

Геном

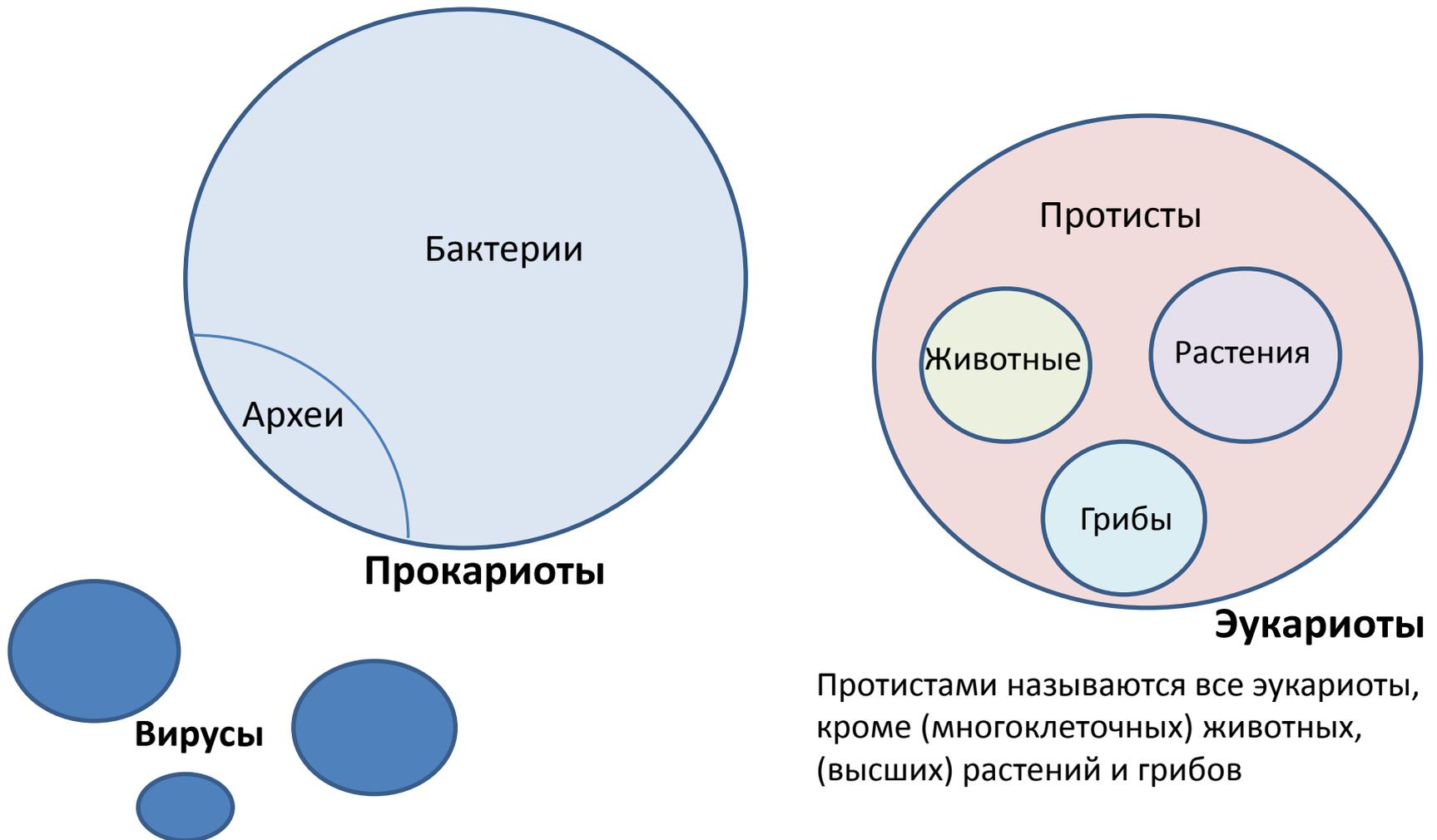
Сергей Александрович Спирин

Институт физико-химической биологии им. А.Н.Белозерского МГУ, Москва

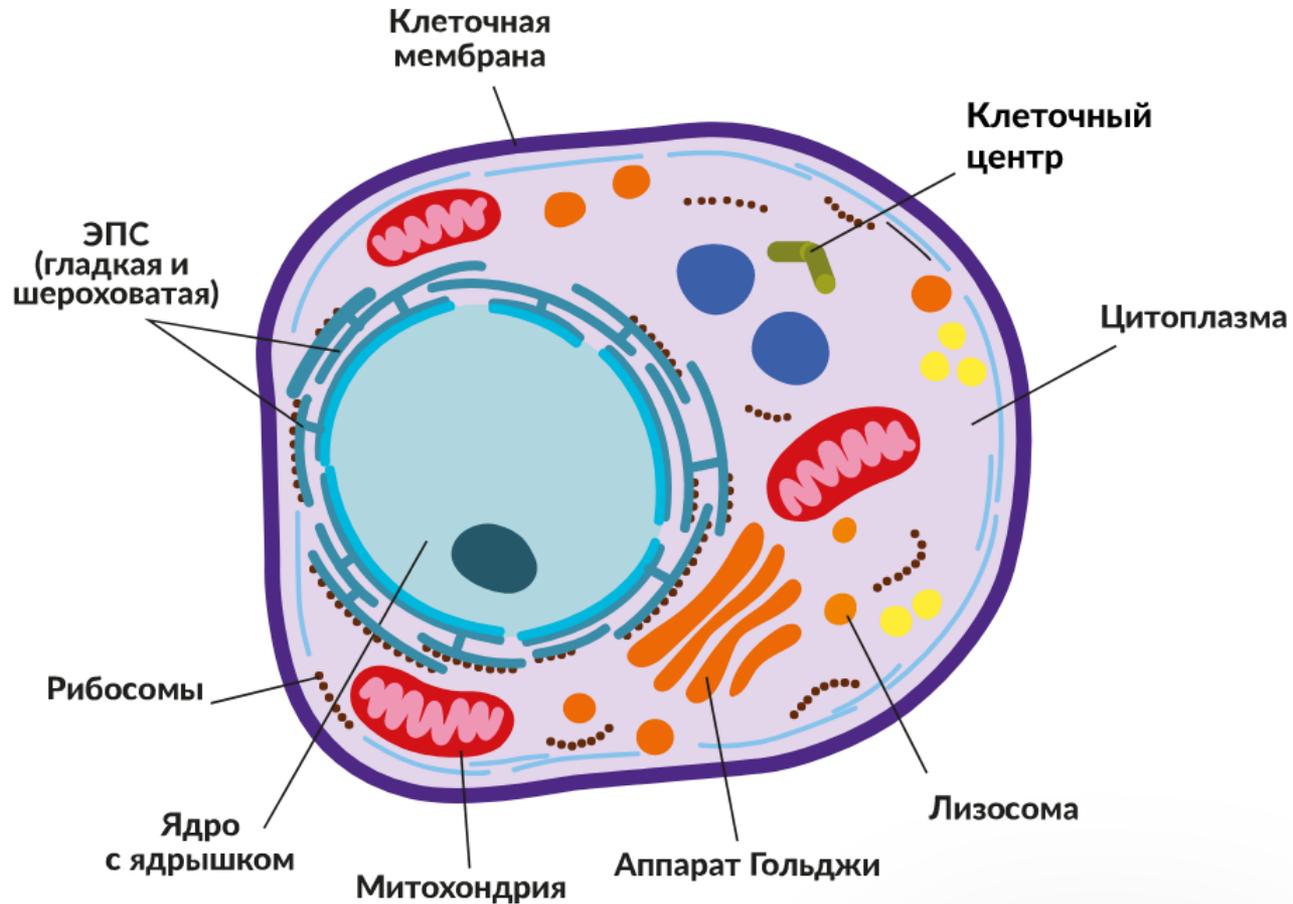
25 февраля 2026

Гено́м — совокупность наследственной информации о живом организме

Отступление: классификация живых организмов

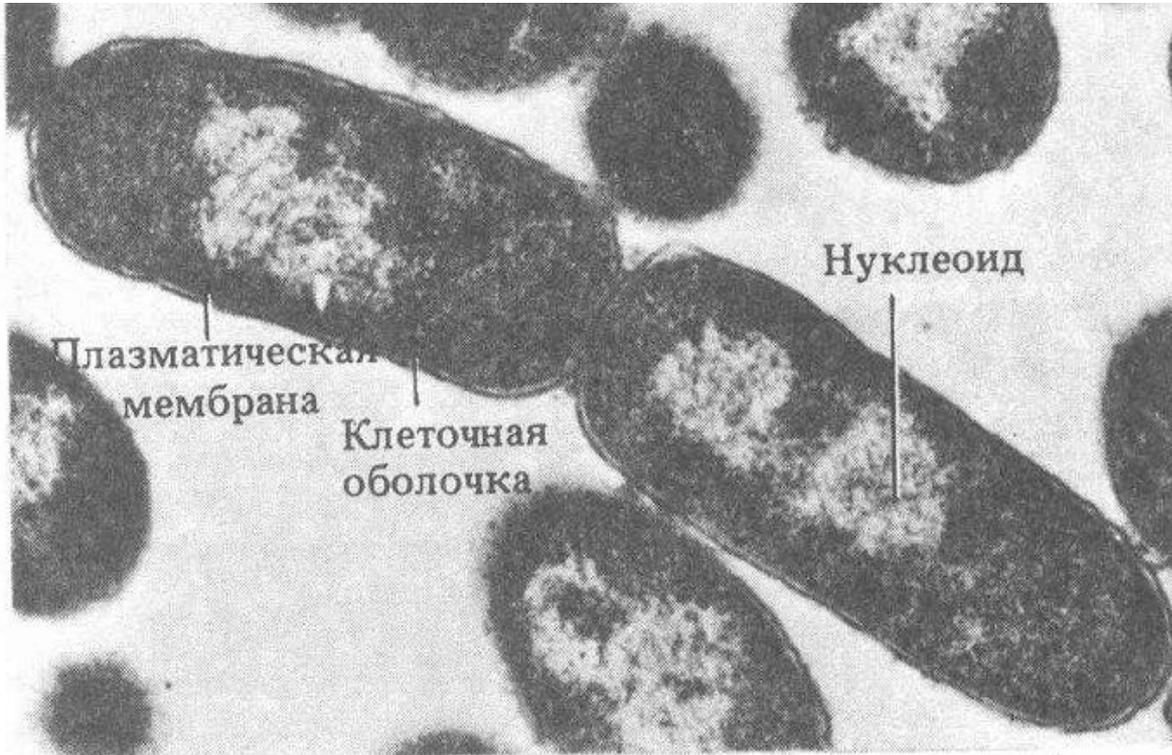


Эукариотическая клетка



<https://obrazavr.ru/biologiya/7-klass-biologiya/vvedenie-v-zoologiyu/znakomstvo-s-mirom-zhivotnyh-vvedenie-v-zoologiyu/kletki-i-tkani-zhivotnyh-organizmov/>

Клетки бактерий

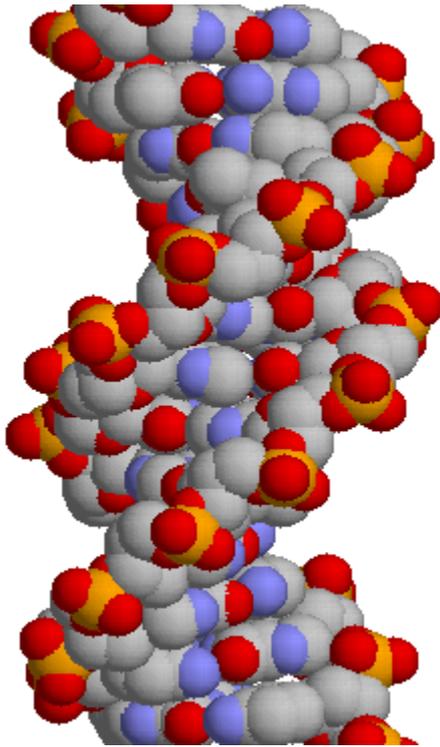


Нуклеоид (т.е. «квази-ядро») состоит в основном из ДНК, с добавлением некоторого количества связанных с ДНК белков

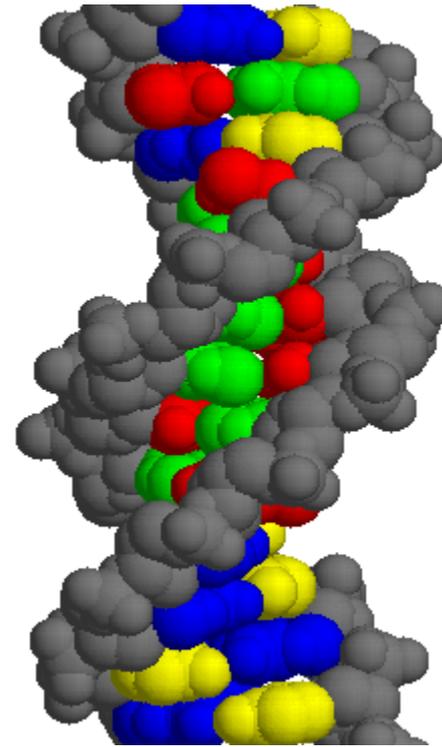
Носители информации

- Носители информации, которой обмениваются люди — это книги, рисунки, песни, разговоры, обычаи, электронные носители, ...
- Носители наследственной информации — нуклеиновые кислоты: ДНК и РНК

ДНК — это молекула



C N O P

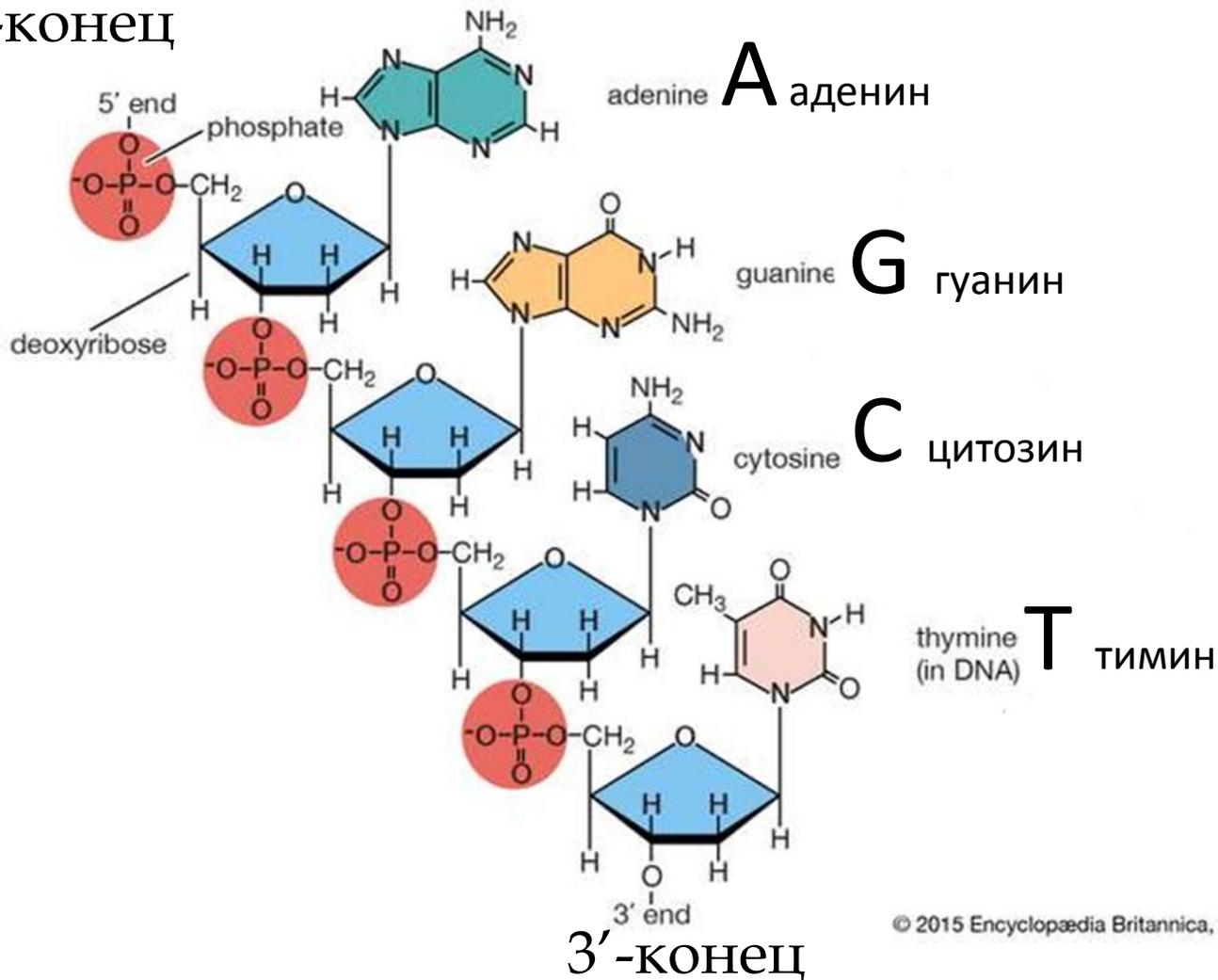


A C G T

В живой клетке ДНК — комплекс из двух молекул, каждая из которых — гетерополимер из мономеров четырёх типов: А, С, G, Т

Химическая формула ДНК

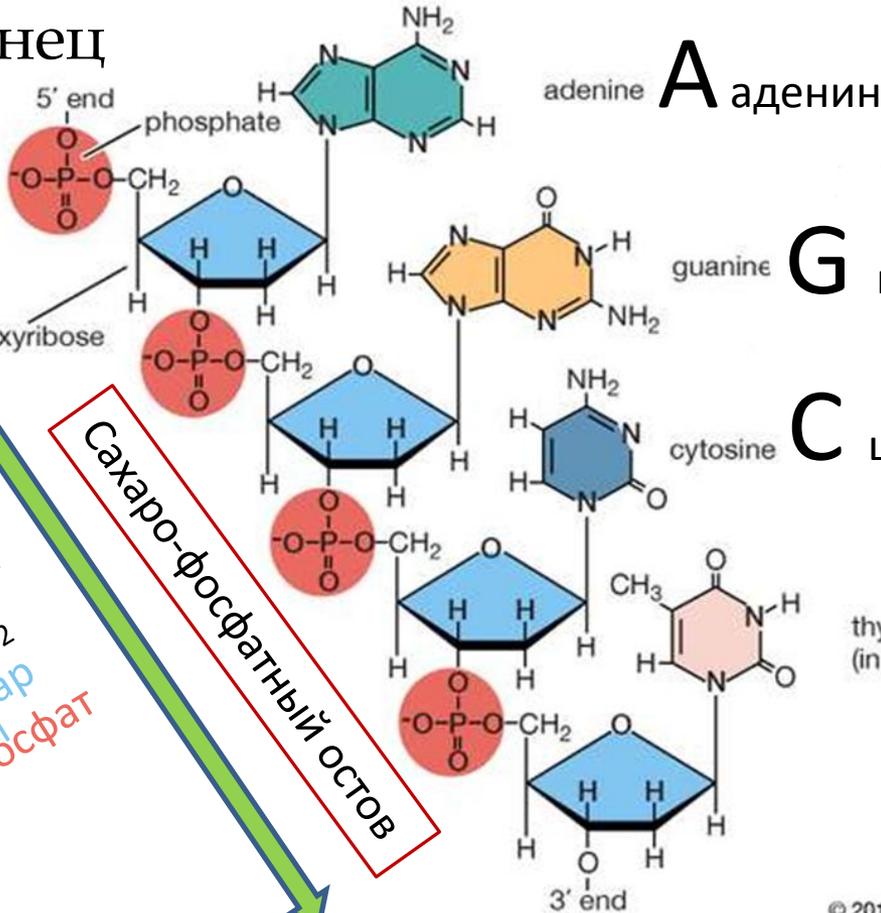
5'-конец



© 2015 Encyclopædia Britannica, Inc.

Формула молекулы ДНК определяется последовательностью букв **A, T, G, C**

5'-конец



Основания ДНК

= **AGCT**

Сахаро-фосфатный остов

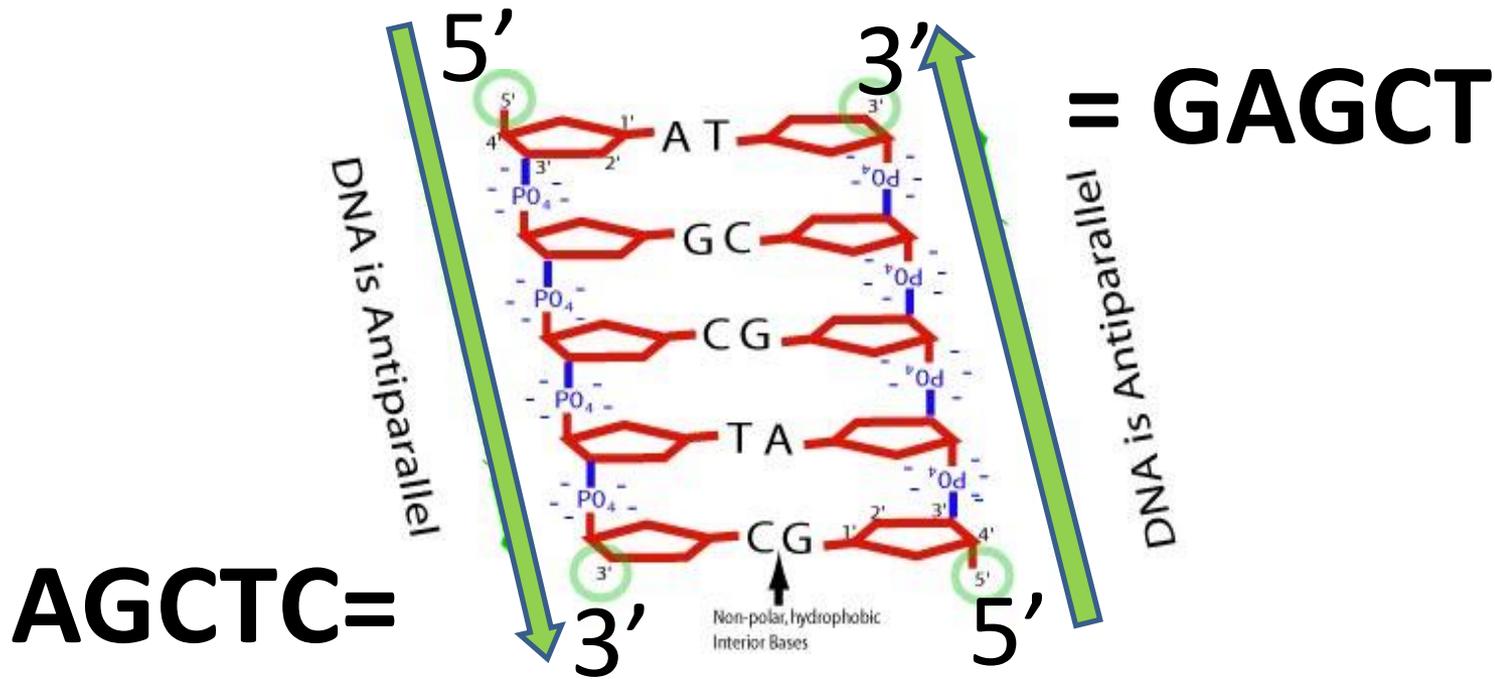
сахар
фосфат
CH₂
сахар
фосфат

3'-конец

© 2015 Encyclopædia Britannica, Inc.

Как определить направление 5' → 3', находясь в середине цепочки?

ДНК состоит из двух **антипараллельных** комплементарных цепочек



Гуанин (G) связан тремя **водородными связями** с цитозином (C)
Аденин (A) связан двумя водородными связями с тиминном (T)

Последовательность одной цепи ДНК

- Принято записывать буквами **A, T, G, C**
- Всегда пишется в направлении от 5'-конца к 3'-концу
- Химическая формула ДНК однозначно определяется последовательностью
- Последовательность несёт всю наследственную информацию.

Последовательность двухцепочечной ДНК

— это последовательность одной из цепей.

Какой?

gatcaacactacttgacttcaagacttaccataaagaaaactatagtgtggtattggcaa
aagacaagacaaatagatcaacataacaaaataaagggccatgaaatagacccat at agt
caattgat t t t t gacaaagaaggattggcaatagaatggggtaagat agtcttctcaac
aaacggtaccagaatgactgaataccacatgcaaaaagaaaaagaaatgaacctagaca
cagatcttatacagttcacaaaaatgtaactcaaaatgaatcatagacctaaatataata
ttcaagactataaaaccctaaaatataacataggggaaaatctaaacaatcttgagtttg
ttaatgact t t t tagatacaataccaaaggcaggatccaggaaagaatcgataagctggg
cttcattaaaattaaaatatttctgctctatgaagccactgtcaagagaaggaaaaggca
agccatagactgggagaaaaatatttacaagaacat acatgat aaaggact at t at ccaa

Последовательность двухцепочечной ДНК

— это последовательность одной из цепочек

Какой?

А всё равно!

Последовательность второй цепочки однозначно
восстанавливается «по комплементарности»

```
gatcaacactacttgacttcaagacttaccataaagaaaactatagtgtggtattggcaa  
aagacaagacaatatagatcaacataacaaaaataaagggccatgaaatagacccat atagt  
caattgatTTTTgacaaagaaggattggcaatagaatggggtaagatagtcttctcaac  
aaacggtaccagaatgactgaataccacatgcaaaaagaaaaagaaatgaacctagaca  
cagatcttatacagttcacaaaaatgtaactcaaaatgaatcatagacctaatataata  
ttcaagactataaaaccctaaaatataacataggggaaaatctaaacaatcttgagtttg  
ttaatgactTTTTtagatacaataccaaaggcaggatccaggaaagaatcgataagctggg  
cttcattaaaattaaaatatttctgctctatgaagccactgtcaagagaaggaaaaggca  
agccatagactgggagaaaaatatttacaagaacat acatgat aaaggact attat ccaa
```

Последовательность ДНК

Последовательность двухцепочечной ДНК можно записать **двумя равноправными способами**

AGCTCT = AGAGCT

(эти две записи описывают одинаковые молекулы)

Упражнение

Последовательность цепи ДНК gatcaacactacttgactt

Напишите последовательность комплементарной цепи

Репликация ДНК

При делении клетки происходит **репликация ДНК**.
Двойная спираль расплетается, и к каждой из двух цепей пристраивается новая комплементарная цепь.
Получаются **две** копии исходной двойной спирали.

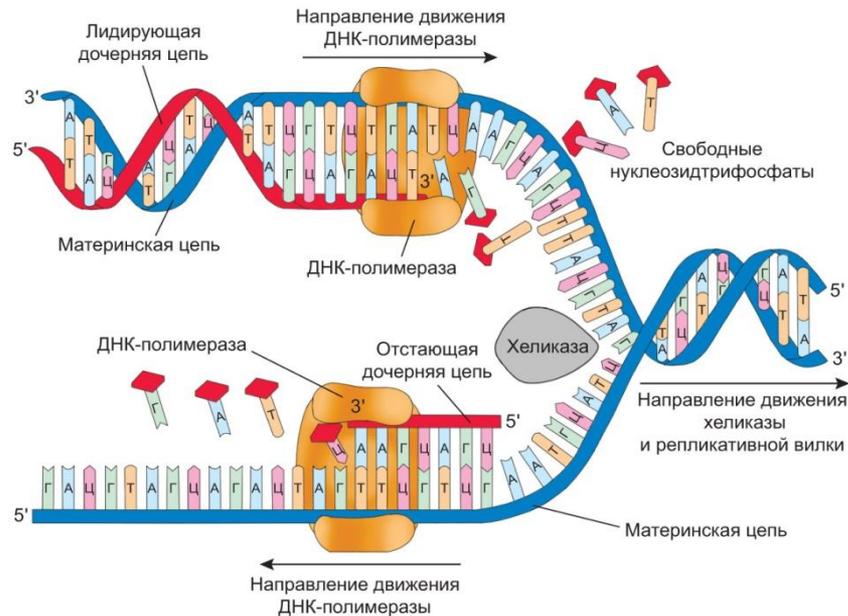


Рис. 16.4. Схема процесса репликации ДНК

Передача информации между поколениями

Благодаря репликации ДНК двух клеток, получившихся при делении, идентична ДНК исходной клетки

Передача информации между поколениями

Благодаря репликации ДНК двух клеток, получившихся при делении, идентична ДНК исходной клетки

... или не совсем идентична:

1. Бывают ошибки репликации (вставляется не тот нуклеотид)
2. Между делениями ДНК может испортиться, в том числе какая-то пара нуклеотидов может замениться из-за ошибок репарации (репарацией называется «починка» неминуемых повреждений)

Можно ли прочитать последовательность ДНК?

- Можно. Но не просто.
 - Секвенаторы (автоматы для прочтения последовательности ДНК) умеют читать только относительно короткие последовательности ДНК (например, 150 п.н.)
 - Другие секвенаторы могут читать больше, но делают много ошибок
 - Задача биоинформатики — из коротких последовательностей составить последовательность всей ДНК
 - Об этом будет лекция

Геном организма с точки зрения биоинформатики — это

1. Набор последовательностей всех ДНК из одного организма. Т.е. число последовательностей равно числу **различных** двойных молекул ДНК в организме.
2. Для **референсного генома** вида (например человека) из немного различающихся последовательностей разных людей берут «самую типичную» (на самом деле первую попавшуюся :)

Бывают различия и внутри организма:

- соматические мутации;
- запланированное многообразие (гены иммуноглобулинов)

Размеры геномов

У **человека**: 3 млрд. пар нуклеотидов (п.н., по-английски base pairs, bp)

У **многоклеточных животных**: от сотен млн. п.н. (круглые черви) до десятков млрд. п.н. (тритоны)

У **высших растений**: от сотен млн. до сотен млрд. п.н.

У **гриб(к)ов**: от 9 млн. до 177 млн. п.н., в среднем около 40 млн. п.н.
Геном дрожжей — около 12 млн. п.н.

У типичных **бактерий**: один-три миллиона п.н. Бывают исключения в обе стороны (от 130 тысяч до 14 миллионов)

У **вирусов**: от двух тысяч до миллиона п.н.

Референсный геном SARS-CoV-2 состоит из 29903 букв.

У многих вирусов (в том числе у коронавирусов) носителем генома является не ДНК, а РНК (немного другая молекула, менее стабильная)

РНК тоже имеет отношение к кодированию информации

- У клеточных организмов (т.е., не вирусов) последовательность любой молекулы РНК является копией участка одной из цепей ДНК
- Ген — это участок генома, с которого копируется функциональная РНК
- С гена белка копируется мРНК, которая служит матрицей для синтеза белка

Информация, записанная в геноме

- **Гены белков** – участки ДНК, кодирующие аминокислотные последовательности белков
- **Гены некодирующих РНК** (т.е., РНК, отличных от мРНК)
- **Регуляторные участки** — сигналы для белков и молекулярных машин, позволяющие регулировать **экспрессию** генов в зависимости от потребностей организма
- Много чего другого ...

Экспрессия и регуляция

Процесс производства белка по «инструкции», записанной в ДНК, называется «экспрессия»

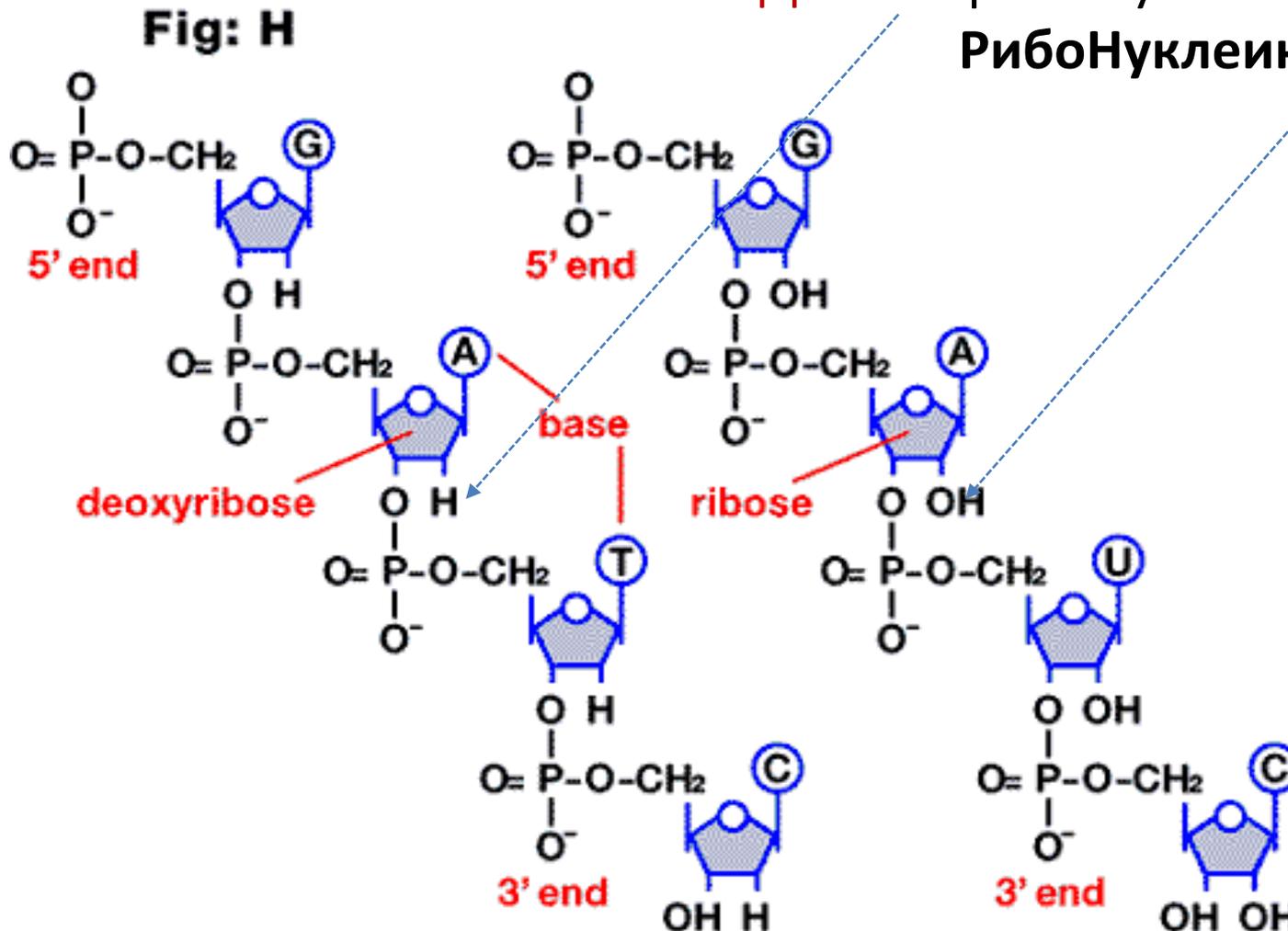
В разных условиях, разных тканях, на разных стадиях развития экспрессируются разные белки.

Инструкции по регуляции экспрессии тоже записаны на ДНК.

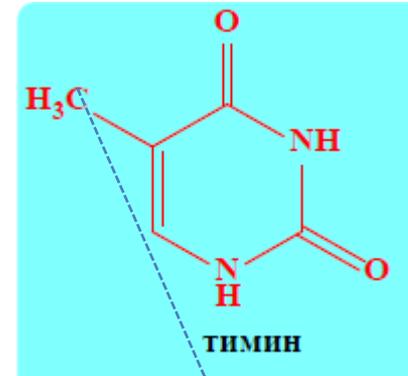
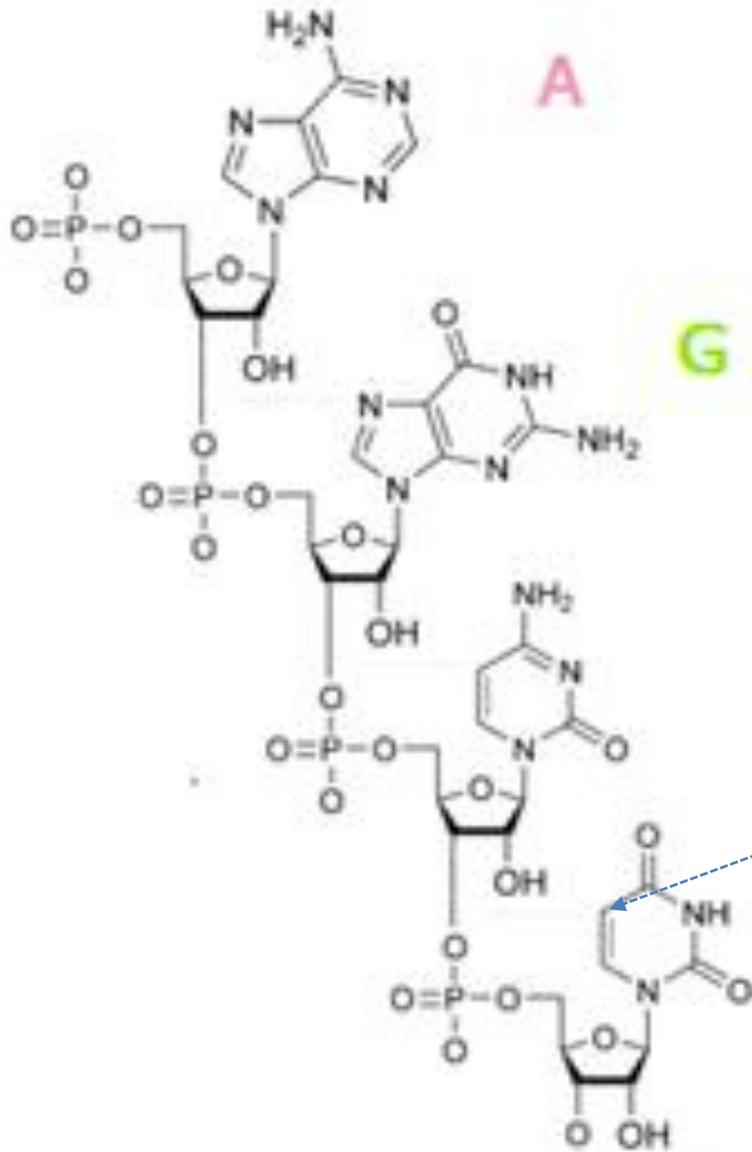
По оценкам, в геноме человека регуляторные участки занимают около 8% генома – примерно в 10 раз больше, чем кодирующие.

Отличия РНК от ДНК

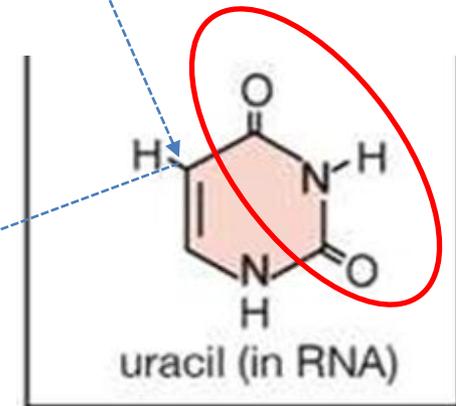
ДезоксирибоНуклеиновая Кислота
РибоНуклеиновая Кислота



Урацил вместо тимина



U урацил
вместо T



U: атомы, участвующие в
образовании водородных
связей те же, что у T

Всего два отличия, а какая разница в биологии!

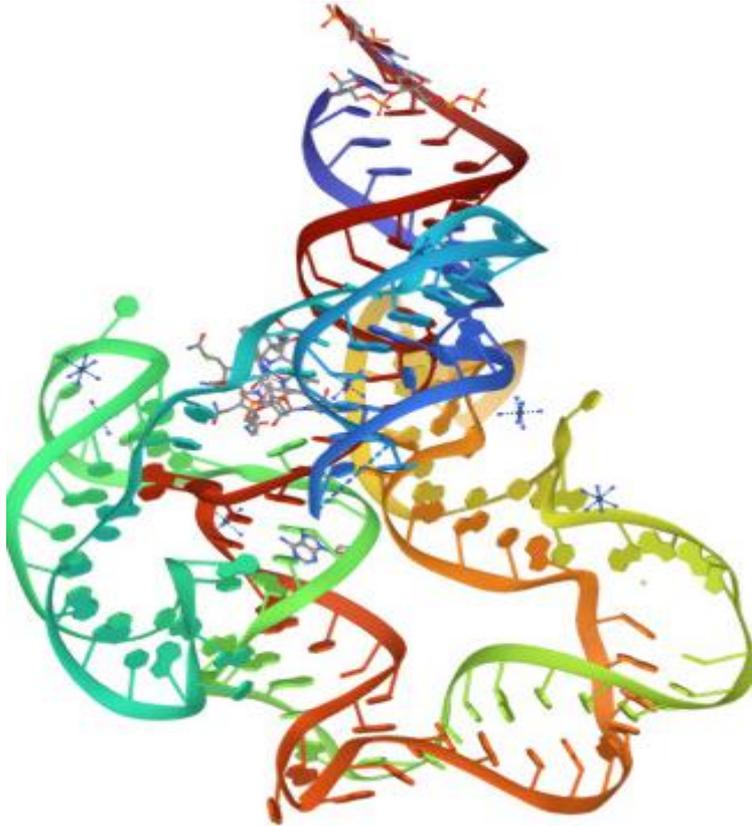
ДНК

- Двухцепочечная, бывает длинной (до сотен млн п.н.), линейной или кольцевой
- Носитель генома у клеточных организмов и многих вирусов

РНК

- Обычно одноцепочечная, линейная и не очень длинная
- Сложная 3D структура
- Разные РНК имеют разные функции
- Последовательность любой РНК клеточного организма представляет собой копию какого-либо участка последовательности ДНК
- Носитель генома у некоторых вирусов

Пространственная структура РНК отличается от таковой у ДНК



PDB [4GXY](#)

Рентгеноструктурная расшифровка 3D структуры РНК, регулирующей экспрессию некоторых генов бактерии. Схематическое изображение, построенное на основе расшифровки координат всех атомов

Видны участки двойных спиралей

Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch.

[Peselis, A.](#), [Serganov, A.](#)

(2012) Nat Struct Mol Biol **19**: 1182-1184

•PubMed: [23064646](#) [Search on PubMed](#)

•DOI: [10.1038/nsmb.2405](#)

•Primary Citation of Related Structures:

[4GXY](#)

•

PubMed Abstract:

Coenzyme B(12) has a key role in various enzymatic reactions and controls expression of bacterial genes through riboswitches. Here we report the crystal structure of the *Symbiobacterium thermophilum* B(12) riboswitch bound to its ligand adenosylcobalamin. The riboswitch forms a unique junctional structure with a large ligand-binding pocket tailored for specific recognition of the adenosyl moiety and flanked by structural elements that stabilize the regulatory region and enable control of gene expression.

Генетический код

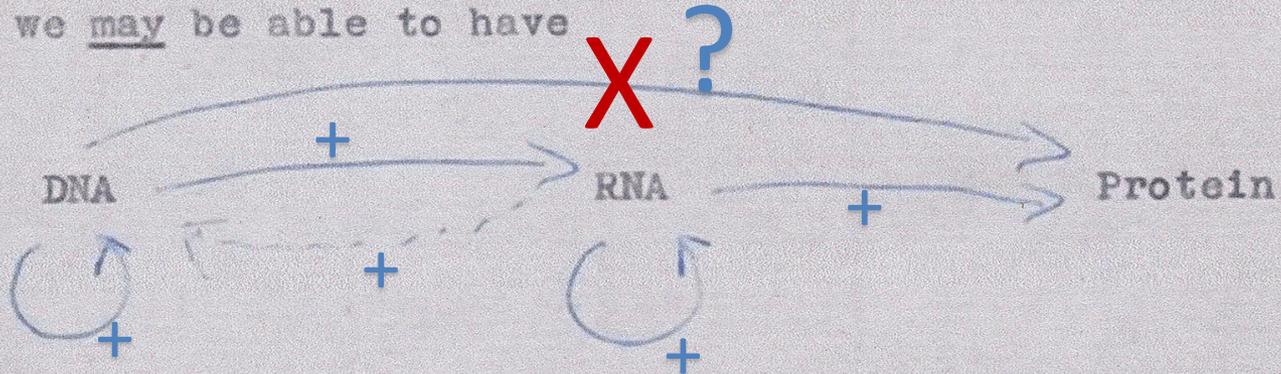
	T(U)	C	A	G
T(U)	TTT Phe TTC Phe TTA Leu TTG Leu	TCT Ser TCC Ser TCA Ser TCG Ser	TAT Tyr TAC Tyr TAA stop TAG stop	TGT Cys TGC Cys TGA stop TGG Trp
C	CTT Leu CTC Leu CTA Leu CTG Leu	CCT Pro CCC Pro CCA Pro CCG Pro	CAT His CAC His CAA Gln CAG Gln	CGT Arg CGC Arg CGA Arg CGG Arg
A	ATT Ile ATC Ile ATA Ile ATG Met	ACT Thr ACC Thr ACA Thr ACG Thr	AAT Asn AAC Asn AAA Lys AAG Lys	AGT Ser AGC Ser AGA Arg AGG Arg
G	GTT Val GTC Val GTA Val GTG Val	GCT Ala GCC Ala GCA Ala GCG Ala	GAT Asp GAC Asp GAA Glu GAG Glu	GGT Gly GGC Gly GGA Gly GGG Gly

Аминокислоты

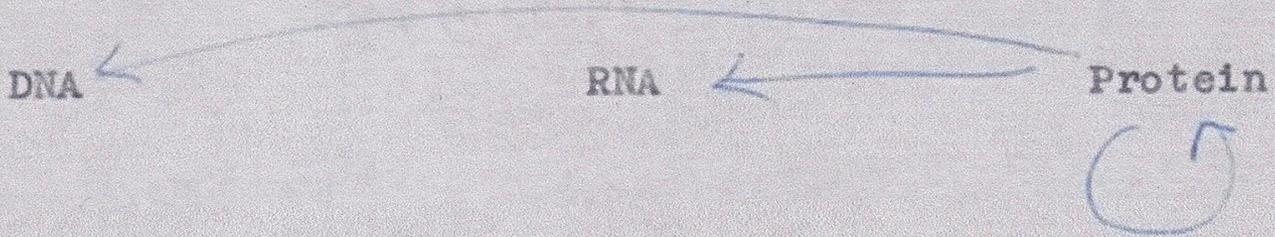
A Ala Alanine Аланин
 R Arg Arginine Аргинин
 N Asn Asparagine Аспарагин
 D Asp Aspartic Acid Аспарагиновая кислота
 C Cys Cysteine Цистеин
 Q Gln Glutamine Глютамин
 E Glu Glutamic Acid Глутаминовая кислота
 G Gly Glycine Глицин
 H His Histidine Гистидин
 I Ile Isoleucine Изолейцин
 L Leu Leucine Лейцин
 K Lys Lysine Лизин
 M Met Methionine Метионин
 F Phe Phenylalanine Фенилаланин
 P Pro Proline Пролин
 S Ser Serine Серин
 T Thr Threonine Треонин
 W Trp Thryptophan Триптофан
 Y Tyr Tyrosine Тирозин
 V Val Valine Валин
 "stop" в таблице кода означает
 стоп-кодон – сигнал окончания трансляции.

Crick's first outline of the central dogma, from an unpublished note made in 1956.

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it. That is, we may be able to have

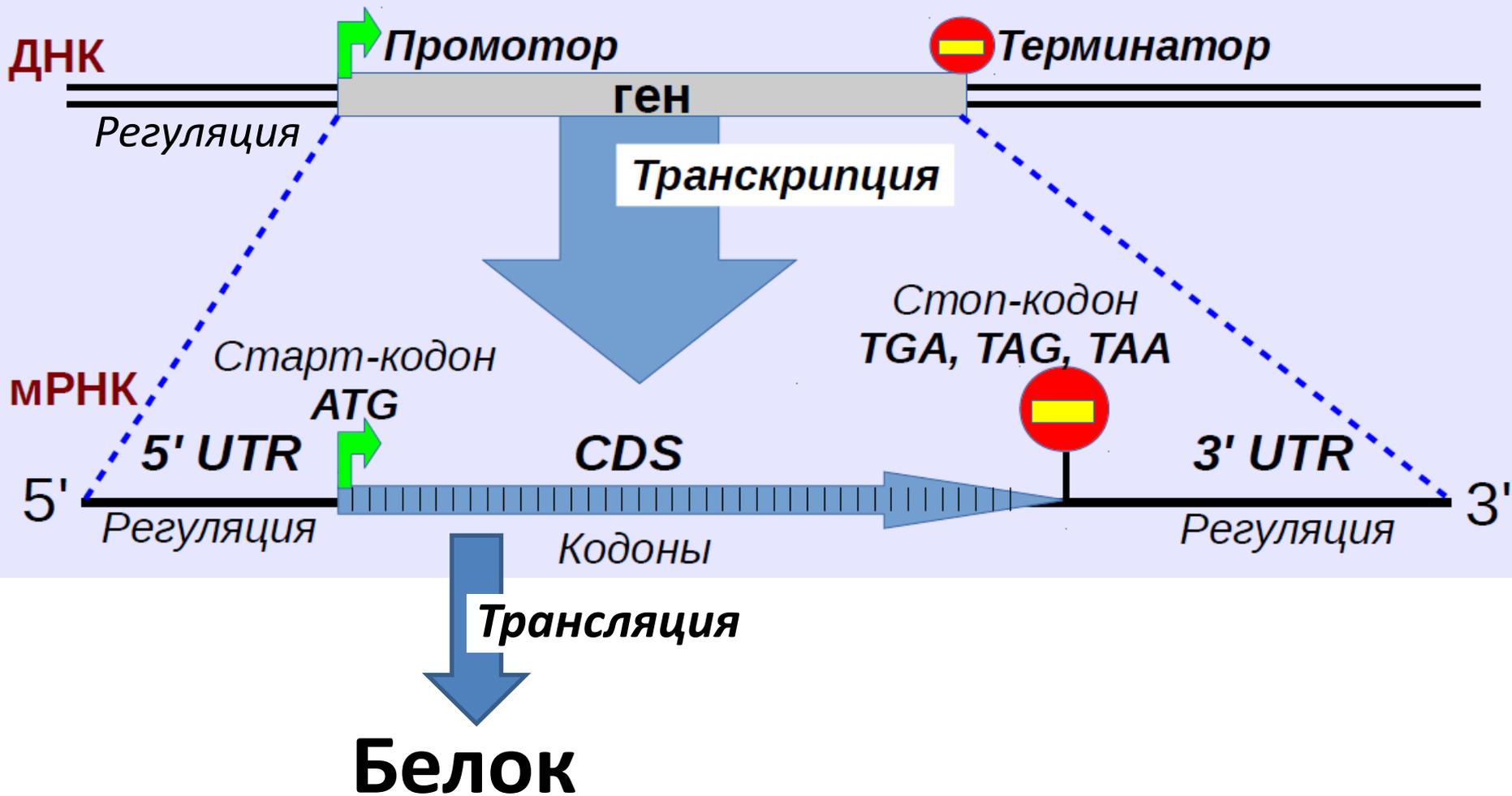


but never

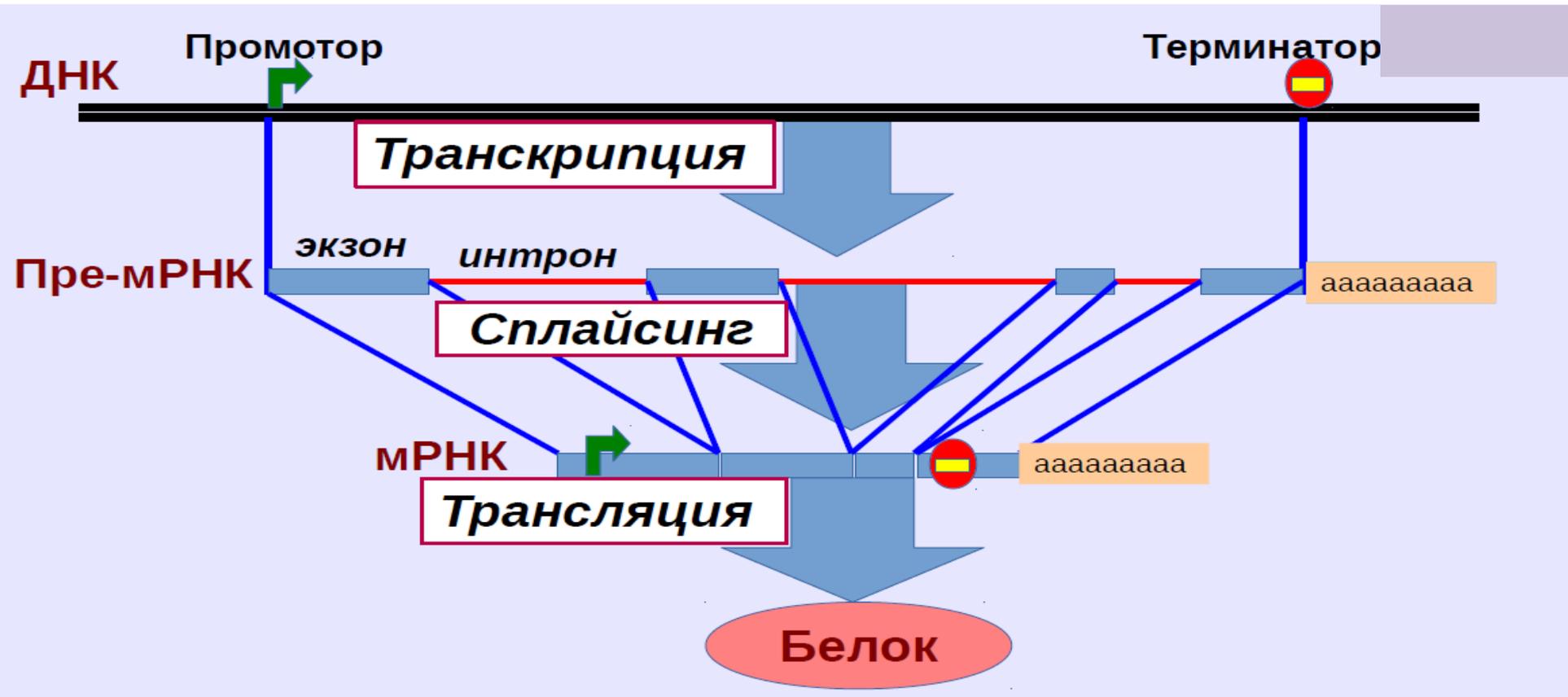


where the arrows show the transfer of information.

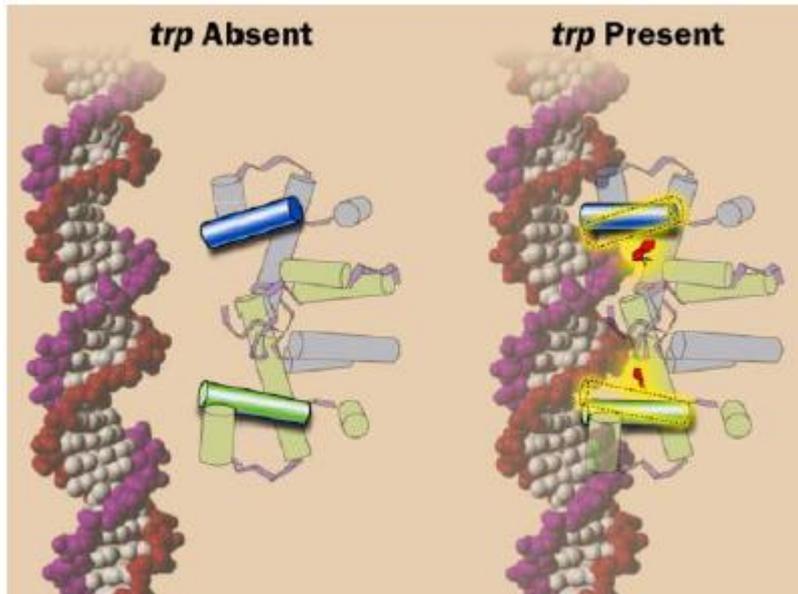
Синтез белка у прокариот



Синтез белка у эукариот



Регуляция у бактерий (пример)

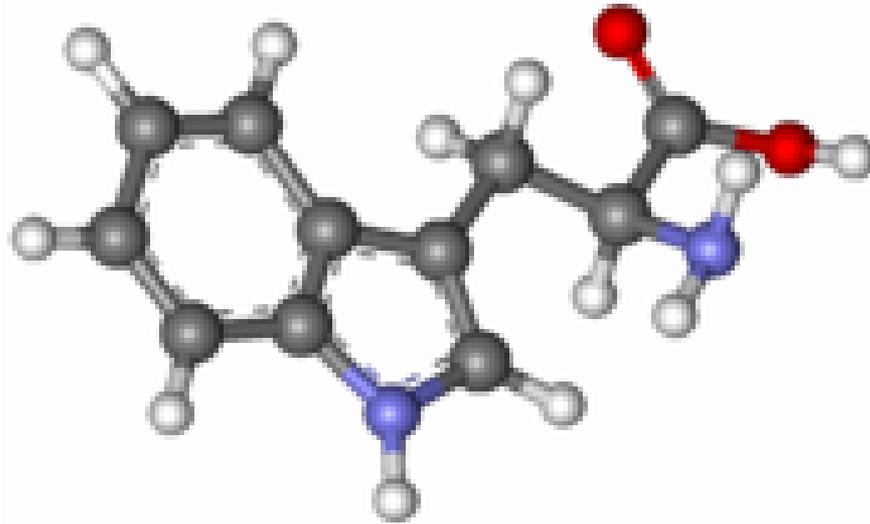


Белок **триптофановый репрессор** регулирует синтез белков – ферментов синтеза триптофана.

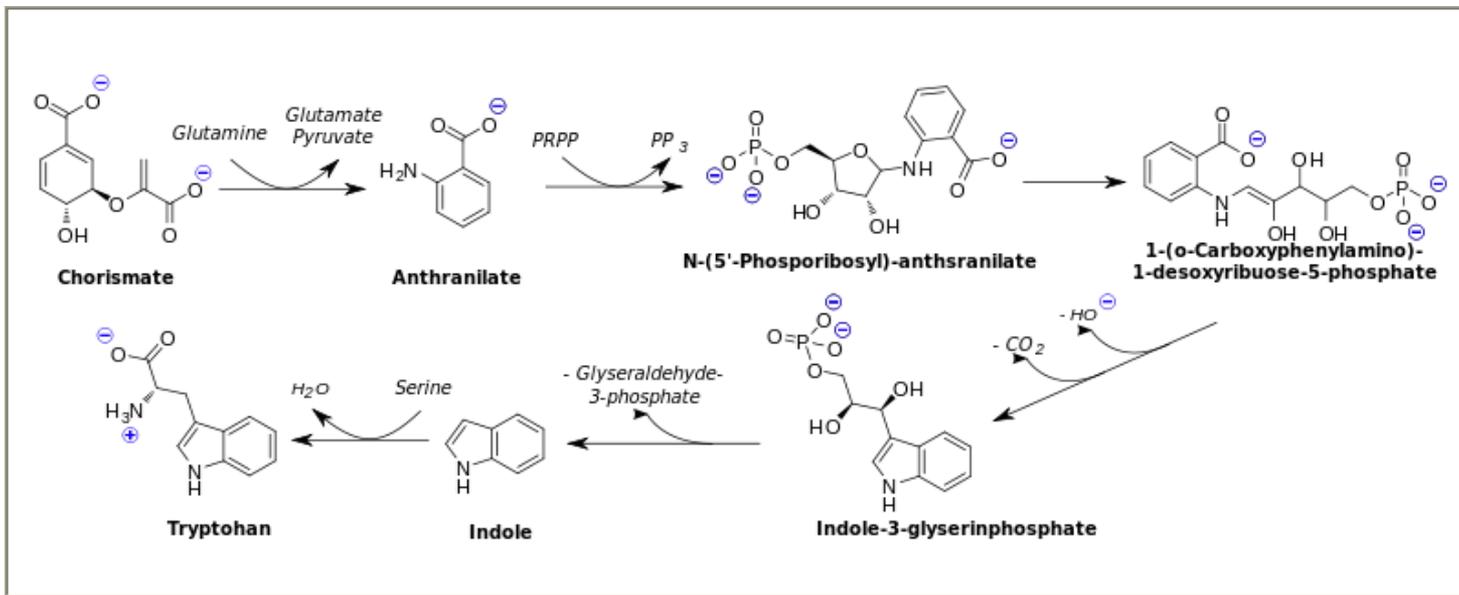
Когда триптофана много, он связывается с репрессором, после чего репрессор получает возможность связаться с **триптофановым оператором** — участком ДНК перед генами ферментов. Пока репрессор сидит на операторе, синтез РНК (а следовательно и ферментов) не происходит.

Картинка взята из презентации https://www.uvm.edu/~dstratto/bcor011/x2012/32_operons.pdf
Там же можете почитать некоторые подробности.

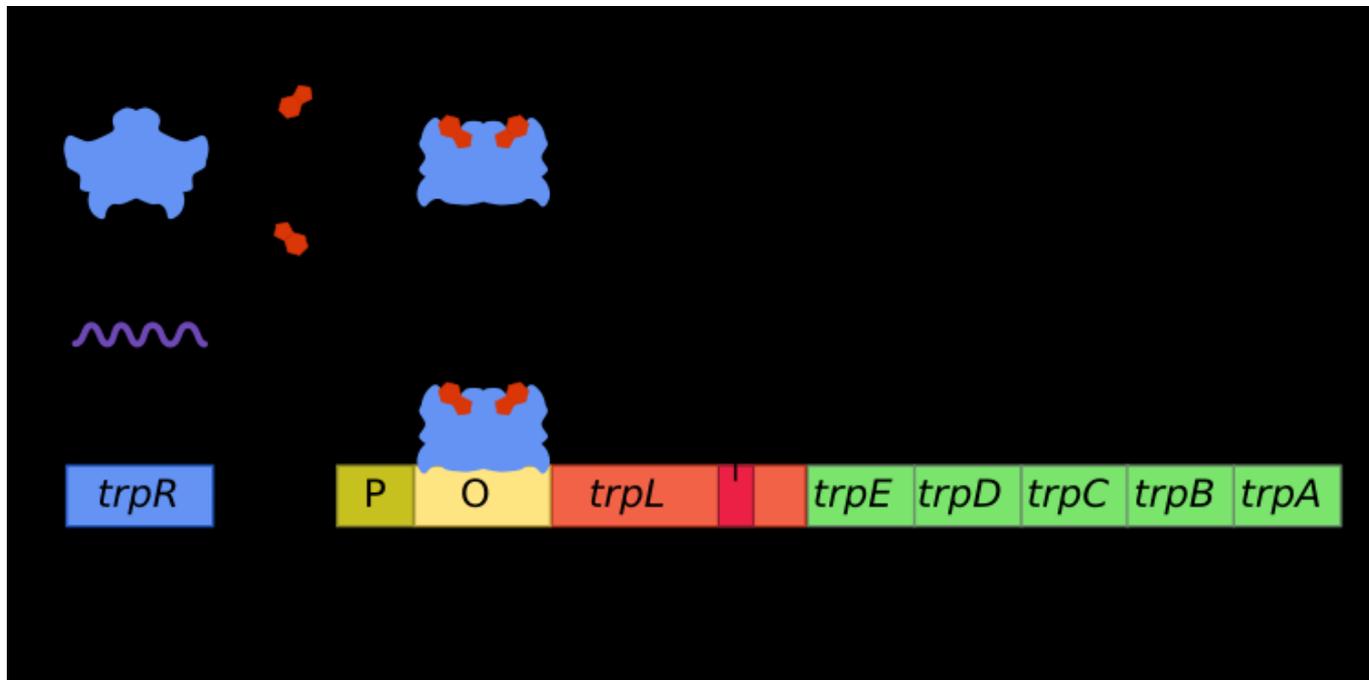
Триптофан — одна из аминокислот, из которых синтезируются белки



Биосинтез триптофана



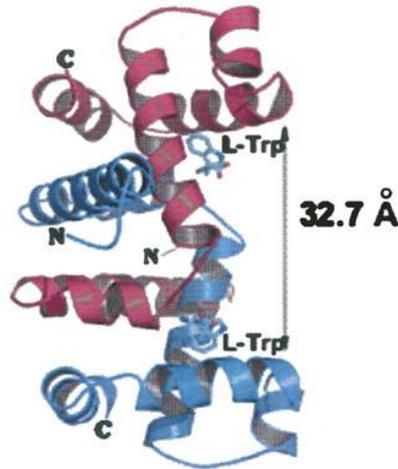
Триптофановый оперон



Белок-регулятор (транскрипционный фактор)



trp aporepressor
(no Trp)



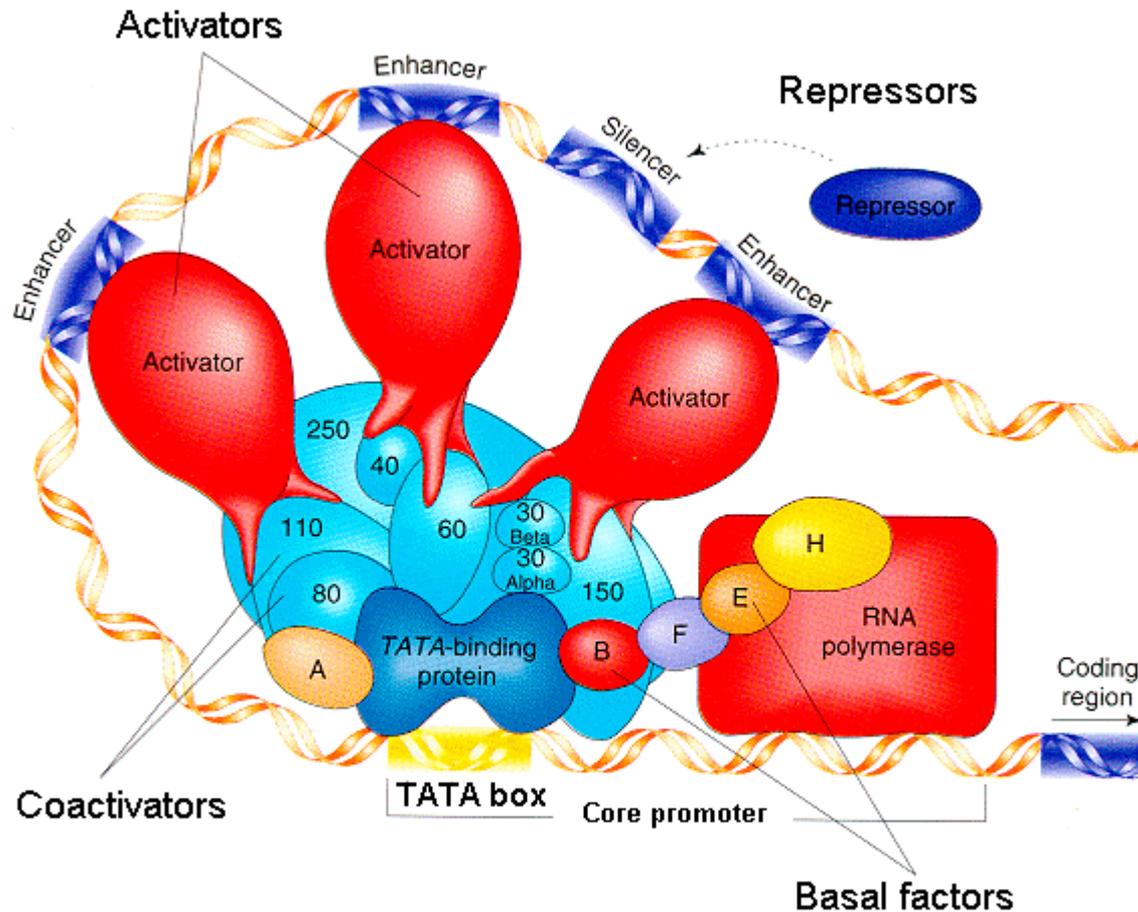
trp repressor
(with Trp)



trp repressor-
operator complex

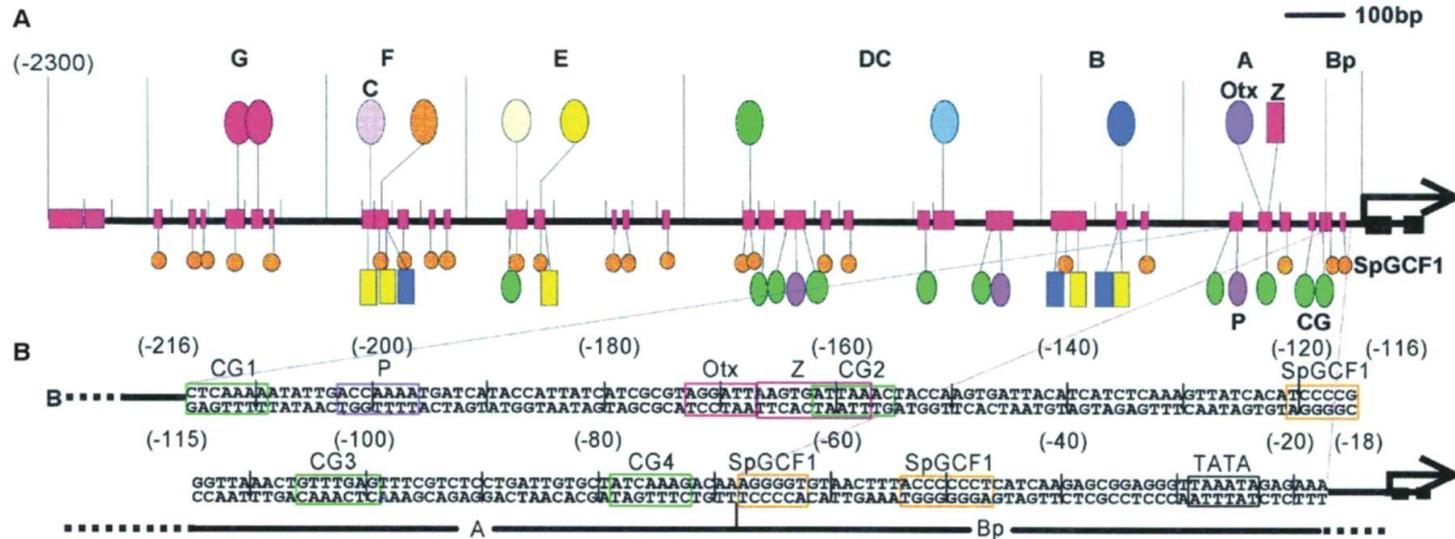
Белки-регуляторы связываются с «регуляторными сайтами» — участками ДНК с определённой последовательностью — и включают или выключают работу соседних генов.

Регуляция у эукариот устроена примерно так...



Eukaryotic transcription initiation, from biology.kenyon.edu (after Tjian)

Регуляция одного гена у животного



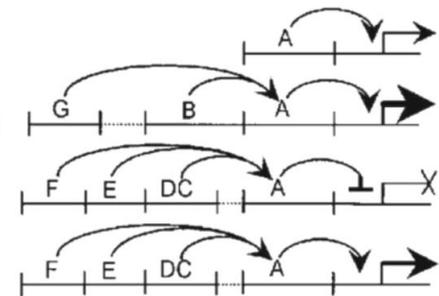
С Функции модуля А:

Экспрессия в вегетативных бляшках на ранних стадиях

Синергизм с модулями В и G - усиление экспрессии в энтодерме на поздних стадиях

Репрессия в энтодерме (модули E и F) и скелетогенной мезенхиме (модуль DC)

Модули E, F и DC при обработке LiCl



Нарушение регуляции может приводить к драматическим последствиям



В результате мутации в регуляторном участке у мухи-дрозофилы вместо усиков выросли ноги...

Человеческий геном

Референсный геном человека

- двадцать две аутосомы
- две половые хромосомы X и Y
- митохондриальная хромосома

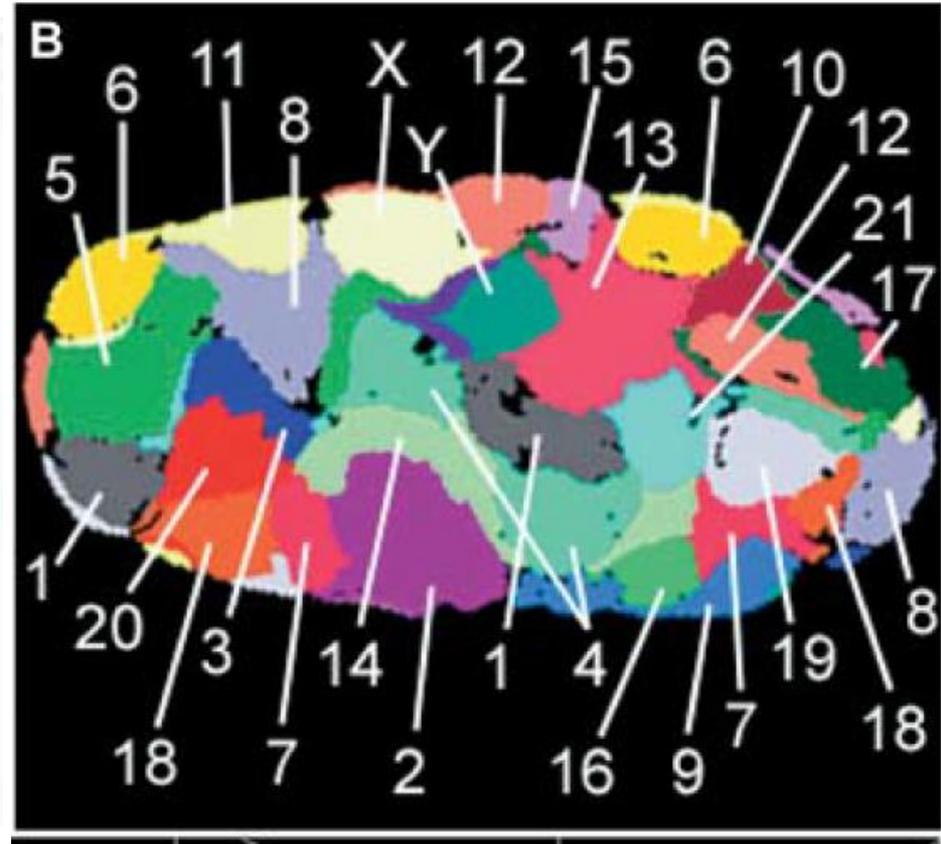
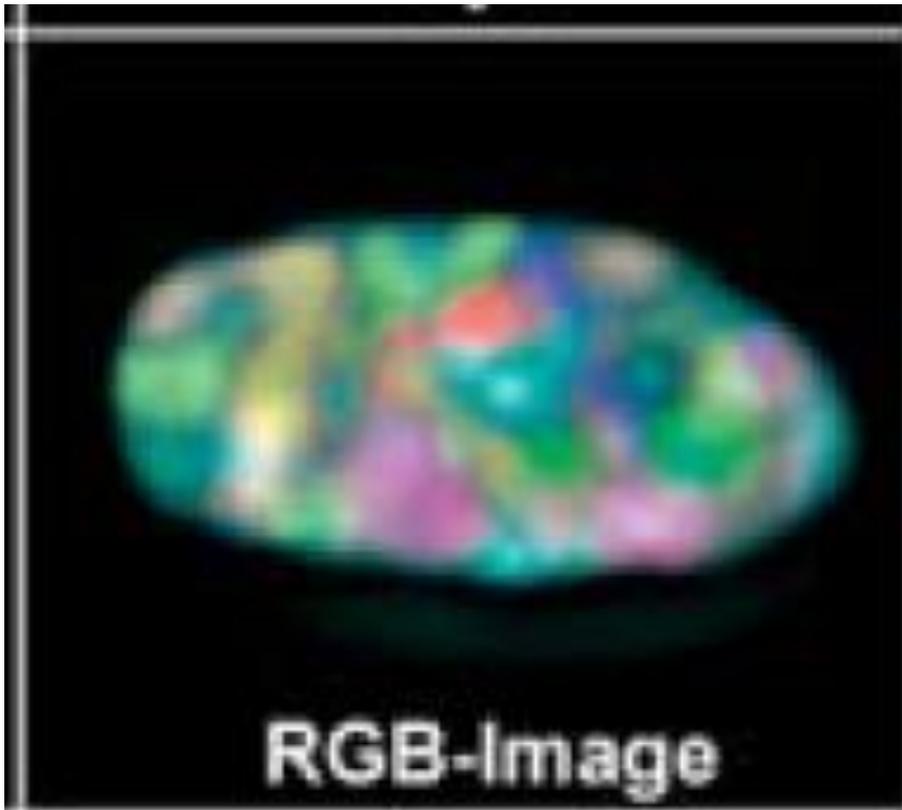
Всего около 3 миллиардов букв А, Т, G, С
в 25 последовательностях

В ядре каждой клетки содержится такой набор хромосом:

М 22 аутосомы и X от мамы, 22 аутосомы и Y от папы = 46

Ж 22 аутосомы и X от мамы, 22 аутосомы и X от папы = 46

ДНК в ядре клетки человека. Разные ДНК покрашены в разные цвета, одинаковые по последовательности (>99%) – в одинаковые цвета



Bolzer A et al. Three-dimensional maps of all chromosomes in fibroblast nuclei and prometaphase rosettes. PLoS Biol. 2005

Как получена микрофотография

К каждой ДНК подобраны многочисленные пробы – кусочки ДНК, полностью комплементарные участку данной ДНК.

Probe size can range from few kb to megabases (Mb), depending on the application.

К концам проб присоединены флюорофоры одного из 7 цветов.

Пропорции проб к одной ДНК с флюорофорами разных цветов подобраны специально, чтобы различать разные ДНК.

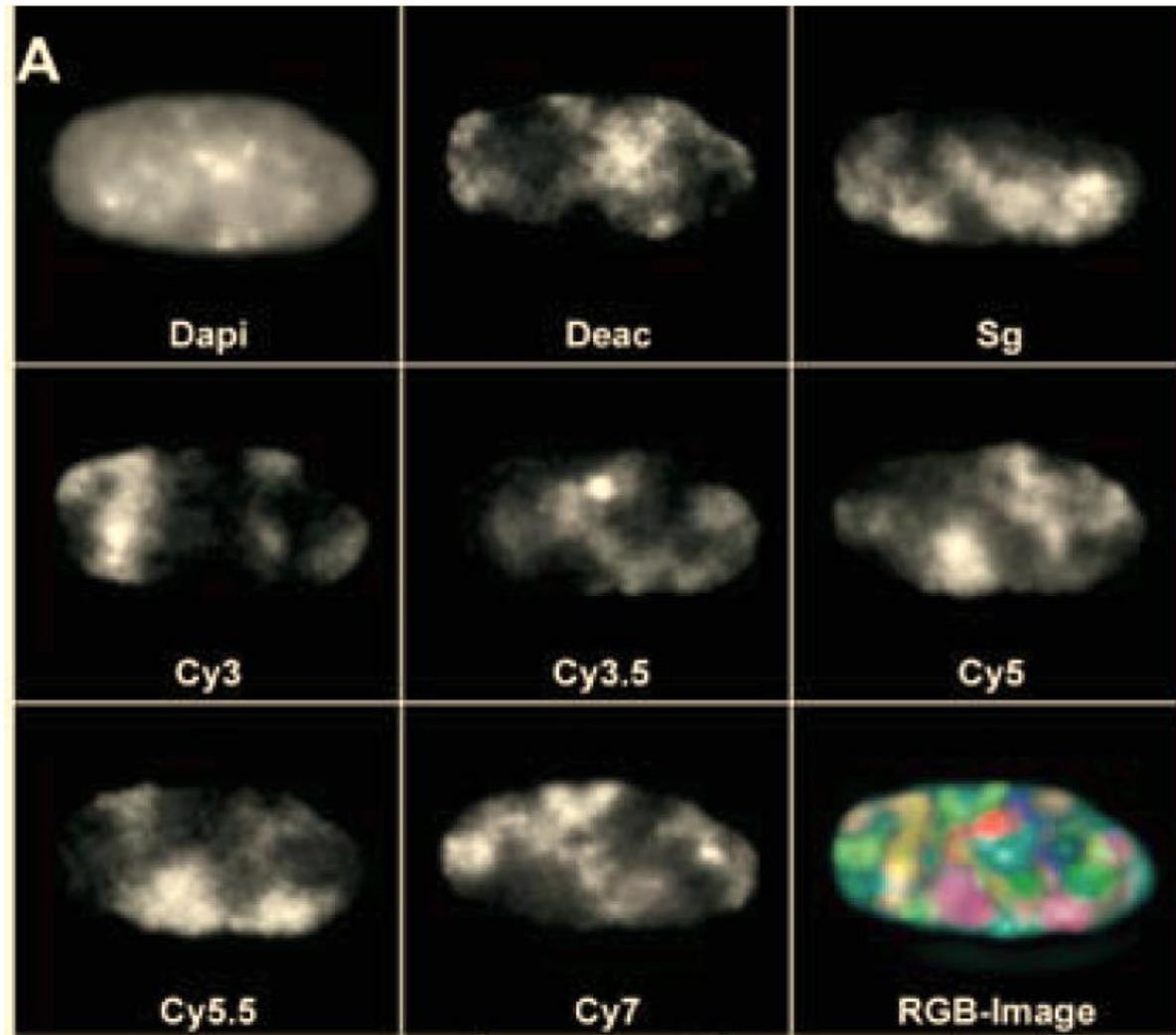
(А) Деконволюция (объединение) флюоресцентных микрофотографий в семи каналах

(one channel for DAPI (DNA counterstain and seven channels for the following fluorochromes: diethylaminocoumarin (Deac), Spectrum Green (SG), and the cyanine dyes Cy3, Cy3.5, Cy5, Cy5.5, and Cy7)

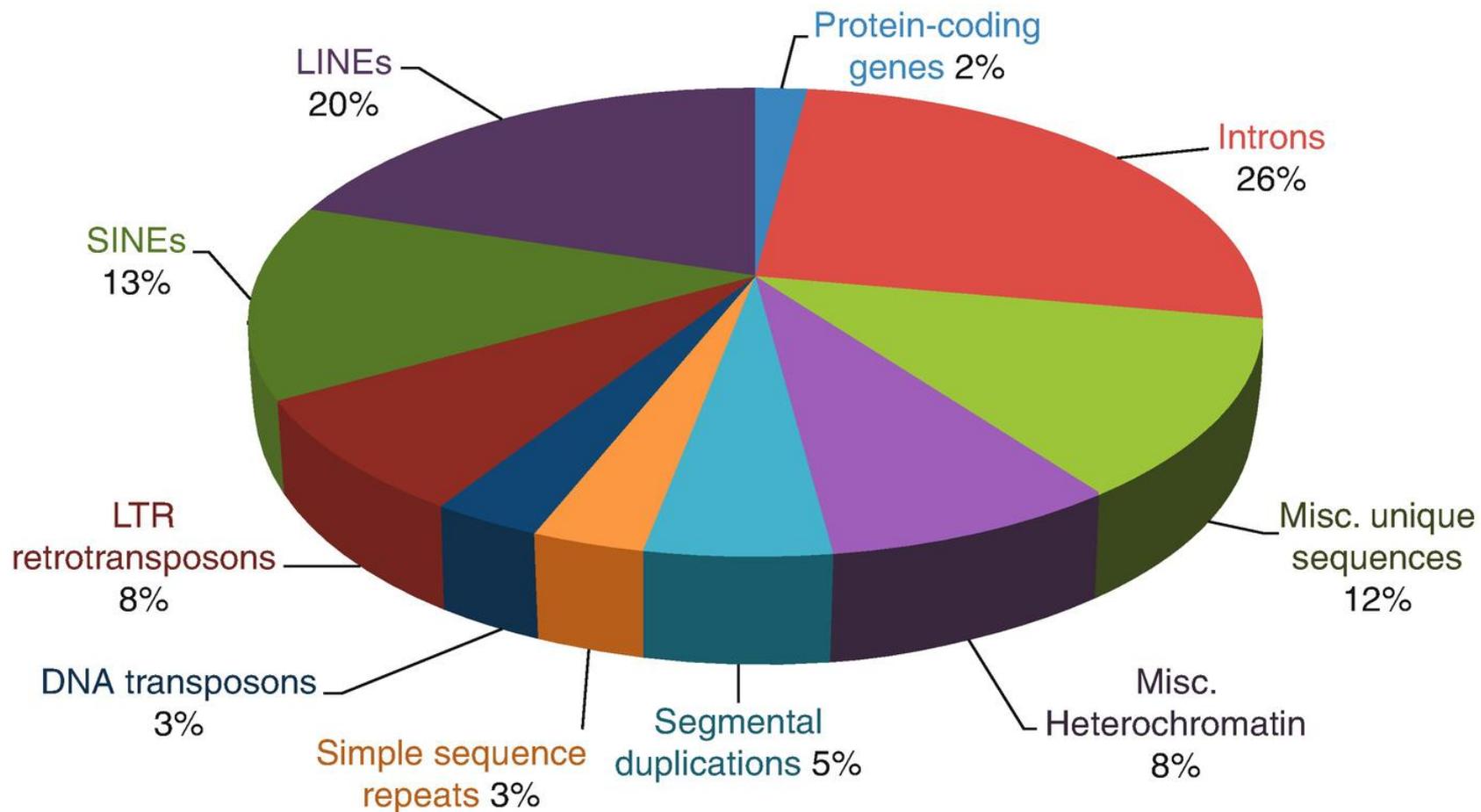
Окраска изображения разных хромосом (1–22, X, and Y) в 24 цвета получена наложением семи каналов

(В) Прорисовка изображения с искусственно подобранными цветами

Исходные микрофотографии в восьми каналах и их совмещение



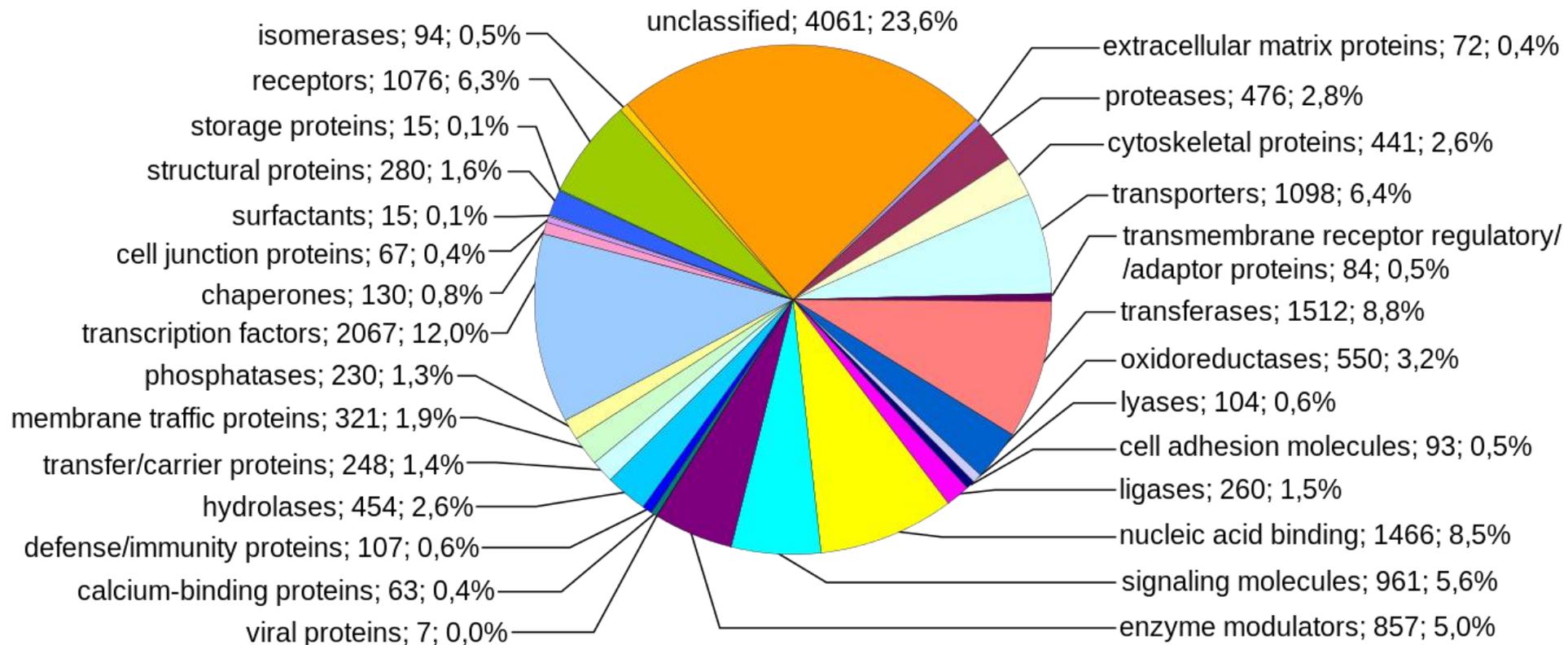
Человеческий геном



https://link.springer.com/chapter/10.1007/978-3-030-73151-9_1/figures/1

Белков у человека – десятки тысяч

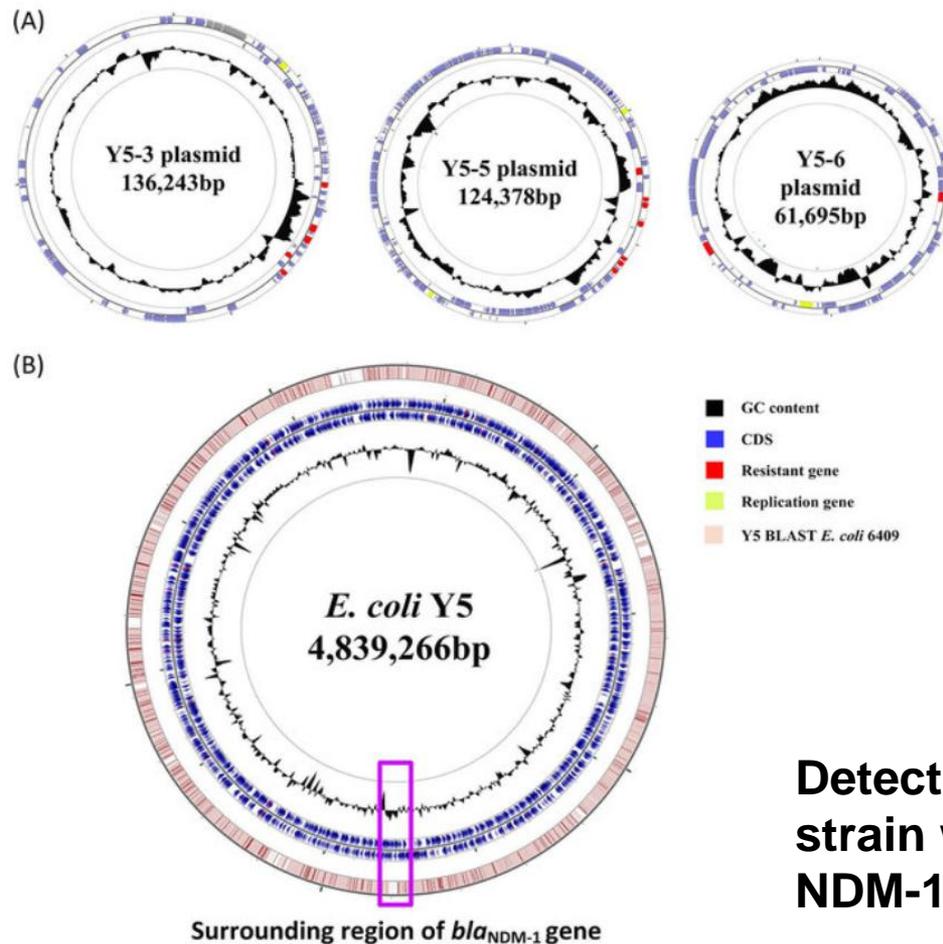
Всего последовательностей белков у каждого млекопитающего около 20 000, у типичной бактерии – одна-две тысячи, у вирусов от одного до сотни.



Геном бактерии

- Одна или несколько **кольцевых** двухцепочечных молекул ДНК – хромосом
- Размер хромосомы порядка миллионов п.н.
- Часто ещё несколько относительно маленьких кольцевых ДНК, называемых плазмидами.
- Родственные бактерии обмениваются плазмидами и фрагментами хромосом. Для эволюции – это способ обмена генетическим материалом. Пример – обмен плазмидами, несущими устойчивость к антибиотикам

Геном кишечной палочки E.coli штамм Y5. Одна хромосома и три плазмиды



Figure

Caption

FIG 1 Circular maps of the *E. coli* Y5 genome and its plasmids. (A) Circular graphs of three plasmids. (B) Circular graph of the Y5 genome sequence and genome alignment. Blue arrows denote coding sequences, red arrows denote resistance genes, and replication genes are denoted by green arrows. Genome alignment between Y5 ... [Read more](#)

This figure was uploaded by [Zhi Ruan](#)

Content may be subject to copyright.

Detection of an Escherichia coli ST167 strain with two tandem copies of *bla* NDM-1 encoded in the chromosome

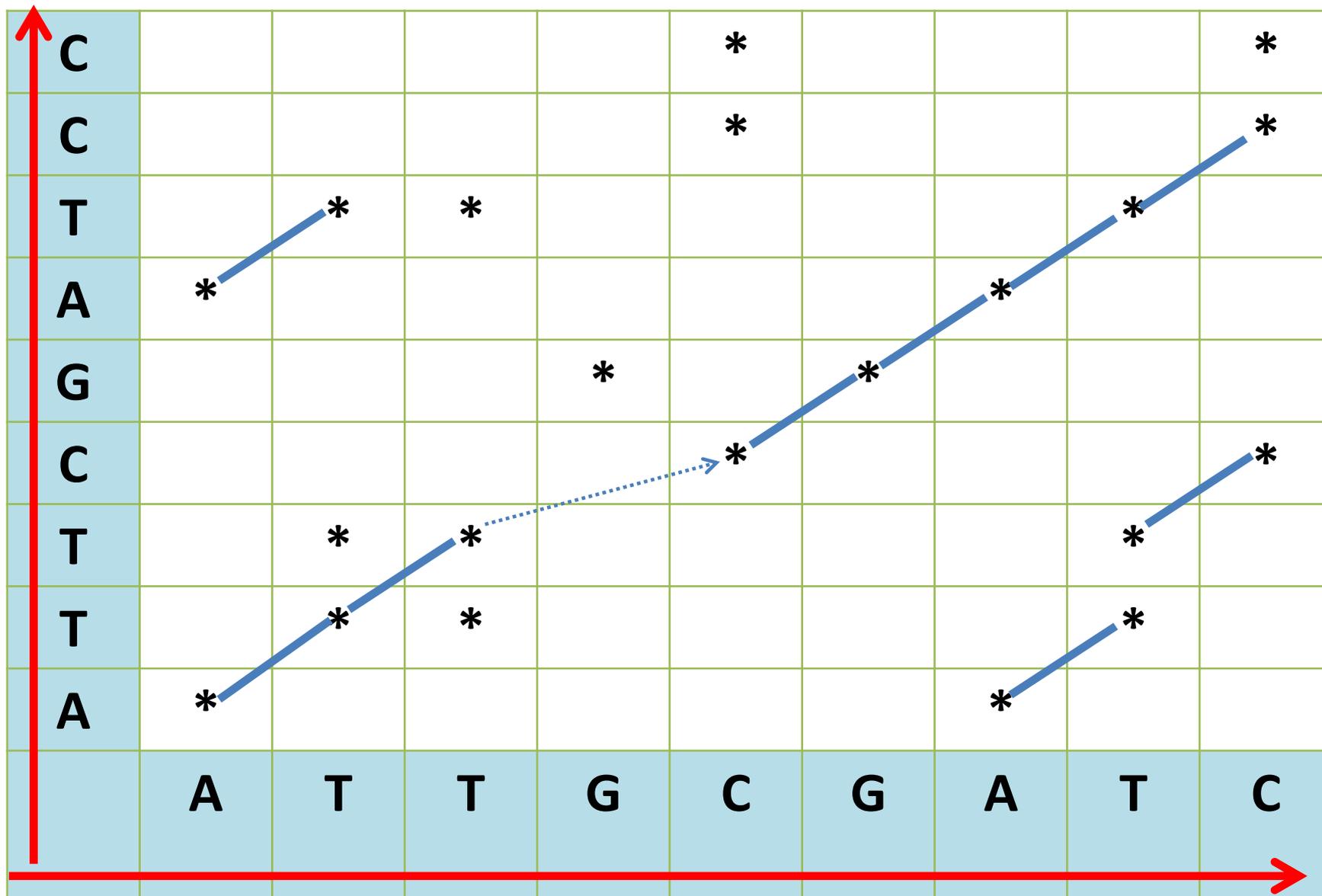
Эволюция

- Геном потомка отличается немножко от генома родителей
- Это происходит из-за случайных мутаций в ДНК
- Мутации бывают точечными (например, вместо С у предка стоит Т у потомка) и крупными (например, потеря сотни нуклеотидов у потомка)
- У эукариот отличие генома потомка получается также за счет получения ДНК и от папы и от мамы
- Вредные и слабо вредные мутации теряются в поколениях из-за отбора

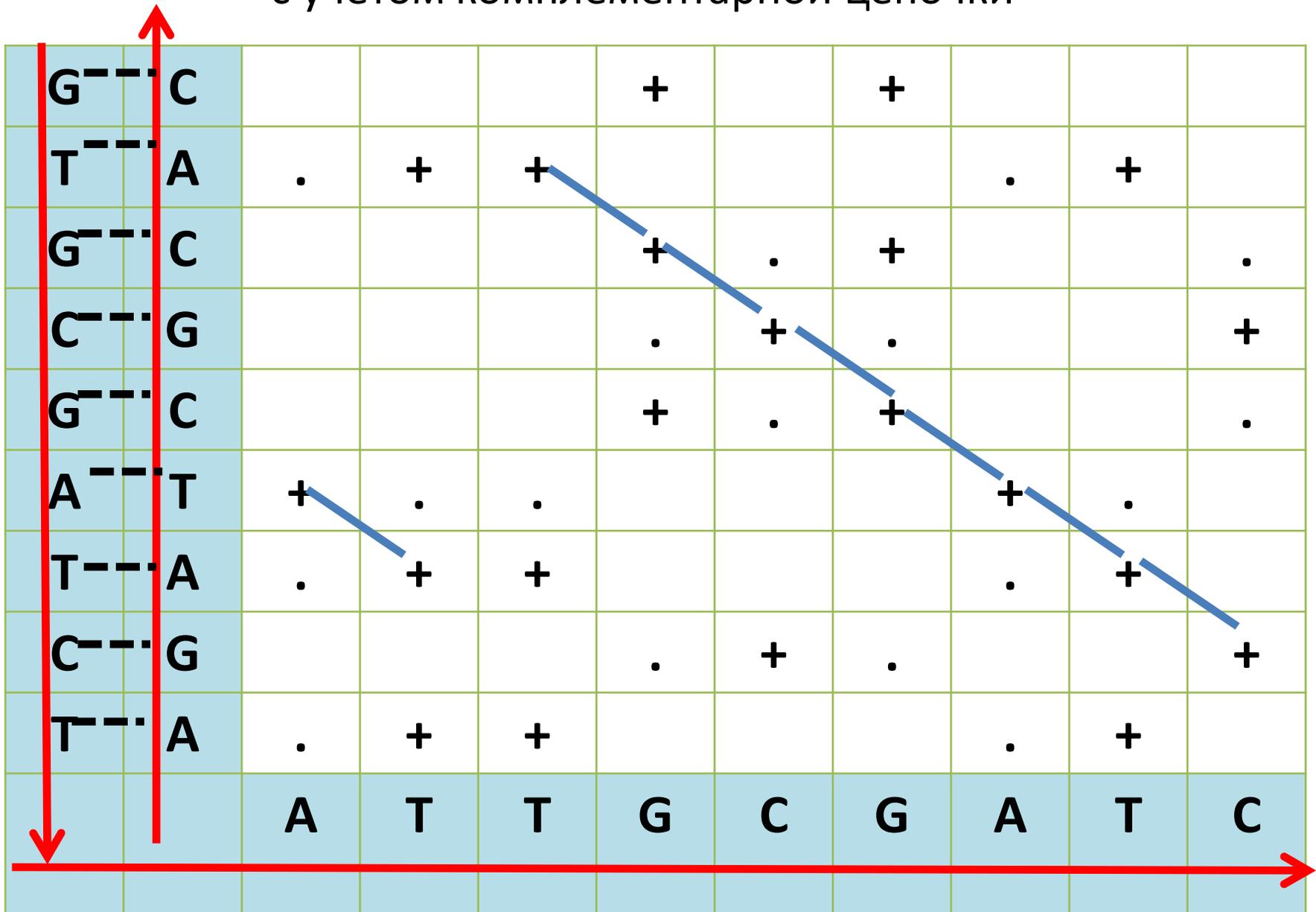
Два метода сравнения последовательностей

- Карта локального сходства. Может быть применена для сравнения геномов целиком
- Выравнивание последовательностей. Может быть применена только для сравнения последовательностей, произошедших из одной и той же предковой последовательности. Такие последовательности называются **ГОМОЛОГИЧНЫМИ**

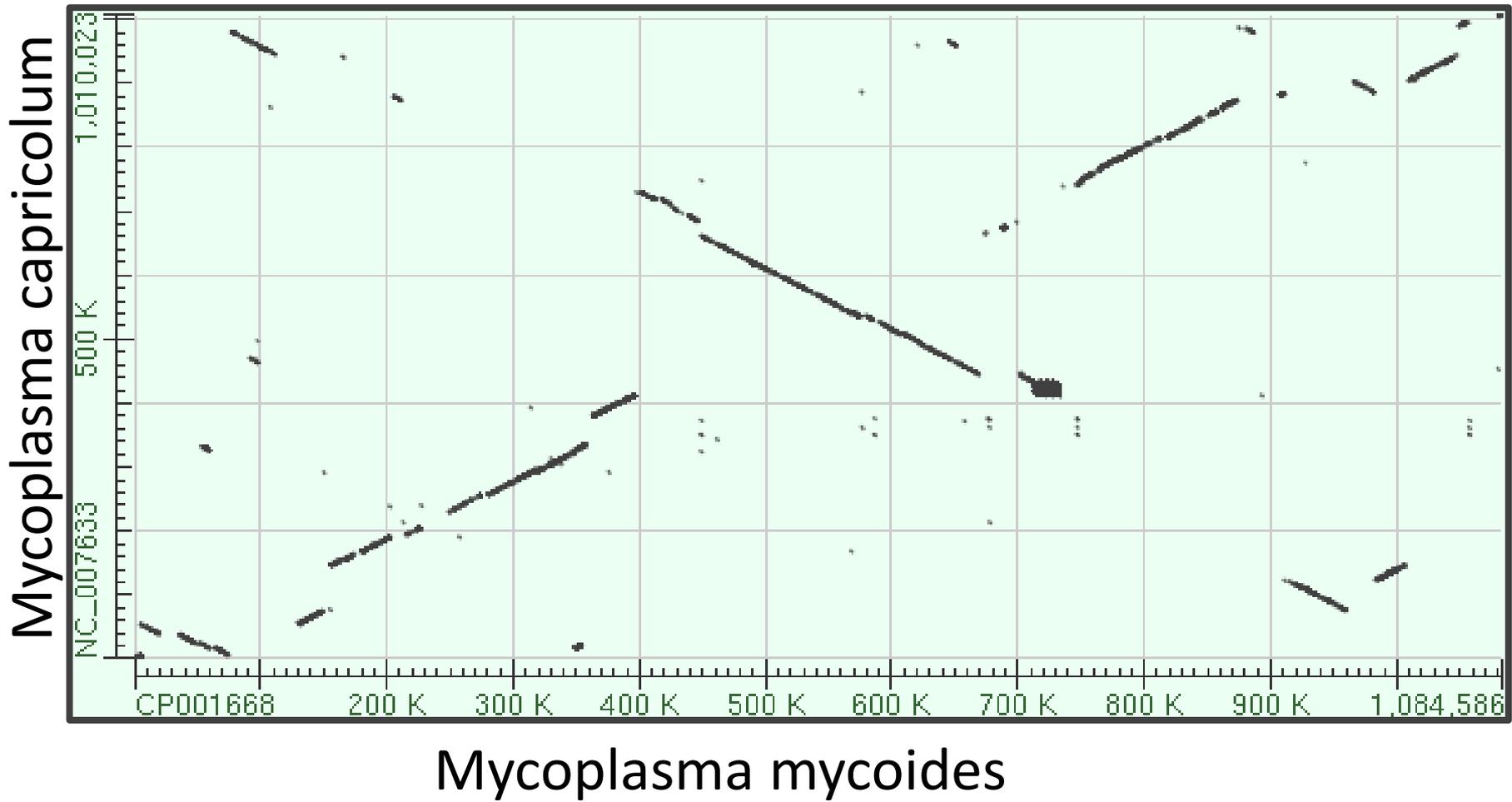
Карта сходства двух последовательностей



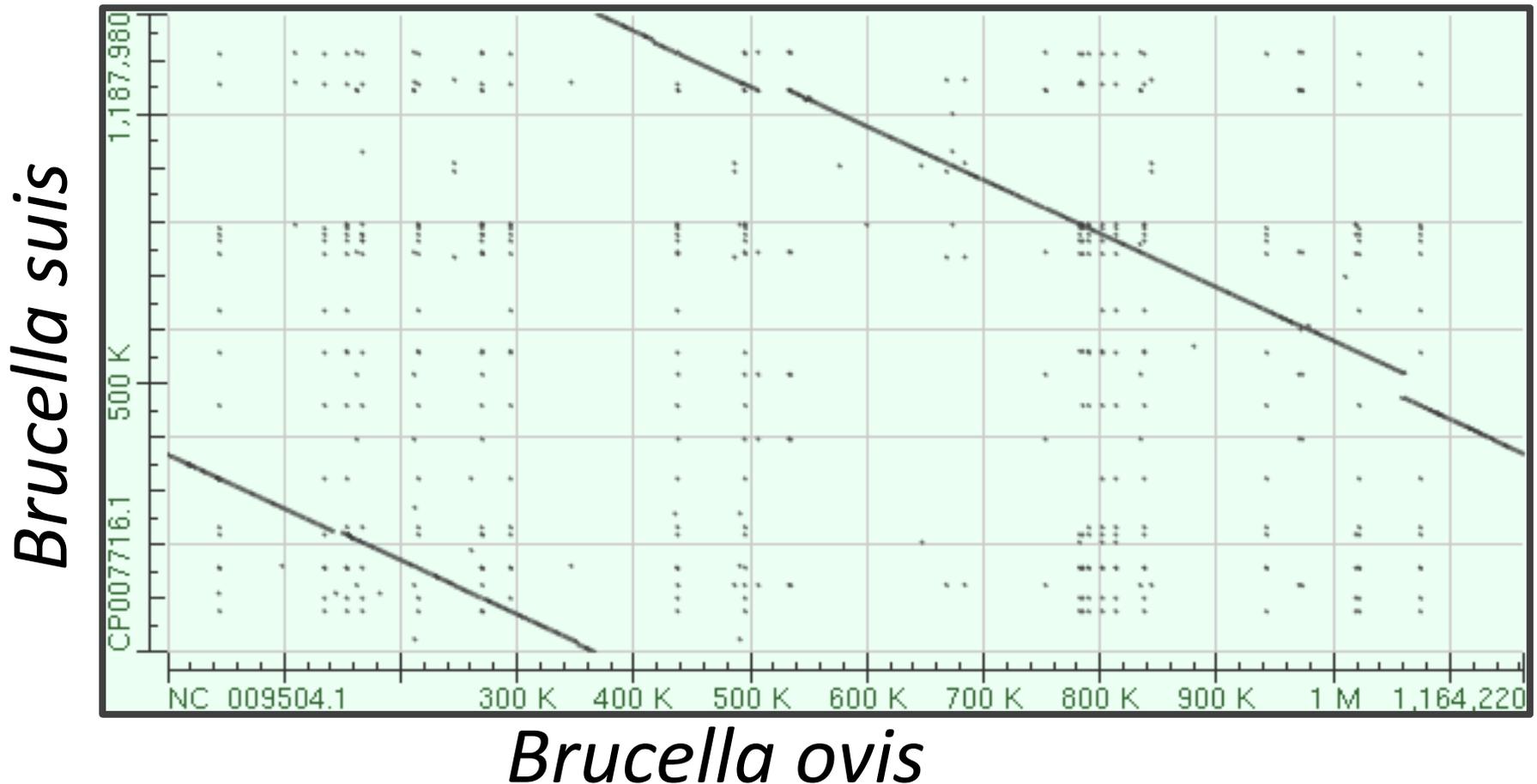
Карта сходства двух последовательностей с учетом комплементарной цепочки



Карта локального сходства геномов *M. capricolum* и *M. mycoides*



Карта сходства геномов двух видов болезнетворных бактерий - бруцелл



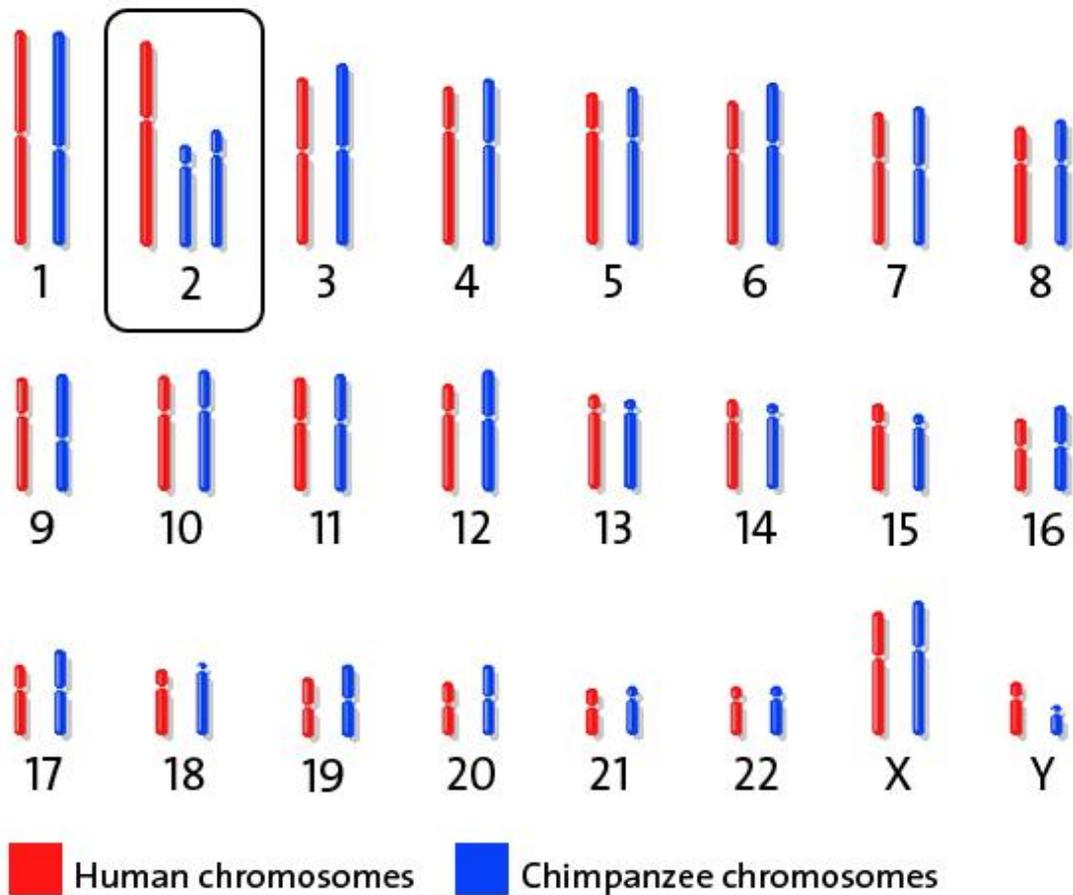
Выравнивание последовательностей геномов *M. capricolum* и *M. mycoides* (маленький фрагмент)

<i>1M. mycoides</i>	1091	t a a - - - t t a a t t a t a a a t t t a t a a a t a t t t t t c a t t a a G T C T G A	1130
<i>1M. capricolum</i>	1116	T A A T T T T T A A T T A T A A A T T T A T A A A T A T T T T T C A T T A A G T C T A A	1158
<i>1M. mycoides</i>	1131	T G T A T T C A C C T T T T T T A A T A T A T A A A A C T C C A G A A A G A A A A T C	1173
<i>1M. capricolum</i>	1159	T A T A T T C A C C T T T T T T A A C A T A T A A A A C T C C A G A A A G A A A A T C	1201
<i>1M. mycoides</i>	1174	T T T A A A A C G T T T A G C T T T A T T A T C A T C T A A G T T T T T A A A A T C T	1216
<i>1M. capricolum</i>	1202	T T T A A A A C G T T T A G C T T T A T T A T C A T C T A A G T T T T T A A A A T C T	1244
<i>1M. mycoides</i>	1217	A C A A C A A C A A C T T T T T G A T C T A A T A A A G T A T C T A C A A T T G A T T	1259
<i>1M. capricolum</i>	1245	A T A A C A A C A A C A T T A T G T T C T A A T A A A G T A T C A A C A A T T G A T T	1287
<i>1M. mycoides</i>	1260	G A A C T T C A G A A A A T T T C A T A G G A C T A A A T A C A T A A G T G T T A A T	1302
<i>1M. capricolum</i>	1288	G A A T T T C A G A A A A T T T C A T A G G A C T A A A A A C A T A T G T A T T A A C	1330

В среднем, 92% совпадающих букв на сходных участках



Человек – шимпанзе: две хромосомы объединились в одну





Человек – шимпанзе: сравнение последовательностей

Human/1-27893

Chimp/1-27866

1939 **A** **T** **T** **A** **T** **T** **T** **G** **C** **C** **T** **T** **C** **T** **A** **C** **A** **T** **T** **T** **C** **C** **T** **C** **G** **A** **T** **G** 1966
 1936 **A** **T** **T** **A** **T** **T** **T** **G** **C** **C** **T** **T** **C** **T** **A** **C** **A** **T** **T** **T** **C** **C** **T** **C** **G** **A** **T** **G** 1963

Human/1-27893

Chimp/1-27866

1967 **A** **A** **T** **T** **T** **C** **A** **G** **G** **G** **T** **T** **G** **T** **T** **T** **G** **T** **T** **C** **T** **T** **G** **A** **A** **T** **C** **T** 1994
 1964 **A** **A** **T** **T** **T** **C** **A** **G** **G** **G** **T** **T** **G** **T** **T** **T** **G** **T** **T** **C** **T** **T** **G** **A** **A** **T** **C** **T** 1991

Human/1-27893

Chimp/1-27866

1995 **T** **T** **A** **A** **A** **T** **G** **A** **A** **T** **T** **A** **A** **T** **A** **A** **T** **A** **C** **T** **T** **C** **A** **C** **A** **T** **A** **T** 2022
 1992 **T** **T** **A** **A** **A** **T** **G** **A** **A** **T** **T** **A** **A** **T** **A** **A** **T** **A** **C** **T** **T** **C** **A** **C** **A** **T** **A** **T** 2019

Human/1-27893

Chimp/1-27866

2023 **C** **A** **C** **T** **G** **T** **A** **C** **C** **A** **A** **A** **G** **G** **C** **C** **C** **C** **A** **A** **A** **G** **G** **C** **A** **G** **G** **T** 2050
 2020 **C** **A** **C** **T** **G** **T** **A** **C** **C** **A** **A** **A** **G** **G** **C** **C** **C** **C** **A** **A** **A** **G** **G** **C** **A** **G** **G** **T** 2047



Человек – шимпанзе

Human/1-27893

Chimp/1-27866

1734 **CGTCTCAAAAAAAAAAAAAA**1750
 1710 **CGTCTCAAAAAAAAAAAAAA**1726

Human/1-27893

Chimp/1-27866

1751 **A** - - - - -
 1727 **AGAAAAGAAAAGAAAAA**1743

Human/1-27893

Chimp/1-27866

1752 - - - - **AGAAAAGAAA**1764
 1744 **GAAAAGAAAAGAGAAA**1760

Human/1-27893

Chimp/1-27866

1765 **AAG** - **TTAGGTCTTG****GGT**1780
 1761 **AAGGTTAGGTCTTTGGT**1777



Человек – шимпанзе

<i>Human/1-27893</i>	9233	AAGCACAAAGGAAGCTGG	9250
<i>Chimp/1-27866</i>	9227	AAGCACAAAGGAAGCTGG	9244
<i>Human/1-27893</i>	9251	TCTCAACCTGAGAAAACC	9268
<i>Chimp/1-27866</i>	9245	TCTCAACCTGAGAAAACC	9262
<i>Human/1-27893</i>	9269	AATTCACCCCTTTGTAAAA	9286
<i>Chimp/1-27866</i>	9263	AATTTAGCCCTTTGTAAAA	9280
<i>Human/1-27893</i>	9287	CCCTCCCTACACCCCCAC	9304
<i>Chimp/1-27866</i>	9281	CCCTCCCTACACCCCCAC	9298



Comparison to Human Genome

	Chimp	Mouse
<i>Overall identity</i> Процент одинаковых букв	95%	28%
<i>Aligned bases</i> Процент выравниваемых участков	97%	40%
<i>Identity at aligned bases</i> Процент одинаковых букв в выровненных участках	99%	69%
<i>Identity in genes</i> Процент одинаковых букв в генах	99%	85%

Из презентации **Katherine S. Pollard**