

МФК

Биоинформатика

февраль – май 2023

Биоинформатика

- Био-
потому, что **результаты биологические**
- -информатика
потому что **методы информатики**
- Какая информация на входе?
Объясним
- Какие результаты на выходе?
Объясним

Биоинформатика. Объекты

- Биологические последовательности
 - ДНК
 - РНК
 - белки
- Пространственные структуры биологических макромолекул
- Другие массовые биологические данные.

Состав записавшихся студентов

год обучения	Число записавшихся	Ф-ты с ≥ 3 студентами
1	38	Биофак, биотех
2	5	
3	4	
4	23	Биофак, почв, вмик, мехм
Всего	70	

Ну и ну! ПОЕХАЛИ)))

Полезная информация

Сайт МФК: «Биоинформатика 2022»

https://kodomo.fbb.msu.ru/wiki/Main/mf_2023s

С орг-вопросами обращайтесь к

Сергею Александровичу Спирин (организатору МФК и лектору)

sas@belozersky.msu.ru

Кроме лекций для того, чтобы разобраться в теме и получить зачёт, предлагается выполнять домашние задания. Бывают задания на поиск в интернет, бывают школьного уровня, бывают посложнее

Зачёт автомат можно получить при зачёте оговоренного числа домашних заданий, выдаваемых после каждой лекции

На сайте найдёте

- Список записавшихся на МФК
- Ссылки на презентации
- Домашние задания (после каждой лекции)
- Форма для вопросов преподавателям

Попозже

- Правила получения зачета
- Ведомость с результатами проверок ДЗ
- Новости

Пока сайт в процессе. Coming soon

ОФИЦИАЛЬНЫЙ список студентов, записавшихся через ректорат дает право на получение зачёта в зачётку и в зачётную ведомость МГУ.

Желающие могут попросить включить их дополнительно в список участников МФК Биоинформатика на нашем сайте.

Это не меняет их официальный статус. Дополнительно записанные студенты – вольнослушатели – не получают права на проставление зачёта по МФК.

Узнавайте что можно сделать в своей учебной части.

Преподаватели не обязан проверять домашние задания вольнослушателей. Но если захочет – может

Л1 ГЕНОМ

**Генóм —
совокупность наследственной информации о
ЖИВОМ**

Андрей Владимирович Алексеевский
aba@belozersky.msu.ru

План

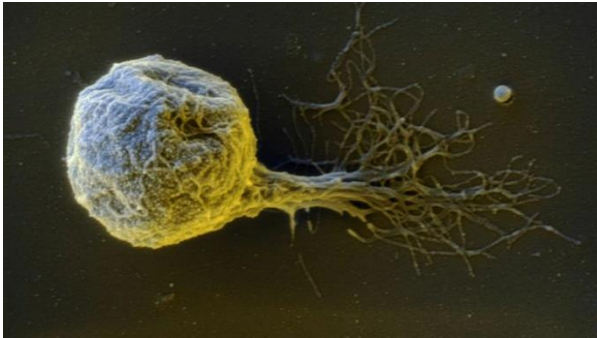
1. Живое умеет
2. Что такое информация? Носители информации
3. Цепочка ДНК – молекула
4. Беда с одной цепочкой ДНК!!!
Последовательность дцДНК=dsDNA
5. Можно ли прочитать последовательность ДНК и РНК?
6. Геном организма с точки зрения биоинформатики это
7. Откуда взялись дцДНК молекулы?
8. РНК тоже имеет отношение к кодированию
9. Что записано в геноме (и б.м. понятно нам)
10. Геном человека
11. Геномы бактерий
12. Геном адено-ассоциированного вируса
13. Геном коронавируса. Надеюсь, что не успею рассказать

1. Живое умеет

- размножаться (это его жизненная установка)
- копировать свой геном потомкам
- **с ошибками => ЭВОЛЮЦИЯ** (этим отличаемся от компьютерных вирусов)))
 - **эукариоты** (от греч. *éu* — полностью и *káρυον* — ядро)
многоклеточные эу это мы: папа + мама => ребёнок
одноклеточные эу: удвоение ядер и деление, несколько способов.
 - **прокариоты** (от др.-греч. *πρό* — 'перед' и *κάρυον* — 'ядро')
бактерии
археи про Энцеладу картинки самый быстрый
одна клетка делится на две: клетка => 2 с тем же геномом
репликация (= удвоение) генома
 - Клетка многоклеточных эукариот – живая!
 - Вирусы – вообще, анекдот)))
и сами не живут, и другим не дают жить здоровими

Археи – древние и экстремофилы

1) Methanocaldococcus villosus



M. villosus способны двигаться со скоростью 589 и 468 мкм в секунду соответственно (B. Hertzog, R. Wirth, 2012. [Swimming Behavior of Selected Species of Archaea](#)). Скорость последней соответствует почти 500 длинам ее тела в секунду. Для сравнения: гепард — самое быстрое млекопитающее на земле — при атаке движется со скоростью всего 20 длин тела в секунду, а сапсан — самое быстрое животное вообще — при пикировании разгоняется до ~190 длин своего тела в секунду. Движение архей также контролируется структурами, выступающими за пределы тела. Только вот жгутик архей принципиально отличается от жгутика бактерий. Для него даже используют отдельное название: [археллум](#) (по аналогии с названием жгутика, *flagellum*). Археи передвигаются вперед, если археллум движется по часовой стрелке, и назад — если против часовой стрелки.

Иллюстрация с сайта <https://ecoportal.su/news/view/115435.html>

2) Methanothermococcus okinawensis

<https://life.ru/p/1093754>

По мнению исследователей из Германии и Австрии, условия на Энцеладе, одном из спутников Сатурна, подходят для жизни архей — одноклеточных микроорганизмов, не имеющих ядра. Под покрытой льдом поверхностью спутника Сатурна находится океан. Выяснилось, что археи вида *Methanothermococcus okinawensis*, производящие метан, оказались прекрасно приспособлены к жизни в условиях, которые ученые создали в лаборатории и которые могут существовать на Энцеладе. Археи *Methanothermococcus okinawensis* живут вблизи гидротермальных источников на морском дне при очень высокой температуре, перерабатывая углекислый газ и водород в метан.

“To date, methanogenic archaea are the only known microorganisms that are capable of performing biological

CH₄ production in the absence of oxygen^{12,13}”

Taubner et al., Biological methane production under putative Enceladus-like conditions, 2018



2. Что такое информация?

Армянское радио

- «Правда ли, что Иштоян выиграл в лотерею машину?»

Иштоян известный футболист Арарата в 70-х

- «**Правда.** Но не Иштоян, а Петросян, не машину, а швейную машинку; не в лотерею, а в карты; и не выиграл, а проиграл»

Петросян чемпион мира по шахматам в 60-х

«Сколько информации в этом сообщении?»

И.М.Гельфанд

Вопросы

- Что такое информация? Нашел такое определение:

ИНФОРМАЦИЯ — сведения независимо от формы их представления))) ¹⁾

Теория информации основанная Шенноном – математическая теория передачи данных²⁾ – используется в биоинформатике, но слишком формализована и проста для объяснения живого:)

- Какая информация закодирована в геноме?
- Сведения: от кого? кому (получатель)?
- Как кодируется информация в геноме?
- Как прочитать её людям и зачем?

¹⁾ Wiki со ссылкой на Когаловского Р.М. специалиста по информационным систем.

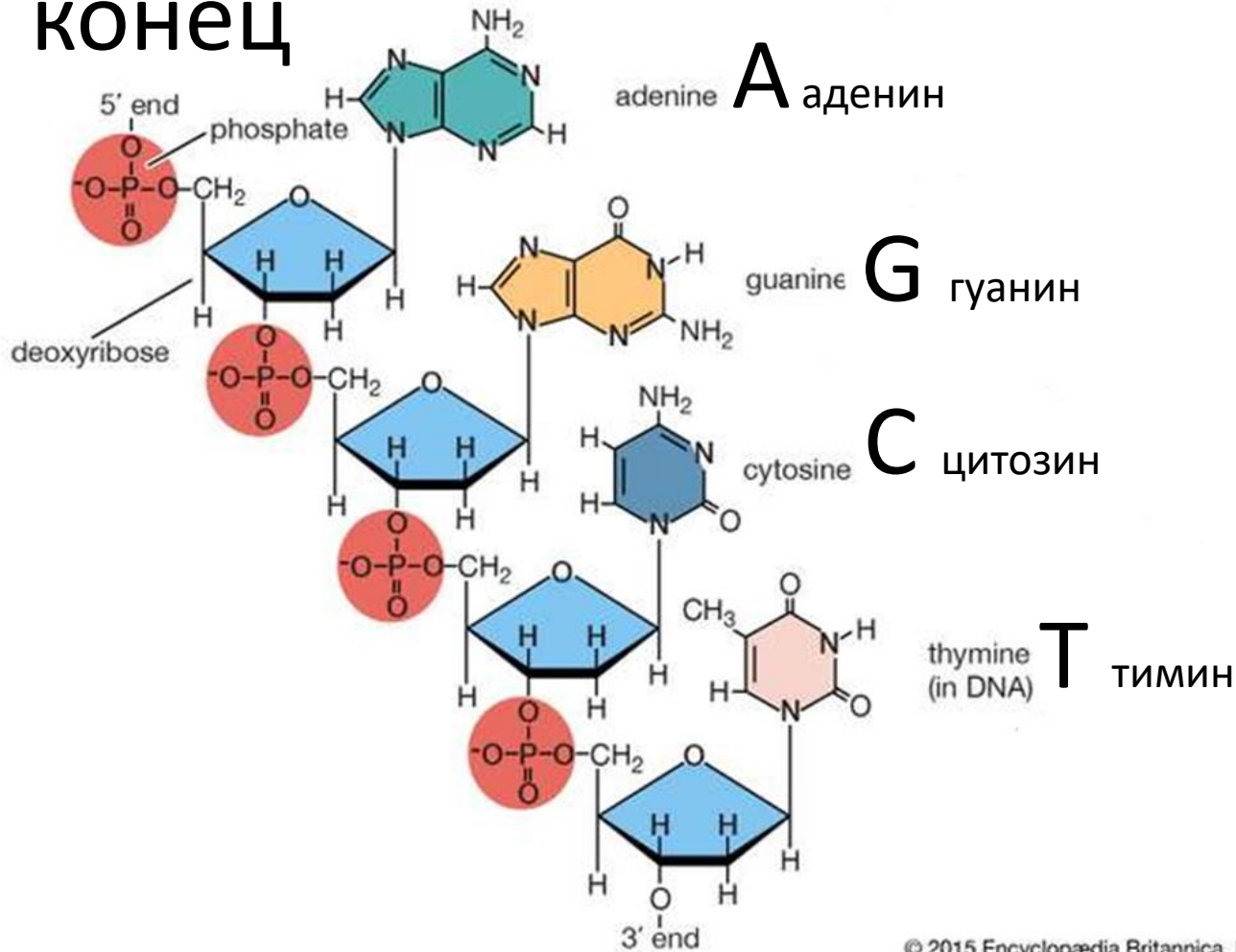
²⁾ С.Shannon, “The Mathematical Theory of Communications” , 1948

Носители информации

- Книги, электронные носители, рисунки, песни, разговоры, обычаи
- Носители наследственной информации ДНК (и РНК)

3. ДНК – молекула.

У неё есть химическая формула
5' конец

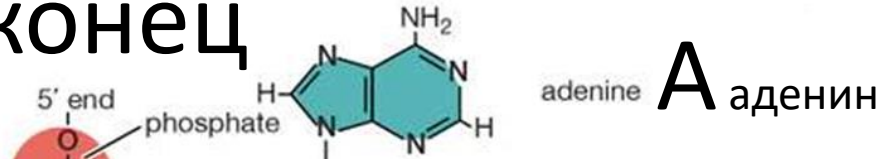


= **AGCT**

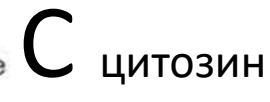
3' конец

Формула молекулы ДНК определяется последовательностью букв А, Т, G, С

5' конец



Основания ДНК



= AGCT

Сахаро-фосфатный остов

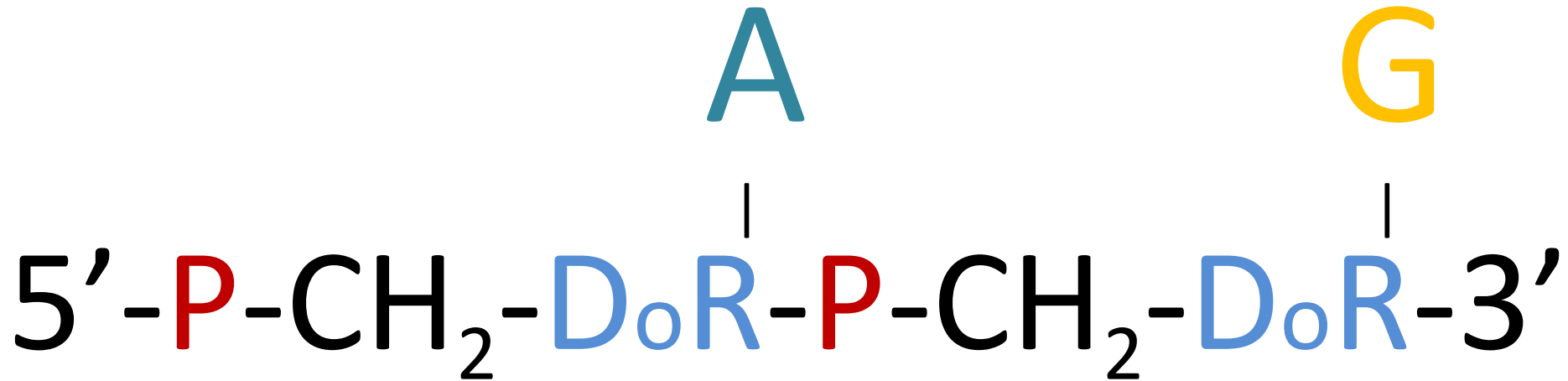
сахар
фосфат
CH₂
сахар
фосфат

3' конец

Как определить направление 5' => 3' находясь в середине цепочки?

© 2015 Encyclopædia Britannica, Inc.

Читать справа-налево или
слева-направо?



AG \neq GA

4. Беда с цепочкой ДНК!!!

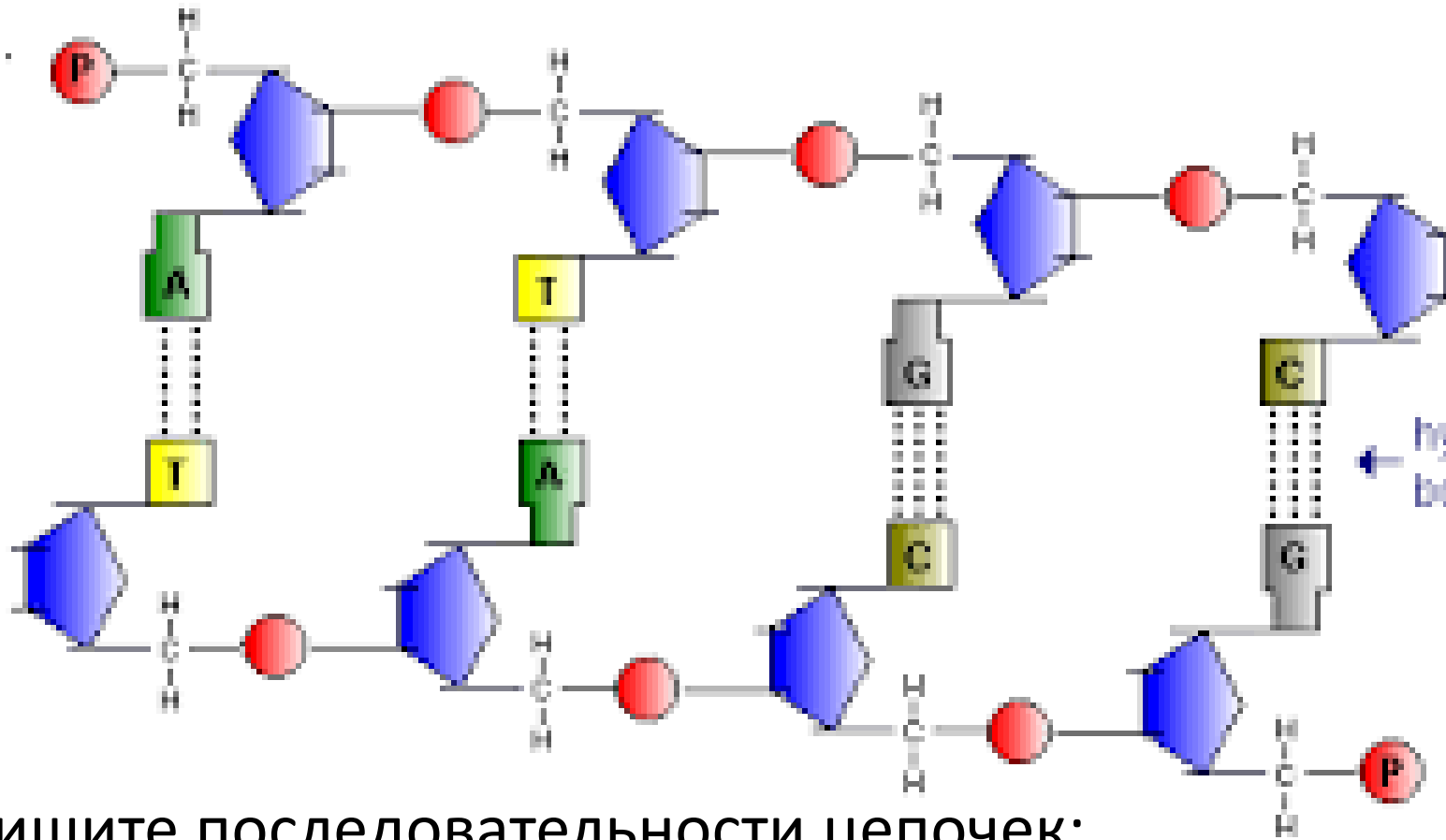
- Одна цепочка ДНК легко расщепляется при разного рода воздействиях(((
Что бы придумать ... природе? Эволюции?
Инопланетянам?
Нам)))

- Эврика! А вот что можно использовать:
G с C умеют соединяться 3мя водородными связями,
A с T – двумя,
T с G одной.

Водородные связи слабые((НО!

- Сделаем так: verte

Схема двухцепочечной ДНК (dsDNA)



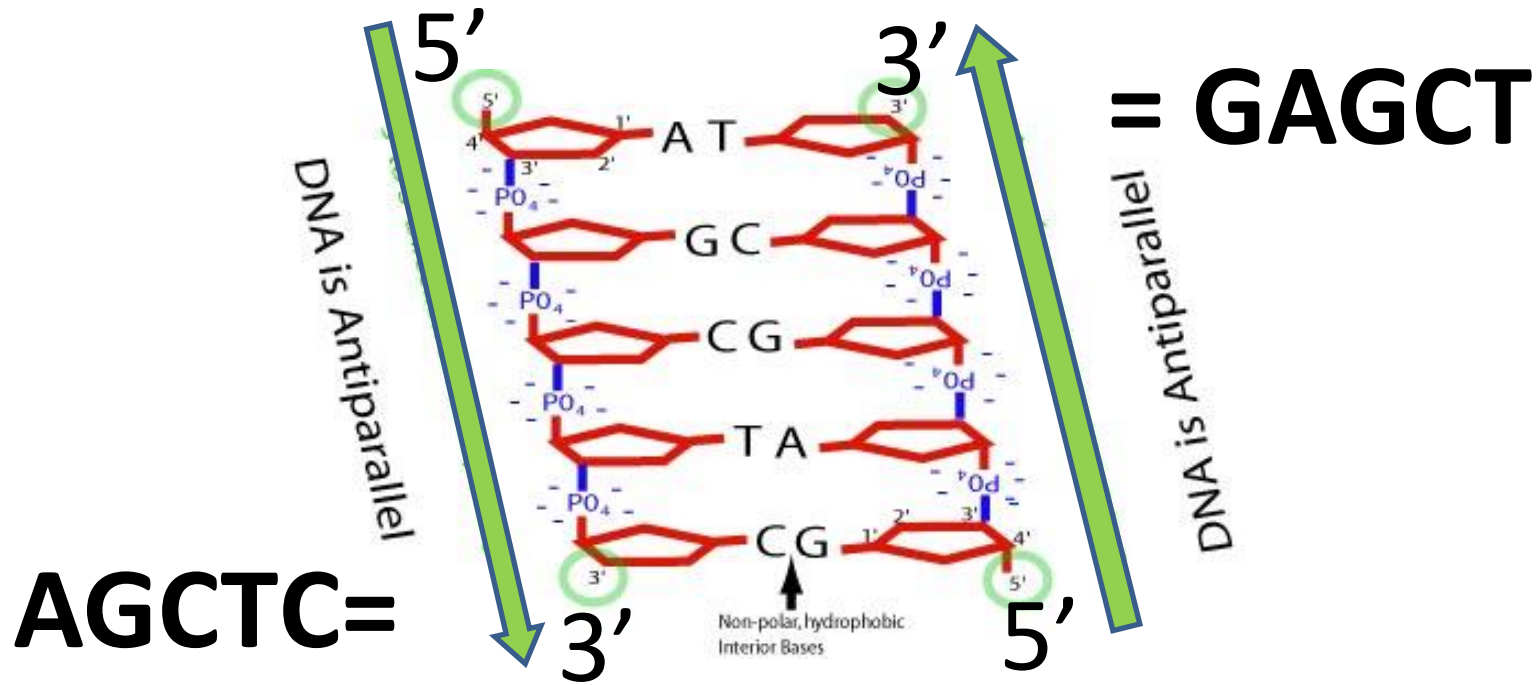
Напишите последовательности цепочек:

ВЕРХНЕЙ

НИЖНЕЙ

Какая цепочка главнее?

ДНК состоит из двух **антипараллельных** комплементарных цепочек



Энергия водородной связи невелика по сравнению с энергией ковалентной связи. GC состав генома = процент комплементарных пар GC от длины генома.

Последовательность дцДНК

1. Состоит из букв **A, T, G, C**
2. ВСЕГДА пишется в направлении ОТ 5' КОНЦА К 3' КОНЦУ
3. Химическая формула ДНК однозначно определяется последовательностью
4. Последовательность несёт всю наследственную информацию.

3. – правда, но не вся:

В клетке бывают химические модификации ДНК

4. – правда, но не вся:

5. Можно ли прочитать последовательность ДНК (и РНК)?

- Можно и ДНК, и РНК. Но не просто.
 - Секвенаторы (автоматы для прочтения последовательности ДНК) умеют читать только относительно короткие последовательности ДНК (например, 150 п.н.)
 - Задача биоинформатики из коротких последовательностей составить последовательность всей ДНК
 - Об этом будет лекция

6. Геном организма с точки зрения биоинформатики это

1. набор последовательностей всех dsDNA из одной клетки. Т.е. число последовательностей равно числу молекул dsDNA в клетке.
2. У многоклеточных (нас с вами) геномы всех клеток совпадают
3. В референсном геноме вида (человека) – из различающихся подпоследовательностей берут самую типичную

2. – правда, но не вся

Бывают различия: соматические мутации;
запланированные различия (иммуноглобулины)

Референсный геном SARS-CoV-2

>NC_045512.2 Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, complete genome

ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAAA
CGAACTTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACCTCACGCAGTATAATTAATAAC
TAATTA CTGTCGTTGACAGGACACGAGTAACTCGTCTATCTTCTGCAGGCTGCTTACGGTTTTCGTCCGTG
TTGCAGCCGATCATCAGCACATCTAGGTTTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTC
CCTGGTTTTCAACGAGAAAACACACGTCCAACCTCAGTTTTGCCTGTTTTTACAGGTTTCGCGACGTGCTCGTAC
GTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGG
CTTAGTAGAAGTTGAAAAAGGCGTTTTTGCCTCAACTTGAACAGCCCTATGTGTTTCATCAAACGTTCCGGAT
GCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCAGAACTCGAAGGCATTCAGTACGGTC
GTAGTGGTGAGACACTTGGTGTCCTTGTCCCTCATGTGGGCGAAATACCAGTGGCTTACCGCAAGGTTCT
TCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCTAAAGTCATTTGACTTA
GGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTCAAGAAAACCTGGAACACTAAACATAGCAGTGGTG
TTACCCGTGAACTCATGCGTGAGCTTAACGGAGGGGCATACACTCGCTATGTCGATAACAACCTTCTGTGG
CCCTGATGGCTACCCTCTTGAGTGCATTAAGACCTTCTAGCACGTGCTGGTAAAGCTTCATGCACTTTG
TCCGAACAACCTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTGAACATGAGCATGAAATTG
CTTGGTACACGGAACGTTCTGAAAAGAGCTATGAATTGCAGACACCTTTTTGAAATTAAATTGGCAAAGAA
ATTTGACACCTTCAATGGGGAATGTCCAAATTTTGTATTTCCCTTAAATTCATAATCAAGACTATTCAA
CCAAGGGTTGAAAAGAAAAGCTTGATGGCTTTATGGGTAGAATTCGATCTGTCTATCCAGTTGCGTCAC

Геном - неизвестный текст, иногда из известного организма, иногда – нет (метагеном).

С чего начнем?

Знаем:

- В алфавите четыре буквы А, Т, G, С (понятно)
НАМ ПОВЕЗЛО С ЭТИМ
(см. след. слайд для неспециалиста)
- Буквы идут неупорядоченно,
похоже на случайную последовательность?

脊髄性筋萎縮症遺伝子治療製品 オナセムノゲンアベパルボベク（ゾルゲンスマ[®]） の薬理学的特性と臨床試験成績

渥美 綾香¹，米田 智廣²，土田 健²，香川 雄輔²，富永 俊輔²，川瀬 一穂¹，菊地 信孝¹

オナセムノゲンアベパルボベク（製品名ゾルゲンスマ[®]，開発コード AVXS-101）は，機能的なヒト Survival motor neuron（SMN）遺伝子を脊髄性筋萎縮症（SMA）患者の運動ニューロンに届けられるよう設計された，アデノ随伴ウイルス9型カプシドを有する非増殖性遺伝子組み換えアデノ随伴ウイルスベクター製品である。2020年3月19日に2歳未満の「SMA（臨床所見は発現していないが，遺伝子検査によりSMAの発症が予測されるものも含む）」を対象に承認された。静脈内に投与された本品は，SMAの根本原因であるSMN1遺伝子の機能欠損を補って運動ニューロンの変性・消失を防ぎ，神経及び筋肉の機能を高め，筋萎縮を防ぐことで，SMA患者の生命予後及び運動機能を改善することが期待される。また，導入されたSMN遺伝子は患者のゲノムDNAに組み込まれることなく細胞の核内にエピソームとして留まり，運動ニューロンのような非分裂細胞に長期間安定して存在するように設計されていることから，1回の静脈注射で治療が完結する。SMAモデルマウスへの本品投与により，SMNタンパク質の持続的発現，体重増加，運動機能の改善，生存期間の延長等が認められた。臨床でのオナセムノゲンアベパルボベクの有効性は，I型SMA患者（CL-101試験）及び未発症のSMA患者（CL-304試験）を対象とした臨床試験にて確認され，両試験において自然歴に比べ有意に「出生から永続的な呼吸補助が必要となる又は死亡までの期間」が延長されることが示された。また，両試験とも自然歴では見られない運動マイルストーンの達成も確認

Лингвистический анализ текста

- Одинаковы ли частоты букв?
G+C – состав = % процент букв G и C вместе = GC-состав.
Чем G и C отличаются от A и T?
- Часто и редко встречающиеся слова
(т.е. короткие последовательности)
- Равномерность частоты букв и слов вдоль текста

Все эти вопросы изучаются и имеют биологически смысл!
Примеры наблюдений:

- $\#C \approx \#G$, $\#T \approx \#A$ ($\#$ = число)
- Слов CG *мало* в определенных геномах
Слов TA *мало* во всех геномах

Мало по сравнению с ожидаемым при случайном расположении букв в тексте.

Самое простое предположение:

Ожидаемое $\#CG = \text{частота}(C) \times \text{частота}(G) \times (\text{число букв в геноме})$. Для др. слов аналогично

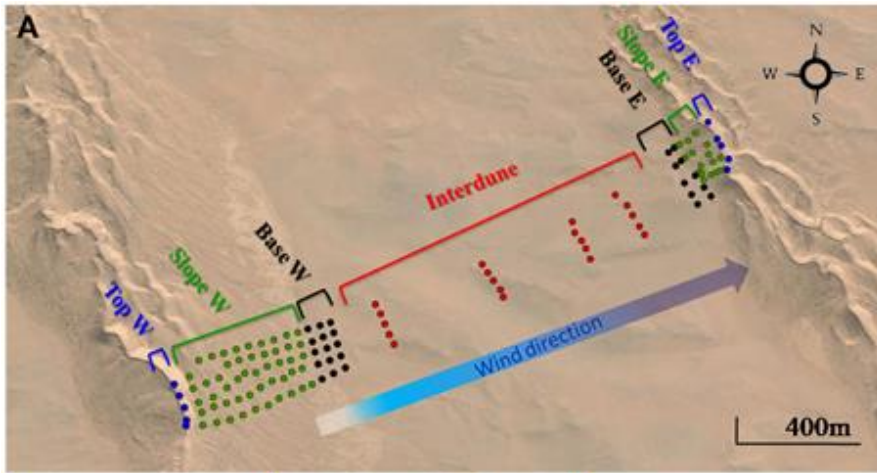
- В некоторых геномах $\#C > \#G$ в одной части и $\#G > \#C$ в другой части («GC skew»)

Всерьёз думают о живых вакцинах, основанных на вирусах с увеличенным числом CG или TA!

Interestingly, most mammalian RNA viruses have low frequencies of CpGs ([45,46](#)). Furthermore, viruses with high CpG frequencies may be more recognizable by pathogen innate immune sensors ([47–50](#)).

Attenuation of the classical oral poliovirus vaccine is based on very few point mutations, which can revert to virulence after a few rounds of viral replication ([144](#)). These pioneering results obtained with recoded polioviruses suggest that codon-usage in recoded viruses may be much more stable than most RNA virus point mutants, and could possibly enable the development of live attenuated RNA virus vaccines with superior genetic stability.

Martinez et al., 2019, NAR



Ronca et al., 2015

ВОПРОС: У какой бактерии G+C состав больше: живущей (i) в пустыне НАМИБ в Африке или (ii) в Антарктиде?

<https://www.sci.news/biology/science-microbes-deep-lake-antarctica-01424.html>



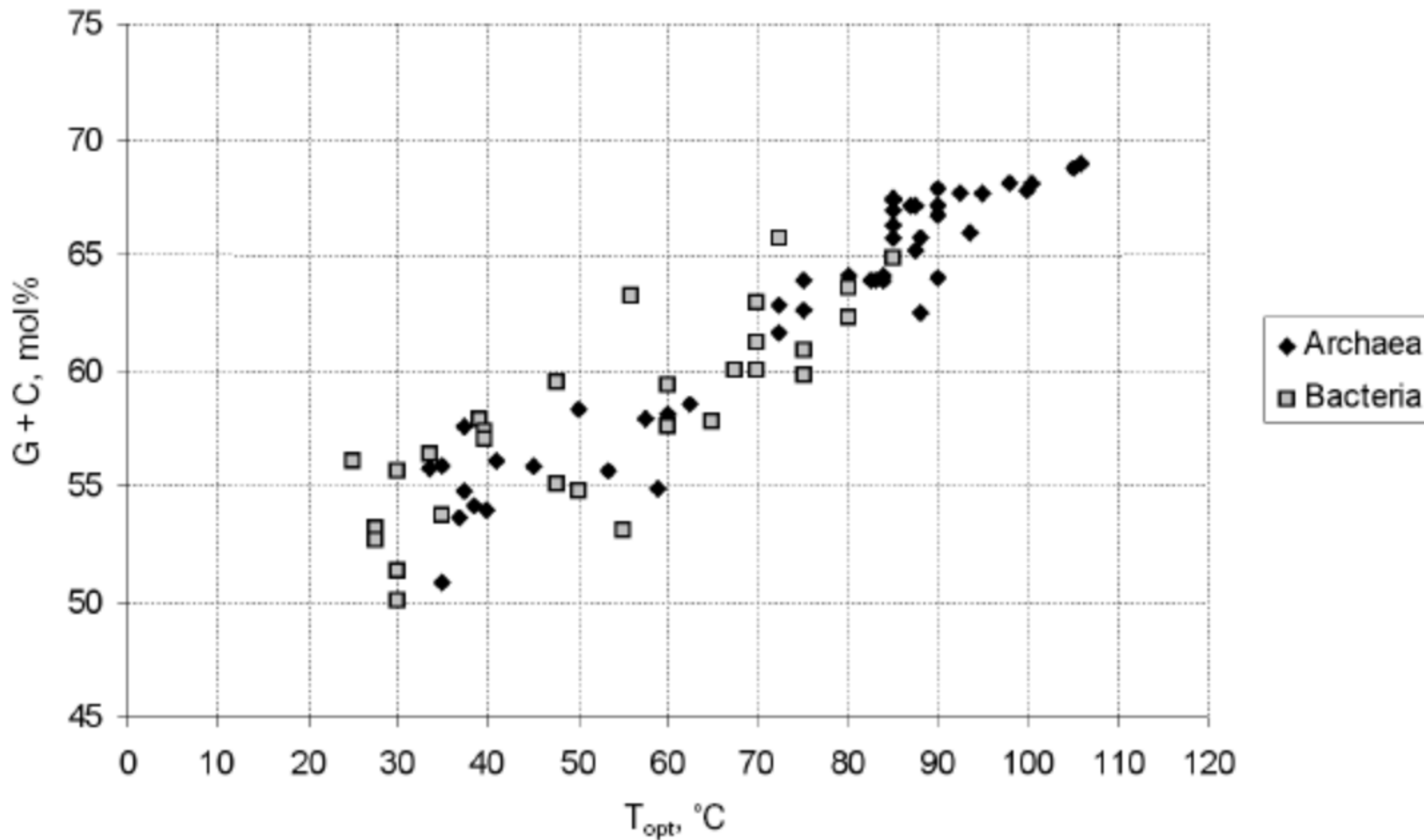


Fig. 1. Relationship between the optimum growth temperature and the G + C content of the 16S rRNAs of archaea and bacteria.

Lebedinsky et al., REVIEW: Phylogenetic Systematics of Microorganisms Inhabiting Thermal Environments, 2007

7. Откуда взялась dsDNA как носитель генома?

- Носителем генома у всех живых клеточных организмов – эукариот и прокариот (архей и бактерий) является dsDNA
- Поэтому считается, что геном **LUCA = Last Universal Common Ancestor** также был записан на dsDNA.
- Носителем генома вирусов бывает и РНК, и ДНК.

*Абсолютно у всех
клеточных организмов
геном состоит
из двуцепочечной ДНК.
А у вирусов –
что угодно!
Любые варианты,
РНК любых форм
и ДНК любых форм.*



Евгений Кунин

Откуда возникла ДНК (или РНК) первого живого организма на Земле?

- НИКТО НЕ ЗНАЕТ
- Есть гипотезы
- Читайте **Евгения Кунина “Логика случая”** и его интервью последних лет
Он умный и пытается рационально подойти к решению этого вопроса.

Пока данных маловато, на мой вкус.



Из Кунина

Интервью: Что нового после издания книги?

“Эпигенетика, метагеномика. Что-то еще нового и важного?

<https://nplus1.ru/material/2018/12/04/kunin>

«Я бы остановился на этом. Для меня это два крайне важных, фундаментальных направления, получивших неожиданно большое развитие за последние семь лет. Способные сильно повлиять на будущую науку и жизнь (на самом деле они уже влияют)»

Эпигеномика это

описание химических модификаций ДНК в клетке и их наследования при делении клеток.

Одна из самых частых, и хорошо изученных модификаций – присоединение метильной группы к определенному основанию в определенной последовательности.

В геноме человека – метилирование цитозина в последовательности из двух букв **CG**
(CpG острова)

Метагеномика это

Секвенирование тотальной ДНК из образца, чтобы определить какие микроорганизмы обитают в определенном месте.

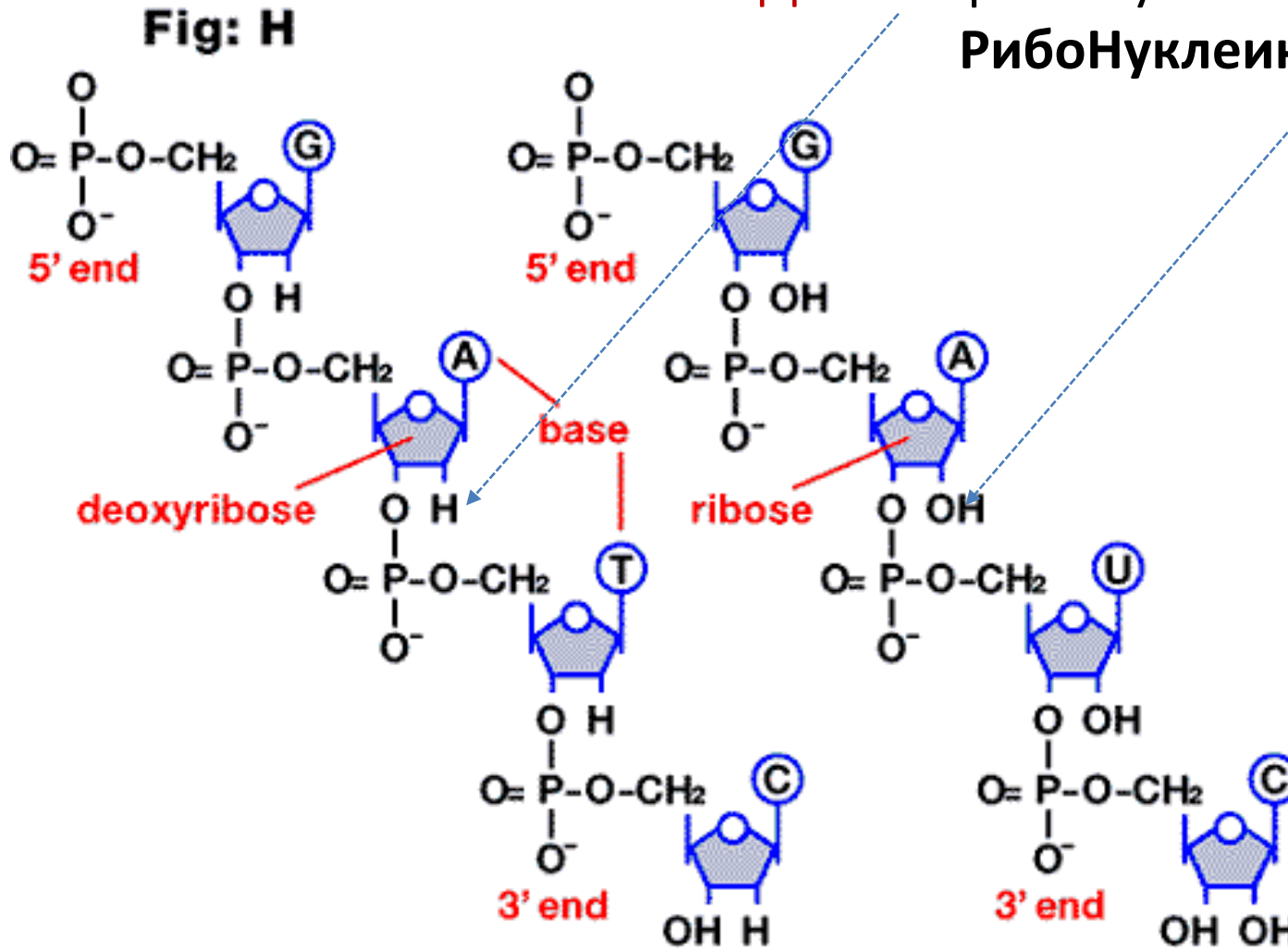
Делают метагеномы микробных сообществ почвы, кишечного тракта или полости рта человека, проб воды и многих-многих других микробиот.

8. РНК тоже имеет отношение к кодированию информации

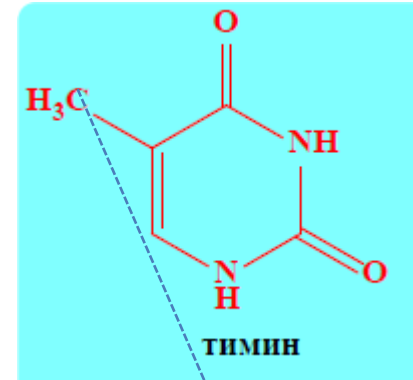
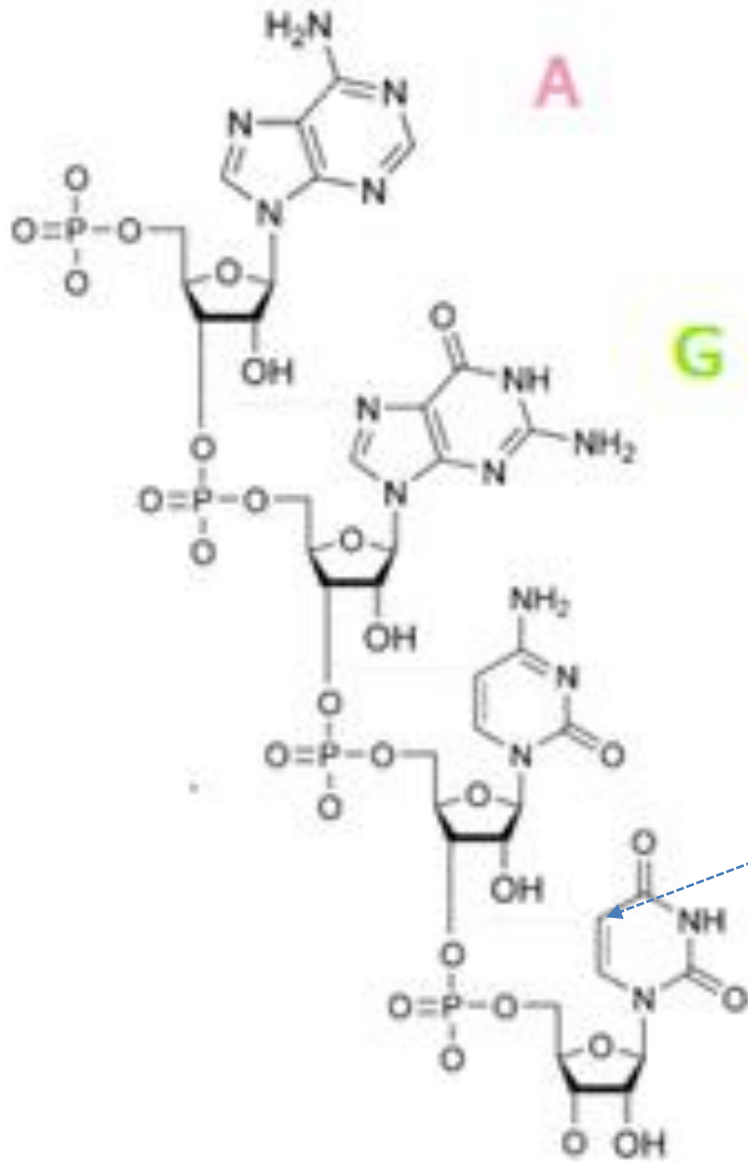
- мРНК копирует кусочек генома на ДНК, содержащего информацию о последовательности белка переносить информацию

6. РНК отличия от ДНК

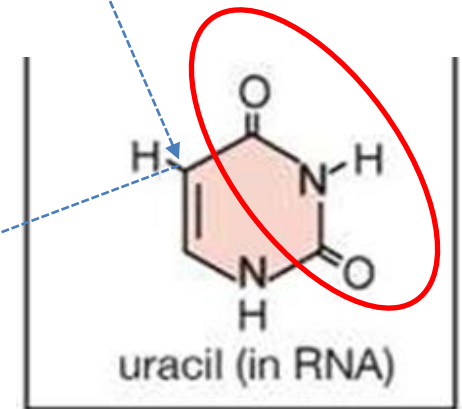
ДезоксирибоНуклеиновая Кислота
РибоНуклеиновая Кислота



Урацил вместо тимина



U урацил
вместо Т



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

Все возможные комплементарные двойные цепочки из ДНК и РНК образуются в клетках

Потому, что урацил сохраняет способность образовывать две водородные связи с аденином, так как отсутствующая у него CH_3 группа (метильная группа) не участвует в образовании водородных связей.

- ДНК-ДНК = dsDNA
- ДНК-РНК гибриды
- РНК-РНК = dsRNA

Всего два отличия, а какая разница в биологии!!!

ДНК

- ДНК 2х-цепочечная, бывает длинной (до сотен млн п.н.), линейной или кольцевой
- Носитель генома
- 3Д структура – след слайд

РНК

- Обычно одноцепочечная и не очень длинная
- Разные РНК имеют разные функции
 - мРНК копирует последовательность ДНК кодирующую белок
 - рРНК
 - тРНК
 - множество других с разными функциям
- 3Д структура разнообразна - след слайды

Двойная спираль ДНК

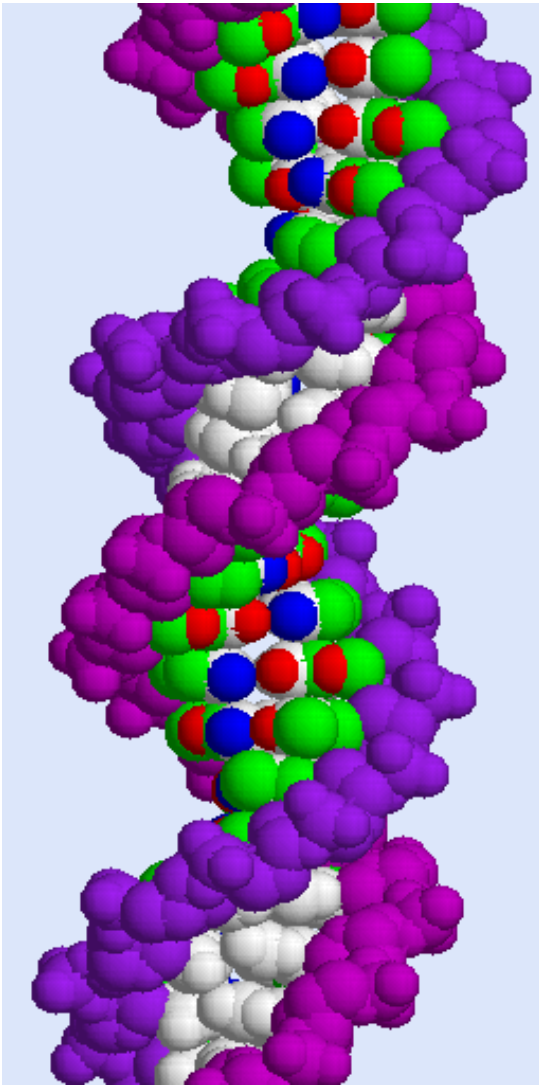


Рис. 3D структура фрагмента ДНК *in vitro*.

Получена с помощью рентгеноструктурного анализа (РСА).

Шарик = атом

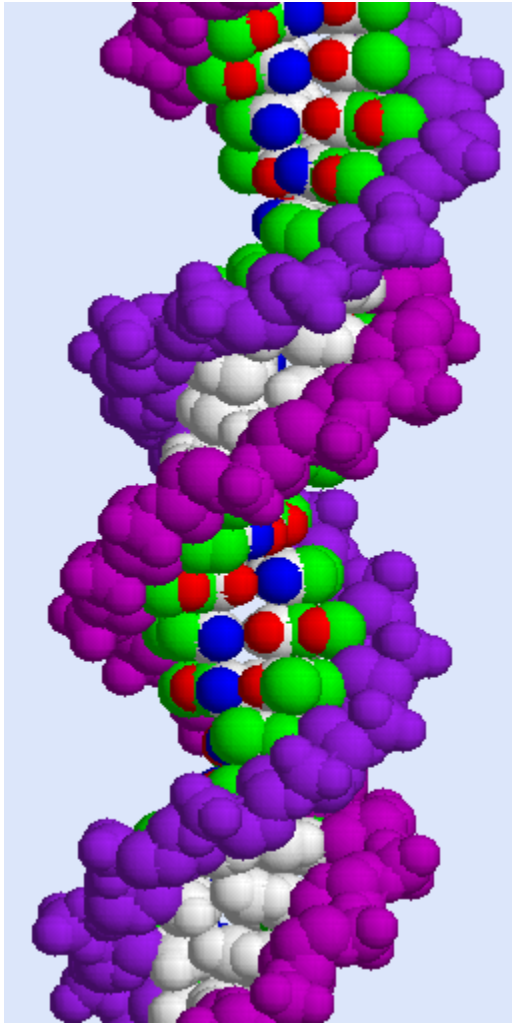
Водороды не видны, т.к. маленькие и РСА их не видит

Атомы остова – фиолетовые двух оттенков чтобы различать цепочки.

В большой бороздке **красный – кислород**, **синий – азот**, **зеленый – углерод**. Малая бороздка не окрашена

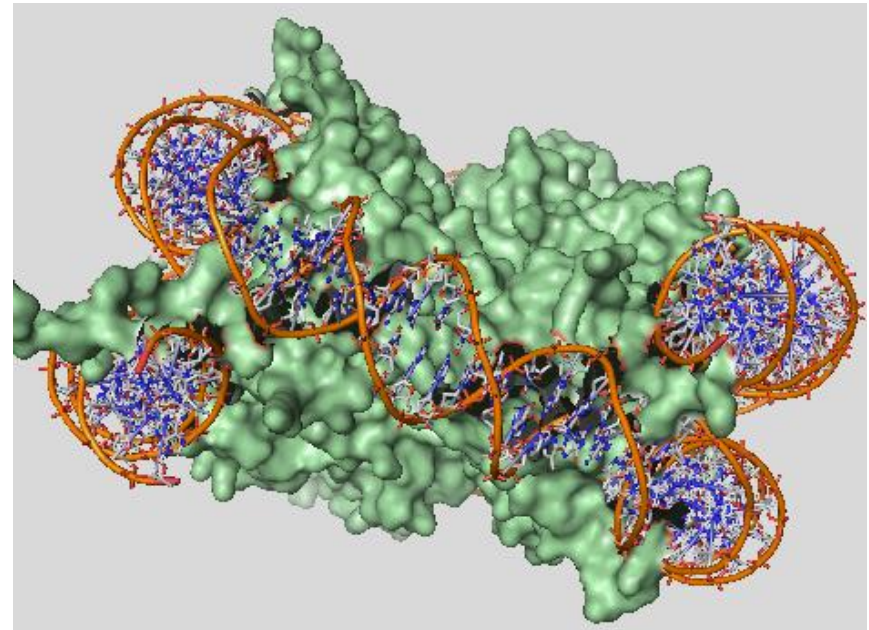
БЕЛКИ (некоторые) МОГУТ УЗНАТЬ ПОСЛЕДОВАТЕЛЬНОСТЬ УЧАСТКА ДНК по расположению N, O, C в большой бороздке. СИГНАЛЫ!

Лучше один раз увидеть



Двойная
спираль
ДНК.

Раскраска
моя ААл 😊



Нуклеосома:
ДНК человека на
"катушке" из гистонов:
вид сбоку (гистоны –
такие белки)

Обе структуры расшифрованы с помощью рентгеноструктурного анализа.

Пространственная структура РНК отличается от таковой у ДНК

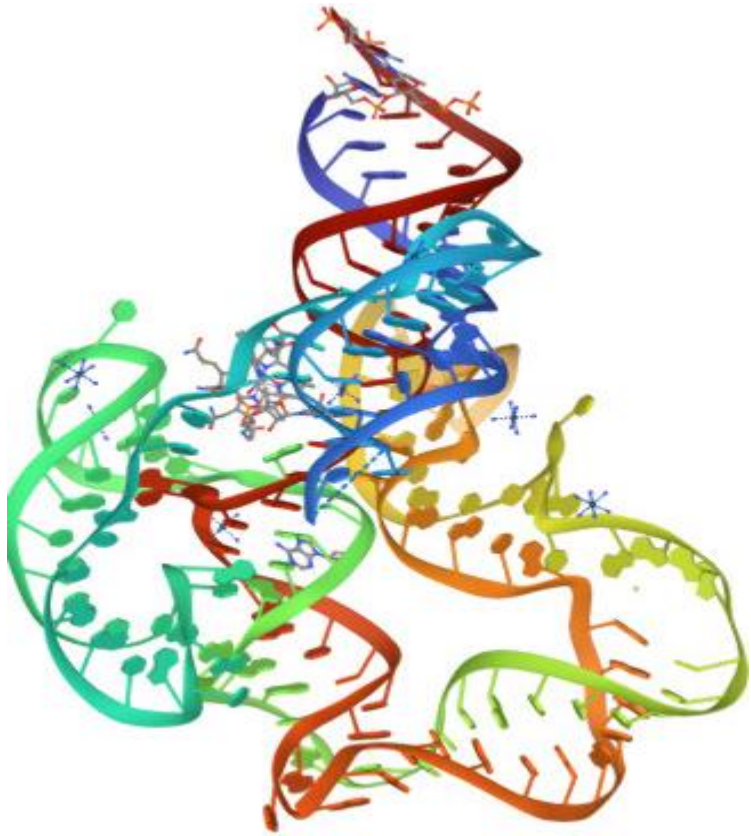


Рис. PDB code [4GXY](#)

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

Видны участки dsRNA!

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.

9. Что записано в геноме (и б.м. понятно нам)

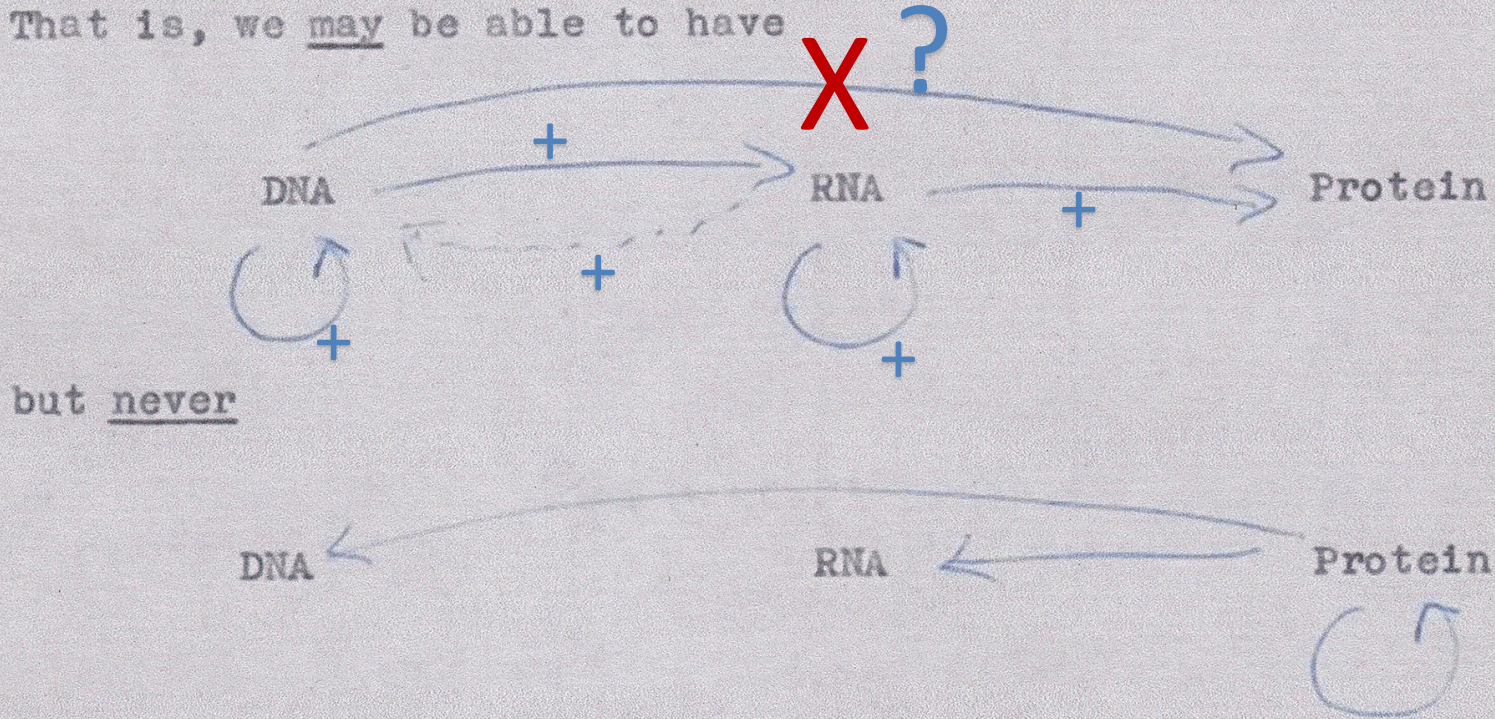
- **Гены белков** – участки ДНК, кодирующие аминокислотную последовательность белка.
 - **ДНК-зависимая РНК-полимераза** переписывает участок ДНК содержащий ген белка в матричную РНК с той же последовательностью оснований (с заменой Т на U)
 - **Рибосома** в соответствии с триплетами мРНК синтезирует белок
- **Гены молекул РНК**, отличных от мРНК
- **Сигналы** для белков и молекулярных машин
- Много чего другого

Таблица генетического кода

AAA	K	CAA	Q	GAA	E	TAA	Stop
AAG	K	CAG	Q	GAG	E	TAG	Stop
AAC	N	CAC	H	GAC	D	TAC	Y
AAT	N	CAT	H	GAT	D	TAT	Y
ACA	T	CCA	P	GCA	A	TCA	S
ACG	T	CCG	P	GCG	A	TCG	S
ACC	T	CCC	P	GCC	A	TCC	S
ACT	T	CCT	P	GCT	A	TCT	S
AGA	R	CGA	R	GGA	G	TGA	Stop
AGG	R	CGG	R	GGG	G	TGG	W
AGC	S	CGC	R	GGC	G	TGC	C
AGT	S	CGT	R	GGT	G	TGT	C
ATA	I	CTA	L	GTA	V	TTA	L
ATG	M	CTG	L	GTG	V	TTG	L
ATC	I	CTC	L	GTC	V	TTC	F
ATT	I	CTT	L	GTT	V	TTT	F

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



where the arrows show the transfer of information.

Что записано в геноме

- **Гены белков** – участки ДНК, кодирующие аминокислотную последовательность белка.
 - **ДНК-зависимая РНК-полимераза** переписывает участок ДНК содержащий ген белка в матричную РНК с той же последовательностью оснований (с заменой Т на U)
 - **Рибосома** в соответствии с триплетами мРНК синтезирует белок
- **Гены молекул РНК**, отличных от мРНК
- **Сигналы для белков и молекулярных машин**

10. Геном человека

Референсный геном человека

- двадцать две аутосомы (в ядре)
- две половые хромосомы X и Y (в ядре)
- митохондриальная ДНК человека (не в ядре – в цитоплазме клетки)
- содержат вместе 3 099 734 149 пар оснований

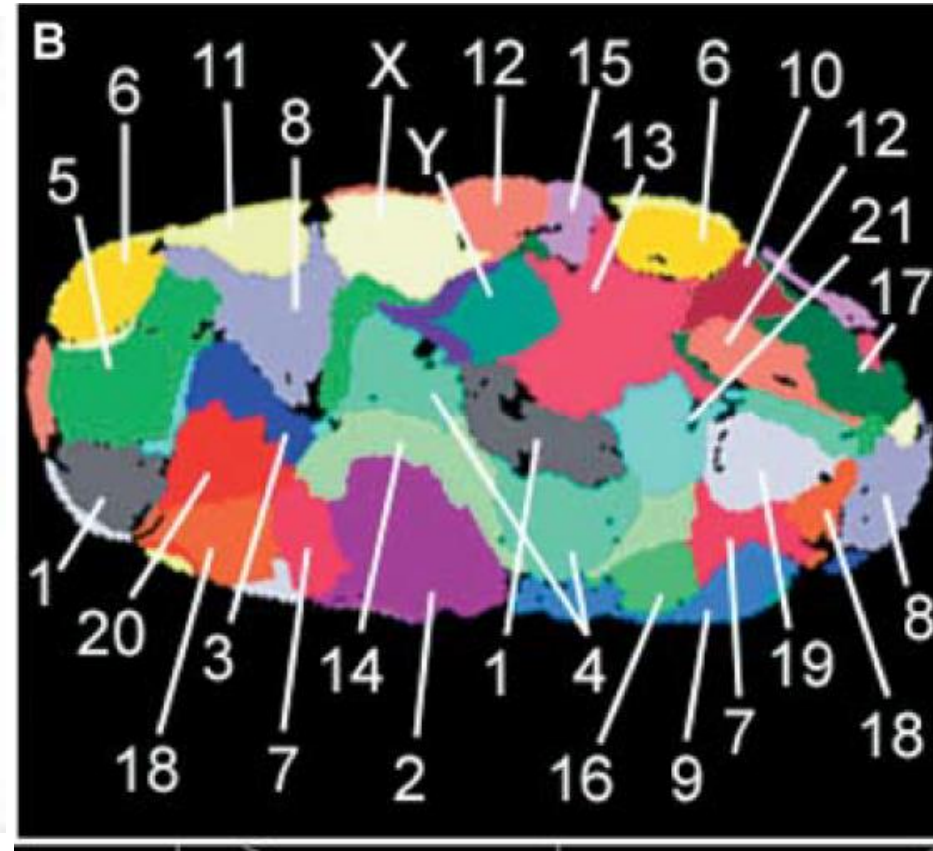
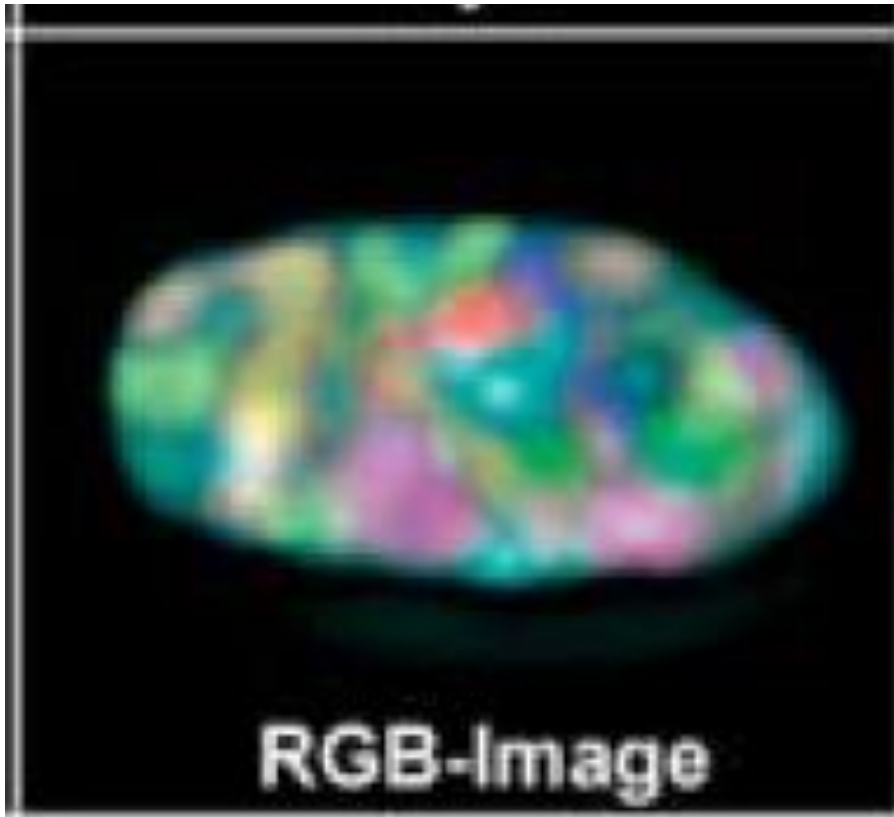
около 3 миллиардов букв А, Т, G, С

В ядре клетки содержится такой набор хромосом

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

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

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



Клетка находится перед стадией деления: каждая хромосома состоит из 2х одинаковых ДНК после удвоения (репликации). При делении клетки он разойдутся в разные дочерние клетки

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 – 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.

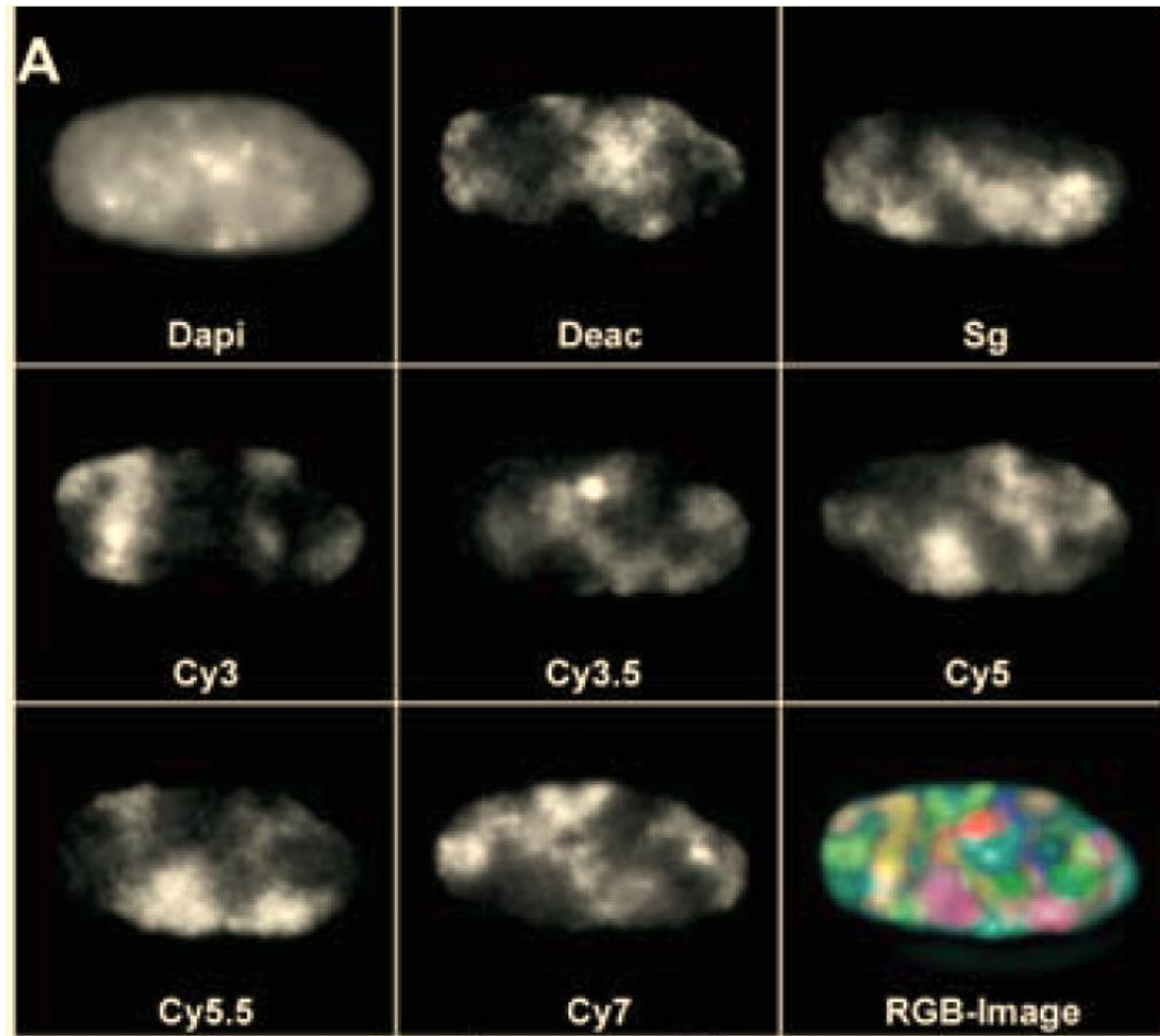
(А) Деконволюция (объединение) флюоресцентных микрофотографий в 7 каналах

(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 7 каналов

(В) Прорисовка с искусственно подобранными цветами изображения (А)

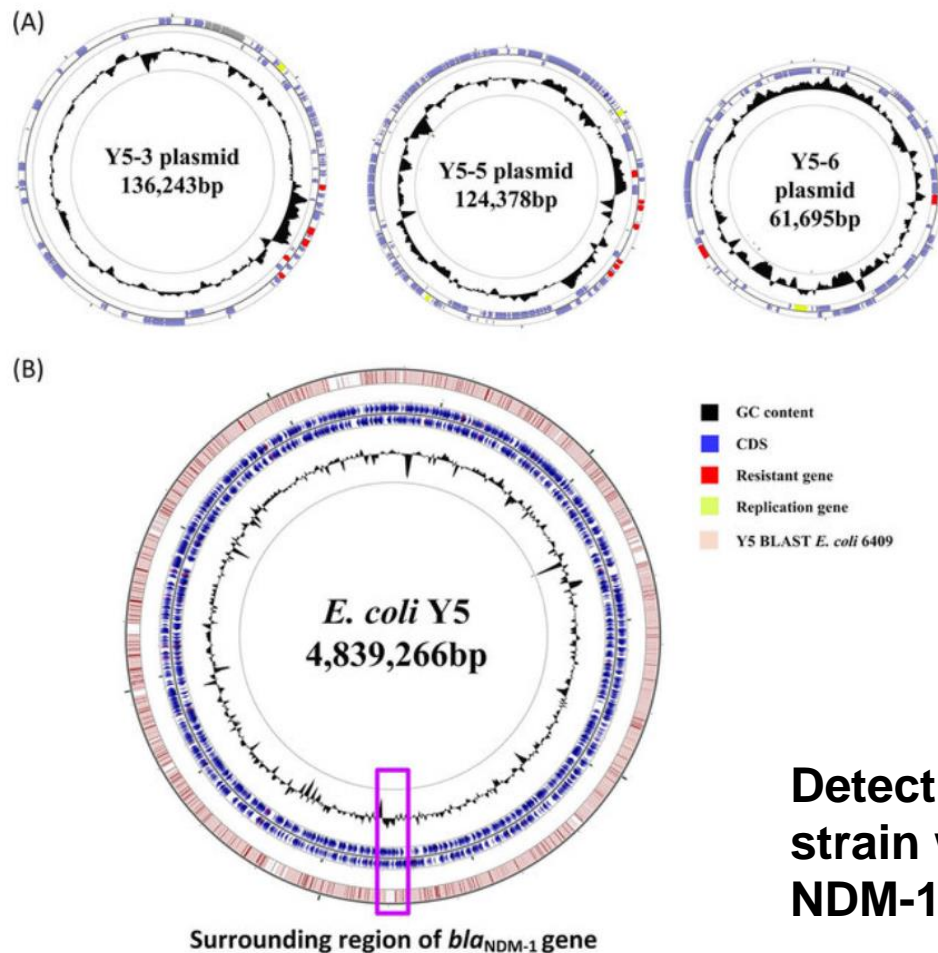
Исходные микрофотографии в восьми каналах и их совмещение



11. Геном бактерии

- Одна или несколько кольцевых молекул ДНК – хромосом
- Размер хромосомы порядка 1 млн п.н.
- Ноль или несколько маленьких ДНК, называемых плазмидами.
- Бактерии обитающие в одном микробиоме обмениваются плазмидами и фрагментами хромосом. Для эволюции – это неполовой способ обмена генетическим материалом
- Пример – обмен плазмидами, несущими устойчивость к антибиотикам

Геном кишечной палочки E.coli штамм Y5. Одна хромосомом 3 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](#)

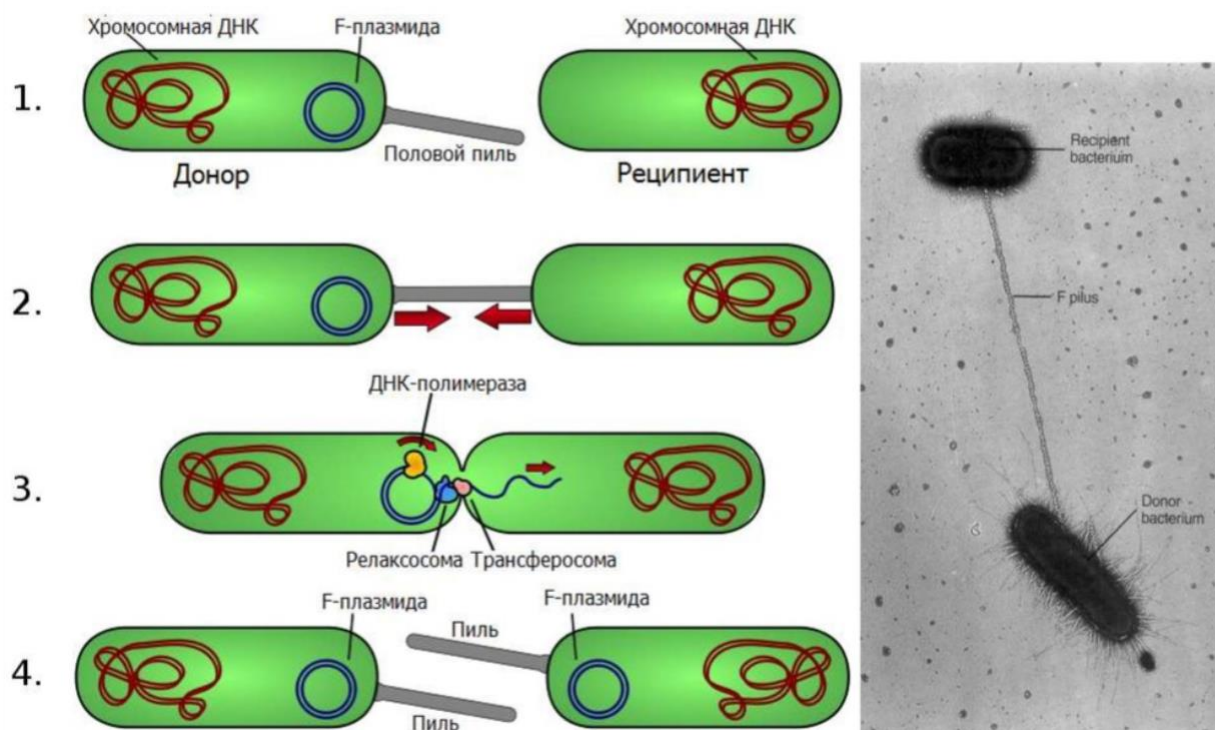
This figure was uploaded by [Zhi Ruan](#)

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

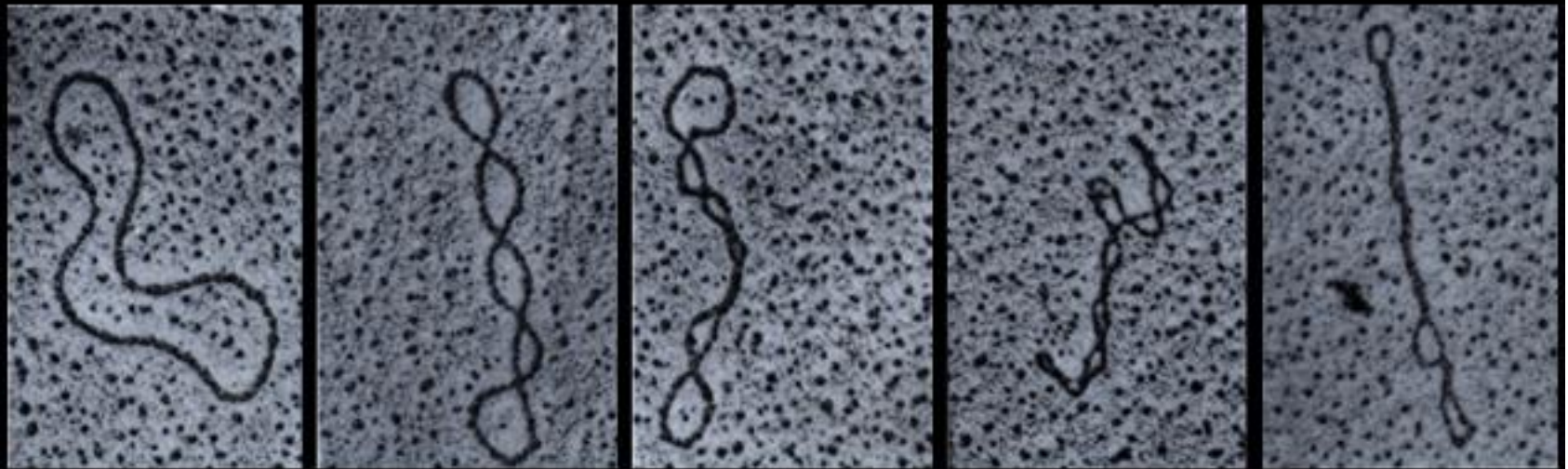
Один из способов обмена ДНК между бактериями

Конъюгация у бактерий



Слайд из презентации <https://ppt-online.org/388476>, посвященной генетике бактерий. Автор мне неизвестен. Понятная подача материала

Микрофотография маленькой кольцевой ДНК бактерии - плазмиды



No
supercoiling



Tightly
supercoiled

Маленькая **плазмида** бактерии
Электронная микроскопия

12. Геном адено-ассоциированного вируса AAV2 используют для генной терапии людей с неподдающимися иному лечению патологиями.

Проходят клинические испытания.

Wang et al. , Adeno-associated virus vector as a platform for gene therapy delivery, 2019

Список заболеваний при которых применяется генная терапия с помощью AAV2. Фазы испытаний.

Primary gene delivery target	Condition	AAV capsid	Transgene product	Strategy	Sponsor	Phase	ClinicalTrials.gov identifier
Brain	AADC deficiency	AAV2	AADC	Replacement	Krystof Bankiewicz, UCSF	Phase I	NCT02852213
		AAV2	AADC	Replacement	National Taiwan University Hospital	Phase II	NCT02926066
	Batten disease (CLN2)	AAVrh.10	CLN2	Replacement	Weill Cornell	Phase I/II	NCT01414985
	Batten disease (CLN6)	AAV9	CLN6	Replacement	Nationwide Children's Hospital	Phase I/II	NCT02725580
	MPS-III B	AAV5	NAGLU	Replacement	uniQure	Phase I/II	NCT03300453
	Parkinson disease	AAV2	AADC	Addition	Jichi Medical University	Phase I/II	NCT02418598
			GDNF	Addition	NINDS	Phase I	NCT01621581
			Neurturin	Addition	Sangamo	Phase I/II	NCT00985517
			AADC	Addition	Voyager	Phase I	NCT03065192
Spinal cord	SMA	AAV9	SMN	Replacement	AveXis	Phase III	NCT03461289
	Giant axonal neuropathy	AAV9	GAN	Replacement	NINDS	Phase I	NCT02362438
Eye	Achromatopsia	AAV2	CNGB3	Replacement	AGTC	Phase I/II	NCT02599922
		AAV8	CNGB3	Replacement	MeiraGTx	Phase I/II	NCT03001310
	Choroideraemia	AAV2	REP1	Replacement	Nightstar	Phase III	NCT03496012
			REP1	Replacement	Spark	Phase I/II	NCT02341807
			REP1	Replacement	STZ eyetrial	Phase II	NCT02671539
			REP1	Replacement	University of Oxford	Phase II	NCT02407678
	LCA	AAV2	RPE65	Replacement	Spark	Phase III	NCT00999609
			AAV5	RPE65	Replacement	MeiraGTx	Phase I/II

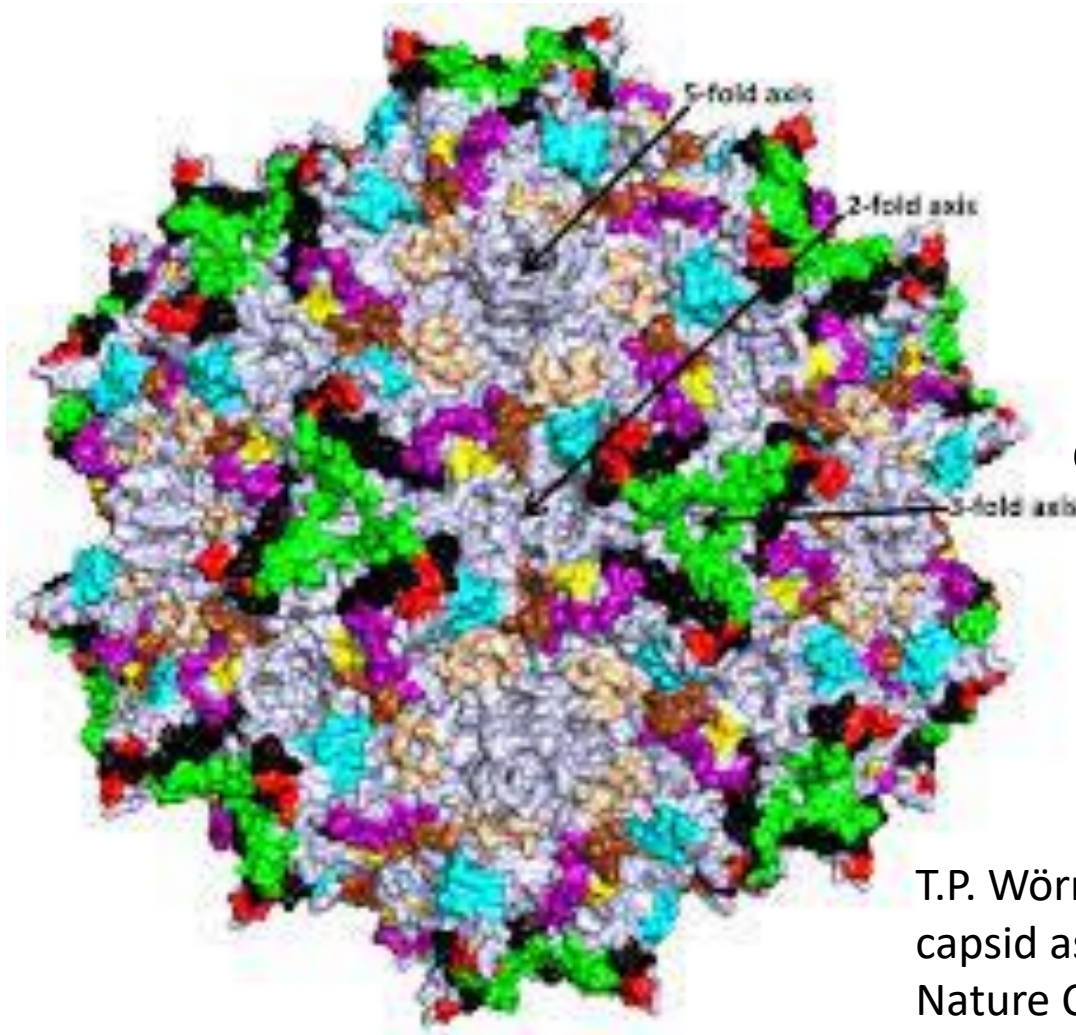
Table 1 (cont.) | A selection of ongoing rAAV interventional clinical trials

Primary gene delivery target	Condition	AAV capsid	Transgene product	Strategy	Sponsor	Phase	ClinicalTrials.gov identifier
Liver	Haemophilia B	AAV8	FIX	Replacement	Shire	Phase I/II	NCT01687608
		ND	FIX	Replacement	Pfizer	Phase II	NCT02484092
		ND	FIX	Replacement	Pfizer	Phase III	NCT03587116
		AAV6	FIX	Replacement	Sangamo	Phase I	NCT02695160
		AAV8	FIX	Replacement	St. Jude Children's Research Hospital	Phase I	NCT00979238
		AAV5	FIX	Replacement	uniQure	Phase III	NCT03569891
		ND	FIX	Replacement	UCL	Phase I	NCT03369444
	MPS-I	AAV6	ZFN1, ZFN2 and IDUA donor	Editing	Sangamo	Phase I	NCT02702115
	MPS-II	AAV6	ZFN1, ZFN2 and IDS donor	Editing	Sangamo	Phase I	NCT03041324
	MPS-IIIa	AAVrh.10	SGSH	Replacement	LYSOGENE	Phase II/III	NCT03612869
	MPS-VI	AAV8	ARSB	Replacement	Fondazione Telethon	Phase I/II	NCT03173521
	OTC deficiency	AAV8	OTC	Replacement	Ultragenyx	Phase I/II	NCT02991144
Muscle	A1AT deficiency	AAV2	A1AT	Replacement	UMMS	Phase I	NCT00377416
	CMT1A	AAV1	NTF3	Addition	Nationwide Children's Hospital	Phase I/II	NCT03520751
	DMD	AAVrh.74	Micro-dystrophin	Replacement	Nationwide Children's Hospital	Phase I/II	NCT03375164
		AAV9	Mini-dystrophin	Replacement	Pfizer	Phase I	NCT03362502
		AAV9	Micro-dystrophin	Replacement	Solid Biosciences	Phase I/II	NCT03368742
	Dysferlinopathy	AAVrh.74	DYSF	Replacement	Nationwide Children's Hospital	Phase I	NCT02710500
	HIV infections	AAV1	PG9 antibody	Addition	International AIDS Vaccine Initiative	Phase I	NCT01937455
		AAV8	VRC07 antibody	Addition	NIAID	Phase I	NCT03374202
	Pompe disease	AAV8	GAA	Replacement	Actus Therapeutics	Phase I/II	NCT03533673
		AAV9	GAA	Replacement	University of Florida	Phase I	NCT02240407
	X-linked MTM	AAV8	MTM1	Replacement	Audentes	Phase I/II	NCT03199469

Adeno-associated viruses (AAV)

- AAV поражают людей и других приматов
- Для быстрого размножения им необходим вирус-помощник: аденовирус (AdV) или некоторые другие
- При отсутствии вируса-помощника они встраиваются в геном зараженной клетки и латентно реплицируются по мере репликации клетки-хозяина неограниченно долго
- Если аденовирус заражает КЛЕТКУ, ЗАРАЖЁННУЮ AAV, то AAV получает возможность быстро размножиться, выходит из клетки и заражает новые клетки

Так выглядит ААV в собранном виде



РСА расшифровка
капсида ААV2

Капсид имеет
симметрию икосаэдра

T.P. Wörner et al., Adeno-associated virus capsid assembly is divergent and stochastic, Nature Communications 12:1642 (2021)

Геном AAV

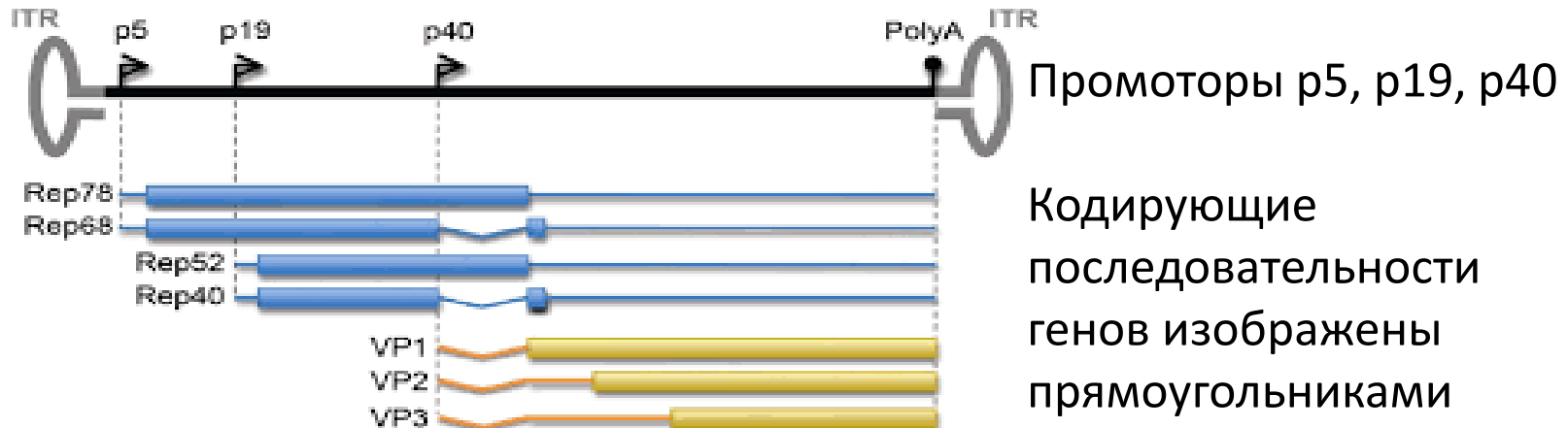
- Геном AAV – ssDNA одна цепочка ДНК!
- В геноме закодировано ДВА гена: **rep** и **cap**
- **rep** участвует в репликации
- **cap** – белок образующий икосаэдральный капсид вириона, содержащий ssDNA
- 5' и 3' концы ДНК содержат инвертированные повторы из 145 нуклеотидов
- Хотя генов два, они кодируют 7 белков благодаря альтернативному сплайсингу.



Придумал непростое задание на дом про этот геном.
Будьте внимательны)))

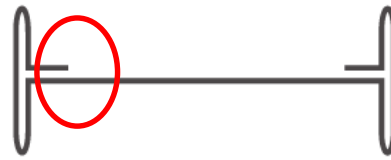
Промоторы и альтернативный сплайсинг

- Промотор – это участок dsDNA содержащий сигнал
 - для ДНК-зависимой РНК полимеразы (которая делает мРНК)
 - о том, что после него (промотора) надо начинать транскрипцию – переписывание гена на мРНК
- Сплайсинг мРНК состоит в том, что удаляются из мРНК определенные участки, называемые интронами. Какие именно – указывается сигналами сплайсинга.
- Альтернативный сплайсинг состоит в том, что какой-то интрон не удаляется, т.е. в белок вставляется фрагмент.
- Хотя генов в геноме AAV два, они кодируют 7 белков благодаря альтернативному сплайсингу.



Для репликации генома AAV нужен хозяйский белок RPA, заделывающий участки ssDNA в dsDNA

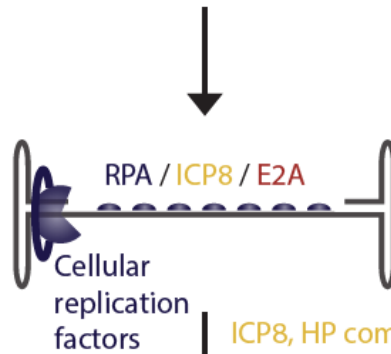
1. Incoming viral genome self-anneals to form terminal hairpins



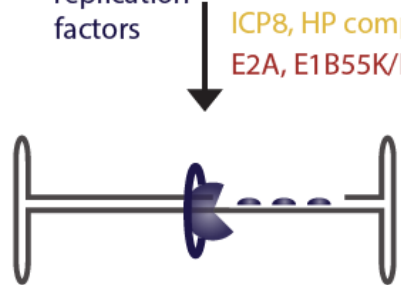
RPA реагирует на обрыв dsDNA как в овале.

Белки Rep78 и Rep68 связываются с тем же местом, препятствуя связыванию RPA

2. Replication factors are recruited



3. A partial dsDNA molecule is synthesized using the ITR structure as 3'-OH primer



Белки вируса помощников (жёлтые и красные) прогоняют Rep78 и Rep68

4. dsAAV genome is transcribed and translated into AAV



... и далее. Получается dsDNA AAV

Векторы rAAV на основе AAV генома для генной терапии

Today, recombinant AAVs (rAAVs) are the leading platform for in vivo delivery of gene therapies. The first rAAV gene therapy product, alipogene tiparvovec (Glybera), was approved by the European Medicines Agency to treat lipoprotein lipase deficiency in 2012

Although the clinical success of rAAV gene therapy is encouraging, we must acknowledge the limitations and challenges of this gene delivery platform, which include issues with rAAV manufacturing and immunological barriers to delivery AAV is found in multiple vertebrate species, including human and non-human primates (NHPs).

The current consensus is that AAV does not cause any human diseases

rAAV не встраиваются в геном человека. Существуют в виде транскрибируемой dsDNA в виде Гантели (см. пред. Слайды)
Для их репликации добавляют белки бз AdV

Wang et al., Adeno-associated virus vector as a platform for gene therapy delivery, 2019

ДАЛЕЕ ПРИВЕДЕНА ПОСЛЕДОВАТЕЛЬНОСТЬ ПОЛНОГО ГЕНОМА AAV2

Она вам пригодится для выполнения
нетривиального домашнего задания

>NC_001401.2 Adeno-associated virus - 2, complete genome 4679 bp
TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGG
GCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCATCACTAGGGG
TTCCTGGAGGGGTGGAGTCGTGACGTGAATTACGTCATAGGGTTAGGGAGGTCTTGTATTAGAGGTCACG
TGAGTGTTTTTGCGACATTTTTGCGACACCATGTGGTCACGCTGGGTATTTAAGCCCGAGTGAGCACGCAGG
GTCTCCATTTTGAAGCGGGAGGTTTTGAACGCGCAGCCGCCATGCCGGGGTTTTACGAGATTGTGATTAAG
GTCCCCAGCGACCTTGACGAGCATCTGCCCGGCATTTCTGACAGCTTTGTGAACTGGGTGGCCGAGAAGG
AATGGGAGTTGCCGCCAGATTCTGACATGGATCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGA
GAAGCTGCAGCGCGACTTTCTGACGGAATGGCGCCGTGTGAGTAAGGCCCCGGAGGCCCTTTTTCTTTGTG
CAATTTGAGAAGGGAGAGAGCTACTTCCACATGCACGTGCTCGTGAAACCACCGGGGTGAAATCCATGG
TTTTGGGACGTTTCTGAGTCAGATTCGCGAAAACTGATTCAGAGAATTTACCGCGGGATCGAGCCGAC
TTTGCCAAACTGGTTCGCGGTCACAAAGACCAGAAATGGCGCCGGAGGCGGGAACAAGGTGGTGGATGAG
TGCTACATCCCCAATTACTTGCTCCCCAAAACCCAGCCTGAGCTCCAGTGGGCGTGGACTAATATGGAAC
AGTATTTAAGCGCCTGTTTGAATCTCACGGAGCGTAAACGGTTGGTGGCGCAGCATCTGACGCACGTGTC
GCAGACGCAGGAGCAGAACAAGAGAATCAGAATCCAATCTGATGCGCCGGTGATCAGATCAAAAACCT
TCAGCCAGGTACATGGAGCTGGTCGGGTGGCTCGTGGACAAGGGGATTACCTCGGAGAAGCAGTGGATCC
AGGAGGACCAGGCCTCATACATCTCCTTCAATGCGGCCTCCAACCTCGCGGTCCCAAATCAAGGCTGCCTT
GGACAATGCGGGAAAGATTATGAGCCTGACTAAAACCGCCCCGACTACCTGGTGGGCCAGCAGCCCGTG
GAGGACATTTCCAGCAATCGGATTTATAAAATTTTGGAACTAAACGGGTACGATCCCAAATATGCGGCTT
CCGTCTTTCTGGGATGGGCCACGAAAAGTTCGGCAAGAGGAACACCATCTGGCTGTTTGGGCCTGCAAC
TACCGGGAAGACCAACATCGCGGAGGCCATAGCCCACACTGTGCCCTTCTACGGGTGCGTAAACTGGACC
AATGAGAACTTTCCCTTCAACGACTGTGTGACAAAGATGGTGATCTGGTGGGAGGAGGGGAAGATGACCG
CCAAGGTCGTGGAGTCGGCCAAAGCCATTCTCGGAGGAAGCAAGGTGCGCGTGGACCAGAAATGCAAGTC
CTCGGCCCAGATAGACCCGACTCCCGTGATCGTCACTCCAACACCAACATGTGCGCCGTGATTGACGGG
AACTCAACGACCTTCGAACACCAGCAGCCGTTGCAAGACCGGATGTTCAAATTTGAACTCACCCGCCGTC
TGGATCATGACTTTGGGAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGATCACGT
GGTTGAGGTGGAGCATGAATTCTACGTCAAAAAGGGTGGAGCCAAGAAAAGACCCGCCCCAGTGACGCA
GATATAAGTGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTCAGACGCGGAAGCTTCGA
TCAACTACGCAGACAGGTACCAAAAACAATGTTCTCGTCACTGGGCATGAATCTGATGCTGTTTCCCTG
CAGACAATGCGAGAGAATGAATCAGAATTCAAATATCTGCTTCACTCACGGACAGAAAGACTGTTTAGAG
TGCTTTCCCGTGTCAGAATCTCAACCCGTTTCTGTGTCGTCAAAAGGCGTATCAGAAACTGTGCTACATTC
ATCATATCATGGGAAAGGTGCCAGACGCTTGCACCTGCCTGCGATCTGGTCAATGTGGATTTGGATGACTG
CATCTTTGAACAATAAATGATTTAAATCAGGTATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGA

CACTCTCTCTGAAGGAATAAGACAGTGGTGGAAAGCTCAAACCTGGCCACCACCACCAAAGCCCGCAGAG
CGGCATAAGGACGACAGCAGGGGTCTTGTGCTTCCTGGGTACAAGTACCTCGGACCCTTCAACGGACTCG
ACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGCACGACAAAGCCTACGACCGGCAGCT
CGACAGCGGAGACAACCCGTACCTCAAGTACAACCACGCCGACGCGGAGTTTCAGGAGCGCCTTAAAGAA
GATACGTCTTTTGGGGGCAACCTCGGACGAGCAGTCTTCCAGGCGAAAAAGAGGGTCTTGAACCTCTGG
GCCTGGTTGAGGAACCTGTAAAGACGGCTCCGGGAAAAAAGAGGGCCGGTAGAGCACTCTCCTGTGGAGCC
AGACTCCTCCTCGGGAACCGGAAAGGCGGGCCAGCAGCCTGCAAGAAAAAGATTGAATTTTGGTCAGACT
GGAGACGCAGACTCAGTACCTGACCCCCAGCCTCTCGGACAGCCACCAGCAGCCCCCTCTGGTCTGGGAA
CTAATACGATGGCTACAGGCAGTGGCGCACCAATGGCAGACAATAACGAGGGCGCCGACGGAGTGGGTAA
TTCCTCGGGAAATTTGGCATTGCGATTCCACATGGATGGGCGACAGAGTCATCACCACCAGCACCCGAACC
TGGGCCCTGCCACCTACAACAACCACCTCTACAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACA
ATCACTACTTTGGCTACAGCACCCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCACC
ACGTGACTGGCAAAGACTCATCAACAACAACCTGGGGATTCCGACCCAAGAGACTCAACTTCAAGCTCTTT
AACATTTCAAGTCAAAGAGGTCACGCAGAATGACGGTACGACGACGATTGCCAATAACCTTACCAGCACGG
TTCAGGTGTTTACTGACTCGGAGTACCAGCTCCCGTACGTCTCGGCTCGGCGCATCAAGGATGCCTCCC
GCCGTTCCCAGCAGACGTCCTTCATGGTGGCACAGTATGGATACTCACCTGAACAACGGGAGTCAGGCA
GTAGGACGCTCTTCATTTTACTGCCTGGAGTACTTTCTTCTCAGATGCTGCGTACCGGAAACAACCTTTA
CCTTCAGCTACACTTTTGGAGACGTTCCCTTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCT
CATGAATCCTCTCATCGACCAGTACCTGTATTACTTGAGCAGAACAACACTCCAAGTGGAAACCACCAG
CAGTCAAGGCTTCAGTTTTCTCAGGCCGGAGCGAGTGACATTCGGGACCAGTCTAGGAACTGGCTTCCTG
GACCCTGTTACCGCCAGCAGCGAGTATCAAAGACATCTGCGGATAACAACAACAGTGAATACTCGTGGAC
TGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCCCGGCCATGGCAAGCCAC
AAGGACGATGAAGAAAAGTTTTTTCCTCAGAGCGGGGTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAA
CAAATGTGGACATTGAAAAGGTCATGATTACAGACGAAGAGGAAATCAGGACAACCAATCCCCTGGCTAC
GGAGCAGTATGGTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAAGCAGCTACCGCAGATGTCAAC
ACACAAGGCGTTCTTCCAGGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGGCCCATCTGGGCAA
AGATTCCACACACGGACGGACATTTTACCCCTCTCCCCTCATGGGTGGATTTCGGACTTAAACACCCTCC
TCCACAGATTTCTCATCAAGAACACCCCGGTACCTGCGAATCCTTCGACCACCTTCAGTGGGCAAAGTTT
GCTTCCCTTCATCACACAGTACTCCACGGGACAGGTGAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAAA
ACAGCAAACGCTGGAATCCCGAAATTCAGTACACTTCCAACCTACAACAAGTCTGTTAATGTGGACTTTAC
TGTGGACACTAATGGCGTGTATTTCAGAGCCTCGCCCCATTGGCACCAGATACTGACTCGTAATCTGTAA
TTGCTTGTTAATCAATAAACCGTTTAATTCGTTTTCAGTTGAACTTTGGTCTCTGCGTATTTCTTTCTTAT
CTAGTTTCCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTAACCTACAAGGAACCCCTAGTGAT
GGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGC
CCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAA

13. Вирус SARS-CoV-2

- Ну его – надоел)))
- Ещё и 13й номер(((

Что записано в файле с геномом SARS-CoV-2 для людей?

- Последовательность
- Аннотация – формализованное описание того, что известно про последовательность

Из аннотации записи NC_045512 в банке данных Refseq на сайте NCBI

LOCUS NC_045512 29903 bp ss-RNA linear VRL 18-JUL-2020
DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1,
complete genome.
ACCESSION NC_045512

[gene](#) 21563..25384 /gene="S" /gene_synonym="spike glycoprotein"
[CDS](#) 21563..25384 /gene="S" spike protein" /product="surface
glycoprotein" /protein_id="[YP_009724390.1](#)"
/translation="MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFR
SSVLHSTQDLFLPFFSNVTWFHAIHVS GTNGTKRFDNPVLPFNDGVYFASTEKSNIIR
GWIFGTTLDSKTQSL LIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY

.....
.....

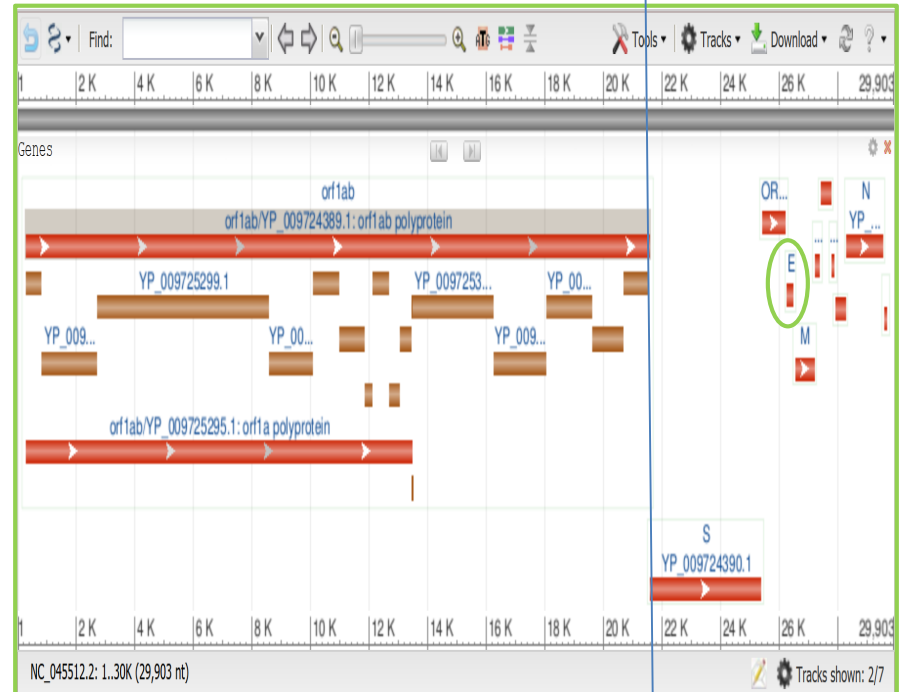
ГЕНЫ ORF1ab и ORF1a

красные полоски до вертикальной линии

А зрелые белки, которые они кодируют, изображены коричневыми полосками.

Как так?

И почему два гена на одном месте, но разной длины?



При заражении covid-19 в клетке хозяина (человека) оказывается РНК вируса.

РИБОСОМА (молекулярная машина для синтеза белков) опознаёт ее как мРНК - матричную РНК гена

Как удаётся коронавирусу выдать свою РНК за мРНК? Нужны соответствующие сигналы...

ТРАНСЛЯЦИЯ: СИНТЕЗ МОЛЕКУЛЫ БЕЛКА

В клетке для синтеза белка нужна матричная РНК

- мРНК это молекула РНК, на которой записана копия гена
(комплементарная к комплементарной цепочке гена!
«минус на минус = плюс»)
- Белок синтезирует рибосома используя мРНК
- В клетке человека много РНК разных типов
(в лекции А.Жариковой будет об этом).
- Рибосома отличает мРНК по двум сигналам
 - Специальная группа атомов **cap** на 5' конце.
 - **ПолиА** на 3'-конце

В клетке хозяина (человека) оказывается РНК коронавируса

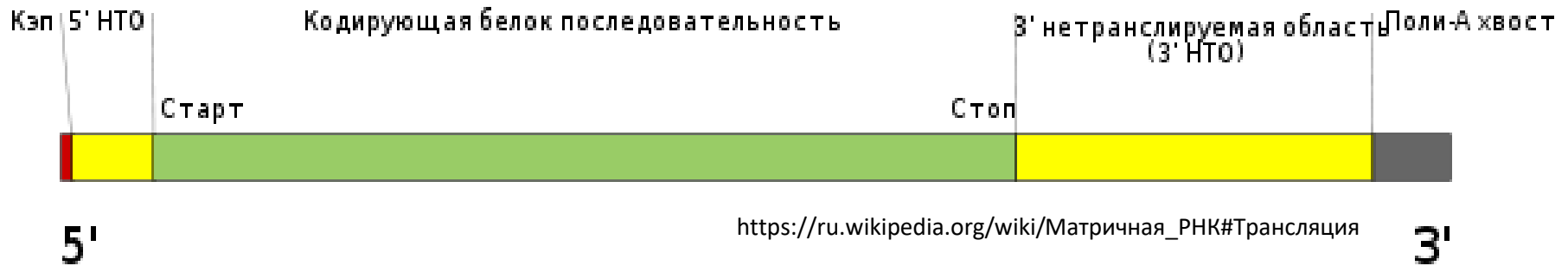
РНК коронавируса несет оба эти сигнала.

- ПолиА на 3'-конце РНК видны в файле с РНК (см)
- Кэп тоже есть

```
.....AAAATTAAATTTTAGTAGTGCTATCCCC  
ATGTGATTTTAAATAGCTTCTTAGGAGAAT  
GACAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAAAAA
```

Одна мРНК – один ген и один белок

У человека и других эукариот. Бывают исключения, но они редки!!!



* Малая субъединица рибосомы узнает кэп и сканирует мРНК до появления кодона ATG.

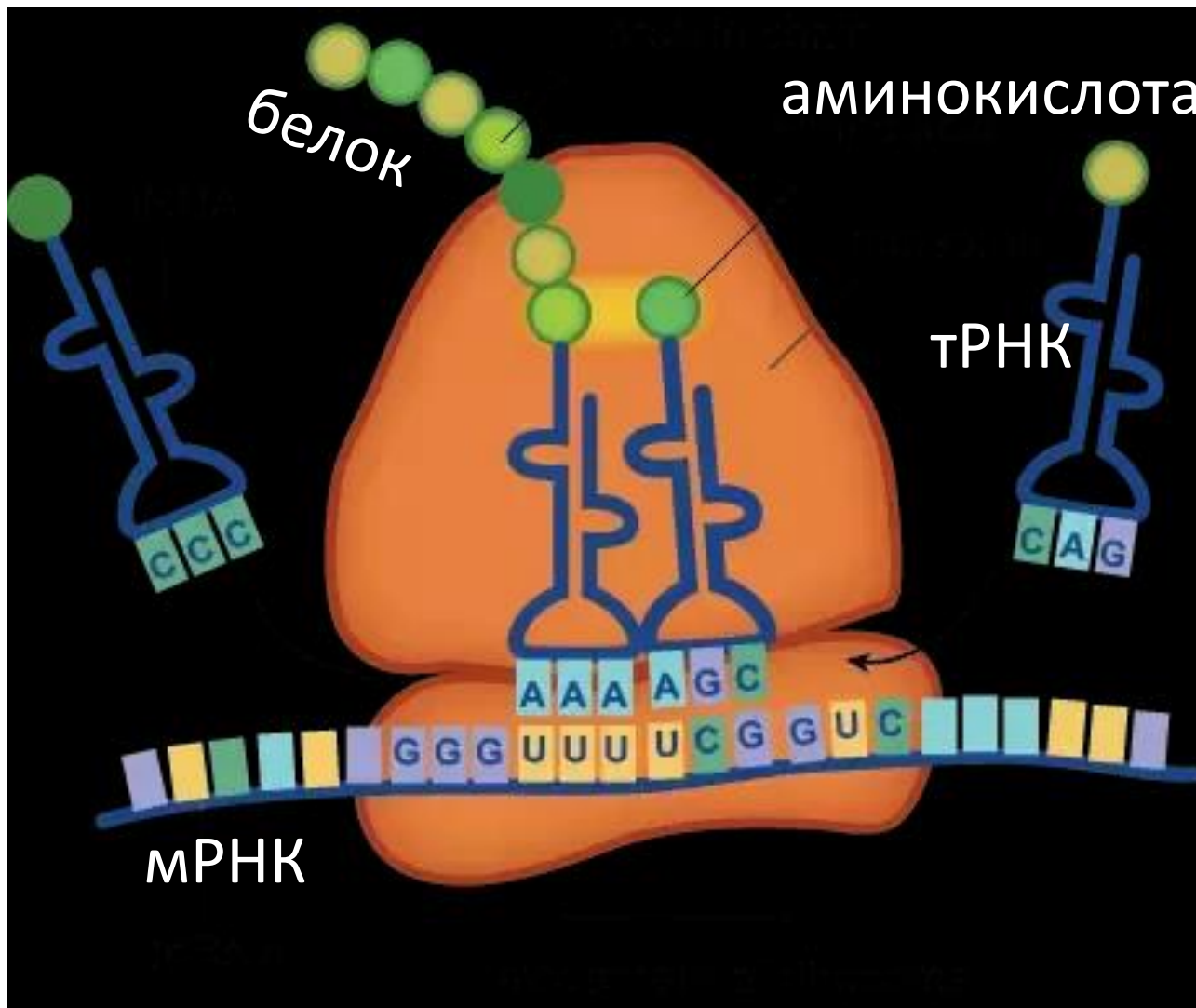
Причем не просто ATG, а ATG в подходящем окружении, так называемая последовательность Кóзак

* С триплета ATG начинается синтез белка рибосомой, по кодонам. Заканчивается на стоп кодоне.

По таблице генетического кода!

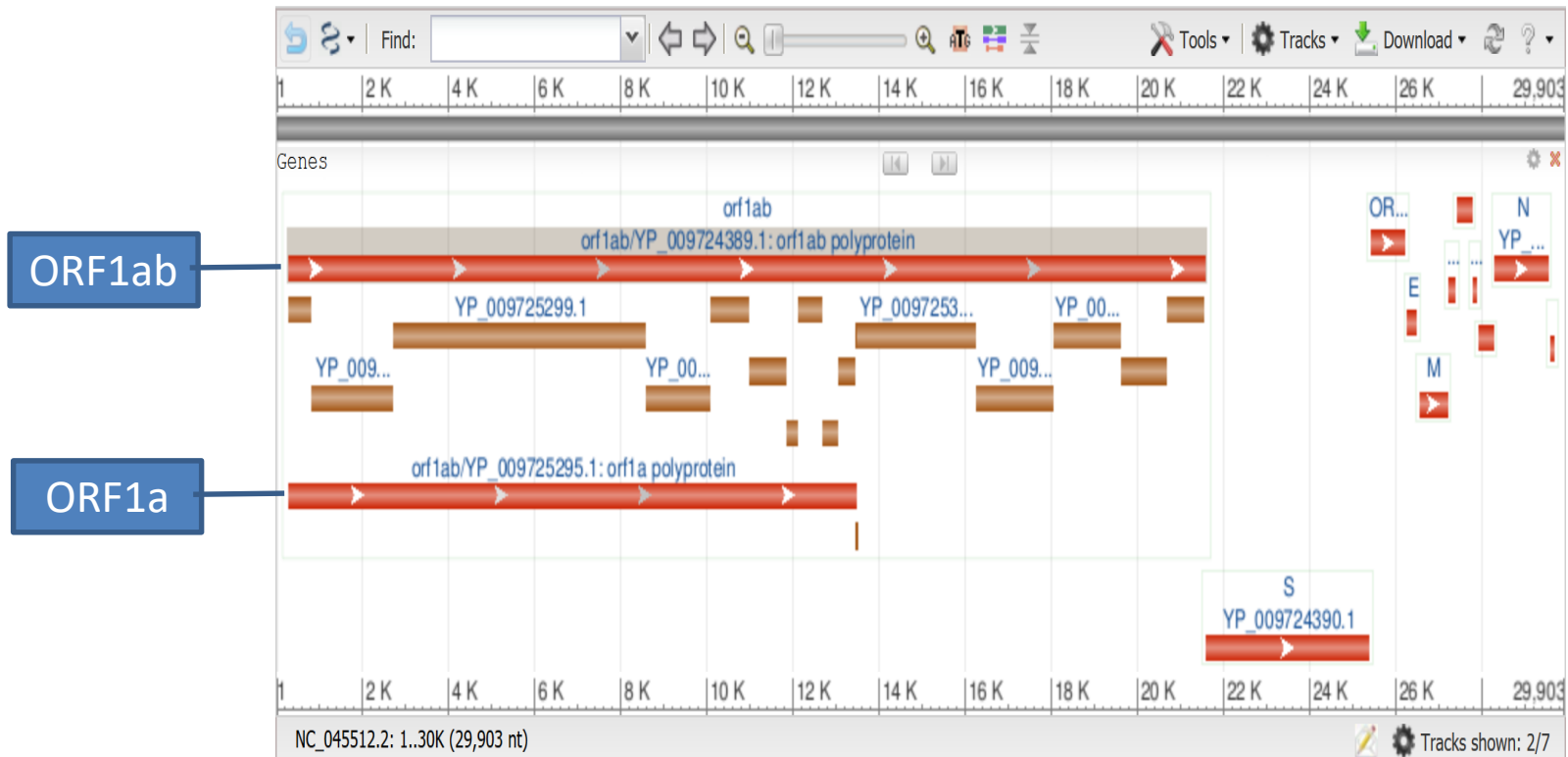
AAA	K	CAA	Q	GAA	E	TAA	Stop
AAG	K	CAG	Q	GAG	E	TAG	Stop
AAC	N	CAC	H	GAC	D	TAC	Y
AAT	N	CAT	H	GAT	D	TAT	Y
ACA	T	CCA	P	GCA	A	TCA	S
ACG	T	CCG	P	GCG	A	TCG	S
ACC	T	CCC	P	GCC	A	TCC	S
ACT	T	CCT	P	GCT	A	TCT	S
AGA	R	CGA	R	GGA	G	TGA	Stop
AGG	R	CGG	R	GGG	G	TGG	W
AGC	S	CGC	R	GGC	G	TGC	C
AGT	S	CGT	R	GGT	G	TGT	C
ATA	I	CTA	L	GTA	V	TTA	L
ATG	M	CTG	L	GTG	V	TTG	L
ATC	I	CTC	L	GTC	V	TTC	F
ATT	I	CTT	L	GTT	V	TTT	F

Так рибосомы делают белки



Ген ORF1a заканчивается стоп-кодом, как положено.

Так и транслируется рибосомой человека

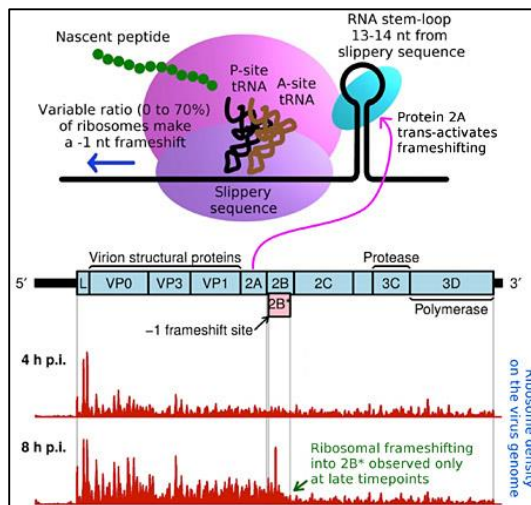
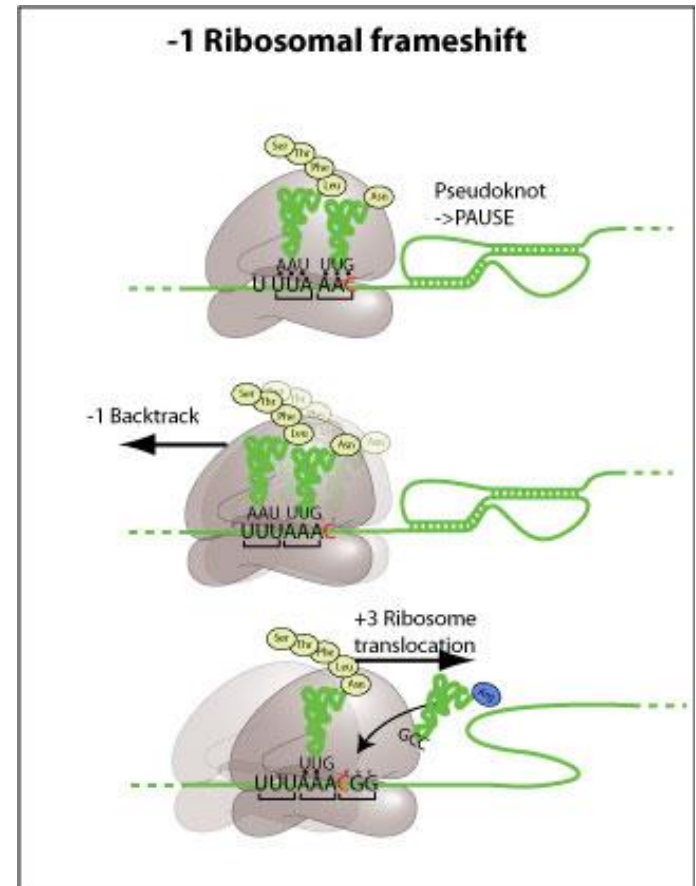
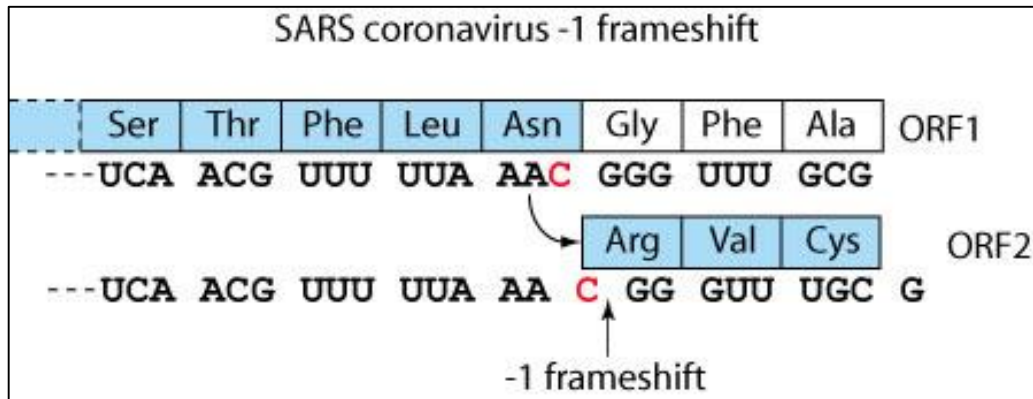


Как же транслируется ген ORF1ab???

Не бывает правил без исключений!

Программируемый рибосомный сдвиг.

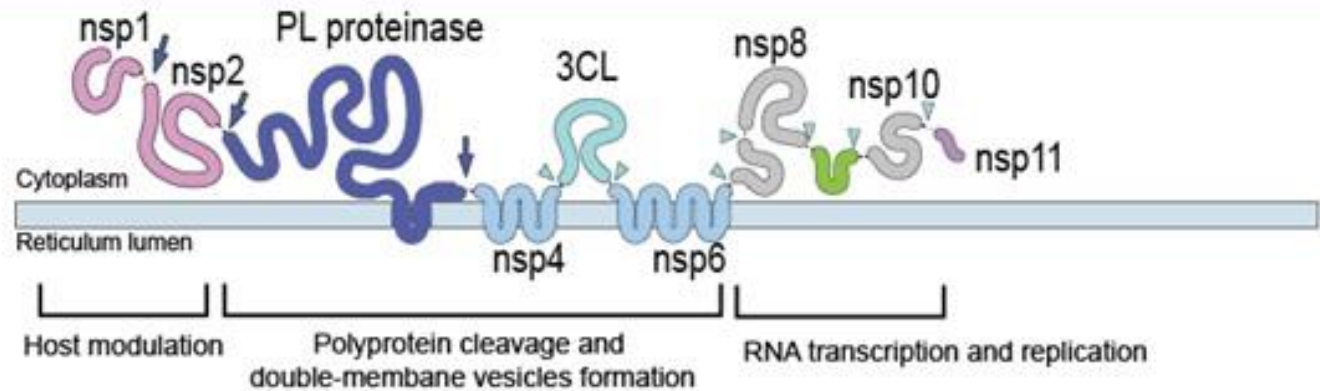
Рибосома останавливается из-за шпильки на РНК и slippery sequence. Отскакивает на ОДИН нуклеотид(букву). И продолжает синтез белка



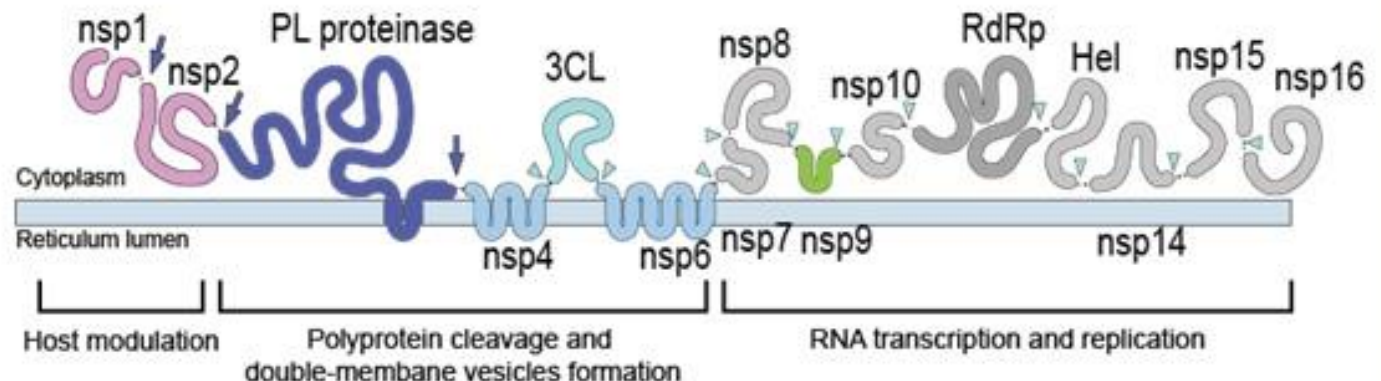
Продукты генов ORF1ab и ORF1a - Большие белки – полипротеины. См. след слайд

Они сами себя разрезают на отдельные белки

ORF1a =
polyprotein 1a
(pp1a)



ORF1b =
polyprotein 1ab
(pp1ab)



© ViralZone 2020
SIB Swiss Institute of Bioinformatics

Figure 2: SARS-Cov-2 polyproteins

There are two major, partially overlapping, ORFs in the 5' two-thirds of the viral genome. ORF1a is frameshifted to ORF1b, and these encode the replicase polyproteins pp1a and pp1ab respectively. Polyproteins pp1a and pp1ab are proteolytically cleaved into 16 putative non-structural proteins (nsps).

Функции некоторых ранних белков

Name	Число а/к	зачем нужен
NSP1	180	Деградирует некоторые хозяйские РНК
NSP3a	1945	Протеиназа, отрезает nsp1, nsp2, nsp3
NSP5a	306	Протеиназа, режет полипротеин в 11 местах
NSP8	198	Помогает при репликации РНК
NSP12a	932	Полимераза – по РНК делает комплементарную РНК (RDRp)
NSP13	601	Хеликаза (расплетает двойную спираль РНК)
NSP14	527	Присоединяет сар к РНК
NSP15a	346	Уклонение от защиты хозяйских клеток

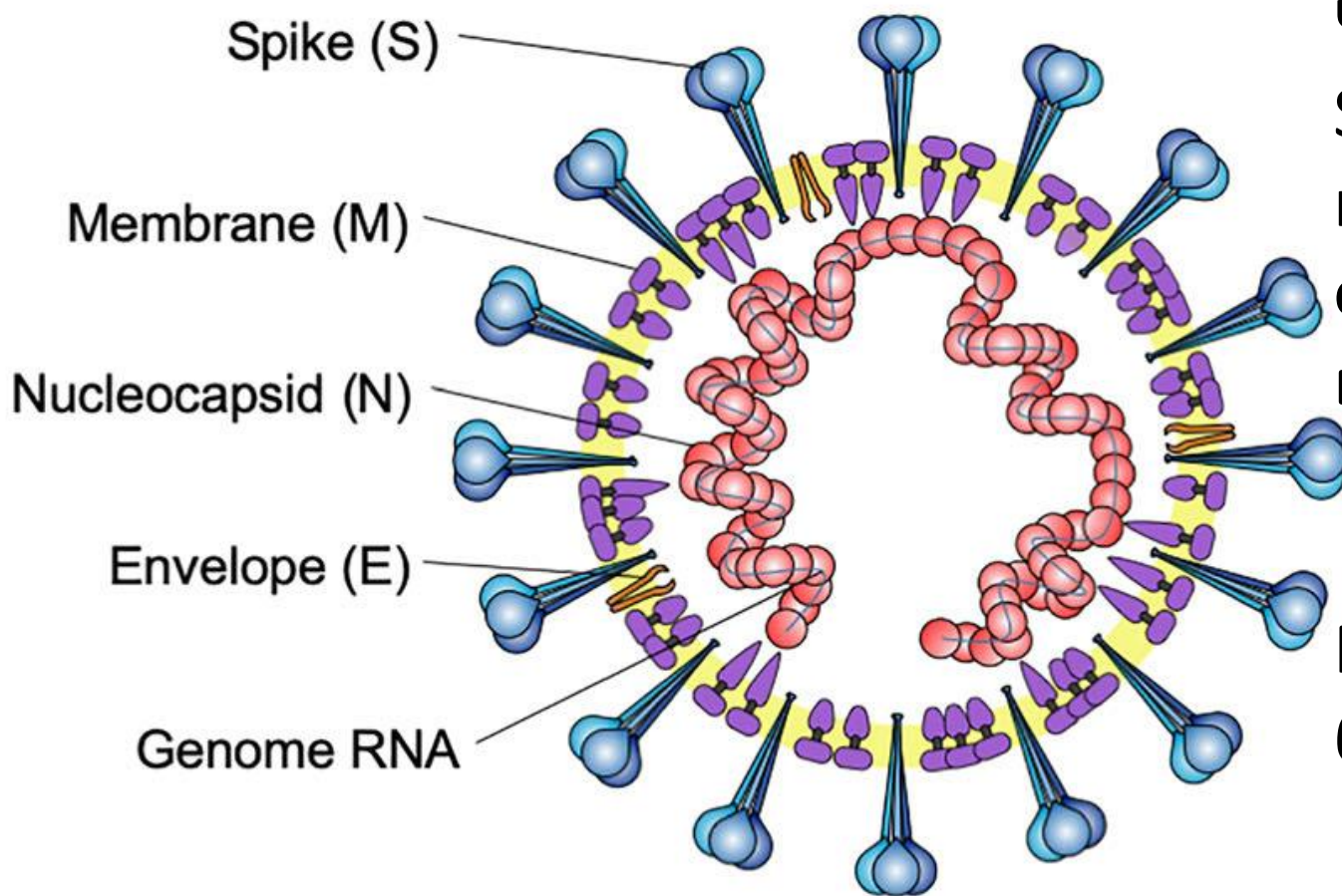
Как транслируются поздние белки?

Среди поздних генов все белки составляющие вирион, который существует пока вирус вне хозяина.

Оно и понятно – вирион собирается в конце заражения, когда есть много РНК – геномов и пора выходить из клетки.

Коронавирус вирус: белки

Схема вириона, существующего между заражениями



Четыре белка:
S, M, E, N
и РНК

оставляют
вирион

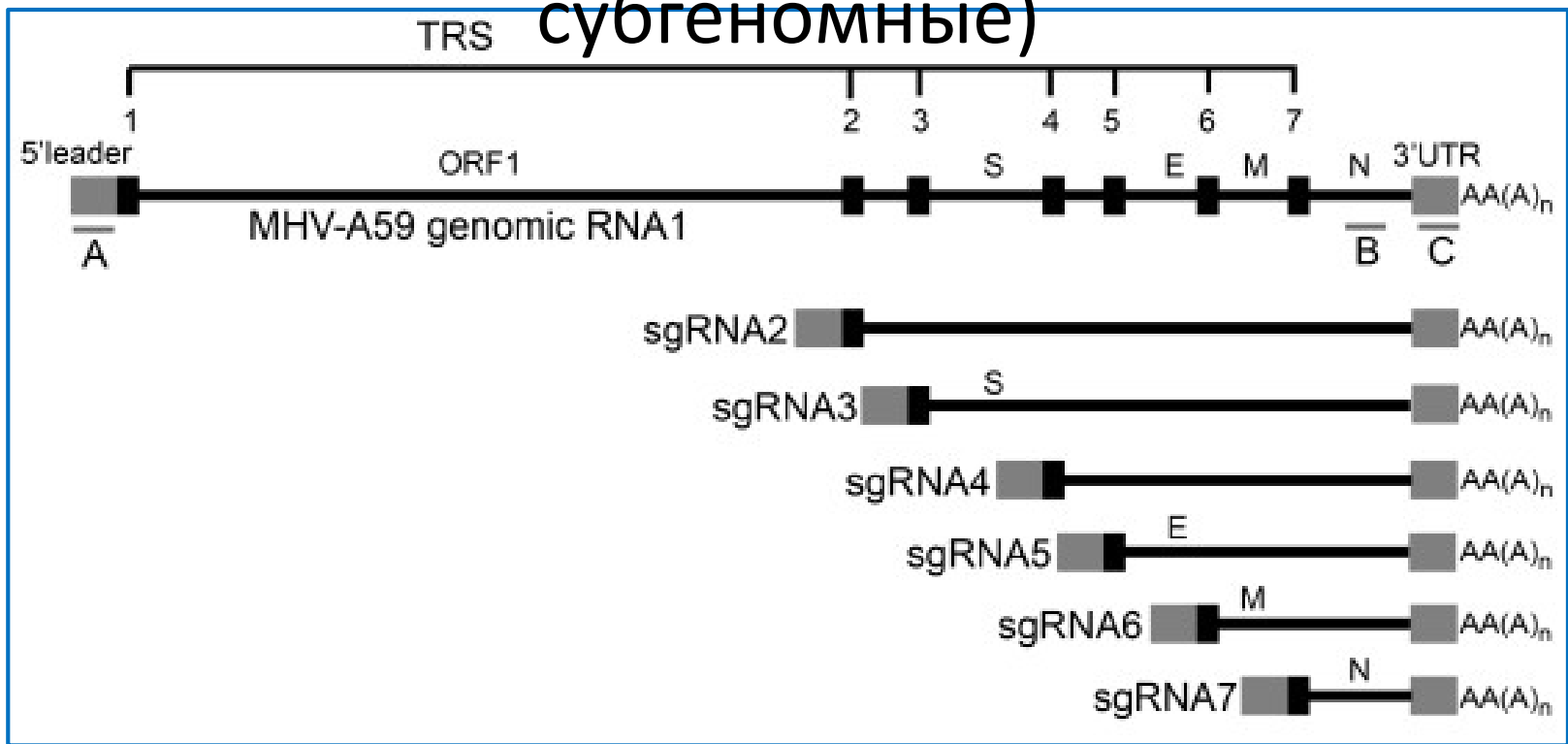
РНК облеплена
белками N

Функции белков

- **M белок** составляет оболочку (**капсид**) вириона, вместе с липидной мембраной (желтая)
- **E белок** нужен **для правильной кривизны капсида**.
2я функция - в хозяйской клетке. Пентамер E является ионным каналом в мембране органеллы ERGIC¹⁾
Коронавирусы, лишённые E, могут размножаться, хотя и менее патогенны
- **N белок** облепляет РНК в конформации **бусы на струне** для сохранности генома. При сборке капсида он обладает удивительной **способностью связываться только с РНК коронавируса!**

1) Между эндоплазматическим ретикулумом и аппаратом Гольджи)

Коронавирус синтезирует отдельные мРНК для поздних генов (называют сгмРНК, субгеномные)

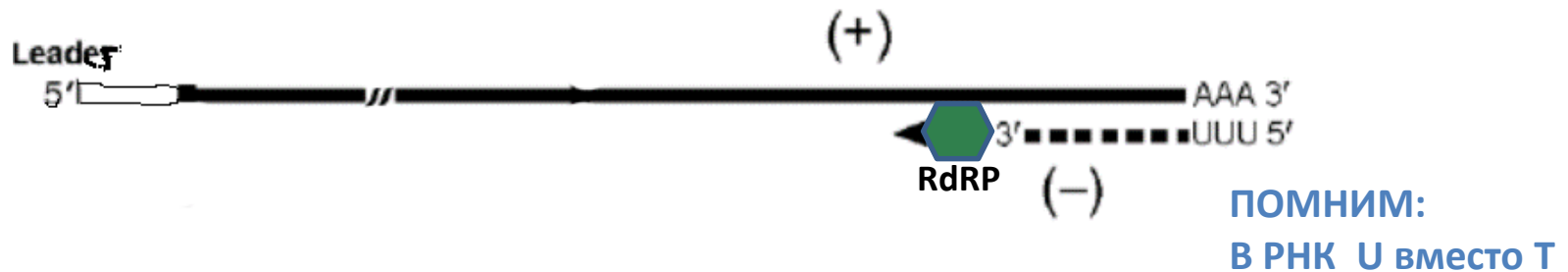


ИДЕЯ коронавируса: лидерную последовательность «склеить» с участком начиная от позднего гена и до конца!

Сохраняются все 5' концевые и 3' концевые сигналы (КЭП, полиА и др.)

СДЕЛАВ ЭТО ПРЕДОК КОРОНАВИРУСОВ ЗАКРИЧАЛ ЭВРИКА! И заразил множество хозяев.

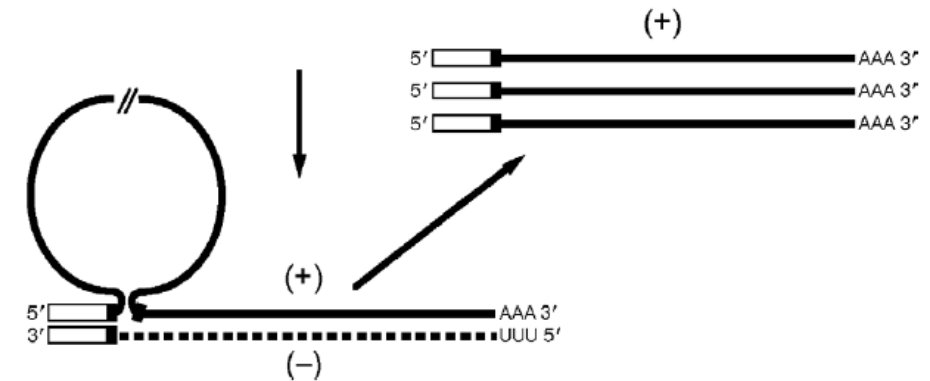
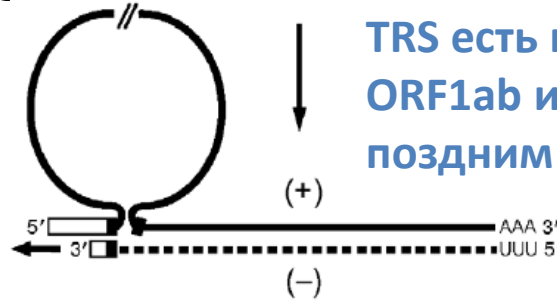
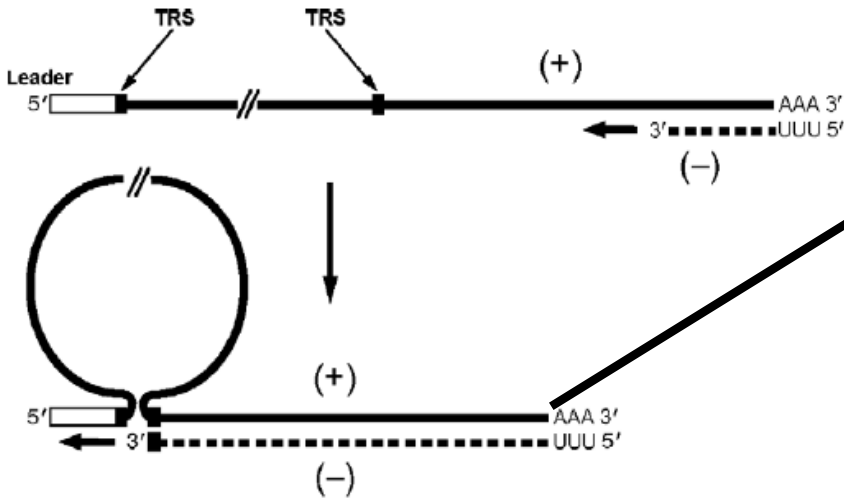
РНК зависимая РНК полимераза коронавируса



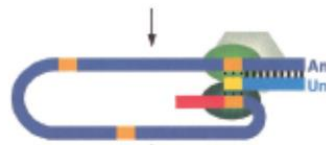
- Для синтеза новых РНК нужных для сборки новых частиц вируса нужен белок «РНК зависимая РНК полимераза, RdRP», его ген 11й среди зрелых ранних белков коронавирусов.
- С РНК коронавируса, которую обозначают +РНК, он делает комплементарную копию, называемую -РНК
- RdRP может копировать любую РНК, в частности, -РНК
- Минус на минус будет плюс!!!

Коронавирусам (эволюции) пришлось долго ломать голову чтобы придумать такое! Сигналы TRS – последовательности похожие на СТАААС – обозначены черным прямоугольником, желтым на цветном рисунке.

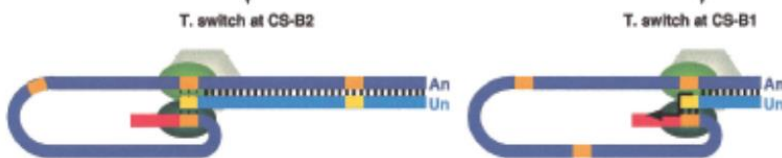
TRS есть перед геном ORF1ab и перед каждым поздним геном



B Base-pairing scanning



C Template switch



По -сгРНК полимераза RdRP делает +сгмРНК.

И рибосома транслирует первый ген на ней в белок

КОНЕЦ ПРЕЗЕНТАЦИИ