



## Всероссийская олимпиада студентов «Я – профессионал»

задания заключительного этапа  
по направлению «Биоинженерия и биоинформатика»

Категория участия: «Магистратура/специалитет»

### Задание 1.

Важность антибиотиков в современной медицине трудно переоценить. Особенно остро строит проблема резистентности микроорганизмов к антибиотическим препаратам. В 2017 году ВОЗ опубликовала список несущих угрозу микроорганизмов, среди которых распространена множественная устойчивость к антибиотикам. Эту группу организмов прозвали грязной дюжиной («Dirty dozen»)

#### WHO priority pathogens list for R&D of new antibiotics

##### Priority 1: CRITICAL

- *Acinetobacter baumannii*, carbapenem-resistant
- *Pseudomonas aeruginosa*, carbapenem-resistant
- *Enterobacteriaceae*, carbapenem-resistant, ESBL-producing

##### Priority 2: HIGH

- *Enterococcus faecium*, vancomycin-resistant
- *Staphylococcus aureus*, methicillin-resistant, vancomycin-intermediate and resistant
- *Helicobacter pylori*, clarithromycin-resistant
- *Campylobacter* spp., fluoroquinolone-resistant
- *Salmonellae*, fluoroquinolone-resistant
- *Neisseria gonorrhoeae*, cephalosporin-resistant, fluoroquinolone-resistant

##### Priority 3: MEDIUM

- *Streptococcus pneumoniae*, penicillin-non-susceptible
- *Haemophilus influenzae*, ampicillin-resistant
- *Shigella* spp., fluoroquinolone-resistant

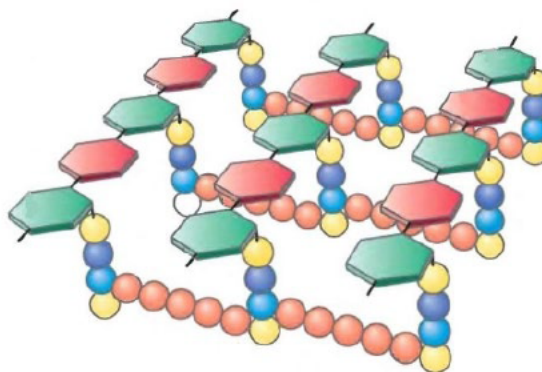
Для того, чтобы снять с конского волоса этот Дамоклов меч, необходимо не только понимать механизм действия различных антибиотиков, но и исследовать молекулярные и биохимические процессы, приводящие к возникновению резистентности.

По механизму действия антибиотики разделяют на 7 классов:

- Ингибиторы синтеза клеточной стенки
- «Ингибиторы» функции цитоплазматической мембраны
- Ингибиторы синтеза белков
- Ингибиторы репликации
- Ингибиторы транскрипции
- Ингибиторы синтеза миколовых кислот (специфичных для микобактерий, например, палочки Коха)
- Ингибиторы синтеза фолиевой кислоты

Два первых класса включают в себя львиную долю существующих антибиотических препаратов.

Первый класс антибиотиков, как следует из названия класса, препятствует синтезу клеточной стенки, что приводит к гибели клетки при её делении.



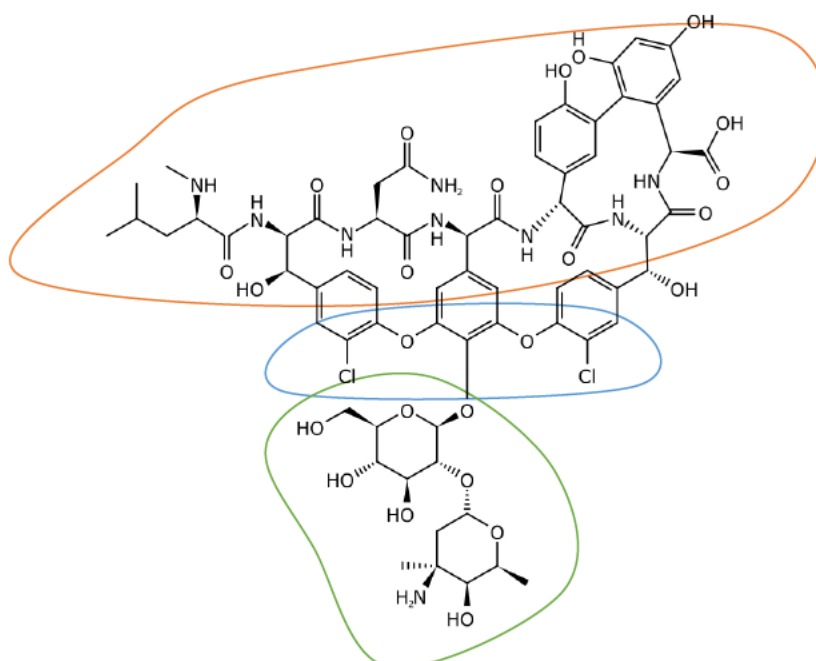
- 1) Что входит в состав клеточной стенки бактерий? (подпишите прямо на схеме) (3 балла).

В клинической практике наиболее важными антибиотиками данного класса являются  $\beta$ -лактамы и гликопептиды.

- 2) Изобразите общую структуру антибиотиков пенициллинового ряда. Каков механизм его действия (с указанием химических реакций)? Каким образом реализуется резистентность в отношении антибиотиков пенициллинового ряда? Напишите уравнение реакции с учетом реакционных групп белковых молекул. Есть ли альтернативные мишени у  $\beta$ -лактамных антибиотиков? Приведите пример. (5 баллов).

Строго говоря,  $\beta$ -лактамы — более широкое понятие, чем антибиотики пенициллинового ряда, они также включают в себя цефалоспорины, карбапенемы и монобактамы. Их механизм действия аналогичен, но за счет значительного отличия в структуре молекулы они демонстрируют высокую устойчивость к процессам, обеспечивающим резистентность к пенициллинам.

- 3) На рисунке изображен ванкомицин, принадлежащий ряду гликопептидов. Каков механизм его действия в отношении микроорганизмов? Какие функции у выделенных фрагментов молекулы? Как устойчивость формируется по отношению к гликопептидам? (4 балла)





Другой мишенью антибиотических препаратов выступают процессы биосинтеза белков: внутри этой группы выделяют антибиотики, действующие на 30S и на 50S субъединицу рибосомы.

К первой подгруппе относят тетрациклины и аминогликозиды. Тетрациклины связываются с 16S рРНК в области А-сайта, препятствуя проникновению аминоацил-тРНК, в то время как аминогликозиды (канамицин, гентамицин, неомицин), также взаимодействующие с 16S рРНК, приводят к неправильному прочтению кодона и/или нарушению транслокации рибосомы, обеспечивая обрыв синтезируемой полипептидной цепи.

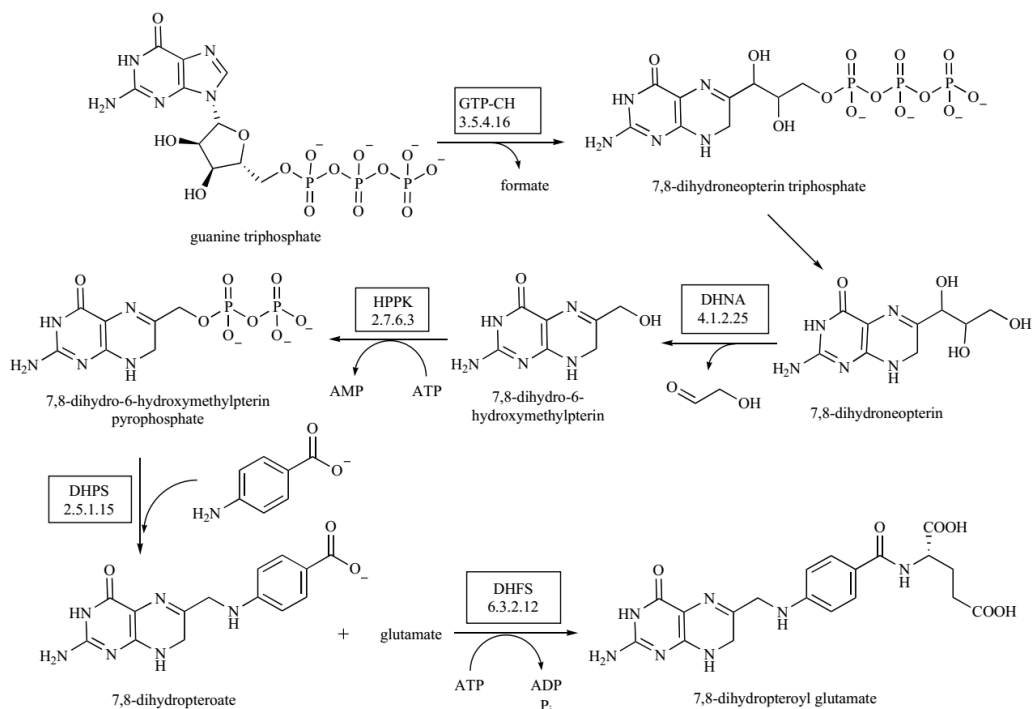
Хлорамфеникол в свою очередь взаимодействует с 50S субъединицей рибосомы, ингибируя пептидил-трансферазную реакцию

- 4) С помощью каких процессов бактерии формируют устойчивость к тетрациклинам, аминогликозидам и хлорамфениколу? (3 балла)

Антибиотики, мишенью которых выступает ДНК, включают две группы веществ: нитроимидазолы и хинолоны. Первые из них действуют исключительно прямолинейно: вступая в реакции с внутриклеточными веществами образуют активные формы азота (нитрозосоединения и нитроксильные радикалы), что приводит к многочисленным повреждениям ДНК, её разрушению. Вторые же, напротив, действуют только на реплицирующуюся ДНК, останавливая репликативную вилку.

- 5) Каким образом действуют хинолоны при ингибировании синтеза ДНК? Как реализуется устойчивость к хинолонам (4 балла)?

Путь биосинтеза фолиевой кислоты изображен на рисунке. Сульфаниламиды относятся к ингибиторам синтеза витамина В9. Антибиотическое действие обусловлено конкурентным ингибированием фермента DHPS (дигидроптероат синтазы).

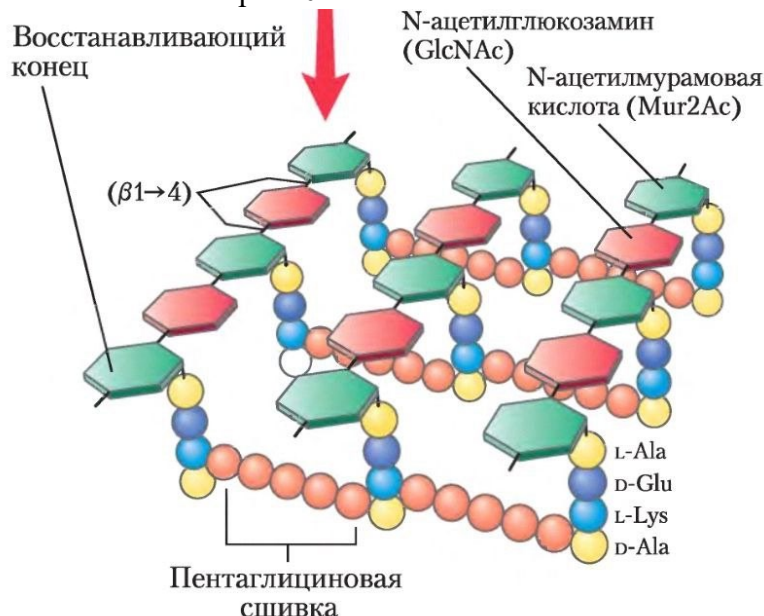


- 6) Напишите уравнение Михаэлиса-Ментен для конкурентного ингибирования (1 балл).

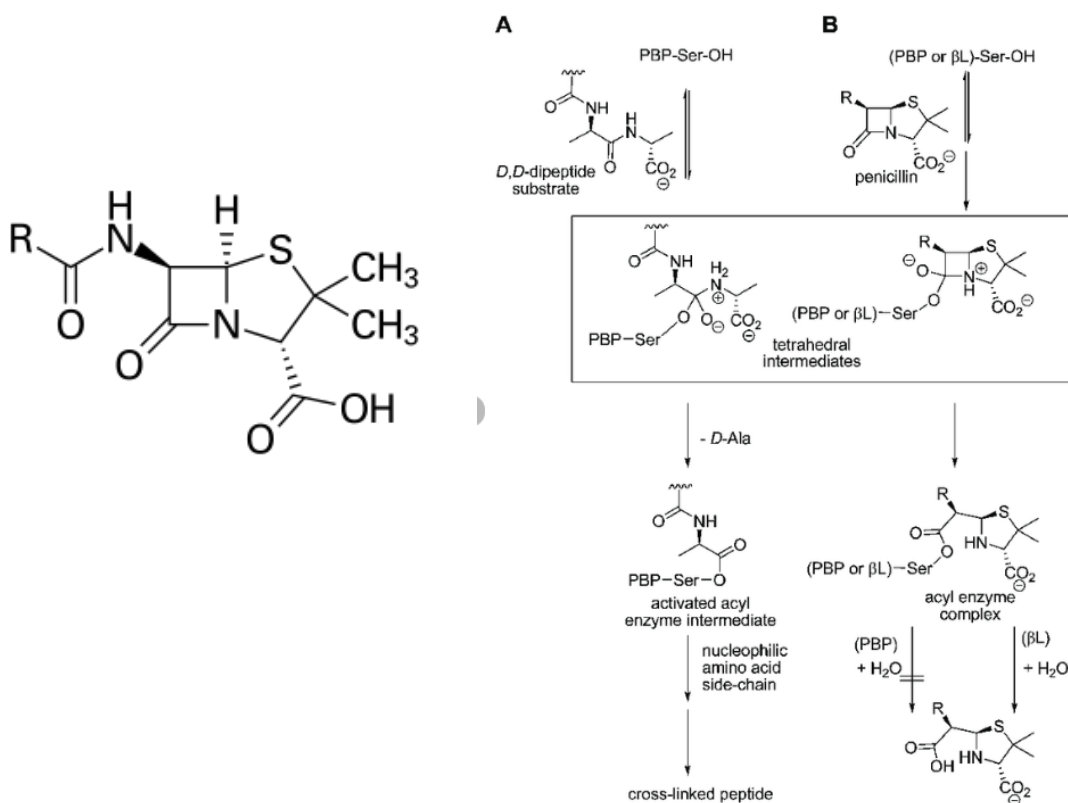


## Критерии оценивания задачи 1

- 1) За каждый компонент (N-ацетилглюкозамин, N-ацетилмурамовая кислота, L-Ala, D-Glu, L-Lys, D-Ala, пентаглициновый пептид) по 0,5 балла. Если указали верно, но перепутали местами углеводные остатки или первые 2 от N-мурамовой кислоты, по 0,25 балла. Если перепутаны местами D-Ala и L-Lys или указан неверный оптический изомер — 0 баллов за компонент.



- 2) Связывается с транспептидазой, катализирующей присоединение пентаглицинового мостика к соседней пептидной цепочке с отщеплением D-Ala. Образуется

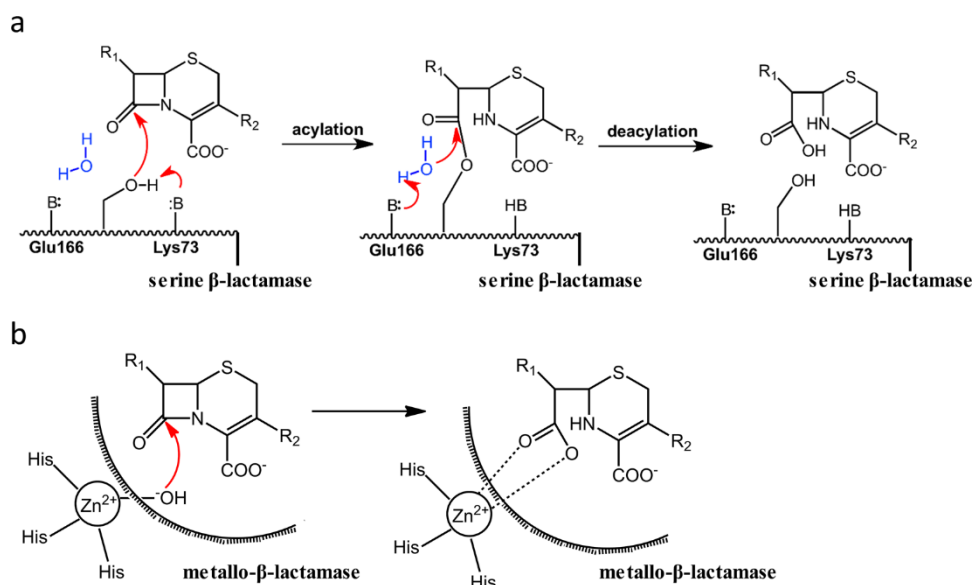




ацилфермент с большим периодом полураспада, что приводит к инактивации фермента.

В клетках функционирует фермент  $\beta$ -лактамаза, расщепляющая амидную связь в структуре  $\beta$ -лактамовых антибиотиков с образованием биологически неактивной пенициллоиновой кислоты.

Альтернативной мишенью выступает сама  $\beta$ -лактамаза, ингибиторы к которой также относятся к этому ряду антибиотиков. Например, клавулановая кислота — конкурентный ингибитор.



(Уравнения такие же только для пенициллина, а не цефалоспорины. И хватит только одной)

За структуру пенициллина, описание его работы, указание  $\beta$ -лактамазы как фермента, обуславливающего резистентность, по 0,5 балла. За уравнения реакции с пенициллином и упоминание ингибитора  $\beta$ -лактамазы по 1 баллу. За 2 механизма реакции с  $\beta$ -лактамазой 1,5 балла, за один — 1 балл.



- 3) Селективно связывает дипептид D-Ala-D-Ala, препятствуя транспептидазной реакции. Фрагменты: связывание дипептида, стабилизация гликозидного хвоста, углеводородные остатки, обуславливающие способность димеризоваться. Устойчивость обусловлена изменением состава дипептида, препятствующим связыванию. Действует только на грамположительные бактерии: другой путь резистентности — уплотнение мембраны для предупреждения проникновения в клетку.

За действие антибиотика и хотя бы один из двух механизмов устойчивости по 1 баллу.

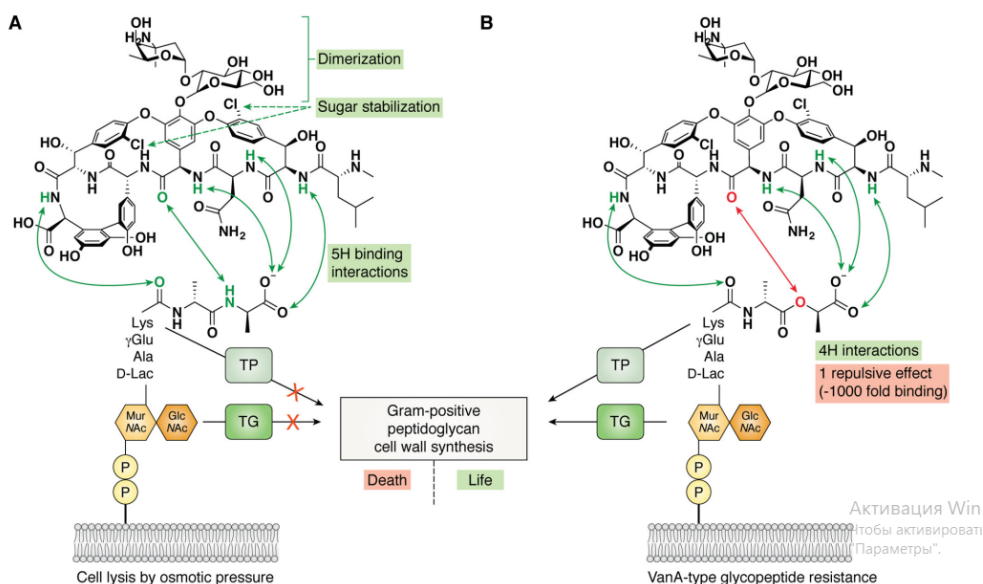
За описание функциональных групп 2 балла: если одна из трех опущена, но указан фрагмент связывающий дипептид, 1,5 баллов, если две не указаны, но указан фрагмент, связывающий дипептид 1 балл, если не указан пептид-связывающий участок — 0 баллов, независимо от наличия функций других фрагментов.

- 4) Для тетрациклина – эффлюксный насос («откачивающий» - я не знаю, как этот транспортер в русскоязычной литературе называют), защита рибосом мутациями в сайтах взаимодействия, окисление множеством тетрациклин деструктаз (кассета генов) до неактивных продуктов (NADH зависимое FAD-опосредованное), Для аминогликозидов основные — фосфотрансферазы, нуклеотидил трансферазы, но есть и ацилтрансферазы.  
Для хлорамфеникола основной путь – ацетилтрансфераза.

За резистентность для каждого антибиотика по 1 баллу. Необходимо указать хотя бы один. Если у аминоглюкозидов указан только ацелитрансферазный путь – 0,5 балла за этот антибиотик.

- 5) Хинолоны подавляют репликацию ДНК на уровне расплетания ДНК. Являются ингибиторами ДНК топоизомераз второго типа: ДНК гиразы и ДНК топоизомеразы IV. Встраивается в ДНК-белковый комплекс, когда ДНК и топоизомераза связаны ковалентно, и существует разрывы сахарофосфатного остова. Стабилизирует такое переходное состояние, препятствуя лигазной реакции ДНК гиразы.

**JBC REVIEWS:** Strategies for GPA modification

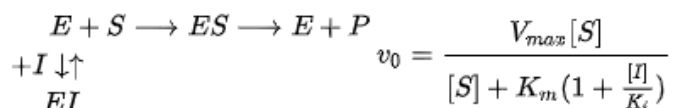




Ацилирование по аминокетогруппам, катализируемое ацетилтрансферазой, или синтез белков, защищающих ДНК и топоизомеразу от взаимодействия с хинолоном, эффлюксные насосы.

За указание ферментов ингибируемых — 1 балл, за препятствование лигазной реакции и встраивание в ДНК-белковый комплекс по 1 баллу. За хотя бы 1 механизм резистентности 1 балл.

- 6) За уравнение 1 балл. Если не написали уравнение, но хотя бы написали кинетическую схему уравнения 0,25 балла.







## Задание 2

После празднования китайского нового года биоинженер Вася видит сон...

И снится ему, что он исследует две чистых линии разноцветных драконов (1 и 2) с помощью чистой линии белых драконов (3).

Сначала он скрещивает разноцветных драконов линии №1 с линией белых драконов №3, драконы первого поколения имеют белую окраску. Во втором поколении Вася наблюдает расщепление 79% драконов белые, а 21% - разноцветные.

Тогда он скрещивает чистую линию разноцветных драконов №2 с линией белых драконов №3, драконы первого поколения имеют белую окраску, во втором поколении – 75 % драконов белые, в 25% - разноцветные.

Не поверив результатам, Василий повторил оба эксперимента увеличив выборку в 100 000 раз. Несмотря на то, что эксперименты во сне воспроизводятся обычно плохо, в данном случае расщепления все-таки воспроизвелись.

Теперь, прямо во сне, Василий ломает голову как согласовать два эксперимента. Ему и вам, дополнительно известно, что драконы – диплоидны, в основе их генетики лежит процесс мейоза. Разноцветная окраска у драконов определяется доминантным аллелем Р. При этом доминантный эпистатический аллель W блокирует транспорт пигмента в наружные покровы и подавляет развитие окраски.

- 1) Какими терминами можно описать подобный тип взаимодействия генов Р и W? (1 балл)
- 2) Изобразите схему первого скрещивания. Найдите генетическое расстояние между генами Р и W во сне. Какова доля рецессивных дигомозигот среди драконов F<sub>2</sub>? (9 баллов)
- 3) Чем отличаются разноцветные чистые линии 1 и 2? Какими причинами можно объяснить различие результатов в первом и втором скрещивании? (8 баллов)
- 4) Какова доля рецессивных дигомозигот среди белых драконов F<sub>2</sub> в втором скрещивании? (1 балл)

## Критерии оценивания задачи 2

1. Доминантный эпистаз - 1 балл
2. Схема скрещивания

Р: PP<sub>ww</sub> × ppWW  
разноцветные      белые  
G: Pw      pW  
F<sub>1</sub>: Pw/pW  
белые

2 балла

Наличие между этими генами генетического расстояния означает расположение этих генов в одной группе сцепления. Обозначим искомое расстояние за х, значит частота кроссоверных гамет будет тоже х.

F<sub>1</sub>: Pw/pW

G: Pw, pW (некроссоверы), pW, PW (кроссоверы)





Тогда для F2 получаем:

	PW, x/2	Pw, (1-x)/2	pW, (1-x)/2	pw, x/2
PW, x/2				
Pw, (1-x)/2		$(1-x)^2/4$		$x(1-x)/4$
pW, (1-x)/2				
Pw, x/2		$x(1-x)/4$		

За решетку Пенета для F2 -3 балла

Доля разноцветных:  $(1-x)^2/4 + x(1-x)/4 + x(1-x)/4 = 0,21$

$(1-x)/4 + x(1-x)/4 = 0,21$

$(1-x)(1+x) = 0,84$

$1-x^2 = 0,84$

$x^2 = 0,16$

$x = 0,4$  или 40 сантиморганов

За верное уравнение - 1 балл

За верный расчет расстояния – 2 балла

Доля рецессивных гомозигот среди всех драконов:  $x^2/4 = 0,04$  1 балл

3.

Схема скрещивания

P: PP<sub>ww</sub> × ppWW  
разноцветные      белые  
G: Pw      pW  
F1: Pw/pW  
белые

у гибридов F2 не происходит кроссинговер, гены P и W наследуются сцеплено.

2 балл

F2	Pw, 0,5	pW, 0,5
Pw, 0,5		
pW, 0,5		pW/pW 0,25

За схему или решетку – 1 балл

Отличия линий - в хромосомной перестройке – инверсии участка хромосомы, где расположены гены P и W 2 балла

В этом случае в мейозе образуются хромосомы с крупными делециями и дупликациями, имеющими летальный эффект, либо хромосомы не могут разойтись из-за дупликации или делеции центромера. 3 балла

Делеции и дупликации подобного размера маловероятны – 1 балл

4. Рецессивных дигомзигот среди белых драконов F2 во втором скрещивании-нет, доля составляет 0. 1 балл



### Задание 3

Одной из функций ДНК выступает реализация генетической информации, состоящая из двух этапов: транскрипция и трансляция. Сложность транскрипционного процесса с точки зрения числа вовлеченных в него белковых молекул отличается между прокариотическими и эукариотическими организмами. ДНК-зависимый синтез РНК у бактерий и архей выполняет единственный фермент – РНК полимераза или РНКП, в то же время в эукариотических клетках с этой целью функционируют от 3 до 5 различных РНК полимераз: Pol I, Pol II, Pol III и обнаруженные только у растений Pol IV и Pol V.

Pol I, Pol II и Pol III демонстрируют различия по ряду критериев: гены, структура промотора, число субъединиц апофермента, число общих транскрипционных факторов (GTFs – general transcriptional factors), а также механизмы инициации, элонгации и терминации. Ключевую роль в регуляции транскрипции играют сборка инициаторного комплекса и взаимодействие РНК полимераз с транскрипционными факторами.

- 1) Какие гены транскрибируются каждой из РНК полимераз человека? (1 балл)

Строгое соответствие пула генов определённой РНК-полимеразе достигается за счет специализированных промоторов. Структура промотора должна способствовать ассоциации инициаторного комплекса с одной полимеразой и препятствовать его сборке с участием другой. При этом структура промотора для Pol III изменяется в значительной степени между тремя типами генов.

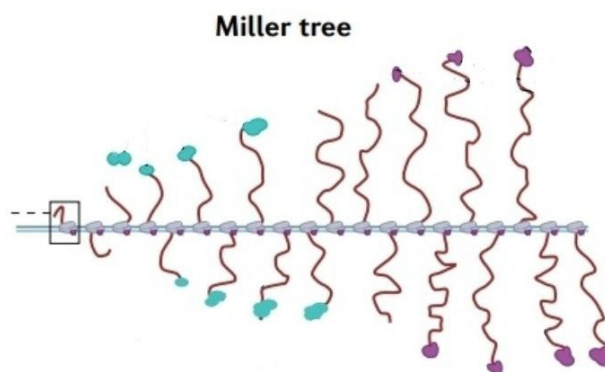
- 2) А) Перечислите основные элементы в составе промотора для Pol I, Pol II и промотора гена рРНК для Pol III (2 балла).  
Б) Приведите схематическое изображение промоторов: обозначьте коровый промотор и прилегающие области. Опишите особенности промотора полимеразы Pol I и Pol III (2 балла), (обязательно укажите точку начала транскрипции).

На этапе инициации транскрипции общие транскрипционные факторы и апофермент РНК-полимеразы образуют единый комплекс, необходимый для формирования транскрипционного «пузыря» и, как следствие, старта транскрипции.

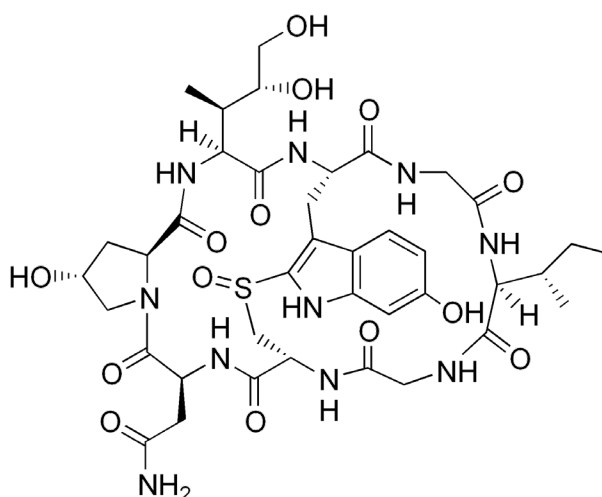
- 3) Сколько субъединиц образуют РНК-полимеразы трех типов? Перечислите факторы инициации, образующих инициаторный комплекс с каждой из полимераз (в случае полимеразы III укажите компоненты инициаторного комплекса только для гена рРНК) (5 баллов).  
4) Опишите постадийный механизм сборки инициаторного комплекса Pol II (5 баллов).

После начала транскрипции на расстоянии 20–120 нуклеотидов Pol II останавливается в результате взаимодействия с белками DSIF и NELF, первый из них выполняет сходную функцию и в случае Pol I. Фактор элонгации Р-TEFb необходим для реактивации РНК-полимеразы II, происходящей за счет фосфорилирования факторов DSIF и NELF, приводящего к их инактивации.

- 5) Какое значение имеет этот процесс при транскрипции у эукариот (1 балл)?



- 6) На рисунке изображено «дерево Миллера». Подпишите структуры и опишите процессы, которые лежат в его основе. (3 балл).
- 7) Какое вещество изображено на картинке? Каковы его свойства (1 балл)?



Для ответа на вопросы 1, 2А и 3 заполните таблицу

РНК полимеразы	Транскрибируемые гены	Элементы промотора	Число субъединиц	Факторы, участвующие при инициации
Pol I				
Pol II				
Pol III				

### Критерии оценивания задачи 3

РНК полимеразы	Транскрибируемые гены	Элементы промотора	Число субъединиц	Факторы, участвующие при инициации
----------------	-----------------------	--------------------	------------------	------------------------------------



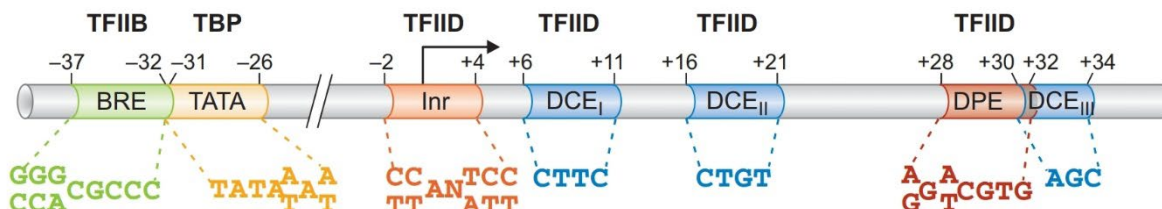
РНК polI	28S рРНК, 5,8S рРНК, 18S рРНК	Коровый промотор (-45 +20) UCE(upstreamcontrol element) или UBE	13	RRN3, SL1 (4 субъединицы), UBF
РНК polII	Все остальные либо мРНК, lncРНК, miРНК	Коровый промотор состоящий из BRE and TATA до TSS, Inr, и после TSS DPE и 3 DCE. UAS элементы, GC и CAAT мотивы перед TATA-боксом	12	TFIIA, TFIIB, TFIID, TFIIE, TFIIH, TFIIK, TFIIF, TFIIE, медиатор
РНК polIII	5S рРНК, тРНК, малые ядерные РНК	Промоторы первого типа состоят из двух участков А-бокса и С-бокса, расположенных после TSS. Разделены IE	17	TFIIA, TFIIB, TFIIC

- 1) Если распределены все группы генов по полимеразам верно — 0,5 баллов, в противном случае 0 баллов.
- 2) А) За каждый тип промотора по 0,5 балла, при этом если к промотору для PolII указано все, кроме UAS и CG-бокса – 0,5 балла.  
Б) Схема промотора для PolII.



Особенности: GC-богатый коровый промотор, промотор разделен на две части, за счет чего при белок-белковом взаимодействии молекул, расположенных на разных частях промотора, при образовании инициаторного комплекса образуется ДНК-петля.

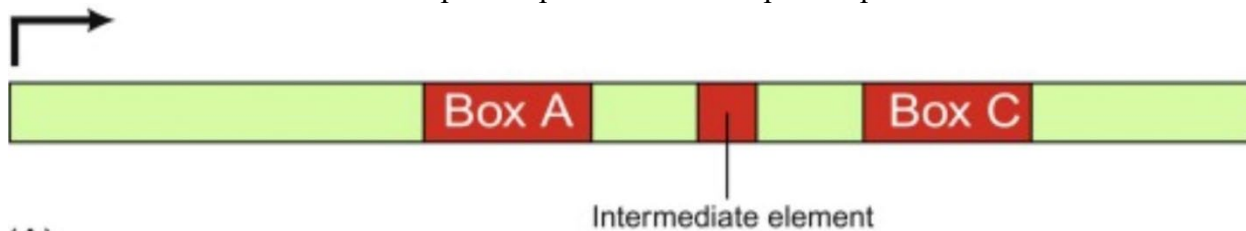
Структура корового промотора для PolII. Перед ним на -75 и -90 располагаются CAAT-мотив и GC-бокс соответственно.





Структура промотора для PolIII тип 1 (5S rRNA)

Особенность 2 из 3 типов промоторов этой полимеразы – расположение после TSS.



(A)

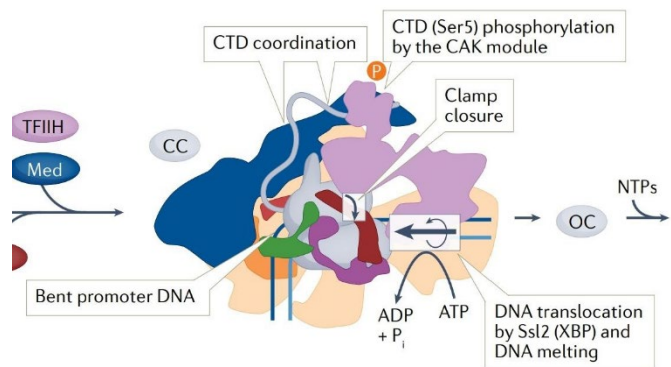
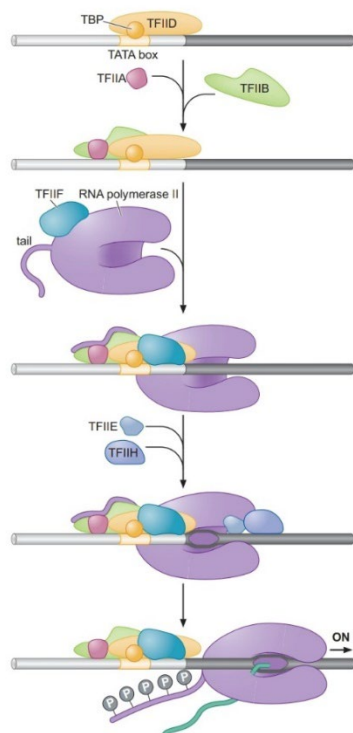
За правильную структуру каждого из трех промотора по 0,5 балла. За 2 особенности промоторов PolI и 1 PolIII, также 0,5 балла. Если не указано одно из трех особенностей 0,25, если 2 и более 0 баллов.

Указывать номер нуклеотидов не нужно, состав самих элементов и указание всех трех DCE для PolIII также избыточны.

3) За число субъединиц всех трех полимеразы 1 балл, за две из них 0,5 балла, за 1 0,25 балла.

За белки инициаторного комплекса по 2 балла за каждую полимеразу. Для PolI и PolIII за 2 из 3 белков 1 балл, за 1 и менее — 0 баллов.

Для PolII за неуказанный медиатор – минус 0,25 балла, за пропуск других факторов: 1 фактор – минус 0,5 баллов, за 2 неуказанных минус 1 балл. Указано меньше 4 – 0 баллов. Если указаны все 7 или 6 (все, кроме медиатора), за каждый лишний фактор – 0,25 баллов.



4) С участием TFIIA TFIIID узнает TATA-бокс и связывается с ним, за чем следует связывание BRE участков TFIIIB, привлечение PolII в комплексе с TFIIF, позиционирующем её на этом белковой платформе, акт ассоциации завершается посадкой TFIIE и TFIIH. Образование комплекса с медиатором приводит к фосфорилированию С-концевого домена полимеразы, что запускает транскрипцию.

Транскрипционный фактор	Функция
-------------------------	---------



TFIIA	Позиционирование TFIID на промоторе
TFIIB	Узнавание BRE, связь между TBP и PolII
TFIID	Узнавание TATA-бокса за счет TBP и других частей корового промотора за счет TAFs
TFIIE	Привлечение TFIIH и регуляция его функций
TFIIF	Образование комплекса с полимеразой перед сборкой, позиционирование полимеразы на промоторе, за счет стабилизации комплекса с другими факторами.
TFIIH	Осуществление перехода комплекса в открытое состояние АТФ-зависимо – медиация плавления промотора, перенос фосфатной группы на С-концевой домен PolII
Mediator	Взаимодействие между активаторами на энхансере и преинициаторным комплексом, сближение киназной субъединицы TFIIH и С-концевого хвоста полимеразы.

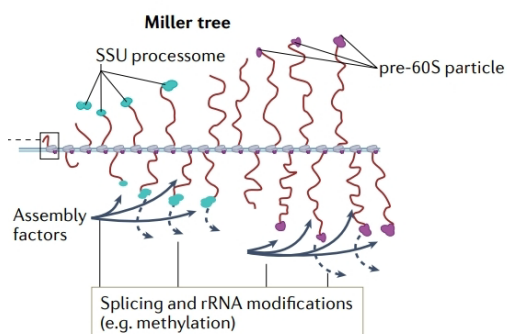
2 балла за механизм, 4 балла за функции.

Если не указана одна из стадий – минус 1 балл, но если пропущена последняя стадия с медиаторным комплексом – минус 0,5 балла, если две стадии слиты в одну – минус 0,5 баллов. Если пропущено более 1 стадии или более двух стадий объединены - 0 баллов. Изменение порядка приравнивается к пропускам стадий.

За функции медиатора, TFIIH по 0,75 балла (если указана одна из двух функций — 0,5 балла), за функции других факторов по 0,5 балла

- 5) Основной механизм контроля и синхронизации транскрипции. Осуществление запуска транскрипции в нужный момент с учетом сложности транскрипционной машинерии эукариот реактивацией гораздо быстрее, чем на этапе сборки инициаторного комплекса.

За утверждение о синхронизации и регуляции 1 балл.



- 6) Дерево Миллера — структура, похожая на структуру «Christmas tree» транскрипционной единицы прокариот. У эукариот формируется в ходе транскрипции рибосомных РНК PolII с параллельным процессингом рождающиеся пре-рРНК.

За указание процесса 1 балл и за указание 2 типов структуры по 1 баллу, то есть 2.

- 7) На рисунке представлен  $\alpha$ -аманитин – токсин бледной поганки, блокирующий PolII, но не блокирующий PolI, PolIII к нему слабо чувствительна. Действие происходит за счет взаимодействия токсина с мостиковой петлей, необходимой для транслокации полимеразы, в результате перемещение фермента





замедляется в несколько тысяч раз, что нарушает процесс транскрипции на этапе элонгации. За подобный ответ – 0,5 баллов. Пояснение механизма необязательно.

#### Задание 4

Метод полимеразной цепной реакции с обратной транскрипцией с детекцией результатов в режиме реального времени (ОТ-ПЦР-РВ) используют для количественного анализа экспрессии генов в клетках эукариот. Значение  $C_q$  (quantification cycle) соответствует циклу ОТ-ПЦР-РВ, на котором уровень флуоресценции превысил критическое значение, задаваемое настройками ПО амплификатора. Из данных, полученных в результате анализа калибровочной кривой, получена формула для перевода значений  $C_q$  в десятичный логарифм:  $C_q = 41.5 - 3,4 * \log_{10}(copy)$ , где  $copy$  – количество копий мРНК в образце. Данным методом был измерен уровень экспрессии пяти генов (“Ген\_1 – Ген\_5”) в двух группах образцов в трех технических повторности в каждый (“Группа\_1” и “Группа\_2”). В ячейках таблицы представлены значения  $C_q$  для данных генов.

	Группа_1_1	Группа_1_2	Группа_1_3	Группа_2_1	Группа_2_2	Группа_2_3
Ген_1	26,11	26,19	25,88	28,16	28,14	28,12
Ген_2	25,98	25,87	25,44	25,88	25,49	25,58
Ген_3	26,97	26,63	26,59	25,75	25,53	25,94
Ген_4	28,02	27,82	27,88	27,26	27,2	27,45
Ген_5	28,84	29,1	28,99	27,09	26,97	26,92

1) Используйте приведенную формулу для вычисления количества копий мРНК для первого гена (Ген\_1) и определите, во сколько раз различается экспрессия данного гена между группами по среднему значению ( $FoldChange = \frac{copy(Группа\ 2)}{copy(Группа\ 1)}$ ). В ответе приведите значение, округленное до второго знака после запятой (5 баллов).

2) Исходя из значений  $C_q$  для пяти приведенных генов в обеих группах, выявите из них самый стабильный ген (характеризующийся наименьшим среднеквадратичным отклонением), который можно использовать в качестве референсного. Вычислите для него среднее значение экспрессии в шести библиотек; ответ округлите до второго знака после запятой (2 балла). Рассчитайте среднеквадратичное отклонение экспрессии данного гена в шести повторностях по формуле  $s = \sqrt{\frac{\sum_{i=1}^n (copy_i - \overline{copy})^2}{N-1}}$ , где  $N(n)$  – общий размер выборки; ответ округлите до второго знака после запятой (3 балла).

Результаты ОТ-ПЦР-РВ подтвердили РНК-секвенированием. После первичной обработки данных, которая включала выравнивание на референсный геном и количественный анализ шести образцов из двух групп, получены значения экспрессии генов. В частности, для идентифицированного стабильного гена (Вопрос 2) известно, что его средняя экспрессия в шести образцах составляет 1200 копий, а выборочная дисперсия  $\sigma^2 = 13225$ .





3) Исходя из приведенных данных, вычислите доверительный интервал экспрессии данного гена по данным РНК-секвенирования, пользуясь правилом  $3\sigma$  (5 баллов). В ответе приведите интервал значений  $3\sigma$ , и округлите его до целых чисел.

4) Для гена со стабильной экспрессией (Вопрос 2) также известно значение экспрессии в образце 1 (780 копий), и для этого образца  $Cq_{O1} = 27,6$ . Рассчитайте число копий данного гена в образце 2, если известно, что его значение цикла выхода  $Cq_{O2} = 27,15$  (5 баллов). Исходите из допущения о том, что зависимость между результатами ОТ-ПЦР-РВ и результатами РНК-секвенирования линейная. Ответ округлите до целого значения.

#### Критерии оценивания задачи 4

1) Количество копий мРНК для первого гена (Ген\_1) вычисляется по приведенной формуле ( $Cq = 41.5 - 3,4 * \log_{10}(copy)$ ), и в первой группе составляет 33610; 31837; 39275; а во второй группе – 8385; 8500; 8615. Следовательно,  $FoldChange = \frac{(8385; 8500; 8615)}{(33610; 31837; 39275)} = 0,24$  (5 баллов).

2) Самый стабильный ген – Ген 2 ( $\frac{Ст. откл (Г2)}{Ср. знач (Г2)} = 15,55\%$ ). Среднее значение экспрессии данного гена в шести библиотеках составляет **44616,65 копий** (2 балла). Среднеквадратичное отклонение экспрессии данного гена в шести повторностях  $s = \sqrt{\frac{62620133 + 25751929 + 68762726 + 28531856 + 42650601 + 12294659}{5}} = 6937,03$  (3 балла).

3) Согласно данным,  $\sigma = 115$ , а доверительный интервал экспрессии данного гена по данным РНК-секвенирования –  $[1200 - 3*\sigma; 1200 + 3*\sigma] = [855, 1545]$  (5 баллов)

4) Согласно приведенным данным, составляем пропорцию и получаем, что количество копий по данным ОТ-ПЦР-РВ составляет 12252 и 16618 для  $Cq_{O1}$  и  $Cq_{O2}$ , соответственно, следовательно в образце O2 экспрессия выше на 35,63%  $\rightarrow 1,3563 * 780 = 1058$  копий (5 баллов).



## Задание 5

Обратная генетика – это подход к изучению функции гена, при котором для выявления связи гена с фенотипом используют модификацию или выключение работы этого гена. Один из самых популярных методов обратной генетики – нокаут генов при помощи системы CRISPR/Cas. Для того, чтобы оценить функциональный эффект нокаута гена на функционирование всего генома, используют метод высокопроизводительного секвенирования РНК (RNA-Seq). Допустим, вы участвуете в следующем эксперименте. В одной клеточной линии выполнили нокаут двух паралогичных генов, отобрали модифицированные клетки и получили две новые клеточные линии – KN1 (с нокаутированным геном 1) и KN2 (с нокаутированным геном 2). Из этих трех линий (Control, KN1 и KN2) выделили РНК и выполнили NGS секвенирование.

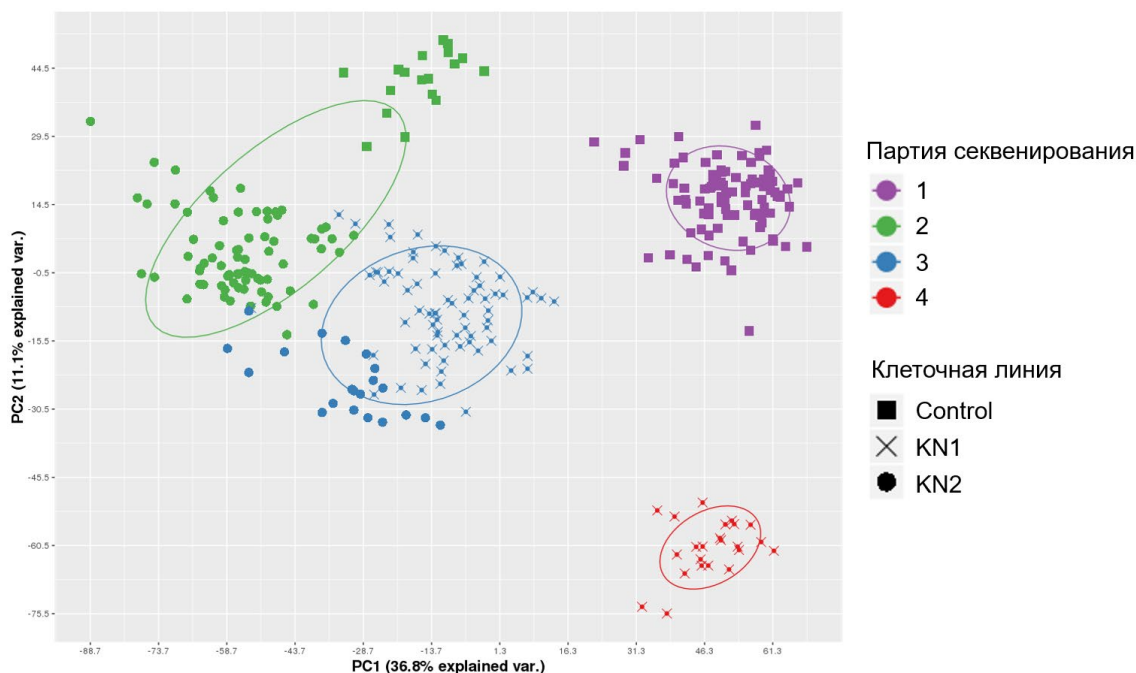
1. В таблице ниже приведено значение экспрессии некоторых генов (в TPM) в трех клеточных линиях (каждая была секвенирована в трех повторностях). Вычислите t-критерий Стьюдента для сравнения среднего значения экспрессии гена ENSG00000234481 в линиях Control и KN1. Используя таблицу из приложения, ответьте на вопрос, являются ли различия в экспрессии гена ENSG00000234481 между линиями Control и KN1 значимыми при уровне значимости 0.05. (4 балла)

Gene	Control-1	Control-2	Control-3	KN1-1	KN1-2	KN1-3	KN2-1	KN2-2	KN2-3
ENSG00000134644	8,8	8,6	4,4	12,1	3,9	5,0	5,2	11,6	5,8
ENSG00000134644	4,6	5,3	4,8	2,1	7,8	4,8	3,8	3,3	5,6
ENSG00000121753	5,5	6,0	4,2	3,2	6,9	4,0	6,4	3,3	4,1
ENSG00000197056	7,4	6,3	4,8	5,0	7,5	7,7	4,1	9,9	2,6
ENSG00000234481	6,3	8,5	12,6	15,2	11,3	16,7	6,0	7,6	7,9
ENSG00000117411	5,5	6,3	4,6	5,8	7,3	8,0	6,2	7,3	5,7
ENSG00000225154	4,9	5,3	6,0	2,9	9,3	6,2	2,9	7,2	4,6
ENSG00000203356	0,5	1,5	0,8	1,7	0,7	0,8	1,4	4,4	1,2
ENSG00000203995	1,2	10,0	2,6	6,1	5,8	5,4	8,7	43,3	20,8
ENSG00000177414	1,1	1,6	2,3	2,4	0,9	0,5	0,6	17,3	4,0
ENSG00000288804	0,3	0,7	0,5	1,0	0,8	0,2	1,1	6,2	1,5

2. В реальных экспериментах, при обработке данных RNA-Seq для выявления дифференциально экспрессирующихся генов, не используется t-тест. Вместо него, как правило, используется метод обобщенных линейных моделей (glm). Объясните, почему t-тест и U-тест Манна-Уитни плохо подходят для обработки таких данных (как в задании 1), и почему используется метод glm. Опишите алгоритм работы метода glm для эксперимента с одним фактором. (6 баллов)
3. Допустим, вы решили выполнить секвенирование РНК из этих клеточных линий не в трех, а в большем количестве повторностей. Выделение РНК, синтез кДНК, подготовка библиотек и секвенирование было выполнено в 4 партии. В первой партии секвенировали только клеточную линию контроля, во второй партии клеточные линии Control, KN2, в третьей партии клеточные линии KN1 и KN2, а в четвертой партии только клеточную линию KN1. После получения значений экспрессии генов, был проведен анализ главных компонент (PCA). График в первой и второй главной компоненте представлен на рисунке ниже. Изучите график и объясните, чем может быть вызвано такое расположение образцов на плоскости двух первых главных компонент. Как называется этот эффект? Какие есть способы



устранения этого эффекта в данных экспрессии генов (приведите не менее 2 способов)? (6 баллов)



4. При анализе данных RNA-Seq, полученных после геномного редактирования, может быть интересно не только анализировать экспрессию генов, но и искать мутации, полученные в результате обработки клеточной линии РНК-направленной эндонуклеазой Cas9. Предложите пайплайн для поиска всех мутаций, полученных в результате обработки эндонуклеазой Cas9, в данных секвенирования РНК клеточных линий KN1 и KN2 (от fastq файлов до файла с мутациями). Приведите конкретные инструменты, которые необходимо использовать на каждом этапе обработки данных. Какие ограничения есть при поиске мутаций в данных RNA-Seq? (4 балла)

## ПРИЛОЖЕНИЕ

Таблица критических значений для t-теста Стьюдента

$\alpha$	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.001
df										
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	318.3
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	22.33
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	10.21
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	7.173
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	5.893
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	5.208
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.785
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	4.501
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	4.297
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	4.144
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	4.025
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.930
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.852
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.787



### Критерии оценивания задачи 5

1. Правильно посчитан t-критерий (2.15) – 2 балла. Правильно определена значимость (результат не является значимым при  $p=0.05$ ) – 2 балла.
2. Объяснено, почему не подходит U-тест (большое количество сравнений, есть зависимые гены, малая выборка, недостаточно мощности критерия) – 1 балл. Объяснено, почему не подходит t-тест (большое количество сравнений, есть зависимые гены, малая выборка, не нормальное распределение) – 1 балл. Объяснено, почему используется glm (подходит для разных распределений, в том числе для экспоненциальных, не требует нормальности распределения) – 2 балла. Правильно описан алгоритм работы метода glm – 2 балла.
3. Правильно объяснена причина (секвенирование в разных партиях) – 1 балл. Правильно назван batch-эффект – 1 балл. Приведены способы коррекции batch-эффекта – максимум 4 балла (по 2 балла за каждый способ)
4. Приведен пайплайн, включающий 3 необходимых этапа – тримминг, картирование и коллинг соматическими коллерами – 1 балл. Приведены адекватные инструменты (например, fastp/trimmomatic/bbduk для тримминга, STAR/hisat2 для картирования, Mutect2 для коллинга соматических мутаций) – 2 балла. Правильный ответ на вопрос про ограничения (только транскрибируемые области генома с достаточно высокой экспрессией) – 1 балл.