

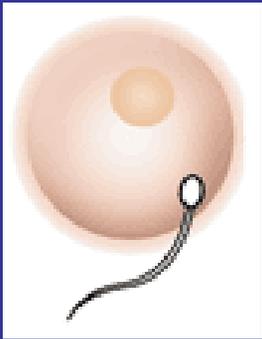


# **Is Alternative Splicing Responsible for Our Higher Level of Complexity?**

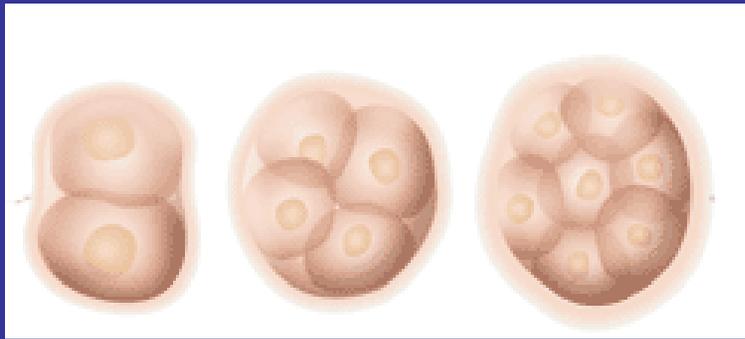
**Gil Ast**

**Department of Human Genetics  
Tel Aviv University Medical School**

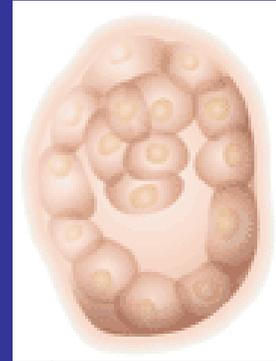
# Embryonic stem cells



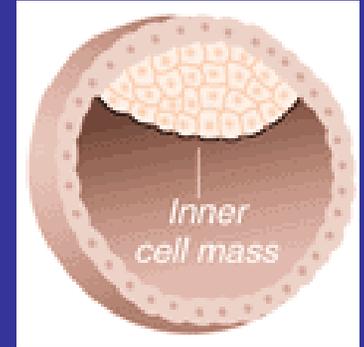
Day 1



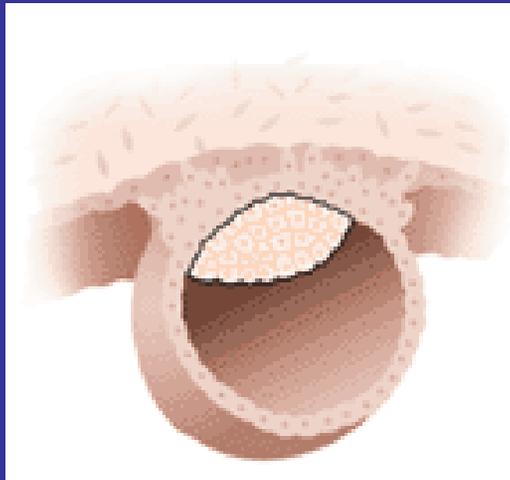
Day 2-3



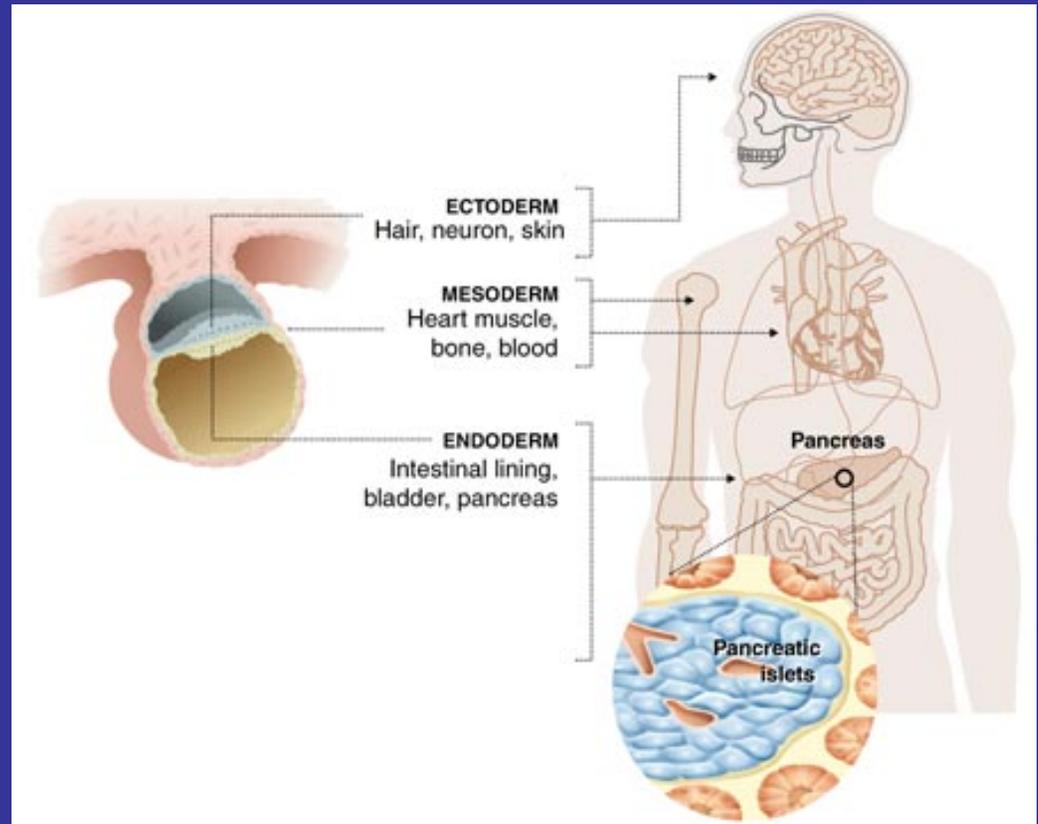
Day 4



Day 5

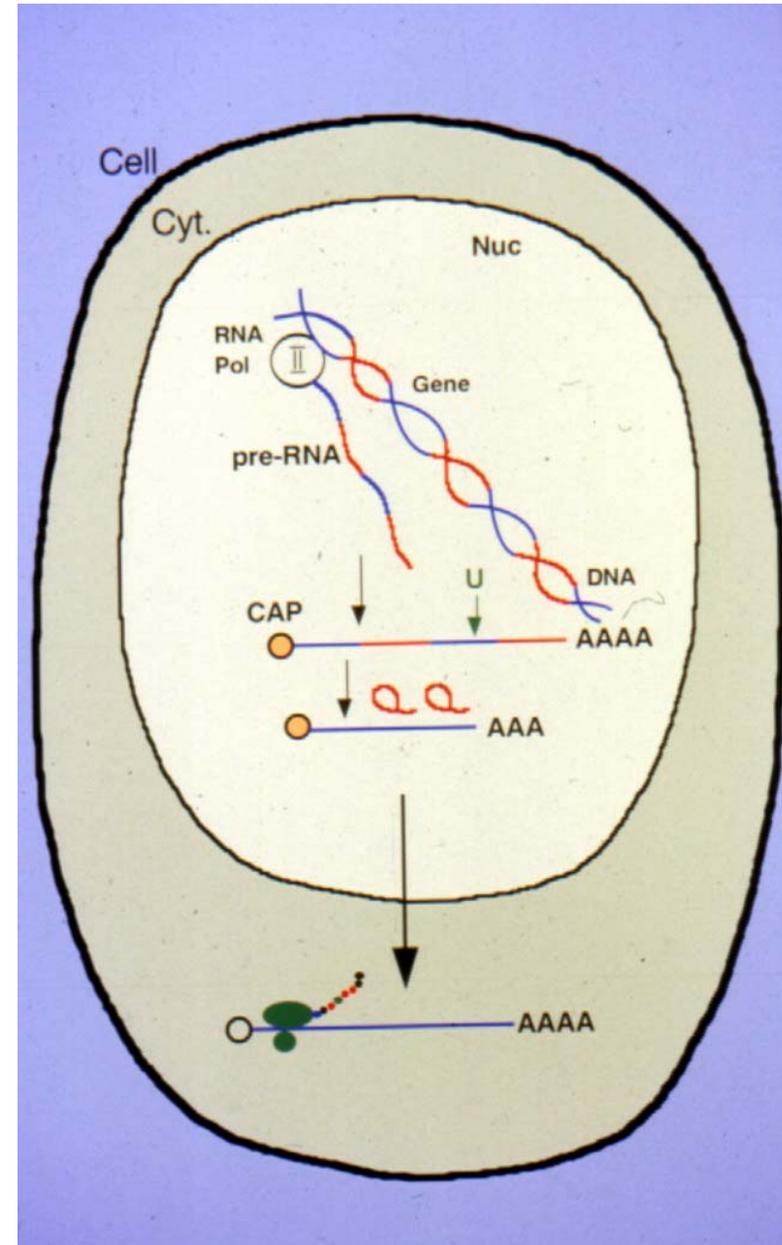
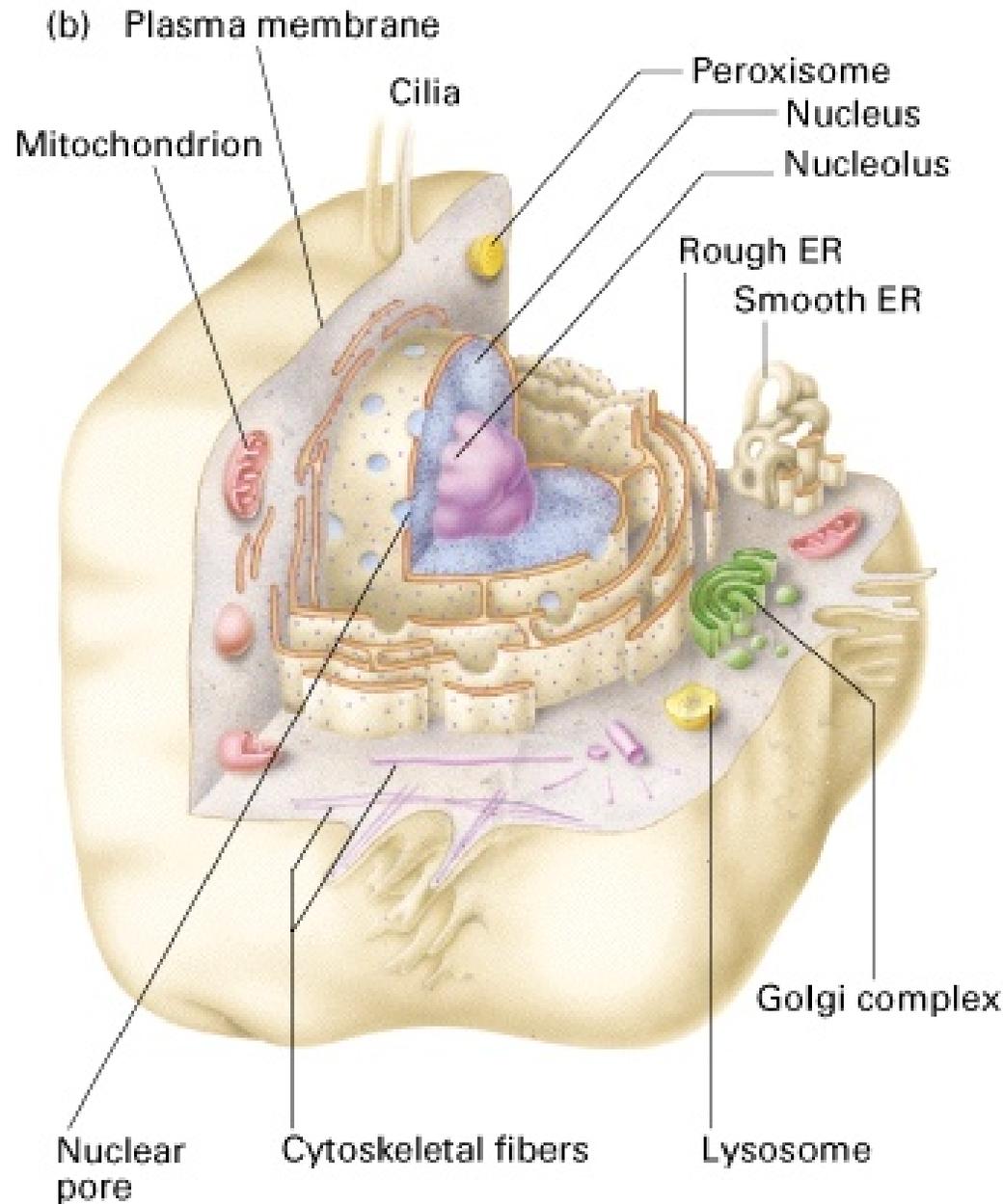


Day 6-9



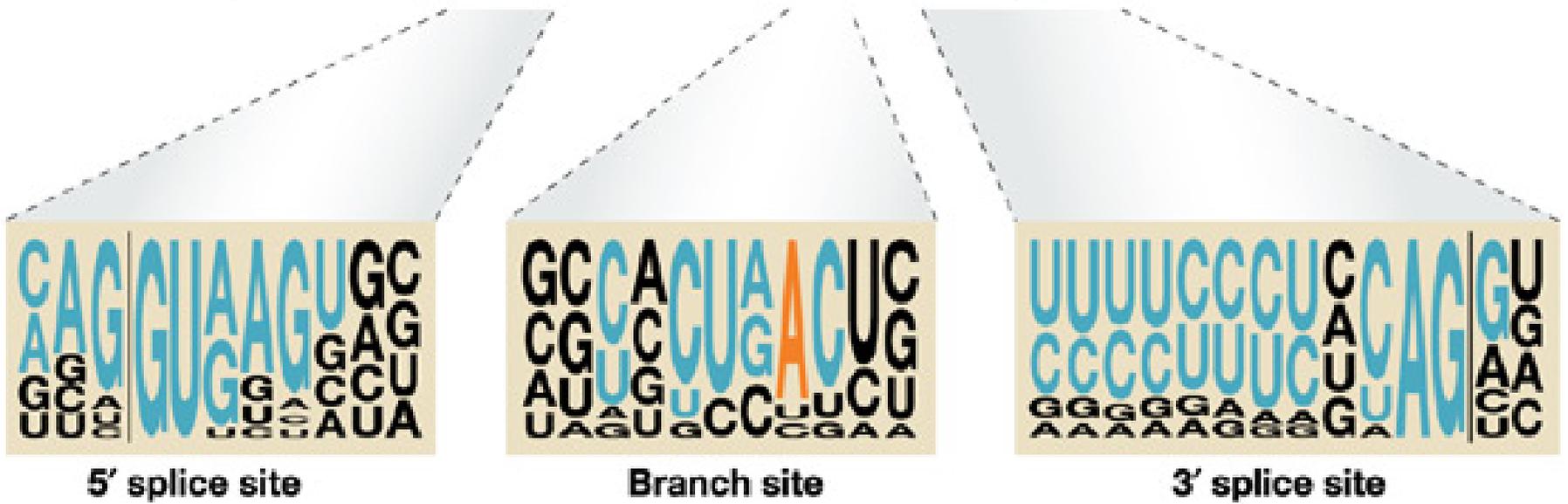
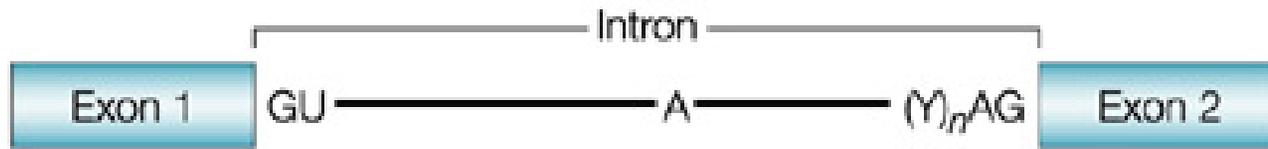
Day 14

# Transcription regulation



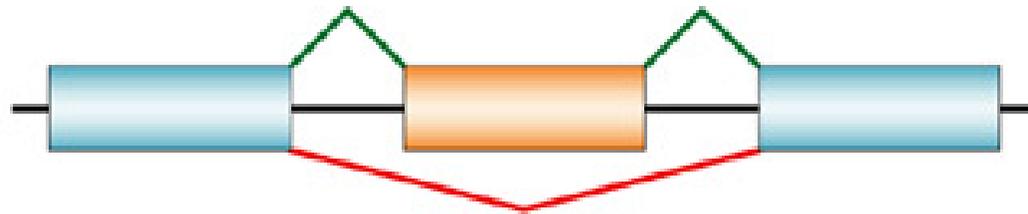
# Consensus sequences

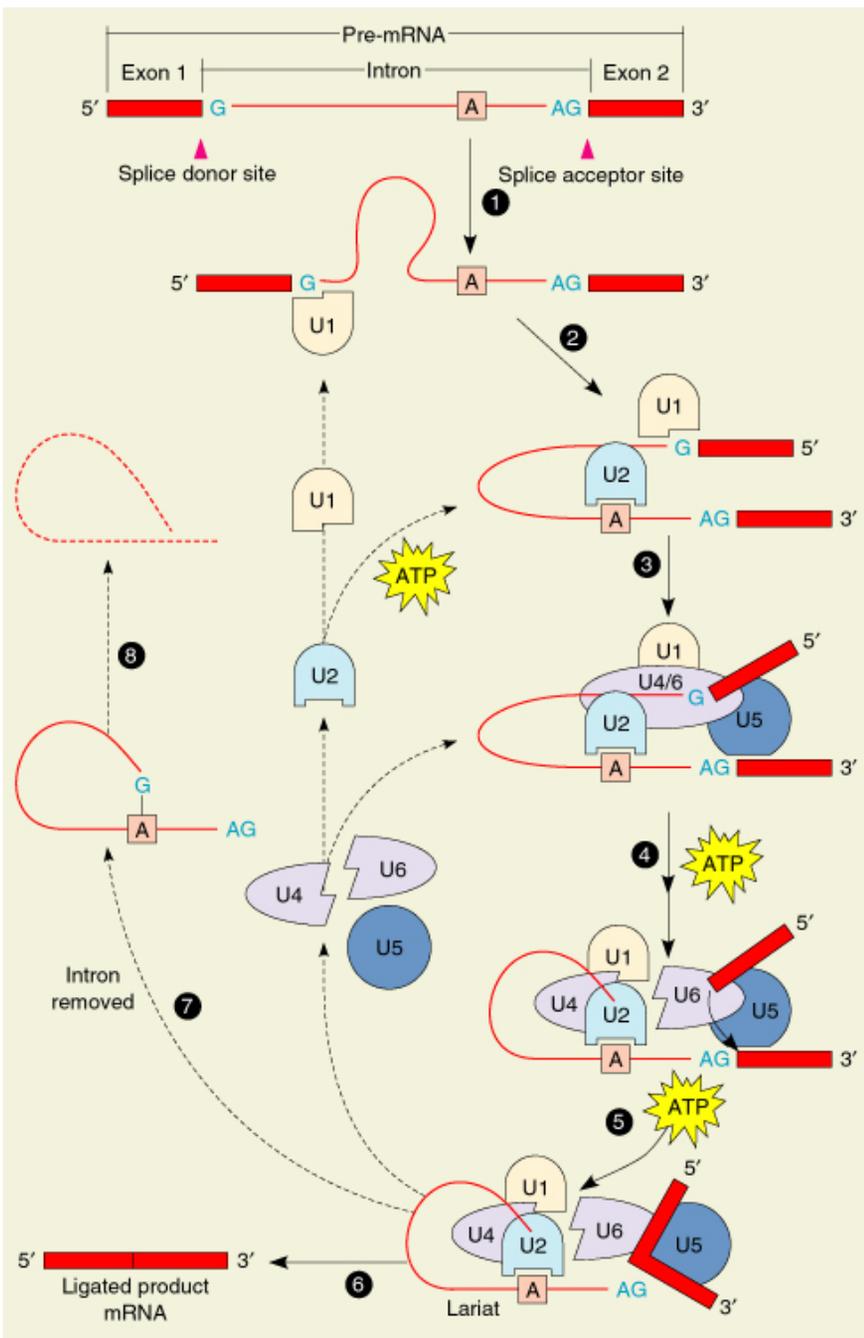
**a**



**b**

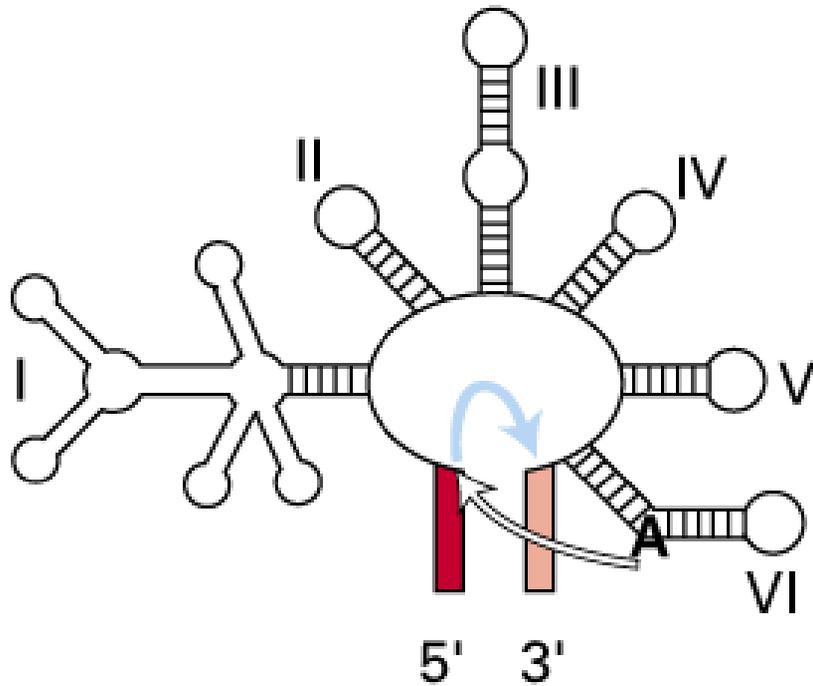
Exon skipping/inclusion





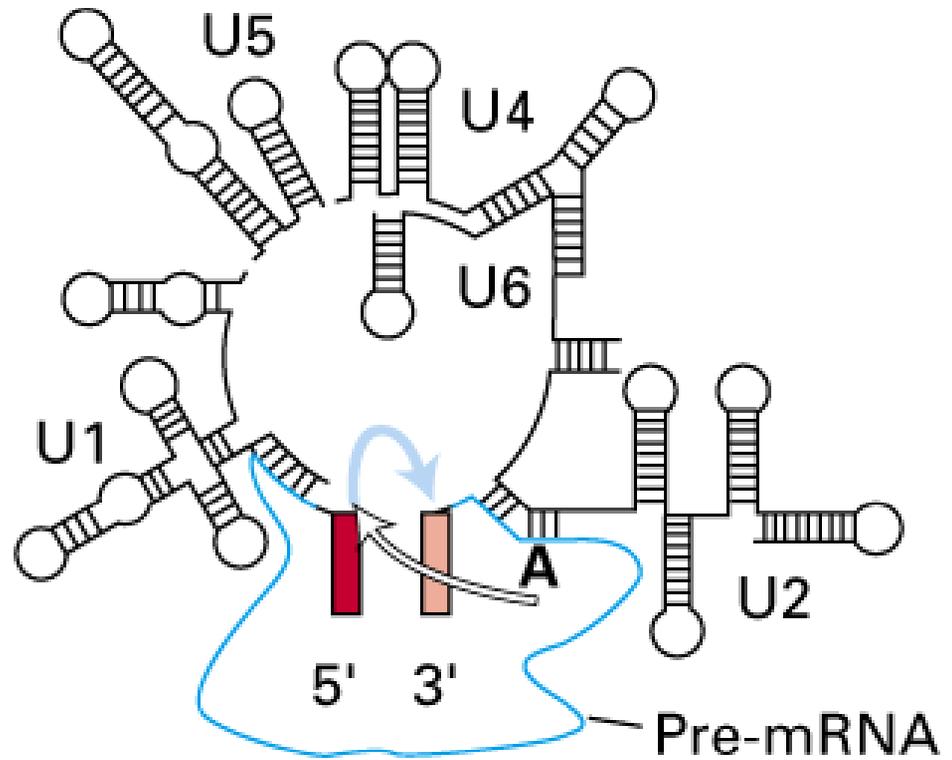
# Evolution of snRNA

(a)



Group II intron

(b)



U snRNAs in spliceosome

# Human genome

2.91 billion base pairs

26,000 protein coding genes

1.5% exons (129 nucleotides)

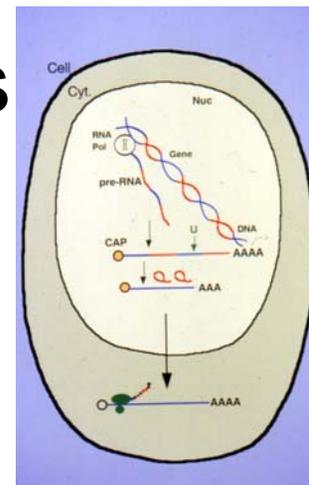
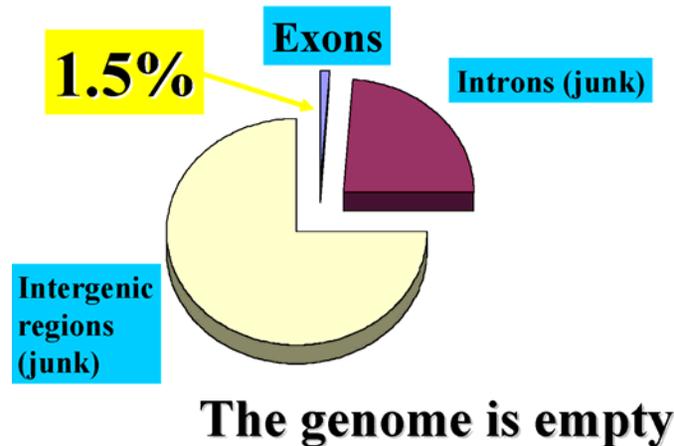
24% introns (~3,000 nucleotides)

75% intergenic (no genes)

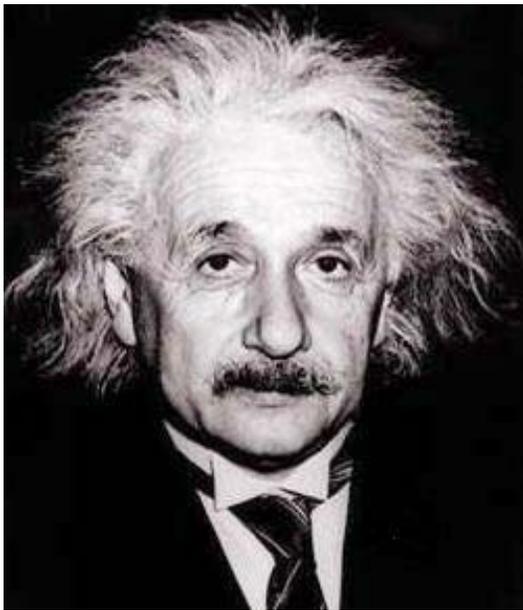
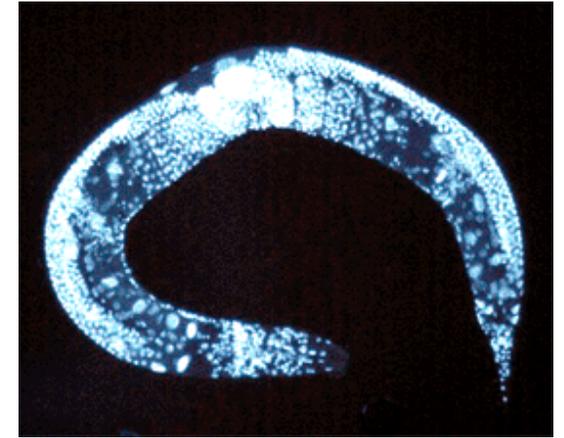
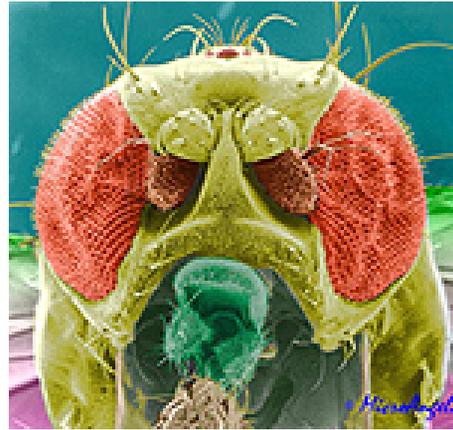
Average size of a gene is 27,894 bases

Contains an average of 11 exons

Titin contains 234 exons.



# Human vs. the rest of the world



19,500 genes  
14,000 genes

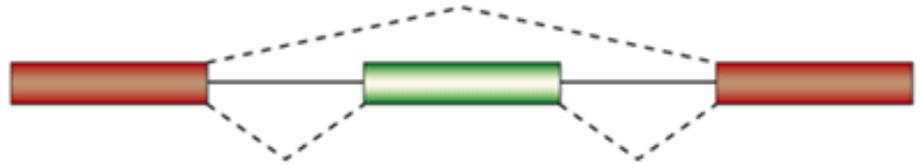
26,000 genes

26,000 genes

40,000 genes

# Different types of alternative splicing

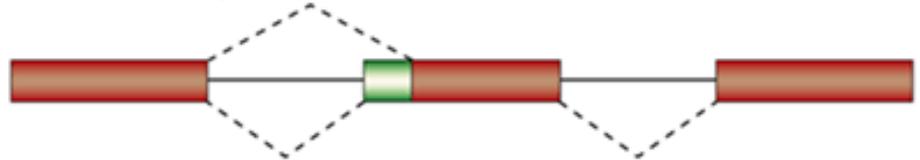
Exon skipping 38%



Alternative 5' splice sites 18%



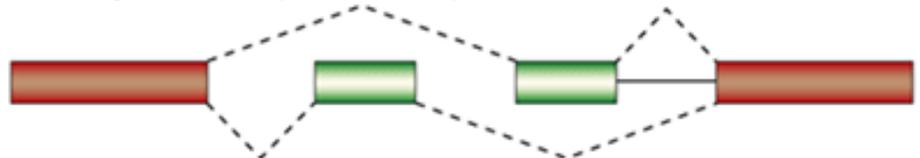
Alternative 3' splice sites 8%



Intron retention 3%

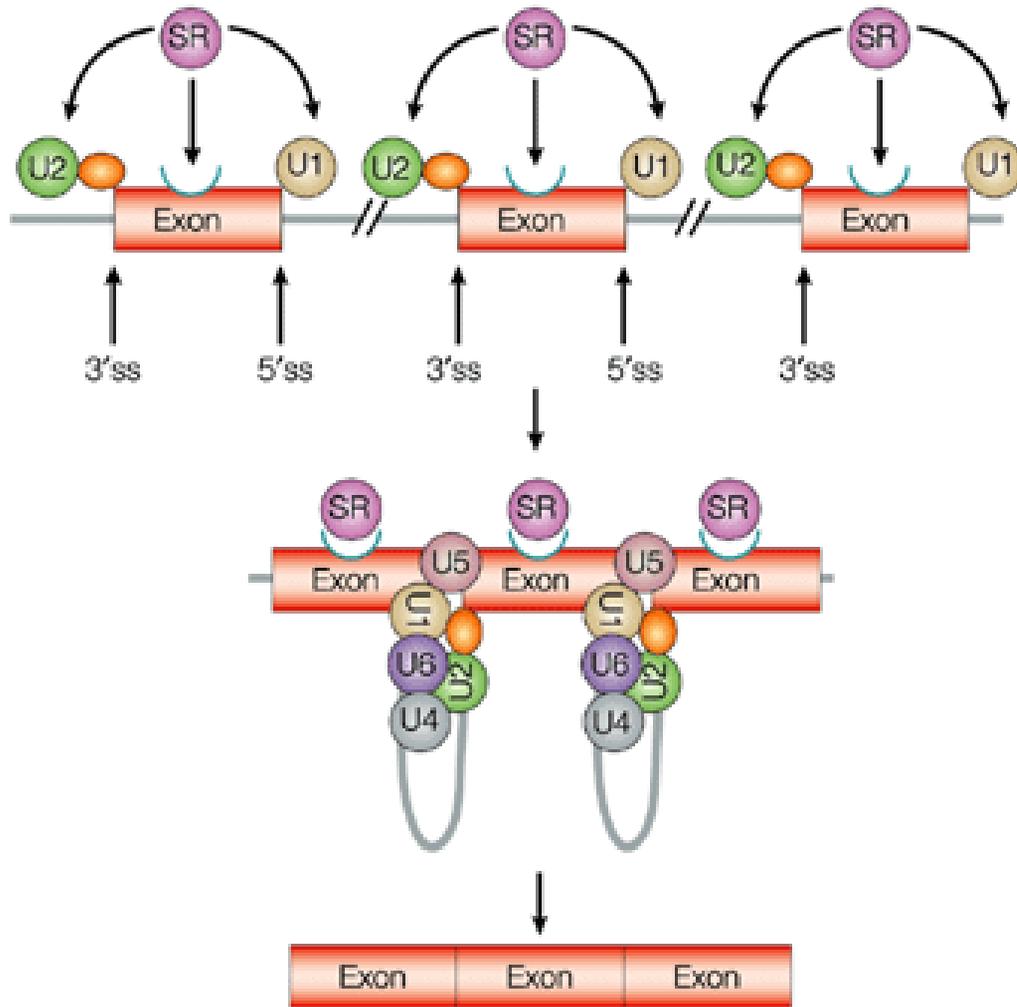


Mutually exclusive (% Unknown)

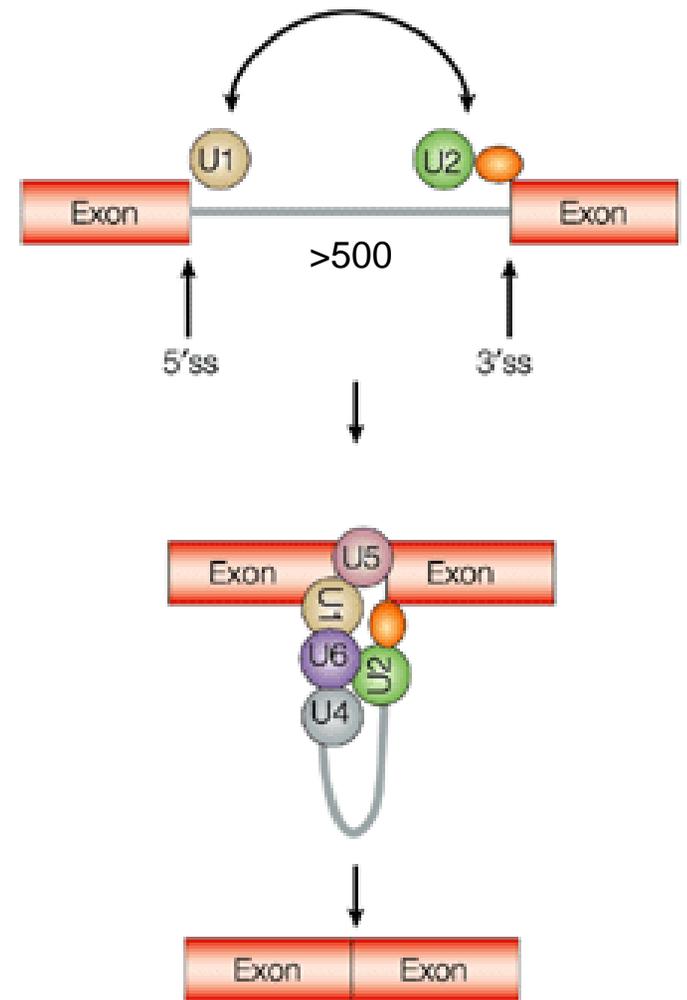


# Exon versus Intron Definition

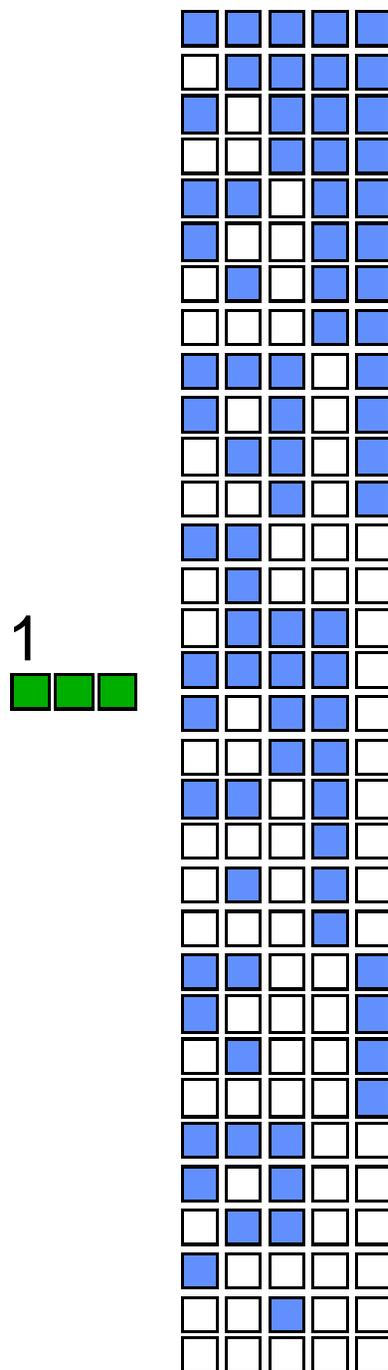
**a Exon definition**



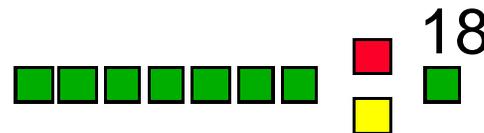
**b Intron definition**



The Troponin T (muscle protein) pre-mRNA is alternatively spliced to give rise to 64 different isoforms of the protein



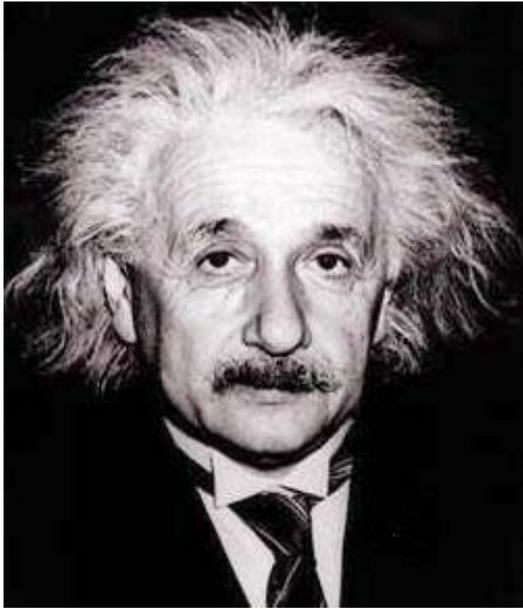
- Constitutively spliced exons (exons 1-3, 9-15, and 18)
- } Mutually exclusive exons (exons 16 and 17)
- Alternatively spliced exons (exons 4-8)



Exons 4-8 are spliced in every possible way giving rise to 32 different possibilities

Exons 16 and 17, which are mutually exclusive, double the possibilities; hence 64 isoforms

# Human vs. the rest of the world



**14,000 genes**



**19,500 genes**

**Condoleezza Rice**



**26,000 genes**

**Mustard**



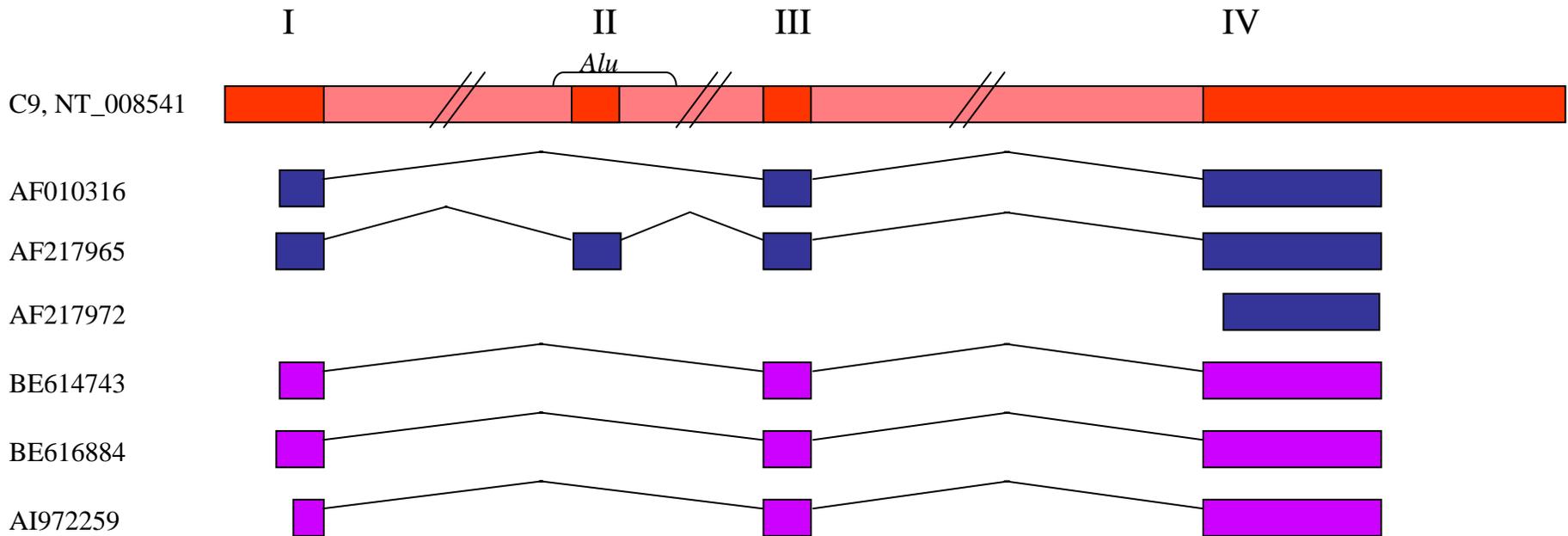
ARTIE.COM

**26,000 genes**



**40,000 genes**

# Identification of alternatively spliced *Alu* exon



**5.2% of the alternatively spliced exons (1182) has a significant hit to an *Alu* sequence ( $E < 10^{-10}$ ). But, none of the constitutively spliced exons (4152) has a significant hit to an *Alu* sequence. That means, that *Alu*-containing exons are alternatively spliced.**

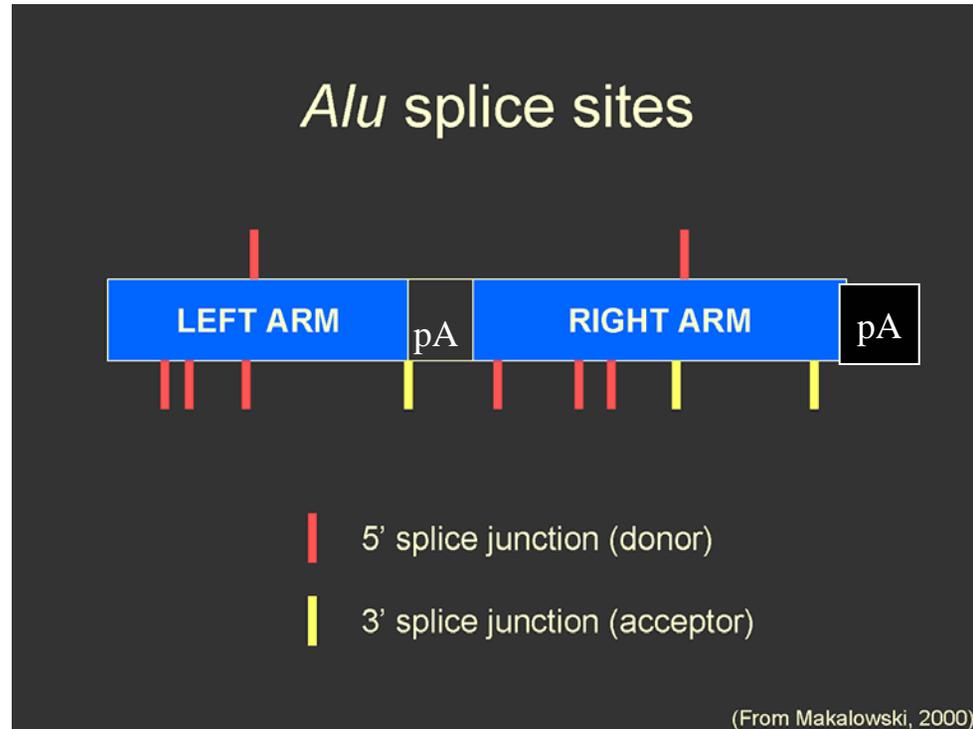
*Alu* are Retrotransposons.

*Alu* sequences both comprise more than 11% of the genome and have reached a copy number of about 1.4 million.

*Alu* elements are short interspersed elements (SINEs), typically 300 nucleotides long, containing two cassettes separated by a polyA tail.

***Alu* elements are unique to primates**

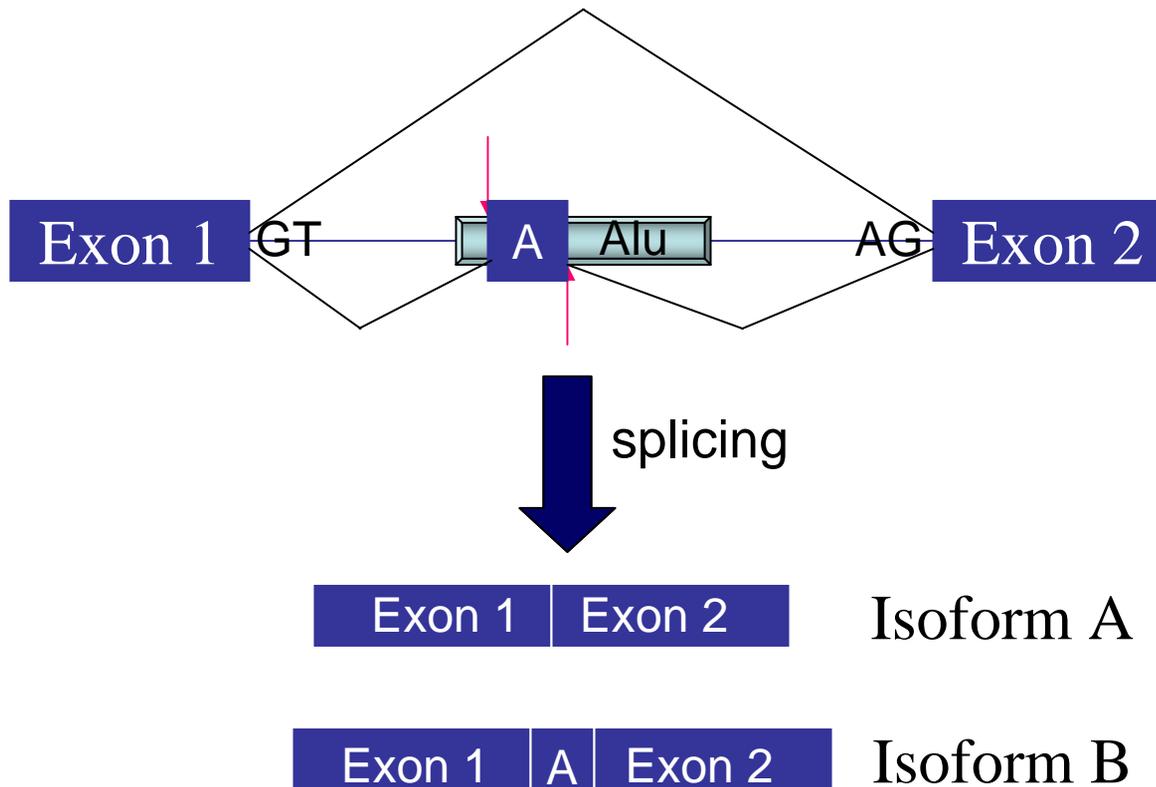
# *Alu* elements



All introns of >1000 bp contain at least 1 *Alu*

AATAAA – Integration site

# Exonization of *Alu* elements



Alport

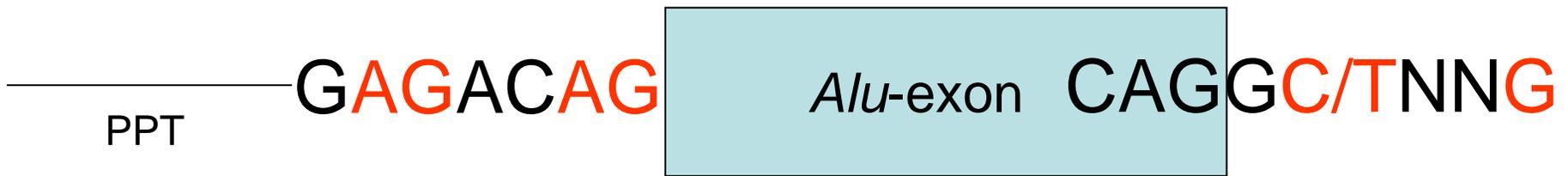
Sly

OAT

The only selective pressure that was found in the exonization of *Alu* elements was in creating or maintaining of weak splice sites

The 3'SS of *Alu*-exons

The 5'SS of *Alu*-exons

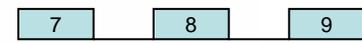


Lev-Maor et al., 2003

Sorek et al., 2004

# The Selection of 3'SS of *Alu*-exons

Gene name	Alu exon number	subfam	intron																exon										
			position relative to distal 3'ss																-7	-6	-5	-4	-3	-2	-1				
			290	289	288	287	286	285	284	283	282	281	280	279	278	277	276	275	274	273	272	271							
		J	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T				
		S	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	G	G	A	G	T	C	T				
1 ADAR2	8	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T				
2 TFB2M	4	Jb	T	C	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T				
3 MVK	4	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T				
4 CBFA2T2	3	Jo	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T				
5 NPD002	7	Jb	T	T	T	C	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	C				
6 MOG	3	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	T	G	T	C	T				
7 n/a	5	Jb	T	T	T	A	T	T	T	T	T	T	T	G	A	G	A	C	A	G	A	G	T	C	T				
8 PTGES	2	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	A	G	T	C	T				
9 DAF*	10	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	T	T	C	T				
10 STK2*	n/a	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	A	G	T	C	T				
11 MLANA	4	Jb	T	T	T	T	T	C	T	T	T	T	T	G	A	A	A	C	A	G	G	A	A	A	T				
12 n/a	24	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	A	A	C	A	G	C	G	T	C	T				
13 ITGB1*	7	Sx	T	T	T	A	T	T	T	T	T	T	T	G	A	G	A	C	A	G	-	-	T	C	T				
14 n/a	2	Sg	T	C	T	T	T	T	T	T	T	T	T	T	G	A	C	A	G	A	G	T	C	C					
15 MBD3	12	Sx	T	T	T	T	T	T	T	T	T	T	T	G	T	G	A	C	A	G	A	G	T	C	T				
16 CNN2	6	Sx	T	T	T	A	T	T	T	T	T	T	T	G	A	G	A	T	A	G	G	A	T	C	T				
17 PGT	12	Sp	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	G	G	A	G	T	T	T				
18 n/a	2	Sg	C	T	T	T	T	T	T	T	T	T	T	G	A	G	A	T	G	G	A	G	T	C	T				
19 RES4-22	18	Sg	T	T	T	A	T	T	T	T	C	G	A	G	A	T	G	G	A	G	T	T	T	T	T				
20 LOC51193	5	Sg	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	T	G	G	A	G	T	C	T				
21 n/a	3	Sx	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	T	G	G	A	G	T	C	C				
22 CHRNA3*	6	Sx	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	T	G	G	A	G	T	C	T				
23 PTD011	2	Sx	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	G	C	C	C	A	G	G				
24 HCA66	18	Sg	T	T	T	T	T	T	T	T	T	T	T	T	A	G	A	C	G	G	A	G	T	C	T				
25 CYP3A43	8	Sg	T	T	T	T	T	T	T	T	T	T	T	T	A	G	A	C	A	G	A	G	T	C	T				
26 LCAT*	6	Jo	T	T	T	T	T	T	T	T	T	T	T	T	A	G	A	G	A	C	A	G	G	T	T				
27 KIAA1169	24	Jo	T	T	T	T	G	T	T	T	T	T	T	T	A	G	A	G	A	T	G	G	T	A	T				
28 SLC3A2	6	Jb	T	T	G	T	T	T	T	T	T	T	T	A	A	G	A	C	A	G	C	A	T	T	T				
29 ICAM2	0	Jb	T	T	T	G	T	T	T	T	T	T	T	A	A	G	A	C	A	G	G	G	T	C	T				
30 COL4A3	6	Sx	T	T	T	T	T	C	T	T	T	T	T	T	A	G	A	T	G	G	A	G	T	C	T				
31 GUSB	9	Sg/x	A	T	T	T	T	T	T	T	T	T	T	T	G	A	T	A	T	G	C	A	G	T	C	T			
32 OAT	4	Jo	T	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	A	G	T	T	T			



12/16 Alu J

9/13 Alu S

PPT

proximal AG      distal AG

Figure 1  
Lev-Maor et al.

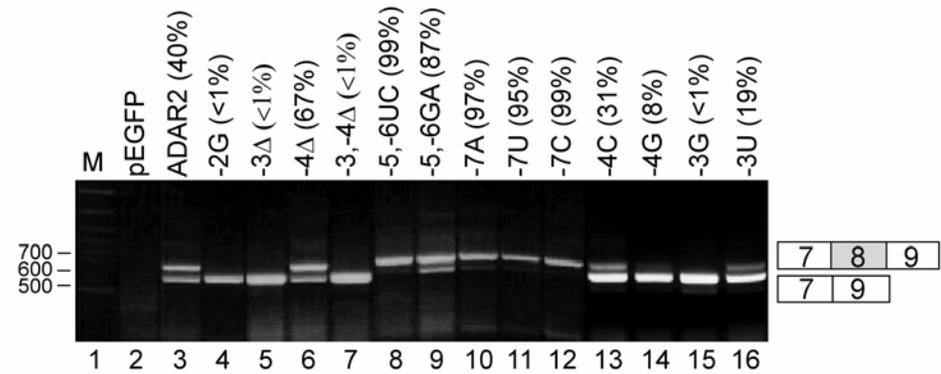
# A weak 3' splice site determine alternative splicing of *Alu*-exons

a

```

-7      -1
ADAR2   TTTTTTTTTTGAGACAGGGGTCTCGCTCTT
-2G     TTTTTTTTTTGAGACGGGGTCTCGCTCTT
-3Δ     TTTTTTTTTTGAGA-AGGGTCTCGCTCTT
-4Δ     TTTTTTTTTTGAG-CAGGGTCTCGCTCTT
-3,-4Δ  TTTTTTTTTTGAG--AGGGTCTCGCTCTT
-5,-6UC TTTTTTTTTTGTCACAGGGTCTCGCTCTT
-5,-6GA TTTTTTTTTTGGAACAGGGTCTCGCTCTT
-7A     TTTTTTTTTTAAGACAGGGTCTCGCTCTT
-7U     TTTTTTTTTTTAGACAGGGTCTCGCTCTT
-7C     TTTTTTTTTTCAGACAGGGTCTCGCTCTT
-4G     TTTTTTTTTTGAGGCAGGGTCTCGCTCTT
-4C     TTTTTTTTTTGAGGCAGGGTCTCGCTCTT
-3G     TTTTTTTTTTGAGAGGAGGGTCTCGCTCTT
-3U     TTTTTTTTTTGAGATTAGGGTCTCGCTCTT
    
```

b



**Fig. 2b**  
Lev-Maor *et al.*

a

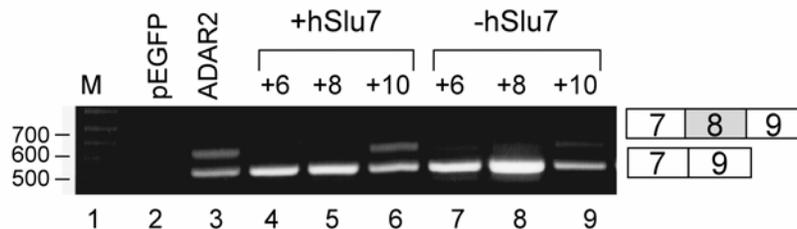
```

ADAR2   TTTTTGAGACAGGGTCT
+6bp    TTTTTGAGACATAGGGTCT
+8bp    TTTTTGAGACATCAAGGGTCT
+10bp   TTTTTGAGACATCAACAGGGTCT
    
```

The 3' splice site of *Alu*-exons



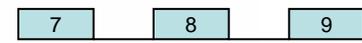
b



**Fig. 3**  
Lev-Maor *et al.*

# The Selection of 3'SS of *Alu*-exons

Gene name	Alu exon number	subfam	intron																exon										
			position relative to distal 3'ss																-7	-6	-5	-4	-3	-2	-1				
			290	289	288	287	286	285	284	283	282	281	280	279	278	277	276	275	274	273	272	271							
		J	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T				
		S	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	G	G	A	G	T	C	T				
1 ADAR2	8	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T				
2 TFB2M	4	Jb	T	C	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T				
3 MVK	4	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T				
4 CBFA2T2	3	Jo	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T				
5 NPD002	7	Jb	T	T	T	C	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	C				
6 MOG	3	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	T	G	T	C	T				
7 n/a	5	Jb	T	T	T	A	T	T	T	T	T	T	T	G	A	G	A	C	A	G	A	G	T	C	T				
8 PTGES	2	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	A	G	T	C	T				
9 DAF*	10	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	T	T	C	T				
10 STK2*	n/a	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	A	G	T	C	T				
11 MLANA	4	Jb	T	T	T	T	T	C	T	T	T	T	T	G	A	A	A	C	A	G	G	A	A	A	T				
12 n/a	24	Jb	T	T	T	T	T	T	T	T	T	T	T	G	A	A	A	C	A	G	C	G	T	C	T				
13 ITGB1*	7	Sx	T	T	T	A	T	T	T	T	T	T	T	G	A	G	A	C	A	G	-	-	T	C	T				
14 n/a	2	Sg	T	C	T	T	T	T	T	T	T	T	T	T	G	A	C	A	G	A	G	T	C	C					
15 MBD3	12	Sx	T	T	T	T	T	T	T	T	T	T	T	G	T	G	A	C	A	G	A	G	T	C	T				
16 CNN2	6	Sx	T	T	T	A	T	T	T	T	T	T	T	G	A	G	A	T	A	G	G	A	T	C	T				
17 PGT	12	Sp	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	G	G	A	G	T	T	T				
18 n/a	2	Sg	C	T	T	T	T	T	T	T	T	T	T	G	A	G	A	T	G	G	A	G	T	C	T				
19 RES4-22	18	Sg	T	T	T	A	T	T	T	T	C	G	A	G	A	T	G	G	A	G	T	T	T	T	T				
20 LOC51193	5	Sg	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	T	G	G	A	G	T	C	T				
21 n/a	3	Sx	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	T	G	G	A	G	T	C	C				
22 CHRNA3*	6	Sx	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	T	G	G	A	G	T	C	T				
23 PTD011	2	Sx	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	G	C	C	C	A	G	G				
24 HCA66	18	Sg	T	T	T	T	T	T	T	T	T	T	T	T	A	G	A	C	G	G	A	G	T	C	T				
25 CYP3A43	8	Sg	T	T	T	T	T	T	T	T	T	T	T	T	A	G	A	C	A	G	A	G	T	C	T				
26 LCAT*	6	Jo	T	T	T	T	T	T	T	T	T	T	T	T	A	G	A	G	A	C	A	G	G	T	T				
27 KIAA1169	24	Jo	T	T	T	T	G	T	T	T	T	T	T	T	A	G	A	G	A	T	G	G	T	A	T				
28 SLC3A2	6	Jb	T	T	G	T	T	T	T	T	T	T	T	A	A	G	A	C	A	G	C	A	T	T	T				
29 ICAM2	0	Jb	T	T	T	G	T	T	T	T	T	T	T	A	A	G	A	C	A	G	G	G	T	C	T				
30 COL4A3	6	Sx	T	T	T	T	T	C	T	T	T	T	T	T	A	G	A	T	G	G	A	G	T	C	T				
31 GUSB	9	Sg/x	A	T	T	T	T	T	T	T	T	T	T	T	G	A	T	A	T	G	C	A	G	T	C	T			
32 OAT	4	Jo	T	T	T	T	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	A	G	T	T	T			



12/16 Alu J

9/13 Alu S

PPT

proximal AG      distal AG

Figure 1  
Lev-Maor et al.

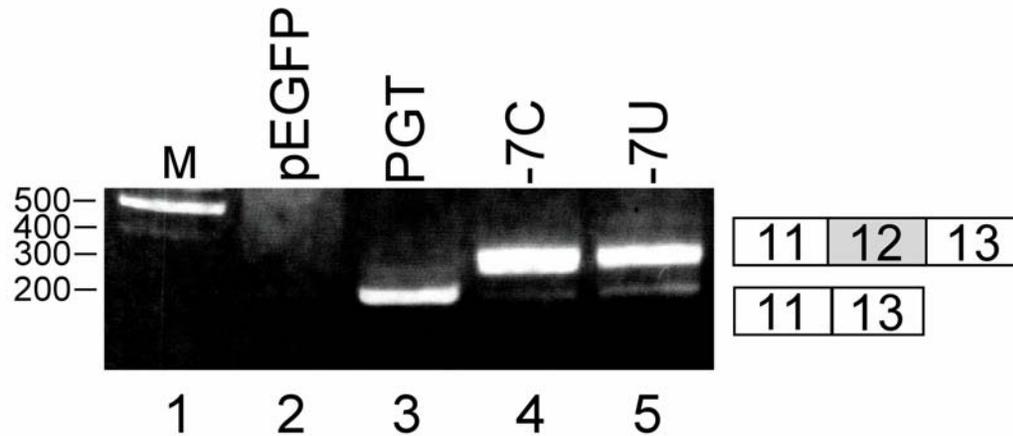
# Exonization of an intronic *Alu*

**a**

```

                -7       -1
ADAR2  TTTTTTTTTTTGAGACAGGGTCTCGCTCTT
PGT    TTTTTTTTTTTGAGACGGAGTTTCGCTCTT
-7C   TTTTTTTTTTTTCAGACGGAGTTTCGCTCTT
-7U   TTTTTTTTTTTTTAGACGGAGTTTCGCTCTT
Alport TTTTTTTTTTTTTAGACGGAGTCTCGCTCTT
    
```

**b**



476,000 Alus in introns; 238,000 in antisense orientation;  
76,000 contain ADAR2 or PGT like 3'ss.

# 5'ss selection of *Alu* exons

Gene name	Alu exon#	subfamily	Position relative to 5'ss																										
			exon									intron																	
			172	171	170	169	168	167	166	165	164	163	162	161	-3	-2	-1	1+	2+	3+	4+	5+	6+	7+					
1 MLANA	4	J	T	A	G	C	T	G	G	G	A	C	C	A	G	G	C	A	C	G	T	G	C	C	A	C	C		
2 RES4-22	18	S	T	A	G	T	T	G	G	A	A	T	T	A	C	A	G	G	C	A	A	G	C	A	A	A	A	C	C
3 HPK1*	31	S	T	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	C	A	T	G	T	G	C	C	A	C	C
4 ADAR2	8	J	G	A	G	C	T	G	G	G	A	C	T	A	C	A	G	G	C	A	T	G	T	A	C	C	A	C	T
5 KIAA1169	24	J	C	A	G	C	T	G	G	A	A	C	T	A	C	A	G	G	T	G	T	G	C	A	C	T	G	T	G
6 LOC51193	5	S	T	A	G	C	T	G	G	G	G	T	T	A	C	A	G	G	T	G	T	G	C	G	C	C	A	C	C
7 ZFX	2	S	T	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	T	G	T	G	C	G	C	C	C	C	C
8 MVK	4	J	T	A	G	C	T	G	G	G	A	C	C	A	C	A	G	G	T	G	T	G	C	G	C	C	A	C	C
9 PTGES	2	J	T	A	G	C	T	G	G	G	A	C	C	A	C	A	G	G	T	G	T	G	T	A	T	C	A	C	C
10 N/A	6	J	T	A	G	C	T	G	G	G	A	C	T	A	C	A	G	G	T	G	T	G	T	G	C	C	A	C	C
11 EVI5	3	J	T	G	G	C	T	G	G	G	A	C	T	A	C	A	G	G	T	G	T	G	T	G	C	C	A	T	C
12 C20orf26	9	J	T	A	G	C	T	G	G	G	A	C	T	A	T	A	G	G	T	A	T	G	T	G	C	C	A	C	C
13 N/A	4	S	T	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	T	G	T	G	T	G	C	C	A	C	C
14 BRCA2	20	S	T	A	G	C	T	G	G	G	A	C	T	A	C	A	G	G	T	G	C	G	T	G	C	C	A	C	C
15 CNN2	6	S	T	A	G	C	T	G	G	G	A	C	T	A	C	A	G	G	T	G	C	A	T	G	C	T	G	C	C
16 BIRC3	2	S	T	A	G	C	T	G	G	A	A	A	T	A	C	A	G	G	T	G	C	G	T	G	C	C	A	C	C
17 CYP3A43	8	S	G	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	T	A	C	A	C	A	C	A	C	C	C
18 N/A	2	S	T	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	T	G	C	C	C	G	C	C	A	C	C
19 MBD3	12	S	T	A	G	C	T	G	G	G	A	T	T	T	C	A	G	G	T	A	C	C	C	G	T	C	A	C	A
20 PLA2G4B	2	S	T	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	T	G	C	C	C	A	C	C	A	C	C
21 BCAS4	5	S	T	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	T	G	C	G	C	G	C	T	A	C	C
22 ICAM2	2	J	T	A	G	C	T	G	G	G	A	T	C	A	C	A	G	G	T	G	A	G	A	G	C	C	A	C	C
23 TGM4*	2	FAM	T	A	A	C	C	G	G	G	A	T	T	A	C	A	G	G	T	A	T	G	T	G	A	C	T	C	C
24 Integrin $\beta 1^*$	7	S	T	A	C	C	T	G	G	G	A	T	T	A	C	A	G	G	T	G	C	C	T	G	C	C	T	C	A
25 CHRNA3*	5	S	T	G	T	C	T	G	G	G	A	C	T	A	C	A	G	G	T	A	C	C	C	G	C	C	C	G	C
26 CTDP1*	7	Y	G	T	G	T	T	G	G	G	A	T	T	A	C	A	G	G	T	A	T	G	A	G	C	C	A	T	T

Fig. 1

# Two positions differ intronic from exonic *Alus*.

A

Position	Alu consensus	Genomic, non-exonized <i>Alus</i> in introns							
		Count per position				Percent per position			
		A	G	T	C	A	G	T	C
172	T	2517	1613	155796	4583	1.5%	1.0%	94.7%	2.8%
171	A	158590	3870	1515	649	96.3%	2.4%	0.9%	0.4%
170	G	5360	153227	2419	3471	3.3%	93.2%	1.5%	2.1%
169	C	1171	2025	6161	155132	0.7%	1.2%	3.7%	94.3%
168	T	1354	1734	159323	2179	0.8%	1.1%	96.8%	1.3%
167	G	9873	152484	1317	630	6.0%	92.8%	0.8%	0.4%
166	G	9879	151737	1819	1122	6.0%	92.2%	1.1%	0.7%
165	G	9679	151024	2073	1426	5.9%	92.0%	1.3%	0.9%
164	A	157893	3364	2273	890	96.0%	2.0%	1.4%	0.5%
163	T	1627	1465	82293	78985	1.0%	0.9%	50.1%	48.1%
162	T	2408	1151	150771	9988	1.5%	0.7%	91.8%	6.1%
161	A	155739	5616	1169	1817	94.8%	3.4%	0.7%	1.1%
-3 160	C	2274	1642	14755	145456	1.4%	1.0%	9.0%	88.6%
-2 159	A	158275	3715	1356	958	96.3%	2.3%	0.8%	0.6%
-1 158	G	6277	155408	1463	877	3.8%	94.7%	0.9%	0.5%
1 157	G	8282	150508	2520	2201	5.1%	92.0%	1.5%	1.3%
2 156	C	3200	2603	49508	108038	2.0%	1.6%	30.3%	66.1%
3 155	G	70291	86492	3762	3155	42.9%	52.8%	2.3%	1.9%
4 154	C	2302	2337	61547	97440	1.4%	1.4%	37.6%	59.6%
5 153	G	32027	71147	4427	55774	19.6%	43.5%	2.7%	34.1%
6 152	C	4294	2941	66256	89861	2.6%	1.8%	40.6%	55.0%
7 151	G	55629	102805	1849	2736	34.1%	63.1%	1.1%	1.7%
150	C	2039	1551	7413	152348	1.2%	0.9%	4.5%	93.3%
149	C	2779	860	9448	150848	1.7%	0.5%	5.8%	92.0%
148	A	156184	4269	1394	2387	95.1%	2.6%	0.8%	1.5%
147	C	2190	2055	8776	151244	1.3%	1.3%	5.3%	92.1%
146	C	3526	1782	11503	147531	2.1%	1.1%	7.0%	89.8%

B

Position	Alu consensus	Exonized <i>Alus</i>								E value
		Observed				Expected				
		A	G	T	C	A	G	T	C	
172	T	0	2	22	1	0.4	0.2	23.7	0.7	
171	A	23	2	0	0	24.1	0.6	0.2	0.1	
170	G	1	22	1	1	0.8	23.3	0.4	0.5	
169	C	0	0	1	24	0.2	0.3	0.9	23.6	
168	T	0	0	24	1	0.2	0.3	24.2	0.3	
167	G	0	25	0	0	1.5	23.2	0.2	0.1	
166	G	0	24	0	0	1.4	22.1	0.3	0.2	
165	G	3	22	0	0	1.5	23.0	0.3	0.2	
164	A	24	1	0	0	24.0	0.5	0.3	0.1	
163	T	1	0	13	11	0.2	0.2	12.5	12.0	
162	T	0	0	21	4	0.4	0.2	22.9	1.5	
161	A	24	0	1	0	23.7	0.9	0.2	0.3	
-3 160	C	0	0	1	24	0.3	0.3	2.2	22.2	
-2 159	A	25	0	0	0	24.1	0.6	0.2	0.1	
-1 158	G	0	25	0	0	1.0	23.7	0.2	0.1	
1 157	G	0	25	0	0	1.3	23.0	0.4	0.3	
2 156	C	0	0	21	4	0.5	0.4	7.6	16.5	1.82E-07
3 155	G	10	15	0	0	10.7	13.2	0.6	0.5	
4 154	C	2	0	12	11	0.4	0.4	9.4	14.9	
5 153	G	2	18	0	5	4.9	10.9	0.7	8.5	3.66E-02
6 152	C	1	0	13	11	0.7	0.5	10.1	13.8	
7 151	G	6	19	0	0	8.5	15.8	0.3	0.4	
150	C	2	0	2	21	0.3	0.2	1.1	23.3	
149	C	1	0	3	21	0.4	0.1	1.4	23.0	
148	A	19	2	2	2	23.8	0.6	0.2	0.4	
147	C	0	1	2	22	0.3	0.3	1.3	23.0	
146	C	1	1	1	22	0.5	0.3	1.7	22.4	

166,276 are full-length antisense intronic *Alus*

25 *Alu* exons

# 5'ss selection of *Alu* exons

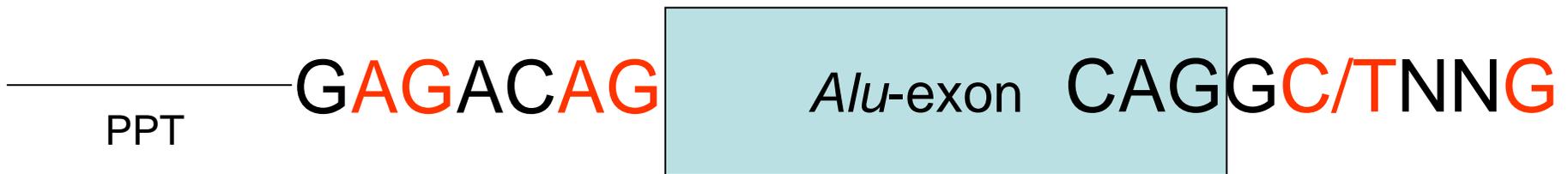
Gene name	Alu exon#	subfamily	Position relative to 5'ss																											
			exon									intron																		
			172	171	170	169	168	167	166	165	164	163	162	161	-3	-2	-1	1+	2+	3+	4+	5+	6+	7+						
1	MLANA	4	J	T	A	G	C	T	G	G	G	A	C	C	A	G	G	C	A	C	G	T	G	C	C	A	C	C		
2	RES4-22	18	S	T	A	G	T	T	G	G	A	A	T	T	A	C	A	G	G	C	A	A	G	C	A	A	A	A	C	C
3	HPK1*	31	S	T	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	C	A	T	G	T	G	C	C	A	C	C
4	ADAR2	8	J	G	A	G	C	T	G	G	G	A	C	T	A	C	A	G	G	C	A	T	G	T	A	C	C	A	C	T
5	KIAA1169	24	J	C	A	G	C	T	G	G	A	A	C	T	A	C	A	G	G	T	G	T	G	C	A	C	T	G	T	G
6	LOC51193	5	S	T	A	G	C	T	G	G	G	G	T	T	A	C	A	G	G	T	G	T	G	C	G	C	C	A	C	C
7	ZFX	2	S	T	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	T	G	T	G	C	G	C	C	C	C	C
8	MVK	4	J	T	A	G	C	T	G	G	G	A	C	C	A	C	A	G	G	T	G	T	G	C	G	C	C	A	C	C
9	PTGES	2	J	T	A	G	C	T	G	G	G	A	C	C	A	C	A	G	G	T	G	T	G	T	A	T	C	A	C	C
10	N/A	6	J	T	A	G	C	T	G	G	G	A	C	T	A	C	A	G	G	T	G	T	G	T	G	C	C	A	C	C
11	EVI5	3	J	T	G	G	C	T	G	G	G	A	C	T	A	C	A	G	G	T	G	T	G	T	G	C	C	A	T	C
12	C20orf26	9	J	T	A	G	C	T	G	G	G	A	C	T	A	T	A	G	G	T	A	T	G	T	G	C	C	A	C	C
13	N/A	4	S	T	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	T	G	T	G	T	G	C	C	A	C	C
14	BRCA2	20	S	T	A	G	C	T	G	G	G	A	C	T	A	C	A	G	G	T	G	C	G	T	G	C	C	A	C	C
15	CNN2	6	S	T	A	G	C	T	G	G	G	A	C	T	A	C	A	G	G	T	G	C	A	T	G	C	T	G	C	C
16	BIRC3	2	S	T	A	G	C	T	G	G	A	A	A	T	A	C	A	G	G	T	G	C	G	T	G	C	C	A	C	C
17	CYP3A43	8	S	G	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	T	A	C	A	C	A	C	A	C	C	C
18	N/A	2	S	T	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	T	G	C	C	C	G	C	C	A	C	C
19	MBD3	12	S	T	A	G	C	T	G	G	G	A	T	T	T	C	A	G	G	T	A	C	C	C	G	T	C	A	C	A
20	PLA2G4B	2	S	T	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	T	G	C	C	C	A	C	C	A	C	C
21	BCAS4	5	S	T	A	G	C	T	G	G	G	A	T	T	A	C	A	G	G	T	G	C	G	C	G	C	T	A	C	C
22	ICAM2	2	J	T	A	G	C	T	G	G	G	A	T	C	A	C	A	G	G	T	G	A	G	A	G	C	C	A	C	C
23	TGM4*	2	FAM	T	A	A	C	C	G	G	G	A	T	T	A	C	A	G	G	T	A	T	G	T	G	A	C	T	C	C
24	Integrin $\beta 1^*$	7	S	T	A	C	C	T	G	G	G	A	T	T	A	C	A	G	G	T	G	C	C	T	G	C	C	T	C	A
25	CHRNA3*	5	S	T	G	T	C	T	G	G	G	A	C	T	A	C	A	G	G	T	A	C	C	C	G	C	C	C	G	C
26	CTDP1*	7	Y	G	T	G	T	T	G	G	G	A	T	T	A	C	A	G	G	T	A	T	G	A	G	C	C	A	T	T

Fig. 1

# The 5' and 3' splice sites are under purifying selection in Alu exons.

The 3' splice site of *Alu*-exons

The 5' splice site of *Alu*-exons



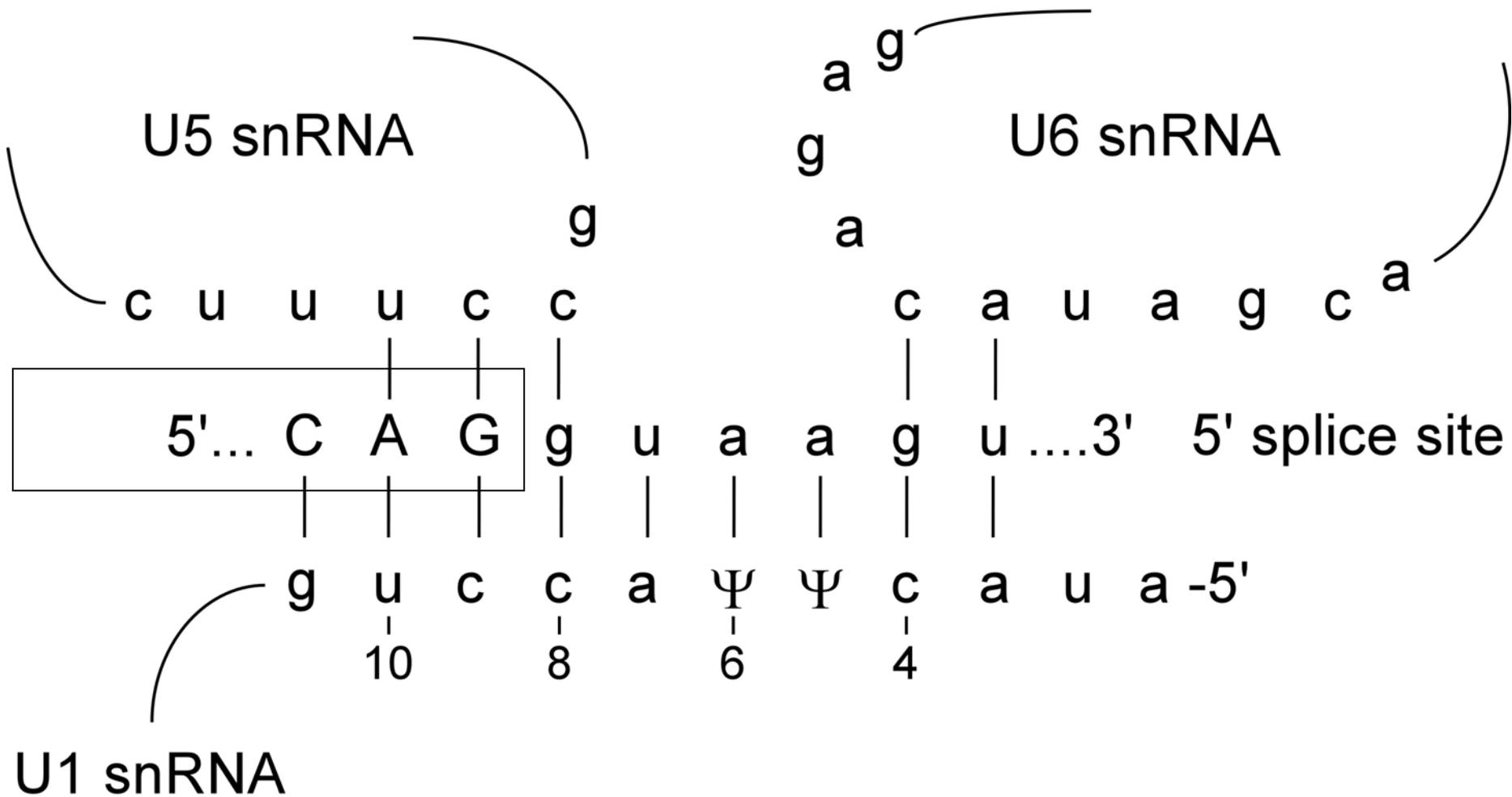
Lev-Maor et al., Science 2003

Sorek et al., Mol. Cell 2004

7,810 have ADAR2 like 3' and 5'ss, yet never exonized!

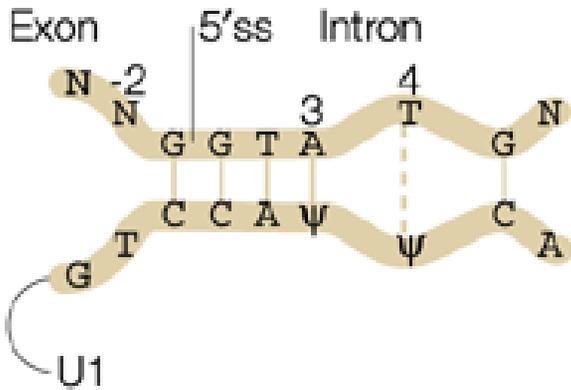
How many are one step from mutation leading to exonization – all!!!

# 5' splice site recognition during mRNA splicing

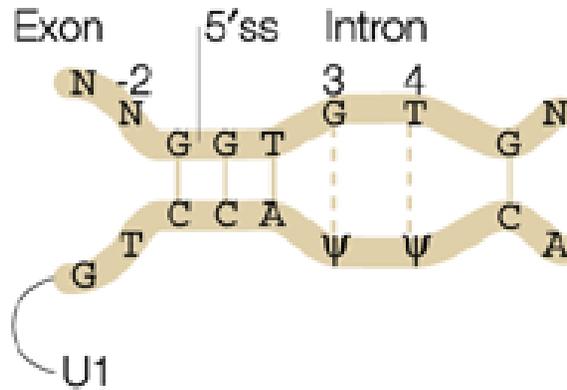


# The base pairing between U1 and the 5' splice site affect alternative splicing

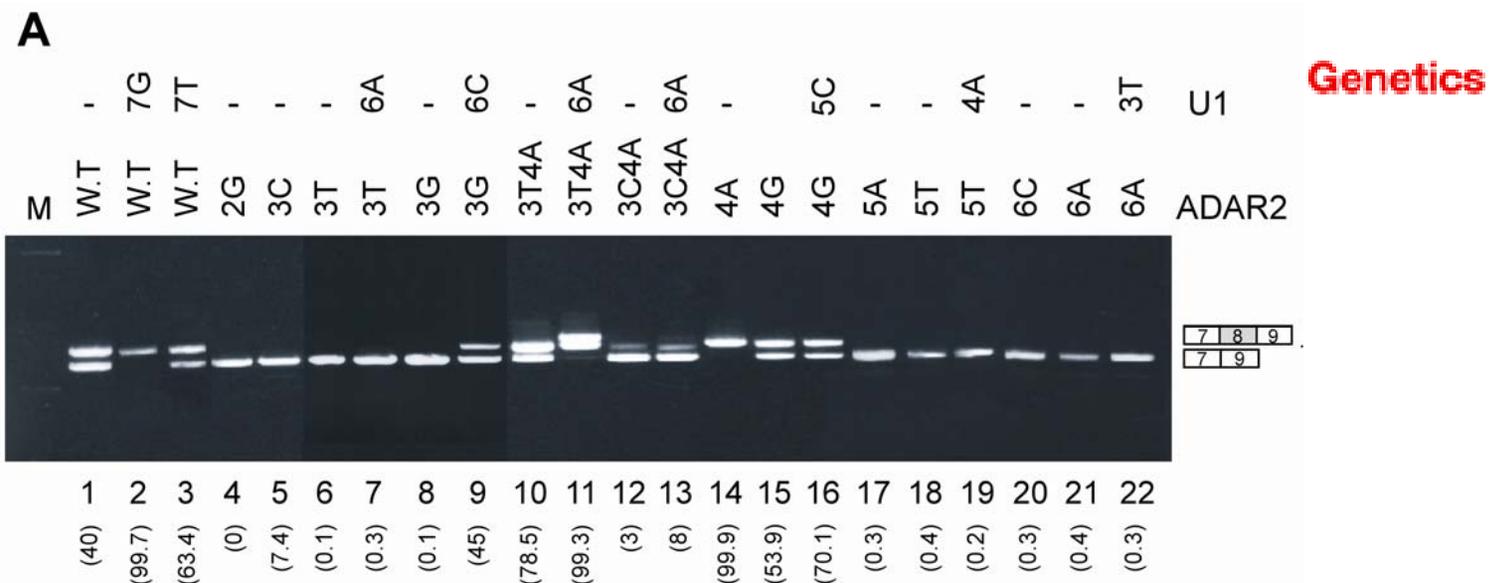
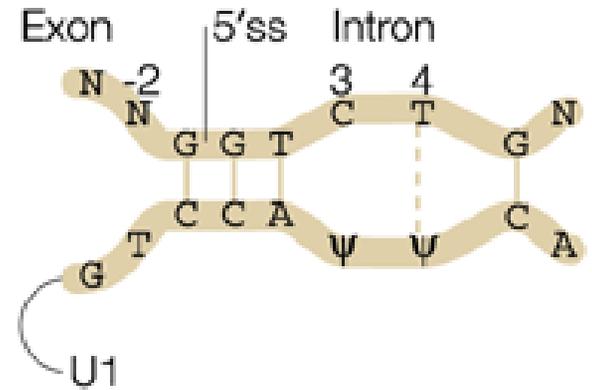
**a Mostly constitutive**



**b Mostly alternative**



**c Exon skipping**



# Comparative analysis of 50,000 human-mouse constitutively spliced exons

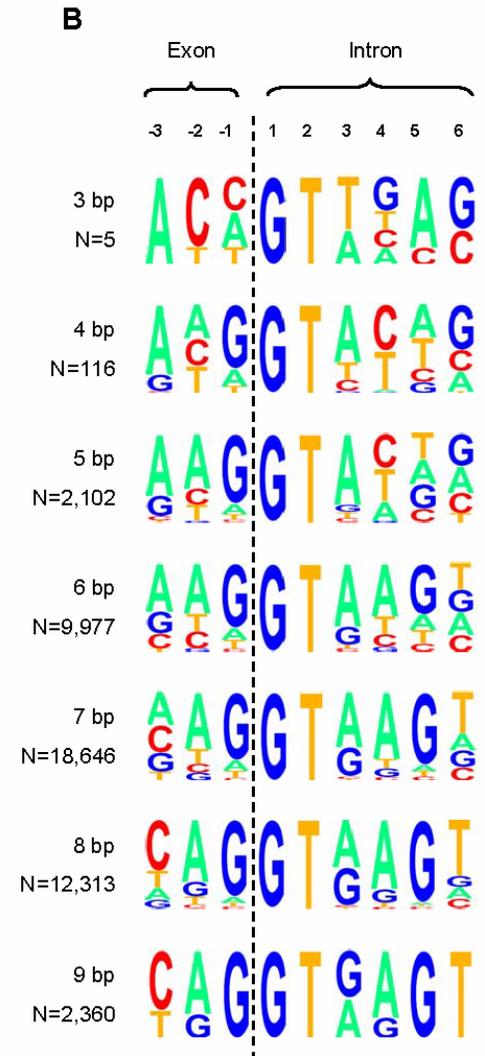
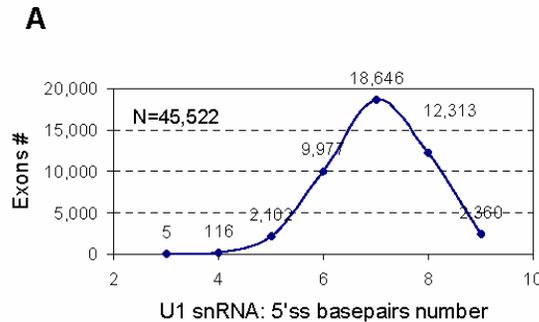
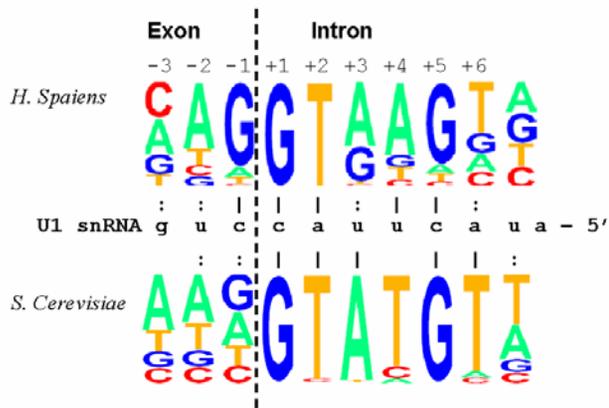
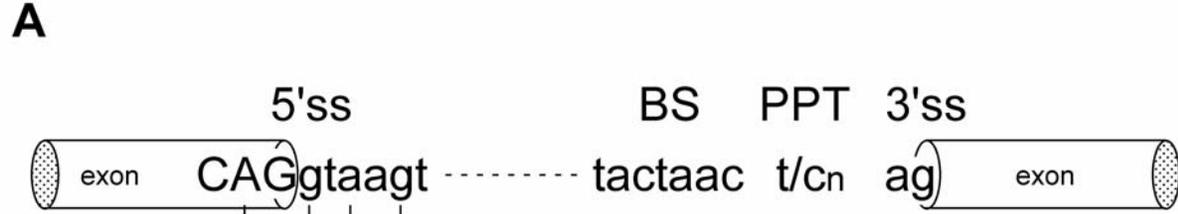


Fig. 1  
Carmel et al.

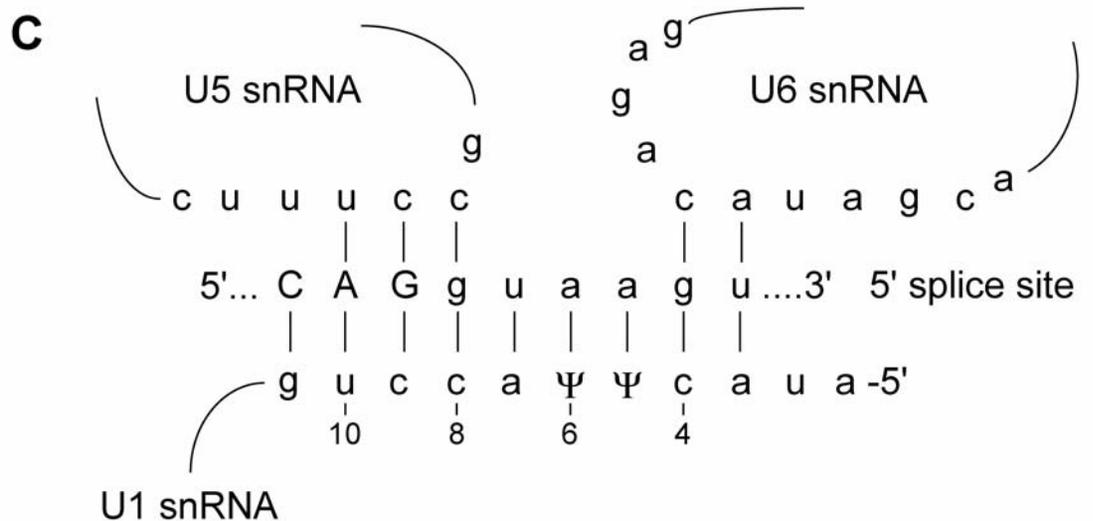
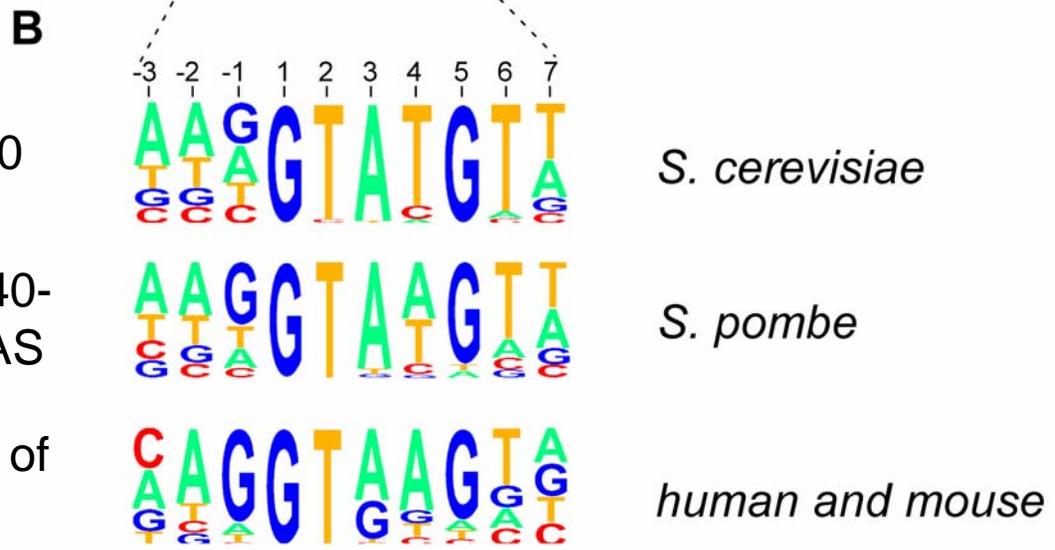
# Comparative analysis of 5'ss



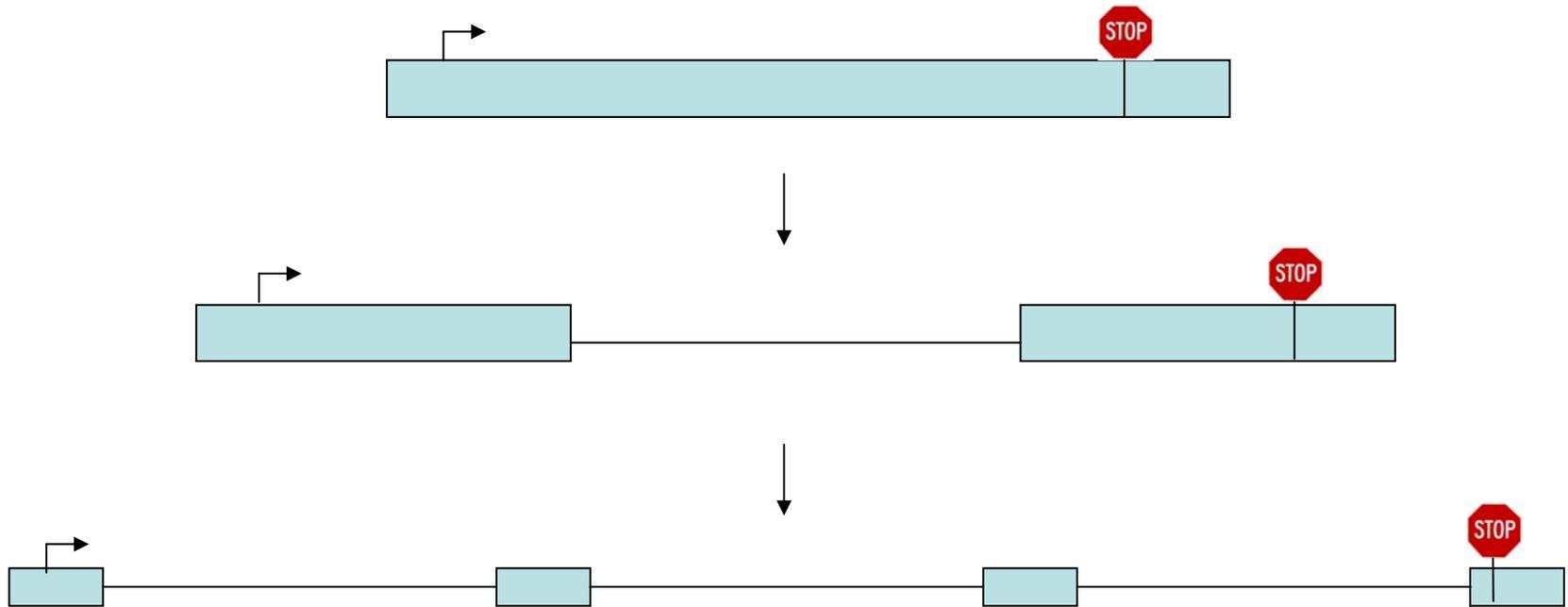
6,140 genes, 234 introns of 270 nt, no AS

4,999 genes, 4,697 introns of 40-70 nt, multi-introns genes, no AS

26,000 genes, 180,000 introns of 3,000 nt, 11 introns per gene, ~70% of the genes are alternatively spliced.



# How did alternative splicing evolve?



During the course of evolution:

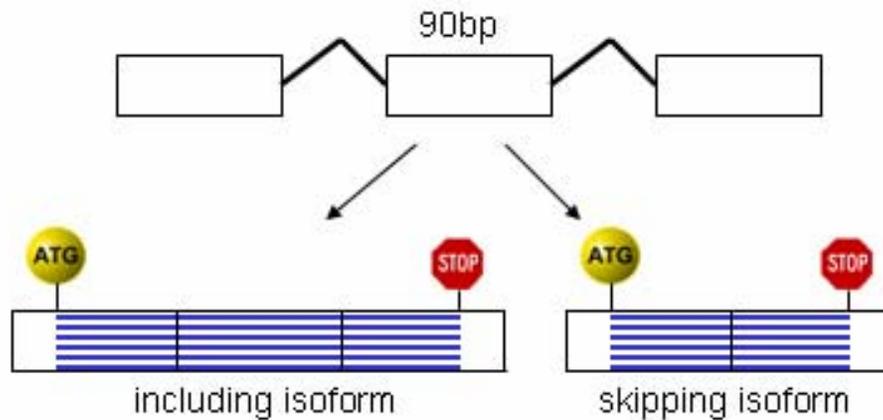
- Appearance of small exons embedded in large introns,
- appearance of exon definition
- and finally relaxation of 5' splice site selection
- allow for sub-optimal recognition of exons, leading to exon skipping.

# Differences between human-mouse conserved alternatively and constitutively spliced exons

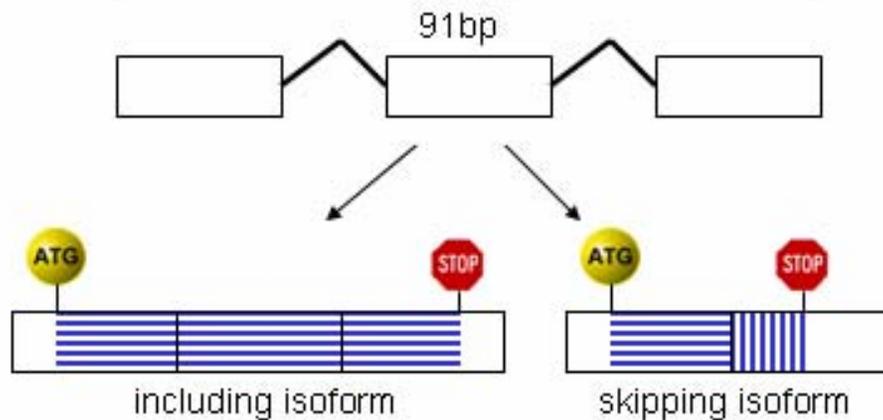
	Alternative	Constitutive
Length	109 nt	132 nt
Conservation	92%	88%
Conservation of flanking intronic sequences	100 nt	10-20 nt
5'ss strength (in kcal mol <sup>-1</sup> )	-5.8	-6.15
Divisibility-by-3	68%	40%

The effect of alternative splicing on the coding sequence.

**A Symmetrical alt. exon (divisible by 3)**

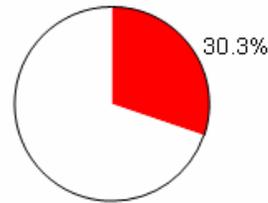


**B Non-symmetrical alt. exon (non-divisible by 3)**

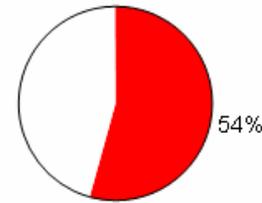


# Symmetrical alternative exons tend not to disrupt protein domain

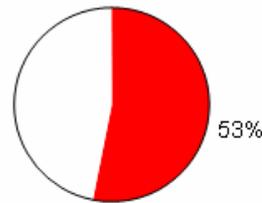
**A** Symmetrical alternative exons



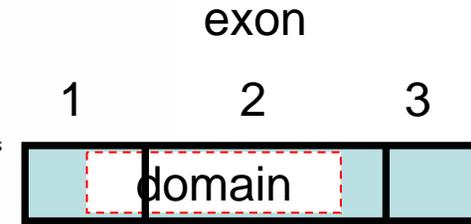
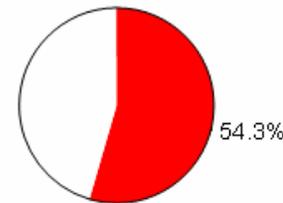
**C** Symmetrical constitutive exons



**B** Non-symmetrical alternative exons



**D** Non-symmetrical constitutive exons

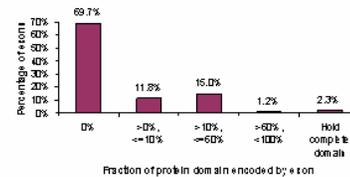


exon

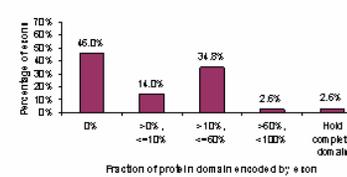


	0%	<10%	<20%	<30%	<40%	<50%	<60%	<70%	<80%	<90%	<100%	100%
Alternative exon (616 exons)	69.7%	11.2%	7.9%	4.3%	3.7%	3.2%	0.2%	0.6%	0.2%	0.0%	0.4%	2.3%
Non-const. (228 exons)	47.0%	11.0%	14.4%	9.7%	5.5%	4.2%	5.5%	1.3%	0.0%	1.4%	0.0%	0.0%
Constitutive exon (16800 exons)	46.0%	14.0%	16.1%	9.0%	4.5%	3.0%	2.2%	1.2%	0.6%	0.5%	0.3%	2.6%
Non-const. (22846 exons)	45.7%	14.2%	17.1%	8.6%	4.6%	3.4%	2.2%	1.3%	0.2%	0.5%	0.4%	0.6%

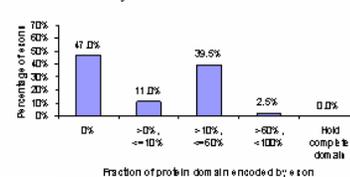
**B** Symmetrical alternative exons



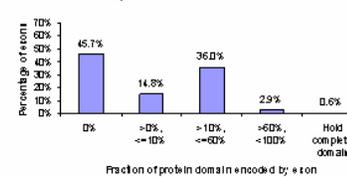
**D** Symmetrical constitutive exons



**C** Non-symmetrical alternative exons

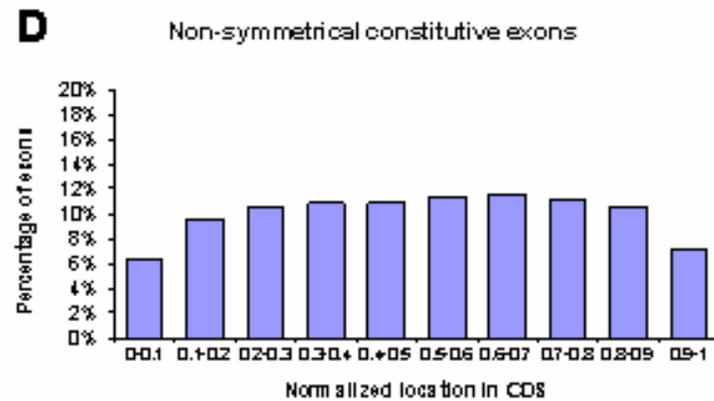
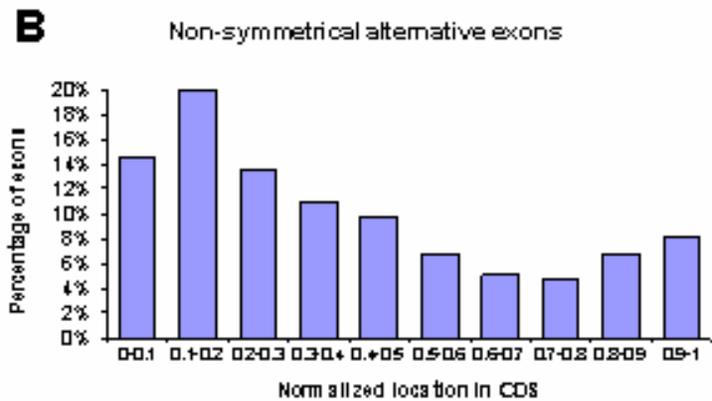
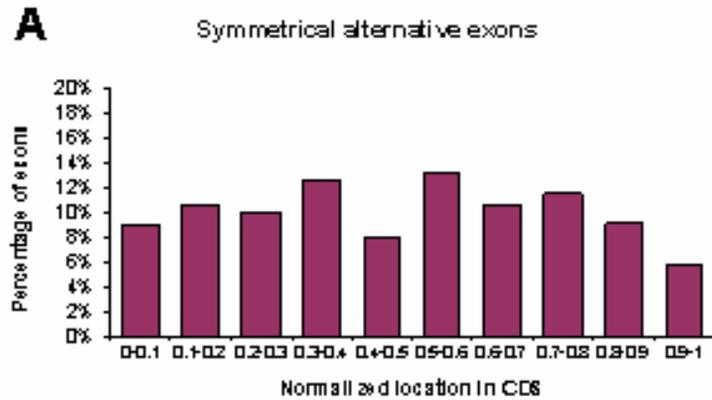


**E** Non-symmetrical constitutive exons



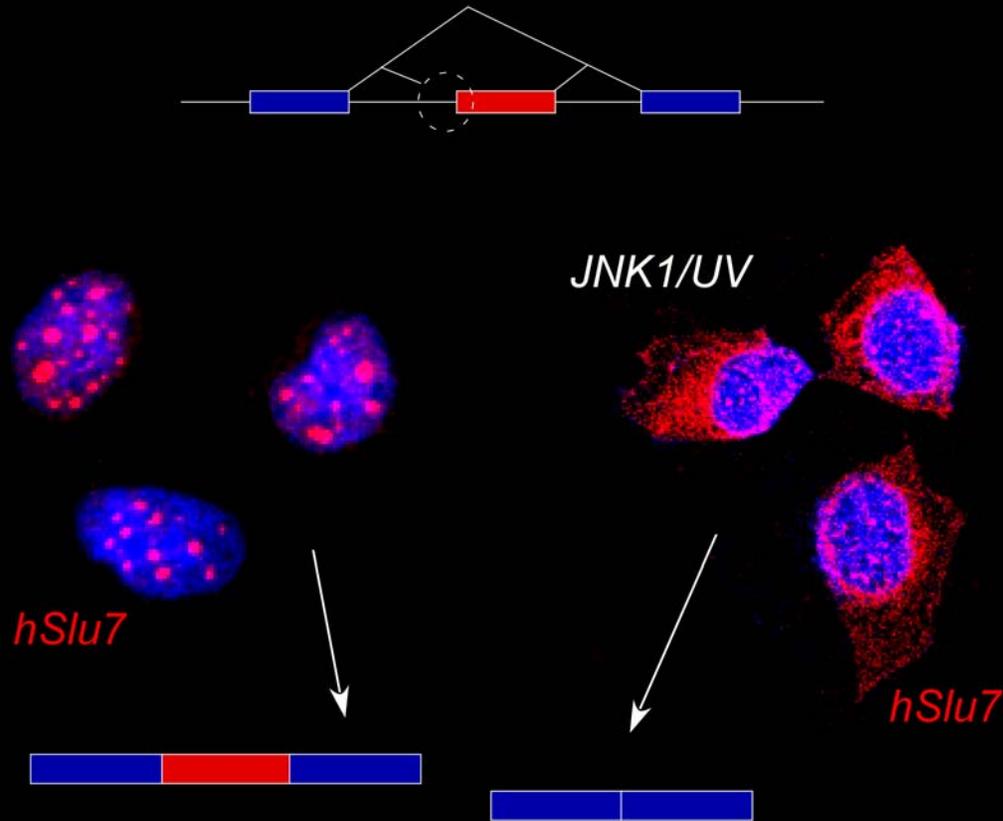
Less than 3% of alternatively and constitutively spliced exons are domain holders!

# Non-symmetrical alternative exons located at the 5' of the coding sequence



**Thus, alternative splicing might have originated from organisms that originally only support constitutive splicing by a relaxation of splice sites recognition.**

# hSlu7 shuttling from the nucleus to the cytoplasm affect alternative splicing



# Acknowledgments



Jun 2004



Oct. 2004

Ida Vig, Galit Lev-Maor, Rotem Sorek, Noam Shomron, Oren Ram, Ido Carmel, Norit Gal, Amir Goren, Noa Sela, Maayan Amit, Alon Magen, Moti Elberstein, Hadas Kerem, Eli Koren.