

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Наумова Наталия Александровна
Должность: Ректор
Дата подписания: 08.07.2026 10:50:03
Уникальный программный государственный идентификатор:
6b5279da4e034bffa79172803da5b710c0f809da

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

Согласовано
и.о. декана факультета естественных наук
« 14 » 04 2026 г.
/Лялина И.Ю./

Рабочая программа дисциплины

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

Согласовано учебно-методической
комиссией факультета естественных наук
Протокол « 14 » 04 2026 г. № 6
Председатель УМКом /Лялина И.Ю./

Рекомендовано кафедрой общей биологии
и биоэкологии
Протокол от « 06 » 03 2026 г. № 8
Зав. кафедрой /Гордеев М.И./

Москва
2026

Авторы-составители:

Гордеев Михаил Иванович, доктор биологических наук, профессор
Москаев Антон Вячеславович, кандидат биологических наук, доцент
Темников Андрей Андреевич, старший преподаватель

Рабочая программа дисциплины «Биоинформационные подходы в геномике» составлена в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки 06.04.01 Биология, утвержденного приказом МИНОБРНАУКИ РОССИИ от 11.08.2020г., № 934

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

Год начала подготовки (по учебному плану) 2026

Содержание

1.	ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ.....	4
1.1.	Цель и задачи дисциплины.....	4
1.2.	Задачи дисциплины:.....	4
1.3.	Планируемые результаты обучения	4
2.	МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.....	4
3.	ОБЪЕМ И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	6
3.1	Объем дисциплины	6
3.2	Содержание дисциплины.....	6
4.	УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ.....	7
5.	ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ	14
5.1	Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы.....	15
5.2	Описание показателей и критериев оценивания компетенций на различных этапах формирования, описание шкал оценивания	15
	Шкала оценивания опроса.....	22
	Шкала оценивания тестирования	22
	Шкала оценивания реферата	22
	Шкала оценивания доклада.....	23
	Шкала оценивания презентации.....	23
	Шкала оценивания курсовой работы	23
5.3	Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы.....	23
	Примерные темы практических работ	24
	Примерные темы рефератов.....	27
	Примерные темы докладов и презентаций	27
	Примерные тестовые задания	29
	Примерный перечень тематик курсовых работ.....	33
	Примерный перечень вопросов к экзамену	35
5.4	Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.....	37
5.5	Итоговая шкала выставления оценки по дисциплине	38
1.4.		39
5.6	Итоговая шкала выставления оценки по курсовой работе	40
6.	УЧЕБНО-МЕТОДИЧЕСКОЕ И РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ	40
6.1	Основная литература	41
6.2	Дополнительная литература.....	41
6.3	Ресурсы информационно-телекоммуникационной сети «Интернет»	41
7.	МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ.....	42
8.	ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ	43
9.	МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	45

1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

1.1. Цель и задачи дисциплины

Цель освоения дисциплины: формирование у магистрантов системного мышления и практических навыков в области анализа полногеномных данных, основанных на понимании алгоритмических принципов биоинформатики, для решения фундаментальных и прикладных задач современной биологии.

1.2. Задачи дисциплины:

- формирование целостного представления о методологии управления проектами, в том числе методическими основами рыночного подхода к системе экономики планирования реализации проектов, методами анализа и синтеза управленческих решений, основанных на идеях достижения максимального результата в условиях ограниченности имеющихся ресурсов и способов повышения рентабельности;
- изучение формы и принципы организации научно-исследовательской деятельности;
- приобретение знания по вопросам планирования, организации и проведения научного проекта;
- формирование способности работы с основными источниками финансирования научной проектной деятельности;
- рассмотрение актуальные проблемы развития биологической науки и определить критерии выбора направления научного исследования.

1.3. Планируемые результаты обучения

В результате освоения данной дисциплины у обучающихся формируются следующие компетенции:

ОПК-2 . Способен творчески использовать в профессиональной деятельности знания фундаментальных и прикладных разделов дисциплин (модулей), определяющих направленность программы магистратуры

ОПК-5. Способен участвовать в создании и реализации новых технологий в сфере профессиональной деятельности и контроле их экологической безопасности с использованием живых объектов

ОПК-6. Способен творчески применять и модифицировать современные компьютерные технологии, работать с профессиональными базами данных, профессионально оформлять и представлять результаты новых разработок

СПК-2.Способен использовать программы для эпигенетического анализа и качества эпигенетических данных

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

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

Учебная дисциплина «Биоинформационные подходы в геномике» опирается на знания, умения и виды деятельности, формируемые в процессе изучения дисциплин: «История и методология биологии», «Управление проектами в научной сфере», «Учение о биосфере, глобальные экологические проблемы».

Дисциплина «Биоинформационные подходы в геномике» является методологической основой для исследовательской и теоретической работы в рамках подготовки магистерской диссертации.

3. ОБЪЕМ И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

3.1 Объем дисциплины

Показатель объема дисциплины	Форма обучения Очная
Объем дисциплины в зачетных единицах	4
Объем дисциплины в часах	144
Контактная работа:	16,6
Лекции	4
Практические занятия	10
Контактные часы на промежуточную аттестацию:	2,3
Предэкзаменационная консультация	2
Экзамен	0,3
Самостоятельная работа	100
Контроль	27,4

Форма промежуточной аттестации- экзамен и курсовая работа в 1 семестре

3.2 Содержание дисциплины

Наименование разделов (тем) Дисциплины с кратким содержанием	Количество часов	
	Лекции	Практические занятия
Тема 1. Вычислительная среда и алгоритмический фундамент геномики Основное содержание: - Инфраструктура данных: архитектура международных банков данных (INSDC, Ensembl, UCSC Genome Browser), форматы хранения геномной информации (FASTA, FASTQ, SAM/BAM, VCF, BED). Принципы курирования и версионирования референсных геномов. - Вычислительный инструментарий: UNIX/Linux как основная среда биоинформатика (файловая система, конвейеры, управление задачами). Основы работы на HPC-кластерах и в облачных средах. - Алгоритмическая база: динамическое программирование для парного и множественного выравнивания, скрытые марковские модели (HMM) для профильного поиска, алгоритмы на графах и их реализация в геномных сборщиках (графы де Брёйна, OLC). - Статистический аппарат: теория вероятностей в биоинформатике (P-value, E-value, поправка на множественное тестирование), байесовские подходы, основы машинного обучения (классификация, кластеризация) в применении к биологическим последовательностям.	1	4
Тема 2. Анализ индивидуальных геномов и транскриптомов Основное содержание: - Препроцессинг данных NGS: контроль качества (FastQC), фильтрация и тримминг ридов (Trimmomatic, cutadapt), стратегии борьбы с техническими артефактами. - Картирование и сборка: алгоритмы быстрого картирования коротких	1	2

ридов (BWA, Bowtie2) и их индексация. Сборка геномов <i>de novo</i> и референс-ориентированная сборка. Визуализация выравниваний (IGV). - Поиск генетических вариантов: детекция однонуклеотидных полиморфизмов (SNP) и коротких инсерций/делеций (Indel) с помощью GATK и bcftools. Аннотация вариантов (VEP, SnpEff) и предсказание их функционального эффекта. - Анализ транскриптома (RNA-seq): алгоритмы картирования сплайсированных ридов (STAR, HISAT2). Количественная оценка экспрессии генов и транскриптов (featureCounts, Salmon). Статистический поиск дифференциально экспрессируемых генов (DESeq2, edgeR).		
Тема 3. Сравнительная и эволюционная геномика. Основное содержание: - Сравнительный анализ на уровне геномов: идентификация ортологов и паралогов, построение филогенетических деревьев (методы максимального правдоподобия и байесовского вывода). Понятие синтении и хромосомных перестроек. - Поиск консервативных элементов: методы выявления участков, находящихся под действием отбора (phastCons, phyloP), предсказание регуляторных районов. - Пангеномика: концепция пангенома (ядро и аксессуарный геном) у прокариот и эукариот, методы построения и анализа пангеномов. - Метагеномика и таксономическое профилирование: маркерные гены (16S rRNA, ITS) и полногеномный метагеномный анализ (metagenome assembly). Алгоритмы биннинга и восстановления геномов (MAGs).	1	2
Тема 4. Интегративная геномика и системная биология Основное содержание: - Функциональная аннотация и обогащение: базы данных онтологий (Gene Ontology) и сигнальных путей (KEGG, Reactome). Статистический анализ перепредставленности (ORA) и методы анализа функциональных классов (GSEA). - Эпигеномика и структура хроматина: анализ данных ChIP-seq (пик-коллинг, мотивный анализ), ATAC-seq, Hi-C. Модели трёхмерной организации генома и её связь с регуляцией транскрипции. - Построение генных регуляторных сетей: методы обратной инженерии сетей из транскриптомных данных (алгоритмы на основе корреляции, взаимной информации, деревьев решений). Интеграция с данными по транскрипционным факторам. - Современные вызовы и рубежи: пространственная транскриптомика, интегративный анализ мультиомиксных данных на уровне единичной клетки (single-cell multiomics), дизайн масштабного вычислительного эксперимента и воспроизводимость научных исследований.	1	2
Итого:	4	10

4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Темы для самостоятельного изучения	Изучаемые вопросы	Количество часов	Формы самостоятельной работы	Методическое обеспечение	Формы отчетности
Тема 1. Вычислительная среда и алгоритмический фундамент геномики	Контейнеризация и воспроизводимость. Принципы работы Docker и Singularity. Создание и запуск готовых биоинформатических образов (biocontainers). Почему контейнеры стали стандартом воспроизводимости в геномике. Менеджеры пайплайнов. Обзор Snakemake и Nextflow: синтаксис, понятие правил и каналов, примеры простых геномных пайплайнов. Преимущества перед линейными скриптами. Современные форматы сжатия. Архитектура формата CRAM (сжатие с учётом референса). Сравнение экономии места с BAM. Практическая работа с утилитами конвертации. Графовые геномные сборщики: детальный разбор. Отличие OLC- и графово-де Брёйновых подходов. Устройство SPAdes и Canu: учёт ошибок, разрешение повторов. Статистические основы	20	Анализ литературных источников, конспектирование Подготовка к опросу	Основная и рекомендуемая учебная и научная литература Интернет-ресурсы	Опрос

	дифференциального анализа. Теоретические модели распределений: отрицательное биномиальное в DESeq2, квази-правдоподобие в edgeR. Почему нельзя использовать t-тест для RNA-seq. Этические и правовые рамки. GDPR, работа с персональными генетическими данными, Европейская и национальные политики депонирования геномной информации человека в контролируемые базы (EGA, dbGaP).				
Тема 2. Анализ индивидуальных геномов и транскриптомов	Выравнивание длинных прочтений (PacBio, ONT). Алгоритмы minimap2 и NGMLR: принцип работы с высокой ошибкоориентированностью. Цепочки конкатенации и зародышевый подход к выравниванию. Выявление структурных вариантов (SV). Классификация методов: на основе парных концов, расщеплённых ридов и глубины покрытия (DELLY, Manta, LUMPY). Сложности при поиске больших инверсий и транслокаций. Сборка транскриптома de	30	Анализ литературных источников, конспектирование Подготовка докладов	Основная и рекомендуемая учебная и научная литература Интернет-ресурсы	Доклад

	novo. Алгоритм Trinity: стадии Inchworm, Chrysalis, Butterfly. Альтернативный подход Oases. Сравнение с референс-зависимым анализом. Анализ альтернативного сплайсинга. Подходы SUPPA и rMATS: оценка процента включённого экзона, статистическая рамка для сравнения изоформ между условиями. Введение в single-cell RNA-seq. Нормализация, снижение размерности (PCA, t-SNE, UMAP) и кластеризация в пакетах Seurat и Scanpy. Практическое задание: воспроизведение части анализа по готовой виньетке. Оценка качества сборки генома. Инструменты QUAST и BUSCO: метрики целостности, контиговости, доли однокопийных ортологов. Интерпретация отчётов.				
--	--	--	--	--	--

Тема 3. Сравнительная и эволюционная геномика.	Филогеномика: от генов к полным геномам. Методы конкатенации и супердеревьев (ASTRAL). Проблема неполного сортирования линий и её решение с помощью коалесцентных моделей. Поиск горизонтального переноса генов (ГПГ). Методы на основе нуклеотидного состава (Alien Hunter), сравнения филогений (TreePuzzle) и несоответствия топологий. Практический анализ в контексте бактериального пангенома. Предсказание генов <i>ab initio</i>. Скрытые марковские модели в Augustus и GeneMark-ES. Автоматическое обучение на геномных данных без референса. Понятие «кодирующего потенциала». Метагеномика вирусов. Построение вирусных метагеномных сборок (VIBRANT, VirSorter). Идентификация профагов в бактериальных геномах. Микросателлиты и минисателлиты. Алгоритмы поиска tandemных повторов (Tandem Repeats	20	Анализ литературных источников, конспектирование Подготовка докладов , подготовка к опросу	Основная и рекомендуемая учебная и научная литература Интернет-ресурсы	Опрос
--	---	----	--	--	-------

	Finder), их эволюционная динамика и использование в популяционной генетике и ДНК-паспортизации. Молекулярные часы и калибровка. Релаксированные часы (BEAST), методы палеонтологической калибровки и проблемы датировки вирусных филогений.				
Тема 4. Интегративная геномика и системная биология	Полногеномное бисульфитное секвенирование (WGBS). Особенности картирования бисульфитно-конвертированных ридов (Bismark, bwa-meth). Извлечение метилирования CpG и визуализация. Сборка генома до хромосомного уровня с помощью Hi-C. Принцип метода, алгоритмы 3D-DNA и SALSA. Оценка качества с помощью контактных карт и визуализация в Juicebox. Глубокое обучение в геномике. Нейросетевые классификаторы патогенности вариантов (CADD, DeepVariant). Принцип работы сверточных сетей на геномных признаках. Обратная инженерия регуляторных сетей. Алгоритмы GENIE3 и ARACNe: случайные леса и	30	Анализ литературных источников, конспектирование Подготовка докладов Написание реферата	Основная и рекомендуемая учебная и научная литература Интернет-ресурсы	Реферат

	взаимная информация для поиска транскрипционных взаимодействий. Интеграция с данными ChIP-seq. Протеогеномика. Использование протеомных данных для улучшения аннотации генов. Понятие протеогеномной базы данных и поиск вариантов пептидов, отсутствующих в референсе. Пространственная транскриптомика. Технологии 10x Visium и MERFISH. Вычислительные методы: SpatialDE для выявления пространственно-вариабельных генов, Giotto для интегрального анализа.				
--	--	--	--	--	--

Итого:		100			
---------------	--	------------	--	--	--

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

5.1 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 2. Самостоятельная работа.
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 2. Самостоятельная работа.
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 2. Самостоятельная работа.
СПК-2. Способен использовать программы для эпигенетического анализа и качества эпигенетических данных	1. Работа на учебных занятиях 2. Самостоятельная работа.

5.2 Описание показателей и критериев оценивания компетенций на различных этапах формирования, описание шкал оценивания

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
ОПК-2	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа.	знать: – фундаментальные принципы организации геномов и методы их аннотации; – алгоритмические основы базовых биоинформатических инструментов (выравнивание, динамическое программирование, скрытые марковские модели); – структуру и логику работы основных	Опрос, тестирование	Шкала оценивания опроса. Шкала оценивания тестирования.

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			международных биологических баз данных. уметь: – грамотно выбирать биоинформатический инструментальный инструментарий под типовую задачу геномного анализа; – интерпретировать форматы FASTA, FASTQ, SAM/BAM, VCF; – проводить выравнивание коротких прочтений на референсный геном и визуализировать результат.		
	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	Доклад, презентация, реферат,	Шкала оценивания доклада. Шкала оценивания презентации. Шкала оценивания реферата.

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			использовать фундаментальные знания молекулярной биологии и эволюции для интерпретации результатов полногеномных исследований; – техникой переноса аналитических подходов из смежных дисциплин (статистика, машинное обучение) в собственную научную работу.		
ОПК- 5	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа.	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	Опрос, тестирование	Шкала оценивания опроса. Шкала оценивания тестирования.

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			данных для дальнейшего анализа; – применять стандартные протоколы таксономического профилирования микробных сообществ (16S, метагеномная сборка).		
	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 метагеномных сборок для описания новых форм живых объектов; –	Доклад, презентация, реферат	Шкала оценивания доклада. Шкала оценивания презентации. Шкала оценивания реферата.

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			приёмами валидации вычислительных предсказаний экспериментальными методами.		
ОПК-6	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа.	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; –	Доклад, презентация, реферат	Шкала оценивания доклада. Шкала оценивания презентации

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			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; – навыками публичного депонирования обработанных данных и метаданных в международные архивы (NCBI, ENA) в соответствии с принятыми стандартами.		и. Шкала оценивания реферата.
СПК-2	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа.	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	Опрос, тестирование	Шкала оценивания опроса. Шкала оценивания тестирования.

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			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.		
	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 транскрипции, ассоциированных с	Доклад, презентация, реферат	Шкала оценивания доклада. Шкала оценивания презентации. Шкала оценивания реферата.

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

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 объему, так и по содержанию (хотя и соответствует теме); аргументация не на соответствующем уровне, некоторые проблемы с употреблением терминологии дисциплины	1

Максимальный балл 20 за 4 опроса

Шкала оценивания тестирования

Для оценки тестовых работ используются следующие критерии:

0-20 % правильных ответов оценивается как «неудовлетворительно» (2-балла);

30-50% - «удовлетворительно» (3-5 баллов);

60-80% - «хорошо» (6-8 баллов);

80-100% – «отлично» (8-10 баллов).

Максимальный балл – 20 за 2 тестирования

Шкала оценивания реферата

Показатель	Баллы
Содержание соответствует поставленным цели и задачам, изложение материала отличается логичностью и смысловой завершенностью, 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

Максимальный балл 20 за 1 реферат

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

Максимальный балл 5 за доклад

Шкала оценивания презентации

Критерии оценивания	Баллы
Представляемая информация систематизирована, последовательна и логически связана. Проблема раскрыта полностью. Широко использованы возможности технологии Power Point.	5
Представляемая информация в целом систематизирована, последовательна и логически связана (возможны небольшие отклонения). Проблема раскрыта. Возможны незначительные ошибки при оформлении в Power Point (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 Power Point использованы лишь частично.	1

Максимальный балл 5 за презентацию

5.3 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Примерные темы практических работ

Практическая работа № 1

Тема: Введение в вычислительную среду биоинформатика и работа с геномными форматами

Содержание занятия:

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 (индексация samtools faidx), FASTQ (структура, кодировка качества Phred), SAM/BAM (флаги, создание bam-файла из sam утилитой samtools view -bS), VCF (поля, чтение заголовка), BED/GFF (колонки, конвертация).
4. Практическое задание: написать скрипт на Bash для подсчёта GC-состава в последовательностях FASTA-файла; конвертировать тестовый SAM в сортированный BAM; извлечь статистику покрытия из BAM-файла при помощи bedtools genomecov.

Оборудование: компьютерный класс с установленной ОС Linux (Ubuntu 20.04+) или доступ к удалённому серверу по SSH, терминал, Conda (Miniconda3), предустановленные пакеты samtools, bedtools, seqtk, gcc. Тестовые файлы предоставляются преподавателем.

Примерные 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?

Практическая работа № 2

Тема: Препроцессинг и картирование данных NGS на референсный геном

Содержание занятия:

1. Контроль качества исходных прочтений: запуск FastQC на тестовых FASTQ-файлах, анализ отчётов (GC-состав, перепредставленность адаптеров, качество по позициям). Создание объединённого отчёта с помощью MultiQC.
2. Тримминг прочтений: удаление адаптеров и низкокачественных концевых нуклеотидов с использованием Trimmomatic или cutadapt. Обоснование выбора параметров (скользящее окно, минимальная длина чтения).
3. Построение индекса референсного генома (на примере E. coli или хромосомы человека) с помощью bwa index или bowtie2-build.
4. Картирование парно-концевых ридов на референс: запуск bwa mem / bowtie2, конвейерная обработка (pipe) с сортировкой samtools sort и преобразованием в BAM.
5. Визуализация выравнивания: загрузка BAM-файла и референса в IGV (Integrative Genomics Viewer). Навигация по геному, просмотр покрытия и отдельных ридов, анализ пиков.

Оборудование: компьютерный класс / сервер с ОС Linux, Conda-окружение с пакетами FastQC, MultiQC, Trimmomatic, cutadapt, bwa, bowtie2, samtools, IGV (локальная установка или через веб-старт). Тестовые FASTQ-файлы и референсный геном предоставляются.

Примерные вопросы для опроса

1. Какие типы артефактов выявляет FastQC? О чём свидетельствует неравномерный GC-состав в риде?
2. В чём разница между жёстким (hard) и мягким (soft) триммингом? Когда предпочтителен каждый из них?
3. Объясните принцип работы алгоритма BWA-MEM. В каких случаях лучше

использовать BWA-ALN?

4. Что означает значение MAPQ в поле SAM? Как оно вычисляется?
5. Как с помощью samtools отфильтровать BAM-файл, оставив только уникально картированные риды ($\text{MAPQ} \geq 30$)?

Практическая работа № 3

Тема: Поиск и аннотация генетических вариантов (SNP/Indel calling)

Содержание занятия:

1. Подготовка BAM-файлов к вариантному анализу: добавление read group информации (samtools addreplacerg), удаление ПЦР-дубликатов (picard MarkDuplicates или samtools markdup).
2. Базовый коллинг вариантов с помощью bcftools mpileup и bcftools call. Получение первичного VCF-файла.
3. Фильтрация вариантов: применение жестких фильтров по качеству (QUAL), глубине покрытия (DP), аллельному дисбалансу. Знакомство с рекомендациями GATK Best Practices для фильтрации.
4. Аннотация отфильтрованных вариантов: использование SnpEff или Ensembl VEP. Построение базы аннотаций для исследуемого организма, запуск аннотирования, интерпретация результатов (влияние на кодирующие последовательности: миссенс, нонсенс, синонимичные замены).
5. Визуализация вариантов в IGV: проверка избранных SNP/Indel вручную, оценка правдоподобия колла.

Оборудование: сервер/компьютеры с bwa, samtools, bcftools, picard, SnpEff (с предустановленной базой для генома, напр. GRCh38 или E. coli), IGV. Тестовые BAM-файлы (например, результаты предыдущего занятия) и референсный геном.

Примерные вопросы для опроса

1. Зачем необходима информация о read group в BAM-файле для коллинга вариантов?
2. Какие параметры качества VCF-записи вы знаете? Что отражает QUAL и INFO-поле DP?
3. В чём различие между герминативным и соматическим коллингом вариантов?
4. Как SnpEff предсказывает эффект варианта? Какие транскрипты используются для аннотации?
5. Опишите алгоритмический подход bcftools mpileup для определения вероятности генотипа в позиции.

Практическая работа № 4

Тема: Анализ дифференциальной экспрессии генов по данным RNA-seq

Содержание занятия:

1. Получение матрицы количества прочтений на ген:
 - Картирование ридов RNA-seq на референсный геном с помощью STAR (учёт сплайсинга).
 - Подсчёт числа ридов, картированных на экзоны генов, с использованием featureCounts или HTSeq-count. Получение count-таблицы.
2. Импорт данных в среду R / RStudio. Установка и загрузка пакетов DESeq2, ggplot2, pheatmap.
3. Создание объекта DESeqDataSet, предварительная фильтрация генов с низкой экспрессией.
4. Статистический анализ дифференциальной экспрессии: выполнение DESeq-анализа, получение списка генов с $\log_2$ fold change и p-value, применение поправки на множественное тестирование (Benjamini-Hochberg).
5. Визуализация результатов:
 - Построение графика «volcano plot» с выделением значимо изменённых генов.
 - Тепловая карта (heatmap) топ-50 дифференциально экспрессируемых генов.

- График MA-plot для оценки зависимости эффекта от средней экспрессии.
- 6. Экспорт списка генов для последующего функционального анализа.

Оборудование: сервер с установленными STAR, samtools, featureCounts; компьютеры с ОС Linux/Windows/Mac, R и RStudio, пакетами DESeq2, ggplot2, pheatmap, dplyr. Тестовые RNA-seq FASTQ-файлы (например, парное сравнение «лечение vs контроль») и файл аннотации генов в формате GTF.

Примерные вопросы для опроса

Почему для картирования RNA-seq ридов используют алгоритмы, учитывающие сплайсинг (STAR, HISAT2), а не обычные выравниватели типа BWA?

1. В чём разница между RPKM, FPKM и TPM? Какую метрику предпочтительнее использовать и почему?
2. Объясните, как DESeq2 моделирует дисперсию биологических повторностей. Зачем нужна оценка размера факторов (size factors)?
3. Что такое поправка на множественное тестирование и почему она критична при анализе тысяч генов?
4. Какие биологические причины могут привести к обнаружению высокого fold change при низкой экспрессии гена? Как интерпретировать такие результаты?

Практическая работа № 5

Тема: Эпигенетический анализ: пик-коллинг ChIP-seq/ATAC-seq и мотивный анализ

Содержание занятия:

1. Контроль качества эпигеномных данных: запуск FastQC для ChIP-seq/ATAC-seq FASTQ, интерпретация специфичных метрик (доля дубликатов, избыточность). Оценка качества библиотек с помощью ChIPQC или ATACseqQC.
2. Картирование на референс (bwa mem) и фильтрация: удаление дубликатов, фильтрация по MAPQ, создание BAM-файлов.
3. Пик-коллинг с помощью MACS2 для ChIP-seq (гистоновые модификации/транскрипционные факторы) или MACS2 с параметром --nomodel --shift -100 --extsize 200 для ATAC-seq. Генерация bedGraph/bigWig файлов для визуализации.
4. Визуализация профилей покрытия в IGV: загрузка bigWig, оценка пиков в промоторных регионах известных генов.
5. Аннотация пиков с помощью bedtools: поиск пересечений с геномными регионами (промоторы, экзоны, интроны, межгенные области). Построение распределения пиков по геномным компартментам в R.
6. Мотивный анализ: выделение последовательностей из пиков (bedtools getfasta), поиск обогащённых мотивов с помощью HOMER findMotifsGenome.pl или онлайн-сервиса MEME-Suite. Интерпретация найденных мотивов (связь с известными факторами транскрипции).

Оборудование: сервер/компьютеры с bwa, samtools, bedtools, MACS2, HOMER (с загруженными геномами), IGV, R/RStudio с пакетами ChIPseeker (для аннотации), ggplot2. Тестовые ChIP-seq FASTQ-файлы (например, H3K27ac) или ATAC-seq и соответствующий контроль (input).

Примерные вопросы для опроса

Для чего в ChIP-seq эксперименте нужен контрольный образец (Input)? Как он используется в MACS2?

1. Чем отличаются профили пиков гистоновых модификаций широкого типа (broad marks) от пиков транскрипционных факторов (narrow marks)? Какие параметры MACS2 следует менять?
2. Как метрика FRiP (Fraction of Reads in Peaks) характеризует качество ChIP-seq эксперимента?
3. Какие статистические модели лежат в основе пик-коллера MACS2?
4. Почему при ATAC-seq анализе важно учитывать сдвиг ридов, связанный с транспозой Tn5? Как это реализовано в параметрах MACS2?

Примерные 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.

Примерные 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. супердеревья — 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.

Примерные темы презентаций

1. Воспроизводимый биоинформатический анализ: как контейнеризация (Docker, Singularity) меняет научную практику.
2. Сравнение языков программирования в биоинформатике: Bash, Python и R — выбор инструмента под задачу.
3. Алгоритмы динамического программирования в действии: глобальное и локальное выравнивание (Needleman-Wunsch, Smith-Waterman).
4. Индексация генома и её влияние на скорость картирования: FM-индекс и его реализация в Bowtie/BWA.
5. Эволюция форматов геномных данных: от FASTA до CRAM — зачем биоинформатикам эффективное сжатие.
6. Контроль качества NGS-данных: от FastQC до MultiQC — как автоматизировать рутинный анализ.
7. Секвенирование единичных клеток: вычислительные проблемы нормализации, кластеризации и траекторного анализа.
8. Поиск генетических вариантов: от BCFtools к GATK — эволюция лучших практик коллинга SNP и Indel.
9. Аннотация вариантов: Ensembl VEP, SnpEff, ANNOVAR — сравнение подходов и

интерпретация клинической значимости.

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

Примерные тестовые задания

Вариант 1

1. Какой формат файла используется для хранения нуклеотидных последовательностей с оценками качества каждого основания?

- А) FASTA
- Б) SAM
- В) FASTQ
- Г) VCF

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

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

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

- А) BED
- Б) GFF
- В) VCF
- Г) FASTA

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

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

- А) SAMtools
- Б) BWA
- В) Trimmomatic
- Г) MACS2

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

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

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

- А) N50
- Б) FRiP
- В) MAPQ
- Г) RPKM

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

11. Что описывает BED-файл?

- А) Последовательности белков
- Б) Геномные координаты регионов (экзонов, пиков и т.д.)
- В) Филогенетические деревья
- Г) Уровни экспрессии генов

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

- А) samtools index
- Б) samtools sort
- В) samtools view
- Г) samtools merge

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

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

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

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

А) BWA aln

Б) Bowtie2

В) STAR

Г) minimap2

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

18. Какой инструмент используется для функциональной аннотации генетических вариантов (например, предсказания влияния на белок)?

А) FastQC

Б) SnpEff / Ensembl VEP

В) Trimmomatic

Г) BCFtools

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

20. Какой метод позволяет изучать трёхмерную организацию хроматина в масштабе всего генома?

А) ATAC-seq

Б) Hi-C

В) ChIP-seq

Г) WGBS

Вариант 2

1. В каком формате хранится информация о выравнивании прочтений на референсный геном в текстовом виде?

А) FASTQ

Б) SAM

В) BED

Г) GFF

2. Какая программа позволяет создать единый отчёт о качестве множества образцов, объединяя результаты FastQC?

А) Trimmomatic

Б) MultiQC

В) MACS2

Г) IGV

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

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

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

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

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

8. Какой из перечисленных инструментов предназначен для сборки генома de novo?

- А) BWA
- Б) SPAdes
- В) DESeq2
- Г) MACS2

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

10. Какой пакет R широко используется для анализа данных секвенирования единичных клеток (scRNA-seq)?

- А) DESeq2
- Б) Seurat
- В) edgeR
- Г) limma

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

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

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

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

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

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

А) MACS2

Б) MEME / HOMER

В) FastQC

Г) BCFtools

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

18. Аббревиатура NGS расшифровывается как:

А) New Generation Sequencing

Б) Next Generation Sequencing

В) Nucleotide Genome Scanning

Г) Nano Geometry Sequencing

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

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

Примерный перечень тематик курсовых работ

1. Сравнительная сборка и аннотация бактериального генома из коротких прочтений. Оценить качество сборки de novo с помощью SPAdes, выполнить предсказание генов ab initio (Prokka) и провести сравнительный анализ с референсным штаммом.

2. Поиск и функциональная аннотация однонуклеотидных полиморфизмов (SNP) у модельного организма.

Выполнить коллинг вариантов (BCFtools/GATK), отфильтровать их по качеству и аннотировать эффект (SnprEff) для выявления потенциально значимых замен.

3. Анализ дифференциальной экспрессии генов по данным RNA-seq при ответе на стрессовое воздействие.

С использованием DESeq2 или edgeR определить гены, значимо меняющие экспрессию, и провести функциональное обогащение (GO, KEGG).

4. Реконструкция и визуализация генных регуляторных сетей на основе транскриптомных данных.

Используя инструменты типа GENIE3 или ARACNe, построить сеть коэкспрессии, выделить модули и соотнести их с биологическими процессами.

5. Идентификация и характеристика пиков связывания транскрипционного фактора по данным ChIP-seq.

С помощью MACS2 найти участки связывания, провести мотивный анализ (HOMER/MEME) и аннотировать целевые гены.

6. Анализ доступности хроматина методом ATAC-seq и её связь с экспрессией генов. Выявить участки открытого хроматина, интегрировать с RNA-seq данными и определить регуляторные элементы, активные в конкретных условиях.

7. Метагеномное профилирование таксономического состава микробного сообщества по данным 16S рРНК.

Используя QIIME2 или DADA2, проанализировать альфа- и бета-разнообразие, выявить дифференциально представленные таксоны.

8. Восстановление геномов (MAGs) из метагеномных сборок и их функциональный потенциал.

Выполнить биннинг контигов (MetaBAT, CONCOCT), оценить качество геномов (CheckM) и проанализировать наличие генов устойчивости к антибиотикам.

9. Филогеномный анализ группы бактерий на основе однокопийных ортологов. Извлечь маркерные гены, построить филогенетическое дерево (IQ-TREE) и сравнить топологию с деревом, полученным по 16S рРНК.

10. Исследование пангенома патогенных штаммов: ядро и аксессуарные гены. С помощью Roary или Panaroo определить пангеном, выявить гены вирулентности и построить кривую размера пангенома.

11. Выявление структурных вариантов (SV) в геноме человека по данным длинных прочтений.

Используя Sniffles или cuteSV, найти делеции, дупликации, инсерции и оценить их возможное влияние на фенотип.

12. Анализ альтернативного сплайсинга по данным RNA-seq при развитии заболевания. С помощью rMATS или SUPPA определить изменения в использовании экзонов и провести аннотацию затронутых изоформ.

13. Эпигенетический анализ полногеномного бисульфитного секвенирования (WGBS). Картировать риды (Bismark), выявить дифференциально метилированные регионы (DMR) и соотнести их с промоторами генов.

14. Интеграция данных ChIP-seq (H3K27ac) и RNA-seq для идентификации активных энхансеров.

Определить суперэнхансеры (ROSE), связать их с генами-мишенями и оценить обогащение функциональных категорий.

15. Исследование трёхмерной организации хроматина (Hi-C) в локусе интереса. Построить контактные матрицы (Juicer, cooler), визуализировать топологически ассоциированные домены (TAD) и петли хроматина, сопоставить с эпигеномными метками.

16. Разработка воспроизводимого пайплайна для анализа RNA-seq с использованием Snakemake/Nextflow.

Реализовать автоматизированный workflow от контроля качества до отчета с дифференциальной экспрессией и визуализацией.

17. Сравнение алгоритмов нормализации и кластеризации для single-cell RNA-seq данных.

На реальном датасете оценить влияние методов (SCTransform, LogNormalize) и алгоритмов кластеризации (Leiden, Louvain) на результат.

18. Оценка качества референсных геномов разных версий с помощью BUSCO и QUAST. Сравнить несколько сборок генома модельного организма, выявить проблемные участки и дать рекомендации по выбору референса для картирования.

19. Прогнозирование патогенности генетических вариантов методами машинного обучения.

С применением CADD, PolyPhen-2 или собственной модели на основе аннотаций классифицировать несинонимичные SNP и сопоставить с клиническими данными.

20. Биоинформатический поиск и аннотация мобильных генетических элементов в геномах прокариот. Используя ISfinder, MITE-Hunter или другие инструменты, выявить вставки IS-элементов и оценить их вклад в геномную пластичность.

Примерный перечень вопросов 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 rRNA) и полногеномный метагеномный анализ; таксономическое профилирование (Kraken, MetaPhlAn).
27. Восстановление геномов из метагеномов (MAGs): дифференциальный биннинг по тетра-нуклеотидным 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 дупликаций, 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.

Примерный перечень практических заданий на экзамен

1. Напишите команду для запуска FastQC на файле sample.fastq. Какие две метрики вы проверите в первую очередь?
2. Напишите команду для тримминга адаптеров программой cutadapt с параметром качества

20 и минимальной длиной 50.

3. Напишите команды для индексации референса ref.fa программой bwa и последующего картирования парных ридов R1.fq и R2.fq.

4. Как преобразовать SAM в BAM, отсортировать и индексировать BAM с помощью samtools? Напишите цепочку из трёх команд.

5. Напишите команду bcftools mpileup и bcftools call для получения вариантов с минимальным качеством картирования 30.

6. Как отфильтровать VCF-файл, оставив только варианты с $QUAL > 100$ и глубиной $DP \geq 10$? Напишите команду bcftools filter.

7. Напишите команду для аннотации VCF-файла с помощью SnpEff для генома GRCh38.

8. В чём разница между missense_variant и nonsense_variant? Приведите по одному примеру влияния на белок.

9. Для RNA-seq: напишите команду картирования программой STAR с указанием индекса и файлов ридов.

10. Как подсчитать число ридов на гены с помощью featureCounts? Напишите команду с указанием файла аннотации GTF.

11. Напишите на R две строки кода: загрузка матрицы counts и запуск дифференциального анализа с помощью DESeq2 (без параметров).

12. Почему для RNA-seq используют отрицательное биномиальное распределение, а не нормальное?

13. Объясните одной фразой, что такое граф де Брёйна для $k=3$.

14. Напишите команду для вызова пиков ChIP-seq программой macs2 callpeak с контролем input.bam.

15. Что измеряет метрика FRiP? Какой уровень FRiP считается хорошим?

16. Напишите команду для таксономической классификации метагеномных ридов с помощью kraken2.

17. Что такое MAG? Как оценить его качество? Назовите одну программу для биннинга.

18. Напишите команду iqtree для построения дерева максимального правдоподобия по выровненному FASTA с моделью GTR+G и 1000 бутстрепов.

19. Как найти перепредставленность GO-терминов по списку генов? Назовите один инструмент (сайт или пакет R).

20. Напишите на R команду для построения PCA-графика по матрице экспрессии с помощью ggplot2 (без настройки осей, только общая конструкция).

5.4 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Программа освоения дисциплины предусматривает следующие формы текущего контроля: активно участвовать в опросах, подготовить доклад, презентацию, реферат, тестирование.

Формой промежуточной аттестации является экзамен, который проходит в форме устного собеседования по вопросам.

Максимальное количество баллов, которое может набрать студент в течение семестра за различные виды работ – 70 баллов. Максимальная сумма баллов, которые может получить студент на экзамене – 30 баллов. Общая сумма баллов по изучаемой дисциплине составляет максимально-100 баллов.

Шкала оценивания экзамена

Критерии оценивания	Баллы
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
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
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 недочётами: скрипт содержит ошибки или реализована только часть пайплайна, нет чёткой интерпретации результатов. Студент может объяснить отдельные шаги, но не демонстрирует целостного понимания процесса.	13–18
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
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

Итоговая шкала выставления оценки по дисциплине

При выставлении итоговой оценки преподавателем учитывается работа магистранта в течение всего срока освоения дисциплины, а также баллы, полученные на промежуточной аттестации.

Баллы, полученные магистрантами в течение освоения дисциплины	Оценка по дисциплине
81-100	Отлично
61-80	Хорошо
41-60	Удовлетворительно

Шкала оценивания курсовой работы

Критерии оценивания	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 присутствуют, 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.	41–60
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	0–20

Критерии оценивания	Баллы
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.	

Итоговая шкала выставления оценки по курсовой работе

При выставлении оценки за курсовую работу преподавателем учитывается работа магистранта над выполнением курсовой работы в течение всего семестра, а также защита самой курсовой работы.

Итоговые баллы, полученные магистрантами на защите курсовой работы	Оценка по курсовой работе
81-100	Отлично
61-80	Хорошо
41-60	Удовлетворительно
0-40	Неудовлетворительно

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1 Основная литература

1. Биоинформатика : учебник для вузов / под редакцией М. С. Гельфанда. — Москва : Юрайт, 2025. — 412 с. — Текст : электронный. — URL: <https://urait.ru/bcode/520001>
2. Компьютерная геномика: анализ данных секвенирования нового поколения : учебное пособие / А. В. Козлов, Е. Н. Соколова. — Москва : КноРус, 2024. — 286 с. — Текст : электронный. — URL: <https://book.ru/book/950123>
3. Вычислительная биология и биоинформатика : учебное пособие для вузов / И. В. Дементьева [и др.]. — Санкт-Петербург : Лань, 2025. — 340 с. — Текст : электронный. — URL: <https://e.lanbook.com/book/340560>

6.2 Дополнительная литература

- 7 Кунин Е. В. Логика случая: природа и происхождение биологической эволюции / пер. с англ. — 3 е изд., перераб. — Москва : Лаборатория знаний, 2025. — 648 с. — Текст : электронный. — URL: <https://e.lanbook.com/book/375001>
- 8 Прикладная биоинформатика: практикум по анализу данных NGS / под ред. Н. А. Колчанова. — Новосибирск : Изд-во НГУ, 2024. — 256 с. — Текст : непосредственный.
- 9 Себастиан В. Прикладная биоинформатика: Python и R для анализа геномных данных / пер. с англ. — Москва : ДМК Пресс, 2025. — 512 с. — Текст : электронный. — URL: <https://dmkpress.com/catalog/computer/data/9785970601235/>
- 10 Бородовский М. Ю., Опарин В. Д. Высокопроизводительные вычисления в биоинформатике: НРС, облака и контейнеризация. — Москва : Издательство МГУ, 2024. — 198 с. — Текст : электронный.
- 11 Певзнер П., Компо Ф. Алгоритмы в биоинформатике: введение для биологов / пер. с англ. — Москва : Юрайт, 2025. — 480 с. — Текст : электронный. — URL: <https://urait.ru/bcode/521234>
- 12 Геномика и постгеномика: методы анализа и интерпретации данных / под ред. А. А. Миронова. — Москва : Лаборатория знаний, 2024. — 368 с. — Текст : электронный. — URL: <https://e.lanbook.com/book/365780>
- 13 Флах П. Машинное обучение: наука и искусство построения алгоритмов, которые извлекают знания из данных. — 2 е изд. — Москва : Вильямс, 2025. — 400 с. — Текст : непосредственный.
- 14 Метагеномика и мультиомиксные подходы в экологии и медицине : учебное пособие / Т. В. Татарина, Н. Н. Лукашова. — Ростов-на-Дону : ЮФУ, 2024. — 212 с. — Текст : электронный. — URL: <https://www.studentlibrary.ru/book/ISBN9785927534562.html>
- 15 Певзнер Д. А. Визуализация и интерпретация геномных данных средствами R/Bioconductor : учебно-методическое пособие. — Москва : ИПК «Логос», 2025. — 144 с. — Текст : непосредственный.
- 16 Тельнов А. А. Репродуцируемые исследования в биоинформатике: workflow менеджеры Snakemake и Nextflow. — Москва : КноРус, 2024. — 178 с. — Текст : электронный. — URL: <https://book.ru/book/951234>

6.3 Ресурсы информационно-телекоммуникационной сети «Интернет»

1. Национальный центр биотехнологической информации США (NCBI) – базы данных GenBank, RefSeq, PubMed, SRA, инструменты BLAST, dbSNP. – URL: <https://www.ncbi.nlm.nih.gov>
2. Европейский институт биоинформатики (EMBL-EBI) – архивы ENA, ArrayExpress, PRIDE, сервисы Clustal Omega, InterPro. – URL: <https://www.ebi.ac.uk>
3. Интегративный геномный браузер Ensembl – геномные аннотации, сравнительная геномика, вариация, регуляция. – URL: <https://www.ensembl.org>
4. Браузер генома UCSC – визуализация, аннотации, консервативные элементы, ENCODE. –

- URL: <https://genome.ucsc.edu>
5. Универсальный белковый ресурс UniProt – последовательности, функциональные аннотации, домены. – URL: <https://www.uniprot.org>
 6. Консорциум Gene Ontology – контролируемый словарь функций генов, обогащение. – URL: <http://geneontology.org>
 7. Платформа для воспроизводимой биоинформатики Galaxy – графический интерфейс для пайплайнов. – URL: <https://usegalaxy.org>
 8. Проект Bioconductor – программные пакеты на R для анализа омиксных данных. – URL: <https://www.bioconductor.org>
 9. Российская платформа «Биоинформатический портал» – курсы, датасеты, вики-ресурсы. – URL: <https://bioinf.me>
 10. Интерактивная образовательная среда Rosalind – задачи по алгоритмам биоинформатики. – URL: <https://rosalind.info>
 11. Платформа Kaggle – соревнования и датасеты по биомедицинской тематике. – URL: <https://www.kaggle.com>
 12. Обучающий ресурс «Вся биология» – раздел современной геномики и биоинформатики. – URL: <https://sbio.info>
 13. Федеральный институт промышленной собственности (ФИПС) – патентные базы, в т.ч. по биотехнологии. – URL: <https://www.fips.ru>
 14. Европейское патентное ведомство (Espacenet) – поиск патентов в области биоинформатики. – URL: <https://worldwide.espacenet.com>
 15. Электронно-библиотечная система Лань <https://e.lanbook.com>
 16. ООО «Электронное издательство Юрайт» <https://urait.ru>

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

1..Методические рекомендации по организации самостоятельной работы

8. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ

Лицензионное программное обеспечение:

Microsoft Windows (версия 10/11 или выше)

Microsoft Office (в составе Word, Excel, PowerPoint)

Kaspersky Endpoint Security для рабочих станций

MATLAB (с пакетами Bioinformatics Toolbox, Statistics and Machine Learning Toolbox) – образовательная лицензия

ПО для удалённого доступа к высокопроизводительным вычислительным кластерам (лицензия на клиентские приложения, например, MobaXterm Professional)

Информационные справочные системы:

Система «КонсультантПлюс»

Специализированная справочная система по биоинформатике – онлайн-документация и форумы Biostars (<https://www.biostars.org>)

Профессиональные базы данных:

fgosvo.ru – Портал Федеральных государственных образовательных стандартов высшего образования

pravo.gov.ru – Официальный интернет-портал правовой информации

www.edu.ru – Федеральный портал Российское образование

Биоинформатические профессиональные базы данных:

NCBI (National Center for Biotechnology Information) – <https://www.ncbi.nlm.nih.gov> (GenBank, RefSeq, SRA, dbSNP, PubMed)

EMBL-EBI (European Bioinformatics Institute) – <https://www.ebi.ac.uk> (ENA, ArrayExpress, PRIDE, InterPro)

Ensembl Genome Browser – <https://www.ensembl.org>

UCSC Genome Browser – <https://genome.ucsc.edu>

UniProt (Universal Protein Resource) – <https://www.uniprot.org>

KEGG (Kyoto Encyclopedia of Genes and Genomes) – <https://www.genome.jp/kegg/>

Gene Ontology (GO) – <http://geneontology.org>

Protein Data Bank (PDB) – <https://www.rcsb.org>

Свободно распространяемое программное обеспечение, в том числе отечественного производства

Системное и общего назначения:

ОМС Плеер (для воспроизведения Электронных Учебных Модулей)

7-zip (архиватор)

Google Chrome (интернет-браузер)

LibreOffice (офисный пакет, альтернатива Microsoft Office)

Виртуализация и контейнеризация:

Oracle VirtualBox (для развёртывания виртуальных машин Linux)

Docker Desktop (для работы с биоинформатическими контейнерами)

WSL (Windows Subsystem for Linux) – встроенный компонент Windows

Операционные системы (свободные версии):

Ubuntu Linux (или аналоги: Debian, CentOS Stream)

Astra Linux Common Edition (отечественная ОС для рабочих станций)

Языки программирования и среды разработки:

Python (интерпретатор) с менеджером пакетов conda (Miniconda3)

R (среда статистической обработки данных)

RStudio Desktop (интегрированная среда для R)

Jupyter Notebook / JupyterLab

Visual Studio Code (редактор кода)

Специализированные биоинформатические пакеты (устанавливаются через conda или из исходных кодов):

FastQC, MultiQC (контроль качества NGS-данных)

Trimomatic, cutadapt (тримминг прочтений)

BWA, Bowtie2, STAR, HISAT2 (картирование коротких и сплайсированных прочтений)

SAMtools, BEDTools (работа с форматами SAM/BAM, BED)

BCFtools, GATK (поиск и фильтрация генетических вариантов)

SnEff, Ensembl VEP (аннотация вариантов)

featureCounts, HTSeq-count, Salmon (количественная оценка экспрессии)

DESeq2, edgeR (анализ дифференциальной экспрессии, пакеты R)

MACS2 (поиск пиков ChIP-seq/ATAC-seq)

HOMER, MEME Suite (мотивный анализ)

SPAdes, Canu (сборка геномов de novo)

Bismark, bwa-meth (анализ бисульфитного секвенирования)

Snakemake, Nextflow (менеджеры пайплайнов)

Интегративные платформы и браузеры (веб-интерфейсы, не требующие установки):

Galaxy (<https://usegalaxy.org>) – графическая платформа для построения биоинформатических конвейеров

Integrative Genomics Viewer (IGV) – локальное приложение для визуализации геномных данных

Отечественное свободное ПО:

Язык программирования lScript (при изучении автоматизации, опционально)

Платформа «Биоинформатический портал» (<https://bioinf.me>) – образовательные материалы и облачные вычисления (русская разработка)

Все перечисленные свободно распространяемые программы могут быть установлены как на рабочие станции учащихся, так и использоваться в виртуализованной среде под управлением Linux. При необходимости все пакеты актуализируются и доустанавливаются администратором учебных классов или удалённых серверов.

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Материально-техническое обеспечение дисциплины включает в себя:

- учебные аудитории для проведения лекционных и семинарских занятий, оснащённые учебной мебелью, доской, демонстрационным оборудованием, персональным компьютером преподавателя с подключением к сети «Интернет», мультимедийным проектором и экраном;
- компьютерный класс (лаборатория) для проведения практических занятий, оснащённый рабочими станциями учащихся с установленным специализированным свободно распространяемым и лицензионным программным обеспечением (операционные системы Linux и/или Windows, биоинформатические пакеты, языки программирования R и Python, среды разработки, доступ к менеджерам пакетов conda и контейнеризации), подключением к локальной сети и сети «Интернет»; при необходимости — доступом к высокопроизводительному вычислительному серверу (HPC-кластеру) или облачной вычислительной инфраструктуре;
- помещения для самостоятельной работы обучающихся, оснащённые компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде, профессиональным биоинформатическим базам данных и вычислительным ресурсам
- помещение для самостоятельной работы для инвалидов и лиц с ограниченными возможностями здоровья, оснащенное компьютерной техникой, подключенной к сети Интернет, обеспечено доступом к электронно-образовательной среде Университета, Комплект учебной мебели, персональные компьютеры с подключением к сети Интернет и обеспечением доступа к электронным библиотекам и в электронную информационно-образовательную среду Университета, доска, проектор подвесной.