

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)
Институт биологии, экологии, почвоведения, сельского и лесного хозяйства (БИОЛОГИЧЕСКИЙ
ИНСТИТУТ)



Директор Биологического института

Д. С. Воробьев

«25» марта 20 22 г.

Рабочая программа дисциплины

Биоинформатика: анализ генома

по направлению подготовки

06.04.01 Биология

Направленность (профиль) подготовки:
«Генетика, геномика и синтетическая биология»

Форма обучения
Очная

Квалификация
Магистр

Год приема
2022

Код дисциплины в учебном плане: Б1.В.ДВ.03.01

СОГЛАСОВАНО:

Руководитель ОП

В. Н. Стегний

Председатель УМК

А. Л. Борисенко

Томск – 2022

1. Цель и планируемые результаты освоения дисциплины (модуля)

Целью освоения дисциплины является формирование следующих компетенций:

- ОПК-2 – способность творчески использовать в профессиональной деятельности знания фундаментальных и прикладных разделов дисциплин (модулей), определяющих направленность программы магистратуры;
- ОПК-6 – способность творчески применять и модифицировать современные компьютерные технологии, работать с профессиональными базами данных, профессионально оформлять и представлять результаты новых разработок
- ОПК-8 – способность использовать современную исследовательскую аппаратуру и вычислительную технику для решения инновационных задач в профессиональной деятельности.
- ПК 2 - Способен проводить основные этапы полевых и лабораторных исследований в соответствии с профилем (направленностью) магистерской программы

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

- ИОПК-2.1. Демонстрирует понимание фундаментальных и прикладных представлений дисциплин, определяющих направленность программы магистратуры.
- ИОПК-2.2. Знает методологические основы дисциплин, определяющих направленность программы магистратуры.
- ИОПК-2.3. Использует фундаментальные знания, практические наработки и методический базис специальных дисциплин, определяющих направленность программы магистратуры, при планировании и реализации профессиональной деятельности.
- ИОПК-6.1. Владеет навыками описывать разнообразие, пути и перспективы применения компьютерных технологий в современной биологии.
- ИОПК-6.2. Знает компьютерные технологии и профессиональные базы данных при планировании профессиональной деятельности, обосновывает их выбор.
- ИОПК-8.2. Знает современную исследовательскую аппаратуру и вычислительную технику при решении стандартных и инновационных задач в профессиональной деятельности
- ИПК 2.4. Описывает, обобщает и делает выводы на основе результатов исследования, в том числе с помощью современных компьютерных технологий

2. Задачи освоения дисциплины

- Освоить аппарат обработки и анализа геномных данных.
- Научиться применять понятийный аппарат биоинформатики для решения практических задач профессиональной деятельности.

3. Место дисциплины (модуля) в структуре образовательной программы

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Семестр 2, зачет.

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: молекулярная биология, теория вероятности и математическая статистика.

6. Язык реализации

Русский

7. Объем дисциплины (модуля)

Общая трудоемкость дисциплины составляет 2 з.е., 72 часа, из которых:

- лекции: 12 ч.;
- семинарские занятия: 18 ч.;
- практические занятия: 0 ч.;
- лабораторные работы: 0 ч.;
- в том числе практическая подготовка: 0 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины (модуля), структурированное по темам

Тема 1. Введение в биоинформатику

Краткое введение в дисциплину на примере различных задач, решаемых биоинформатикой.

Тема 2. Картирование и референсные геномы

Описание процесса выравнивания последовательностей на референс. Контроль качества входных данных, а также результатов выравнивания.

Тема 3. Поиск герминальных и соматических мутаций

Освоение набора инструментов для поиска мутаций, а также их аннотирования и интерпретации.

Тема 4. Поиск структурных вариаций

Освоение набора инструментов для поиска структурных вариаций.

Тема 5. Введение в секвенирование единичных клеток

Описание технологий и задач, решаемых с помощью технологии единичных клеток.

Тема 6. Контроль качества и нормализация данных единичных клеток. Контроль качества и требования к предобработке данных.

Тема 7. Кластеризация данных и дифференциальная экспрессия

Решение задачи кластеризации в рамках набора данных секвенирования единичных клеток.

Тема 8. Работа с Loupe Browser. Освоение инструмента для обработки и визуализации данных единичных клеток.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля выполнения заданий в рамках каждой отдельной темы курса.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет во втором семестре проводится устно в форме защиты отчёта по итоговому заданию по вариантам. Дополнительно каждый студент получает билет с 4 теоретическими вопросами. Продолжительность зачета 1,5 часа.

Примерный перечень теоретических вопросов

1. Объясните, что такое картирование.
2. Объясните, чем герминальные мутации отличаются от соматических.
3. Опишите разницу между FASTQ и uBAM форматами. Где и для чего они применяются?
4. Какое преимущество и за счёт чего даёт секвенирование на основе парных прочтений?

Примеры задач:

1. Задача 1.

Дано: файлы коротких прочтений

Требуется: выравнять прочтения на референс

2. Задача 2.

Дано: файлы выравненных на референс прочтений секвенирования одной хромосомы человека из образца его крови

Требуется: найти и проаннотировать список герминальных мутаций

Результаты **зачета с оценкой** определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Текущий контроль на итоговую оценку никак не влияет и используется лишь в качестве индикатора освоения студентом материала на текущий момент времени. Итоговое задание предполагает применение навыков, полученных при выполнении предыдущих заданий, а также публичную презентацию и защиту своих результатов в рамках обучаемой группы студентов. Итоговая оценка ставится на основании средневзвешенного значения по всем задачам в зависимости от их комбинации в рамках варианта.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=33097>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

в) Методические указания по проведению лабораторных работ.

г) Методические указания по организации самостоятельной работы студентов.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

– Стефанов, В. Е. Биоинформатика : учебник для академического бакалавриата / В. Е. Стефанов, А. А. Тулуб, Г. Р. Мавропуло-Столяренко. — Москва : Издательство Юрайт, 2019. — 252 с.

– Часовских, Н. Ю. Биоинформатика : учебник / Н. Ю. Часовских. - Москва: ГЭОТАР-Медиа, 2020. - 352 с.

б) дополнительная литература:

– Arthur Lesk. Introduction to Bioinformatics (fifth edition) — Oxford University Press, 2019. — 432 pages

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:
– Ubuntu 20.04, Microsoft Visual Code, Anaconda, R, RStudio, GATK 4, samtools, vcftools, bwa, FastQC, Fastp, MultiQC, Annovar, ClinCNV, CNVKit, IGV, Loupe Browser;
– публично доступные веб-технологии (BLAST: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>; IGV Web App: <https://igv.org/>; Genome Browser Gateway: <https://genome-euro.ucsc.edu/cgi-bin/hgGateway>).

б) информационные справочные системы:
– GATK Best Practices Workflows – <https://gatk.broadinstitute.org/hc/en-us/sections/360007226651-Best-Practices-Workflows>
– IGV User Guide – <https://software.broadinstitute.org/software/igv/UserGuide>

в) профессиональные базы данных:
– GenBank – <https://www.ncbi.nlm.nih.gov/genbank/>
– Набор различных баз данных для аннотации найденных мутаций с помощью инструмента Annovar – <https://annovar.openbioinformatics.org/en/latest/user-guide/download/>
– Genome Aggregation Database – <https://gnomad.broadinstitute.org/>

14. Материально-техническое обеспечение

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

Аудитории для проведения занятий лекционного и семинарского типа индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации в смешанном формате, оснащенные компьютерной техникой.

15. Информация о разработчиках

Артемов Глеб Николаевич, канд. биол. наук, доцент кафедры генетики и клеточной биологии ТГУ. Воробьев Ростислав Сергеевич, младший научный сотрудник лаборатории биоразнообразия и экологии ТГУ.