

## Задача (Б.Н. Гарифуллин)



В последнее время особую популярность приобрели добровольные вычисления (volunteer computing)—распределённые расчеты с использованием ресурсов, добровольно предоставленных частными пользователями сети Интернет. Подобные проекты были запущены и в биохимической сфере: например, [Folding@home](#) (расчет третичной структуры белков) или [DNA@Home](#) (поиск последовательностей в молекулах ДНК, соответствующих различным генам). В этом ряду особое место занимает свободно распространяемая мобильная игра Playto Cure: Genes in Space, разработанная в интересах CancerResearchUK и использующая огромный коллектив игроков для анализа генетических данных, полученных в реальных клинических исследованиях рака молочной железы. В рамках игры, пользователь собирая вымышленную субстанцию Элемент Альфа и двигаясь по межгалактическому пути, полному астероидов, фактически производит анализ паттерна генетического материала. Или более правильно: игрокам в Playto Cure: Genes in Space необходимо уточнить индивидуальные вариации числа копий некоторых генов в хромосомном наборе.

1. У Вас есть уникальный шанс в рамках олимпиады реально помочь исследователям из CancerResearchUK. Для это Вам необходимо достичь в игре минимум 25-го уровня, а в качестве ответа предоставить скриншот соответствующего раздела.

Одним из изучаемых онкогенов является ген трансмембранного белка-рецептора EGFR (ответствен за развитие порядка 20 видов злокачественных новообразований). Гликопротеин EGFR в виде мономера имеет молекулярную массу порядка 135 кДа, 30% которой приходится на углеводный компонент.

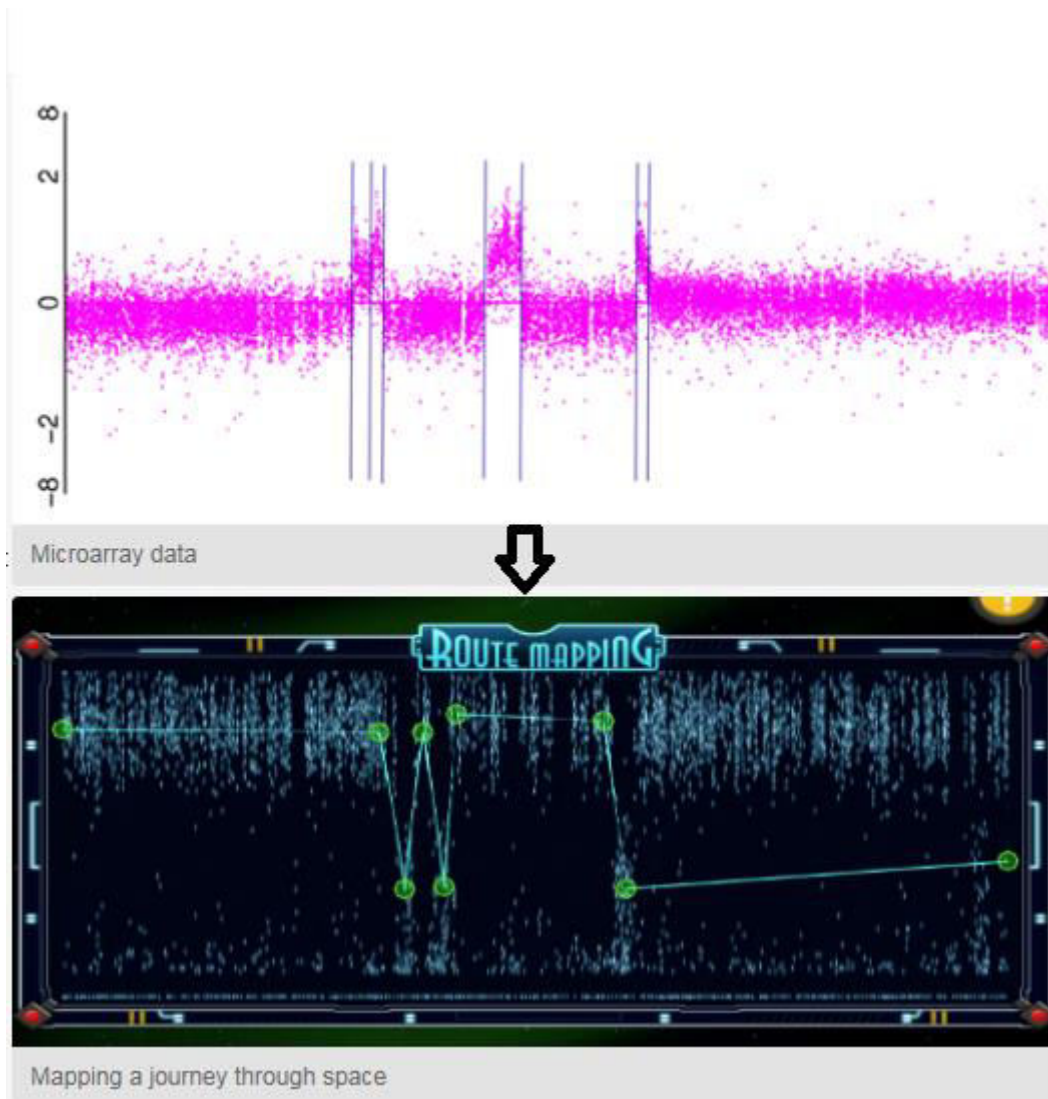
2. Оцените суммарную массу (в кДа) 4 копий гена EGFR в хромосоме человека, указав все сделанные допущения.

3. Оцените линейный размер (в нм) гликопротеина EGFR, исходя из длин отдельных аминокислотных остатков.

Предположим, что ген EGFR встроен в геном археобактерии, помещенную в питательную среду, единственным источником углерода в которой выступает меченный изотопом углерода  $^{14}\text{C}$  метиламин.

4. Оцените массу (в кДа) гена EGFR в ДНК, полученной в ходе однократной репликации исходного генетического материала археобактерии.

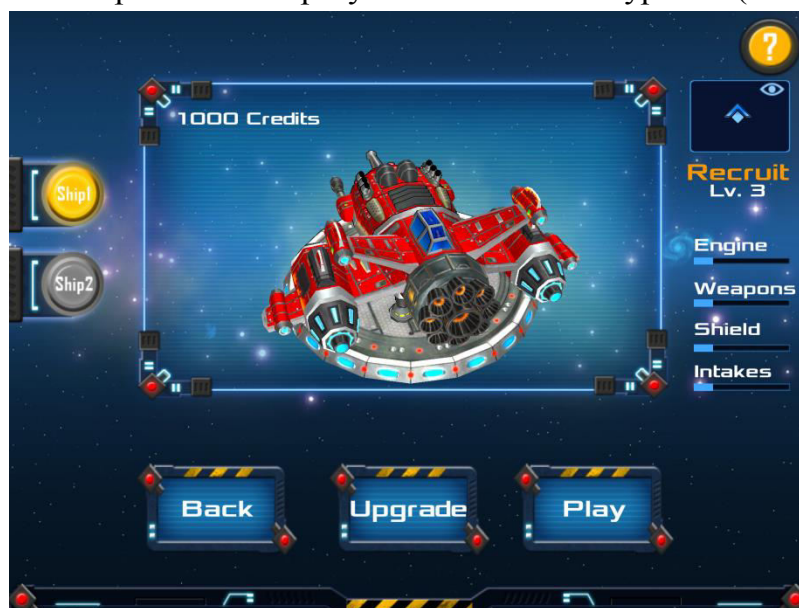
В верхней половине рисунка, представленного ниже, окрашенная розовым линия – это хромосома определенной длины, а пики на графике представляют собой добавочные копии определенно сегмента ДНК (данные получены по перспективной нанотехнологии ДНК-микрочип). Нижняя половина рисунка отображает трансформацию программных научных данных в игровой процесс. При этом для установления точного начала и конца добавочного фрагмента ДНК и требуется человеческий глаз.



5. Изобразите в рамках рассмотренной концепции 7-ю хромосому, несущую 4 копии EGFR.

## Решение (10 баллов)

1. Максимальный балл получает участник, приславший скриншот страницы игры, где виден его персональный результат не ниже 25 уровня (2 балла):



2. На аминокислотную часть EGFR приходится молярной массы  $135 \cdot 0,7 = 95$  кДа. Исходя из того, что для крупного белка распределение канонических аминокислот близко к эквимольному, можно предположить, что средняя молярная масса аминокислотного остатка составляет 110 г/моль. Тогда:

$$n(\text{ам. ост.}) = 95000 \text{ Да} / 110 \text{ Да} = 860$$

Структуру каждой аминокислоты кодирует кодон, состоящий из трех нуклеотидов, что с учетом наличия двух цепей ДНК приводит к значению числа нуклеотидов:

$$n(\text{нукл.}) = 860 \cdot 3 \cdot 2 = 5160$$

Средняя молярная масса остатка нуклеотида, содержащего каноническое азотистое основание, близка к 310 Да, отсюда:

$$M(4 \text{ код. ген}) = 5160 \cdot 310 \cdot 4 \approx 6400 \text{ кДа}$$

Хорошо известно, что кодирующей является не более 2% генома человека, и если применить это значение к гену EGFR получаем  $6400 \cdot 50 \approx 320000$  кДа

Полученное значение крайне приблизительное, особенно из-за неизвестного соотношения кодирующей и не кодирующей части гена EGFR.

(4 балла)

3. Гликопротеин EGFR, как следует из условия, является трансмембранным белком, выполняющим рецепторную функцию. Тем самым, он содержит, как высокорегулярную вытянутую часть, проходящую через мембрану, так и глобулярные, обращенные внутрь и снаружи от клетки фрагменты. Оценить длину белка, исходя из длин отдельных аминокислотных остатков, в данном случае невозможно, так как третичная структура глобулярной части трудно просчитываема (смотри проект [Folding@home](http://Folding@home), упомянутый в условии). Размер белка можно оценочно получить, например, путем сопоставления с толщиной мембраны эукариотической клетки.

(1,5 балла)

4. Производить расчет в данном случае не имеет никакого смысла, так как уровень радиоактивности, создаваемый единственным источником углерода -

изотопом  $^{14}\text{C}$ , будет препятствовать сколь-нибудь возможной биохимической активности и устойчивости молекул ДНК.

(1,5 балла)

5. Так как копии генов должны следовать друг за другом, то в рамках рассматриваемого графика сформируется прямая с одним пиком, соотношение ширины плато которого к длине плеча самой хромосомы (справочная информация: содержит 158 миллионов пар оснований) составит оценочно:

$$n = \frac{860 * 3 * 50 * 4}{158 \cdot 10^6} * 100\% \approx 0,3\%$$

На графике, отображенном на листе формата А4, это будет узкий заостренный пик. Картинка из условия адаптирована [отсюда](#).

(1 балл)