS-CLIP profiles mapped to genome and compared to RNA-Seq profiles.



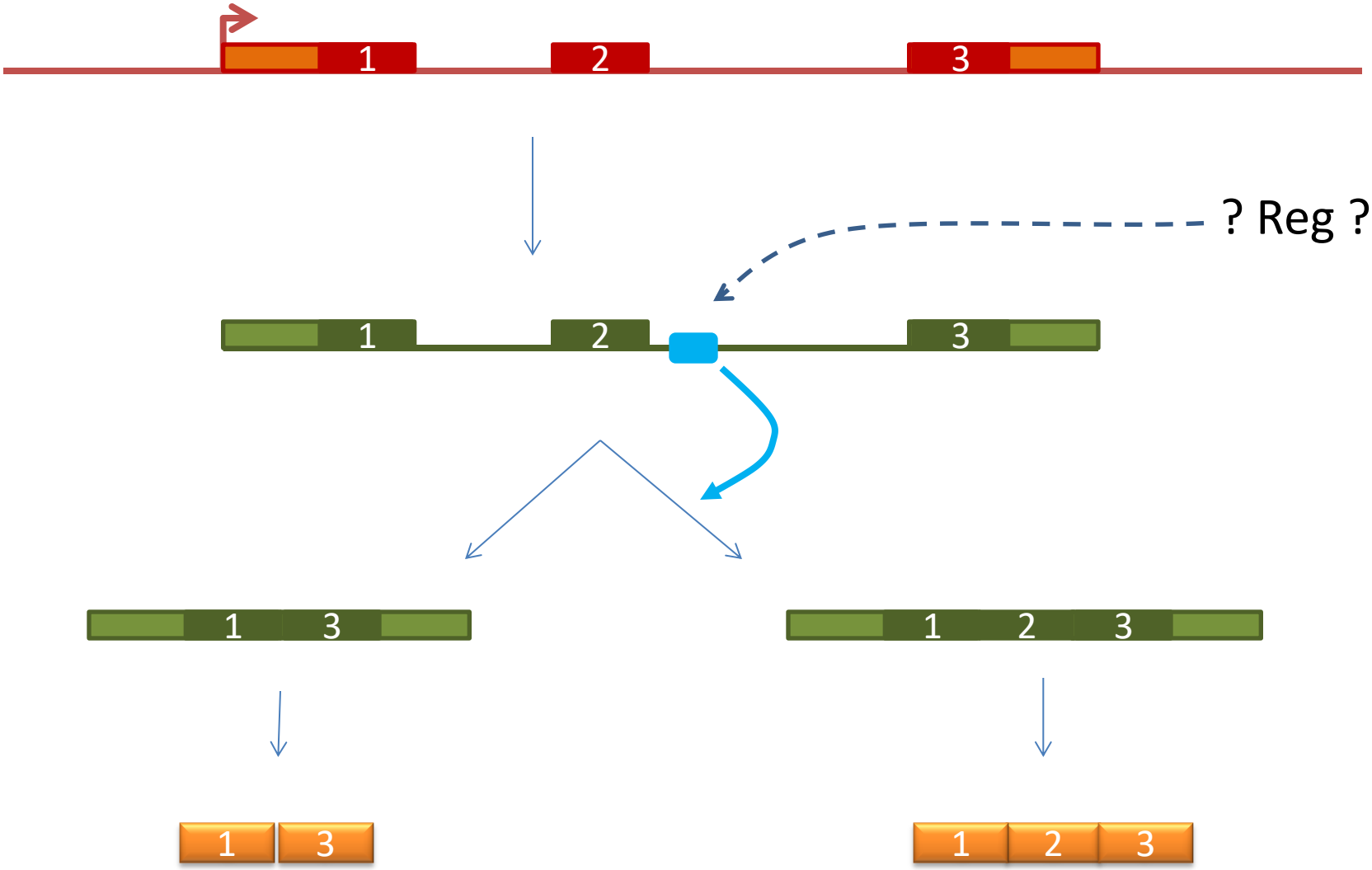
2.3 Sequence elements that promote or repress splicing at adjacent «**ss**»

cis elements

Approach 1 : random select sequences that drive either exon inclusion or exon skipping in experimental model systems

Approach 2: RNA immunoprecipitation using Abs against Splicing Factors (SR, hnRNP, other) and then NGS (RIP, CLIP, HITS-CLIP)

Note: Studies based on pure search of conserved motifs was not successfull: experimental + bioinformatics approaches needed


1
2
3
? Reg ?
1
2
3
1
3
1
2
3
1
3
1
2
3

AS regulatory sequences characterized by de-novo functional assays
Example:

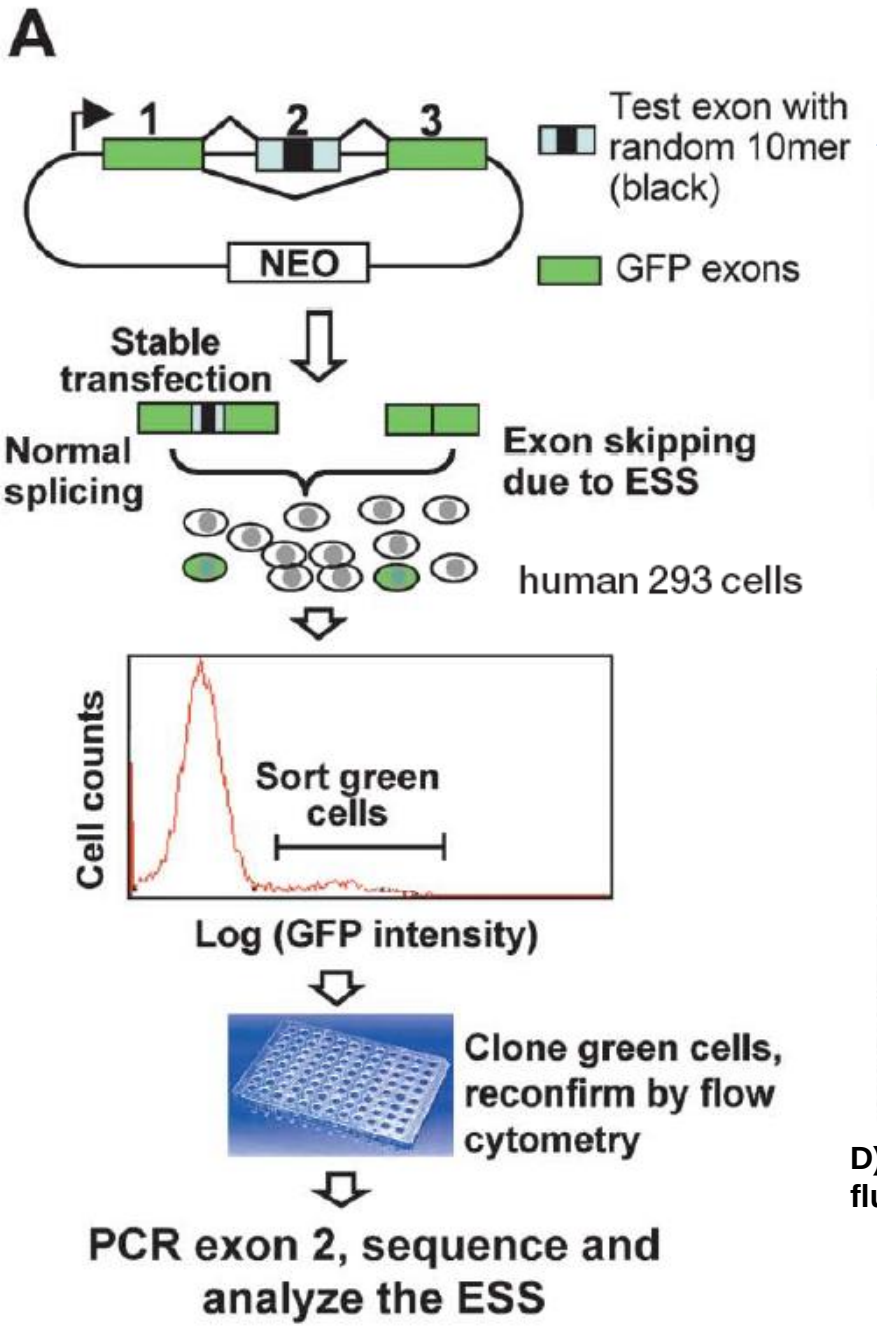
Cell, Vol. 119, 831–845, December 17, 2004, Copyright ©2004 by Cell Press

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

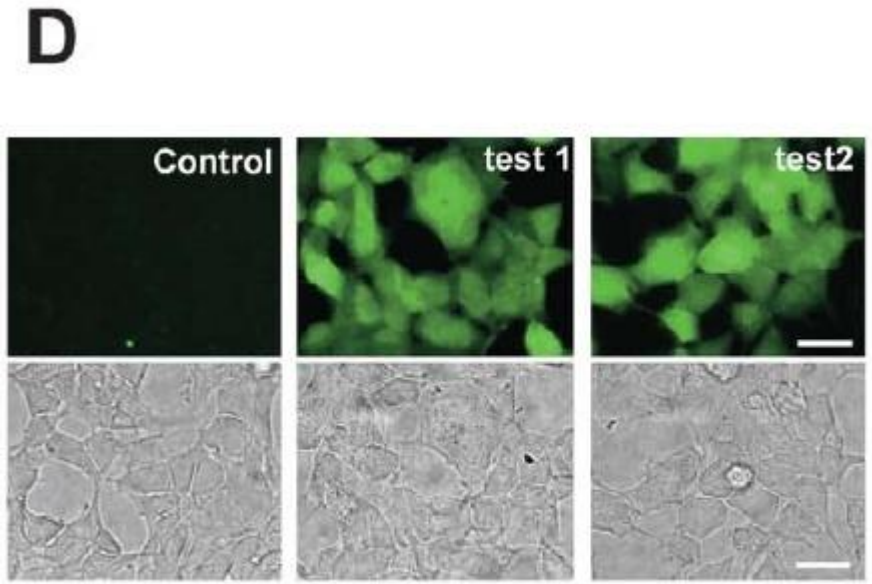
Exonic splicing silencers (ESSs) are *cis*-regulatory elements that inhibit the use of adjacent splice sites, often contributing to alternative splicing (AS). To systematically identify ESSs, an in vivo splicing reporter system was developed to screen a library of random decanucleotides. The screen yielded 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.



random

.....GATCCGNNNNNNNNNNTCTTCG.....

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



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

Wang et al., 2004

Following selection, all positive cell clones are extracted

→RNA

RT-PCR using primers pairs flanking the interrogated exon → cDNA library

The library is sequenced (clone-and-sequence or NGS) → a collection of sequences

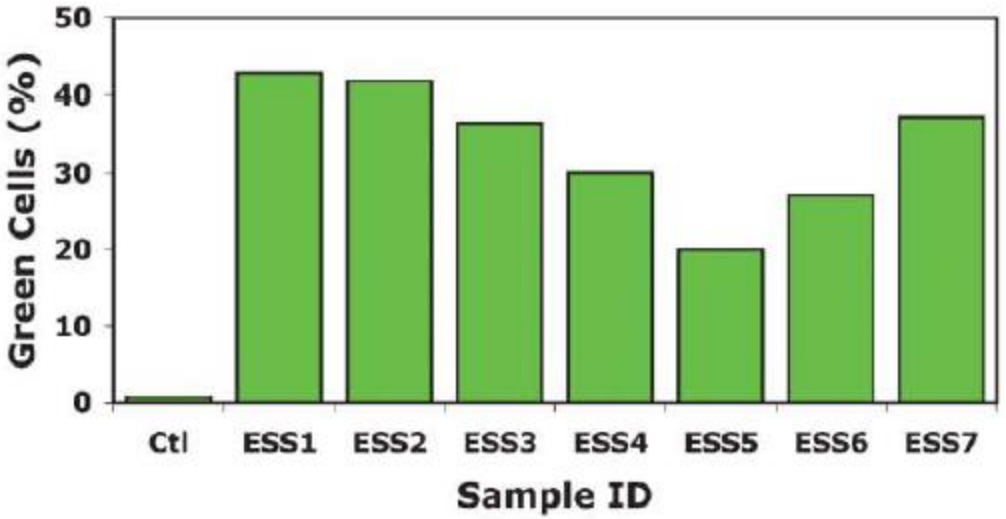
Sequences are bioinformatically examined to find one or more «consensus» motifs

Consensus motifs inserted in test exon and validated.

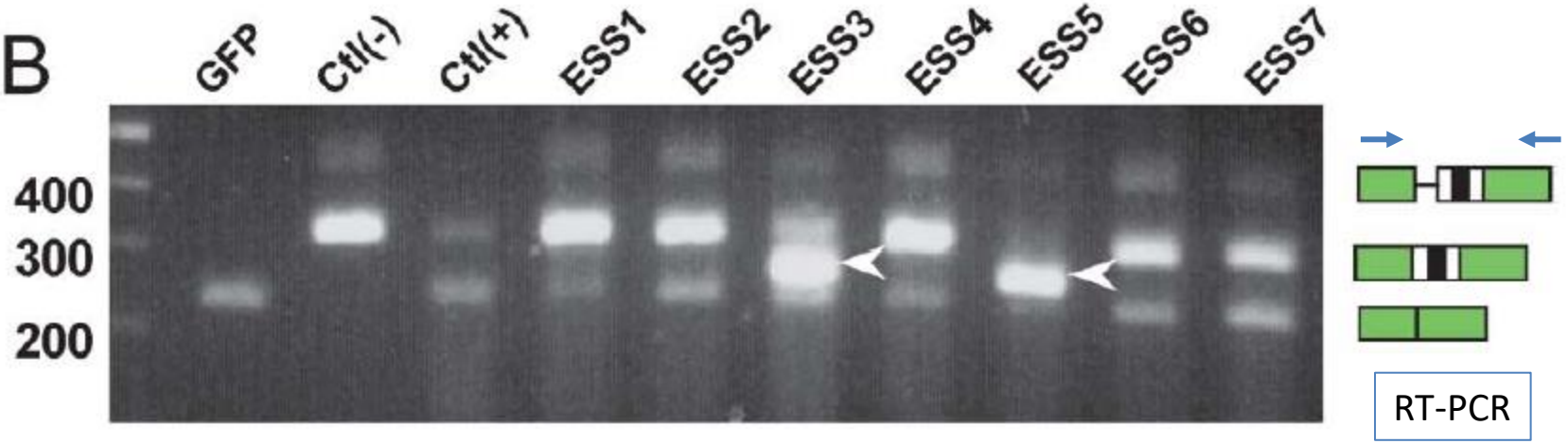
A

consensus

ESS1	TTTGTTCCGT
ESS2	GGGTGGTTTA
ESS3	GTAGGTAGGT
ESS4	TTCGTTCTGC
ESS5	GGTAAGTAGG
ESS6	GGTTAGTTTA
ESS7	TTCGTAGGTA



B



A number of similar studies identified putative **motifs for positive and negative regulation** of splicing, classified on effect and position:

ESE = exonic splicing enhancer
ESS = exonic splicing silencer
ISE = intronic splicing enhancer
ISS = intronic splicing silencer

ESE (exonic splicing enhancers) were discovered and described first

Note that the «code» is superimposed to coding sequences

ESE mutations were identified in various human diseases

When identified motifs are mapped to reference genes collection, it is observed that regulatory motifs spread a little distance from exons into introns, as predicted by mouse-human conservation analysis seen before.

Maps of potential regulatory sequences are obtained.

Regulatory sequences are found primarily close to the 5’-ss and 3’-ss i.e. around exons

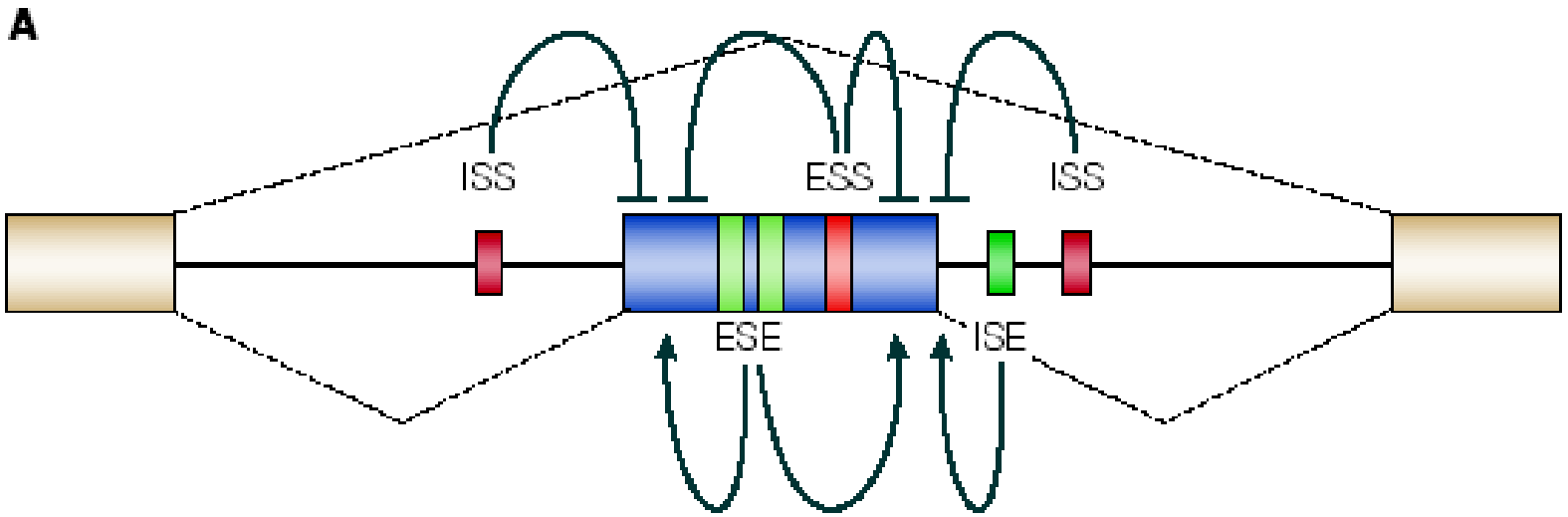


Figure 1 | **Elementary alternative splicing events and regulatory elements.**
A | In addition to the splice-site consensus sequences, a number of auxiliary elements can influence alternative splicing. These are categorized by their location and activity as exon splicing enhancers and silencers (ESEs 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.

Regulatory factors

Tissue-specific splicing factors