

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Наумова Наталья Александровна
Должность: Ректор
Дата подписания: 17.08.2026 18:05:25
Уникальный программный ключ:
6b5279da4e034bff679172803da5b7b5559f619e

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ»
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ)

Физико-математический факультет

Кафедра вычислительной математики и информационных технологий

УТВЕРЖДЕН

на заседании кафедры

Протокол от «18» марта 2026 г., № 9

Зав. кафедрой Шевчук М.В. /Шевчук М.В./

**ФОНД
ОЦЕНОЧНЫХ СРЕДСТВ**

по дисциплине

Языки программирования для биоинформатиков

Направление подготовки (специальности) 06.04.01 Биология

Программа подготовки Биоинформатика и анализ больших данных в экологическом мониторинге

Москва
2026

Содержание

1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы.....	3
2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания	3
3. Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы.....	7
4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.....	32

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

Изучение дисциплины «Языки программирования для биоинформатиков» позволяет сформировать следующие компетенции:

Код и наименование компетенции	Этапы формирования
СПК-2. Способен использовать программы для эпигенетического анализа и качества эпигенетических данных	1. Работа на учебных занятиях 2. Самостоятельная работа
СПК-4. Способен проводить экспертно-аналитическую работу при проведении научных исследований и экспериментальных работ	1. Работа на учебных занятиях 2. Самостоятельная работа

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
СПК-2	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа	Знать: - основные понятия эпигенетики и типы эпигенетических данных; принципы работы программных средств эпигенетического анализа; - методы оценки качества и достоверности эпигенетических данных; источники ошибок и искажений в эпигенетических наборах данных. Уметь: - использовать программные	Лабораторные работы Конспект Тестирование	Шкала оценивания лабораторной работы Шкала оценивания конспекта Шкала оценивания тестирования

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			инструменты для обработки эпигенетических данных; - выполнять первичную фильтрацию и подготовку данных к анализу; - проводить оценку качества эпигенетических данных; - интерпретировать результаты эпигенетического анализа.		
	Продвинутый	1. Работа на учебных занятиях 2. Самостоятельная работа	Знать: - основные понятия эпигенетики и типы эпигенетических данных; принципы работы программных средств эпигенетического анализа; - методы оценки качества и достоверности эпигенетических данных; источники ошибок и искажений в эпигенетических наборах данных. Уметь: - использовать программные инструменты для обработки эпигенетических данных; - выполнять первичную фильтрацию и подготовку данных к анализу; - проводить оценку качества эпигенетических данных; - интерпретировать результаты эпигенетического анализа. Владеть: - навыками работы с программами эпигенетического анализа; - методами контроля качества биологических данных; - инструментами визуализации эпигенетических	Лабораторные работы Конспект Тестирование	Шкала оценивания лабораторной работы Шкала оценивания конспекта Шкала оценивания тестирования

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			профилей; - практическими навыками обработки и интерпретации эпигенетических данных.		
СПК-4	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа	Знать: - методы экспертного анализа научных и экспериментальных данных; -принципы организации и проведения научных исследований; -основы статистической обработки биологических данных; -требования к оформлению научно-аналитических материалов. Уметь: -проводить анализ и интерпретацию результатов научных исследований; -применять методы статистической обработки данных; -выявлять закономерности и аномалии в экспериментальных данных; -оформлять аналитические отчёты по результатам исследований.	Лабораторные работы Конспект Тестирование	Шкала оценивания лабораторной работы Шкала оценивания конспекта Шкала оценивания тестирования
	Продвинутый	1. Работа на учебных занятиях 2. Самостоятельная работа	Знать: - методы экспертного анализа научных и экспериментальных данных; -принципы организации и проведения научных исследований; -основы статистической обработки биологических данных; -требования к оформлению научно-аналитических материалов. Уметь: -проводить анализ и интерпретацию результатов научных исследований; -применять методы	Лабораторные работы Конспект Тестирование	Шкала оценивания лабораторной работы Шкала оценивания конспекта Шкала оценивания тестирования

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			статистической обработки данных; -выявлять закономерности и аномалии в экспериментальных данных; -оформлять аналитические отчёты по результатам исследований. Владеть: -навыками экспертной оценки научных данных; -методами анализа экспериментальных результатов; -инструментами статистической обработки информации; -технологиями подготовки научно-аналитических заключений.		

Описание шкал оценивания

Критерии и шкала оценивания лабораторной работы

Критерий оценивания	Баллы
Задание выполнено полностью, соответствует предъявляемым требованиям (к каждому заданию предъявляются свои требования, прописанные перед каждым заданием)	6
Задание выполнено полностью, но есть неточности в оформлении материала или совсем не соответствует требованиям, предъявляемым к оформлению	3
Задание выполнено не полностью или есть неточности в выполнении, есть неточности в оформлении материала или совсем не соответствует требованиям, предъявляемым к оформлению	2
Максимальное количество баллов	6

Критерии и шкала оценивания конспекта

Критерии оценивания	Баллы
Текст конспекта логически выстроен и точно изложен, ясен весь ход рассуждения	1
Даны ответы на все поставленные вопросы, изложены научным	1

языком, с применением терминологии	
Ответ на каждый вопрос заканчиваться выводом, сокращения слов в тексте отсутствуют (или использованы общепринятые)	0,5
Оформление соответствует образцу. Представлены необходимые таблицы и схемы	0,5
Максимальное количество баллов	3

Критерии и шкала оценивания тестирования

Критерии оценивания	Баллы за один правильный ответ
На вопрос дан правильный ответ	1
На вопрос дан неправильный ответ	0
Максимальное количество баллов за тест (15 вопросов)	15

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

Текущий контроль

СПК-4. Способен проводить экспертно-аналитическую работу при проведении научных исследований и экспериментальных работ

Знать: методы экспертного анализа научных и экспериментальных данных, принципы организации и проведения научных исследований, основы статистической обработки биологических данных, требования к оформлению научно-аналитических материалов.

Задания, необходимые для оценивания сформированности СПК-4 на пороговом уровне

Перечень тестовых заданий

1. **Что является основным критерием достоверности результатов научного эксперимента в экспертно-аналитической работе?**
 - а) Соответствие результатов гипотезе исследователя
 - б) Воспроизводимость результатов при повторении эксперимента**
 - в) Количество процитированных источников в обзоре
 - г) Сложность использованного оборудования
2. **Какой метод экспертной оценки минимизирует субъективность мнений и основывается на сборе и обработке индивидуальных анонимных оценок независимых специалистов?**
 - а) Метод Дельфи**
 - б) Мозговой штурм (Brainstorming)
 - в) SWOT-анализ
 - г) Метод фокус-групп
3. **Что представляет собой первичный этап проведения аналитического исследования в науке?**
 - а) Проведение экспериментальных тестов
 - б) Оформление патента на изобретение
 - в) Формулировка проблемы, цели и задач исследования**
 - г) Расчет статистической значимости данных
4. **Какая шкала измерения используется для ранжирования экспертами объектов (например, приоритетность научных направлений) от наиболее важного к наименее важному?**
 - а) Номинальная
 - б) Порядковая (ординальная)**
 - в) Интервальная
 - г) Абсолютная
5. **Какая статистическая ошибка возникает, когда исследователь признает наличие эффекта (или связи), которого в реальности не существует (отвергнута верная нулевая гипотеза)?**

- а) Ошибка I рода (альфа-ошибка)
- б) Ошибка II рода (бета-ошибка)
- в) Систематическая ошибка отбора
- г) Ошибка измерения

Задания, необходимые для оценивания сформированности СПК-4 на продвинутом уровне

Перечень тем конспектов

1. **Планирование эксперимента:** Методология построения гипотез и выбор оптимальных методов исследования.
2. **Анализ научной литературы:** Техники быстрого критического разбора и структурирования источников (рецензируемые статьи, патенты, обзоры).
3. **Оценка погрешностей и неопределенностей:** Анализ систематических и случайных ошибок при проведении измерений.
4. **Статистическая обработка данных:** Выбор и применение базовых математических инструментов для подтверждения достоверности результатов.
5. **Методы визуализации данных:** Правила построения графиков, диаграмм и таблиц для наглядной интерпретации итогов.
6. **Оценка применимости оборудования:** Анализ технических характеристик приборов и их влияние на точность эксперимента.
7. **Анализ полноты эксперимента:** Методы проверки достаточности выборки и отсутствия пропущенных переменных.
8. **Систематизация качественных данных:** Подходы к анализу результатов фокус-групп, опросов или интервью в прикладных науках.
9. **Преимственность и воспроизводимость:** Анализ условий, необходимых для повторения эксперимента другими исследователями.
10. **Оценка выводов и экстраполяция:** Способы проверки логической связи между полученными результатами и финальным заключением работы.

Уметь: проводить анализ и интерпретацию результатов научных исследований, применять методы статистической обработки данных, выявлять закономерности и аномалии в экспериментальных данных, оформлять аналитические отчёты по результатам исследований.

Задания, необходимые для оценивания сформированности СПК-4 на пороговом уровне¹

Типовые задания лабораторных работ

Задание 1. Анализ источников и формирование гипотезы

Суть задания: Студентам предлагается выбрать узкую научную проблему, изучить 3–5 актуальных статей (например, в Google Академии) и сформулировать первичную проверяемую гипотезу.

Аналитический фокус: Оценка достоверности источников, выявление «белых пятен» в текущих исследованиях и логическое обоснование своей гипотезы.

Задание 2. Планирование эксперимента и выявление переменных

Суть задания: Разработать детальный план модельного эксперимента (например, влияние освещения на скорость роста микрорзелени или температуры на время растворения вещества).

Аналитический фокус: Классификация переменных (независимые, зависимые, контролируемые), выбор методов измерения и оценка возможных систематических или случайных погрешностей.

Задание 3. Обработка данных и графическая визуализация

Суть задания: На основе предоставленного набора «сырых» экспериментальных данных (таблица с измерениями) требуется очистить данные от аномалий, провести базовые статистические расчеты (среднее значение, стандартное отклонение) и построить информативный график.

¹ Указываются отдельно по уровням, в случае если формулировки ЗУВ различаются в зависимости от уровней сформированности компетенций.

Аналитический фокус: Выбор оптимального типа диаграммы для конкретного типа данных и умение интерпретировать визуальные тренды без математической подгонки.

Задание 4. Поиск артефактов и оценка достоверности результатов

Суть задания: Предоставить студентам результаты гипотетического или реального эксперимента с заведомо скрытой ошибкой (например, нарушением методики или выбросом из-за неисправного прибора).

Аналитический фокус: Обнаружение аномалий, критический анализ причин их возникновения, принятие экспертного решения об исключении данных из выборки и обоснование этого шага.

Задание 5. Структурирование выводов и оценка погрешностей

Суть задания: Написать краткий отчет по готовому набору вводных данных и результатов. Вывод должен быть разделен на три блока: ответ на гипотезу, анализ источников погрешностей и практическая значимость.

Аналитический фокус: Отделение корреляции от причинно-следственной связи, умение признавать ограничения проведенного исследования и формулировать дальнейшие шаги для науки.

Задания, необходимые для оценивания сформированности СПК-4 на продвинутом уровне

Типовые задания лабораторных работ

Задание 1. Метод Дельфи для оценки перспективности научных направлений

Цель: Исключение субъективизма и достижение консенсуса при оценке инновационных проектов.

Задание: Студентам предлагается выступить в роли независимых экспертов для оценки 5 гипотетических научных проектов. Задача — провести $N=3$ раунда анонимного анкетирования с обоснованием своих оценок. После каждого раунда студенты получают сводную статистику (медиану и квартили оценок).

Анализ: Снижение дисперсии оценок и выявление причин расхождения мнений экспертов по ключевым критериям (новизна, реализуемость).

Задание 2. Матричный анализ и ранжирование экспериментальных факторов

Цель: Выявление приоритетных направлений для изменения условий эксперимента.

Задание: На основе заданного набора экспериментальных данных (например, температуры, давления, концентрации реагента) и их влияния на выход продукта, составить матрицу взвешенных оценок.

Анализ: Применение метода анализа иерархий (МАИ). Определение весовых коэффициентов методом попарного сравнения и обоснование оптимального вектора управляющих параметров для следующей серии опытов.

Задание 3. Экспертиза на достоверность и выявление артефактов в массивах данных

Цель: Очистка результатов эксперимента от случайных и систематических погрешностей.

Задание: Предоставляется массив «сырых» данных (например, спектроскопии или хроматографии) с искусственно встроенными аномалиями (выбросами) и скрытым дрейфом базовой линии.

Анализ: Оценка с помощью критериев Романовского или Шовене. Внешняя верификация данных, построение калибровочных графиков, исключение некорректных результатов и оценка остаточной погрешности с помощью пакетов статистического анализа.

Задание 4. Метод «Черного ящика» для факторного анализа сложных систем

Цель: Определение функциональных связей в системах без доступа к их внутреннему устройству.

Задание: Моделируется физический процесс с неизвестными внутренними параметрами. Варьируя входные данные и фиксируя выходные значения, требуется построить математическую модель процесса.

Анализ: Использование регрессионного анализа. Проверка адекватности построенной модели через коэффициент детерминации и составление экспертного заключения о поведении системы.

Задание 5. Оценка неопределенностей и рисков в планировании экспериментов

Цель: Определение пределов доверия к полученным научным результатам.

Задание: Оценить систематические и случайные погрешности в серии многократных измерений. Рассчитать стандартную и расширенную неопределенность измерений согласно руководству (GUM).

Анализ: Построение бюджета неопределенности. Ранжирование источников ошибок (приборные, методические, внешние факторы) и выработка экспертных рекомендаций по модернизации лабораторной установки.

Выполнение этих работ развивает критическое мышление, учит обосновывать выводы и принимать решения в условиях информационной неопределенности.

Владеть: навыками использования вычислительной техники в биоинформатике, методами обработки и анализа экспериментальных данных, инструментами автоматизации обработки биологических данных.

Задания, необходимые для оценивания сформированности СПК-4 на пороговом уровне

Перечень тем конспектов

1. Методология критического анализа научной литературы

Оценка достоверности источников, выявление методологических ошибок и предвзятости в опубликованных исследованиях.

2. Планирование и дизайн научных экспериментов

Основы теории планирования эксперимента, выделение переменных, контроль факторов и обеспечение воспроизводимости результатов.

3. Статистические методы обработки экспериментальных данных

Применение описательной и математической статистики: проверка гипотез, корреляционный и дисперсионный анализы для подтверждения значимости результатов.

4. Методы сбора, валидации и очистки данных

Практики работы с «сырыми» данными, обнаружение аномалий и обеспечение надежности датасетов для анализа.

5. Оценка неопределенности и погрешностей измерений

Классификация и расчет систематических и случайных погрешностей, анализ пределов точности измерительного оборудования.

6. Оценка рисков и обоснование безопасности эксперимента

Идентификация потенциальных экологических, технических и этических рисков, а также разработка протоколов минимизации.

7. Инструменты автоматизации и визуализации аналитических данных

Обзор современных программных продуктов для анализа (например, R, Python) и принципы создания информативных графиков для отчетов.

8. Коммерциализация исследований и защита интеллектуальной собственности

Основы патентного поиска, анализ патентной чистоты и базовые стратегии защиты научных разработок перед внедрением.

9. Структурирование и написание экспертно-аналитических резюме

Правила составления научных отчетов, рефератов и аналитических справок для целевой аудитории (от грантодателей до профильных специалистов).

Задания, необходимые для оценивания сформированности СПК-4 на продвинутом уровне

Типовые задания лабораторных работ

Задание 1. Валидация и очистка экспериментальных данных

- **Суть работы:** Выявление аномалий, пропусков и систематических ошибок в «сырых» данных, полученных с датчиков или в ходе анкетирования.
- **Задачи:**
 - Провести первичный разведочный анализ данных (EDA).
 - Найти и обосновать удаление аномальных выбросов с помощью межквартильного размаха (IQR) или критерия Шовене.

- Применить методы интерполяции или множественного *imputation* для восстановления пропущенных значений.
- **Аналитический фокус:** Оценка влияния методов очистки на итоговую статистическую значимость результатов.

Задание 2. Сравнительный анализ моделей для аппроксимации экспериментальных кривых

- **Суть работы:** Выбор математической модели, наиболее точно описывающей физический, химический или биологический процесс.
- **Задачи:**
 - Подобрать не менее трех теоретических моделей (например, линейную, полиномиальную или экспоненциальную) для описания заданного массива точек.
 - Выполнить подгонку кривых (*curve fitting*) с расчетом коэффициентов.
 - Сравнить модели между собой с использованием метрик R^2 , скорректированного R^2 и критерия Акайке (AIC).
- **Аналитический фокус:** Выбор оптимальной модели по принципу бритвы Оккама (баланс между точностью и сложностью во избежание переобучения).

Задание 3. Анализ многомерных данных методом главных компонент (PCA)

- **Суть работы:** Снижение размерности многомерного массива данных для выявления скрытых закономерностей и кластеров.
- **Задачи:**
 - Нормализовать исходный многомерный датасет.
 - Рассчитать матрицу ковариации, собственные значения и собственные векторы.
 - Построить график «каменистой осыпи» (*scree plot*) и выбрать оптимальное количество главных компонент.
 - Проинтерпретировать физический или смысловой смысл выделенных компонент.

- **Аналитический фокус:** Экспертная интерпретация того, какие именно комбинации исходных параметров вносят наибольший вклад в дисперсию результатов.

Задание 4. Оценка неопределенности измерений и анализ источников погрешностей

- **Суть работы:** Комплексный анализ точности экспериментальной установки или методики.
- **Задачи:**
 - о Рассчитать стандартную и расширенную неопределенность прямых измерений (тип A и тип B).
 - о Выполнить закон распространения неопределенности для косвенных измерений (например, для вычисления плотности $\rho=m/V$).
 - о Построить бюджет неопределенности в виде таблицы.
- **Аналитический фокус:** Определение главного источника погрешности и разработка рекомендаций по модернизации эксперимента для повышения его точности.

Задание 5. Планирование многофакторного эксперимента (DOE)

- **Суть работы:** Оптимизация процесса проведения эксперимента для минимизации количества опытов при оценке влияния множества факторов.
- **Задачи:**
 - о Разработать матрицу планирования для полнофакторного эксперимента 2^k (например, для 3 независимых факторов).
 - о Провести эксперимент на виртуальной модели или по готовым табличным данным.
 - о Рассчитать коэффициенты регрессии и построить поверхность отклика.
 - о Определить оптимальные значения факторов для достижения целевого результата.
- **Аналитический фокус:** Выявление эффектов взаимодействия между факторами (например, как температура и давление влияют на выход продукта совместно, а не по отдельности).

Совет по выполнению: Для реализации этих работ рекомендуется использовать среду программирования Python (библиотеки `pandas`, `numpy`, `scipy`, `scikit-learn`, `matplotlib`) или специализированные статистические пакеты.

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

Знать: основные понятия эпигенетики и типы эпигенетических данных, принципы работы программных средств эпигенетического анализа, методы оценки качества и достоверности эпигенетических данных, источники ошибок и искажений в эпигенетических наборах данных.

Задания, необходимые для оценивания сформированности *СПК-2* на пороговом уровне

Перечень тестовых заданий

1. Какой формат файла используется для хранения трехмерной (3D) структуры белков и нуклеиновых кислот?

- А) FASTA.
- Б) FASTQ.
- В) PDB.
- Г) VCF.

2. Процесс трансляции — это синтез:

- А) РНК на матрице ДНК.
- Б) Белка на матрице мРНК.
- В) ДНК на матрице ДНК.
- Г) РНК на матрице РНК.

3. В биоинформатике под "множественным выравниванием" понимают:

- А) Сравнение двух последовательностей между собой.
- Б) Сопоставление трех и более биологических последовательностей для поиска консервативных участков.

- В) Анализ структуры одного белка.
- Г) Статистический анализ экспрессии генов.

4. Что обозначает аббревиатура SNP (СНП) в геномике?

- А) Синтетический нуклеотидный полиморфизм.
- Б) Структурный нуклеотидный повтор.
- В) Однонуклеотидный полиморфизм.
- Г) Специальный нуклеотидный белок.

5. В чем заключается главная цель транскриптомики?

- А) Определение последовательности всего генома организма.
- Б) Изучение всех метаболитов в клетке.
- В) Анализ структуры белков.
- Г) Измерение уровня экспрессии всех молекул РНК в клетке или ткани в конкретных условиях.

Задания, необходимые для оценивания сформированности СПК-2 на продвинутом уровне

Перечень тем конспектов

- **Тема 1. Биоинформатика бисульфитного секвенирования (Bisulfite Sequencing)**
 - *Принципы:* Дезаминирование неметилированного цитозина в урацил.
 - *Программы:* Bismark для выравнивания прочтений, methylKit для статистического анализа дифференциально метилированных регионов (DMRs).
- **Тема 2. Применение программ для дизайна экспериментов метилирования**
 - *Принципы:* Подбор праймеров для целевого анализа сайтов.
 - *Программы:* MethMarker для дизайна методов COBRA, пиросеквенирования и MSP.
- **Тема 3. Интегративный анализ метилирования и экспрессии генов**
 - *Принципы:* Поиск взаимосвязи между метилированием промоторов/энхансеров и транскрипционной активностью.

- *Программы:* Пакет [EpiMix](#) в R.
- **Тема 4. Обработка данных иммунопреципитации хроматина (ChIP-seq)**
 - *Принципы:* Картирование участков связывания транскрипционных факторов и гистоновых модификаций.
 - *Программы:* [MACS \(Model-based Analysis of ChIP-Seq\)](#) для поиска пиков обогащения (peak calling).
- **Тема 5. Анализ открытого хроматина и доступности ДНК (ATAC-seq)**
 - *Принципы:* Определение регуляторных элементов генома (промоторов, энхансеров) с использованием ферментной транспозазы.
 - *Программы:* Пайплайн nf-core/atacseq или пакеты Genrich.
- **Тема 6. Визуализация третичной структуры хроматина (Hi-C)**
 - *Принципы:* Изучение пространственной укладки генома и взаимодействия между удаленными локусами.
 - *Программы:* Пакет Juicer и браузер HiGlass для интерактивного просмотра контактов.
- **Тема 7. Исследование эпигенома в интерактивных геномных браузерах**
 - *Принципы:* Визуализация различных эпигенетических меток (ChIP-seq, ATAC-seq, метилирование) на референсном геноме.
 - *Программы:* [IGV \(Integrative Genomics Viewer\)](#) и браузер EpiExplorer.
- **Тема 8. Многомерный и обобщенный анализ эпигеномных датасетов**
 - *Принципы:* Выявление скрытых ассоциаций между эпигенетическими факторами и свойствами участка генома.
 - *Программы:* Веб-сервис EpiGRAPH.
- **Тема 9. Эпигеномные ассоциации (EWAS) и функциональная аннотация**
 - *Принципы:* Статистический поиск связи эпигеномных модификаций с фенотипами и заболеваниями.
 - *Программы:* Пакеты для R minfi и ChAMP.

Уметь: использовать программные инструменты для обработки эпигенетических данных, выполнять первичную фильтрацию и подготовку данных к анализу, проводить оценку качества эпигенетических данных, интерпретировать результаты эпигенетического анализа.

Задания, необходимые для оценивания сформированности СПК-2 на пороговом уровне²

Типовые задания лабораторных работ

Задание 1. Визуализация эпигеномных меток в UCSC Genome Browser

- **Задание:** Найти в браузере генома (например, hg38) промотор гена *BRCA1*. Сопоставить пики ChIP-seq для гистоновых модификаций (H3K4me3) (активный промотор) и (H3K27me3) (репрессивный хроматин) с уровнем экспрессии гена в разных клеточных линиях.

Задание 2. Анализ данных бисульфитного секвенирования (WGBS)

- **Задание:** Используя программу *Bismark* или R-пакет *methyKit*, оценить статус метилирования цитозинов (CG-динуклеотидов) в заданных регуляторных областях. Выявить дифференциально метилированные регионы (DMR) между образцами нормы и патологии.

Задание 3. Идентификация открытого хроматина (ATAC-seq)

- **Задание:** Загрузить данные секвенирования ATAC-seq. Выполнить картирование ридов, определить участки гиперчувствительности хроматина (peak calling) с помощью *MACS2* и соотнести открытые регионы с сайтами связывания известных транскрипционных факторов.

Задание 4. Профилирование экспрессии и метилирования

- **Задание:** Загрузить датасеты из базы *GEO* для образцов раковой и здоровой ткани. Провести интегративный анализ с помощью *DESeq2*, чтобы выяснить, приводит ли гиперметилирование промоторов к подавлению транскрипции определенных генов.

² Указываются отдельно по уровням, в случае если формулировки ЗУВ различаются в зависимости от уровней сформированности компетенций.

Задание 5. Аннотация эпигенетических данных

- **Задание:** Используя инструмент *ChIPseeker*, взять готовый набор данных ChIP-seq для белка CTCF (фактор изоляции) и программно аннотировать полученные пики (определить, лежат ли они в промоторах, экзонах, интронах или межгенных областях).

Задания, необходимые для оценивания сформированности СПК-2 на продвинутом уровне

Типовые задания лабораторных работ

Задание 1. Анализ профилей метилирования ДНК

Инструменты: Язык R, пакеты ChAMP или minfi.

Задание: Загрузить набор данных бисульфитного микрочипа Illumina HumanMethylation450/EPIC. Выполнить контроль качества (QC), фильтрацию данных и нормализацию. Идентифицировать дифференциально метилированные CpG-сайты (DMS) и дифференциально метилированные регионы (DMR) между контрольной и экспериментальной группами клеток.

Задание 2. Идентификация сайтов связывания белков и гистоновых меток (ChIP-seq)

Инструменты: Консольные утилиты FastQC, BWA, MACS3, deepTools.

Задание: Провести картирование сырых прочтений (FASTQ) на референсный геном, выполнить фильтрацию и удаление дубликатов. Вызвать пики (peak calling) для заданной модификации гистонов (например, H3K4me3 или H3K27me3) и построить тепловые карты (heatmap), а также профили распределения вокруг промоторов генов.

Задание 3. Анализ доступности хроматина и регуляторных элементов (ATAC-seq)

Инструменты: Trim Galore, Bowtie2, MACS3, HOMER.

Задание: Выполнить предобработку и выравнивание данных ATAC-seq. Идентифицировать открытые участки хроматина в геноме. Провести анализ обогащения (motif enrichment) в найденных пиках для поиска потенциальных транскрипционных факторов, отвечающих за регуляцию экспрессии в заданном типе ткани.

Задание 4. Визуализация эпигеномных данных и интеграция с транскриптомом

Инструменты: Геномный браузер IGV (Integrative Genomics Viewer).

Задание: Загрузить и визуализировать треки различных эпигенетических меток (H3K4me1, H3K4me3, H3K27ac) и уровней метилирования ДНК для заданного локуса. Сопоставить эти данные с треком РНК-seq для оценки того, как эпигенетический ландшафт (активные и репрессивные метки) коррелирует с уровнем экспрессии целевого гена.

Задание 5. Поиск энхансеров и функциональная аннотация эпигенома

Инструменты: R, пакет ChIPseeker, базы данных ENCODE / Roadmap Epigenomics.

Задание: Аннотировать набором данных эпигенетических пиков (полученных в предыдущих работах), чтобы определить, в каких функциональных областях генома (промоторы, интроны, межгенные пространства) они располагаются. Идентифицировать суперэнхансеры (super-enhancers) для данного типа клеток с использованием аннотации гистоновых меток.

Владеть: навыками работы с программами эпигенетического анализа, методами контроля качества биологических данных, инструментами визуализации эпигенетических профилей, практическими навыками обработки и интерпретации эпигенетических данных.

Задания, необходимые для оценивания сформированности СПК-2 на пороговом уровне

Перечень тем конспектов

1. Контроль качества и предобработка данных (Pre-processing)

- **Инструменты:** FastQC, Trim Galore, MultiQC.
- **Суть конспекта:** Изучение алгоритмов оценки качества прочтений, очистки от адаптеров и удаления низкокачественных нуклеотидов в форматах FASTQ для предотвращения ошибок на последующих этапах выравнивания.

2. Картирование и выравнивание эпигеномных данных

- **Инструменты:** Bismark (для метилирования), Bowtie2, BWA.
- **Суть конспекта:** Принципы индексации референсного генома и подходы к специфическому выравниванию бисульфитно-конвертированной ДНК (или

других эпигенетических меток) с учетом неполной гомологии последовательностей.

3. Анализ метилирования ДНК (DNA Methylation)

- **Инструменты:** methylKit, BiSeq, DMRcate.
- **Суть конспекта:** Определение уровня метилирования на уровне отдельных CpG-динуклеотидов, выявление дифференциально метилированных областей (DMRs), а также статистические методы поиска биологически значимых изменений.

4. Анализ гистоновых модификаций и транскрипционных факторов (ChIP-seq)

- **Инструменты:** MACS2, HOMER, DiffBind.
- **Суть конспекта:** Поиск сайтов связывания белков (пиков), вычитание фонового шума (control), а также сравнительный анализ дифференциального обогащения гистоновых меток между различными клеточными типами.

5. Изучение открытого хроматина (ATAC-seq и DNase-seq)

- **Инструменты:** Deeptools, ATACseqQC, HINT.
- **Суть конспекта:** Картирование профилей доступности хроматина, идентификация нуклеосомных позиций, обнаружение сайтов связывания регуляторных факторов и оценка активности промоторов.

6. Интеграция эпигеномных и транскриптомных данных

- **Инструменты:** EpiMix, epidecodeR.
- **Суть конспекта:** Изучение программ для совместного анализа профилей метилирования и уровней экспрессии генов с целью выявления функционального влияния эпигенетических меток на транскрипцию.

7. Поиск паттернов и машинное обучение в эпигенетике

- **Инструменты:** EpiGRAPH, ChromHMM, EpiExplorer.
- **Суть конспекта:** Использование методов кластеризации и скрытых марковских моделей (HMM) для аннотации состояний хроматина (например, разделение генома на активные, репрессированные и промоторные домены).

8. Эпигенетический анализ на уровне единичных клеток (scEpigenomics)

- **Инструменты:** Seurat, Signac, Cicero.

- **Суть конспекта:** Работа с массивами данных single-cell ATAC-seq и scRNA-seq, методы снижения размерности (UMAP, t-SNE) для кластеризации гетерогенных популяций клеток на основе их эпигенетического ландшафта.

9. Визуализация и интерпретация эпигеномных данных

- **Инструменты:** IGV (Integrative Genomics Viewer), UCSC Genome Browser.
- **Суть конспекта:** Практика визуализации эпигенетических треков, поиск перекрытий с известными регуляторными элементами (энхансеры, CTCF-сайты) и интеграция с базами данных ENCODE.

Задания, необходимые для оценивания сформированности СПК-2 на продвинутом уровне

Типовые задания лабораторных работ

1. **Первичный аудит сырых ридов и оценка качества (FastQC и MultiQC)**
 - **Задание:** Используя [FastQC](#), проанализировать FASTQ-файлы с результатами секвенирования библиотеки ChIP-seq (до и после иммунопреципитации). Построить сводный отчет с помощью MultiQC.
 - **Навык:** Выявление артефактов пробоподготовки (чрезмерная дупликация, адаптерные последовательности, низкое качество чтения на 3'-концах) для определения параметров обрезки (trimming).
2. **Оценка эффективности картирования и удаления дубликатов (BWA и Picard)**
 - **Задание:** Провести выравнивание ридов на эталонный геном (например, с помощью BWA-MEM) и удалить PCR-дубликаты с использованием утилиты [Picard MarkDuplicates](#). Рассчитать долю картированных ридов, уникальных фрагментов и коэффициент heredundancy (NRF).
 - **Навык:** Понимание того, как клональная амплификация влияет на сложность библиотеки и ограничивает реальное разрешение эпигенетического сигнала.
3. **Контроль качества обогащения и анализ соотношения сигнал/шум (FRiP и MACS3)**
 - **Задание:** Выполнить пик-коллинг с помощью [MACS3](#) для данных гистонных модификаций (например, H3K4me3). Вычислить долю ридов в пиках (FRiP score) и построить профили распределения (геномные треки).

- **Навык:** Оценка качества самого биологического эксперимента (IP-эффективности) — в высококачественных данных FRiP score должен быть не менее 1%, а для транскрипционных факторов — более 2%.
4. **Анализ профилей распределения и оценка геномного контекста (deepTools)**
- **Задание:** Используя пакет [deepTools](#), сгенерировать профили (heatmaps и метапрофили) распределения сигнала вокруг транскрипционных стадий (TSS) для разных меток.
 - **Навык:** Верификация биологической логики данных: например, обогащение H3K4me3 должно строго локализоваться вблизи промоторов, а H3K36me3 — в телах активных транскрибируемых генов.
5. **Оценка биологической репликации и геномной аннотации (ChIPseeker)**
- **Задание:** Загрузить данные из реплик одного эксперимента в R, используя пакет ChIPseeker, и рассчитать коэффициент корреляции между повторностями. Аннотировать выявленные пики относительно известных геномных структур (экзоны, интроны, промоторы, межгенные области).
 - **Навык:** Проверка воспроизводимости (согласованности) биологических реплик согласно стандартам [ENCODE](#) и подтверждение того, что пики картируются в ожидаемых регуляторных элементах.

Промежуточная аттестация

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

Знать: основные понятия эпигенетики и типы эпигенетических данных, принципы работы программных средств эпигенетического анализа, методы оценки качества и достоверности эпигенетических данных, источники ошибок и искажений в эпигенетических наборах данных.

Уметь: использовать программные инструменты для обработки эпигенетических данных, выполнять первичную фильтрацию и подготовку данных к анализу, проводить оценку качества эпигенетических данных, интерпретировать результаты эпигенетического анализа.

Владеть: навыками работы с программами эпигенетического анализа, методами контроля качества биологических данных, инструментами визуализации эпигенетических профилей, практическими навыками обработки и интерпретации эпигенетических данных.

Примеры заданий, необходимые для оценивания сформированности СПК-2

Задание 1. Первичный контроль качества (QC) данных секвенирования

Сценарий: Вы получили сырые файлы с результатами бисульфитного секвенирования (BS-seq) или ChIP-seq и должны оценить их пригодность.

- **Цель:** Найти низкокачественные риски, дубликаты и адаптеры для их последующей обрезки.
- **Инструменты:** [FastQC](#) (оценка) и [Trim Galore!](#) (очистка).
- **Практический шаг:** Запустите FastQC для исходного файла в формате FASTQ. Изучите отчет. Обратите внимание на графики *Per base sequence quality* и *Adapter Content*. Напишите скрипт для удаления адаптеров и ридов с качеством ниже Q30.
- **Оценка качества:** Сравните отчеты до и после обработки, чтобы убедиться, что артефакты устранены, а длина ридов осталась достаточной.

Задание 2. Анализ качества и нормализация данных микрочипов

Сценарий: У вас есть данные ДНК-метилирования на базе чипов Illumina EPIC. Вам нужно отфильтровать плохие пробы и технические шумы до начала поиска дифференциально метилированных позиций (DMP).

- **Цель:** Очистка данных массива от «грязных» образцов и зондов.
- **Инструменты:** Пакеты [Minfi](#) или [RnBeads](#) в среде R.
- **Практический шаг:** Загрузите набор сырых данных (IDAT-файлы). Вычислите интенсивность контрольных сигналов (бисульфитная конверсия, фоновый шум) с помощью встроенных функций `qc()`. Удалите пробы, где процент провалившихся зондов превышает порог 5% ($p\text{-value} > 0.01$).

Выполните квантильную нормализацию и оцените график плотности распределения бета-значений до и после нормализации.

Задание 3. Выявление дифференциально метилированных участков (DMR) и их аннотация

Сценарий: Требуется найти эпигенетические изменения (например, связанные со старением) между двумя группами образцов.

- **Цель:** Найти участки ДНК с достоверно разным уровнем метилирования и связать их с генами.
- **Инструменты:** Пакет [DMRcate](#) или [ChIPseeker](#) в R.
- **Практический шаг:** После завершения шагов контроля качества (из Задания 2) примените линейные модели для выявления позиций с изменением метилирования. Используйте метод DMRcate для объединения этих позиций в блоки (DMR). Проведите аннотацию, чтобы узнать, в каких именно функциональных областях генома (промоторы, экзоны, энхансеры) произошли изменения.

Совет: Для отработки навыков используйте публичные эпигенетические датасеты, которые легко скачать из базы данных [NCBI GEO](#). Дополнительно можно использовать пакет EpiCompare для быстрого визуального бенчмаркинга качества разных датасетов.

СПК-4. Способен проводить экспертно-аналитическую работу при проведении научных исследований и экспериментальных работ

Знать: методы экспертного анализа научных и экспериментальных данных, принципы организации и проведения научных исследований, основы статистической обработки биологических данных, требования к оформлению научно-аналитических материалов.

Уметь: проводить анализ и интерпретацию результатов научных исследований, применять методы статистической обработки данных, выявлять закономерности и

аномалии в экспериментальных данных, оформлять аналитические отчёты по результатам исследований.

Владеть: навыками использования вычислительной техники в биоинформатике, методами обработки и анализа экспериментальных данных, инструментами автоматизации обработки биологических данных.

Примеры заданий, необходимые для оценивания сформированности СПК-4

Задание 1. Поиск сайтов связывания с транскрипционными факторами (Position Weight Matrices)

- **Тема:** Статистические модели в биоинформатике.
- **Суть:** Напишите программу, которая принимает на вход выравнивание известных сайтов связывания белка (например, промоторов) и строит позиционную весовую матрицу (PWM). Программа должна сканировать этой матрицей новую геномную последовательность, рассчитывать p -value для каждого окна и находить наиболее вероятные сайты связывания с учетом фонового распределения нуклеотидов.
- **Навыки:** Работа с матрицами и двумерными массивами, вероятностные расчеты, функции скользящего окна (sliding window).

Задание 2. Реализация алгоритма локального выравнивания (Смита-Ватермана)

- **Тема:** Динамическое программирование.
- **Суть:** Реализуйте алгоритм Смита-Ватермана для поиска наиболее похожих участков между двумя длинными строками (ДНК или аминокислотными последовательностями). Программа должна заполнять матрицу штрафов/наград (например, используя матрицу BLOSUM62 или простую схему: +2 за совпадение, -1 за несовпадение, -2 за гэп) и выполнять обратный трассинг (traceback) для получения оптимального локального выравнивания.
- **Навыки:** Вложенные циклы, оптимизация памяти (расчет текущей строки через предыдущую), работа с файлами в формате FASTA.

Задание 3. Эвристический поиск: базовый аналог BLAST (k-мерные индексы)

- **Тема:** Эвристики и структуры данных.
- **Суть:** Базовые алгоритмы выравнивания слишком медленные для больших баз данных. Создайте программу-индексатор, которая разбивает геном на короткие слова фиксированной длины (k -меры, например, $k=11$) и строит хеш-таблицу с позициями этих слов в геноме. Затем реализуйте алгоритм поиска совпадений (seeds) для быстрого нахождения потенциально похожих последовательностей перед их полноценным выравниванием.
- **Навыки:** Использование хеш-таблиц (словей/словарей), работа с памятью, базовые принципы эвристического поиска.

Задание 4. Парсинг, фильтрация и анализ BAM-файлов

- **Тема:** Форматы биологических данных и работа с библиотеками.
- **Суть:** Используя специализированную библиотеку (например, `pysam` для Python), напишите скрипт для обработки высокопроизводительных данных секвенирования (NGS). Программа должна открывать BAM-файл, отфильтровывать прочтения (reads) с низким качеством картирования (mapping quality) и идентифицировать регионы с аномально высокой или низкой глубиной покрытия (coverage). Дополнительно можно реализовать выявление участков с высокой плотностью SNPs (однонуклеотидных полиморфизмов).
- **Навыки:** Работа с библиотеками для биоинформатики, парсинг специфических форматов файлов (SAM/BAM), агрегация и статистический анализ данных.

Для выполнения данных задач студентам лучше всего порекомендовать использование **Python** (из-за развитой экосистемы библиотек, таких как `Biopython` и `pysam`)

Примерные вопросы к зачету

1. Каковы основные различия между интерпретируемыми (Python) и компилируемыми (C/C++) языками с точки зрения биоинформатических вычислений?
2. Какие структуры данных в Python наиболее эффективны для хранения табличных данных (например, аннотации генов) и почему?
3. Объясните концепцию ленивых вычислений (generators) на примере обработки больших файлов формата FASTQ.
4. Как регулярные выражения (RegEx) применяются для поиска сайтов рестрикции или мотивов в белковых последовательностях?
5. Напишите код для эффективного построчного чтения FASTA-файла, чтобы не перегружать оперативную память.
6. В чем разница между list, set и dict в Python при поиске уникальных идентификаторов генов? Приведите примеры использования.
7. Как осуществляется взаимодействие скрипта на Python с базами данных NCBI через Entrez (модуль Bio.Entrez)?
8. Опишите структуру объекта SeqRecord в библиотеке Biopython и приведите пример извлечения информации о координатах гена.
9. Как в Python реализуется парсинг файлов формата GFF3/GTF для анализа структуры транскриптов?
10. Какие библиотеки Python используются для визуализации геномных данных (например, для построения heatmap или геномных треков) и каковы их особенности?
11. Почему язык R является стандартом де-факто для статистического анализа в биоинформатике?
12. Объясните разницу между объектами типа vector, matrix, list и data.frame в R.
13. Что такое tibble и чем он удобнее классического data.frame при работе с таблицами экспрессии генов?

14. Как работает концепция "сплит-применение-комбинирование" (split-apply-combine) при использовании пакета dplyr?
15. Какие функции в R используются для транспонирования и объединения матриц экспрессии (например, merge, join)?
16. Как написать функцию apply в R для вычисления среднего уровня экспрессии по всем генам (строкам матрицы)?
17. Каким образом ggplot2 использует грамматику графиков для создания точечных диаграмм (scatterplot) с наложенными линиями тренда?
18. Как выполнить множественное тестирование гипотез (Multiple Testing Correction) в R для контроля уровня ложноположительных результатов (FDR) при RNA-seq?
19. Объясните процесс кластеризации данных экспрессии (иерархическая кластеризация и метод k-средних) с помощью встроенных функций R.
20. Как осуществляется визуализация результатов дифференциальной экспрессии с помощью графика (Volcano plot) в R?
21. Какие алгоритмы лежат в основе парсинга файлов формата BAM и как извлечь информацию о выравнивании прочтений в Python (библиотека pysam)?
22. Какова архитектура создания воспроизводимых биоинформатических пайплайнов с использованием систем управления рабочими процессами (Nextflow или Snakemake)?
23. Как в рамках пайплайна обрабатываются пропуски, дубликаты и нулевые значения (zero-inflated data) в матрицах RNA-seq?
24. Объясните принцип работы алгоритма поиска гомологов с помощью BLAST. Как автоматизировать запуск локального BLAST и парсить XML-вывод в Python?
25. Как реализовать поиск SNPs (однонуклеотидных полиморфизмов) по файлу формата VCF с использованием командной строки и Python/R?
26. В чем заключаются особенности объектной модели gRanges в R/Bioconductor и как с ее помощью выполнять пересечение геномных интервалов (bedtools-подобные операции)?

27. Какие существуют способы аннотации генов (например, конвертация Ensembl ID в Gene Symbol) с использованием пакетов `org.hs.eg.db` в R?
28. Как осуществляется контроль качества (Quality Control) прочтений FASTQ (например, с использованием FastQC) и фильтрация данных в автоматизированном скрипте?
29. Что такое функциональное обогащение (Gene Ontology Enrichment analysis) и как его выполнить средствами пакета `clusterProfiler` в R?
30. Опишите базовый алгоритм трансляции ДНК в белок с учетом открытых рамок считывания (ORF) и возможность его реализации на Python.

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

Формами текущего контроля являются тестирование, лабораторная работа, конспект.

Общее количество баллов по дисциплине - 100 баллов.

Максимальное количество баллов, которое можно набрать в течение семестра за выполнение практических работ, написание конспектов и тестирования - 80 баллов.

Формой промежуточной аттестации является зачет. Зачет проходит в устной форме по вопросам.

Структура оценивания зачета

Критерии оценивания	Баллы
Ставится, если студент обнаруживает всестороннее, систематическое и глубокое знание программного материала по дисциплине; обстоятельно анализирует структурную взаимосвязь рассматриваемых тем и разделов дисциплины; усвоил основную и знаком с дополнительной литературой, рекомендованной программой, а также усвоил взаимосвязь основных	16-20

Критерии оценивания	Баллы
понятий дисциплины в их значении для приобретаемой профессии; проявил творческие способности в понимании, изложении и использовании учебного материала.	
Ставится, если студент, обнаруживает полное знание программного материала, успешно выполняет предусмотренные в программе задания; усвоил основную литературу, рекомендованную в программе; показал систематический характер знаний по дисциплине и способен к их самостоятельному пополнению и обновлению в ходе дальнейшей образовательной деятельности.	11-15
Ставится, если студент обнаруживает знание основного программного материала в объеме, необходимом для дальнейшего обучения и профессиональной деятельности; справляется с выполнением заданий, предусмотренных программой; знаком с основной литературой, рекомендованной программой; допускает погрешности не принципиального характера в ответе на экзамене.	5-10
Ставится в том случае, если студент обнаруживает пробелы в знаниях основного программного материала, допускает принципиальные ошибки в выполнении предусмотренных программой заданий.	0-4

Итоговая шкала оценивания результатов освоения дисциплины

Итоговая оценка по дисциплине формируется из суммы баллов по результатам текущего контроля и промежуточной аттестации и выставляется в соответствии с приведенной ниже таблицей.

Оценка по 100-балльной системе	Оценка по традиционной системе
81 – 100	отлично
61 - 80	хорошо
41 - 60	удовлетворительной
0 - 40	неудовлетворительно