

Геном пуст ? o_o

10 ноября 16:00

Конференц-зал ИЦИГ

Публичная лекция Владимира Бабенко

«Сплайсинг,

или

куда исчезли 80 тысяч

предсказанных

генов человека»

Экзоны
1.5%



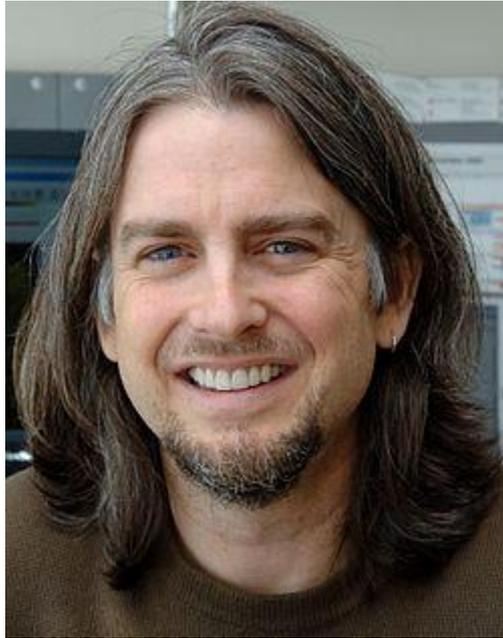
Интроны
(мусор)

Межгенные
участки
(мусор)

May 2000 Cold Spring Harbor Symposia on Genome sequencing

- So far, 165 bets have been placed by scientists in more than 50 countries. The mean prediction is 61,170 genes, with the lowest guess at 27,462 and the highest at **153,478** (www.genomebiology.com)
- Familiar people who have already placed bets include James Watson, Eric Lander, and Francis Collins. "[Craig] Venter was at the 2000 meeting, but he flew off to get the King Faisal International Prize for Science in Saudi Arabia, so he didn't bet," Stewart said. "He's due to show up here in June. I'm sure he can spare \$20."

EST assembling projects: TIGR (US) vs STACK (SA)



John Quackenbush vs Winston Hide

Число генов?

- **Gene Index analysis of the human genome estimates approximately 120,000 genes**
- Feng Liang¹, Ingeborg Holt¹, Geo Pertea¹, Svetlana Karamycheva¹, Steven L. Salzberg¹ & John Quackenbush¹
- Nature Genetics 25, 239 - 240 (2000)



the human genome

February 2001

.....

Nuclear fission

Five-dimensional
energy landscapes

.....

Seafloor spreading

The view from under
the Arctic ice

.....

Career prospects

Новые факты о геномах

- 60% транскрибируемой функциональной РНК в геноме человека – **белок – некодирующая**
- Исключительно насыщенный альтернативный сплайсинг
- миРНК – новый регулятор экспрессии
- Регуляторные функции интронов
- Эпигенетика

Gene number



C. elegans

~



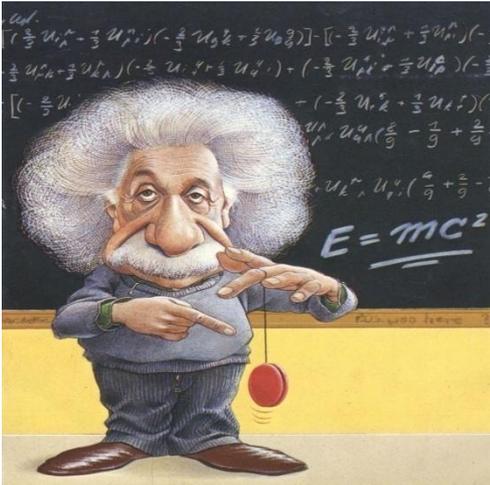
Drosophila melanogaster

~14,100



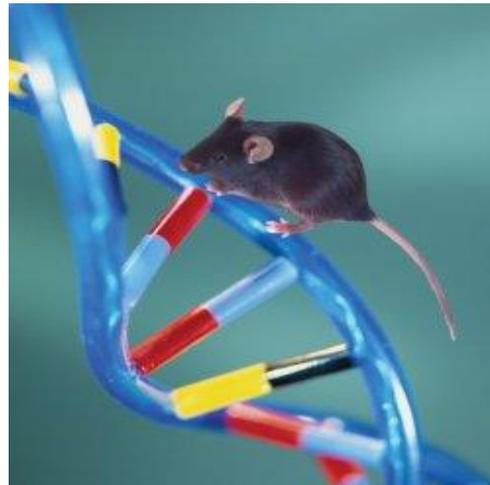
Arabidopsis thaliana

~



Homo Sapiens

~24,000



Mouse

~24,000



Rice (Oryza sativa)

~50,000

Гипотеза «один ген – один продукт» приказала «долго жить»

- Более **99%** мультиэкзонных генов человека производят **более 1 продукта**
- Исследование протеома (2011) с помощью высокоразрешающего метода показало, что на **1000** генов (локусов) производится **3000** белковых вариантов, с учетом посттранскрипционной модификации, альтернативного сплайсинга и протеолиза.

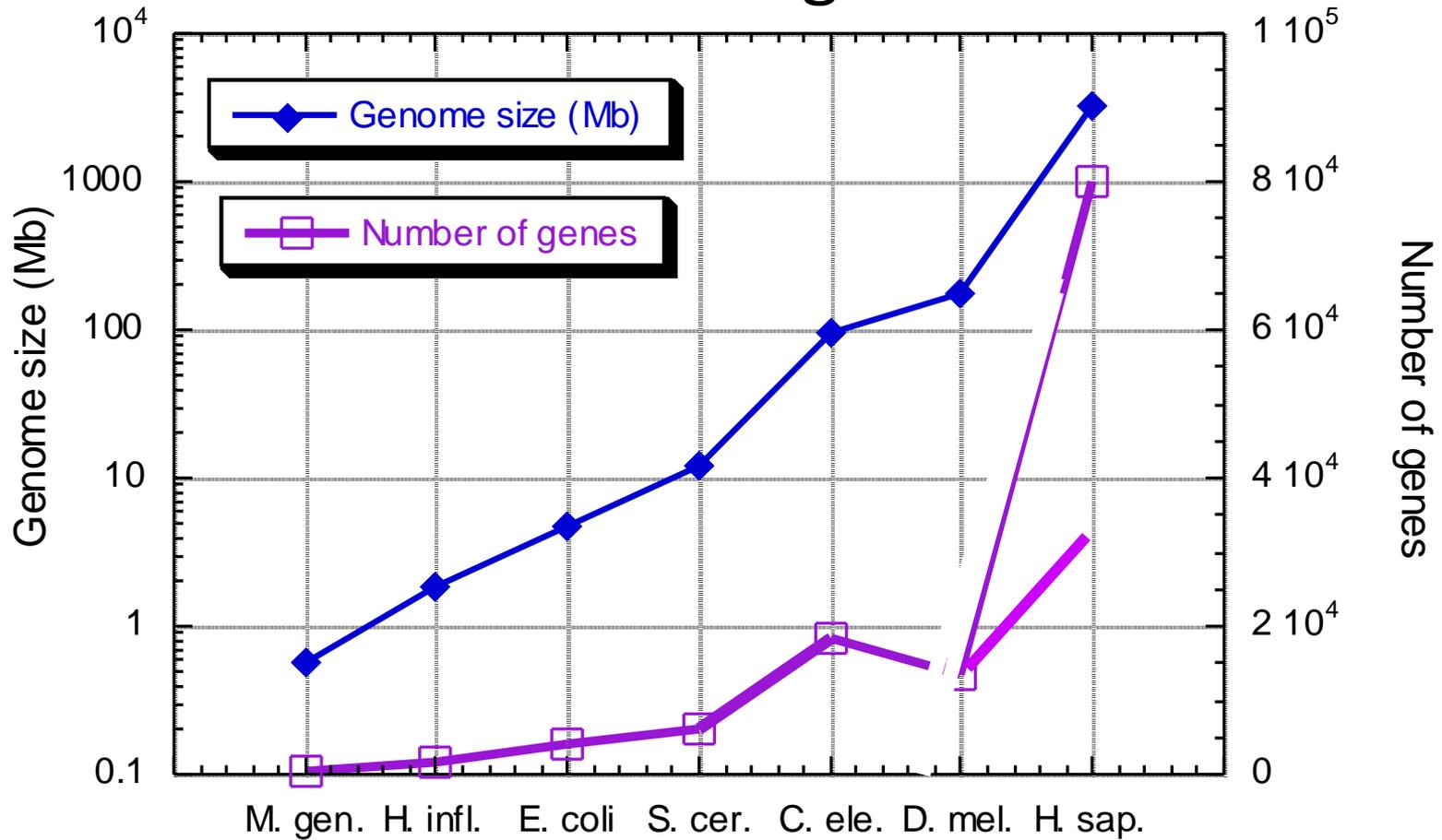
Основные темы доклада

- Эволюция экзон – интронной структуры.
- Альтернативный сплайсинг – режим эволюции генома
- Сплайсосома – сложный многоступенчатый каскад рибонуклеопротеинов
- Феномен сплайсинга на примерах.
- Хроматин и альтернативный сплайсинг.

Эволюция экзон – интронной структуры гена

- Размер экзонов (интронов)
- Число экзонов (интронов)
- Позиция интронов (экзонов)

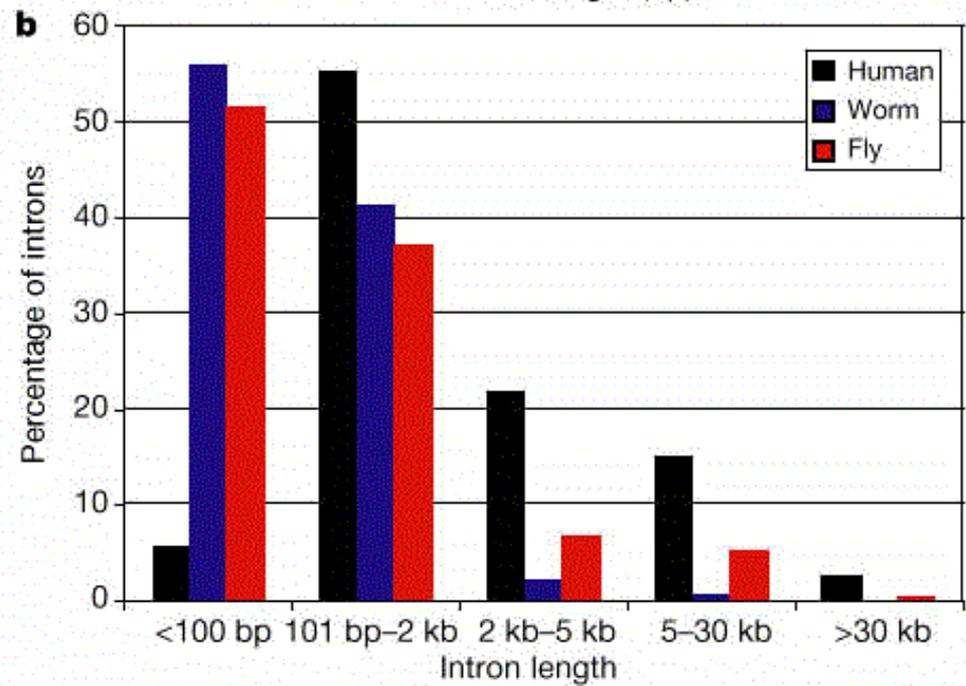
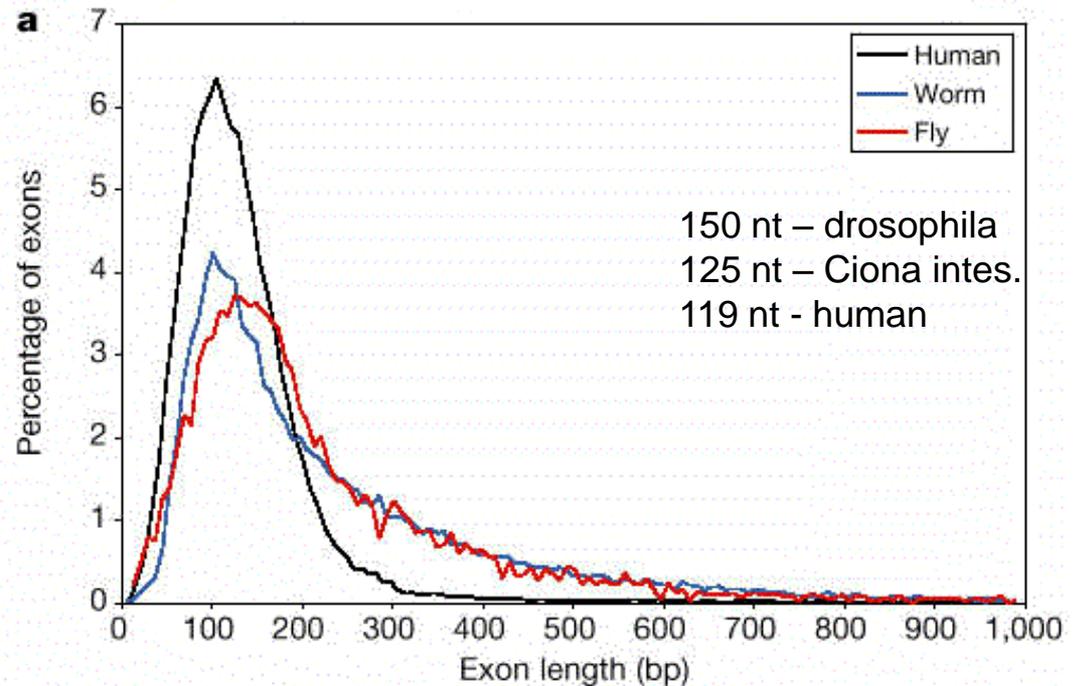
Genome size increases exponentially, but not number of genes



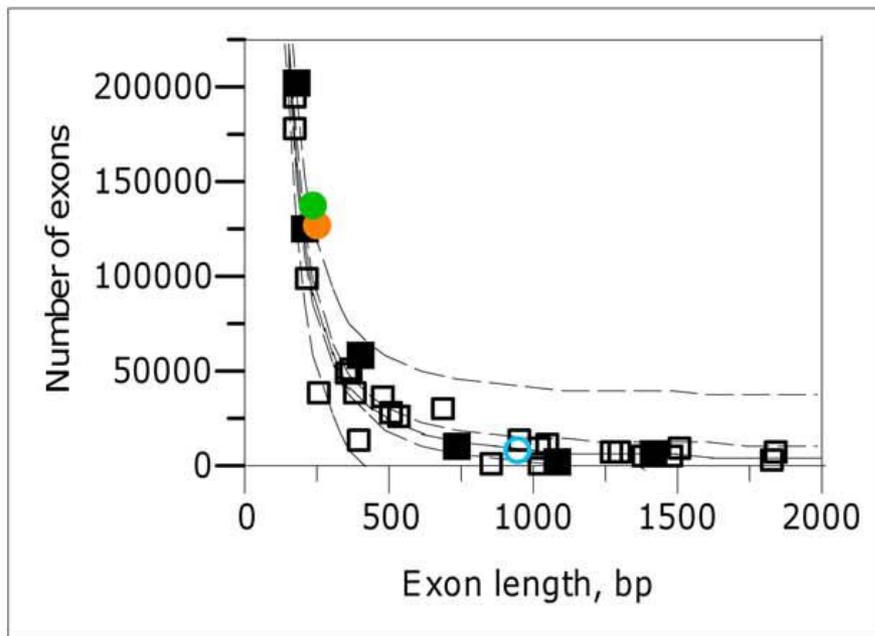
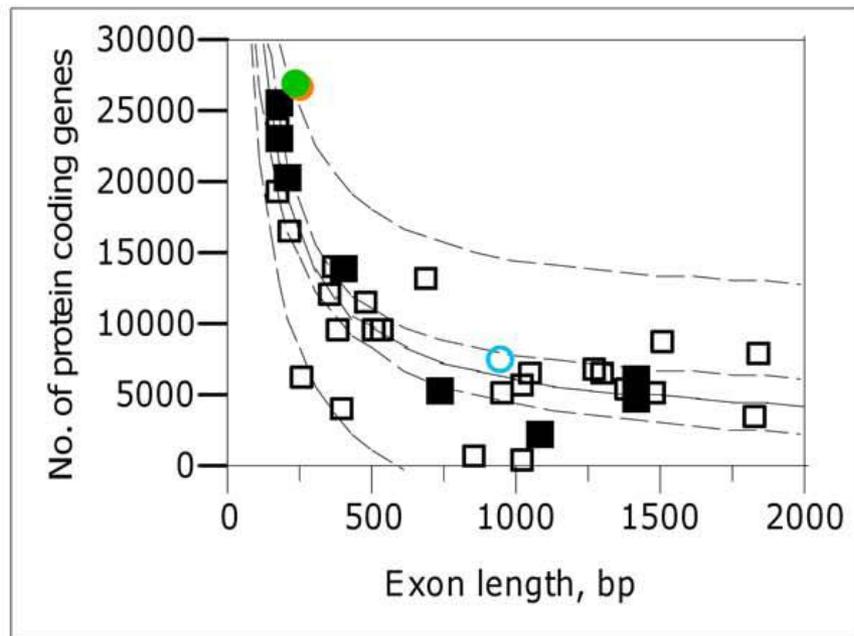
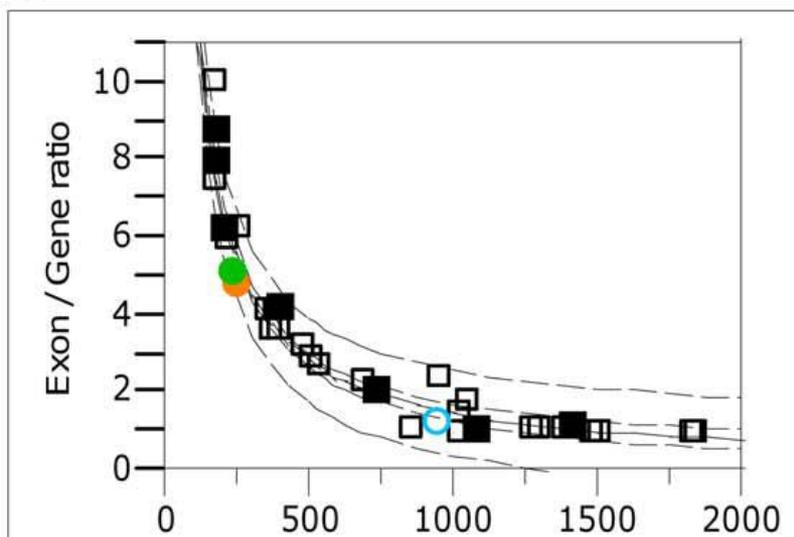
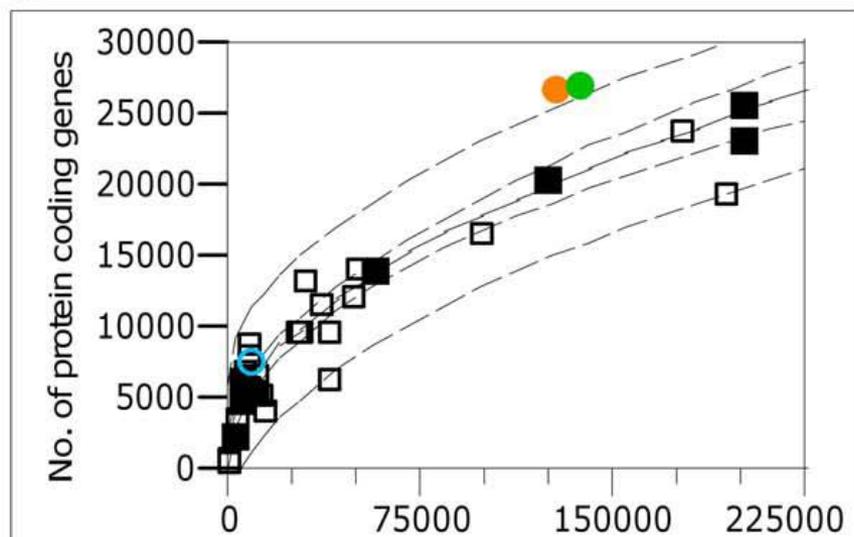
Drosophila – 14,000
 Nematodes – 19,000
 Mouse – 30,000

Species
 Ciona – 16,000
 Pufferfish – 31,000
 Human – 26,000 + ~12,000 RNA

Compared to worm and fly, human has shorter exons and longer introns on the extremes of the distribution



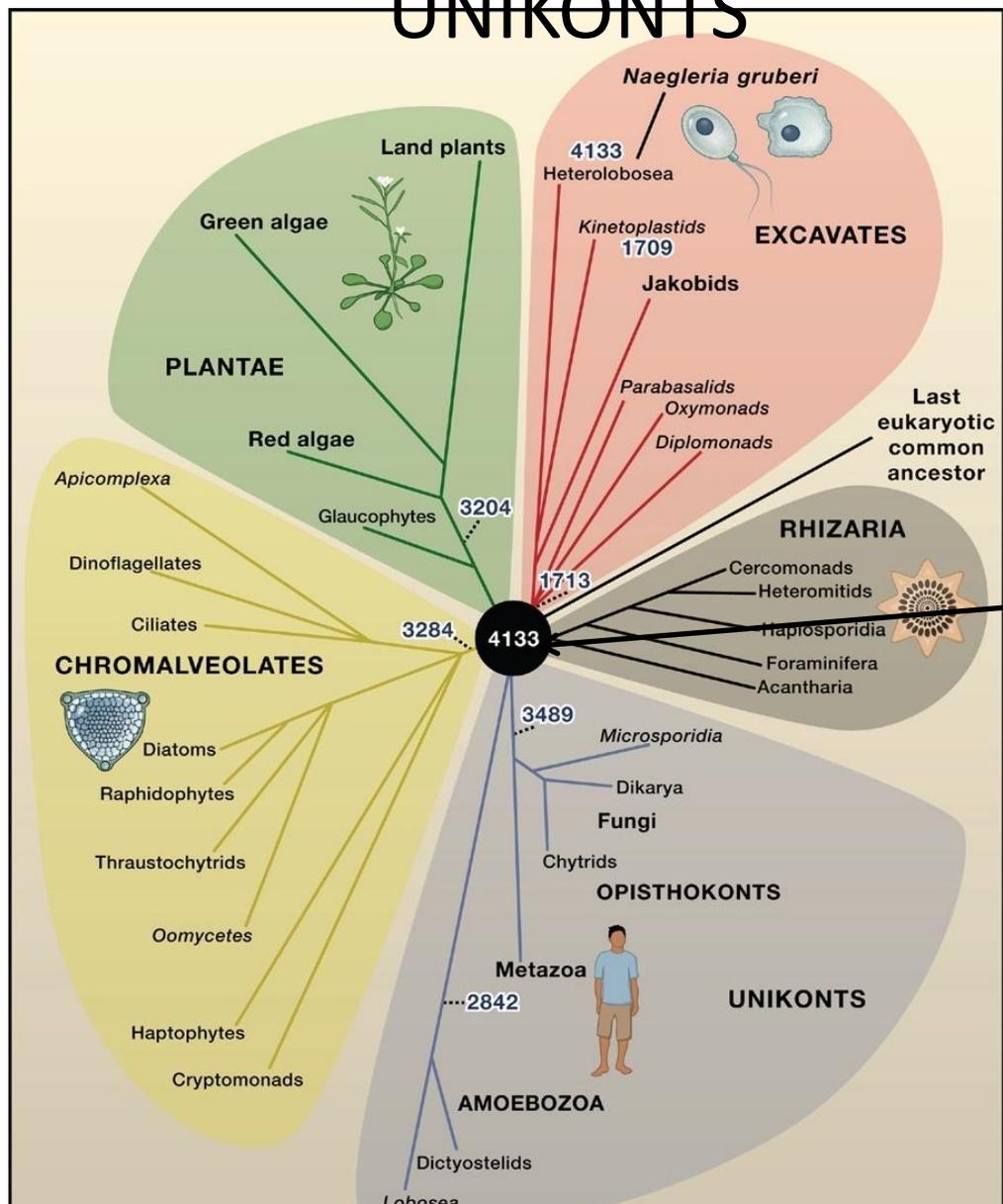
Обратная корреляция длины экзона и числа генов

A**B****C****D**

Парадигма интронов жива до настоящего времени

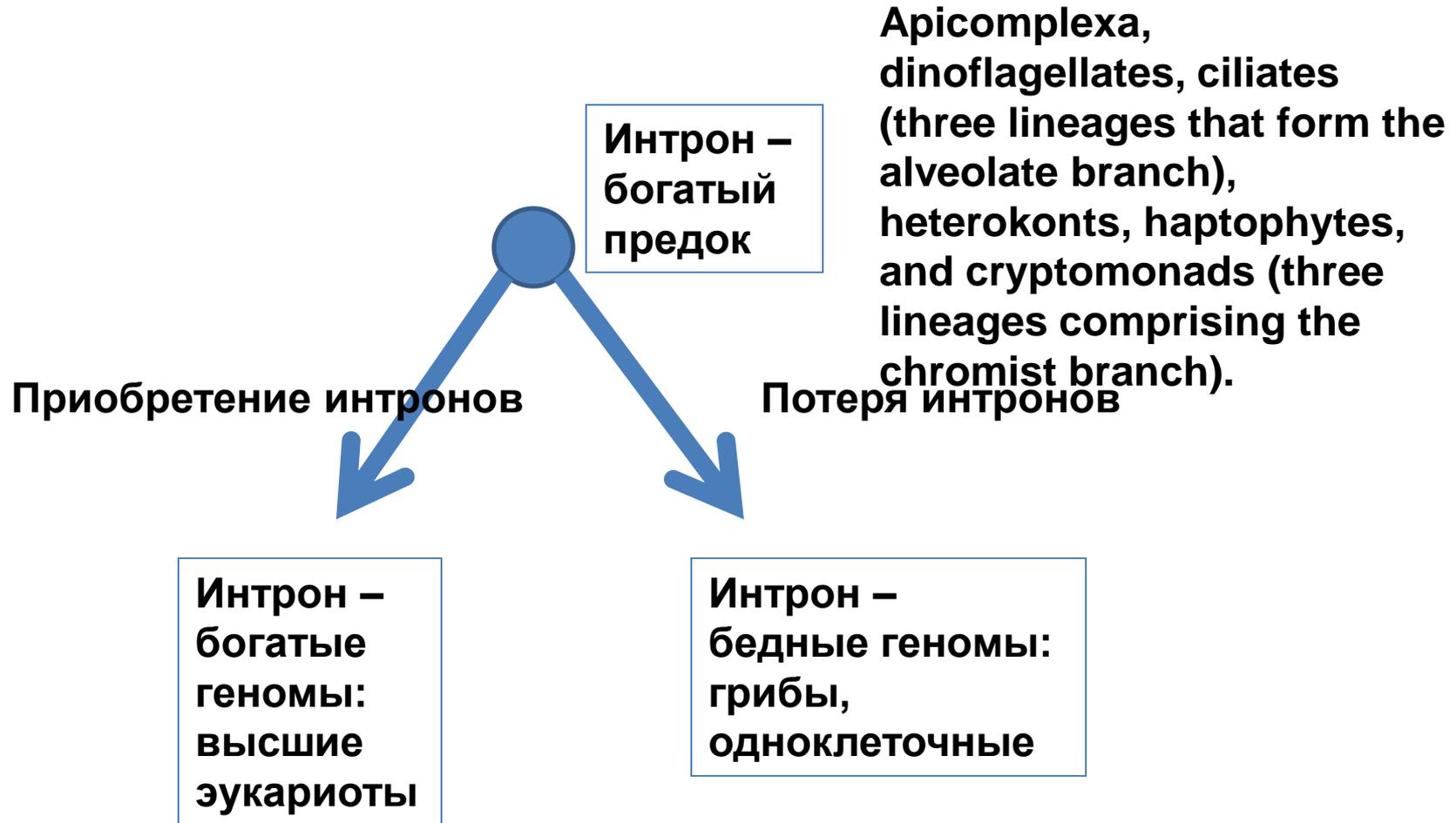
- Nature 1978 W. Gilbert “Why genes in pieces?”

5 эукариотических супергрупп: PLANTAE, EXCAVATES, RHIZARIA, CHROMALVEOLATES, UNIKONTS



Предок имел
4133 интрона

«Парадокс» эволюции количества интронов в геномах эукариот



Сплайсинг

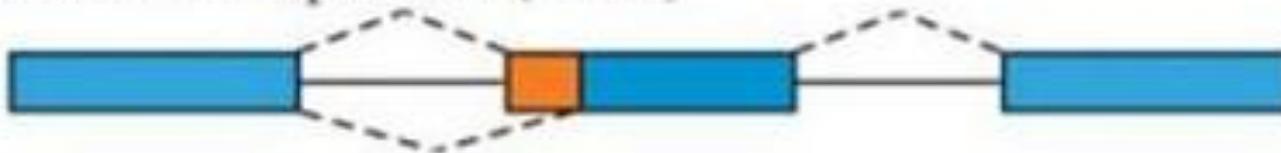
- Процесс формирования структуры зрелой мРНК путем вырезания сегментов РНК.

Основные типы альтернативного сплайсинга

exon skipping (38.4%)



alternative acceptor site (18.4%)



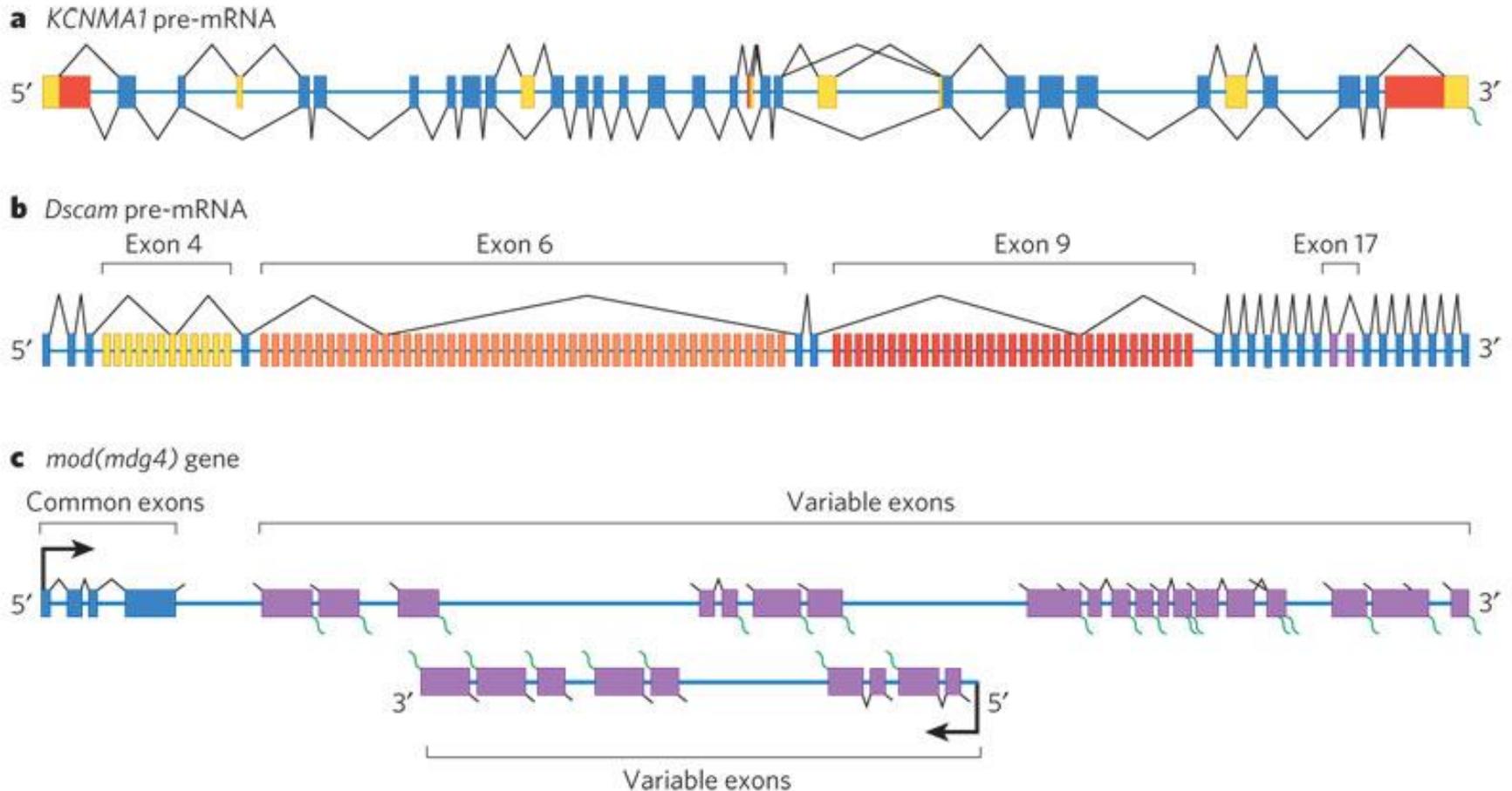
alternative donor site (7.9%)

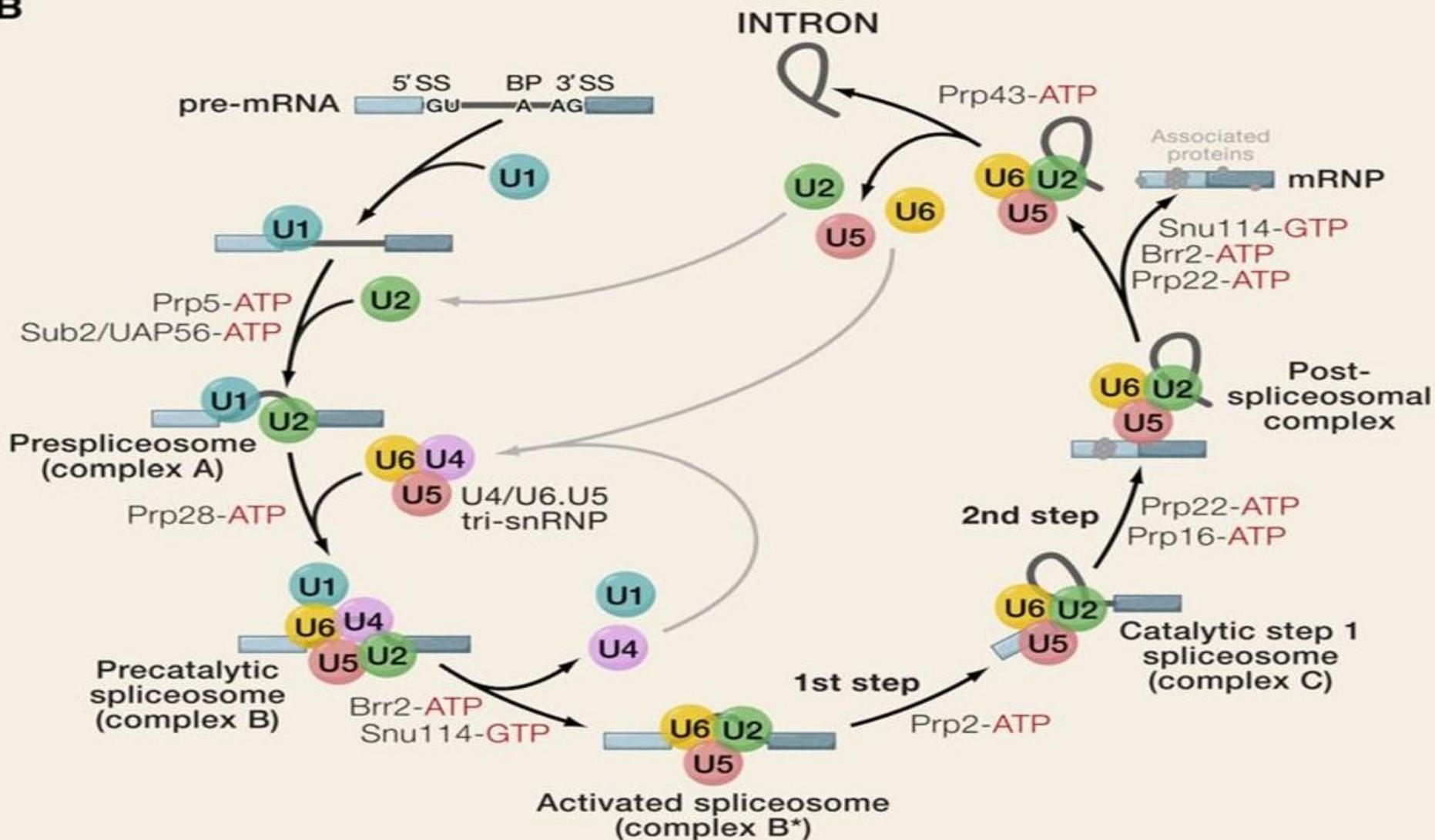


intron retention (2.8%)



Классические примеры альтернативного сплайсинга



A**B**

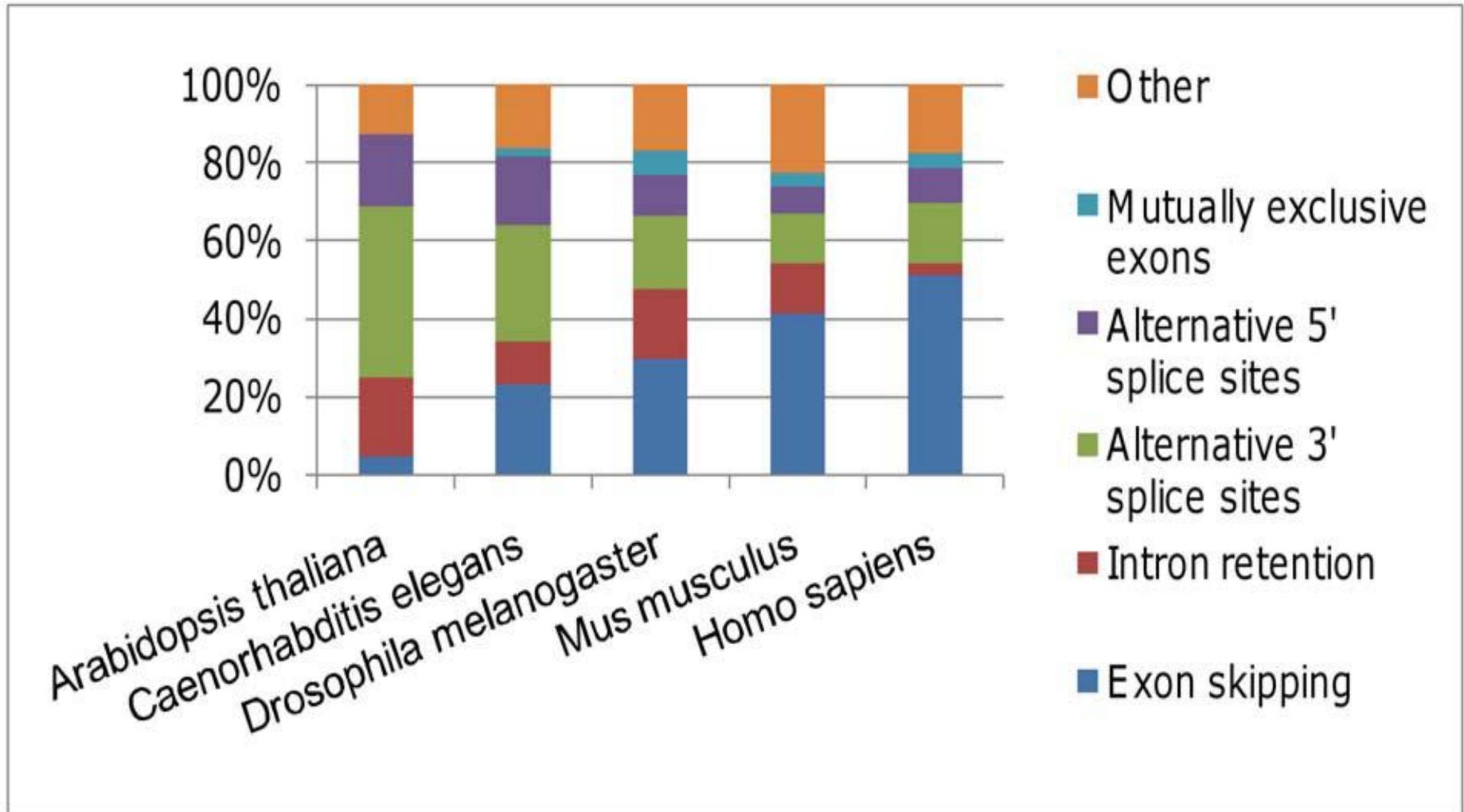
Эволюционные аспекты сплайсинга

- Существует во всех организмах, но развит в многоклеточных
- Связан с тканеспецифичностью, в частности с нейрогенезом.
- **Эволюция не количеством (генные дубликации), но качеством (вариации транскриптов).**
- Эффект альтернативного транскрипта различается по клеточной локализации (мембранный, растворимый), функции

Режим эволюции интронов - сплайсинга

- Nonsense mediated decay – широко используемый способ регуляции экспрессии

Relative frequency of alternative splicing (AS) types in five species.



Сплайсинг: прерогатива эукариот? Нет

- Ген CBF5 (дискерин в *H.S.*) в геномах **архей** **содержит интроны**: *Aeropyrum pernix*, *Sulfolobus solfataricus*, *S. tokodaii*, and *S. acidocaldarius*
- тРНК гены в прокариотах

! Сплайсинг катализируется с помощью белков и (ИЛИ) U мяРНК

Сплайсинг(интроны) – это:

- **Источник разнообразия белков (2-200 раз)**
- **Источник регуляции экспрессии, в т.ч. стадия – и тканеспецифичной**
- **Источник ко-регуляции генов с помощью интронных миРНК**
- **Возможность регулировать хроматиновую структуру ДНК районов генов**

Сплайсинг – это

- Источник аллель-специфического полиморфизма
- Источник иммунной вариации (в *D.melanogaster*)
- Активация мРНК теломеразы
- Транслятор жизненно важной информации, содержащейся в интронах

Использование генно-инженерных конструкций сплайсинга для

- Замещения мутантных аллелей на аллель дикого типа
- Формирования трансгенных конструкций (транс – сплайсинг, формирования химер)
- Основа для 63 патентов в области биотехнологии на основе энзиматических конструкций в последние 2 года.

Трудности изучения:

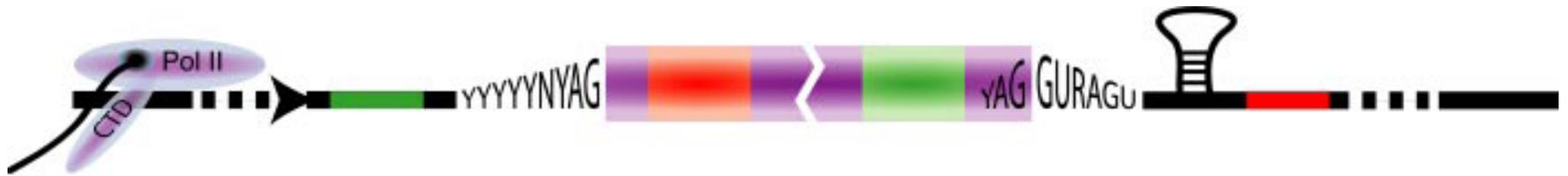
- Сплайсосома – рибонуклеопротеин, включающий в себя 5 мРНК и более 150 белков
- Чрезвычайно сложная и многоступенчатая кинетика катализа
- Вырожденность сигналов сайтов сплайсинга

Факторы, влияющие на сплайсинг (как альтернативный, так и конститутивный)

Вторичная структура мРНК в 5' и 3' районах интрона;

- Один из вариантов альтернативного сплайсинга – **кассетный**: соединение «докерного» экзона с только одним экзоном из кассеты («акцептор») путем формирования вторичной структуры

Сайты связывания факторов: энхансеры и сайленсеры (интронные ISE (ISS) и экзонные (ESE (ESS)))



Splice-site strength

YYYYYNYAG 3' ss
YAG GURAGU 5' ss

Binding sites for exonic splicing regulators

 Exonic Splicing Enhancer (ESE)
 Exonic Splicing Silencer (ESS)

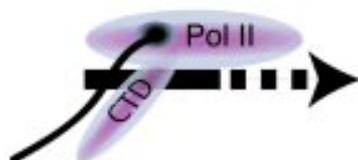
Exon/intron architecture

 variable exon length
 variable intron length

Binding sites for intronic splicing regulators

 Intronic Splicing Enhancer (ISE)
 Intronic Splicing Silencer (ISS)

Pre-mRNA synthesis



Local secondary structures



**Сплайсинг – критический
процесс: мутация, нарушающая
сайт сплайсинга ведет к
развитию заболевания**

Survival Motor Neuron gene:

пропуск 7 экзона ведет к

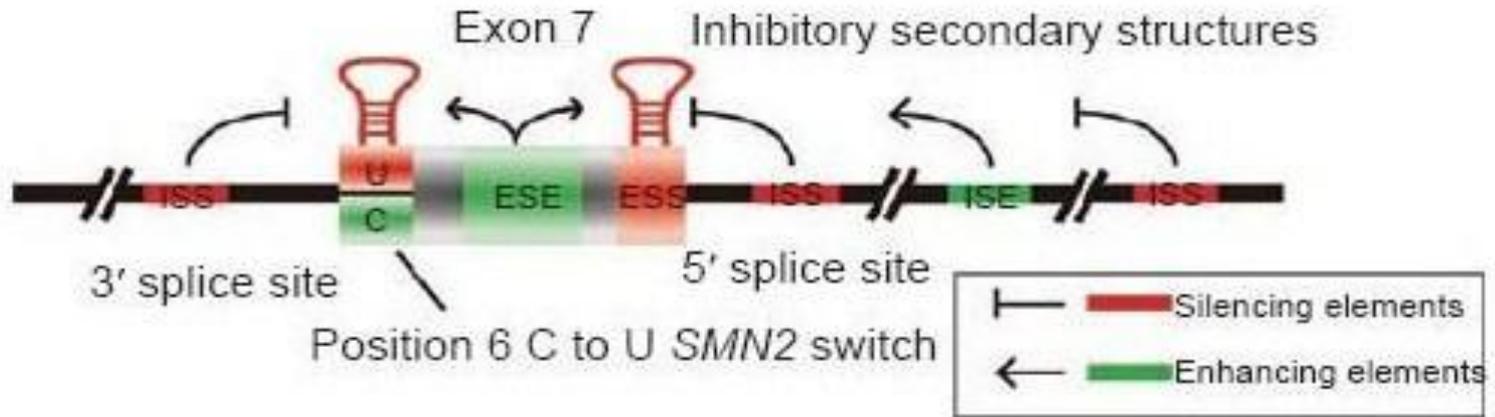
Spinal Muscular Atrophy; пропуск
экзона определяется мутацией

C->U

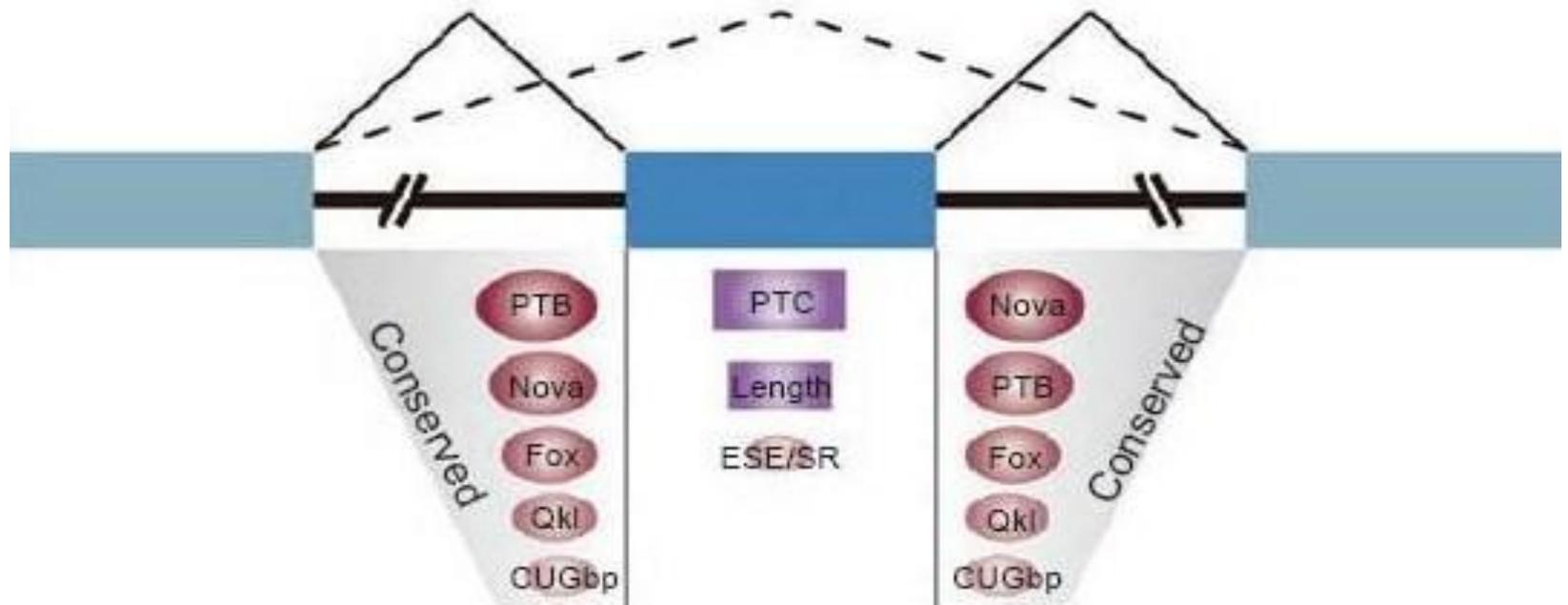
Энхансеры и сайленсеры альтернативного сплайсинга

Факторы регуляции, известные на текущий момент, на примере экзона 7 гена SMN

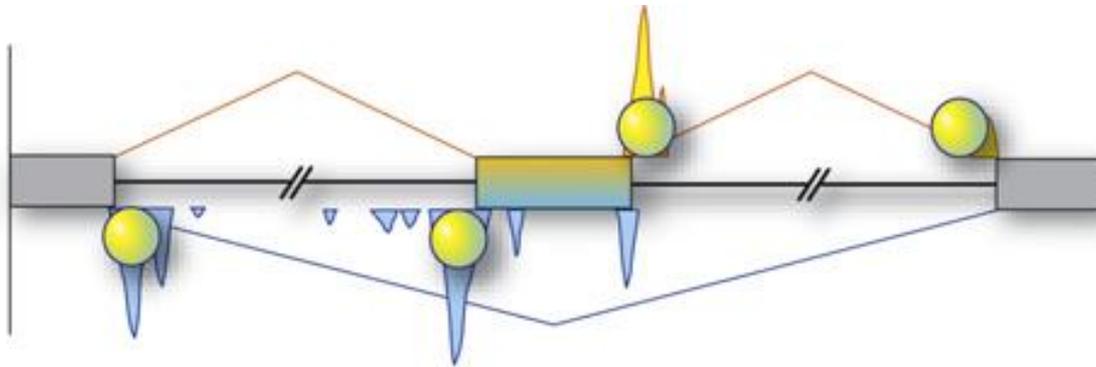
A



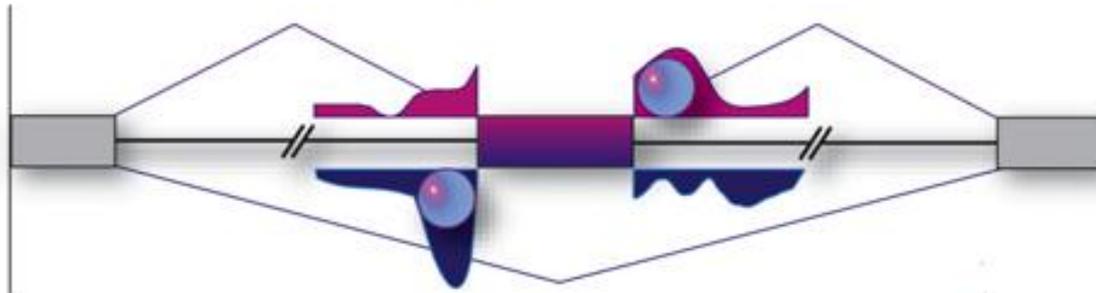
B



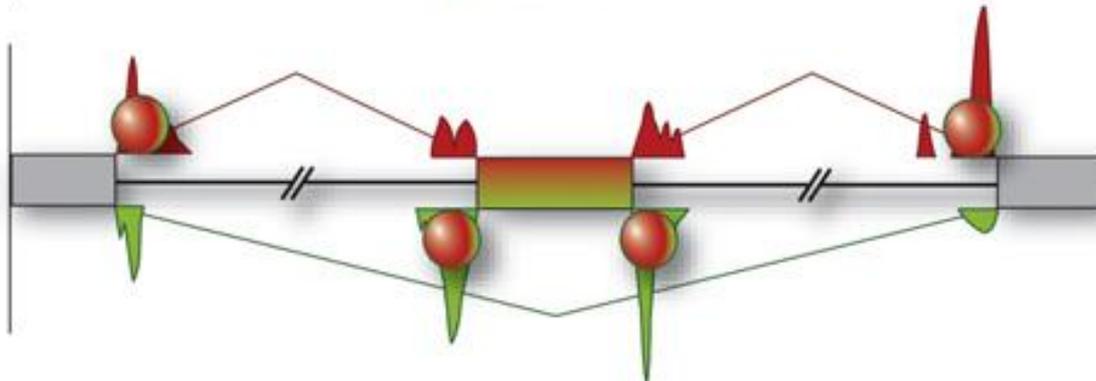
RNA Maps of Three Splicing Regulators



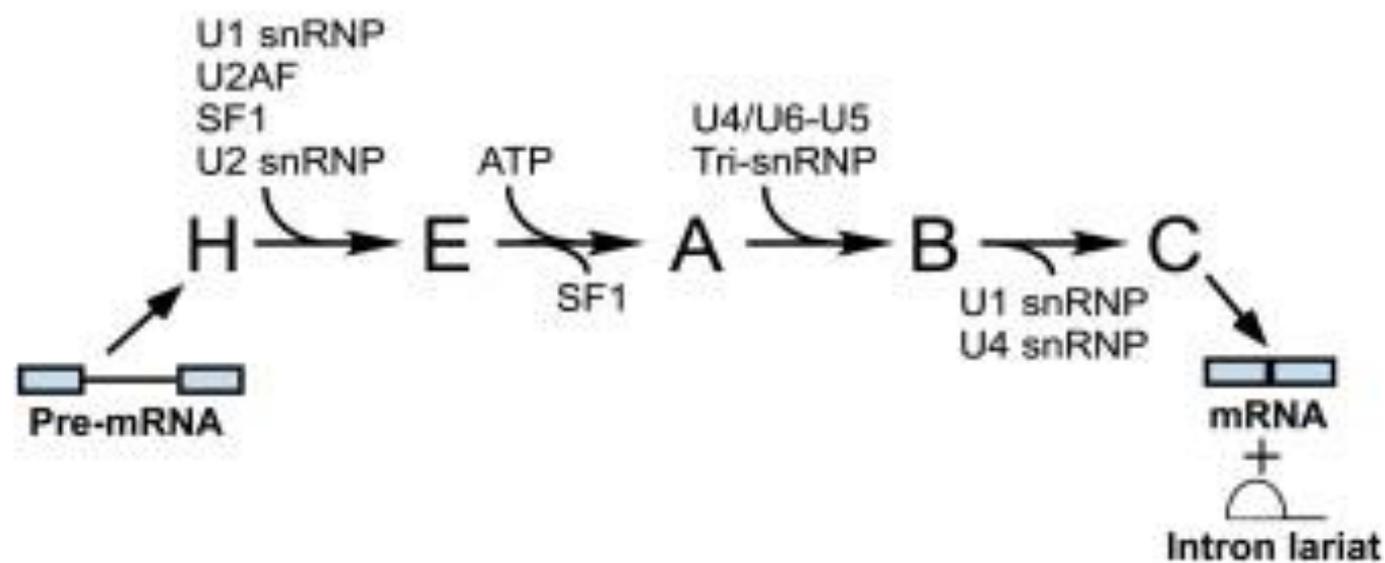
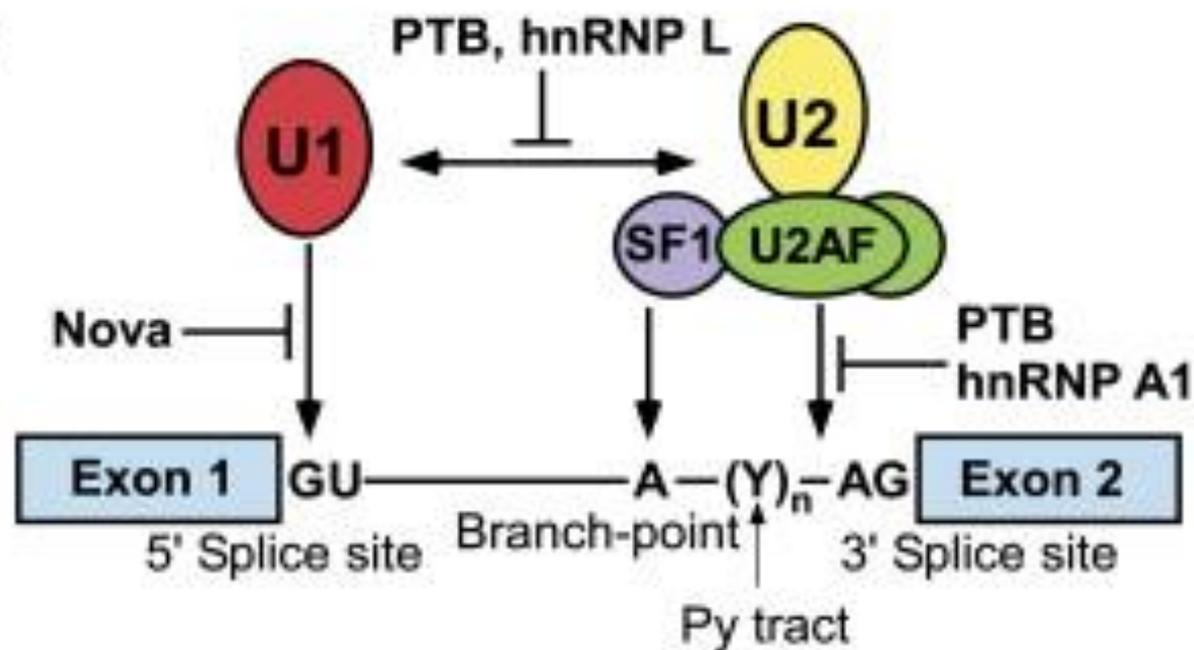
(A) Nova RNA map



(B) Fox2 RNA map.

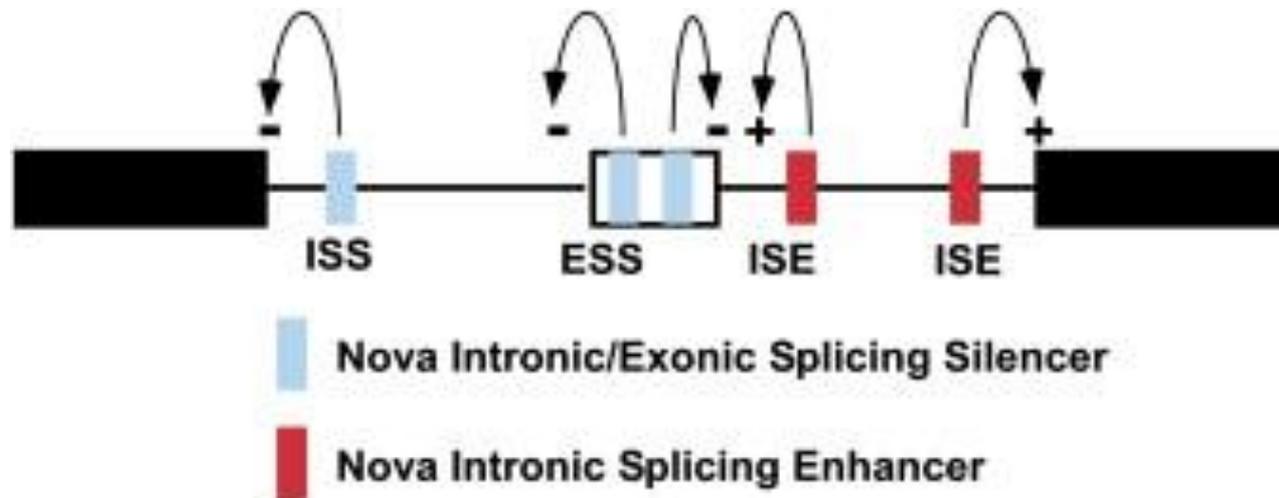


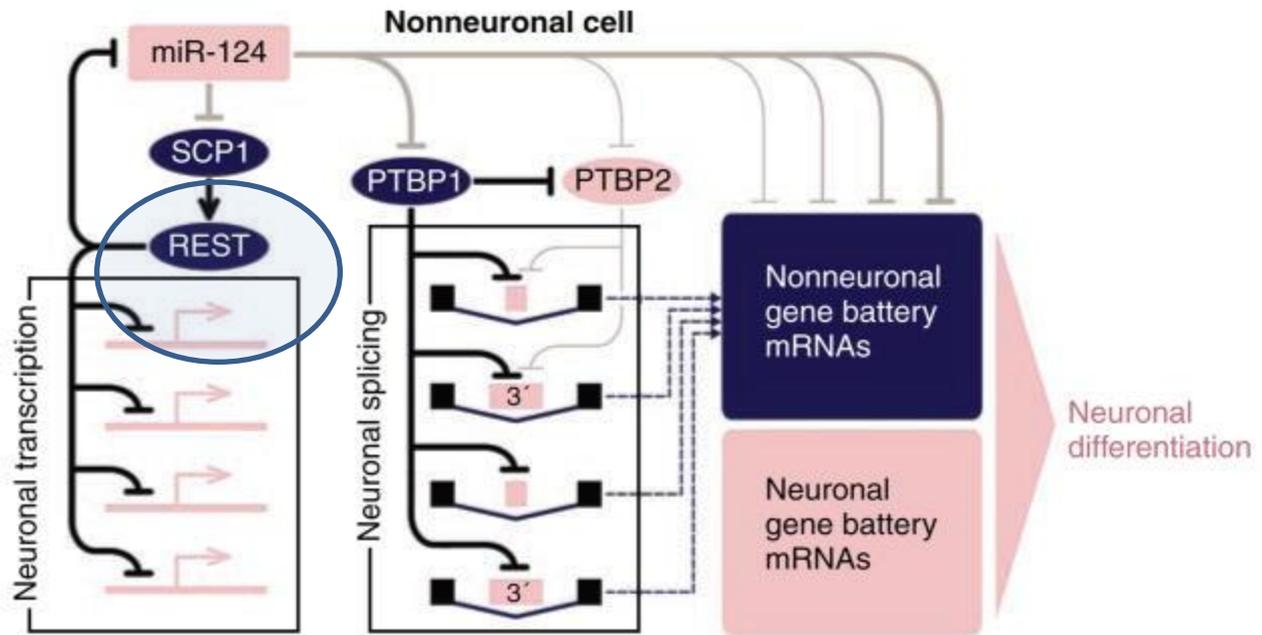
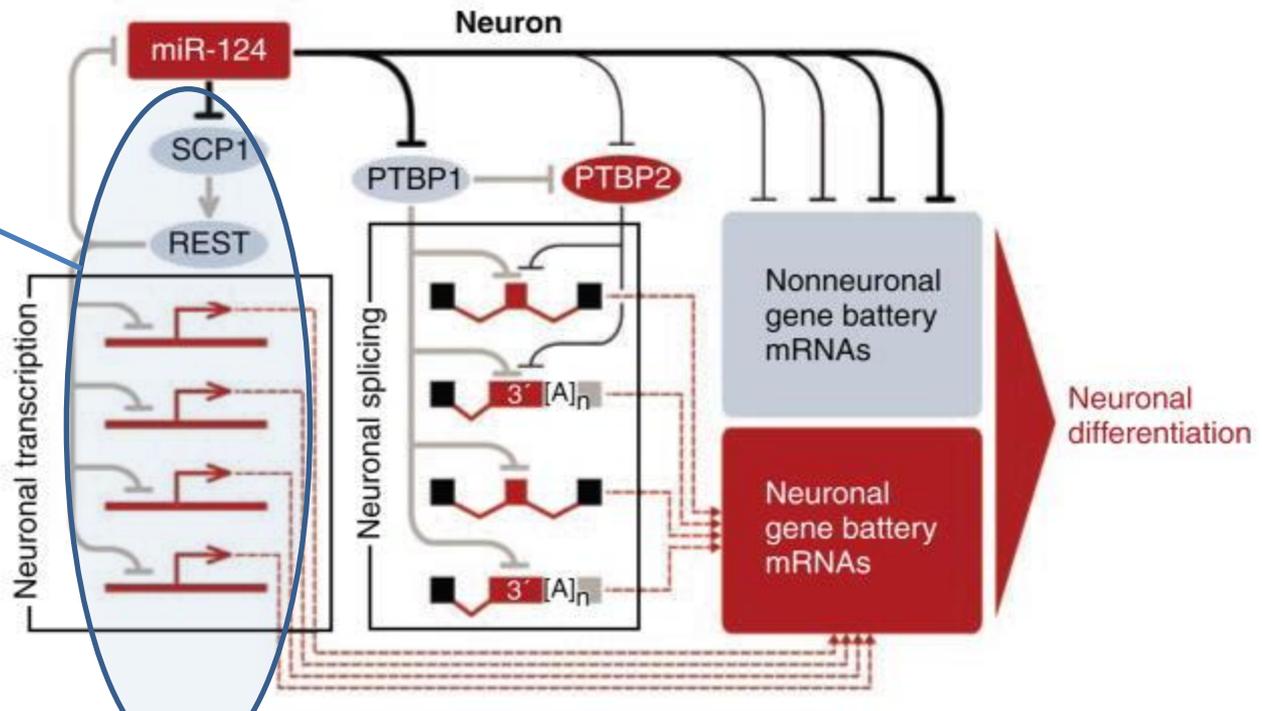
(C) PTB RNA map.

A**B**

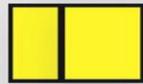
An RNA map predicting Nova-dependent splicing regulation (Nature 2006)

- Nova binding sites (YCAAY clusters)

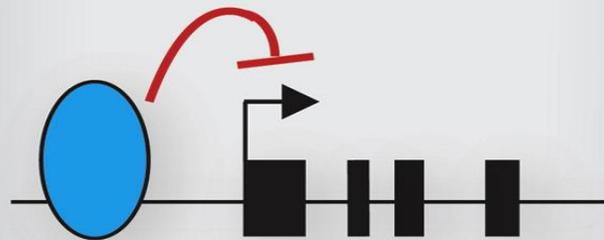


A**REST regulation**

Neural Precursor Cells



REST



Repression of
Neurogenesis genes

Neurons



**Stop
codon**



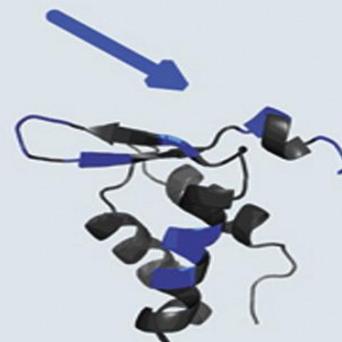
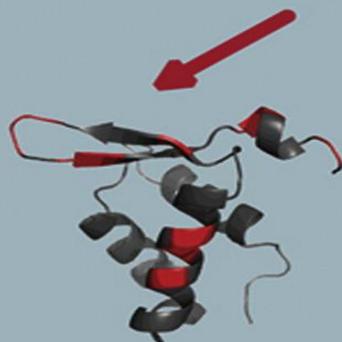
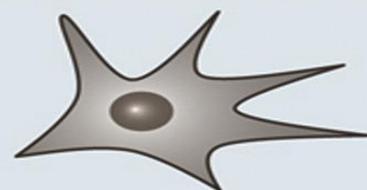
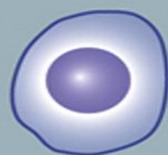
REST4



Expression of
Neurogenesis genes

ESC, iPSC

Differentiated cells

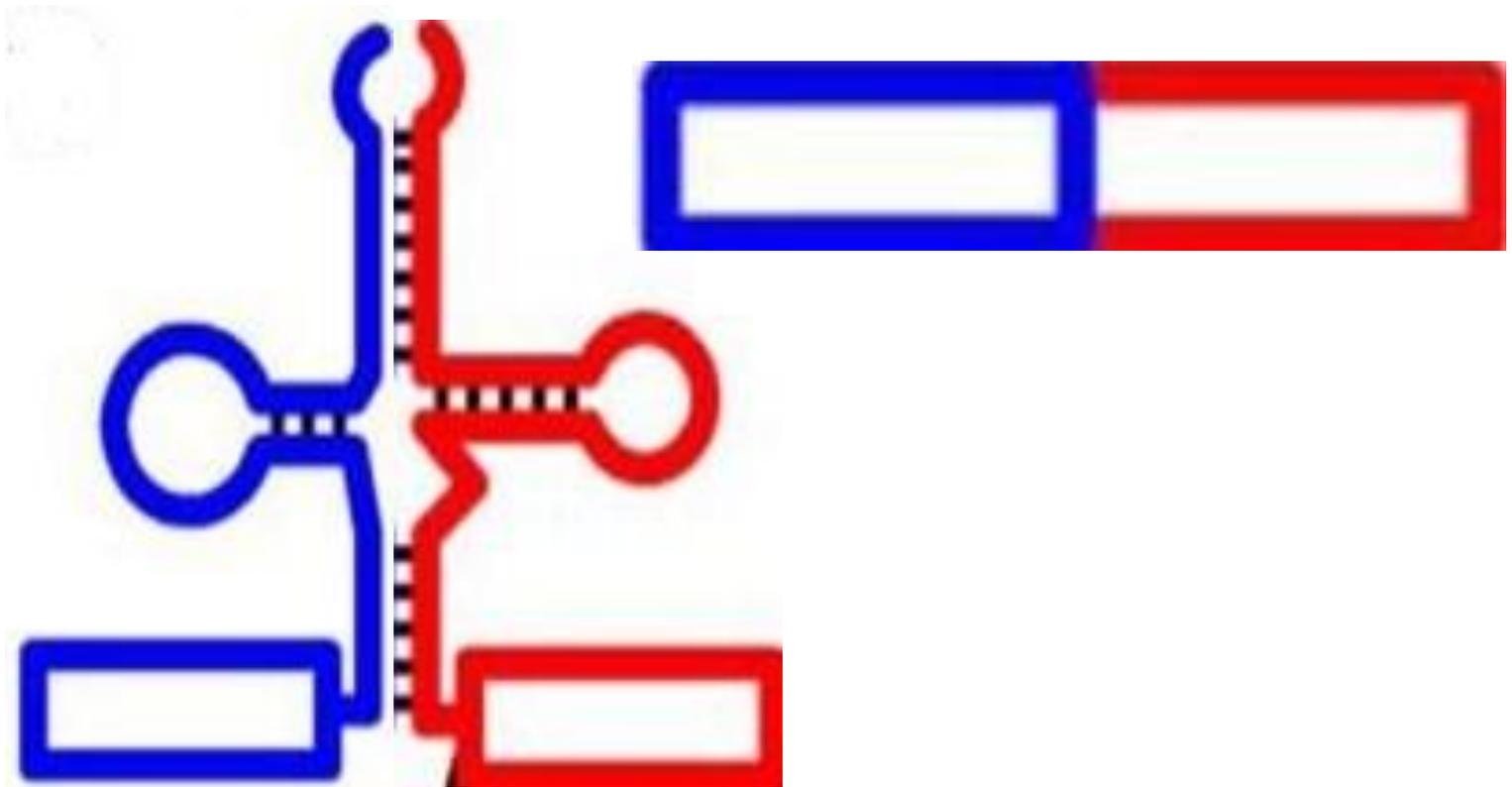


Pluripotency genes
e.g. *OCT4*, *NANOG*



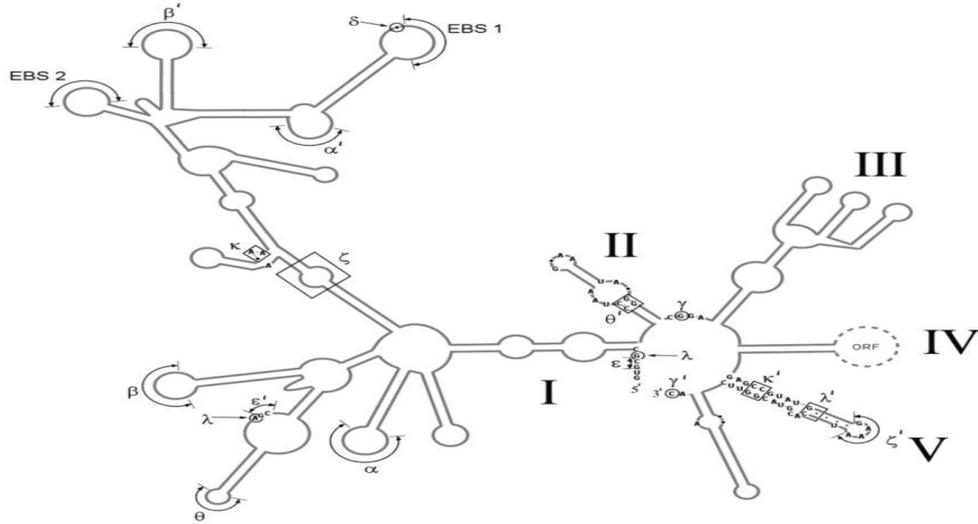
Differentiation genes

тРНК – опосредованный сплайсинг – перспектива генной инженерии

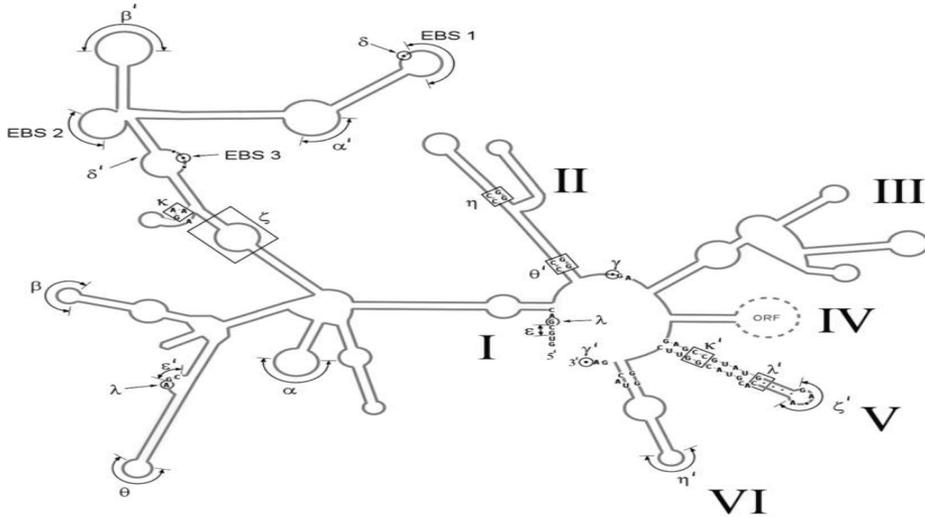


Само-слайсирующие интроны (только РНК)

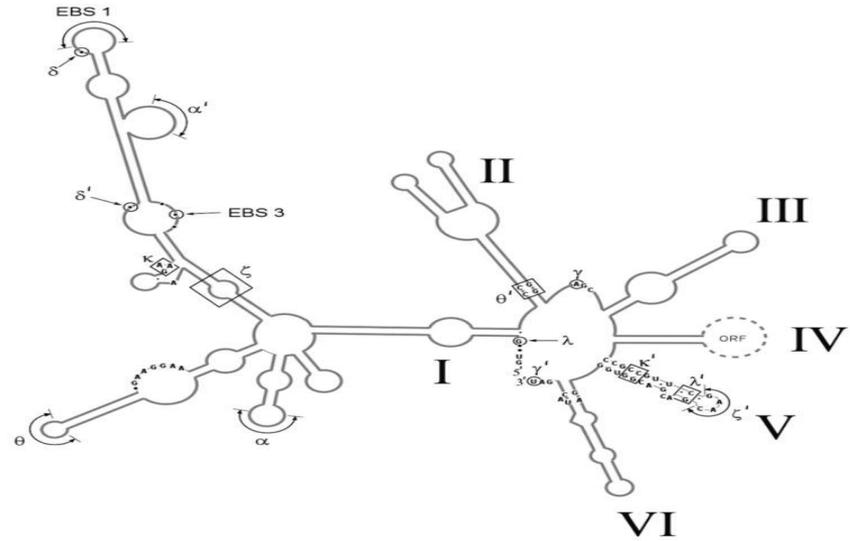
IIA

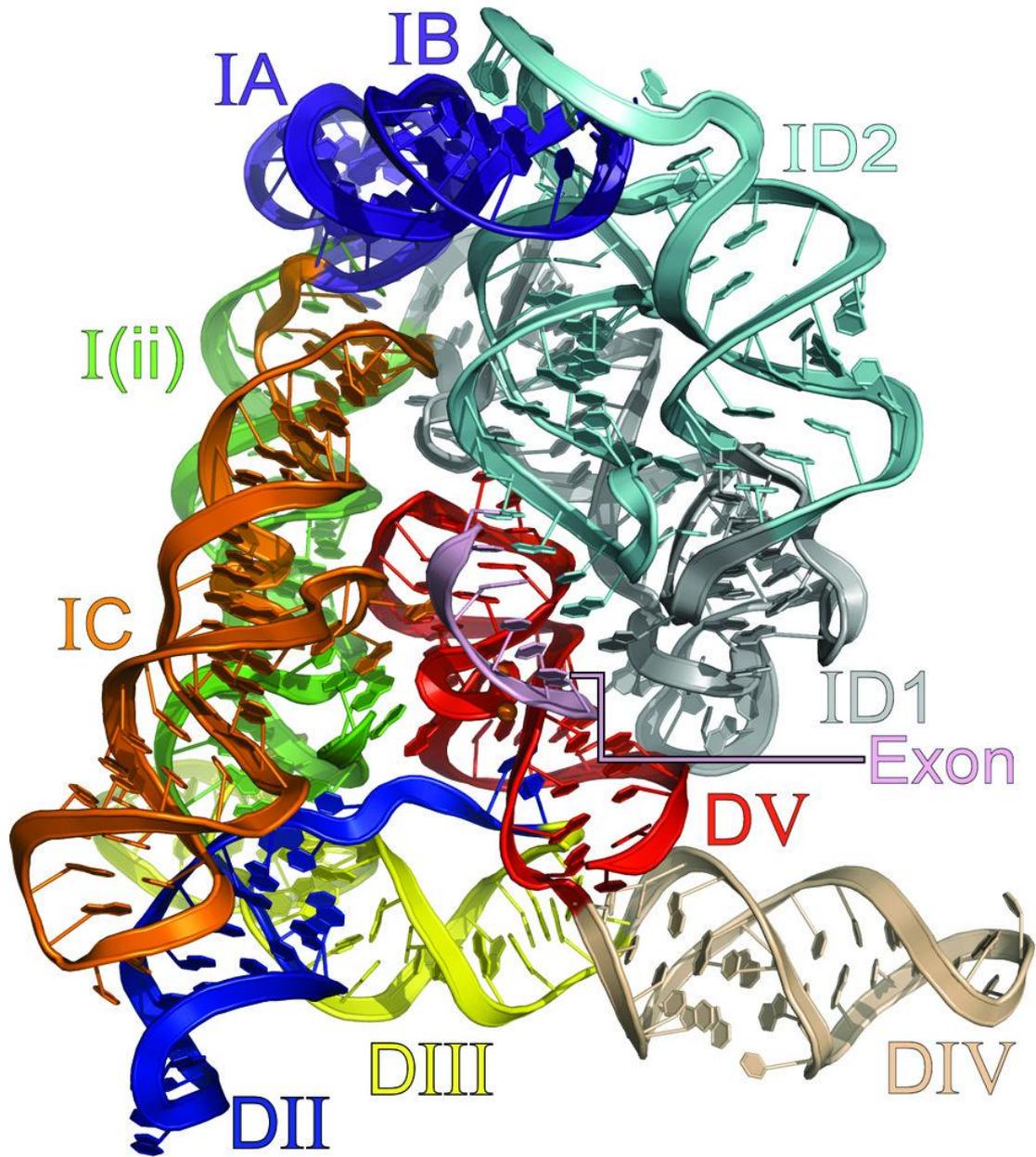


IIB



IIC



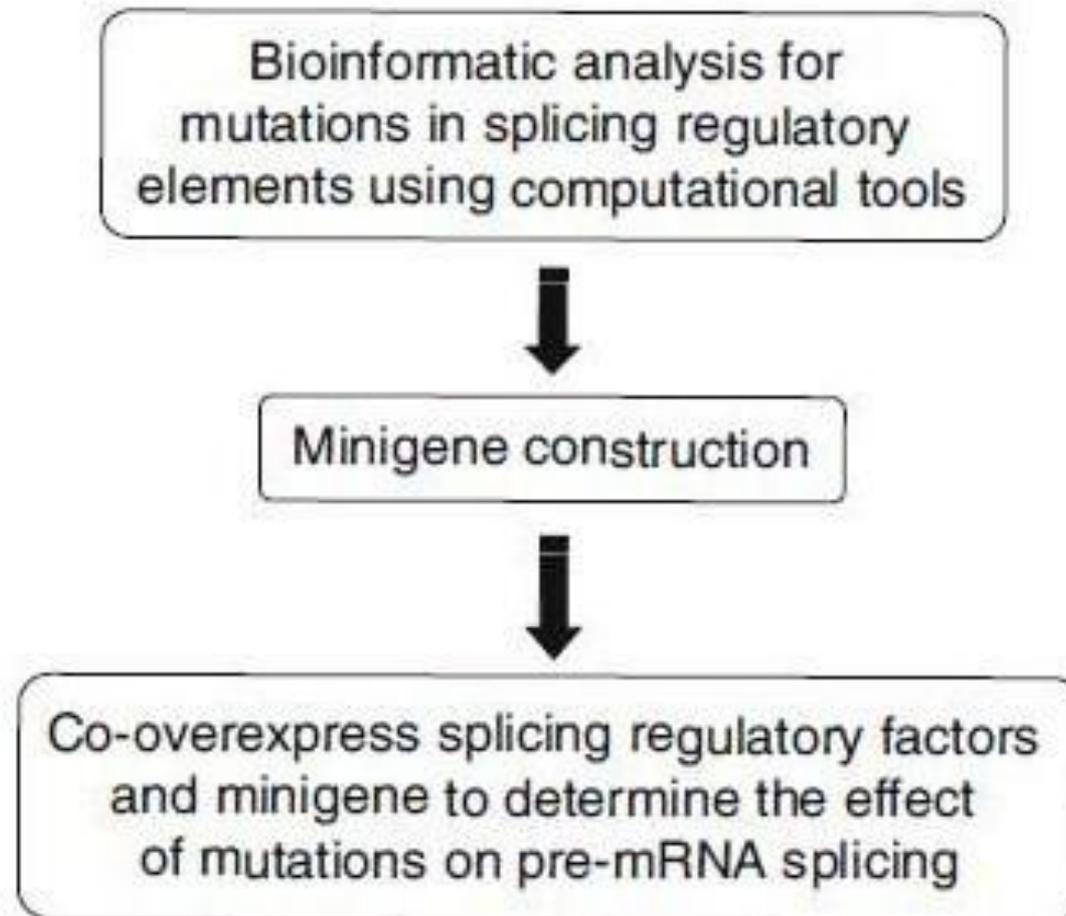


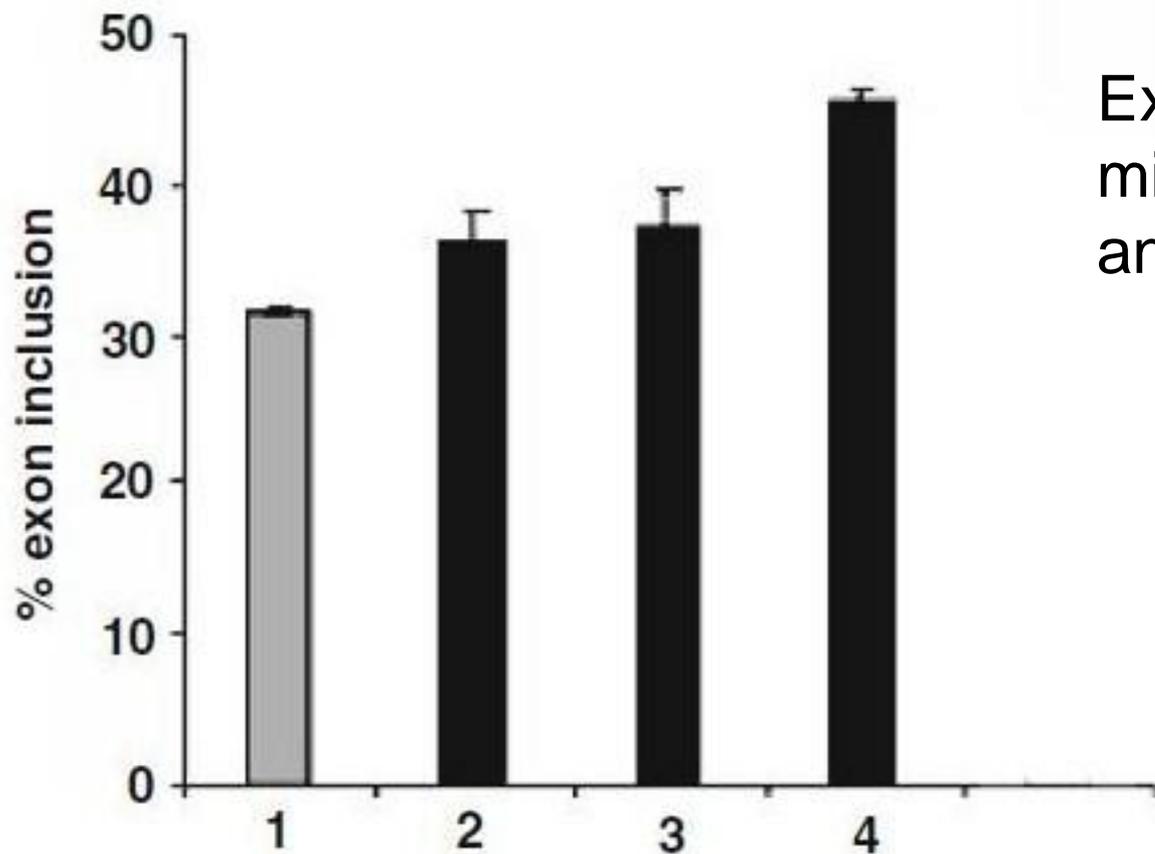
Черты сплайсинга интронов группы I

- Участвуют 5 малых ядерных (мя) РНК:
U1, U2, U4, U5, U6
- Пре-мРНК также принимает активное участие в сплайсинге
- Вторичная структура (дуплексы) (мя) РНК существуют на всех стадиях

Примеры сплайсинга,
методы анализа сплайсинга
ресурсы анализа

Схема изучения сигналов сплайсинга в конкретном гене на панели мутаций минигена



A**B****C**

Example for a minigene analysis.

Проблемы анализа с помощью минигена

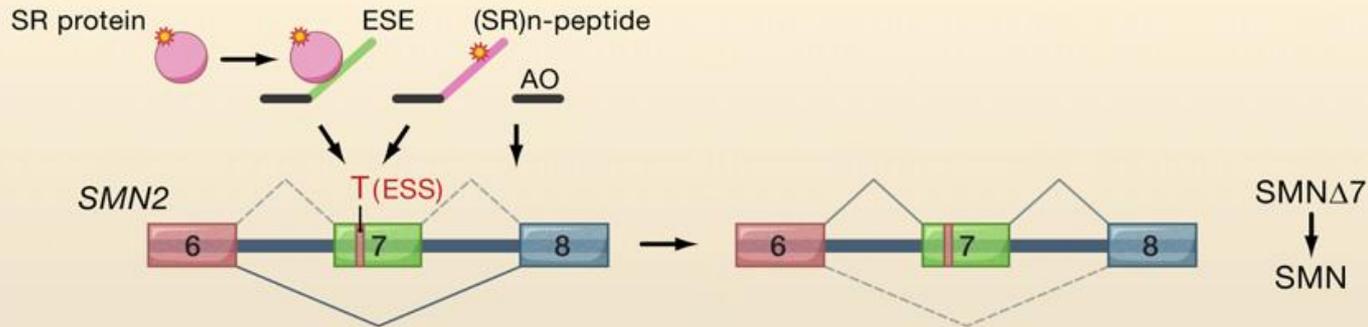
- Интронные энхансеры
- Далекие энхансеры
- Вторичная структура интронов
- Длинные интроны

Databases to explore alt splicing

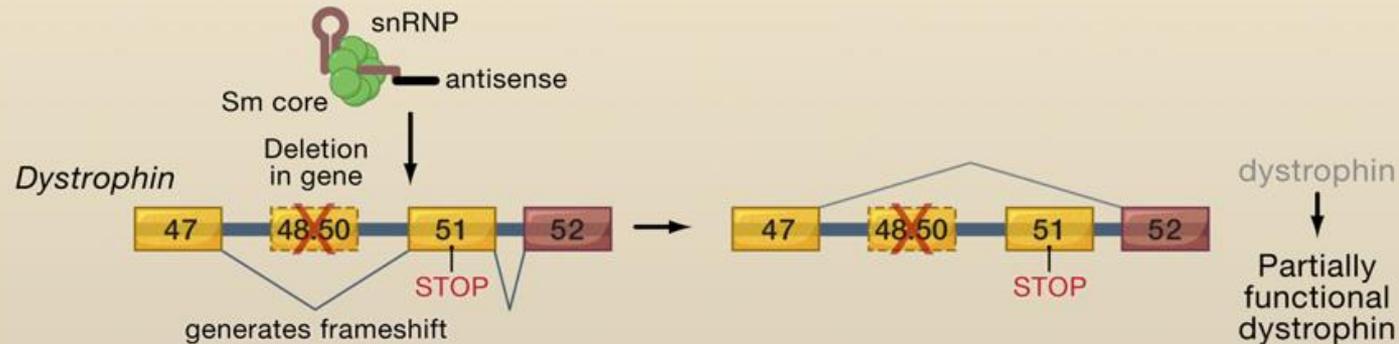
Database	Description	Reference sequence	Species covered	Additional analysis tool	Integration with other resources	URL
ASAP	Database of human tissue specific alt splicing events	ESTs	Human		Microarray	http://bioinfo.mbi.ucla.edu/ASAP/
ASD	Database of manually curated alt events	Genomic	Human and	Intron analysis,	Ensemble	http://www.ebi.ac.uk/asd/
ASTRA	Database of elementary splicing events	Genomic	Human, mouse,		Genbank database	http://alterna.cbrc.jp/index.php?sp=m#search
ECgene	Database that combined	Genomic	Human	Gene structure	Genbank,	http://genome.ewha.ac.kr/ECgene/
H-DBAS	Database of genome-wide screen of alt events	Genomic	Human	Motif	HUGO gene	http://www.hinvital.jp/h-dbas/
Hollywood	Database built upon	Genomic	Human and	Tool for	Ensemble and genbank	http://hollywood.mit.edu/Login.php
MAASE	Manually annotated	Genomic	Human,	Oligonucleotides		http://maase.genomics.purdue.edu/
PALS db	A collection of all isoforms	mRNA	Human,	Putative alternative	Unigene, HUGO	http://ymbc.ym.edu.tw/palsdb/

Therapeutic Approaches that Use or Target RNAs

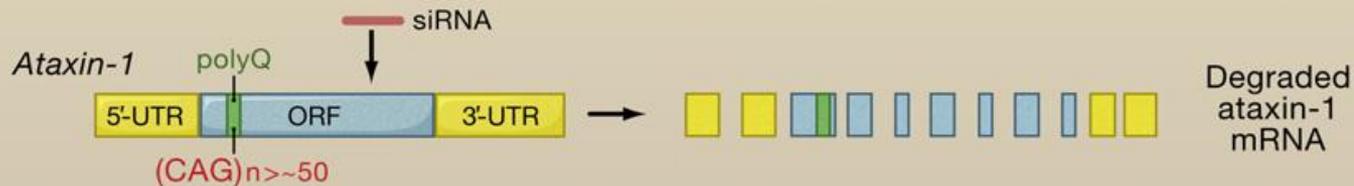
A Antisense oligonucleotides (AOs)



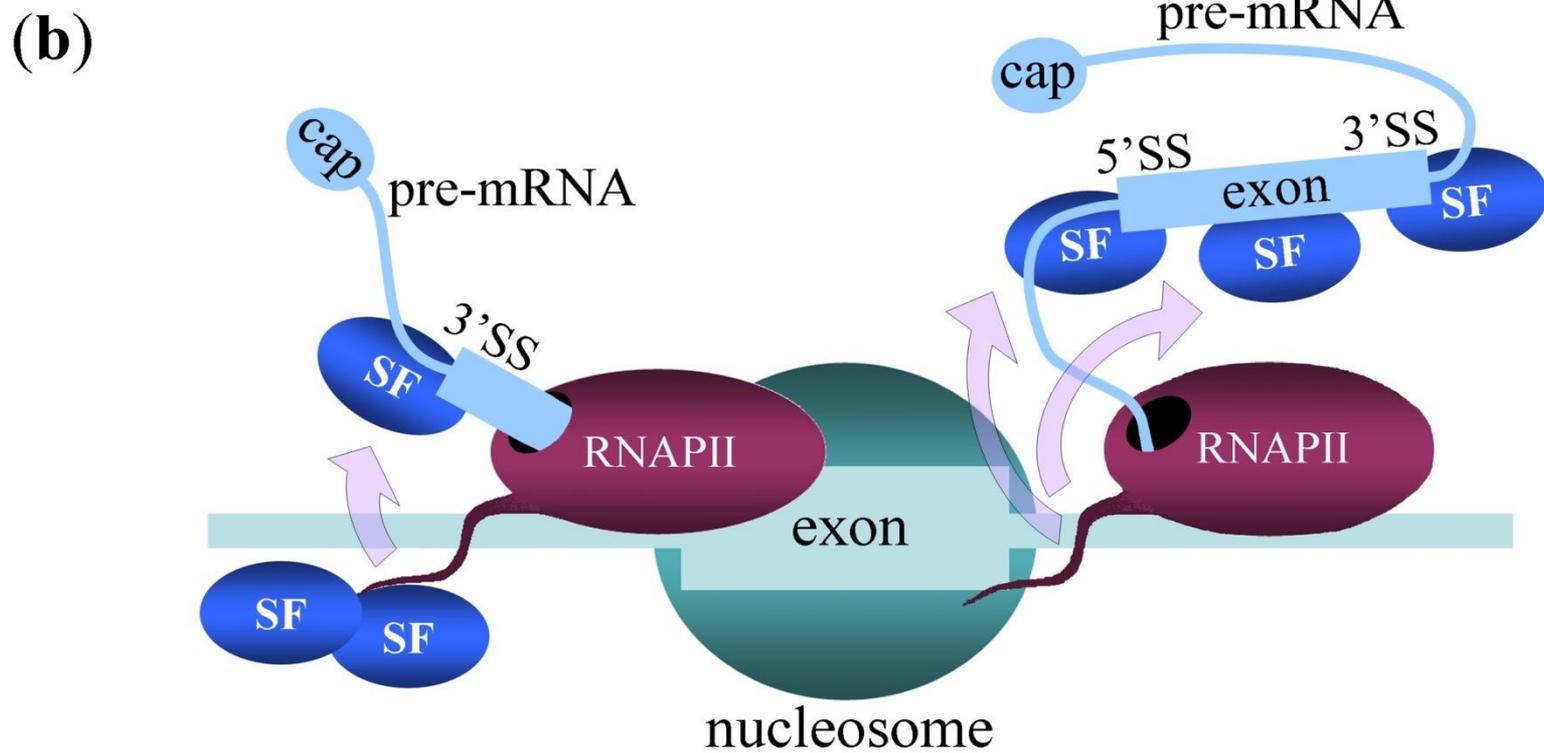
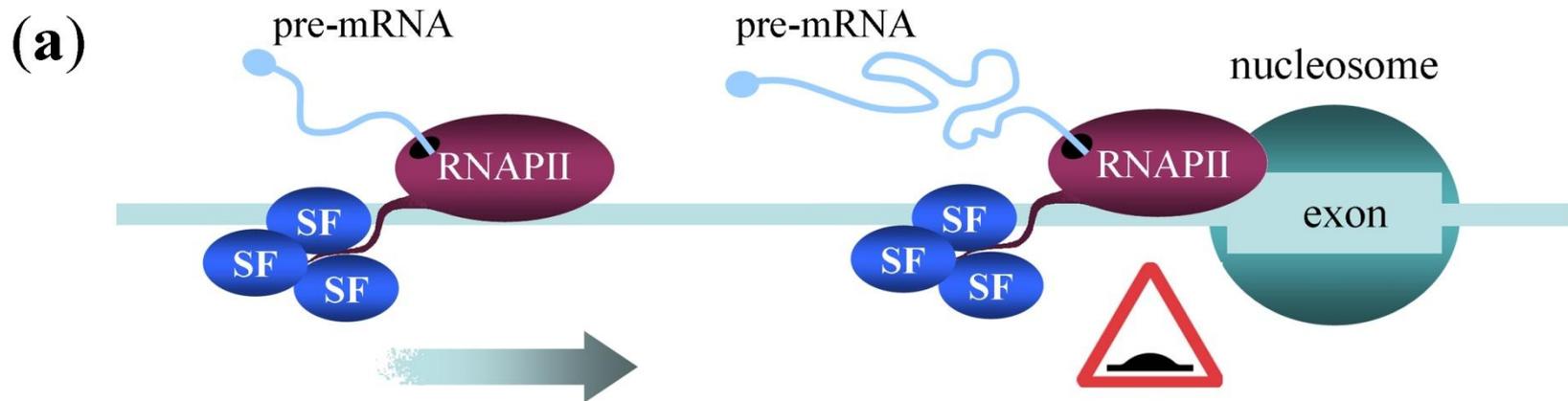
B snRNAs as vehicles for antisense RNA



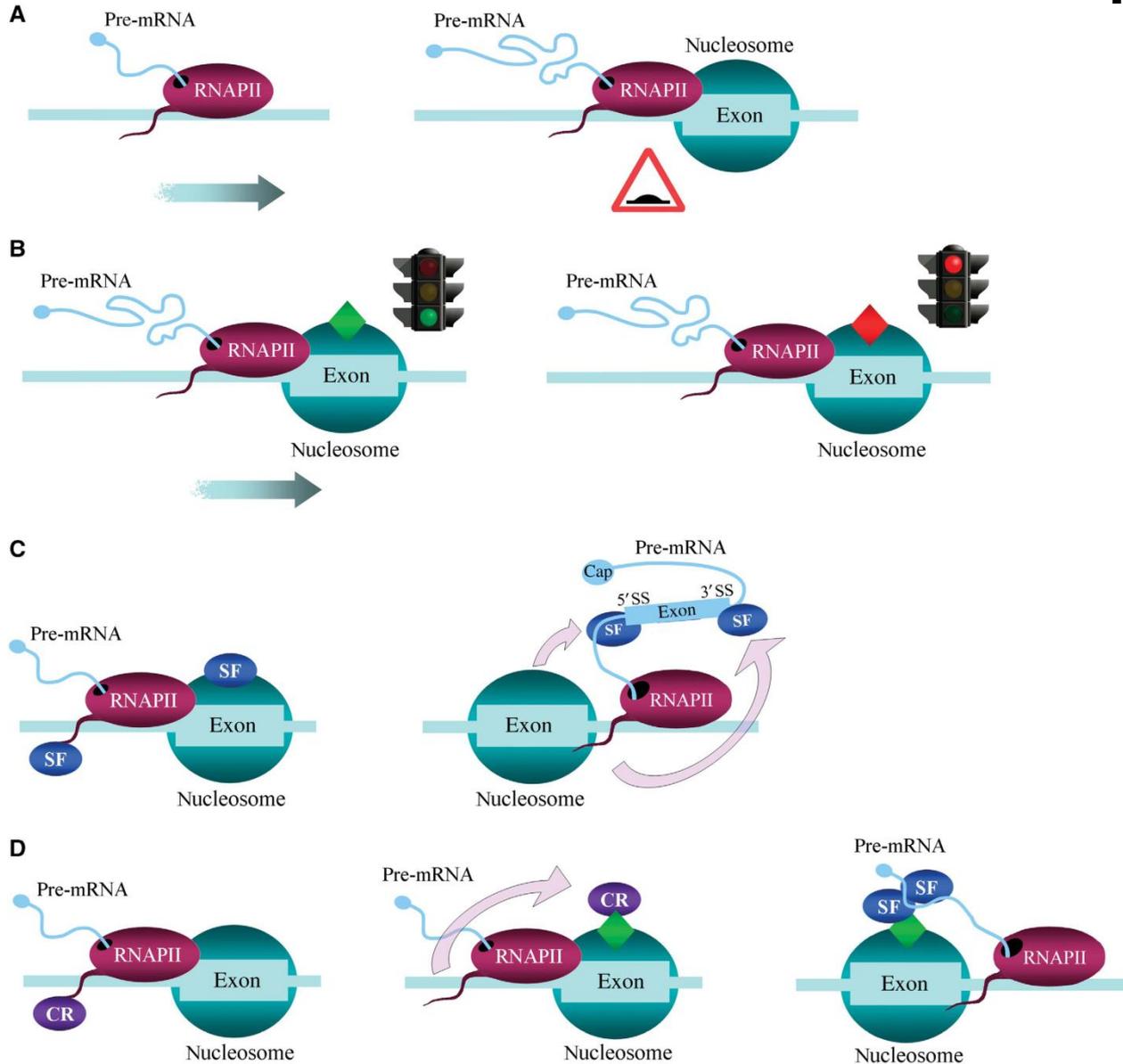
C RNA interference (RNAi)



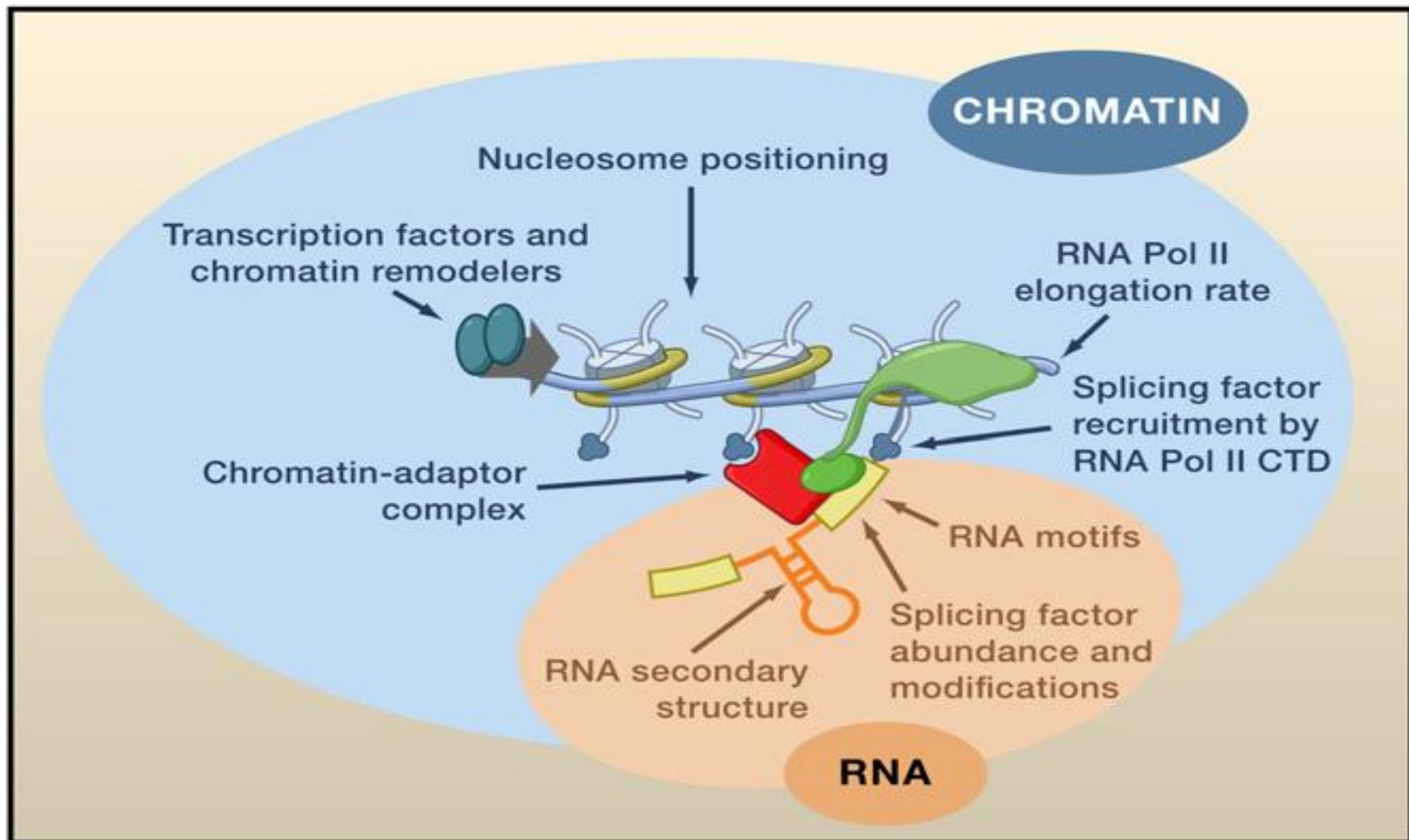
Эпигенетические состояния хроматина и сплайсинг



Nucleosomes may be linked with exon–intron architecture and splicing



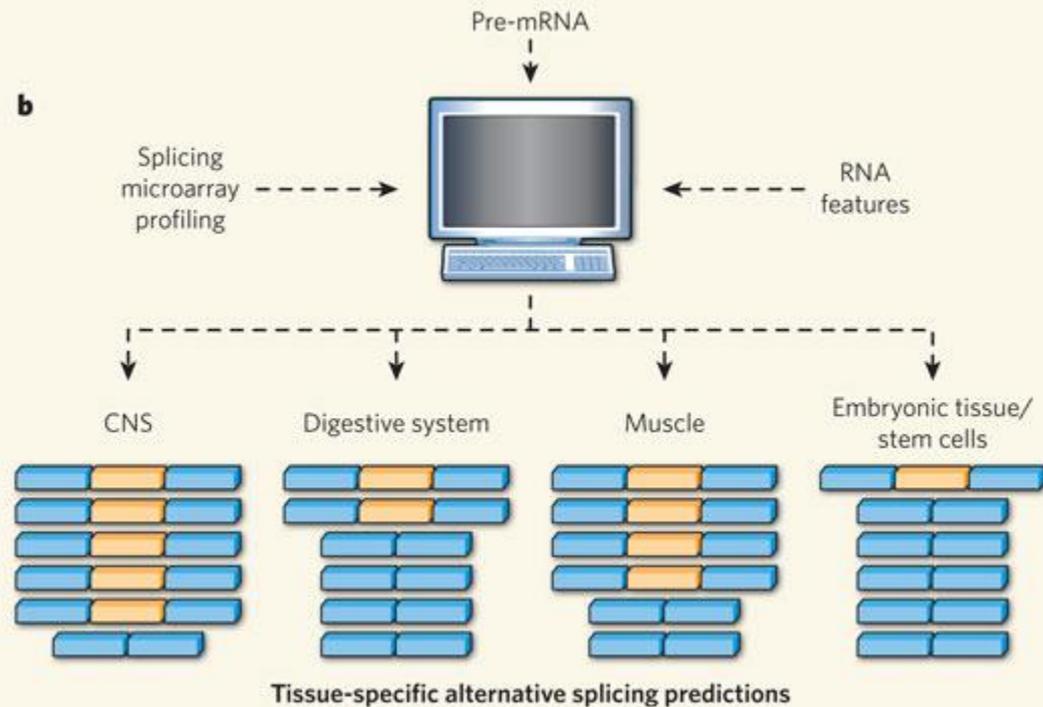
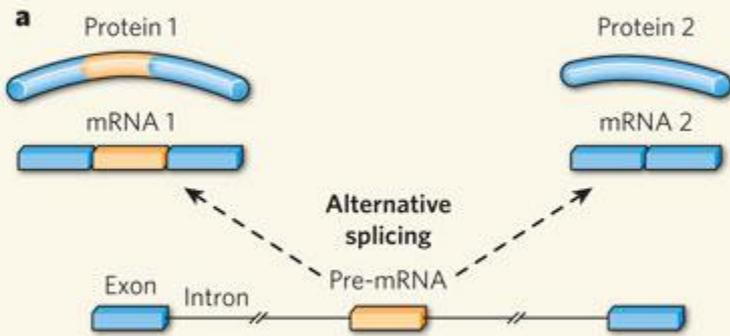
Интегральная модель регуляции (альтернативного) сплайсинга на уровне хроматина



P.S.

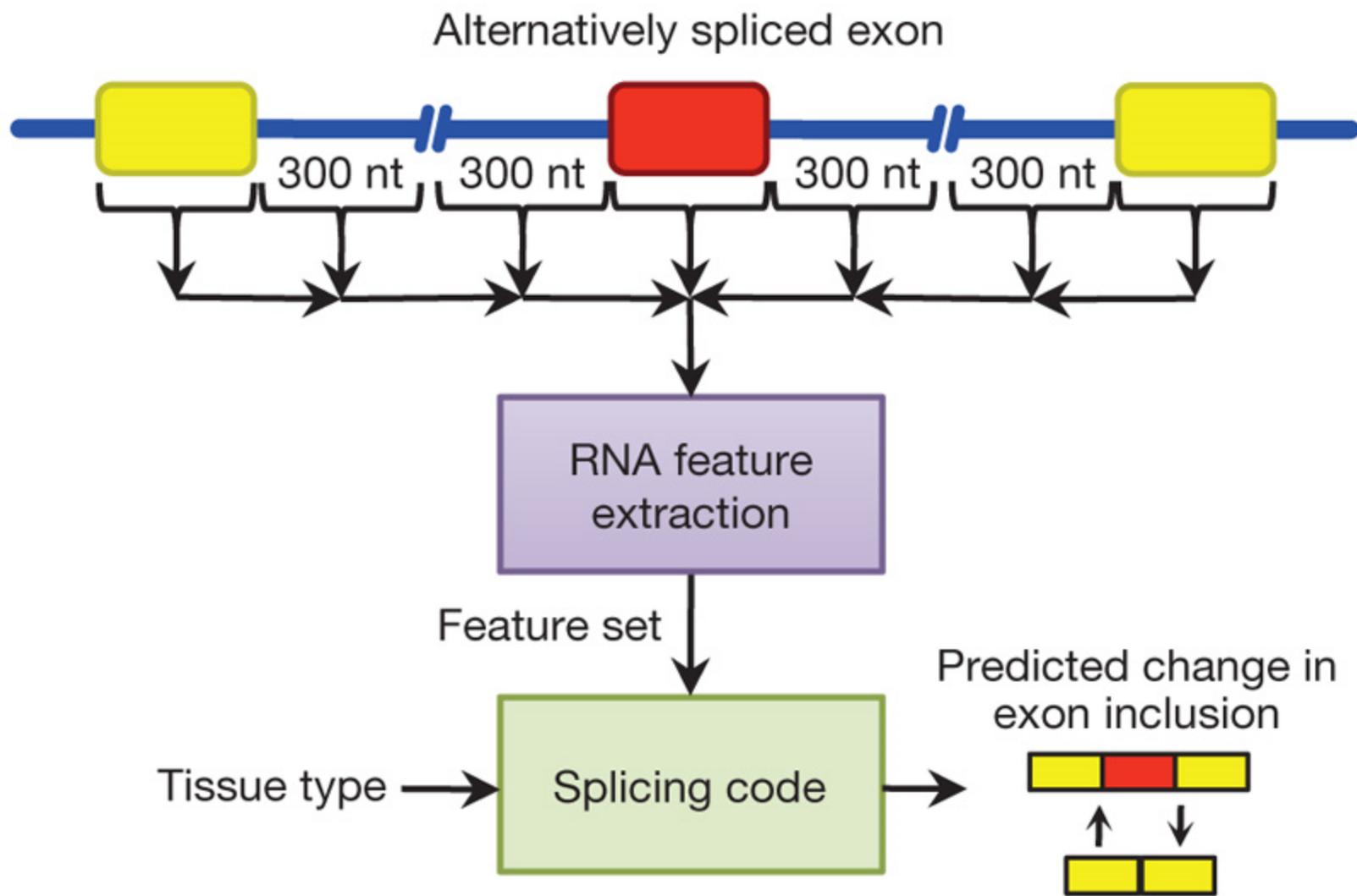
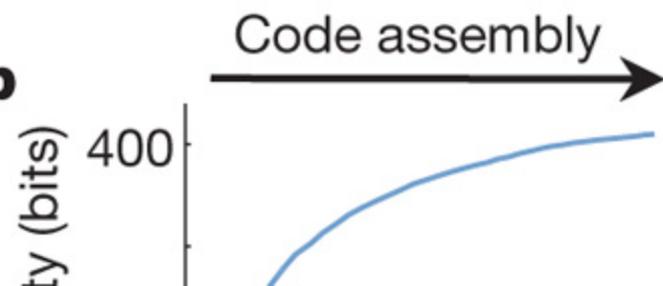


Predicting pre-mRNA fate – splicing code.



Производительность кода сплайсинга

- Код корректно предсказывает включение экзона в тканеспецифичный транскрипт в 82.4% процентов случаях при использовании данных микроэррея, и в 93.3% случаях при использовании данных RT-PCR
- Метод способен отличить альтернативный экзон от конститутивного в 60% (ошибка I рода 40%) при ошибке II рода 1%.

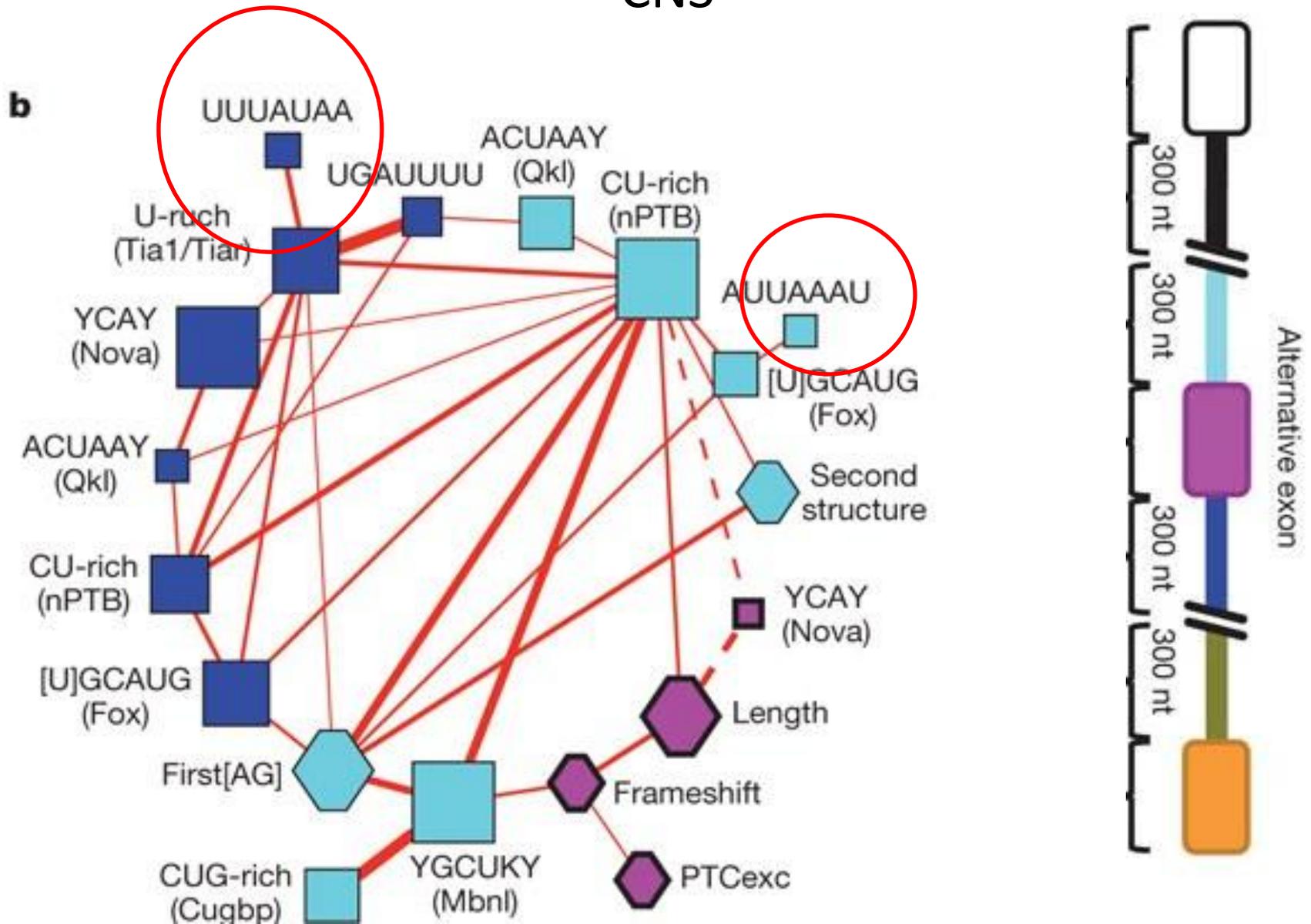
a**b****d**

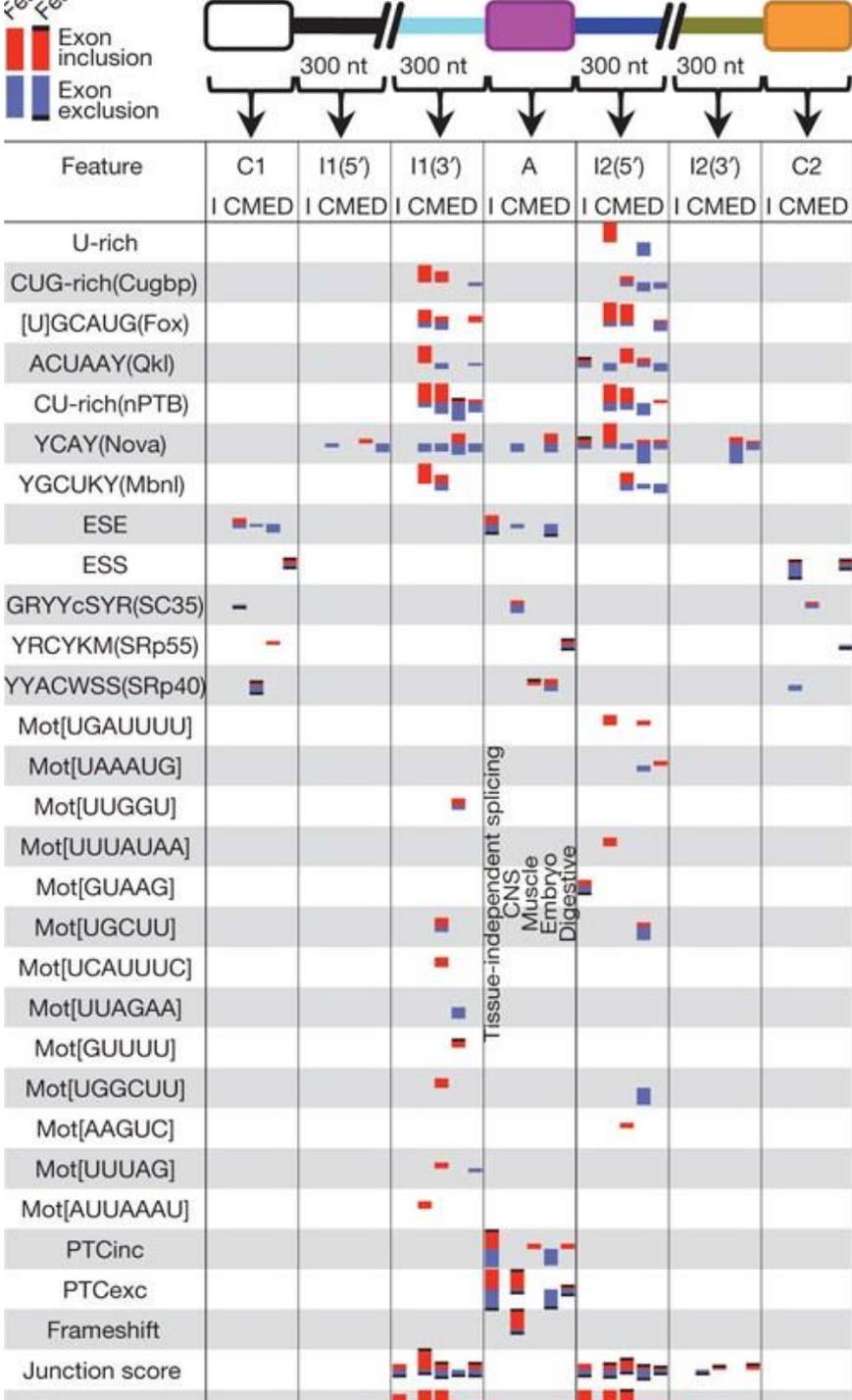
Final assembled code

T

W/o

Unexpectedly frequent feature pairs were identified and used to generate feature interaction networks for CNS





The region-specific activity

of each feature in increased exon inclusion (red bar) or exclusion (blue bar) is shown for CNS (C), muscle (M), embryo (E) and digestive (D) tissues, plus a tissue-independent mixture (I).

Features (Barash et al., 2009)

- The alternative splicing analysis makes the additional important point that inclusion or exclusion of exons relies on multiple RNA feature elements ⁹. On average, 12 tissue-specific RNA features were necessary to define exon inclusion or exclusion in central nervous system tissues and 19 RNA features define embryo-specific alternative splicing. While certain tissue-specific exons share a subset of the same RNA feature elements, the final set of RNA features that make up an exon's identity is likely to be unique for each exon. From now on, it should be the expectation that alternative splicing is modulated by multiple RNA features. This expectation is fully supported when evaluating one of the biochemically most analyzed alternative splicing events, the inclusion of the *Survival of Motor Neuron (SMN)* exon 7, a gene associated with Spinal Muscular Atrophy ¹⁰. Several splicing enhancers and silencers, either intronic or exonic have been identified through mutational analyses in addition to RNA secondary structure elements ([Figure 1A](#)), suggesting that even the splicing of constitutive exons, such as *SMN* exon 7 in its normal context, is mediated by multiple splicing RNA elements. For alternatively spliced exons it can be expected that additional RNA features participate to ensure regulated exon selection. It should also be anticipated that tissue-specific alternative splicing is under the control of splicing networks that consist of unique RNA feature elements. The identity of these splicing networks varies between tissues, thus promoting tissue-specific alternative splicing.