

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

УТВЕРЖДАЮ
Директор ИБСиБ
_____ А.В. Васин
«30» мая 2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

«Вычислительная геномика»

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 "Вычислительная геномика и протеомика" направлено на формирование современных представлений о предмете и основных концепциях биоинформатики, задачах вычислительной геномики и протеомики, 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
ПК-1	Способен разрабатывать новые алгоритмы обработки цифровой биологической информации, в том числе больших данных
ИД-1 ПК-1	Разрабатывает новые алгоритмы обработки цифровой биологической информации, при анализе геномов.

Планируемые результаты изучения дисциплины

знания:

- Знание новых алгоритмов обработки цифровой биологической информации, при анализе геномов.

умения:

- Умение разрабатывать новые алгоритмы обработки цифровой биологической информации, при анализе геномов.

навыки:

- Владение методиками разработки новых алгоритмов обработки цифровой биологической информации, при анализе геномов.

2. Место дисциплины в структуре ООП

В учебном плане дисциплина «Вычислительная геномика» относится к модулю «Модуль цифровых компетенций (Digital)».

Изучение дисциплины базируется на результатах освоения следующих дисциплин:

- Высокопроизводительные методы получения данных
- Вычислительная математика
- Компьютерная метаболомика

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	30
Лабораторные занятия	30
Самостоятельная работа	80
Промежуточная аттестация (зачет)	4
Общая трудоёмкость освоения дисциплины	144, ач
	4, 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 вычислительную геномику и протеомику	3	3	10
2.	Биоинформационные базы данных	3	3	10
3.	Сравнение полинуклеотидных и полипептидных последовательностей	4	4	10
4.	Автоматическое аннотирование последовательности. Онтология.	4	4	10

5.	Молекулярная эволюция.	4	4	10
6.	Статистика последовательностей ДНК	4	4	10
7.	Методы вычислительной геномики	4	4	10
8.	Биоинформационная поддержка эксперимента	4	4	10
Итого по видам учебной работы:		30	30	80
Зачеты, ач				0
Часы на контроль, ач				0
Промежуточная аттестация (зачет)		4		
Общая трудоёмкость освоения: ач / зет		144 / 4		

4.2. Содержание разделов и результаты изучения дисциплины

Раздел дисциплины	Содержание
1. Введение в вычислительную геномику и протеомику	История развития биоинформационных методов. Основные задачи и методы биоинформатики в исследовании генома и протеома
2. Биоинформационные базы данных	Основы структур баз данных (записи, поля, объекты) классификация баз по способу заполнения (автоматические, архивные, курируемые) основные базы данных: GenBank, EMBL SwissProt, TrEMBL, PIR PDB базы, содержащие результаты глобальных экспериментов по анализу экспрессии, протеомике, и т.п. банки белковых семейств (SCOP, Prosite, ProDom, PFAM, InterPro) метаболические базы данных генетические банки (физические карты, OMIM) специализированные банки данных конкретные белковые семейства, РНК и т.д. конкретные геномы функциональные сайты в белках и ДНК
3. Сравнение полинуклеотидных и полипептидных последовательностей	Сравнение последовательностей выравнивание двух последовательностей глобальное и локальное выравнивание, вес выравнивания, матрицы аминокислотных замен дот-матрицы глобальное выравнивание: алгоритм Нидельмана-Вунша локальное выравнивание: алгоритм Смита-Ватермана другие алгоритмы локального выравнивания другие варианты выравнивания (fitting, overlaps, блочное выравнивание, сплайсированное выравнивание) статистическая значимость выравниваний и ее зависимость от вероятностной модели последовательности (в т.ч. сегменты малой сложности) зависимость выравнивания от параметров множественное выравнивание динамическое программирование последовательное выравнивание (Clustal) другие алгоритмы множественного выравнивания (DIALIGN, Match-Box, алгоритм Леонтовича-Бродского) профили, скрытые марковские модели поиск блоков (максимизация ожидания, Gibbs sampler, имитация теплового отжига и т.д.)
4. Автоматическое аннотирование последовательности. Онтология.	Поиск гомологов в базах данных, интерпретация результатов, сравнение алгоритмов, зависимость от параметров (матрицы, фильтры и т.п.)

5. Молекулярная эволюция.	Эволюция молекул и организмов, ортологи и паралоги, горизонтальный перенос, деревья видов и деревья генов Филогенетическое дерево как математический объект Модели эволюции алгоритмы построения филогенетических деревьев, матрица расстояний методы, основанные на матрице расстояний (UPGMA, neighbour-joining, minimal evolution, топологические инварианты и др.), другие методы (максимальная экономия, максимальное правдоподобие) алгоритмические проблемы поиска оптимального дерева, bootstrapping, согласование деревьев Эволюция на уровне генома (синтения, хромосомные перестройки)
6. Статистика последовательностей ДНК	Нуклеотидный состав (изохоры, GC-острова, картирование старта репликации) частые и редкие слова (вероятностные проблемы) статистика ДНК как характеристика генома
7. Методы вычислительной геномики	Метаболическая реконструкция (в т.ч. неортологичные замещения), позиционный анализ, эволюция регуляторных взаимодействий, эволюция белковых семейств, их доля в геноме
8. Биоинформационная поддержка эксперимента	Подбор праймеров для ПЦР Анализ данных двумерного фореа (обработка изображений), масс-спектрометрия белков Анализ данных по экспрессии генов (microarrays и др.), обработка изображений, кластеризация профилей экспрессии, диагностика по экспрессии генов

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	5
2.	Практикум по выравниваниям: построение выравниваний (написание и использование программ), сравнение локальных и глобальных выравниваний, зависимость выравнивания от параметров, оценка статистической значимости	5
3.	Поиск по сходству в базах данных. Smith-Waterman, хэширование (lookup table), BLAST, оценка значимости (E-value, P-value), фильтрация повторов и обработка участков малой сложности (фильтрация, пересчет значимости), паттерны (Prosite), профили, Psi-BLAST, HMM (PFAM)	6
4.	Анализ популяционных данных SNP тандемные повторы митохондрии и Y-хромосомы данные по рестрикции	5
5.	Практикум в String и Prosite	5
6.	Практическая работа по выявлению влияния степени метилирования промоторов на экспрессию генов. Работа с банками данных по экспрессии	4
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)	0
самостоятельное изучение разделов дисциплины	10
выполнение домашних заданий, домашних контрольных работ	15
подготовка к лабораторным работам, к практическим и семинарским занятиям	10
подготовка к контрольным работам, коллоквиумам	3
Итого текущей СР:	40
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	20
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	20
Итого творческой СР:	40
Общая трудоемкость СР:	80

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 «Вычислительная геномика» 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.