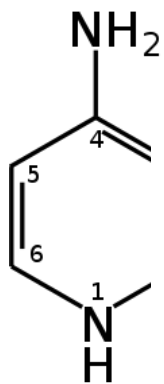
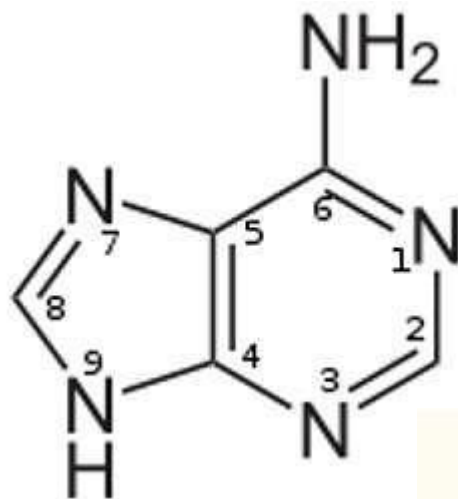


# Строение и функционирование генома

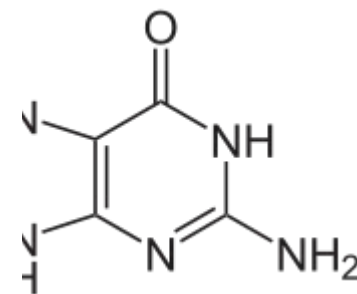
# Нуклеиновые кислоты



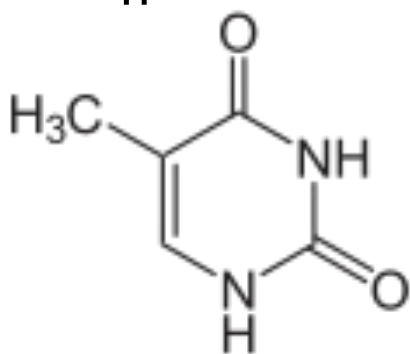
ЦИТОЗИН



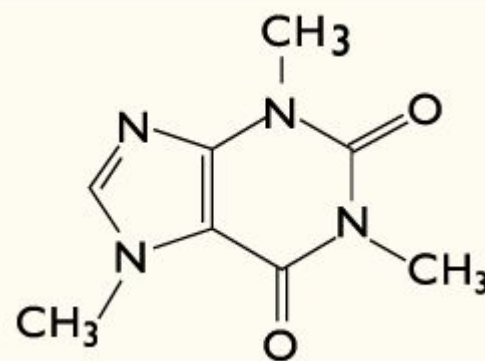
АДЕНИН



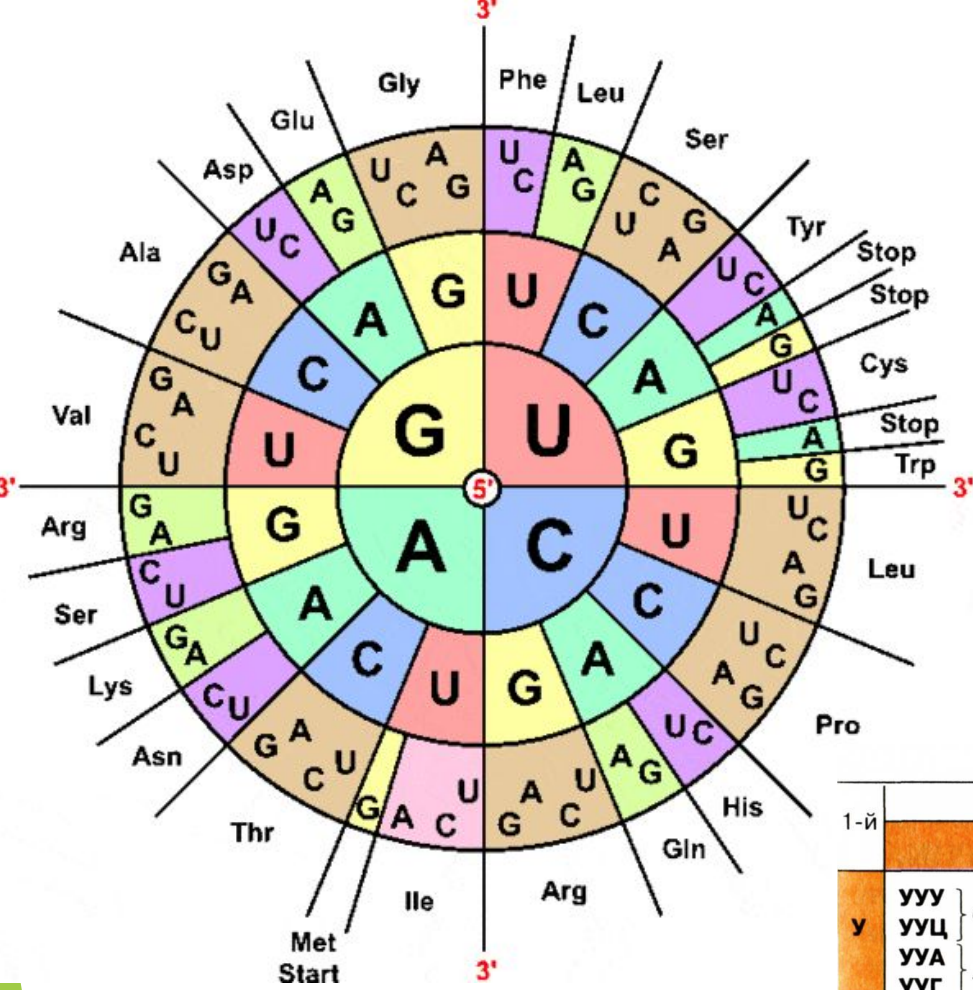
ГУАНИН



ТИМИ́Н



КОФЕИН



		Нуклеотид							
		2-й				3-й			
1-й		У	Ц	А	Г				
У	УУУ	Фенилаланин	УЦУ	Серин	УАУ	Тирозин	УГУ	Цистеин	У
	УУЦ		УЦЦ		УАЦ		УГЦ		Ц
	УУА		УЦА		УАА		УГА		А
	УУГ		УЦГ		УАГ		УГГ		Г
Ц	ЦУУ	Лейцин	ЦЦУ	Пролин	ЦАУ	Гистидин	ЦГУ	Аргинин	У
	ЦУЦ		ЦЦЦ		ЦАЦ		ЦГЦ		Ц
	ЦУА		ЦЦА		ЦАА		ЦГА		А
	ЦУГ		ЦЦГ		ЦАГ		ЦГГ		Г
А	АУУ	Изолейцин	АЦУ	Треонин	ААУ	Аспарагин	АГУ	Серин	У
	АУЦ		АЦЦ		ААЦ		АГЦ		Ц
	АУА		АЦА		ААА		АГА		А
	АУГ		АЦГ		ААГ		АГГ		Г
Г	ГУУ	Валин	ГЦУ	Аланин	ГАУ	Аспарагиновая кислота	ГГУ	Глицин	У
	ГУЦ		ГЦЦ		ГАЦ		ГГЦ		Ц
	ГУА		ГЦА		ГАА		ГГА		А
	ГУГ		ГЦГ		ГАГ		ГГГ		Г

# Основные события

- ▶ Проект *Геном Человека* (1990 – 2003гг)
- ▶ Проект *ENCODE* (2003г)

Цель – провести полный анализ функций элементов генома человека.

На данный момент показано, что по крайней мере 80 % генома человека является биологически активным.



# Уровни компактиза- ции ДНК

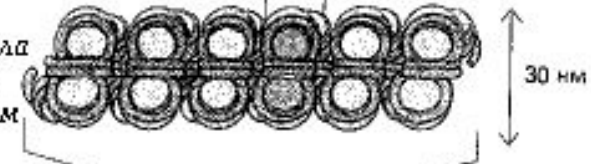
Короткий участок  
двойной спирали ДНК



Хроматин  
в форме «бусин на нити»



Хроматиновая фибрилла  
30 нм, состоящая из  
упакованных нуклеосом



Часть хромосомы  
в растянутом виде



Конденсированный  
участок метафазной  
хромосомы

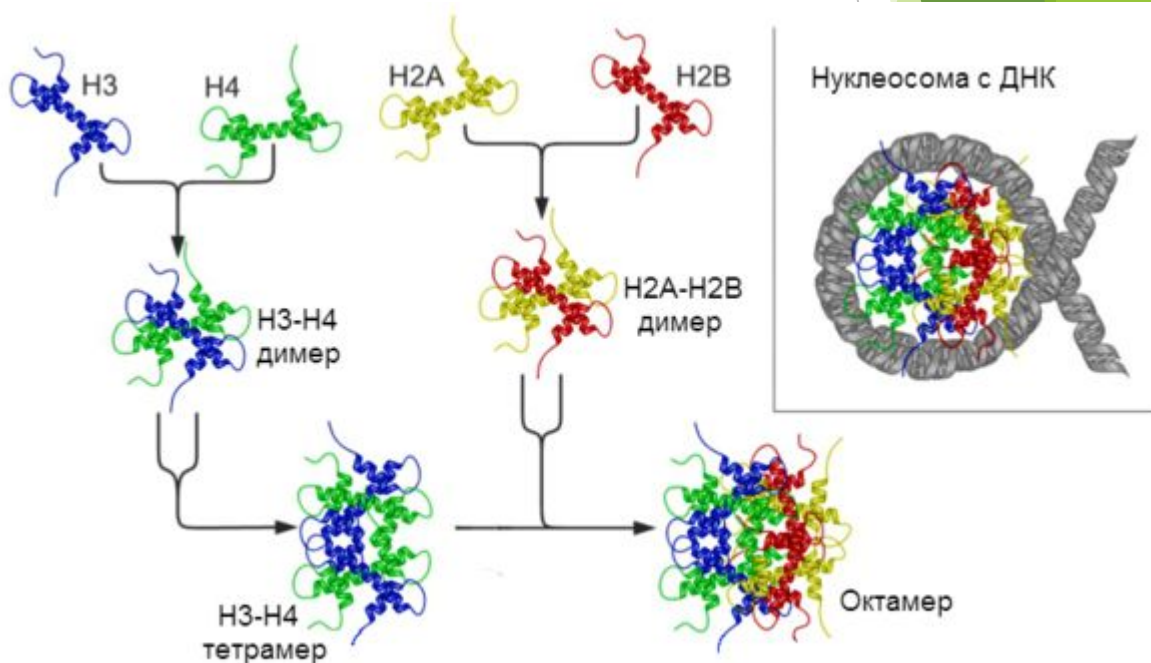


Целая  
метафазная  
хромосома



# ГИСТОНЫ

- ▶ Гистоны – обширный класс ядерных белков, выполняющих две основные функции: они участвуют в упаковке нитей ДНК в ядре и в эпигенетической регуляции ядерных процессов.
- ▶ Существует пять различных типов гистонов: H1/H5 (линкерный); H2A, H2B, H3, H4 (коровые).
- ▶ Большое содержание лизина и аргинина.



# Хроматин

## Эухроматин

- ▶ Деспирализованный хроматин, готовый к экспрессии.
- ▶ Имеет большее содержание негистоновых белков, по сравнению с гетерохроматином.

## Гетерохроматин

протяжённые участки повторяющихся и высоко конденсированных последовательностей, которые не кодируют никаких белков.



# Хромосомы

- ▶ 23 пары хромосом: 22 пары — аутосомы и 1 пара — половые хромосомы.
- ▶ От 45 миллионов п.о. до 280 миллионов п.о.

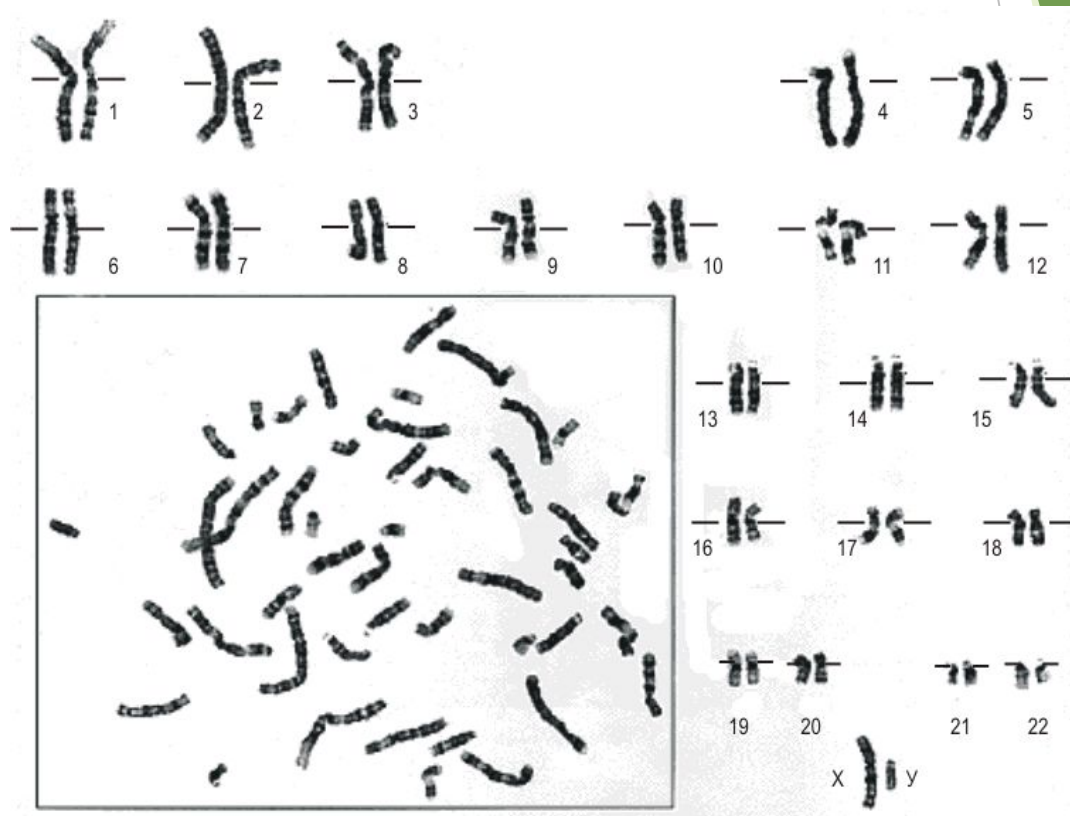


# Методы окрашивания

Кариотип — полный набор хромосом человека.

В 1960 г. были разработаны методы окрашивания, которые позволяют выявить чередующиеся светлые и окрашенные поперечные полосы (диски, «бэнды»). G-диски (в честь Гизмы).

Нумерация: от самой большой к самой маленькой.  
Исключение: **21**

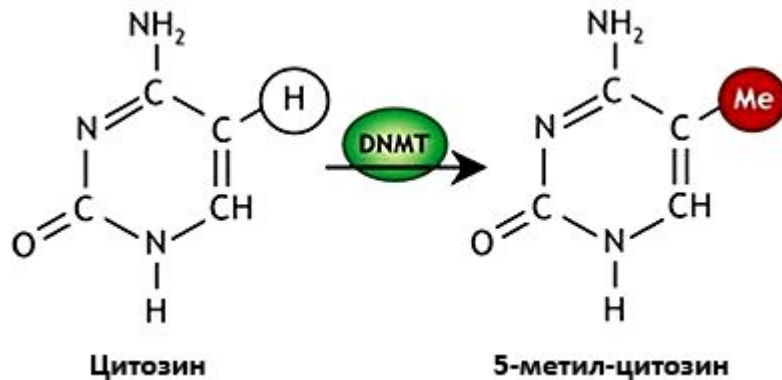


идиограмм  
а

- ▶ Тёмные диски относительно более богаты парами оснований А=Т, в них довольно мало генов, они позднее других реплицируются в клеточном цикле.
- ▶ Светлые G-диски более обогащены парами оснований G C, содержат большую часть генов и реплицируются раньше.
- ▶ Общее содержание GC в геноме человека составляет 41%, но оно неравномерно: 37% в тёмных и в 45% светлых.

# Эпигенетика

- ▶ Метилирование ДНК
- ▶ Осуществляется ДНК-метилтрансферазами.
- ▶ Приводит к подавлению активности гена.



# Ацетилирование гистонов

- ▶ Ацетилирование ацетилтрансферазой 14-го и 9-го лизинов гистона H3 (H3K14ac и H3K9ac, соответственно) коррелирует с транскрипционной активностью в данном районе хромосомы.
- ▶ Это происходит из-за того, что ацетилирование лизина меняет его положительный заряд на нейтральный, что делает невозможным его связь с негативно заряженными фосфатными группами в ДНК.
- ▶ В результате, происходит отсоединение гистонов от ДНК, что приводит к посадке ДНК других транскрипционных факторов которые запускают транскрипцию.

- ▶ **Ремоделирование хроматина** – процесс перемещения нуклеосом по ДНК, приводящий к изменению плотности нуклеосом или к расположению их на определенном расстоянии друг от друга.
- ▶ Ремоделирование осуществляется специальными белковыми комплексами, при этом затрачивается энергия в виде АТФ.

# Цифры

- ▶ 3,2 миллиарда пар нуклеотидов
- ▶ 20 000 – 35 000 генов
- ▶ Интроны составляют 20-25% генома

Ген – участок ДНК, который транскрибируется в РНК-копию одной из нитей ДНК.

## В среднем

Ген человека содержит 27 т.п.н.

Число экзонов, приходящееся на ген — 8.

Размер экзона — 145 п.н.

Размер интрона — 3365 п.н.

## В частности

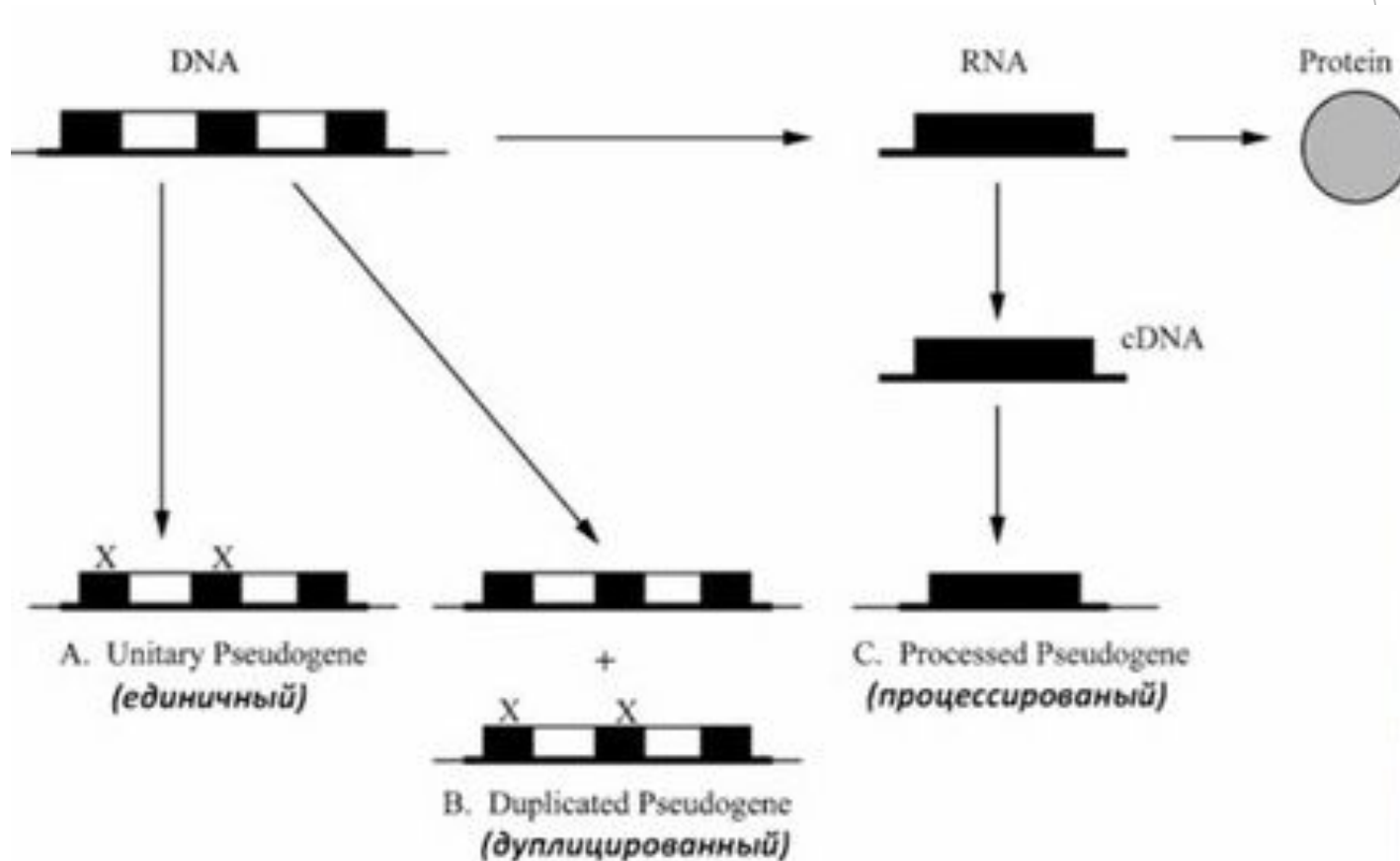
Ген 30 т.п.н. — 2,4 млн.п.н.

Длина интронов может быть более 30 т.п.н

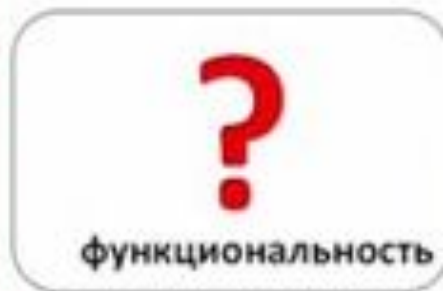
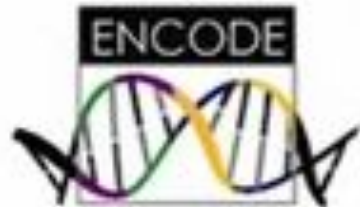


# Псевдогены?

- ▶ Образуются из дублированных генов, одна из копий которых приобретает такие мутации, которые делают невозможным её экспрессию.

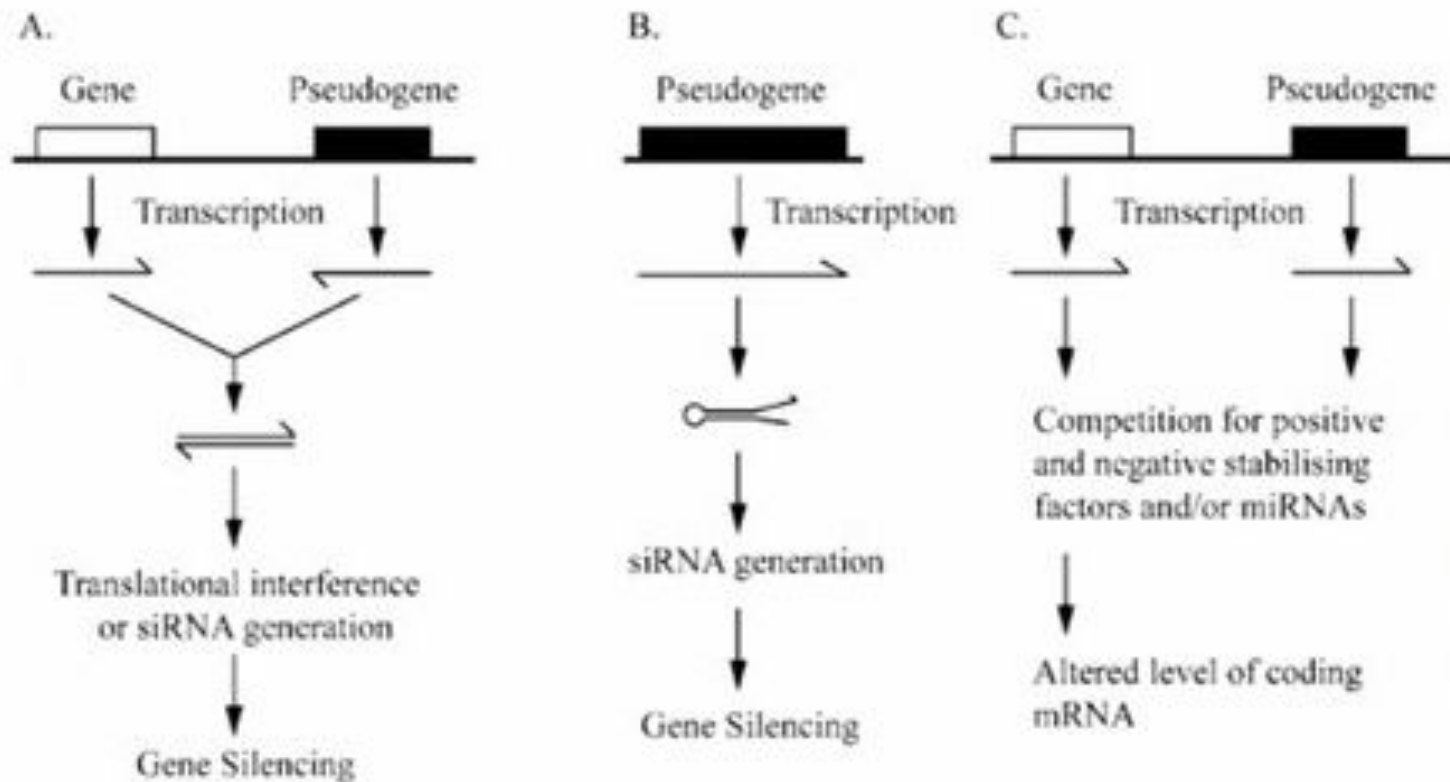


Виды	Количество генов	Количество процессированных псевдогенов	%
<i>Arabidopsis</i>	33583	4260	13
<i>Caenorhabditis elegans</i>	21187	2445	12
<i>Drosophila melanogaster</i>	22372	2208	10
<i>Danio rerio</i>	34291	16357	48
<i>Gallus gallus</i>	22720	5539	24
<i>Canis lupus familiaris</i>	24953	12852	52
<i>Rattus norvegicus</i>	42743	13962	33
<i>Mus musculus</i>	58433	19119	33
<i>Pan troglodytes</i>	32989	16785	51
<b>Homo sapiens</b>	<b>42000</b>	<b>17609</b>	<b>42</b>



- Образованы 10% кодирующих генов
- Около 80% примато-специфичные
- Значительная фракция псевдогенов (до 20%) транскрипционно активна...

# Механизмы действия процессированных псевдогенов



# Копийность

## Однокопийная ДНК

Интроны

Экзоны

Некодирующие  
последовательности

## Повторяющаяся ДНК

- ▶ Мусорная «Junk» ДНК
  1. Тандемно повторяющиеся последовательности
  2. Диспергированные последовательности

Центромерная ДНК

170 п.н.

Образует ряды в 250 т.п.н. — 5 млн.

п.н.

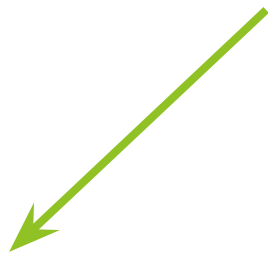
Образует центромеры.

Теломерная ДНК

GGGTTA

Потворяются 250-1500

# Диспергированные повторяющиеся последовательности (ретротранспозоны)



## LINE (Long Interspersed Elements)

Представлена в геноме до 500 000 копий — ~15% всего генома.

Только 40-50 из них функционально активны (кодируют белки, которые способны вызывать транспозицию либо самого элемента L1, либо др мобильных элементов)



SIN

E

Последовательности Alu  
300 п.н.

~1 млн. последовательностей  
Alu

~10-12% всей ДНК

В основном они находятся между генами и внутри интронов, но изредка могут быть включены в мРНК

# Повторяющиеся последовательности в ДНК

## Диспергированные повторы

### Транспозоны

мобильные элементы, репликация которых включает перемещение своей ДНК последовательности на новое место в геноме



- 2-3% генома человека
- Примеры: *MER1*, *MER2*

### Ретротранспозоны

мобильные элементы, которые могут самовоспроизводиться в геноме, осуществляя реакцию обратной транскрипции.

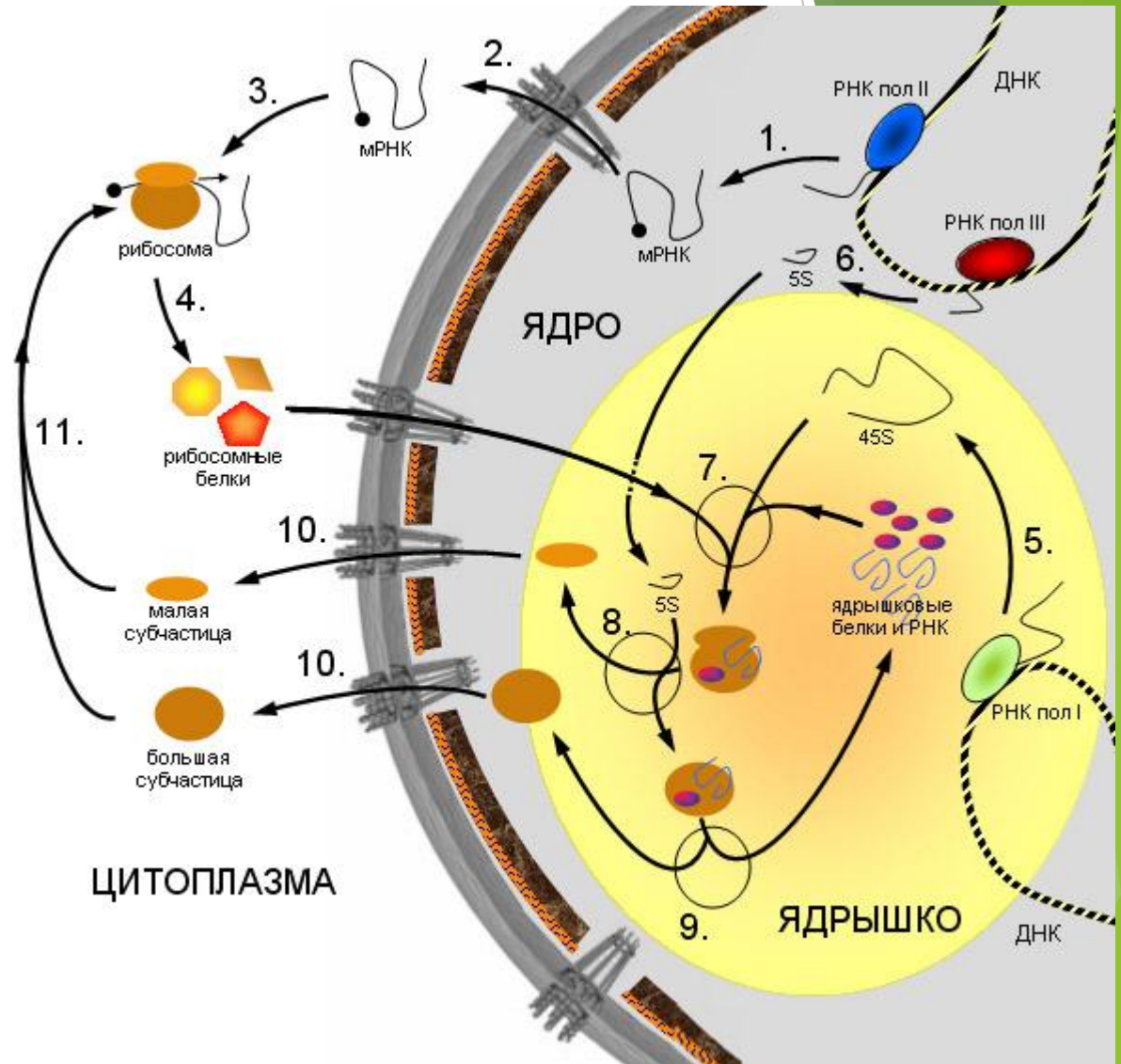


- 42% генома человека
- Примеры: *Alu*, *MIR*, *LINE1*, *LINE2*.....



Ядрышковые  
организаторы?

Синтез рРНК РНК  
полимеразой I





# Регуляция

- ▶ Фактически все метаболические функции живых клеток опосредованы белками.
- ▶ Со всеми генами связаны регуляторные последовательности ДНК, к которым присоединяются белки-регуляторы.
- ▶ Регуляторные последовательности делятся на транс- и цис- регуляторные.

- ▶ Цис-элементы обычно находятся поблизости от сайта инициации транскрипции в 5'-направлении, но могут располагаться и на значительном отдалении от него.
- ▶ подразделяются на промоторы, энхансеры и сайленсеры
- ▶ Как правило, транс-регуляторные элементы кодируют факторы транскрипции.

## A map of the cis-regulatory sequences in the mouse genome

Yin Shen<sup>1\*</sup>, Jiong Yao<sup>1\*</sup>, David P. McQuay<sup>2</sup>, Zhao Ye<sup>2</sup>, Lei Fu<sup>2,3</sup>, Jianxin Guo<sup>2</sup>, Ulrich Wagner<sup>2</sup>, James Street<sup>1,4</sup>, Leonard Lee<sup>1</sup>, Victor V. Lobanenkov<sup>2</sup> & Bing Ren<sup>1,2</sup>

The laboratory mouse is the most widely used mammalian model organism in biomedical research. The  $2.4 \times 10^7$  bases of the mouse genome possess a high degree of conservation with the human genome<sup>1</sup>, so a thorough annotation of the mouse genome will be of significant value to understanding the function of the human genome. So far, most of the functional sequences in the mouse genome have yet to be found, and the cis-regulatory sequences in particular are still poorly annotated. Comparative genomics has become a powerful tool for the discovery of these sequences<sup>2</sup>, but on its own it cannot resolve their temporal and spatial functions. Recently, ChIP-seq has been developed to identify cis-regulatory elements in the genomes of several organisms including humans, *Drosophila melanogaster* and *Caenorhabditis elegans*<sup>3,4</sup>. Here we apply the same experimental approach to a diverse set of 19 tissues and cell types in the mouse to produce a map of nearly 300,000 mouse cis-regulatory sequences. The annotated sequences add up to 17% of the mouse genome, and include more than 70% of conserved non-coding sequences. We define tissue-specific enhancers and identify potential transcription factors regulating gene expression in each tissue or cell type. Finally, we show that much of the mouse

genome is organized into domains of coordinately regulated enhancers and promoters. Our results provide a resource for the annotation of functional elements in the mammalian genome and for the study of mechanisms regulating tissue-specific gene expression.

We identified the genomic localization of RNA polymerase II (pII), the initiator-binding protein (CCCTC-binding factor (CTCF)) and three chromatin modification marks, histone H3 lysine 4 trimethylation (H3K4me3), histone H3 lysine 4 monomethylation (H3K4me1) and H3 lysine 27 acetylation (H3K27ac), in 13 adult tissues, four embryonic tissues and two primary cell lines (Fig. 1a,b) by performing chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) (Supplementary Tables 1 and 2). Enrichment of H3K4me1 or pII binding signals is indicative of an active promoter, whereas the presence of H3K4me3 or H3K27ac marks promoter regions can be used as marks for enhancers<sup>5,6</sup>. CTCF binding is considered a mark for potential insulator elements<sup>7</sup>. In a subset of tissue and cell types, we also performed ChIP-seq on the coactivator protein p300 and used its promoter-distal binding sites to train an enhancer predictive tool on the basis of chromatin signatures<sup>8</sup>. We determined the transcription in each tissue and cell

- Проанализировано 19 тканей и типов клеток мыши
- Найдено около 300 000 цис-регуляторных последовательностей
- Они составляют до 11% генома мыши
- И включают в себя более 70% консервативных не-кодирующих последовательностей

Энхансеры – небольшой участок ДНК, способный связываться с факторами транскрипции

- ▶ действуют опосредованно через белки-активаторы
- ▶ Взаимодействуя с комплексом-медиатором, они активируют РНК-полимеразу II и общие факторы транскрипции, что ведет к транскрибированию генов.
- ▶ Могут располагаться на расстоянии  $> 60$  т.п.о., а так же в интронах.