

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Наумова Наталия Александровна

Должность: Ректор

Дата подписания: 17.08.2026 15:03:23

Уникальный программный ключ:

6b5279da4e034bffa79172803da5b7b559fc69e2

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования

«ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ»

(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ)

Факультет естественных наук
Кафедра общей биологии и биоэкологии

УТВЕРЖДЁН

на заседании кафедры общей биологии и
биоэкологии

Протокол от «28» апреля 2026 г. № 9

Заведующий кафедрой

 Гордеев М.И./

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

По дисциплине

БИОИНФОРМАЦИОННЫЕ ПОДХОДЫ В ГЕНОМИКЕ

Направление подготовки

06.04.01 Биология

Программа подготовки

Биоинформатика и анализ больших данных в экологическом мониторинге

Квалификация

Магистр

Форма обучения

Очная

Москва

2026

Авторы-составители:

Гордеев М.И., д.б.н., профессор кафедры общей биологии и биоэкологии ФГАОУ ВО «ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ»

Москаев А.В., к.б.н., доцент кафедры общей биологии и биоэкологии ФГАОУ ВО «ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ»

Темников А.А., старший преподаватель кафедры общей биологии и биоэкологии ФГАОУ ВО «ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ»

Фонд оценочных средств по дисциплине «Биоинформационные подходы в геномике» составлен в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки 06.04.01 Биология, утверждённого приказом МИНОБРНАУКИ России от 11.08.2020 № 934.

Дисциплина входит в обязательную часть Блока 1 «Дисциплины (модули)» и является обязательной для изучения.

Оглавление

1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы	5
2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания.....	5
3. Типовые контрольные задания	13
3.1. Тестовые задания	13
3.2. Ситуационные задачи (краткий ответ)	19
3.3. Открытые вопросы (краткий ответ).....	19
3.4. Задания для практических занятий	20
3.5. Вопросы для опроса.....	20
3.6. Темы рефератов	21
3.7. Темы докладов	22
3.8. Темы презентаций.....	23
3.9. Темы курсовых работ	24
3.10. Вопросы к экзамену.....	24
4. Методические материалы, определяющие процедуры оценивания	26
4.1. Общие положения.....	26
4.2. Формы текущего контроля и их балльная оценка	26
4.3. Процедура проведения экзамена.....	27
4.4. Порядок ликвидации академической задолженности	28

1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы

Код и наименование компетенции	Этапы формирования
ОПК-2. Способен творчески использовать в профессиональной деятельности знания фундаментальных и прикладных разделов дисциплин (модулей), определяющих направленность программы магистратуры.	1. Работа на учебных занятиях (лекции, практические занятия). 2. Самостоятельная работа (подготовка к опросам, выполнение рефератов, докладов, презентаций, курсовой работы).
ОПК-5. Способен участвовать в создании и реализации новых технологий в сфере профессиональной деятельности и контроле их экологической безопасности с использованием живых объектов.	1. Работа на учебных занятиях (лекции, практические занятия). 2. Самостоятельная работа (подготовка к опросам, выполнение рефератов, докладов, презентаций, курсовой работы).
ОПК-6. Способен творчески применять и модифицировать современные компьютерные технологии, работать с профессиональными базами данных, профессионально оформлять и представлять результаты новых разработок.	1. Работа на учебных занятиях (лекции, практические занятия). 2. Самостоятельная работа (подготовка к опросам, выполнение рефератов, докладов, презентаций, курсовой работы).
СПК-2. Способен использовать программы для эпигенетического анализа и качества эпигенетических данных.	1. Работа на учебных занятиях (лекции, практические занятия). 2. Самостоятельная работа (подготовка к опросам, выполнение рефератов, докладов, презентаций, курсовой работы).

2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
ОПК-2 – Способен творчески использо-	Пороговый	1. Работа на учебных	знать: – фундаментальные принципы организации геномов и методы их аннотации; – алгоритмические ос-	Опрос, тестирование	Шкала оценивания опроса. Шкала оце-

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
вать в профессиональной деятельности знания фундаментальных и прикладных разделов дисциплин (модулей), определяющих направленность программы магистратуры		занятиях	новы базовых биоинформатических инструментов (выравнивание, динамическое программирование, скрытые марковские модели); – структуру и логику работы основных международных биологических баз данных. уметь: – грамотно выбирать биоинформатический инструментарий под типовую задачу геномного анализа; – интерпретировать форматы FASTA, FASTQ, SAM/BAM, VCF; – проводить выравнивание коротких прочтений на референсный геном и визуализировать результат.		нивания тестирования.
	Продвинутый	1. Работа на учебных занятиях 2. Самостоятельная работа	уметь: – творчески комбинировать методы геномики, транскриптомики и эпигеномики для проверки нестандартных биологических гипотез; – критически оценивать ограничения алгоритмов и адаптировать вычислительные пайплайны под нетиповые экспериментальные данные; – самостоятельно проектировать воспроизводимый конвейер анализа NGS-данных в виде скрипта или workflow. владеть: – способностью использовать фундаментальные знания молекулярной биологии и эволюции для интерпретации результатов полногеномных исследований; – техникой переноса аналитических подходов из смежных дисциплин (статистика, машинное обучение)	Доклад, презентация, реферат, курсовая работа	Шкала оценивания доклада. Шкала оценивания презентации. Шкала оценивания реферата. Шкала оценивания курсовой работы

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			в собственную научную работу.		
ОПК-5 – Способен участвовать в создании и реализации новых технологий в сфере профессиональной деятельности и контроле их экологической безопасности с использованием живых объектов	Пороговый	1. Работа на учебных занятиях	знать: – технологическую цепочку получения геномных данных от пробоподготовки до инструментального анализа; – современные платформы секвенирования (Illumina, PacBio, ONT) и области их применения; – принципы метагеномного анализа и оценки биоразнообразия с использованием молекулярных маркеров. уметь: – выполнять контроль качества исходных прочтений и оценивать пригодность данных для дальнейшего анализа; – применять стандартные протоколы таксономического профилирования микробных сообществ (16S, метагеномная сборка).	Опрос, тестирование	Шкала оценивания опроса. Шкала оценивания тестирования.
	Продвинутый	1. Работа на учебных занятиях 2. Самостоятельная работа	уметь: – разрабатывать элементы новых технологических решений для анализа живых систем (например, дизайн пайплайна для выявления новых биосинтетических кластеров); – интерпретировать метагеномные данные в контексте экологической безопасности и биоремедиации; – оценивать риски горизонтального переноса генов в природных и техногенных нишах с помощью биоинформатических методов. владеть: – навыками биннинга и восстановления геномов (MAGs) из метагеномныхборок для	Доклад, презентация, реферат, курсовая работа	Шкала оценивания доклада. Шкала оценивания презентации. Шкала оценивания реферата. Шкала оценивания курсовой работы

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			описания новых форм живых объектов; – приёмами валидации вычислительных предсказаний экспериментальными методами.		
ОПК-6 – Способен творчески применять и модифицировать современные компьютерные технологии, работать с профессиональными базами данных, профессионально оформлять и представлять результаты новых разработок	Пороговый	1. Работа на учебных занятиях	знать: – основные языки и среды программирования в биоинформатике (Bash, Python, R); – принципы работы реляционных и нереляционных биологических баз данных; – стандарты представления геномных данных (GFF, BED, VCF, FASTA). уметь: – писать простые скрипты для парсинга и преобразования биологических форматов; – осуществлять программный доступ к данным через API и запросы к Entrez/SRA; – подготавливать графики и диаграммы, отражающие результаты вычислительного анализа.	Опрос, тестирование	Шкала оценивания опроса. Шкала оценивания тестирования.
	Продвинутый	1. Работа на учебных занятиях 2. Самостоятельная работа	уметь: – модифицировать существующие алгоритмы и пайплайны под нестандартные научные задачи; – создавать интерактивные отчёты (R Markdown, Jupyter Notebook) с полным циклом воспроизводимого анализа; – внедрять системы управления workflow (Snakemake/Nextflow) в собственную исследовательскую практику. владеть: – техникой профессиональной визуализации многомерных омиксных данных; – навы-	Доклад, презентация, реферат, курсовая работа	Шкала оценивания доклада. Шкала оценивания презентации. Шкала оценивания реферата. Шкала оценивания курсовой работы

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			ками публичного депонирования обработанных данных и метаданных в международные архивы (NCBI, ENA) в соответствии с принятыми стандартами.		
СПК-2 – Способен использовать программы для эпигенетического анализа и качества эпигенетических данных	Пороговый	1. Работа на учебных занятиях	знать: – типы эпигенетических модификаций и соответствующие высокопроизводительные методы их детекции (ChIP-seq, ATAC-seq, WGBS, Hi-C); – специализированные форматы данных (bigWig, bedGraph, pairs) и инструменты их визуализации (IGV, UCSC Browser); – стандартные метрики качества эпигеномных экспериментов (FRiP, NSC, RSC). уметь: – запускать программы контроля качества для ChIP-seq и ATAC-seq (FastQC, MultiQC, ChIPQC); – выполнять базовый пик-коллинг с помощью MACS2 и визуализировать профили покрытия.	Опрос, тестирование	Шкала оценивания опроса. Шкала оценивания тестирования.
	Продвинутый	1. Работа на учебных занятиях 2. Самостоятельная работа	уметь: – проводить дифференциальный анализ участков связывания и доступности хроматина (DiffBind, csaw); – выполнять интеграцию эпигеномных данных с транскриптомными (RNA-seq + ChIP-seq) для реконструкции регуляторных сетей; – оценивать качество конформационных данных Hi-C и строить контактные матрицы. владеть: – навыками мотивного анализа и предсказания факторов	Доклад, презентация, реферат, курсовая работа	Шкала оценивания доклада. Шкала оценивания презентации. Шкала оценивания реферата. Шкала оценивания курсовой работы

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			транскрипции, ассоциированных с выявленными регионами (HOMER, MEME); – техникой аннотации геномных районов на основе интегрированных эпигеномных разметок (ChromHMM, Segway).		

Шкала оценивания тестирования

Критерии оценивания	Баллы
80–100% правильных ответов – «отлично»	8–10
60–79% правильных ответов – «хорошо»	6–8
30–59% правильных ответов – «удовлетворительно»	3–5
0–29% правильных ответов – «неудовлетворительно»	0–2

Максимальное количество баллов – 20 (за два теста по 10 баллов).

Шкала оценивания опроса

Критерии оценивания	Баллы
Ответ полный и содержательный, соответствует теме; магистрант умеет аргументировано отстаивать свою точку зрения, демонстрирует знание терминологии дисциплины.	5
Ответ в целом соответствует теме (не отражены некоторые аспекты); магистрант умеет отстаивать свою точку (хотя аргументация не всегда на должном уровне); демонстрирует удовлетворительное знание терминологии дисциплины.	2
Ответ неполный как по объёму, так и по содержанию (хотя и соответствует теме); аргументация не на соответствующем уровне, некоторые проблемы с употреблением терминологии дисциплины.	1

Максимальное количество баллов – 20 (4 опроса по 5 баллов).

Шкала оценивания реферата

Показатель	Баллы
Содержание соответствует поставленным цели и задачам, изложение материала отличается логичностью и смысловой завершённостью, студент показал владение материалом, умение четко, аргументировано и корректно отвечать на поставленные вопросы, отстаивать собственную точку зрения.	10–20
Содержание недостаточно полно соответствует поставленным цели и задачам исследования, работа выполнена на недостаточно широкой источниковой базе и не учитывает новейшие достижения науки, изложение материала носит преимущественно описательный характер, студент показал достаточно уверенное владение материалом, однако недостаточное умение четко, аргументировано и корректно отвечать на поставленные вопросы и отстаивать собственную точку зрения.	6–9
Содержание не отражает особенности проблематики избранной темы; содержание работы не полностью соответствует поставленным задачам, источниковая база является фрагментарной и не позволяет качественно решить все поставленные в работе задачи, работа не учитывает новейшие достижения историографии темы, студент показал неуверенное владение материалом, неумение отстаивать собственную позицию и отвечать на вопросы.	3–5
Работа не имеет логичной структуры, содержание работы в основном не соответствует теме, источниковая база исследования является недостаточной для решения поставленных задач, студент показал неуверенное владение материалом, неумение формулировать собственную позицию.	0–2

Максимальное количество баллов – 20.

Шкала оценивания доклада

Критерии оценивания	Баллы
Доклад соответствует заявленной теме, выполнен с привлечением достаточного количества научных и практических источников по теме, магистрант в состоянии ответить на вопросы по теме доклада.	5
Доклад в целом соответствует заявленной теме, выполнен с привлечением нескольких научных и практических источников по теме, студент в состоянии ответить на часть вопросов по теме доклада.	2
Доклад не совсем соответствует заявленной теме, выполнен с использованием только 1 или 2 источников, студент допускает ошибки при изложении материала, не в состоянии ответить на вопросы по теме доклада.	1

Максимальное количество баллов – 5.

Шкала оценивания презентации

Критерии оценивания	Баллы
Представляемая информация систематизирована, последовательна и логически связана. Проблема раскрыта полностью. Широко использованы возможности технологии PowerPoint.	5
Представляемая информация в целом систематизирована, последовательна и логически связана (возможны небольшие отклонения). Проблема раскрыта. Возможны незначительные ошибки при оформлении в PowerPoint (не более двух).	2
Представляемая информация не систематизирована и/или не совсем последовательна. Проблема раскрыта не полностью. Выводы не сделаны или не обоснованы. Возможности технологии PowerPoint использованы лишь частично.	1

Максимальное количество баллов – 5.

Шкала оценивания курсовой работы

Показатель	Баллы
Содержание полностью соответствует целям и задачам, тема раскрыта глубоко и всесторонне. Изложение отличается безупречной логикой, смысловой завершенностью и высоким научным стилем. Работа выполнена на широкой источниковой базе, включающей новейшие отечественные и зарубежные публикации; присутствует оригинальный анализ и корректная биоинформатическая/статистическая обработка данных. Вычислительная часть (пайплайн, скрипты) полностью воспроизводима, код задокументирован. Оформление строго соответствует требованиям. На защите студент демонстрирует свободное владение материалом, уверенно и аргументированно отвечает на все вопросы, отстаивает собственную точку зрения, подкрепляя её данными. Доклад структурирован, презентация визуально информативна.	81–100
Содержание полностью соответствует целям и задачам, тема раскрыта достаточно полно. Изложение логично, выводы обоснованы. Источниковая база современна и репрезентативна, но имеются незначительные пробелы в обзоре отдельных узких аспектов. Биоинформатическая часть выполнена корректно, однако могут присутствовать мелкие недочёты в оформлении кода или воспроизводимости. Оформление работы в целом соответствует требованиям, есть единичные технические погрешности. На защите студент показывает хорошее владение материалом, отвечает на большинство вопросов, но в ответах на самые сложные вопросы может испытывать незначительные затруднения; свою позицию аргументирует убедительно.	61–80
Содержание недостаточно полно соответствует поставленным цели и задачам: тема в целом раскрыта, но отдельные её аспекты освещены поверхностно. Работа опирается на ограниченный круг источников, новейшие достижения науки учтены слабо. Изложение материала носит преимущественно описательный характер, элементы самостоятельного анализа присут-	41–60

Показатель	Баллы
ствуют, но неглубоки. Биоинформатическая часть выполнена с погрешностями (не все шаги обоснованы, код плохо читаем или не воспроизводится). Оформление содержит заметные отклонения от требований. На защите студент демонстрирует уверенное, но не всегда глубокое владение материалом; на часть вопросов отвечает с помощью наводящих уточнений, испытывает затруднения в аргументации собственной позиции.	
Содержание не отражает проблематику темы или не полностью соответствует задачам. Источниковая база фрагментарна и не позволяет качественно решить поставленные задачи; многие важные современные публикации проигнорированы. Материал изложен непоследовательно, преобладает компиляция без осмысления. Вычислительная часть либо отсутствует, либо выполнена с грубыми ошибками. Оформление работы имеет серьёзные недостатки. На защите студент демонстрирует неуверенное владение материалом, путается в основных понятиях, не умеет чётко отвечать на вопросы и отстаивать собственную позицию. Ответы односложны и неглубоки.	21–40
Работа не имеет логичной структуры, содержание в основном не соответствует теме и поставленным задачам. Источниковая база крайне скудна или использована некорректно (плагиат, подмена темы). Отсутствует самостоятельная биоинформатическая проработка. Оформление грубо нарушает все основные требования. Студент практически не владеет материалом, не способен ответить на базовые вопросы по теме, не может сформулировать ни цели, ни выводы, ни собственную позицию. Защита сводится к чтению фрагментов текста без понимания сути.	0–20

Итоговая шкала оценивания по курсовой работе

Сумма баллов	Оценка
81–100	отлично
61–80	хорошо
41–60	удовлетворительно
0–40	неудовлетворительно

3. Типовые контрольные задания

3.1. Тестовые задания

Вариант 1

Выберите один правильный ответ из четырёх предложенных.

- Какой формат файла используется для хранения нуклеотидных последовательностей с оценками качества каждого основания?
 - FASTA
 - SAM
 - FASTQ

г) VCF

Ответ: в

2. Какой алгоритм лежит в основе работы выравнивателя BWA-MEM?

а) Динамическое программирование по Смит–Ватерману

б) Поиск точных совпадений с помощью FM-индекса и последующее расширение

в) Скрытые марковские модели

г) Хеширование k-меров

Ответ: б

3. Что означает значение MAPQ в столбце 5 SAM/BAM-файла?

а) Количество картированных пар

б) Качество выравнивания чтения, логарифмически связанное с вероятностью ошибочного картирования

в) Процент идентичности чтения референсу

г) Количество совпадающих нуклеотидов

Ответ: б

4. Какой формат используют для хранения информации о генетических вариантах?

а) BED

б) GFF

в) VCF

г) FASTA

Ответ: в

5. Какую основную задачу решает программа FastQC?

а) Тримминг адаптеров

б) Контроль качества прочтений и визуализация метрик

в) Картирование ридов на геном

г) Поиск генетических вариантов

Ответ: б

6. Какой инструмент предназначен для удаления адаптерных последовательностей из FASTQ-файлов?

а) SAMtools

б) BWA

в) Trimmomatic

г) MACS2

Ответ: в

7. Какой пакет языка R является стандартом для анализа дифференциальной экспрессии генов по данным RNA-seq?

а) ggplot2

б) DESeq2

в) limma (без voom)

г) Seurat

Ответ: б

8. Какую структуру данных используют графовые сборщики генома de novo (например, SPAdes)?

а) Двоичное дерево поиска

б) Граф де Брёйна

в) Хеш-таблицу k-меров

г) Нейронную сеть

Ответ: б

9. Какая метрика качества ChIP-seq эксперимента показывает долю прочтений, попадающих в пики?

а) N50

б) FRiP

в) MAPQ

г) RPKM

Ответ: б

10. Какая поправка на множественное тестирование обычно используется при анализе дифференциальной экспрессии тысяч генов?

а) Поправка Бонферрони

б) Метод Бенджамини–Хохберга

в) Критерий Стьюдента

г) Точный тест Фишера

Ответ: б

11. Что описывает BED-файл?

а) Последовательности белков

б) Геномные координаты регионов (экзонов, пиков и т.д.)

в) Филогенетические деревья

г) Уровни экспрессии генов

Ответ: б

12. Какой командой можно отсортировать BAM-файл по координатам?

а) samtools index

б) samtools sort

в) samtools view

г) samtools merge

Ответ: б

13. Что такое скрытая марковская модель (НММ) в биоинформатике?

а) Графический метод визуализации геномных данных

б) Статистическая модель, широко применяемая для предсказания генов и профильного анализа

в) Алгоритм выравнивания последовательностей

г) Формат хранения аннотаций

Ответ: б

14. Для чего в первую очередь используется программа MACS2?

а) Картирование ридов

б) Поиск пиков в ChIP-seq данных

в) Сборка транскриптома

г) Аннотация вариантов

Ответ: б

15. Пангеном бактерий включает:

а) Только гены, общие для всех штаммов вида

б) Совокупность всех генов, встречающихся у представителей вида (core + accessory)

в) Только гены, полученные горизонтальным переносом

г) Только гены, находящиеся на плазмидах

Ответ: б

16. Какой из перечисленных выравнивателей специально разработан для картирования сплайсированных ридов RNA-seq?

а) BWA aln

б) Bowtie2

в) STAR

г) minimap2

Ответ: в

17. Что измеряет метрика TPM (Transcripts Per Million) в RNA-seq данных?

а) Общее число картированных фрагментов

б) Нормализованную экспрессию с учётом длины гена и глубины библиотеки

в) Количество нуклеотидов в транскрипте

- г) Процент дублированных ридов
Ответ: б
18. Какой инструмент используется для функциональной аннотации генетических вариантов (например, предсказания влияния на белок)?
а) FastQC
б) SnpEff / Ensembl VEP
в) Trimmomatic
г) BCFtools
Ответ: б
19. Что называют биннингом в метагеномике?
а) Удаление химерных ридов
б) Группировку контигов, предположительно принадлежащих одному геному
в) Подсчёт таксономического состава по 16S рРНК
г) Аннотацию функций генов
Ответ: б
20. Какой метод позволяет изучать трёхмерную организацию хроматина в масштабе всего генома?
а) ATAC-seq
б) Hi-C
в) ChIP-seq
г) WGBS
Ответ: б

Вариант 2

Выберите один правильный ответ из четырёх предложенных.

1. В каком формате хранится информация о выравнивании прочтений на референсный геном в текстовом виде?
а) FASTQ
б) SAM
в) BED
г) GFF
Ответ: б
2. Какая программа позволяет создать единый отчёт о качестве множества образцов, объединяя результаты FastQC?
а) Trimmomatic
б) MultiQC
в) MACS2
г) IGV
Ответ: б
3. Что показывает параметр E-value в результатах поиска BLAST?
а) Длину выравнивания
б) Ожидаемое количество выравниваний с такой же или лучшей оценкой, возникающих случайно
в) Процент идентичности
г) Вероятность ошибки картирования
Ответ: б
4. Для какой основной цели используется пакет bedtools?
а) Поиск генетических вариантов
б) Манипуляции с геномными интервалами (пересечение, объединение, извлечение последовательностей)
в) Статистический анализ дифференциальной экспрессии

- г) Визуализация филогенетических деревьев
Ответ: б
5. На какой структуре индексации основана быстрая работа Bowtie2?
а) Суффиксное дерево
б) FM-индекс, построенный на борроу-вилеровском преобразовании
в) Хеш-таблица всех k-меров
г) Индекс на основе сортированных k-меров minimap
Ответ: б
6. Что такое контиг в контексте сборки генома?
а) Отдельное прочтение после фильтрации
б) Непрерывная последовательность, полученная в результате сборки перекрывающихся ридов
в) Генетический вариант
г) Пара концов одного фрагмента ДНК
Ответ: б
7. Какой формат обычно используется для хранения аннотаций генов (координаты экзонов, кодирующих областей и т.д.)?
а) FASTA
б) GTF/GFF
в) VCF
г) BAM
Ответ: б
8. Какой из перечисленных инструментов предназначен для сборки генома de novo?
а) BWA
б) SPAdes
в) DESeq2
г) MACS2
Ответ: б
9. SNP (Single Nucleotide Polymorphism) — это:
а) Структурная вариация длиной более 1 т.п.н.
б) Вставка мобильного элемента
в) Однонуклеотидная замена в геномной последовательности
г) Изменение числа копий гена
Ответ: в
10. Какой пакет R широко используется для анализа данных секвенирования единичных клеток (scRNA-seq)?
а) DESeq2
б) Seurat
в) edgeR
г) limma
Ответ: б
11. Филогенетическое дерево отображает:
а) Уровни экспрессии генов в разных тканях
б) Эволюционные отношения между организмами или генами
в) Пространственную структуру белка
г) Контактную матрицу хроматина
Ответ: б
12. При анализе ATAC-seq данных в MACS2 часто добавляют параметр --nomodel --shift - 100 --extsize 200. Это делается для:
а) Увеличения чувствительности к точкам связывания транскрипционных факторов
б) Учёта сдвига, вызванного транспозазой Tn5
в) Игнорирования модели фрагментов и фиксации ожидаемого размера фрагмента для

открытого хроматина

г) Все перечисленное

Ответ: г

13. Что такое GC-состав генома?

а) Доля гуанина и цитозина среди всех нуклеотидов

б) Среднее качество прочтений

в) Глубина покрытия

г) Количество генов на миллион пар оснований

Ответ: а

14. Какой командой SAMtools можно преобразовать SAM-файл в отсортированный BAM-файл?

а) `samtools view -bS sample.sam | samtools sort -o sample.sorted.bam`

б) `samtools index sample.sam`

в) `samtools mpileup sample.sam`

г) `samtools merge sample.sam`

Ответ: а

15. Дифференциальная экспрессия генов — это:

а) Различие в количестве копий гена между организмами

б) Статистически значимое изменение уровня транскрипции гена между условиями эксперимента

в) Различие в нуклеотидной последовательности гена

г) Изменение локализации белка в клетке

Ответ: б

16. Какая программа или веб-сервис используется для поиска обогащённых мотивов в наборе последовательностей ДНК?

а) MACS2

б) MEME / HOMER

в) FastQC

г) BCFtools

Ответ: б

17. Технология Hi-C позволяет определить:

а) Уровень метилирования ДНК

б) Пространственные контакты между удалёнными участками хроматина

в) Места связывания транскрипционных факторов

г) Экспрессию генов в единичных клетках

Ответ: б

18. Аббревиатура NGS расшифровывается как:

а) New Generation Sequencing

б) Next Generation Sequencing

в) Nucleotide Genome Scanning

г) Nano Geometry Sequencing

Ответ: б

19. Какой формат документов позволяет создавать полностью воспроизводимые отчёты, объединяя код (Python/R), результаты и текст?

а) Microsoft Word

б) Jupyter Notebook / R Markdown

в) PDF

г) HTML

Ответ: б

20. Какой тип данных получают в результате RNA-seq эксперимента?

а) Последовательности всех белков клетки

б) Фрагменты кДНК, отражающие транскриптом

- в) Полную геномную ДНК
г) Метилированные CpG-островки
Ответ: б

3.2. Ситуационные задачи (краткий ответ)

1. В FASTQ-файле качество основания составляет 30. Какова вероятность ошибки секвенирования для этого основания?
Ответ: 0,001
2. В SAM-файле установлен флаг 0x10 (16). Что это означает для рида?
Ответ: обратное картирование
3. При выравнивании коротких прочтений на референсный геном получен MAPQ = 0. Что это означает?
Ответ: неоднозначное картирование
4. В VCF-файле для варианта указано QUAL = 100. Как интерпретируется это значение?
Ответ: качество варианта
5. В RNA-seq анализе после обработки DESeq2 получен $\log_2FC = 2$ и $p\text{-adj} = 0,01$. Что это означает?
Ответ: значимое повышение экспрессии
6. В ChIP-seq эксперименте FRiP = 0,05. Является ли это значение хорошим?
Ответ: да (низкое)
7. При анализе ATAC-seq данные показали пик в промоторной области гена. Что это может означать?
Ответ: открытый хроматин
8. В метагеномном образце обнаружены контиги, принадлежащие неизвестному микроорганизму. Каким методом можно восстановить его геном?
Ответ: биннинг
9. В пангеноме бактерий core-гены составляют 60% от общего числа генов. Что это говорит о виде?
Ответ: консервативный пангеном
10. При сборке генома de novo получено N50 = 100 кб. Что означает этот показатель?
Ответ: качество сборки

3.3. Открытые вопросы (краткий ответ)

1. Как называется алгоритм, лежащий в основе быстрого картирования коротких прочтений с использованием FM-индекса?
Ответ: борроу-вилеровское преобразование
2. Какой тип данных получают в результате секвенирования РНК?
Ответ: транскриптом
3. Как называется программа для обнаружения пиков в ChIP-seq данных?
Ответ: MACS2
4. Какой формат файла используется для хранения аннотированных генетических вариантов?
Ответ: VCF
5. Как называется структура данных, используемая в сборщиках генома de novo на основе k-меров?
Ответ: граф де Брёйна
6. Какой метод позволяет изучать доступность хроматина в масштабе всего генома?
Ответ: ATAC-seq
7. Как называется статистическая поправка на множественное тестирование, контролирующая долю ложных открытий?
Ответ: Бенджамини-Хохберга (FDR)

8. Как называется процесс группировки контигов в метагеномике для восстановления геномов отдельных организмов?
Ответ: биннинг
9. Какой инструмент используется для создания воспроизводимых пайплайнов на основе правил?
Ответ: Snakemake
10. Как называется метод, позволяющий изучать трёхмерную организацию хроматина?
Ответ: Hi-C

3.4. Задания для практических занятий

1. Знакомство с интерфейсом командной строки ОС Linux: навигация по файловой системе, создание/просмотр/редактирование файлов, права доступа.
2. Установка и использование менеджера пакетов Conda. Создание окружения для биоинформатического анализа, установка пакетов samtools, bedtools, seqtk.
3. Изучение основных форматов геномных данных: FASTA, FASTQ, SAM/BAM, VCF, BED/GFF. Написание скрипта на Bash для подсчёта GC-состава в последовательностях FASTA-файла.
4. Контроль качества исходных прочтений: запуск FastQC на тестовых FASTQ-файлах, анализ отчётов. Создание объединённого отчёта с помощью MultiQC.
5. Тримминг прочтений: удаление адаптеров и низкокачественных концевых нуклеотидов с использованием Trimmomatic или cutadapt.
6. Построение индекса референсного генома с помощью bwa index. Картирование парно-концевых ридов на референс с помощью bwa mem, конвейерная обработка с сортировкой samtools sort.
7. Визуализация выравнивания в IGV. Навигация по геному, просмотр покрытия и отдельных ридов.
8. Подготовка BAM-файлов к вариантому анализу: добавление read group, удаление ПЦР-дубликатов.
9. Коллинг вариантов с помощью bcftools mpileup и bcftools call. Фильтрация вариантов по качеству, глубине покрытия.
10. Аннотация вариантов с помощью SnpEff или Ensembl VEP.
11. Анализ дифференциальной экспрессии генов по данным RNA-seq с использованием DESeq2. Построение volcano plot и тепловой карты.
12. Пик-коллинг ChIP-seq/ATAC-seq с помощью MACS2, визуализация профилей покрытия в IGV.
13. Мотивный анализ с помощью HOMER или MEME-Suite.
14. Метагеномное профилирование с использованием Kraken2.
15. Построение филогенетического дерева с помощью IQ-TREE.

3.5. Вопросы для опроса

1. Опишите структуру FASTQ-записи. Что обозначает символ качества Q30?
2. Какую информацию содержит SAM-флаг 0x10 (16)? Приведите примеры флагов для парно-концевого чтения, картированного в правильной ориентации.
3. В чём разница между форматами BED и GFF? Где они применяются?
4. Зачем нужна индексация FASTA-файла? Какие файлы создаёт samtools faidx?
5. Каким образом вы можете быстро проверить количество ридов в FASTQ-файле, не загружая его целиком в текстовый редактор?
6. Какие типы артефактов выявляет FastQC? О чём свидетельствует неравномерный GC-состав в ридовых парках?
7. В чём разница между жёстким (hard) и мягким (soft) триммингом? Когда предпочтителен каждый из них?

8. Объясните принцип работы алгоритма BWA-MEM. В каких случаях лучше использовать BWA-ALN?
9. Что означает значение MAPQ в поле SAM? Как оно вычисляется?
10. Как с помощью samtools отфильтровать BAM-файл, оставив только уникально картированные риды ($MAPQ \geq 30$)?
11. Зачем необходима информация о read group в BAM-файле для коллинга вариантов?
12. Какие параметры качества VCF-записи вы знаете? Что отражает QUAL и INFO-поле DP?
13. В чём различие между герминативным и соматическим коллингом вариантов?
14. Как SnpEff предсказывает эффект варианта? Какие транскрипты используются для аннотации?
15. Опишите алгоритмический подход bcftools mpileup для определения вероятности генотипа в позиции.
16. Почему для картирования RNA-seq ридов используют алгоритмы, учитывающие сплайсинг (STAR, HISAT2), а не обычные выравниватели типа BWA?
17. В чём разница между RPKM, FPKM и TPM? Какую метрику предпочтительнее использовать и почему?
18. Объясните, как DESeq2 моделирует дисперсию биологических повторностей. Зачем нужна оценка размера факторов (size factors)?
19. Что такое поправка на множественное тестирование и почему она критична при анализе тысяч генов?
20. Какие биологические причины могут привести к обнаружению высокого fold change при низкой экспрессии гена? Как интерпретировать такие результаты?
21. Для чего в ChIP-seq эксперименте нужен контрольный образец (Input)? Как он используется в MACS2?
22. Чем отличаются профили пиков гистоновых модификаций широкого типа (broad marks) от пиков транскрипционных факторов (narrow marks)? Какие параметры MACS2 следует менять?
23. Как метрика FRiP (Fraction of Reads in Peaks) характеризует качество ChIP-seq эксперимента?
24. Какие статистические модели лежат в основе пик-коллера MACS2?
25. Почему при ATAC-seq анализе важно учитывать сдвиг ридов, связанный с транспозазой Tn5? Как это реализовано в параметрах MACS2?
26. Что такое контиг в сборке генома?
27. Какова роль графа де Брёйна в сборке de novo?
28. Какие форматы геномных данных вы знаете и для чего каждый используется?
29. Что такое метагеномный биннинг и для чего он применяется?
30. Как оценивается качество сборки генома? Назовите основные метрики.
31. Что такое пангеном и как он характеризует вид?
32. Какие методы используются для поиска горизонтального переноса генов?
33. Что такое синтения и как она выявляется?
34. Как проводится филогенетический анализ по геномным данным?
35. Какие базы данных используются для функциональной аннотации генов?
36. Что такое мотивный анализ и для чего он нужен?
37. Как анализируются данные полногеномного бисульфитного секвенирования?
38. Что такое топологически ассоциированные домены (TAD) и как их выявляют?
39. Какие методы машинного обучения применяются в геномике?
40. Что такое воспроизводимость научных исследований и как она обеспечивается в биоинформатике?

3.6. Темы рефератов

1. Эволюция методов секвенирования: от Сэнгера до одномолекулярного секвенирования в реальном времени.
2. Сравнительный анализ алгоритмов картирования коротких прочтений (BWA, Bowtie2, STAR): принципы, ограничения и области применения.
3. Математические основы динамического программирования в задачах парного и множественного выравнивания биологических последовательностей.
4. Скрытые марковские модели в аннотации генов и предсказании кодирующих областей.
5. Графы де Брёйна как фундаментальная структура данных в сборке геномов de novo.
6. Проблема поиска структурных вариантов генома по данным коротких и длинных прочтений: сравнительный обзор методов.
7. Воспроизводимость и переносимость биоинформатических пайплайнов: обзор систем управления workflow (Snakemake, Nextflow, CWL).
8. Статистические подходы к анализу дифференциальной экспрессии генов: от микрочипов до RNA-seq (DESeq2, edgeR, limma-voom).
9. Применение методов машинного обучения для поиска патогенных мутаций и предсказания их функциональной значимости.
10. Метагеномика как инструмент оценки биоразнообразия: от маркерных генов до полногеномной сборки и биннинга.
11. Пангеномика прокариот: концепция, методы построения и анализ профиля генов устойчивости к антибиотикам.
12. Филогеномика и коалесцентные модели: разрешение конфликтов генного и видового деревьев.
13. Интегративный анализ данных ChIP-seq и RNA-seq для реконструкции транскрипционных регуляторных сетей.
14. Современные вычислительные подходы к анализу трёхмерной организации генома (Hi-C и производные методы).
15. Пространственная транскриптомика: технологии, вычислительные методы и перспективы для системной биологии.
16. Эпигенетические часы старения: биоинформатические алгоритмы и их применение в геронтологии.
17. Глубокое обучение в геномике: от предсказания функциональных элементов до классификации изображений хроматина.
18. Этические и правовые аспекты хранения и анализа полногеномных данных человека в эпоху персонализированной медицины.
19. Сравнение форматов CRAM и BAM: преимущества и недостатки.
20. Высокопроизводительные вычисления в биоинформатике: HPC-кластеры и облачные платформы.

3.7. Темы докладов

1. Принципы организации международных нуклеотидных банков данных: от INSDC до специализированных ресурсов (EGA, dbGaP).
2. Воспроизводимый биоинформатический анализ: как контейнеризация (Docker, Singularity) меняет научную практику.
3. Сравнение языков программирования в биоинформатике: Bash, Python и R — выбор инструмента под задачу.
4. Алгоритмы динамического программирования в действии: глобальное и локальное выравнивание (Needleman-Wunsch, Smith-Waterman).
5. Индексация генома и её влияние на скорость картирования: FM-индекс и его реализация в Bowtie/BWA.
6. Эволюция форматов геномных данных: от FASTA до CRAM — зачем биоинформатикам эффективное сжатие.

7. Контроль качества NGS-данных: от FastQC до MultiQC — как автоматизировать рутинный анализ.
8. Секвенирование единичных клеток: вычислительные проблемы нормализации, кластеризации и траекторного анализа.
9. Поиск генетических вариантов: от BCFtools к GATK — эволюция лучших практик коллинга SNP и Indel.
10. Аннотация вариантов: Ensembl VEP, SnpEff, ANNOVAR — сравнение подходов и интерпретация клинической значимости.
11. Сборка транскриптома de novo: алгоритм Trinity — от графа к изоформам.
12. Дифференциальный анализ экспрессии: почему DESeq2 и edgeR стали стандартом, а t-тест — нет.
13. Пангеномы бактерий: как core- и accessory-гены формируют экологическую пластичность патогенов.
14. Метагеномная сборка и биннинг: восстановление геномов некультивируемых микроорганизмов из метагеномов.
15. Филогеномика на практике: конкатенация vs. супердерева — как выбрать метод для реконструкции эволюционной истории.
16. Горизонтальный перенос генов: биоинформатические методы обнаружения и его роль в распространении антибиотикорезистентности.
17. ChIP-seq анализ: от пик-коллинга с MACS2 до функциональной аннотации и мотивного анализа.
18. Анализ открытого хроматина: ATAC-seq — принцип метода, обработка данных и интерпретация.
19. Полногеномное бисульфитное секвенирование (WGBS): вычислительные подходы к картированию и оценке метилирования.
20. Hi-C и трёхмерная организация генома: как контактные карты превращаются в топологически ассоциированные домены (TAD).
21. Пространственная транскриптомика: технологии Visium и MERFISH — вызовы для вычислительного анализа.
22. Глубокое обучение в геномике: свёрточные и рекуррентные сети для предсказания функциональных элементов.
23. Системы управления workflow в биоинформатике: Snakemake и Nextflow — преимущества декларативного подхода.
24. Интеграция мультиомиксных данных: стратегии объединения транскриптомики, протеомики и метаболомики для системного понимания клетки.
25. Редактирование генома и биоинформатика: дизайн гидовых РНК для CRISPR-Cas9 и предсказание нецелевых эффектов.

3.8. Темы презентаций

1. Принципы организации международных нуклеотидных банков данных.
2. Воспроизводимый биоинформатический анализ: Docker и Singularity.
3. Сравнение языков программирования в биоинформатике: Bash, Python, R.
4. Алгоритмы динамического программирования: Needleman-Wunsch и Smith-Waterman.
5. FM-индекс и его реализация в Bowtie/BWA.
6. Эволюция форматов геномных данных: от FASTA до CRAM.
7. Контроль качества NGS-данных: FastQC и MultiQC.
8. Секвенирование единичных клеток: вычислительные проблемы.
9. Поиск генетических вариантов: от BCFtools к GATK.
10. Аннотация вариантов: Ensembl VEP, SnpEff, ANNOVAR.
11. Сборка транскриптома de novo: алгоритм Trinity.
12. Дифференциальный анализ экспрессии: DESeq2 и edgeR.
13. Пангеномы бактерий: core и accessory гены.

14. Метагеномная сборка и биннинг.
15. Филогеномика: конкатенация vs. супердеревья.
16. Горизонтальный перенос генов: методы обнаружения.
17. ChIP-seq анализ: пик-коллинг и мотивный анализ.
18. ATAC-seq: анализ открытого хроматина.
19. WGBS: полногеномное бисульфитное секвенирование.
20. Hi-C: трёхмерная организация генома.
21. Пространственная транскриптомика: Visium и MERFISH.
22. Глубокое обучение в геномике.
23. Системы управления workflow: Snakemake и Nextflow.
24. Интеграция мультиомиксных данных.
25. Редактирование генома и биоинформатика.

3.9. Темы курсовых работ

1. Сравнительная сборка и аннотация бактериального генома из коротких прочтений.
2. Поиск и функциональная аннотация однонуклеотидных полиморфизмов (SNP) у модельного организма.
3. Анализ дифференциальной экспрессии генов по данным RNA-seq при ответе на стрессовое воздействие.
4. Реконструкция и визуализация генных регуляторных сетей на основе транскриптомных данных.
5. Идентификация и характеристика пиков связывания транскрипционного фактора по данным ChIP-seq.
6. Анализ доступности хроматина методом ATAC-seq и её связь с экспрессией генов.
7. Метагеномное профилирование таксономического состава микробного сообщества по данным 16S рРНК.
8. Восстановление геномов (MAGs) из метагеномныхборок и их функциональный потенциал.
9. Филогеномный анализ группы бактерий на основе однокопийных ортологов.
10. Исследование пангенома патогенных штаммов: ядро и акцессорные гены.
11. Выявление структурных вариантов (SV) в геноме человека по данным длинных прочтений.
12. Анализ альтернативного сплайсинга по данным RNA-seq при развитии заболевания.
13. Эпигенетический анализ полногеномного бисульфитного секвенирования (WGBS).
14. Интеграция данных ChIP-seq (H3K27ac) и RNA-seq для идентификации активных энхансеров.
15. Исследование трёхмерной организации хроматина (Hi-C) в локусе интереса.
16. Разработка воспроизводимого пайплайна для анализа RNA-seq с использованием Snakemake/Nextflow.
17. Сравнение алгоритмов нормализации и кластеризации для single-cell RNA-seq данных.
18. Оценка качества референсных геномов разных версий с помощью BUSCO и QUAST.
19. Прогнозирование патогенности генетических вариантов методами машинного обучения.
20. Биоинформатический поиск и аннотация мобильных генетических элементов в геномах прокариот.

3.10. Вопросы к экзамену

1. Основные форматы геномных данных (FASTA, FASTQ, SAM/BAM, VCF, BED, GFF/GTF): структура, назначение, примеры использования.
2. Принцип кодирования качества оснований в формате FASTQ. Шкала Phred: определение, расчёт вероятности ошибки.

3. Парное и множественное выравнивание: постановка задачи, матрицы замен (PAM, BLOSUM), штрафы за открытие и продолжение гэта.
4. Алгоритм динамического программирования Нидлмана–Вунша (глобальное выравнивание) и Смита–Ватермана (локальное выравнивание): схемы, временная и пространственная сложность.
5. Скрытые марковские модели (HMM): определение, элементы, применение в профильном анализе белковых семейств и предсказании генов.
6. Принцип борроу-вилеровского преобразования (BWT) и FM-индекса. Как они используются для быстрого картирования коротких прочтений в программах Bowtie и BWA.
7. Графы де Брёйна в геномной сборке: представление k-меров, построение графа, разрешение повторов, эйлеров обход.
8. Сравнение стратегий сборки генома: OLC (Overlap-Layout-Consensus) против графов де Брёйна; преимущества и недостатки при работе с короткими и длинными прочтениями.
9. Контроль качества данных NGS: ключевые метрики FastQC, типы артефактов, способы фильтрации и тримминга.
10. Статистические основы выравнивания: E-value и bit score в BLAST; учёт множественного тестирования в геномном поиске (поправка Бонферрони, метод Бенджамини–Хохберга, FDR).
11. Основные этапы пайплайна обработки данных ДНК-секвенирования: от FASTQ до аннотированного списка вариантов.
12. Выравниватели коротких прочтений (BWA-MEM, Bowtie2): алгоритмические основы, понятие сита, настройка чувствительности и специфичности.
13. SAM/BAM формат: поля FLAG, CIGAR, MAPQ; их интерпретация. Генерация и сортировка BAM-файлов с помощью SAMtools.
14. Поиск генетических вариантов (SNP/Indel): вероятностная модель BCFtools/GATK, понятие prior и posterior вероятности генотипа, фильтрация вариантов по качественным метрикам.
15. Аннотация генетических вариантов: функциональные классы (миссенс, нонсенс, синонимичные, сплайсинговые и др.), инструменты SnpEff и VEP, базы данных популяционных частот (gnomAD).
16. Картирование RNA-seq данных: необходимость учёта сплайсинга; алгоритмы STAR и HISAT2, построение индекса с учётом экзон-интронных границ.
17. Количественная оценка экспрессии: методы подсчёта ридов (featureCounts, HTSeq-count) и методы псевдовыравнивания (Salmon, Kallisto). Сравнение TPM и FPKM/RPKM.
18. Статистический анализ дифференциальной экспрессии: отрицательное биномиальное распределение в DESeq2, модель edgeR, квази-правдоподобие, понятие log2 fold change.
19. Визуализация дифференциальной экспрессии: MA-plot, volcano plot, тепловые карты, PCA-график; их биологическая интерпретация.
20. Сборка транскриптома de novo (Trinity): этапы работы, преимущества и ограничения по сравнению с референс-зависимым анализом.
21. Понятие ортологии и паралолии; методы предсказания ортологов (OrthoFinder, OrthoMCL, InParanoid); значение для аннотации генов.
22. Филогенетический анализ: дистанционные методы (NJ, UPGMA), методы максимальной экономии, максимального правдоподобия и байесовского вывода.
23. Множественное выравнивание целых геномов и синтения: алгоритмы сравнения (MUMmer, LASTZ), выявление хромосомных перестроек.
24. Поиск участков, находящихся под действием отбора: методы на основе нуклеотидных замен (dN/dS), консервативных элементов (phastCons, phyloP).

25. Пангеномика прокариот: концепция открытого и закрытого пангенома, построение пангенома (Roary, Panaroo), анализ генов резистентности.
26. Метагеномика: ампликонное секвенирование (16S rPHK) и полногеномный метагеномный анализ; таксономическое профилирование (Kraken, MetaPhlAn).
27. Восстановление геномов из метагеномов (MAGs): дифференциальный биннинг по тетрапуклеотидным частотам и покрытию (MetaBAT, CONCOCT), оценка качества с CheckM.
28. Горизонтальный перенос генов: биоинформатические методы обнаружения (аномалии нуклеотидного состава, филогенетические несоответствия), его роль в эволюции.
29. Молекулярные часы и датировка филогений: релаксированные часы, калибровка, программные реализации (BEAST).
30. Сравнительная геномика эукариот: особенности выравнивания, учёт дубликаций, потеря и приобретение семейств генов.
31. Функциональная аннотация генов: базы данных Gene Ontology, KEGG, Reactome; статистический анализ перепредставленности (ORA) и GSEA.
32. ChIP-seq анализ: цели, подготовка библиотеки, контроль Input; принцип пик-коллинга MACS2 (модель сдвига фрагментов, пуассоновское распределение).
33. Контроль качества ChIP-seq: FRiP, NSC, RSC, перекрестная корреляция; как отличить успешный эксперимент от неудачного.
34. Мотивный анализ: поиск консенсусных последовательностей, позиционные весовые матрицы (PWM), алгоритмы HOMER и MEME, понятие обогащения.
35. Анализ ATAC-seq: принцип метода, особенности выравнивания и пик-коллинга (Tn5-сдвиг), интерпретация профилей доступности хроматина.
36. Полногеномное бисульфитное секвенирование (WGBS): конверсия цитозина, картирование (Bismark, bwa-meth), метилирование в CpG-контексте, поиск DMR.
37. Конформационная геномика (Hi-C): получение контактных карт, нормализация, идентификация топологически ассоциированных доменов (TAD), петель и компартментов.
38. Интеграция омиксных данных: стратегии объединения RNA-seq, ChIP-seq и WGBS для реконструкции регуляторных сетей; программные средства (RegulomeDB, ChIPseeker).
39. Глубокое обучение в геномике: основные архитектуры (CNN, RNN), примеры применения (DeepVariant, Basenji, Enformer); преимущества и проблемы интерпретации.
40. Пространственная транскриптомика и single-cell мультиомика: вычислительные подходы к интеграции (Seurat v3, Harmony, LIGER), методы выявления пространственно-вариабельных генов.

4. Методические материалы, определяющие процедуры оценивания

4.1. Общие положения

Текущий контроль и промежуточная аттестация по дисциплине «Биоинформационные подходы в геномике» проводятся в соответствии с балльно-рейтинговой системой Университета. Цель оценочных процедур – проверка уровня сформированности компетенций **ОПК-2, ОПК-5, ОПК-6, СПК-2** на различных этапах обучения.

4.2. Формы текущего контроля и их балльная оценка

Форма контроля	Количество за семестр	Максимальный балл за единицу	Максимальный балл за семестр
Опрос	4	5	20
Тестирование	2	10	20
Реферат	1	20	20

Форма контроля	Количество за семестр	Максимальный балл за единицу	Максимальный балл за семестр
Доклад	1	5	5
Презентация	1	5	5
Итого за текущий контроль			70

4.3. Процедура проведения экзамена

Экзамен проводится в период сессии в устной форме. Билет содержит два теоретических вопроса и одно практическое задание (написание команды или фрагмента скрипта). Время на подготовку – 40 минут. Ответ должен быть полным, аргументированным, с использованием биоинформатической терминологии.

Шкала оценивания экзамена

Показатель	Баллы
Отлично (высокий уровень). Студент демонстрирует глубокое, систематическое знание алгоритмических основ биоинформатики, архитектуры баз данных, форматов файлов и статистических принципов анализа NGS. Ответы на теоретические вопросы полные, точные, иллюстрируются схемами алгоритмов (динамическое программирование, BWT, графы де Брёйна), сравнением методов (BWA vs Bowtie, GATK vs BCFtools). Практическое задание выполнено безупречно: скрипт или пайплайн корректен, эффективен, оформлен с комментариями, учтены все этапы обработки данных; выбор инструментов обоснован, интерпретация результатов глубока и биологически содержательна. Студент свободно обсуждает ограничения алгоритмов, предлагает модификации, отвечает на дополнительные вопросы о воспроизводимости и масштабировании.	25–30
Хорошо (базовый уровень). Студент хорошо знает ключевые темы: форматы (FASTQ, SAM/BAM, VCF), основные алгоритмы картирования, сборки, коллинга вариантов и анализа RNA-seq. Ответы на теоретические вопросы достаточно полны, но могут содержать незначительные неточности в деталях (например, параметры скрытых марковских моделей, особенности работы конкретного выравнивателя). Практическое задание выполнено в целом правильно: код работает, основные шаги реализованы, однако возможны небольшие недочёты в стиле оформления, неполная обработка ошибок или недостаточно оптимальная визуализация. Студент отвечает на большинство дополнительных вопросов, понимает связь между этапами пайплайна.	19–24
Удовлетворительно (пороговый уровень). Студент владеет материалом фрагментарно: может описать общие принципы NGS-анализа, но путается в конкретных алгоритмах или форматах. Теоретические ответы поверхностны, требуют наводящих вопросов; допускаются ошибки в интерпретации статистических метрик (E-value, MAPQ, FDR). Практическое задание выполнено с	13–18

Показатель	Баллы
недочётами: скрипт содержит ошибки или реализована только часть пайплайна, нет чёткой интерпретации результатов. Студент может объяснить отдельные шаги, но не демонстрирует целостного понимания процесса.	
Неудовлетворительно (низкий уровень). Знания отрывочны, содержат грубые ошибки. Студент не может объяснить ключевые концепции (динамическое программирование, BWT, граф де Брёйна, отличие SAM от BAM). Практическое задание либо не выполнено, либо содержит принципиальные ошибки (неверный выбор инструмента, синтаксические ошибки, ведущие к неработоспособности кода). Студент не ориентируется в базовой терминологии, не способен связать теорию с практикой.	6–12
Крайне неудовлетворительно (отсутствие знаний). Студент не может ответить ни на один теоретический вопрос; практическое задание не выполнено. Отсутствует понимание даже элементарных основ работы с данными NGS и инструментарием биоинформатики.	0–5

Максимальное количество баллов на экзамене – 30.

Итоговая шкала оценивания по дисциплине

Сумма баллов	Оценка
81–100	отлично
61–80	хорошо
41–60	удовлетворительно
0–40	неудовлетворительно

4.4. Порядок ликвидации академической задолженности

Студенты, набравшие менее 11 баллов текущего контроля или не сдавшие экзамен, проходят пересдачу в порядке, установленном локальными актами Университета. Пересдача включает выполнение дополнительных заданий (повторное тестирование, подготовка нового реферата, собеседование). Курсовая работа, оценённая на «неудовлетворительно», подлежит доработке и повторной защите в установленные сроки.

Ключи ко всем заданиям приведены непосредственно в тексте (для тестов, ситуационных задач, открытых вопросов). Для теоретических вопросов (опрос, экзамен) критерии оценивания указаны в разделе 2; преподаватель оценивает полноту и правильность ответа.