

**Федеральное государственное автономное образовательное учреждение
высшего образования
«Санкт-Петербургский политехнический университет Петра Великого»**

УТВЕРЖДАЮ
Директор ИБСиБ
_____ А.В. Васин
«14» июня 2024 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

«Высокопроизводительные методы получения данных»

Разработчик	Высшая школа биомедицинских систем и технологий
Направление (специальность) подготовки	06.05.01 Биоинженерия и биоинформатика
Наименование ООП	06.05.01_01 Биоинженерия и биоинформатика
Квалификация (степень) выпускника	биоинженер и биоинформатик
Образовательный стандарт	СУОС
Форма обучения	Очная

СОГЛАСОВАНО
Руководитель ОП
_____ Д.И. Богомаз
«15» апреля 2025 г.

Соответствует СУОС
Утверждена протоколом заседания
высшей школы "ВШБСиТ"
от «15» апреля 2025 г. № 6

РПД разработал:
Доцент, к.б.н. Д.И. Богомаз

1. Цели и планируемые результаты изучения дисциплины

Цели освоения дисциплины

Изучение дисциплины "Высокопроизводительные методы получения данных" направлено на формирование современных представлений о предмете и основных концепциях биоинформатики, задачах высокопроизводительных методов анализа данных, возможностях приложения биоинформационных подходов к решению фундаментальных и прикладных проблем молекулярной биологии, молекулярной генетики, клеточной биологии, биомедицины, фармакологии.

Результаты обучения выпускника

Код	Результат обучения (компетенция) выпускника ООП
ОПК-6	Способен создавать компьютерные программы, используемые в биоинформатике и биоинженерии
ИД-4 ОПК-6	Создает компьютерные программы на языке R, используемые в биоинформатике и биоинженерии
ОПК-7	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности
ИД-1 ОПК-7	Применяет принципы работы современных информационных технологий в профессиональной деятельности.
ИД-2 ОПК-7	Создает и управляет современными информационными технологиями и потоками информации в том числе и для решения задач профессиональной деятельности.

Планируемые результаты изучения дисциплины

знания:

- Знание способов создания компьютерных программ на языке R, используемые в биоинформатике и биоинженерии
- Знание принципов работы современных информационных технологий в профессиональной деятельности.
- Знание методов создания и управления современными информационными технологиями и потоками информации в том числе и для решения задач профессиональной деятельности.

умения:

- Умение создавать компьютерные программы на языке R, используемые в биоинформатике и биоинженерии

- Умение использовать принципы работы современных информационных технологий в профессиональной деятельности.
- Умение создавать и управлять современными информационными технологиями и потоками информации в том числе и для решения задач профессиональной деятельности.

навыки:

- Владение навыками создания компьютерных программ на языке R, используемых в биоинформатике и биоинженерии
- Владение способами применения принципов работы современных информационных технологий в профессиональной деятельности.
- Владение способами создания и управления современными информационными технологиями и потоками информации, в том числе и для решения задач профессиональной деятельности.

2. Место дисциплины в структуре ООП

В учебном плане дисциплина «Высокопроизводительные методы получения данных» относится к модулю «Модуль цифровых компетенций (Digital)».

Изучение дисциплины базируется на результатах освоения следующих дисциплин:

- Вычислительная математика

3. Распределение трудоёмкости освоения дисциплины по видам учебной работы и формы текущего контроля и промежуточной аттестации

3.1. Виды учебной работы

Виды учебной работы	Трудоёмкость по семестрам
	Очная форма
Лекционные занятия	14
Лабораторные занятия	30
Самостоятельная работа	55
Часы на контроль	5
Промежуточная аттестация (зачет)	4
Общая трудоёмкость освоения дисциплины	108, ач
	3, зет

3.2. Формы текущего контроля и промежуточной аттестации

Формы текущего контроля и промежуточной аттестации	Количество по семестрам
	Очная форма
Промежуточная аттестация	
Зачеты, шт.	1

4. Содержание и результаты обучения

4.1 Разделы дисциплины и виды учебной работы

№ раздела	Разделы дисциплины, мероприятия текущего контроля	Очная форма		
		Лек, ач	Лаб, ач	СР, ач
1.	Введение в дисциплину	2	2	9
2.	Методы получения и анализа геномных данных	2	4	8
3.	Методы получения и анализа метагеномных данных.	2	4	6
4.	Методы получения и анализа транскриптомных данных.	2	4	8

5.	Методы получения и анализа протеомных данных.	2	4	8
6.	Методы получения и анализа метаболомных данных.	2	6	8
7.	Биоинформационная поддержка эксперимента	2	6	8
Итого по видам учебной работы:		14	30	55
Зачеты, ач				5
Часы на контроль, ач				5
Промежуточная аттестация (зачет)		4		
Общая трудоёмкость освоения: ач / зет		108 / 3		

4.2. Содержание разделов и результаты изучения дисциплины

Раздел дисциплины	Содержание
1. Введение в дисциплину	История развития высокопроизводительных методов получения данных . Основные задачи и высокопроизводительных методов получения данных .
2. Методы получения и анализа геномных данных	Основное приборное обеспечение для получения и анализа геномных данных, способы биоинформатической обработки и хранения этих данных.
3. Методы получения и анализа метагеномных данных.	<p>Основное приборное обеспечение для получения и анализа метагеномных данных, способы биоинформатической обработки и хранения этих данных. Микробные сообщества, показатели альфа-, бета-разнообразия. Маркерные гены идентификации микроорганизмов. Методы оценки микробного разнообразия: классические (посевы на селективные среды), ПЦР-анализ, применение технологий секвенирования первого и второго поколений. Метагеном. Шотган-секвенирование и секвенирование ампликонов 16S rRNA/ITS. Алгоритмы метагеномного анализа. Базы референсных последовательностей (Greengenes, SILVA, MetaPhlan). QIIME - инструмент для анализа ампликонов маркерных генов. Понятие "OTU", индексы альфа-разнообразия, матрицы расстояний бета-разнообразия.</p> <p>Визуализация разнообразия методом MDS. Анализ данных шотган-секвенирования. Таксономическое разнообразие (MetaPhlan), функциональное разнообразие (ChocoPhlan). Сборка метагеномов. Задачи, решаемые путем метагеномного анализа. Проект "Human Microbiome Project". Резистом микробиоты кишечника человека. Взаимосвязь микробного сообщества с аутоиммунными заболеваниями.</p>

4. Методы получения и анализа транскриптомных данных.	<p>Основное приборное обеспечение для получения и анализа транскриптомных данных, способы биоинформатической обработки и хранения этих данных. Анализ экспрессии генов при помощи микрочипов и RNA-seq. Типы РНК, область применения разных RNA-seq.</p> <p>Пайплайн транскриптомного анализа (выравнивание, сборка транскриптов, анализ дифференциальной экспрессии).</p> <p>Нормализация (RPKM, FPKM, CPM). Статистический анализ полученных профилей экспрессии в R (FDR, MDS, Volcano plot).</p> <p>Обогащение терминами Gene Ontology (Web Gestalt), KEGG.</p> <p>Визуализация генных сетей в Strings.</p>
5. Методы получения и анализа протеомных данных.	<p>Основное приборное обеспечение для получения и анализа протеомных данных, способы биоинформатической обработки и хранения этих данных. Масс-спектрометрия и чипы. Программы, позволяющие из фрагментарных данных масс-спектрометрии и чипов</p> <p>выдавать данные о почти полностью собранных из этих фрагментов белков. Программы, основанные на построении выравниваний фрагментов с известными белками из баз данных UniProt и PROSITE. Существующие инструменты, определяющие посттрансляционные модификации.</p> <p>Вычислительные методы биоинформатики для изучения белков-биомаркеров. Протеогеномика - методы протеомики для подтверждения данных, полученных из геномных последовательностей. Структурная протеомика - широкомасштабные исследования структур белков на основе данных рентгеноструктурного анализа и ЯМР-спектроскопии.</p>
6. Методы получения и анализа метаболомных данных.	<p>Основное приборное обеспечение для получения и анализа метаболомных данных, способы биоинформатической обработки и хранения этих данных.</p>
7. Биоинформационная поддержка эксперимента	<p>Систематизация и сравнение разных типов данных. Алгоритмы моделирования генных и метаболомных сетей. Особенности реализации генетической информации и отражение этого процесса в разных типах данных.</p>

5. Образовательные технологии

При освоении дисциплины предусматривается широкое использование активных и интерактивных форм приобретения новых знаний. В курсе запланирована активная работа

магистрантов в компьютерном классе с доступом в Интернет для работы с биоинформационными банками данных, на биоинформационных порталах с программным инструментарием извлечения информации и ее анализа. Магистранты на примерах конкретных исследовательских задач геномики и протеомики разбирают применение вычислительных методов, отрабатывают навыки применения компьютерных методов в биологии.

6. Лабораторный практикум

№ раздела	Наименование лабораторных работ	Трудоемкость, ач
		Очная форма
1.	Средства работы с банками данных. Практикум в Entrez	6
2.	Практикум по выравниваниям: построение выравниваний (написание и использование программ), сравнение локальных и глобальных выравниваний, зависимость выравнивания от параметров, оценка статистической значимости	6
3.	Анализ протеомных баз данных, получение протеомных данных.	6
4.	Анализ метагеномов, выводы о структуре и численности.	6
5.	Анализ экзомов, методы и подходы. Базы данных экзомов.	6
Итого часов		30

7. Практические занятия

Не предусмотрено

8. Организация и учебно-методическое обеспечение самостоятельной работы

Примерное распределение времени самостоятельной работы студентов

Вид самостоятельной работы	Примерная трудоемкость, ач
	Очная форма
Текущая СР	
работа с лекционным материалом, с учебной литературой	2
опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	2
самостоятельное изучение разделов дисциплины	8
выполнение домашних заданий, домашних контрольных работ	8
подготовка к лабораторным работам, к практическим и семинарским занятиям	8
подготовка к контрольным работам, коллоквиумам	3
Итого текущей СР:	31
Творческая проблемно-ориентированная СР	
выполнение расчётно-графических работ	0
выполнение курсового проекта или курсовой работы	0
поиск, изучение и презентация информации по заданной проблеме, анализ научных публикаций по заданной теме	8
работа над междисциплинарным проектом	0
исследовательская работа, участие в конференциях, семинарах, олимпиадах	0
анализ данных по заданной теме, выполнение расчётов, составление схем и моделей на основе собранных данных	16
Итого творческой СР:	24
Общая трудоемкость СР:	55

9. Учебно-методическое обеспечение дисциплины

9.1. Адрес сайта курса

<https://dl-ibmst.spbstu.ru/>

9.2. Рекомендуемая литература

Основная литература

№	Автор, название, место издания, издательство, год (годы) издания	Год изд.	Источник
1	Осенняя Д.И., Дробинцев П.Д. Адаптация инструмента поиска мутаций в геноме для технологии Spark, 2016. URL: http://elib.spbstu.ru/dl/2/v16-2299.pdf	2016	ЭБ СПбПУ

Дополнительная литература

№	Автор, название, место издания, издательство, год (годы) издания	Год изд.	Источник
1	Власьев И.М. Разработка веб-сервиса анализа и визуализации геномных данных, 2015. URL: http://elib.spbstu.ru/dl/2/8594.pdf	2015	ЭБ СПбПУ

Ресурсы Интернета

1. <https://www.bioinformatics.org/>: <https://www.bioinformatics.org/>

9.3. Технические средства обеспечения дисциплины

Освоение дисциплины "Вычислительная геномика и протеомика" предполагает использование следующего материально-технического обеспечения:

Мультимедийная аудитория, вместимостью более 60 человек. Мультимедийная аудитория состоит из интегрированных инженерных систем с единой системой управления, оснащенная современными средствами воспроизведения и визуализации любой видео и аудио информации, получения и передачи электронных документов, Мультимедийная аудитория также оснащена широкополосным доступом в сеть интернет. Компьютерное оборудование имеет соответствующее лицензионное программное обеспечение.

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Для проведения занятий необходима мультимедийная аудитория и компьютерный класс с выходом в Интернет.

10. Материально-техническое обеспечение дисциплины

Освоение дисциплины "Вычислительная геномика и протеомика" предполагает использование следующего материально-технического обеспечения:

Мультимедийная аудитория, вместимостью более 60 человек. Мультимедийная аудитория состоит из интегрированных инженерных систем с единой системой управления, оснащенная современными средствами воспроизведения и визуализации любой видео и аудио информации, получения и передачи электронных документов, Мультимедийная аудитория также оснащена широкополосным доступом в сеть интернет. Компьютерное оборудование имеет соответствующее лицензионное программное обеспечение.

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Для проведения занятий необходима мультимедийная аудитория и компьютерный класс с выходом в Интернет.

11. Критерии оценивания и оценочные средства

11.1. Критерии оценивания

Для дисциплины «Высокопроизводительные методы получения данных» формой аттестации является зачёт. Оценивание качества освоения дисциплины производится с использованием рейтинговой системы.

Зачёт

Для получения зачёта необходимо набрать минимум 50 баллов из 100.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

11.2. Оценочные средства

Оценочные средства по дисциплине представлены в фонде оценочных средств, который является резервной частью основной образовательной программы и размещается в электронной информационно-образовательной среде СПбПУ на портале etk.spbstu.ru

12. Методические рекомендации по организации изучения дисциплины

Дисциплина, прежде всего, подразумевает наряду с теоретическими знаниями формирование у студентов устойчивых практических навыков. Преподаватель должен ориентировать студентов

на полное решение практических задач, которые возникают у биоинформатика-практика, от сборки геномов, до анализа клинически значимых SNP.

Целью курса не является создание и разработка новых алгоритмов или построения теоретической биоинформатики, однако в практических реализациях "простых" биоинформатических задач, курс должен быть самодостаточным.

Курс строится на знаниях, полученных студентами на дисциплинах, как биологического, так и биоинформатического профиля. В ходе прохождения курса необходимо акцентировать внимание студентов на знания и сведения, из смежных дисциплин.

В ходе курса требуется проводить контроль текущей успеваемости, проводить контрольные и проверочные работы, коллоквиумы по наиболее сложным темам, однако не надо забывать, что главной целью этого курса является формирование целостных навыков работы в системе биологические данные-ПО- база последовательностей для получения принципиально новых данных.

13. Адаптация рабочей программы для лиц с ОВЗ

Адаптированная программа разрабатывается при наличии заявления со стороны обучающегося (родителей, законных представителей) и медицинских показаний (рекомендациями психолого-медико-педагогической комиссии). Для инвалидов адаптированная образовательная программа разрабатывается в соответствии с индивидуальной программой реабилитации.