

**АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ
ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАУЧНО-ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «СИРИУС»
(АНОО ВО «УНИВЕРСИТЕТ «СИРИУС»)**

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

«Программные средства анализа и визуализации данных»

Уровень образования: высшее образование – программа специалитета

Специальность: 06.05.01 Биоинженерия и биоинформатика

Направленность (профиль): Биоинженерия

1. Трудоемкость дисциплины (модуля): 2 з.е.

2. Место дисциплины в учебном плане: дисциплина «Программные средства анализа и визуализации данных» входит в Блок 1. «Дисциплины (модули)», обязательную часть, раздел «Профессиональная подготовка» и изучается в 9-10 модулях (5 семестр).

3. Цель дисциплины (модуля): сформировать у обучающихся представления о принципах эффективной визуализации биологических и табличных данных, научить выбирать и создавать графики, диаграммы, дашборды и иллюстрации, соответствующие научному контексту, аудитории и типу данных.

4. Задачи дисциплины (модуля):

- Освоение инструментов для работы с данными: изучение специализированного ПО для анализа биологических данных.
- Развитие умений обработки и анализа биомедицинских данных.
- Формирование навыков создания информативных графиков и дашбордов для визуализации.

5. Перечень разделов (тем) дисциплины и их краткое содержание:

В результате освоения дисциплины обучающиеся получат навыки работы с различным программным обеспечением для визуализации биологических данных и результатов анализа.

Раздел	Содержание
Раздел 1. Основы визуализации и дизайна	Введение в цели визуализации, визуальное восприятие, оценка качества графиков, принципы минимализма, выбор цвета и формы
Раздел 2. Визуализация одномерных и табличных данных	Способы представления числовых и категориальных данных, гистограммы, боксплоты, линейные и точечные графики, визуализация временных рядов
Раздел 3. Многомерные данные и кластеризация	Методы представления многомерных данных (PCA, UMAP, t-SNE), построение тепловых карт, кластеризация образцов, визуализация матриц расстояний
Раздел 4. Визуализация графов и деревьев	Основы визуализации связей и взаимодействий между объектами (гены, белки, клетки), построение и аннотирование филогенетических деревьев
Раздел 5. Интерактивные графики и дашборды	Изучение инструментов для создания интерактивных визуализаций данных. Визуализация с помощью Plotly как основы интерактивных графиков. Построение полноценных дашбордов с помощью R Shiny/Dash/Tableau для комплексной демонстрации результатов анализа
Раздел 6. Научно-графический дизайн и визуальные тезисы (visual abstract)	Создание иллюстраций и инфографики для научной коммуникации с помощью BioRender и других средств. Обсуждение критериев качества постера для конференции

6. Образовательные результаты освоения дисциплины (модуля):

Формируемые компетенции (код компетенции, формулировка)	Планируемые результаты обучения по дисциплине (индикаторы достижения компетенций)
ОПК-5. Способен находить и использовать информацию, накопленную в базах данных по биологическим объектам, включая нуклеиновые кислоты и белки, владеть основными биоинформационическими средствами анализа	ИОПК-5.1 Умеет использовать биологические профессиональные базы данных и справочные системы
	ИОПК-5.2 Применяет принципы анализа научной информации, извлеченной из баз данных по биологическим объектам при решении профессиональных задач деятельности в области биологических наук

ОПК-7. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ИОПК-7.1 Знает основные принципы работы современных ИТ ИОПК-7.2 Применяет ИТ для решения задач в области биоинформатики и биоинженерии
--	---

7. Оценочные и методические материалы

7.1. Оценочные материалы для организации текущего контроля

Контрольные работы (КР1 – КР8)

- Форма: письменная, синхронная
- Место и время проведения: во время контактной работы в аудитории, согласно расписанию

- Примеры заданий (по парам):

КР1:

1. Какая библиотека Python используется для чтения файлов в формате FASTA?

- Pandas
- Biopython
- NumPy
- Matplotlib

2. Какой командой в R можно удалить строки с пропущенными значениями из датафрейма df?

- drop.na(df)
- na.omit(df)
- clean(df)
- remove.na(df)

2. Для чего используется инструмент FastQC?

- Контроль качества данных NGS
- Выравнивание последовательностей
- Визуализация белковых структур
- Кластеризация генов

3. Какая функция в R используется для поиска дифференциально экспрессируемых генов?

- prcomp()
- DESeq()
- heatmap()
- t.test()

КР2:

1. Визуализация 3D-структуры белка

- Задание:
- Загрузите PDB-файл белка (например, 1CRN).
- Подсветьте активный сайт и лиганд.
- Экспортируйте изображение в высоком разрешении.
- Инструменты: PyMOL, ChimeraX
- Выходные данные: 3D-изображение в PNG.

2. Анализ RNA-seq данных

- Задание:
- Загрузите данные RNA-seq из GEO (например, GSE12345).

- Проведите нормализацию (TPM, FPKM).
 - Постройте heatmap для топ-100 дифференциально экспрессируемых генов.
 - Инструменты: DESeq2 (R), Scanpy (Python), ComplexHeatmap
 - Выходные данные: Heatmap, таблица с DEG.
4. Какой инструмент используется для аннотации генов?

Варианты:

- DESeq2,
- ANNOVAR,
- Bowtie2.

5. Как визуализировать сеть белок-белковых взаимодействий?

Варианты:

- Matplotlib,
- Cytoscape,
- ggplot2.

Критерии оценки:

1. Корректность выполнения заданий — до 4 баллов.
2. Полнота и логика — до 4 баллов.

7.2. Оценочные материалы для организации промежуточной аттестации

- Форма проведения: письменная (синхронная), в очном формате в зависимости от расписания. Проведение осуществляется в 5 семестре (9 и 10 модуль дисциплины соответственно).

- Место проведения: учебная аудитория

Пример экзаменационного задания:

1. Какой тип графика лучше всего подходит для отображения корреляции между двумя переменными?

- Столбчатая диаграмма
- Точечный график (scatter plot)
- Круговая диаграмма
- Боксплот

2. Какая библиотека Python позволяет создавать интерактивные графики?

- Matplotlib
- Plotly
- NumPy
- SciPy

3. Какой пакет Python используется для кластеризации одноклеточных данных?

- Biopython
- Scanpy
- PyMOL
- Seaborn

4. Какой инструмент используется для визуализации белковых структур?

- DESeq2
- PyMOL
- Pandas
- FastQC

5. Какой командой в R можно создать боксплот для данных в dataфрейме df?

- `plot(df)`
- `boxplot(df$variable ~ df$group)`
- `hist(df$variable)`

- `barplot(df$variable)`

Критерии оценивания:

- Корректность выполнения заданий — до 10 баллов.

7.3. Методические рекомендации

Обучение по дисциплине предполагает изучение курса на аудиторных занятиях (практические занятия) и в ходе самостоятельной работы студентов. Студентам необходимо ознакомиться с содержанием рабочей программы дисциплины, с целями и задачами дисциплины, ее связями с другими дисциплинами образовательной программы, методическими разработками по данной дисциплине.

Обучение по дисциплине проводится последовательно путем проведения практических занятий с углублением и закреплением полученных знаний в ходе самостоятельной работы с последующим переводом знаний в умения в ходе практических занятий. Получение углубленных знаний по изучаемой дисциплине достигается за счет дополнительных часов к аудиторной работе самостоятельной работы студентов. Выделяемые часы целесообразно использовать для знакомства с дополнительной научной литературой по проблематике дисциплины, анализа научных концепций и современных подходов к осмысливанию рассматриваемых проблем. К самостоятельному виду работы студентов относится работа в библиотеках, в электронных поисковых системах и т.п. по сбору материалов, необходимых для проведения практических занятий или выполнения конкретных заданий преподавателя по изучаемым темам. Обучающиеся могут установить электронный диалог с преподавателем, выполнять посредством него контрольные задания.