

Министерство науки и высшего образования Российской Федерации  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт биологии, экологии, почвоведения, сельского и лесного хозяйства  
(Биологический институт)

УТВЕРЖДЕНО:  
Директор  
Д. С. Воробьев

Оценочные материалы по дисциплине  
Биоинформатика и компьютерная биология  
по направлению подготовки

**06.03.01 Биология**

Направленность (профиль) подготовки:  
**Биология**

Форма обучения  
**Очная**

Квалификация  
**Бакалавр**

Год приема  
**2024**

СОГЛАСОВАНО:  
Руководитель ОП  
Д.С. Воробьев

Председатель УМК  
А.Л. Борисенко

Томск – 2025

## **1. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами**

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-6 Способен использовать в профессиональной деятельности основные законы физики, химии, наук о Земле и биологии, применять методы математического анализа и моделирования, теоретических и экспериментальных исследований, приобретать новые математические и естественнонаучные знания, используя современные образовательные и информационные технологии.

ОПК-7 Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности.

ОПК-8 Способен использовать методы сбора, обработки, систематизации и представления полевой и лабораторной информации, применять навыки работы с современным оборудованием, анализировать полученные результаты.

ПК-1 Способен участвовать в исследовании биологических систем и их компонентов, планировать этапы научного исследования, проводить исследования по разработанным программам и методикам, оптимизировать методики под конкретные задачи.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-6.2 Применяет методы математического анализа и моделирования, теоретических и экспериментальных исследований

ИОПК-7.2 Применяет современные информационно-коммуникационные технологии для решения стандартных профессиональных задач

ИОПК-8.1 Формулирует принципы сбора, обработки, систематизации и представления полевой и лабораторной информации

ИПК-1.2 Проводит анализ и теоретическое обобщение научных данных, результатов экспериментов и наблюдений в соответствии с задачами исследования

## **2. Оценочные материалы текущего контроля и критерии оценивания**

Элементы текущего контроля:

- Тесты;
- Контрольная работа;
- Реферат;

Тест (ИОПК-7.2, ИОПК-8.1)

Основной целью проекта «Геном человека» являлось:

- а) Клонирование первого млекопитающего
- б) Определение полной последовательности нуклеотидов, составляющих ДНК человека
- в) Создание первого персонального компьютера для биологов
- г) Разработка алгоритма BLAST

К международным базам данных биологической информации НЕ относится:

- а) NCBI (National Center for Biotechnology Information)
- б) UniProt (Universal Protein Resource)
- в) Microsoft Excel
- г) Protein Data Bank (PDB)

Алгоритм BLAST предназначен для:

- а) Проведения статистического анализа данных RNA-seq
- б) Поиска гомологичных (сходных) последовательностей в базах данных
- в) Визуализации трехмерной структуры белков

г) Хранения образцов ДНК

Понятие «онтология генов» (Gene Ontology) подразумевает:

- а) Местоположение гена на конкретной хромосоме
- б) Процесс секвенирования генома
- в) Стандартизированную систему описания функций генов и их продуктов
- г) Правила поведения в лаборатории

Основная проблема, решаемая на этапе контроля качества (QC) данных высокопроизводительного секвенирования, — это:

- а) Определение эволюционного возраста образца
- б) Интерпретация биологического значения найденных различий
- в) Идентификация и фильтрация технических артефактов и последовательностей низкого качества
- г) Поиск публикаций по теме исследования в научных журналах

Формат файла FASTA используется для хранения:

- а) Трёхмерных структур белков
- б) Параметров микроскопа
- в) Нуклеотидных или аминокислотных последовательностей
- г) Результатов статистического тестирования

Ключи: 1 - б), 2 - в), 3 - б), 4 - в), 5 - в), 6 - в)

Критерии оценивания: тест считается пройденным, если обучающий ответил правильно как минимум на половину вопросов.

Контрольная работа (ИОПК-7.2, ИОПК-8.1, ИПК-1.2)

Контрольная работа состоит из 2 теоретических вопросов и 1 задачи.

1. Дайте определение биоинформатики как научной дисциплины. Назовите ее основные задачи и место в системе биологических наук.
2. Опишите основные этапы анализа данных полногеномного секвенирования (Whole Genome Sequencing). Кратно охарактеризуйте цель каждого этапа.
3. Что такое эволюционная консервативность последовательности? Объясните, как этот принцип используется для предсказания функциональной значимости участков генома.
4. Опишите принцип работы и сравните алгоритмы глобального (Needleman-Wunsch) и локального (Smith-Waterman) выравнивания последовательностей. В каких случаях применяется каждый из них?
5. Что такое Gene Ontology (GO)? Опишите три основные составляющие онтологии и принцип их использования для функциональной аннотации генов.
6. Каковы основные проблемы и сложности сборки геномов *de novo* из коротких прочтений?
7. Объясните, что такое *e-value* и *query coverage* в контексте анализа результатов BLAST. Как эти параметры помогают оценить достоверность найденного гомолога?
8. Что такое филогенетическое дерево? По какому принципу строится и как интерпретируются его элементы (узлы, ветви, длина ветвей)?

Задачи:

Задача 1. Дана последовательность в формате FASTA:

>contig\_5|organism=Mus musculus|length=150 bp

ATGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGC

Рассчитайте длину последовательности и процентное содержание нуклеотида «С». Какую информацию, помимо самой последовательности, несет эта запись?

Задача 2. BLAST-поиск белка дал лучший результат с  $e\text{-value} = 1e-45$  и  $\text{percent identity} = 78\%$ . Второй лучший результат имеет  $e\text{-value} = 0.003$  и  $\text{percent identity} = 40\%$ . Объясните значения этих параметров. Белок с какой последовательностью является более надежным гомологом и почему?

Задача 3. При анализе метаболического пути дрожжей обнаружен фермент, катализирующий реакцию: Вещество А → Вещество В. Гомолог этого фермента у человека катализирует реакцию: Вещество В → Вещество С. Можно ли на этом основании считать, что эти ферменты имеют одинаковую функцию? Ответ обоснуйте.

Задача 4. В аннотированной последовательности гена указано: `function: ATP-binding; GO:0005524`. Объясните, что означает эта запись. Какой метод, скорее всего, был использован для такой функциональной аннотации?

Задача 5. На приведенной упрощенной схеме филогенетического дерева (A,B),(C,D) расстояние от общего предка до таксона А составляет 0.1, до таксона В — 0.1, до таксона С — 0.4, до таксона D — 0.4. Какие таксоны (A/B или C/D) эволюционно ближе друг к другу? Какие таксоны больше отличаются от общего предка?

Задача 6. При анализе RNA-seq для гена X в контрольной группе было получено 1000 reads, в опытной группе — 5000 reads. Общее количество прочитанных ридов в обеих библиотеках примерно одинаково. Сделайте предварительное предположение об изменении экспрессии гена X. Какой основной статистический тест необходимо применить для подтверждения достоверности этого изменения?

Задача 7. На этапе контроля качества данных NGS получены следующие показатели для рида: `Read length: 150 bp, % of bases with quality ≥ Q30: 85%`. Дайте оценку техническому качеству этого рида. Достаточно ли оно для дальнейшего анализа?

Задача 8. Вам необходимо найти информацию о белке-актине человека (nucleotide sequence, protein sequence, known domains, 3D-structure если доступно). Перечислите основные базы данных, которые вы будете использовать для каждого типа информации, и последовательность ваших действий.

Ответы:

Задача 1. Длина последовательности: 45 п.о. (подсчет символов ATGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGC).

Содержание нуклеотида «С»:  $10/45 \times 100\% \approx 22.2\%$  (в последовательности 10 символов «С»). Дополнительная информация: Идентификатор контига (`contig_5`). Организм (*Mus musculus* — домовая мышь). Ожидаемая длина контига (150 bp) — фактическая длина отличается от заявленной.

Задача 2.  $E\text{-value} = 1e-45$  указывает на чрезвычайно низкую вероятность случайного совпадения (высокая значимость).  $\text{Percent identity} = 78\%$  показывает высокую степень схожести последовательностей. Более надежный гомолог — первый белок ( $e\text{-value} \approx 0$  и высокая идентичность), так как  $e\text{-value}$  на 42 порядка величины лучше, а идентичность почти в 2 раза выше.

Задача 3. Нет, ферменты не имеют одинаковой функции. Обоснование: они катализируют разные реакции (A→B vs B→C) и вероятно принадлежат к разным метаболическим путям. Гомология указывает на общее происхождение, но не гарантирует идентичность функции в связи с возможной неофункционализацией или субфункционализацией.

Задача 4. Запись означает: белок обладает АТФ-связывающей функцией (молекулярная функция), и это аннотировано стандартным термином Gene Ontology с ID

GO:0005524. Метод аннотации: предсказание на основе гомологии или предсказание АТР-связывающих мотивов.

Задача 5. Таксоны А и В ближе друг к другу (расстояние между ними =  $0.1 + 0.1 = 0.2$ ), чем С и D ( $0.4 + 0.4 = 0.8$ ). Таксоны С и D больше отличаются от общего предка (расстояние 0.4 каждый), чем А и В (0.1 каждый).

Задача 6. Предположение: экспрессия гена X увеличена в 5 раз (5000/1000) в опытной группе. Статистический тест: тест Фишера для count data или DESeq2/edgeR (для RNA-seq с поправкой на дисперсию и множественное тестирование).

Задача 7. Качество рида хорошее: Длина 150 bp типична для Illumina. 85% оснований с качеством  $\geq Q30$  (точность 99.9%) превышает порог 80%. Качество достаточно для большинства видов анализа (e.g., выравнивание, variant calling).

Задача 8. Nucleotide sequence: NCBI Nucleotide или ENSEMBL. Protein sequence: UniProt или NCBI Protein. Known domains: InterPro, Pfam, или SMART (через UniProt). 3D-structure: PDB (Protein Data Bank). Последовательность действий: Поиск в UniProt по запросу "human actin". Использование ссылок на соответствующие базы данных из карточки белка.

#### Критерии оценивания контрольной работы

Результаты контрольной работы определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Оценка «отлично» выставляется, если:

- Даны полные, развернутые и правильные ответы на 2 теоретических вопроса.
- Верно решена задача, приведены необходимые расчеты и аргументированные пояснения.
- Продемонстрировано глубокое понимание принципов работы методов и умение их применять для анализа.

Оценка «хорошо» выставляется, если:

- Даны в основном правильные ответы на 2 теоретических вопроса, возможны незначительные неточности или неполнота.
- Верно решена задача, в решениях допущены незначительные ошибки в расчетах или формулировках, но общий принцип решения верен.

Оценка «удовлетворительно» выставляется, если:

- Теоретические вопросы раскрыты поверхностно, с существенными неточностями, дан правильный ответ только на 1 вопрос.
- В решении задач допущены существенные ошибки, но продемонстрировано частичное понимание подхода.

Оценка «неудовлетворительно» выставляется, если:

- Ответы на теоретические вопросы неверны или отсутствуют.
- Задача не решена или решение фундаментально неверно.
- Продемонстрировано полное непонимание принципов работы методов.

#### Реферат (ИОПК-6.2, ИПК-1.2)

Формулировка задания: в рамках самостоятельной работы подготовить реферат на одну из предложенных тем. Объем: 10–15 страниц. Реферат должен содержать титульный лист, содержание, введение, основную часть (разделенную на главы), заключение и список литературы (не менее 10–15 источников, включая научные статьи, монографии и ресурсы международных баз данных).

Темы для реферата:

1. Математические основы и алгоритмы выравнивания биологических последовательностей (на примере алгоритмов Needleman-Wunsch и Smith-Waterman).

2. Применение методов математической статистики и теории вероятностей в анализе данных высокопроизводительного секвенирования (RNA-seq, DNA-seq).
3. Биоинформатические подходы к филогенетическому анализу: от построения множественных выравниваний до реконструкции эволюционных деревьев.
4. Сравнительная характеристика методов и инструментов для предсказания пространственной структуры белка.
5. Системная биология: математическое моделирование метаболических путей и генетических сетей.
6. Роль машинного обучения и искусственного интеллекта в решении задач современной геномики и протеомики.
7. Анализ полногеномных ассоциативных исследований (GWAS) как инструмент поиска генетических основ мультифакториальных заболеваний.
8. Методы сборки (ассемблирования) геномов *de novo* и проблемы, возникающие при их применении.
9. Биоинформатические стратегии аннотирования вновь секвенированных геномов.
10. Эпигенетика и биоинформатика: анализ данных ChIP-seq, bisulfite sequencing для изучения механизмов регуляции экспрессии генов.

Критерии оценивания:

Результаты реферативной работы определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

«Отлично» (85–100%):

- Актуальность и глубина раскрытия темы: тема раскрыта глубоко, полно и всесторонне. Продемонстрировано свободное владение сложным понятийным аппаратом и понимание методологических основ.
- Структура и логика: работа имеет четкую логическую структуру. Все разделы взаимосвязаны, изложение последовательное и аргументированное.
- Анализ и обобщение: продемонстрированы ярко выраженные аналитические способности. Проведен критический анализ и обобщение информации из различных источников, представлена собственная оценка рассматриваемых проблем и методов.
- Научная новизна и самостоятельность: содержатся элементы научного творчества и самостоятельности в обобщении материала, формулировке выводов.
- Оформление и источники: список литературы соответствует заявленной теме, включает современные и релевантные научные публикации. Соблюдены все требования к оформлению текста, цитированию и библиографии.

«Хорошо» (70–84%):

- Актуальность и глубина раскрытия темы: тема раскрыта, однако глубина анализа недостаточна. Имеются незначительные фактические неточности.
- Структура и логика: структура работы в основном соблюдена, но логические связи между разделами могут быть проработаны лучше.
- Анализ и обобщение: информация в основном систематизирована и обобщена, но анализ носит в большей степени описательный характер с элементами критического осмысления.
- Научная новизна и самостоятельность: работа демонстрирует самостоятельность в подборе и компоновке материала, но выводы носят в основном компилятивный характер.

- Оформление и источники: список литературы достаточен, но может содержать источники, не в полной мере соответствующие теме, или устаревшие работы. Имеются незначительные погрешности в оформлении.

«Удовлетворительно» (55–69%):

- Актуальность и глубина раскрытия темы: тема раскрыта поверхностно, фрагментарно. Содержатся фактические ошибки.
- Структура и логика: нарушена логическая структура работы. Связь между разделами слабая.
- Анализ и обобщение: анализ отсутствует или подменен пересказом источников. Обобщения носят декларативный характер и не аргументированы.
- Научная новизна и самостоятельность: работа представляет собой компиляцию фрагментов из источников без их глубокого осмысления и анализа.
- Оформление и источники: список литературы неполный, содержит неавторитетные или устаревшие источники. Имеются существенные нарушения в оформлении.

«Неудовлетворительно» (0–54%):

- Актуальность и глубина раскрытия темы: тема не раскрыта.
- Структура и логика: структура работы не выдержана.
- Анализ и обобщение: анализ и обобщение отсутствуют.
- Научная новизна и самостоятельность: признаки самостоятельной работы отсутствуют, имеются признаки плагиата.
- Оформление и источники: список литературы отсутствует или совершенно не соответствует теме. Требования к оформлению не выполнены.

### **3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания**

Форма промежуточной аттестации: зачет с оценкой.

Структура билета: Билет состоит из трех частей и проверяет достижение всех запланированных индикаторов компетенций.

Тест. Проверяет индикатор ИОПК-7.2. Содержит 5 вопросов с выбором одного или нескольких правильных ответов.

Теоретический вопрос. Проверяет индикатор ИОПК-8.1. Содержит 1 вопрос, требующий развернутого ответа с опорой на методологические принципы.

Практико-ориентированная задача. Проверяет индикаторы ИОПК-6.2 и ИПК-1.2. Содержит 1 задачу, оформленную в виде кейса или ситуации, требующей решения и краткой интерпретации полученных результатов.

Основной целью проекта «Геном человека» являлось:

- а) Клонирование первого млекопитающего
- б) Определение полной последовательности нуклеотидов ДНК человека
- в) Разработка алгоритма BLAST
- г) Создание первого персонального компьютера для биологов

К международным базам данных биологической информации относятся:

- а) NCBI (National Center for Biotechnology Information)
- б) UniProt (Universal Protein Resource)
- в) Microsoft Excel
- г) Protein Data Bank (PDB)

Алгоритм BLAST предназначен для:

- а) Визуализации трехмерной структуры белков
- б) Поиска гомологичных последовательностей в базах данных
- в) Проведения статистического анализа данных RNA-seq
- г) Хранения образцов ДНК

Понятие «онтология генов» (Gene Ontology) подразумевает:

- а) Местоположение гена на конкретной хромосоме
- б) Стандартизированную систему описания функций генов и их продуктов
- в) Процесс секвенирования генома
- г) Правила поведения в лаборатории

Формат файла FASTA используется для хранения:

- а) Трехмерных структур белков
- б) Параметров микроскопа
- в) Нуклеотидных или аминокислотных последовательностей
- г) Результатов статистического тестирования

Ключи: 1 - б), 2 – а), б), г), 3 - б), 4 - б), 5 - в)

Перечень теоретических вопросов:

1. Опишите основные этапы анализа данных RNA-seq. Дайте краткую характеристику цели и основным методам каждого этапа.
2. Дайте определение биоинформатики. Опишите ее основные задачи и место в системе современных биологических наук.
3. Что такое эволюционная консервативность последовательности? Объясните, как этот принцип используется для предсказания функциональной значимости участков генома.
4. Опишите принцип работы алгоритма BLAST. Какие типы BLAST существуют и для решения каких задач они применяются?
5. Что такое Gene Ontology (GO)? Опишите три основные составляющие онтологии и принцип их использования для функциональной аннотации генов.
6. Опишите основные проблемы и сложности сборки геномов *de novo* из коротких прочтений.
7. Объясните, что такое e-value и percent identity в контексте анализа результатов BLAST. Как эти параметры помогают оценить достоверность найденного гомолога?
8. Что такое филогенетическое дерево? По какому принципу строится и как интерпретируются его элементы (узлы, ветви, длина ветвей)?
9. Опишите стратегию и основные этапы функциональной аннотации вновь секвенированного гена.
10. Дайте характеристику основным международным базам данных биологической информации (NCBI, EBI, UniProt, PDB). Какова их основная задача?
11. В чем заключаются основные проблемы хранения и обработки Big Data в биоинформатике?
12. Опишите принципы и методы, используемые для контроля качества (QC) данных высокопроизводительного секвенирования.

Задачи:

Задача 1. BLAST-поиск белка дал два лучших результата: 1) с  $e\text{-value} = 3e-50$ , percent identity = 80%, query coverage = 95%; 2) с  $e\text{-value} = 0.75$ , percent identity = 35%, query coverage = 98%. Объясните значения этих параметров. Белок с какой последовательностью является более надежным гомологом и почему? Сформулируйте предположение о функции исследуемого белка.

Задача 2. При анализе RNA-seq для гена X в контрольной группе (3 повтора) было получено в среднем 1000 reads, в опытной группе (3 повтора) — в среднем 5000 reads. Общее количество прочитанных ридов (library size) во всех библиотеках было примерно одинаковым. Рассчитайте примерное кратность изменения ( $\log_2\text{FoldChange}$ ) экспрессии гена X. Сделайте предварительное предположение об изменении его экспрессии. Какой основной статистический тест необходимо применить для подтверждения достоверности этого изменения и почему?

Задача 3. Вам необходимо найти информацию о белке-инсулине человека (*Homo sapiens*): нуклеотидную последовательность, белковую последовательность, известные домены и, если есть, 3D-структуру. Составьте краткий алгоритм ваших действий: перечислите основные базы данных, которые вы будете использовать для каждого типа информации, и последовательность шагов.

Критерии оценивания итогового контроля:

Результаты зачета определяются оценками «зачтено» или «не зачтено».

Оценка «зачтено» выставляется, если:

- Дан правильный ответ на 3 и более вопроса теста.
- Дан как минимум правильный ответ на теоретический вопрос с несущественными неточностями.
- Задача решена в целом верно с несущественными ошибками в расчетах или интерпретации.

В ином случае выставляется оценка «не зачтено».

#### **4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций)**

Выберите один или несколько правильных ответов.

1. Основной целью проекта «Геном человека» являлось (ИОПК-7.2):

- а) Клонирование первого млекопитающего
- б) Определение полной последовательности нуклеотидов ДНК человека
- в) Разработка алгоритма BLAST
- г) Создание первого персонального компьютера для биологов

2. К международным базам данных биологической информации относятся (ИОПК-7.2):

- а) NCBI (National Center for Biotechnology Information)
- б) UniProt (Universal Protein Resource)
- в) Microsoft Excel
- г) Protein Data Bank (PDB)

3. Алгоритм BLAST предназначен для (ИОПК-7.2):

- а) Визуализации трехмерной структуры белков
- б) Поиска гомологичных последовательностей в базах данных
- в) Проведения статистического анализа данных RNA-seq
- г) Хранения образцов ДНК

4. Понятие «онтология генов» (Gene Ontology) подразумевает (ИОПК-8.1):

- а) Местоположение гена на конкретной хромосоме

- б) Стандартизированную систему описания функций генов и их продуктов
- в) Процесс секвенирования генома
- г) Правила поведения в лаборатории

5. Формат файла FASTA используется для хранения (ИОПК-7.2):

- а) Трехмерных структур белков
- б) Параметров микроскопа
- в) Нуклеотидных или аминокислотных последовательностей
- г) Результатов статистического тестирования

6. Основная проблема, решаемая на этапе контроля качества (QC) данных NGS, — это (ИОПК-8.1):

- а) Определение эволюционного возраста образца
- б) Интерпретация биологического значения найденных различий
- в) Идентификация технических артефактов и последовательностей низкого качества
- г) Поиск публикаций по теме исследования в научных журналах

7. Принцип выравнивания последовательностей по алгоритму Needleman-Wunsch является (ИОПК-6.2):

- а) Локальным
- б) Глобальным
- в) Статистическим
- г) Эвристическим

8. Параметр E-value в результате BLAST указывает на (ИОПК-6.2):

- а) Процент идентичности последовательностей
- б) Ожидаемое количество случайных совпадений
- в) Длину выравнивания
- г) Эволюционное расстояние

9. Для хранения трехмерных структур биологических макромолекул используется база данных (ИОПК-7.2):

- а) GenBank
- б) PDB (Protein Data Bank)
- в) UniProt
- г) PubMed

10. Основным инструментом для визуализации аннотированных геномов является (ИОПК-7.2):

- а) Microsoft Word
- б) GenBank
- в) Геномный браузер (например, UCSC Genome Browser)
- г) BLAST

11. Математическая основа построения филогенетических деревьев включает (ИОПК-6.2):

- а) Только морфологические признаки
- б) Статистические методы анализа множественных выравниваний
- в) Исключительно данные палеонтологии
- г) Субъективные оценки исследователей

12. Этап сборки (ассемблирования) генома *de novo* предназначен для (ИОПК-8.1):

- а) Сравнения с референсным геномом
- б) Восстановления полной последовательности генома из коротких прочтений
- в) Аннотации генов
- г) Визуализации хромосом

13. Метод RNA-seq позволяет определить (ИОПК-8.1):

- а) Только последовательность ДНК
- б) Уровень экспрессии генов
- в) Трехмерную структуру белков
- г) Скорость эволюции

14. Для функциональной аннотации неизвестного гена НЕ используется (ИОПК-8.1):

- а) Поиск гомологов в базах данных
- б) Анализ онтологии генов (GO)
- в) Определение плазмидной карты
- г) Предсказание доменной структуры белка

15. При анализе данных RNA-seq для идентификации дифференциально экспрессируемых генов применяется (ИОПК-6.2):

- а) Только визуальный осмотр
- б) Статистические тесты (например, t-тест, ANOVA)
- в) Исключительно качественные оценки
- г) Сравнение без повторностей

16. Понятие «Big Data» в биоинформатике связано с (ИОПК-8.1):

- а) Малыми объемами данных
- б) Необходимостью обработки терабайтов информации с секвенаторов
- в) Отсутствием необходимости в специализированных методах анализа
- г) Преобладанием ручной обработки

17. Этап предобработки данных NGS включает (ИОПК-8.1):

- а) Только интерпретацию результатов
- б) Фильтрацию низкокачественных прочтений и адаптеров
- в) Исключительно проведение экспериментов
- г) Публикацию статей

18. Для проверки гипотезы о дифференциальной экспрессии гена по данным RNA-seq используется (ИОПК-6.2):

- а) Только значение количества ридов
- б) Статистическая значимость (p-value) и кратность изменения (fold change)
- в) Исключительно цвет графиков
- г) Мнение эксперта

19. Инструмент для работы с геномными данными, который НЕ является базой данных (ИОПК-7.2):

- а) NCBI Nucleotide
- б) UniProt
- в) Microsoft Excel
- г) ENSEMBL

20. Основная цель использования референсного генома при анализе NGS данных — это (ИОПК-8.1):

- а) Увеличение стоимости анализа
- б) Картирование прочтений для определения их позиции
- в) Уменьшение точности
- г) Игнорирование индивидуальных особенностей организма

Ключ для проверки:

1 – б), 2 – а), б), г), 3 – б), 4 – б), 5 – в), 6 – в), 7 – б), 8 – б), 9 – б), 10 – в), 11 – б), 12 – б), 13 – б), 14 – в), 15 – б), 16 – б), 17 – б), 18 – б), 19 – в), 20 – б).

Теоретические вопросы:

1. Определение и основные задачи биоинформатики как дисциплины (ИОПК-7.2, ИОПК-8.1).

Ответ должен содержать: определение биоинформатики; задачи, связанные с хранением, обработкой и анализом биологических данных; примеры практического применения (например, в медицине или эволюционной биологии).

2. Принципы работы и основные этапы алгоритма BLAST (ИОПК-7.2, ИОПК-6.2).

Ответ должен содержать: назначение алгоритма; описание этапов (формирование запроса, поиск слов-затравок, расширение выравнивания, статистическая оценка); типы BLAST (nucleotide, protein) и их применение.

3. Онтология генов (Gene Ontology): структура и применение (ИОПК-8.1).

Ответ должен содержать: определение GO; описание трех аспектов (молекулярная функция, биологический процесс, клеточная компонента); роль в функциональной аннотации генов.

4. Этапы анализа данных RNA-seq (ИОПК-8.1, ИОПК-6.2).

Ответ должен содержать: описание этапов (контроль качества, выравнивание на референсный геном, подсчет ридов, статистический анализ дифференциальной экспрессии); методы и инструменты для каждого этапа.

5. Проблемы сборки геномов *de novo* (ИОПК-8.1).

Ответ должен содержать: основные сложности (повторы, гетерозиготность, ошибки секвенирования); методы преодоления (использование длинных прочтений, оценка качества сборки).

6. Принципы филогенетического анализа (ИОПК-6.2).

Ответ должен содержать: цели филогенетического анализа; этапы (построение множественного выравнивания, выбор модели эволюции, реконструкция дерева); интерпретация элементов дерева.

7. Базы данных в биоинформатике: классификация и применение (ИОПК-7.2).

Ответ должен содержать: классификацию по типу данных (геномные, протеомные, структурные); примеры (NCBI, UniProt, PDB); принципы поиска и анализа информации.

8. Контроль качества (QC) данных NGS (ИОПК-8.1).

Ответ должен содержать: цели QC; параметры оценки качества ридов (длина, quality scores, содержание адаптеров); инструменты для анализа (FastQC).

9. Методы предсказания функции белка (ИОПК-6.2, ИОПК-8.1).

Ответ должен содержать: подходы на основе гомологии, доменной структуры, экспрессии; использование машинного обучения; оценка точности предсказаний.

10. Анализ дифференциальной экспрессии генов: методы и интерпретация (ИОПК-6.2, ИПК-1.2).

Ответ должен содержать: постановку задачи; статистические методы (DESeq2, edgeR); интерпретацию результатов (fold change, p-value, FDR).

### **Информация о разработчиках**

Слепцов Алексей Анатольевич, кандидат медицинских наук, кафедра физиологии растений, биотехнологии и биоинформатики Биологического института Национального исследовательского Томского государственного университета, доцент.