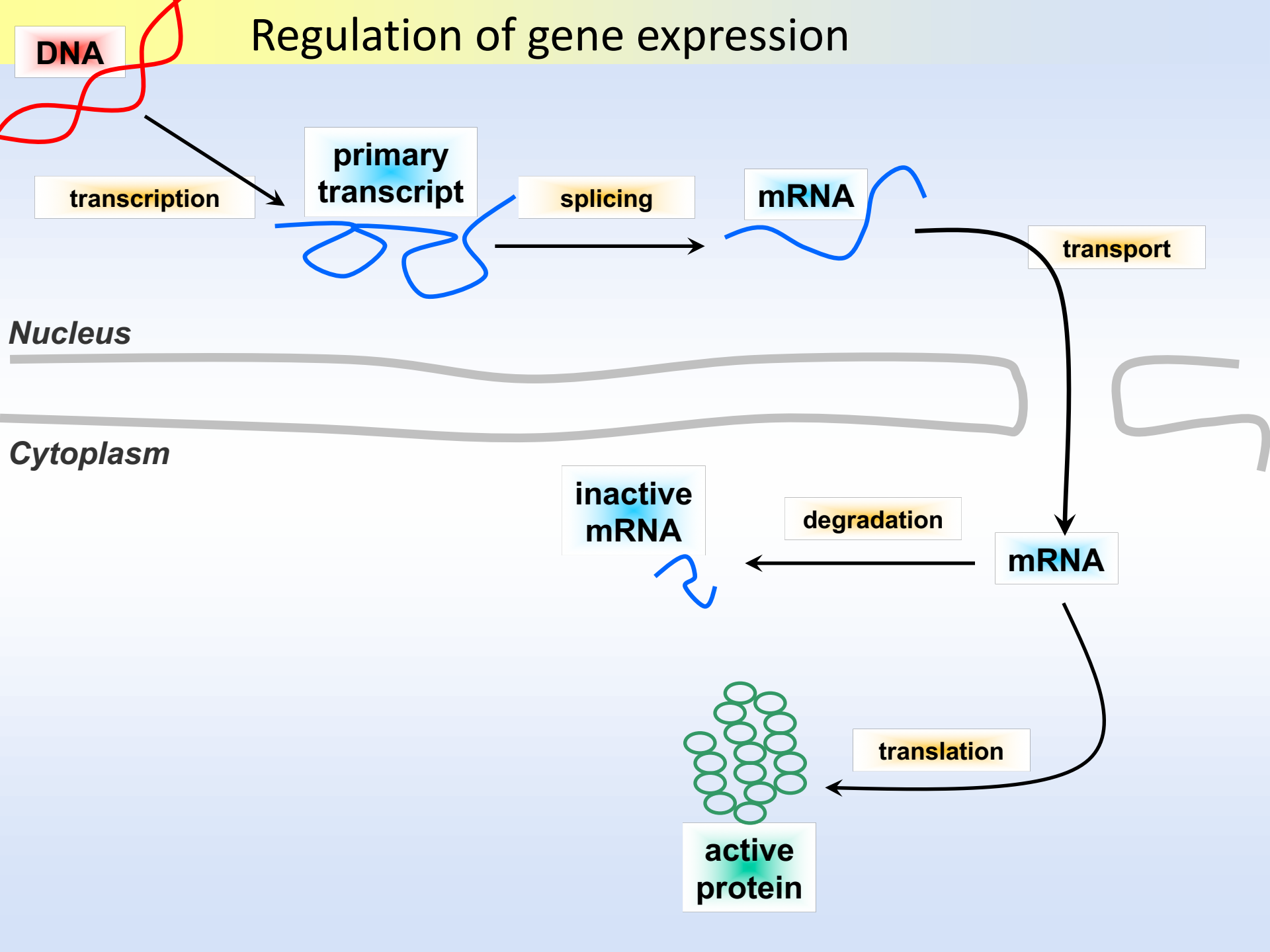
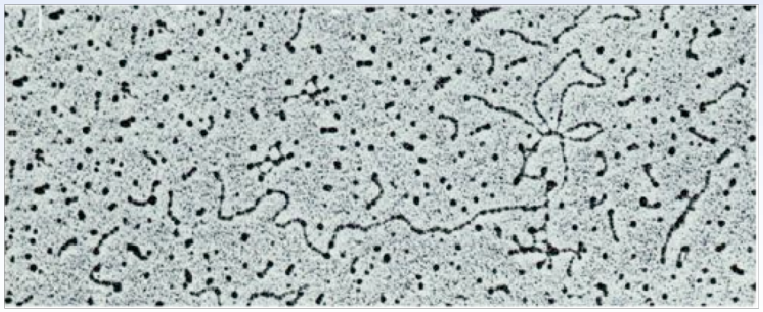
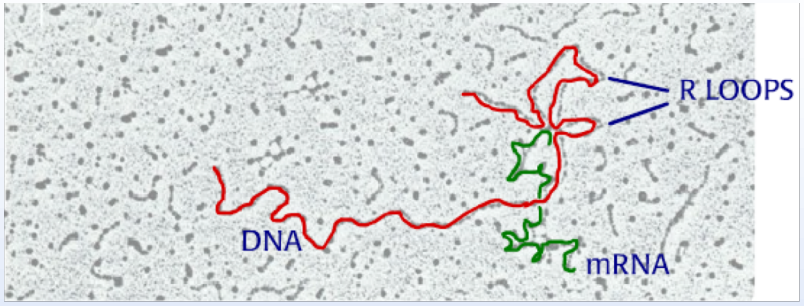
Regulation of gene expression



An Amazing Sequence Arrangement…….

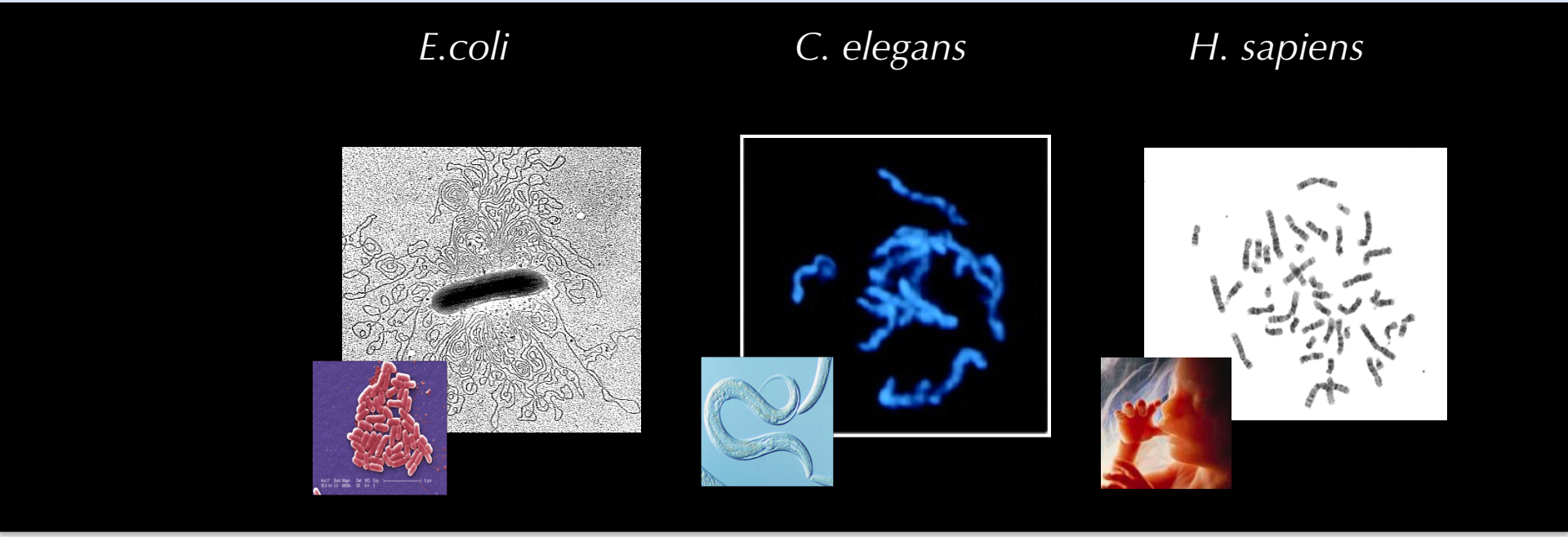
In 1976, the first isolated fragments of mammalian mRNA just didn't match the genes. Then in 1977….





'This startling fact [splicing] *has abruptly changed our conception of what a eukaryotic gene is and how its expression is controlled'* **Broker & Chow**

Genome size and organism complexity



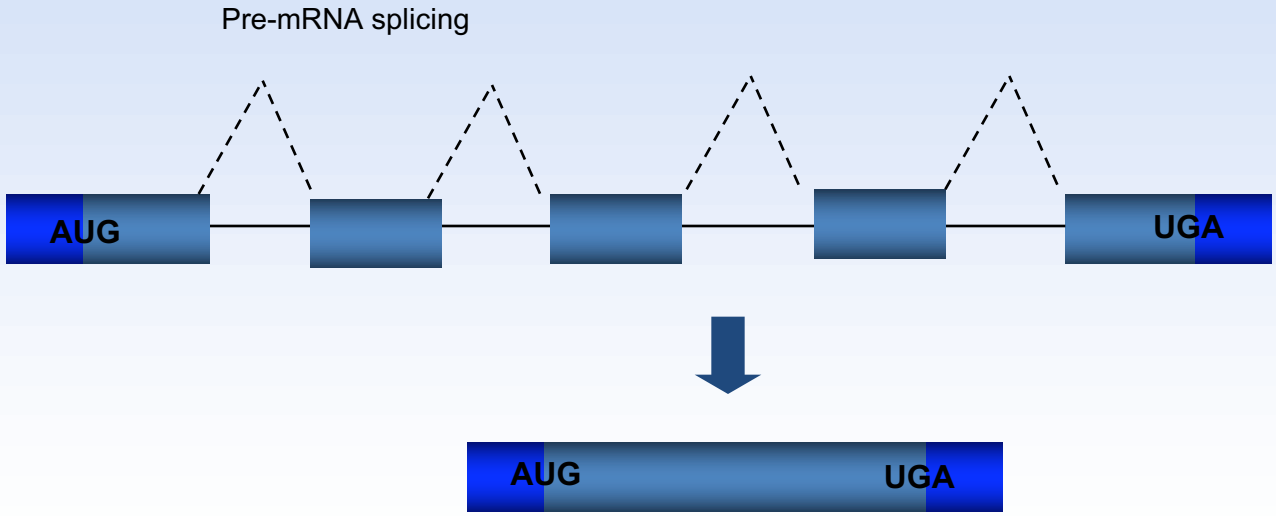
Genome	$5x10^6$ bp	$1x10^8$ bp	$3x10^9$ bp
Chromosomes	1	6	23
Coding genes	6692	20541	21995
ncDNA	5%	60%	98%
non-coding RNA genes	15	23136	ca. 40000
miRNAs	0	224	4274
pseudogenes	21	1522	10616

- Exons and introns

The most common pattern studied is for a gene with 2 exons and one intron

But the mean gene: is 28,000 base pairs and 7 exons amounting to 2,400 bp

The gene encoding dystrophin is 2,400,000 bp long, contains 79 exons, and takes 16 hr to transcribe.

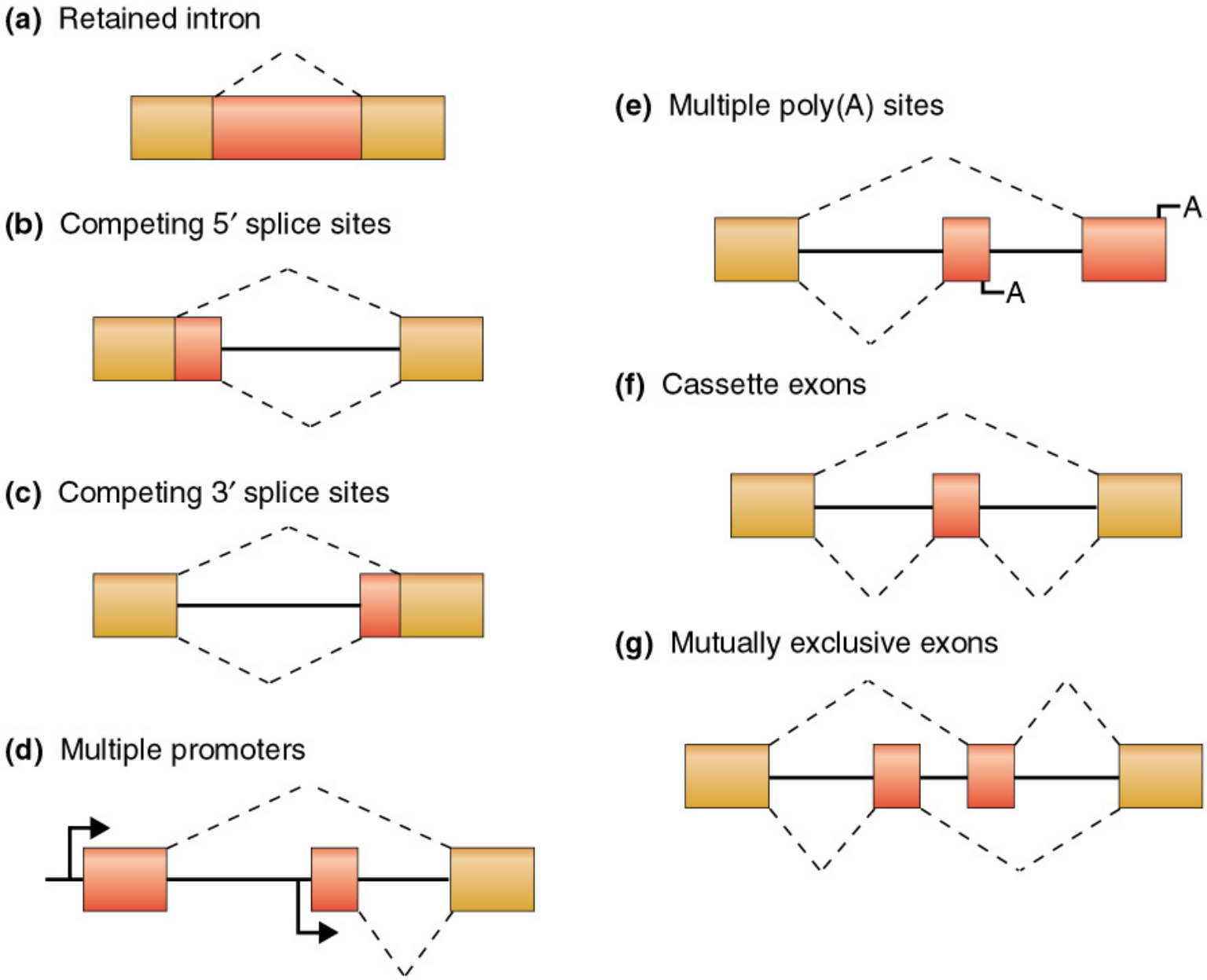


- Essential step in gene expression
- >15-50% of human genetic diseases involve splicing errors

Alternative splicing

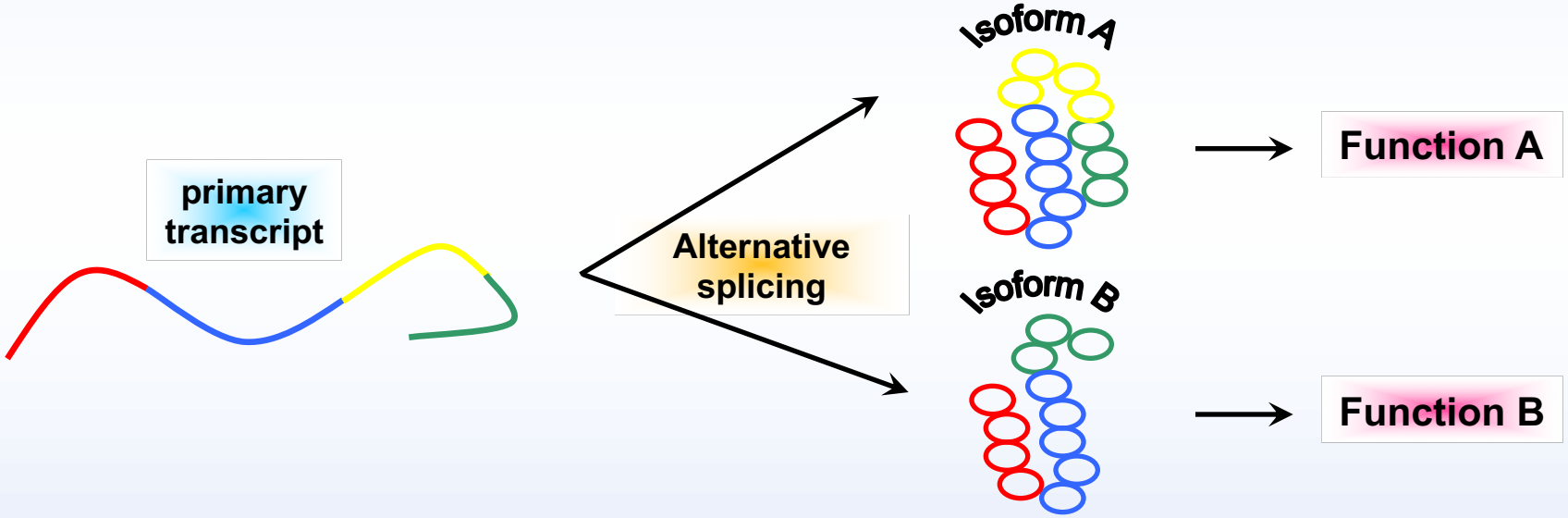
- Important *regulatory* step in gene expression

Types of Alternative Splicing Event



- Regulation of gene expression -

increase the complexity by increasing the combinatorial use of exons by alternative splicing



- RNA splicing gets the most out of genes

In animals, complexity depends less upon the number of genes than upon the number of different ways they can splice the RNA

So even if you found an exon, you couldn't tell how or when it might be used. Splicing exhibits extraordinary fidelity but selective promiscuity

Alternative splicing: functional consequences

Protein - Substitution/deletion of domains, change of reading frame, termination of reading frame

- Altered localization
- Antagonistic isoforms
- Modulation of function
- Unrelated proteins
- On/off switch for gene expression

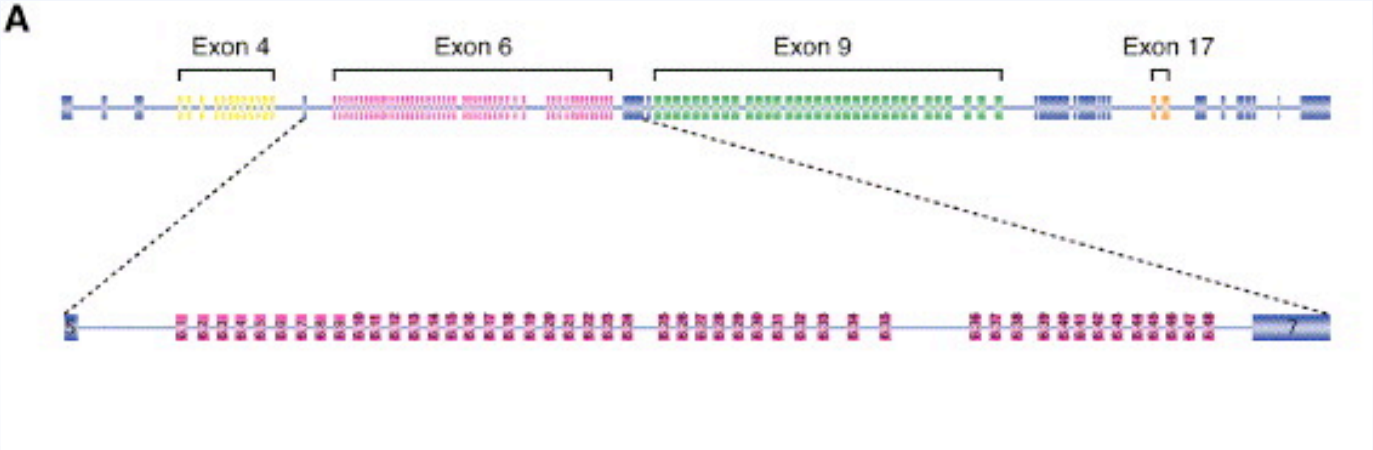
Developmental -

- Sex determination in *Drosophila*
- Apoptosis

Pathology - misregulation of alternative splicing and disease.

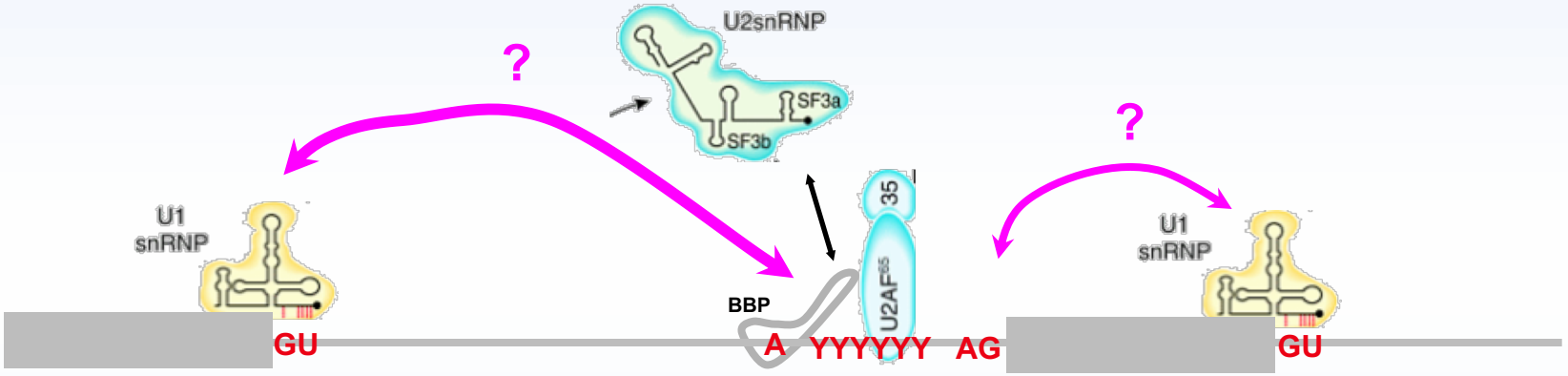
- Cystic fibrosis
- Myotonic Dystrophy, Spinal Muscular Atrophy
- FTDP1 Fronto-temporal dementia
- WT-1 Wilms tumors

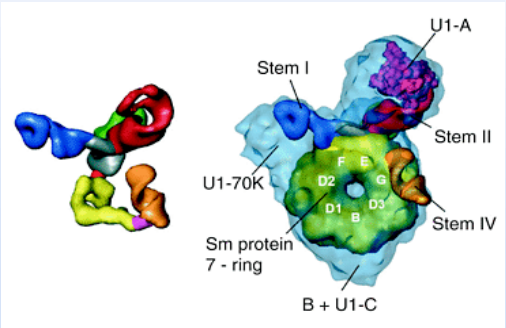
Drosophila Dscam encodes 38,016 distinct axon guidance receptors through the mutually exclusive alternative splicing of 95 variable exons.



Organization of the *D. melanogaster Dscam* gene. *Dscam* contains 115 exons, 95 of which are alternatively spliced. The exon 4, 6, and 9 clusters contain 12, 48, and 33 alternative exons, respectively, that each encode variable immunoglobulin domains. The exon 17 cluster contains two exons that encode alternate versions of the transmembrane domain. The exons within each cluster are alternatively spliced in a mutually exclusive manner. The exon 6 cluster is enlarged to depict its organization.

- Recognition of canonical splice sites

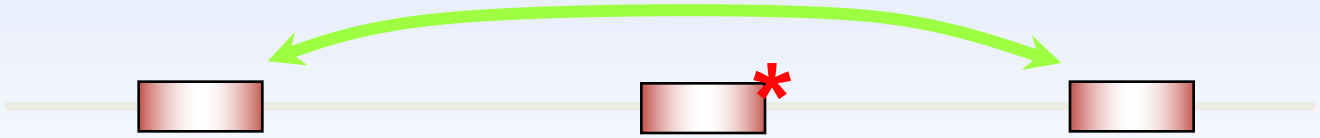


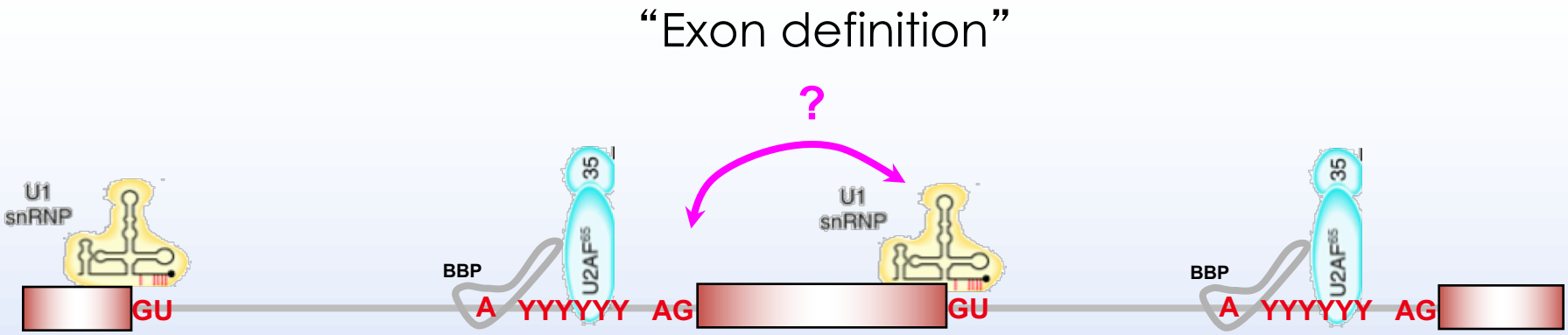


10-15% of single nucleotide mutations causing disease affect splice sites

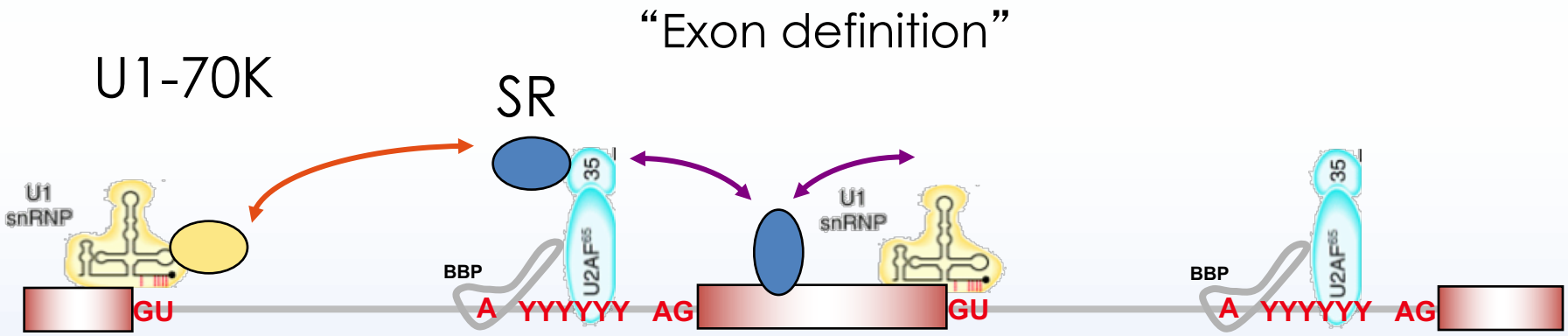
From: H. Stark, P. Dube, R. Luhrmann & B. Kastner, *Nature* 409, 539 - 542 (2001)

- Mutations at constitutive splice sites: exon skipping

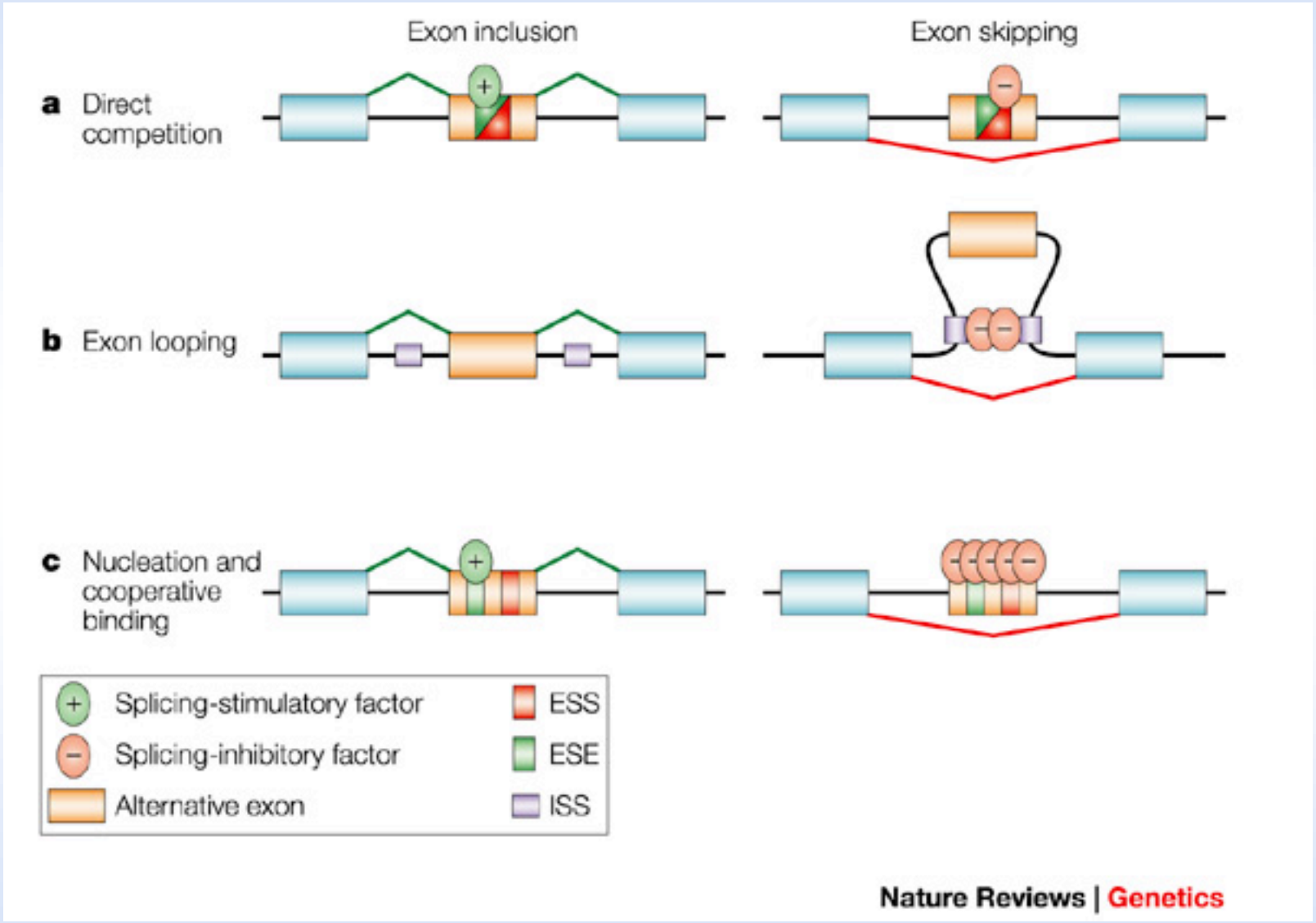




- Mechanistic models connecting ESEs and SRs



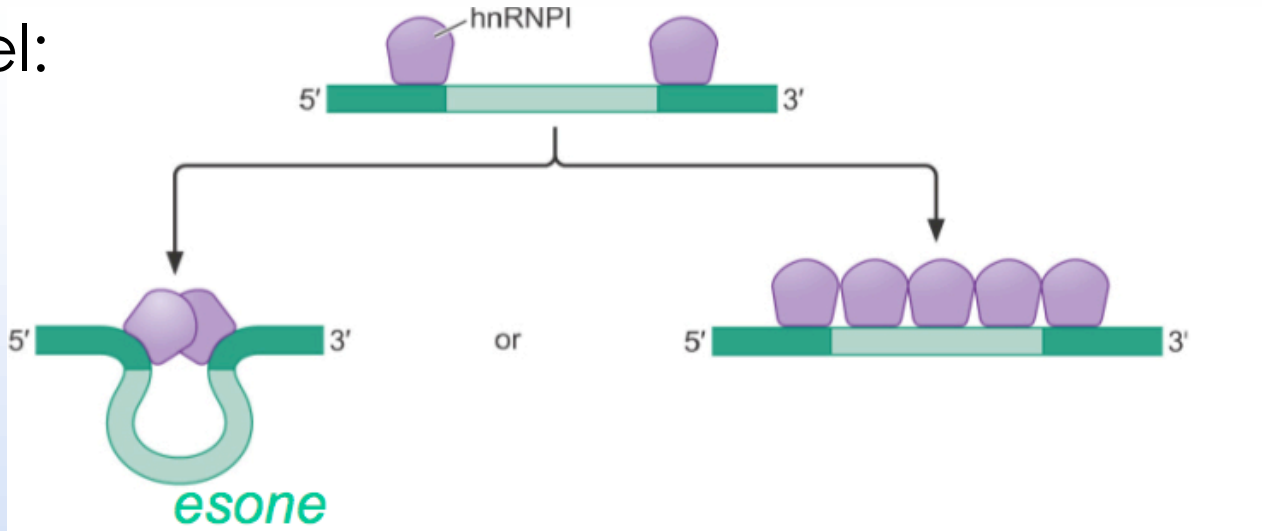
Both enhancer (ESE) and silencer (ESS) of splicing exist. They activate or repress splicing on the neibouring splce sites


Exon inclusion
Exon skipping
a Direct competition
b Exon looping
c Nucleation and cooperative binding
Splicing-stimulatory factor
Splicing-inhibitory factor
Alternative exon
ESS
ESE
ISS
Nature Reviews | Genetics

Both enhancer (ESE) and silencer (ESS) of splicing exist. They activate or repress splicing on the neibouring splce sites

An example of silencers are the HnRNPs (heterogeneous ribonuclear proteins) which bind RNA but do not recruit the splicing machinery due to the lack of the SR domain

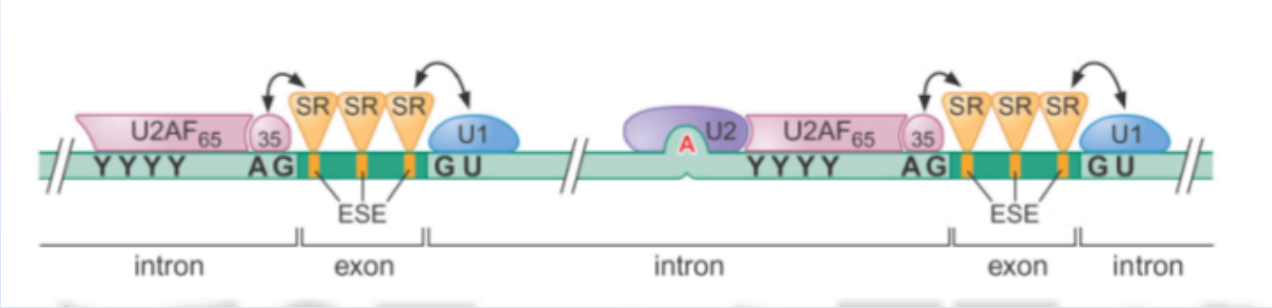
Model:



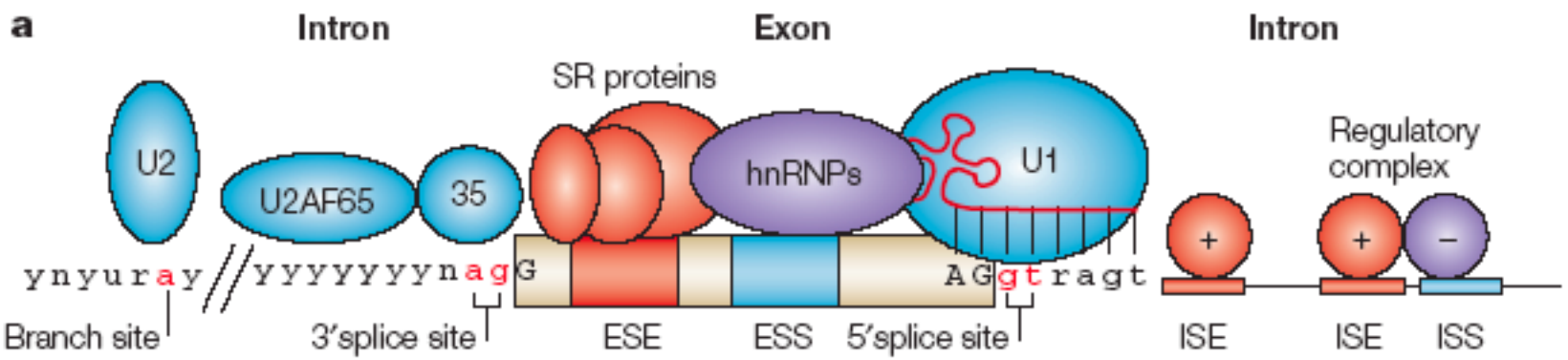
How splice sites are recognized in alternative splicing?

SR proteins bind inside exons on enhancer sites (ESE) and recruit the splicing apparatus on the 5' and 3' splice sites

SR proteins have a RNA binding domain and a arginine (R)-serine (S)-rich domain responsible for recruitment of the splicing apparatus



Regulatory elements in pre-mRNA splicing



Cis acting regulatory elements

ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
ISS	Intronic Splicing Enhancer
ISE	Intronic Splicing Silencer
CERES	Composite Exonic Regulatory Elements of Splicing

Trans-acting factors

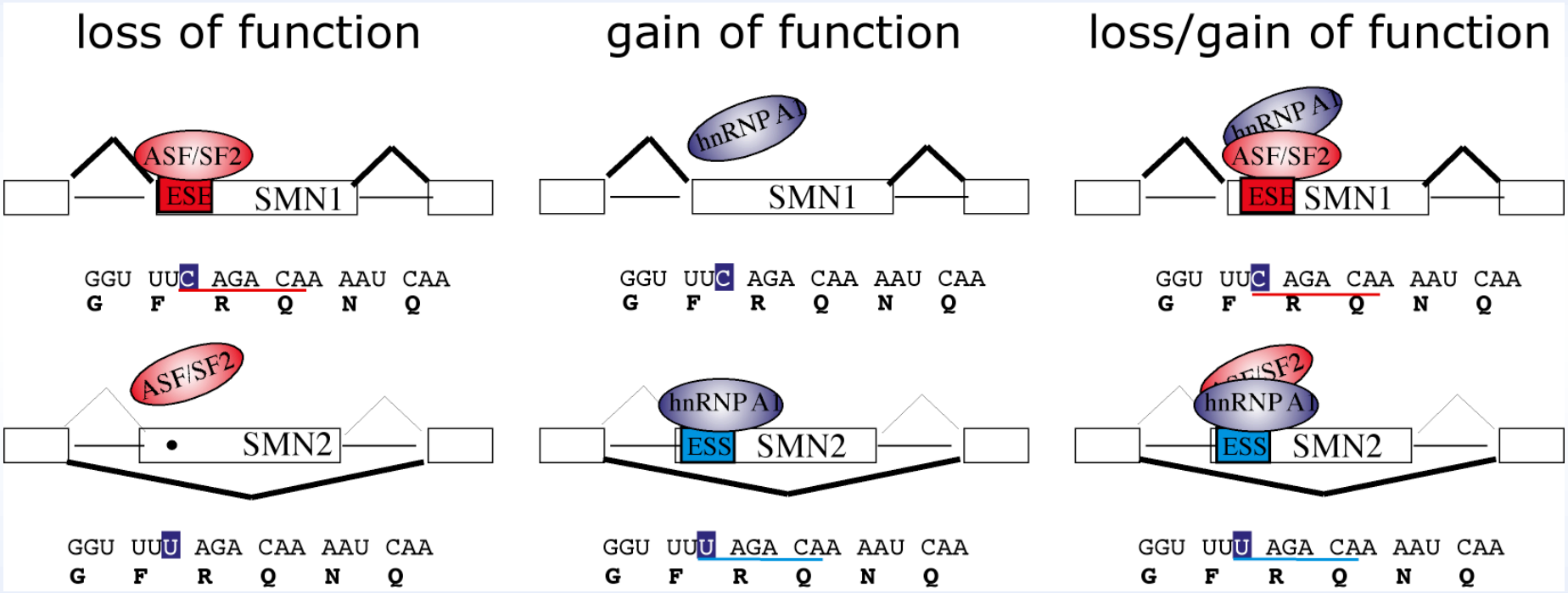
SR proteins	Serine arginine rich proteins (SF2/ASF)
hnRNPs	heterogeneous nuclear RiboNucleoprotein Particles (hnRNPA1)
snRNPs	small nuclear RiboNucleoProteins (U1 snRNP, U2 snRNP)

Mutations in sequences involved in the splicing reaction may cause disease onset

Es. one silent mutation in the ESE of the BRCA1 gene, which predispose to breast cancer, induces exon skipping.

Es. some mutations in the B-globin gene produce slpicing alterations in thalassemic patients

Exon skipping: models



Cartegni L, KrainerAR.
Nat Genet. 2002

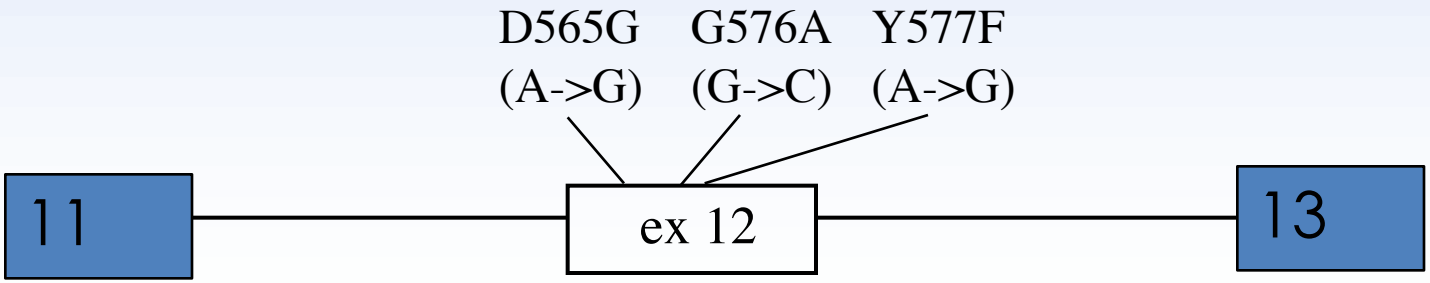
Kashima T, Manley JL.
Nat Genet. 2003

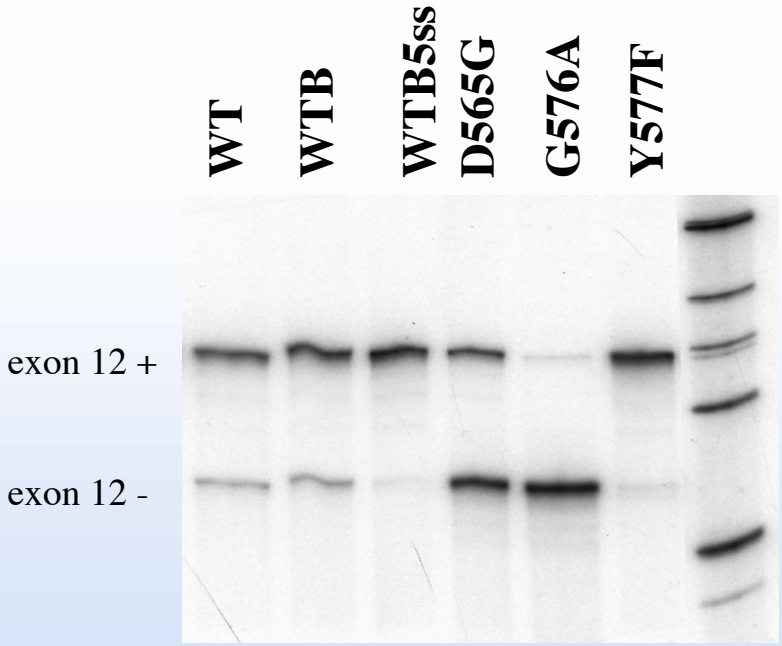
- A considerable number of disease causing mutations in exons or introns may disrupt previously unrecognised splicing regulatory elements
- Variability in the basal splicing machinery among different cell types cause cell-specific sensitivities to individual splicing mutations
- Exon sequence variation at CERES elements may represent a frequent disease-causing mechanism
- Even the most benign looking polymorphism in an exon (or in an intron) cannot be ignored as it may affect the splicing process

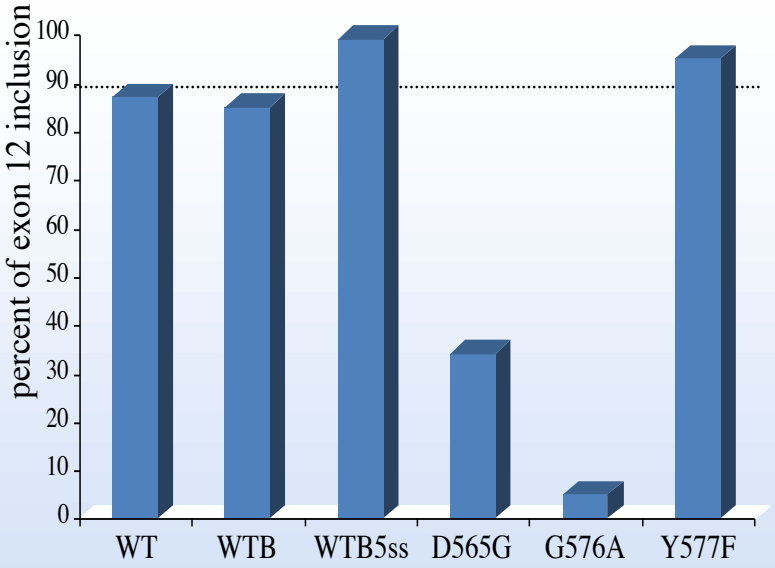
• Single point mutations in coding exons

	Probability of effect on encoded protein	Probability of nonsense-mediated decay	Probability of effects on splicing
Nonsense mutations	+	+	+/-
Missense mutations	+/-	-	+/-
Silent mutations	-	-	++

Missense mutations in CFTR exon 12 induce exon skipping in hybrid minigene
Cystic Fibrosis Transmembrane Conductance Regulator





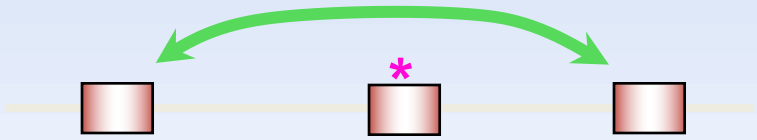


Pagani et al Hum Mol Genet *2003*

• Exon mutations associated with disease

Gene	Mutation	Exon
Missense mutations		
ADA	A215T	7
ATM	E2032K	44
ATP7A	G1302R	4
BRCA1	E1694K	18
CFTR	G58E	9
	D565G	12
F8	R1997W	19
FAH	Q279R*	9
FBN2	D1114H‡	25
FECH	A155P‡	4
HEXB	P404L	11
HMBS	E29L‡	3
HPRT1	G40V	2
	R48H	3
	A161E	6
	G180E	8
	G180V	8
	E182K	8
	P184L	8
	D194Y	8
	E197K	8
	E197V	8
	D201V	8
IL2RG	R285Q‡	6
IVD	R21C	2
	R21P	2
	D20N	2
MAPT	N279K§	10
	S305N*§	10
MLH1	R659P	17
	R659L	17

Gene	Mutation	Exon
PDHA1	A175T	6
PMM2	E139K	5
RHAG	G380V	9
Silent mutations		
APC	R623R	14
AR	S888S	8
ATM	S706S‡	16
	S1135S‡	26
CYP27A1	G112G	2
FAH	N232N	8
FBN1	I2118I	51
HEXA	L187L‡	5
HMBS	R28R	3
HPRT1	F199F	8
ITGB3	T420T	9
LIPA	Q277Q‡	8
MAPT	L284L§	10
	N296N§	10
	S305S‡§	10
MLH1	S577S‡	16
NF1	K354K	8
PAH	V399V	11
PDHA1	G185G‡	6
PKLR	A423A	9
PTPRC	P48P	4
PTS	E81E‡	4
RET	I647I	11
SMN1	F280F	7
TNFRSF5	T136T	5
UROD	E314E	9



Cartegni et al. (2002).
Nature Reviews Genetics 3, 285-298.

Splicing modulators: the SR proteins

FIGURE 2. Schematic diagram of human SR proteins and SR related proteins. **A**: The domain structures of the known members of the human SR protein family are depicted. RRM: RNA recognition motif; RRMH: RRM homology; Z: zinc knuckle, RS: arginine/serine-rich domain. **B**: The domain structures for some of the human SRrps that participate in pre-mRNA splicing are depicted. All proteins, with the exception of SRm300, are drawn to scale. RRM: RNA recognition motif; RS: arginine/serine-rich domain; Zn: zinc finger; DEXD/H Box: motif characteristic of RNA helicases.

From: B. Graveley (2000). RNA 6, 1197-1211

• Challenges for the future

The basic distinction between exons and introns might be laid down by such proteins long before the specific splicing components bind.

There are over 50 proteins that are not part of the splicing complexes but bind to the RNA independently and affect the choice of sites.

The ***BRCA1 gene***, located on chromosome band 17q21, has a coding sequence of 5,592 nucleotides scattered on 22 exons. However, the exact function of the gene is still unclear, although the encoded proteins are involved in various cellular processes, including transcriptional regulation and DNA repair pathways. Germ-line mutations in this gene lead to the predisposition to breast and ovarian cancer. More than 1,250 different mutations of the *BRCA1* gene have been reported worldwide, spread over the entire length of the gene; most of them lead to premature-termination codons, as a result of frameshift, nonsense, or splice-site mutations or large rearrangements

The power of point mutations
L E Maquat
Nature Genetics **27**, 5 - 6 (2001).

Studies of human diseases have shown that nonsense mutations can alter pre-mRNA splicing. A new study, focusing on the breast cancer susceptibility gene, *BRCA1*, demonstrates that one explanation lies in a disrupted exonic splicing enhancer rather than a disrupted translational reading frame. Despite the growing appreciation of the frequent occurrence of exonic sequences that affect the use of splice sites, the prevalence of single-nucleotide polymorphisms that alter splicing has probably been vastly underestimated.

Nat Genet. 2001, 1:55-8.

A mechanism for exon skipping caused by nonsense or missense mutations in *BRCA1* and other genes

Hong-Xiang Liu[1, 2], Luca Cartegni[1], Michael Q. Zhang[1] & Adrian R. Krainer[1]

Point mutations can generate defective and sometimes harmful proteins. The nonsense-mediated mRNA decay **(NMD)** pathway minimizes the potential damage caused by nonsense mutations. In-frame nonsense codons located at a minimum distance upstream of the last exon-exon junction are recognized as premature termination codons (PTCs), targeting the mRNA for degradation. Some nonsense mutations cause skipping of one or more exons; this phenomenon is termed nonsense-mediated altered splicing **(NAS)**. By analyzing NAS in *BRCA1*, we show here that inappropriate exon skipping can be reproduced *in vitro*, and results from disruption of a splicing enhancer in the coding sequence. Enhancers can be disrupted by single nonsense, missense and translationally silent point mutations, without recognition of an open reading frame as such. These results argue against a nuclear reading-frame scanning mechanism for NAS. Coding-region single-nucleotide polymorphisms (cSNPs) within exonic splicing enhancers or silencers may affect the patterns or efficiency of mRNA splicing, which may in turn cause phenotypic variability and variable penetrance of mutations elsewhere in a gene.

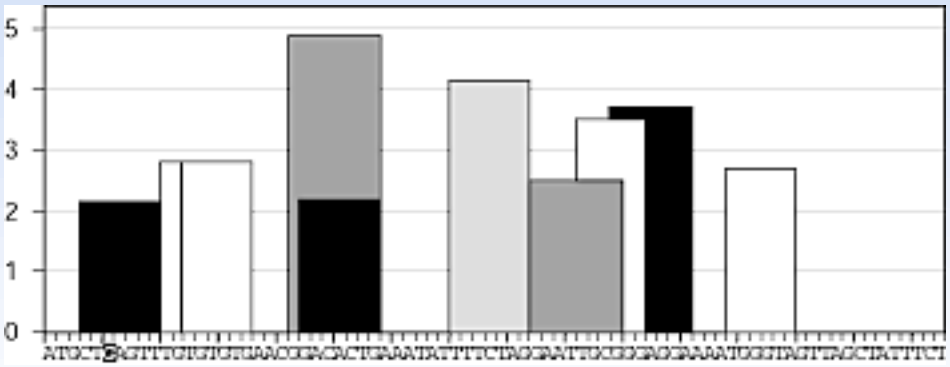
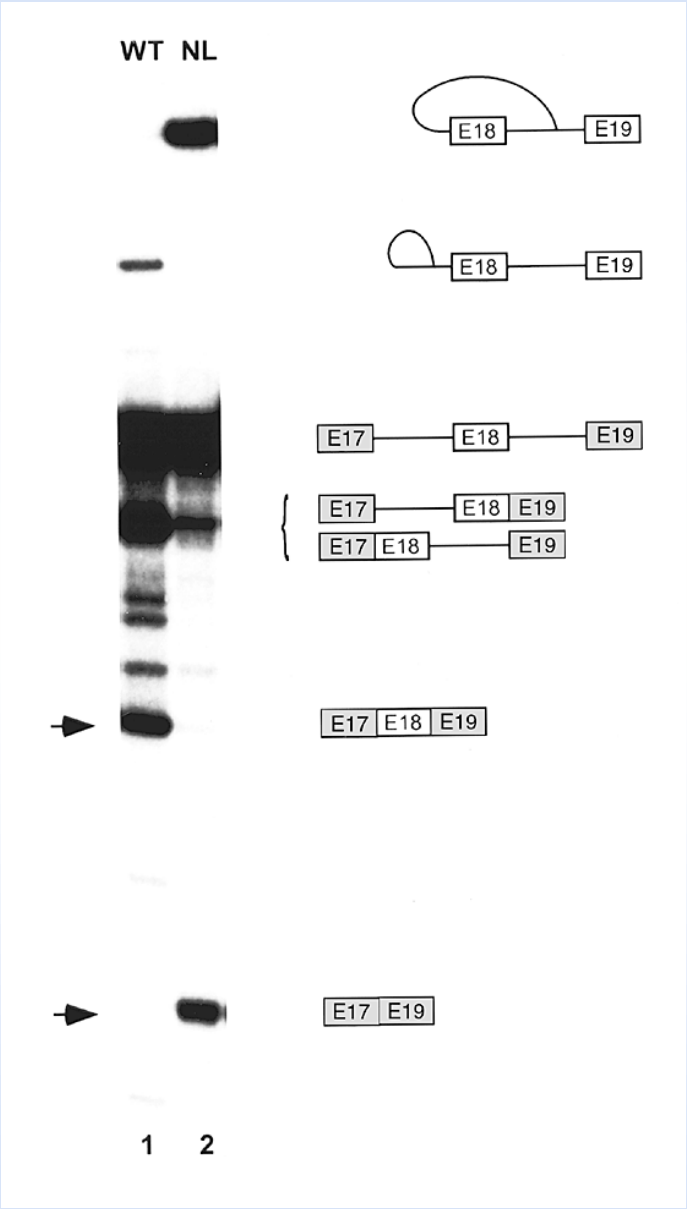


Figure 1: High-score SR protein motifs in *BRCA1* exon 18.

We searched the 78-nt sequence of wild-type exon 18 with four nucleotide-frequency matrices derived from pools of functional enhancer sequences selected *in vitro*. Motif scores reflect the extent of matching to a degenerate consensus, and only the scores above the threshold for each SR protein are shown. High-score motifs are shown in black for SF2/ASF, dark gray for SC35, light gray for SRp40 and white for SRp55. The width of each bar reflects the length of the motif (6, 7 or 8 nt); the placement of each bar along the *x* axis, the position of a motif along the wild-type exon DNA sequence; and the height of the bar, the numerical score on the *y* axis. The thresholds and maximal values are different for each SR protein. The *G* at position 6 (wild type) is highlighted. The nonsense mutation that changes this *G* to a T only affects the first SF2/ASF motif, reducing the score from 2.143 to 0.079 (below the threshold).



In vitro splicing of *BRCA1* minigene transcripts reproduces the exon-skipping phenotype of a nonsense mutation.

We generated **wild-type and mutant *BRCA1* minigene** transcripts by PCR and *in vitro* transcription. We deleted an internal portion of each intron—away from the splice sites and branch site—to generate pre-mRNAs of adequate length for *in vitro* splicing. We spliced wild-type (WT, lane 1) **and nonsense mutant with low SF2/ASF score** (NL, lane 2) radiolabeled transcripts in HeLa cell nuclear extract, and analyzed the products of the reaction by denaturing PAGE and autoradiography. The identity of each band is indicated schematically on the right. Exons 17 and 19 are shown as gray boxes, exon 18 as a white box, and the shortened introns as lines. The arrows indicate the mRNAs generated by exon 18 inclusion or skipping. Although the extent of exon inclusion and skipping varied with different extract preparations or buffer conditions, the ratio of exon skipping over inclusion was reproducibly greater with the mutant pre-mRNA.