

# Биоинформатика. Белки и ДНК. Почему дети похожи на родителей?

С древних времён люди замечали, что у тигров рождаются тигрята, у птиц — птенцы, у людей — дети. Мало того, дети обычно внешне похожи на родителей — имеют тот же цвет глаз, цвет волос или форму носа. Новорождённый младенец часто не похож ни на мать, ни на отца, но со временем он приобретает черты внешнего сходства с матерью и отцом. Получается, что любой организм уже при рождении (а на самом деле ещё до рождения) «знает», какие у него во взрослом состоянии будут глаза, рост или голос. Значит, вся эта информация уже заложена с рождения, она где-то хранится, а по мере роста организм считывает эту информацию и приобретает те черты, которые ему предписаны наследственной программой.

Где же и как хранится наследственная информация? Как организм считывает её и понимает? Как он использует эту информацию по мере развития и роста? Этим вопросам много веков. Но ответы учёные стали находить только в последние 50 лет.

Всё в мире состоит из отдельных частей — и предметы, и живые организмы. В свою очередь, эти части сами состоят из частей и т. д. Некоторые более крупные части мы видим, а есть такие



Почему дети похожи на родителей?

маленькие, которые наш глаз не способен различить (они в миллионы раз тоньше человеческого волоса). Мы видим, например, что дом сложен из кирпичей, но мы не можем различить, из каких частичек состоит вода в стакане. И в воде, и в кирпичах, из которых построен дом, и в клетках тела человека можно выделить мельчайшие частицы, которые называются *молекулами*. Всё вокруг нас — и живое, и неживое — построено из молекул. В разных веществах молекулы различны. Они очень малы, их невозможно увидеть без специальных приборов. К счастью, такие приборы существуют — они помогают изучать молекулы.



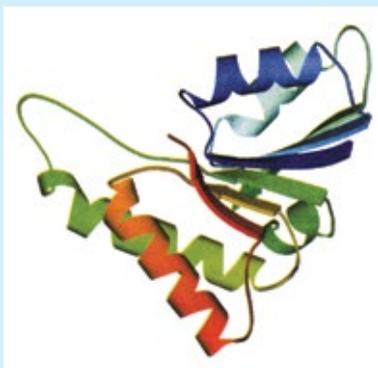
Среди молекул любого живого организма на нашей планете основную часть составляют молекулы особых веществ — **белков**.

Различных белков, т. е. различных видов молекул белков, очень много — несколько миллионов. Например, только в организме человека встречается около 30 000 различных белков. Но при этом молекулы всех белков устроены похожим образом.



Молекула любого белка — это **цепочка (последовательность)**, состоящая из сотен, а иногда и тысяч звеньев. При этом во всех известных молекулах белков **встречается только 20 видов звеньев!** (Можно сказать, что книга жизни написана в 20-буквенном алфавите.)

Звенья белковых цепей называют *аминокислотными остатками*. Каждый аминокислотный остаток имеет своё название и обозначение (одной буквой латинского алфавита), поэтому белки часто описывают словами (последовательностями)



Компьютерная модель молекулы белка. Радужная расцветка позволяет проследить ход звеньев цепочки. Молекула на рисунке увеличена примерно в миллиард раз

букв) в 20-буквенном алфавите (см. таблицу на обороте обложки учебника).

Именно набором белков один организм отличается от другого. Наборы белков у двух людей (у двух кошек, у двух берёз) очень похожи, только мелкие различия определяют, например, разный цвет глаз у разных людей или разную расцветку шерсти двух кошек. Наборы белков у организмов разных видов разные, но чем более родственны эти виды, тем более похожи наборы белков. Например, белки человека и шимпанзе совпадают на 99 % (т. е. различается только одно звено из 100). А у человека и мыши степень сходства около 80 % (различаются 20 звеньев из 100).

Где же хранится наследственная информация?



За хранение и передачу наследственной информации в живых организмах отвечают специальные молекулы — **молекулы ДНК**.

ДНК — это сокращение, полное название — *д*езоксирибонуклеиновая *к*ислота. Молекулы ДНК во всех клетках одного живого организма одинаковы. Но при этом молекулы ДНК разных организмов разные: у каждого человека свои молекулы ДНК, у каждой мышки свои.

Все молекулы ДНК, как и молекулы белков, — это цепочки, но звенья в молекулах ДНК отличаются от звеньев белков.



Звенья ДНК называются **нуклеотидами**. В молекулах ДНК встречается всего 4 вида нуклеотидов.

Молекулы ДНК в клетках живых организмов гораздо длиннее молекул белков. Даже самые короткие молекулы ДНК (ДНК вирусов) содержат сотни тысяч звеньев (нуклеотидов). А ДНК человека содержит около трёх миллиардов нуклеотидов. То есть молекула ДНК — это целая книга, написанная в 4-буквенном алфавите. Нуклеотиды обозначаются латинскими буквами А, С, G и Т.



Молекулы ДНК каждого живого организма **полностью определяют**, какие белки будут в этом организме.

Но как именно это происходит? Иными словами, как в ДНК закодированы (т. е. зашифрованы) белки? На этот вопрос мы ответим позже. А пока займёмся просто шифрованием.

# Шифрование

С древних времён люди использовали шифрование для секретной передачи и хранения информации. Шифрование выглядит как увлекательная игра, но преследует серьёзные цели. Шифры используются в военных целях, для передачи секретных сообщений, для хранения тайного знания и во многих других случаях.

Первые зашифрованные сообщения использовались ещё в Древнем Египте. Способ шифрования тогда был очень прост, сейчас он называется «шифрование простой подстановкой»: каждый иероглиф исходного сообщения заменялся в зашифрованном сообщении другим. При этом одинаковые иероглифы заменялись одинаковыми, а разные — разными.

Сегодня существует много способов шифрования и *шифров*. Мы будем пользоваться лишь одним способом, который построен по следующим правилам:

1. Каждая русская буква, а также пробел или знак препинания заменяется последовательностью латинских букв длины 3. Такая последовательность называется **кодом**. При этом используются только четыре латинские буквы — А, С, G, T.
2. Каждый код всегда заменяет (**кодирует**) одну и ту же букву или знак. При этом одна буква или знак необязательно всегда заменяется одним и тем же кодом.
3. При замене букв и знаков кодами порядок букв и знаков не меняется.



Замена каждой буквы её кодом называется **шифрованием**. Обратная замена каждого кода на соответствующую ему букву называется **расшифровкой**.

Для удобства шифрования и расшифровки строят *шифровальную таблицу*, в которой указан код (или все коды, если их несколько) для каждой буквы, для пробела и знаков препинания. Незаполненная часть такой таблицы помещена ниже.

Буква/знак	Код
А	
Б	
В	
Г	
Д	
Е	
Ё	

 **Полный шифр** — это заполненная шифровальная таблица, указывающая соответствие каждой буквы или знака и каждого кода.

Вот примеры шифровок:

Код буквы Я — САТ.

Шифровка слова ТЫ — ССССГГ.

Зашифруем слово ОНИ — получим шифровку АСТАГСТАА.

Закодируем слово КОМПЬЮТЕР — получим шифровку АГГАСТАСГАТСТТТСГАСССАГААГТ.

После раскодирования шифровки ГАААААСТСТААССГАГТ-АСТААГААААГСАГСССГАСС получаем слово ЗАШИФРОВАННЫЙ.

Обратите внимание, что знак переноса нам помогает удобно расположить на странице слова и их шифровки. При этом знак переноса при шифровании и расшифровке не учитывается.

Полный шифр вы построите, решая задачи.

**182**

Вырежи из вкладыша тетради проектов заготовку шифровальной таблицы и вложи в тетрадь (лучше прикрепить её скрепкой, чтобы потом не выпала). Используя коды из примеров на этой странице выше, заполни в своей таблице все строки, которые сможешь.



Проверь: в таблице должны быть коды для 19 букв.

183

Зашифруй те слова множества F, для шифрования которых в твоей таблице имеются все необходимые коды. Запиши шифровки в тетрадь.

ШАР	БАНАН	ШАРФ	КАРТОШКА
ОНА	ОН	ВЬЮНОК	ДЕДУШКА

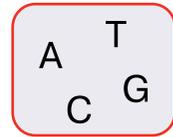
F

184



Сколько различных последовательностей длины 3 можно составить из букв множества M (конечно, буквы могут повторяться)?

Построй дерево перебора вариантов. Можно ли было использовать для шифрования букв русского алфавита не тройки, а пары, составленные из букв множества M? Поясни свой ответ.



M



При расшифровке удобно пользоваться обратной шифровой таблицей — *таблицей расшифровки*. В такой таблице в левой колонке выписаны все возможные коды (в словарном порядке), а в правой — буквы или знаки, которые этим кодам соответствуют. В заготовке таблицы расшифровки заполни все строки, которые сможешь.

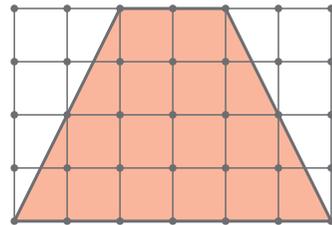
185

Раскодируй шифровки: перепиши их в тетрадь и напиши рядом с каждой шифровкой зашифрованное в ней слово.

A	G	G	A	C	T	G	A	A	A	A							
A	G	G	A	C	T	G	A	A	C	G	G						
A	T	C	A	G	A	A	G	T	A	A	G	A	A	C	A	T	
A	A	A	G	C	C	C	C	A	G	T	A	A	A	G	G	C	C

186

Нарисуй, как разрезать четырёхугольник на части, чтобы из этих частей можно было собрать прямоугольник на сетке. Нарисуй этот прямоугольник и покажи штриховыми линиями, из каких частей он составлен.



187

Множество В — множество шифровок всех слов из множества А. Запиши для каждого слова его шифровку и заполни пустые клетки шифровальной таблицы и таблицы расшифровки.



Проверь себя — в каждой таблице теперь должны быть коды для всех русских букв.

**А**

УХ ЭХ  
 ЧАС УЖ  
 БАЦ  
 БАС ЛУГА  
 ДУГА  
 ЛУЖА  
 ОБЩИХ  
 ОБЪЁМ

**В**

CAGCCT AACAAACAC  
 CCACCT AACAAAATG  
 CCACAA CGCAAAATG

ACACCAAATAAA  
 ATTCCACAAAA  
 ATTCCAAATAAA

ACTAACGCCTAACCT  
 ACTAACTCCATAACG

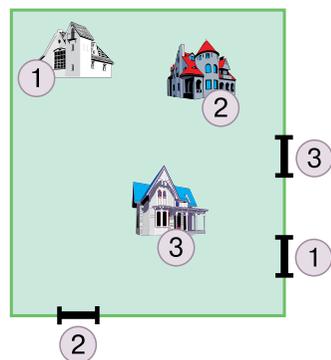
188

Дано зашифрованное предложение. Слова в этой шифровке разделены тройкой символов, кодирующей пробел. Расшифруй и запиши в тетрадь предложение. Дополни таблицы шифровки и расшифровки кодом пробела.

A C G A A A C G A A A C T A A C G C G G  
 A T T A A A C T A A G T A A A A C G C C A

189

На квадратном участке расположены три дома, а в ограде сделаны три калитки. Проложи дорожку от каждого дома к калитке с тем же номером так, чтобы дорожки не пересекались. Нарисуй схему участка и дорожек в тетради. Дома перерисовывать не надо — достаточно поставить номера.



120

190

При помощи таблицы расшифровки раскодируй следующие шифровки, запиши получившиеся слова.

```

АТСААААГТААААСААСТАГГАТГ
С С Г А Г А А А Г А Г Т А А А А Т Т С Т Т
А С Г С С А Г А А С Г Г А Г Г А А А
А Г С А Г А А С Т С С С Т С С А Г А А С Г -
А Т Т А Г А А С Г С Г Г А С С

```

191

При помощи шифровальной таблицы зашифруй слова: ПАРОМ, ВОЗДУХ. Теперь, не обращаясь к шифровальной таблице, зашифруй слова: ПАРОВОЗ, ДУХОМ.

192

Пользуясь шифровальной таблицей, зашифруй предложение: ЛЮБЛЮ ГРОЗУ В НАЧАЛЕ МАЯ.

193

Раскодируй зашифрованное предложение.

```

А Т С А Г Т Т А А С С Т А С Т А С А Т -
А А С Т А А Т Г А Г А А А Т А С Т А С -
А А Г С С А Т С Т А А Г С А А А С Т А -
А Г Г А А А С С С А С Т А Г Г С Т А А -
А Г С Т А А Т Г А Г А А С Г С Т Т

```

194

Найди выигрышную стратегию для игры *Двадцать пять*.

### **Правила игры *Двадцать пять***

Начальная позиция. Число 0.

Возможные ходы. На каждом ходу игрок прибавляет к имеющемуся числу 1, 2, 3 или 4.

Как определить победителя. Игра заканчивается, если позиция оказывается равной 25. Выигрывает тот, кто добавил последнее число.

121

# Биоинформатика.

## Как кодируются белки



Раскрашенная модель молекулы ДНК. Молекула ДНК очень длинная — здесь представлена только небольшая её часть. На модели видно, что молекула ДНК состоит из двух цепочек (одна раскрашена жёлтым, а другая — красным). В этих цепочках нуклеотиды расположены друг против друга и связаны особыми химическими связями (они показаны голубым)

Ты уже знаешь, что в состав любого живого организма входят молекулы белков. Именно набор белков определяет, например, почему у одного человека глаза карие, а у другого — голубые. Каждая молекула белка — это цепочка (последовательность). Звенья этой цепочки называются *аминокислотными остатками* или просто *остатками*. Все белки каждого живого существа закодированы в особой молекуле — молекуле ДНК. Молекула ДНК — это тоже цепочка, состоящая из звеньев другого вида — *нуклеотидов*. Таким образом, в одной последовательности (молекуле ДНК) закодированы другие последовательности (молекулы белков). А как именно, каким шифром в ДНК закодированы белки?

Оказывается, это происходит примерно так же, как в задачах на шифрование, которые ты решал.

Аминокислотные остатки (звенья молекул белка) могут быть только двадцати видов. Это значит, что молекула белка похожа на длинное слово, написанное в 20-буквенном алфавите. Каждый из двадцати возможных остатков имеет своё название и обозначается одной латинской буквой (см. таблицу на обороте обложки учебника).

Все молекулы ДНК построены только из четырёх видов нуклеотидов. Вот их русские и английские названия и буквы, которыми они обозначаются:

<b>аденин</b> ( <i>Adenine</i> , <b>A</b> ),	<b>гуанин</b> ( <i>Guanine</i> , <b>G</b> ),
<b>цитозин</b> ( <i>Cytosine</i> , <b>C</b> ),	<b>тимин</b> ( <i>Thymine</i> , <b>T</b> ).

Молекулу ДНК можно сравнить с очень длинным словом, написанным в 4-буквенном алфавите.

Не вся молекула ДНК кодирует белки, а только некоторые её участки, которые называются **генами**. Молекула ДНК простейших организмов (вирусов, бактерий) почти вся состоит из генов, а в молекуле ДНК человека гены составляют только около 3 % всей длины. Зачем нужны остальные 97 % ДНК человека, науке пока известно не очень хорошо.

В гене (как и в наших шифровках) каждый остаток белка кодируется тройкой нуклеотидов. Такие тройки биологи называют *кодонами*. Например, тройка ААG (аденин — аденин — гуанин) кодирует остаток лизин. При этом один и тот же остаток может кодироваться разными тройками нуклеотидов. Специальные тройки кодируют начало и конец гена. В таблице на обороте обложки учебника вы найдёте полный список остатков и всех шифрующих их кодонов. Эта таблица называется *таблицей генетического кода*.

О том, что белки могут кодироваться тройками нуклеотидов, догадался в середине XX века выдающийся физик и биолог Георгий Гамов. Позже с помощью экспериментов учёные подтвердили эту догадку и определили, какой остаток кодируется каждым кодоном. Замечательно, что *генетический код един для всех известных живых организмов!*

195

В этой шифровке закодировано слово, но по ошибке то ли одну букву пропустили, то ли вставили лишнюю. Не расшифровывая слово, определи, какую именно ошибку допустили при шифровании: пропустили букву или вставили лишнюю. Объясни свой ответ.

А А А А А С А С А G T A G A A A G T -  
А А А А А С С С С С А A G T A A A

123

196

Одно из слов множества Д зашифровали, но при этом допустили ошибки — одну букву в шифровке пропустили и вставили одну лишнюю (необязательно в то же место). Определи, какое слово пытались зашифровать, запиши в тетрадь это слово и его правильную шифровку.

С С А А Т-  
А А Г Г

ТОК      ТАК      ШОК  
ШИК      ТИК

Д

197

Разведчик зашифровал два слова — важный пароль и ответ на него. В начале и в конце, а также между словами разведчик вставил лишние тройки латинских букв, которые не входят в послание, — они служат для того, чтобы замаскировать эти важные слова в тексте (на случай перехвата послания противником). При этом начало и конец каждой части послания разведчик закодировал специальными тройками латинских букв, чтобы в штабе послание всё же смогли расшифровать. Этими тройками разведчик всегда кодирует начало и конец предложения. Расшифруй и запиши в тетрадь пароль и ответ на него. Дополни таблицы шифровки и расшифровки тройками, кодирующими начало и конец предложения.

Т Т Г Г С А С Г Т А Т С А Г А С С Т-  
А С Т С С С А А А С Т Г Т С Т Г А Т-  
С Г Т А А А А А Г Т А А А А С А С-  
Т А А С А Т С Т Г Г Т Т Г Г А

198

Реши задачу.

Известно, что К. М. Петров, В. Д. Петров, П. Б. Петров, Н. В. Петров, М. С. Петров, И. В. Петров, А. К. Петров, Д. М. Петров, Р. Б. Петров, Г. Д. Петров, Б. К. Петров — представители одного рода, причём один из них — основатель рода, остальные — его сыновья, внуки и правнуки. Других сыновей, внуков и правнуков у основателя рода не было. Построй дерево родства Петровых, если известно, что у каждого отца было по два сына, внуков у основателя рода — четыре, а у его сыновей — по два.

199

Дано три шифровки одного слова с ошибками. В одной пропустили одну букву, в другой вставили лишнюю, в третьей произошло и то и другое. Запиши слово и его правильную шифровку.

```

A C A G T C C C A A C A
A T C A G G T C C A A C A
A T C A G T C A A C A

```

200

Наиболее важные послания разведчик шифрует более сложным шифром, чем обычно. Он использует те же коды для тех же букв, что и раньше, но использует также и новые, дополнительные коды (тоже тройки букв из того же 4-буквенного алфавита). Таким образом, одну и ту же русскую букву он может шифровать разными кодами (при этом один и тот же код, как обычно, всегда кодирует одну и ту же букву). Множество W состоит из разных шифровок двух слов: СМЕШНО и ГЛАДИТ. Выясни, какая шифровка к какому слову относится, дополни новыми кодами букв свою шифровальную таблицу (в ней станет 52 кода) и таблицу расшифровки.

```

GGAGTTGAGGGCGTGTGT
TCTGCAGGTTTCAGTATTA
AATATTGACGGCGTGCCC
TGCTTTGCTTCAAGCTTC
GGAGTTGATGGCTACTGT

```

W

201

Молекула ДНК человека состоит из трёх миллиардов нуклеотидов. При этом гены составляют только 3 % всей длины последовательности. Сколько нуклеотидов содержится во всех генах человека?

202

На поле *Робота* стен нет. *Робот* находится в левом верхнем углу прямоугольника из закрашенных клеток. Составь алгоритм, переводящий *Робота* в правый нижний угол прямоугольника.

125

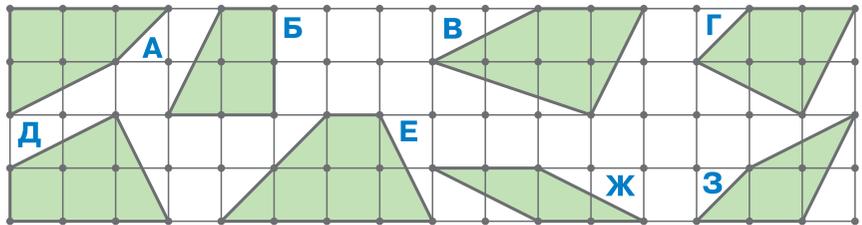
**203**

Разведчик зашифровал два предложения и вставил лишние тройки латинских букв, которые не входят в послание, в начале, в конце и между предложениями. Расшифруй послание.

G C G T C T C G T A T G C C C C A -  
 T A A T T A A A A G A A A A G C T A -  
 A A G A C T A A A G A C T A C C A T -  
 G A G G C T A A G C A A A C T A G A -  
 A A A A A T C A A A A C A C T A A C T -  
 A A C A C G A A A A G C C T G T A T -  
 A G T G G G C G T A A C A G A A G T -  
 A G A A A T T A A C C C A G A C T A -  
 C G A C A A A G C C G G A G A C T A -  
 A A T A G T A A A A G C T A A C A C -  
 C G G C T G T T T G T A T T C G G T

**204**

Найди два четырёхугольника одинаковой площади.

**205**

Участок ДНК

ATG CCA GCC ACA GAC ACA  
 AAC AGC ACC CAC ACC ACG CCG  
 ATG CAC CCA GAC GCC CAA CAC

кодирует такую последовательность остатков:

M P A T D T N S T H T T P M H P D A Q H

Пользуясь таблицами на обороте обложки учебника, найди такой остаток, который в этом примере кодируется тремя разными кодами. Выпиши его название и его кодоны.

206

Дано множество R слов и множество Q шифровок этих слов. Напиши рядом с каждым словом его шифровку. Дополни новыми кодами букв шифровальную таблицу и таблицу расшифровки (должен быть 61 код).

R

ПАР	
	ВОР
ЗРЯ	
	УХО
РОК	
	ПОЛ
КОК	

Q

GGGTTATAT
TATACTTGG
GCGTATTCG
TGATTCGTT
TTGGAGTAT
GTCTTATGG
TAGCCTTTC

207

Реши задачу.

Каждый из четырёх гномов — Ваня, Даня, Женя и Саня — либо всегда говорит правду, либо всегда врёт. Мы услышали такие отрывки разговора: Ваня — Дане: «Ты врун». Даня — Жене: «Ты врун!» Женя — Сане: «Оба они вруны». Потом Женя подумал и добавил: «Да и ты тоже врун». Кто из этих четырёх гномов всегда говорит правду?



208

Реши задачу.

При приготовлении пиццы выпекается большой хлебный корж и посыпается тёртым сыром. К сыру добавляются разные продукты, обеспечивающие тот или иной вкус. В распоряжении Марчелло имеются сладкий перец, репчатый лук, маринованные грибы, свежие помидоры, маринованная морковь и анчоусы (мелкая рыбка — хамса специального посола). По мнению Марчелло, имея один или несколько из этих продуктов (в любых сочетаниях), а также корж и сыр, можно приготовить пиццу. Сколько типов пиццы можно приготовить?

127