

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Директор физтех-школы
биологической и медицинской
физики**

Д.В. Кузьмин

по дисциплине: **Рабочая программа дисциплины (модуля)**
Биоинформатика

программа аспирантуры: Компьютерные науки и информатика

курс: кафедра молекулярной и трансляционной медицины
1

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

Аудиторных часов: 30 всего, в том числе:

лекции: 30 час.

семинары: 0 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 48 час.

Всего часов: 78, всего зач. ед.: 2

Количество контрольных работ, заданий: 2

Программу составил: В.М. Говорун, д-р биол. наук, профессор

Программа обсуждена на заседании кафедры молекулярной и трансляционной медицины 08.06.2023

Аннотация

Целью данной дисциплины является приобретение практических навыков анализа данных протеомных и геномных экспериментов для построения системных моделей биологических процессов. Студент после освоения курса будет понимать фундаментальные понятия, законы, теории постгеномной биологии, задачи биоинформатического анализа и его связь с другими науками, принципы работы современных баз данных по структуре геномов, белков и другой биологической информации.

1. Цели и задачи

Цель дисциплины

- приобретение практических навыков анализа данных протеомных и геномных экспериментов для построения системных моделей биологических процессов.

Задачи дисциплины

- освоение студентами основных средств анализа геномной, структурной и другой биологической информации;
- применение методов биоинформатики для получения новых знаний в области живых систем;
- оказание консультаций и помощи студентам в проведении собственных теоретических и экспериментальных исследований в области биоинформатики.

2. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- фундаментальные понятия, законы, теории постгеномной биологии;
- задачи биоинформатического анализа и его связь с другими науками;
- принципы работы современных баз данных по структуре геномов, белков и другой биологической информации.

уметь:

- абстрагироваться от несущественного при моделировании реальных биологических процессов;
- пользоваться своими знаниями для решения фундаментальных и прикладных задач постгеномной биологии;
- создавать компьютерные программы, используемые в биоинженерии и биоинформатике, и самостоятельно осваивать новые ресурсы (базы данных и программы) и экспериментальные методы;
- делать качественные выводы при переходе к предельным условиям в изучаемых проблемах;
- осваивать новые предметные области, теоретические подходы и экспериментальные методики;
- определять актуальность целей и задач и практическую значимость исследования;
- проводить анализ результатов и методического опыта исследования применительно к общей фундаментальной проблеме в избранной области;
- работать на современном, в том числе и уникальном вычислительном оборудовании;
- эффективно использовать информационные технологии и компьютерную технику для достижения необходимых теоретических и прикладных результатов.

владеть:

- навыками освоения большого объема информации;
- навыками самостоятельной работы в Интернете;
- культурой моделирования биологических задач;
- навыками грамотной обработки результатов опыта и сопоставления с теоретическими данными;
- практикой исследования и решения теоретических и прикладных задач молекулярной медицины;
- навыками теоретического анализа задач геномики, транскриптомики, протеомики и метаболомики, связанных с изучением свойств биологических систем на молекулярном и субклеточном уровнях структурной организации.

3. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

3.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Сборка геномов de novo	3			5
2	Биоинформатический анализ масс-спектрометрической информации в протеомике	3			5
3	Введение в биоинформатику	3			5
4	Визуализация экспериментальных данных в постгеномной биологии	3			5
5	Геномика	3			3
6	Геномное картирование	3			3
7	Обработка транскриптомных данных	2			3
8	Представление геномной информации	2			3
9	Протеомика	2			3
10	Статистический анализ геномных, протеомных и транскриптомных данных	3			10
11	Технологии чтения биологических текстов	3			3
Итого часов		30			48
Подготовка к экзамену		0 час.			
Общая трудоёмкость		78 час., 2 зач.ед.			

3.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 2 (Весенний)

1. Сборка геномов de novo

Программы-ассемблеры. Различие алгоритмов сборки коротких и длинных ДНК-прочтений.

2. Биоинформатический анализ масс-спектрометрической информации в протеомике

Задача сравнения генетических и белковых последовательностей. Методы выравнивания: парное и множественное, локальное и глобальное. Алгоритм глобального выравнивания Нидльма-на-Вунша (Needleman-Wunsh). Алгоритм локального выравнивания Смита-Уотермана (Smith-Waterman). Gibbs sampling.

3. Введение в биоинформатику

История развития компьютерной обработки биологических данных. Определение биоинформатики. Базовые понятия. Общее представление о задачах биоинформатического анализа и его связи с другими науками. Области применения.

4. Визуализация экспериментальных данных в постгеномной биологии

Визуализация данных биологического эксперимента. Программные продукты для визуализации.

5. Геномика

Определение геномики. Структурная и функциональная геномика. Распространенные технологии секвенирования и форматы результатов. Программное обеспечение: Bowtie, samtools, MUMmer.

6. Геномное картирование

Однонуклеотидные полиморфизмы и методы их детекции. Поиск геномных транслокаций. Поиск повторов, комплементарностей и симметрий в последовательностях.

7. Обработка транскриптомных данных

Методы распознавания промоторов. Сборка последовательностей. Картирование сайтов начала транскрипции.

8. Представление геномной информации

Форматы результатов секвенирования. Распознавание структурно-функциональных мотивов в генетических текстах. Понятие консенсуса, весовой матрицы. Оценка точности распознавания.

9. Протеомика

Алгоритмы идентификации белков по масс-спектрам. Программные пакеты для протеомного анализа.

10. Статистический анализ геномных, протеомных и транскриптомных данных

Статистический анализ геномных, протеомных и транскриптомных данных

11. Технологии чтения биологических текстов

Методы определения последовательности ДНК. Секвенаторы нового поколения. Подходы к высокопроизводительному секвенированию ДНК.

4. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Учебная аудитория, оснащенная компьютером и мультимедийным оборудованием (проектор, звуковая система).

5. Перечень рекомендуемой литературы

Основная литература

Предоставляется на кафедре:

1. А.М. Campbell. Discovering Genomics, Proteomics and Bioinformatics. 2002, ISBN-10: 0805347224
2. М. Ридли. Геном: автобиография вида в 23 главах. Эксмо, 2008, ISBN 978-5-699-30682-4

Дополнительная литература

Предоставляется на кафедре:

1. J.M. Claverie, C. Notredame: Bioinformatics for Dummies. 2006, ISBN: 978-0-470-08985-9

6. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

Не используются

7. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

Для части занятий потребуется Zoom. Google Drive для доступа к материалам курса. Приветствуется наличие во время занятий смартфонов/ноутбуков для участия в интерактивных упражнениях.

8. Методические указания для обучающихся по освоению дисциплины (модуля)

Студент, прослушавший курс, должен с одной стороны, овладеть теоретическим аппаратом биоинформатики, а с другой стороны, должен научиться применять полученные знания на практике. Успешное освоение курса требует самостоятельной работы студента. В программе курса для самостоятельной работы студента над темой отводится минимальное время.

Самостоятельная работа включает в себя:

- проработку учебного материала (по конспектам лекций, учебной и научной литературе),
- чтение и конспектирование дополнительной литературы,
- подготовку ответов на вопросы, предназначенных для самостоятельного изучения,
- решение задач, предлагаемых студентам на занятиях,
- подготовку к дифференцированному зачёту.

Руководство и контроль самостоятельной работы студента осуществляется в форме индивидуальных консультаций. Показателем владения материалом служит умение решать задачи. Для формирования умения применять теоретические знания на практике студенту необходимо решать как можно больше задач. При решении задач каждое действие необходимо аргументировать, ссылаясь на рассмотренный ранее теоретический аппарат.

Обычно придерживаются следующей схемы: изучение материала лекции по конспекту в тот же день, когда была прослушана лекция (10-15 минут); повторение материала накануне следующей лекции (10-15 минут), проработка учебного материала по конспектам лекций, учебной и научной литературе, подготовка ответов на вопросы, решение задач (1 час).

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к лектору. Обязательным требованием является выполнение домашних работ, которые систематически сдаются на проверку.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

программа аспирантуры: Компьютерные науки и информатика
Физтех-школа Прикладной Математики и Информатики
кафедра молекулярной и трансляционной медицины

курс: 1

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

Разработчик: В.М. Говорун, д-р биол. наук, профессор

1. Показатели оценивания компетенций

В результате изучения дисциплины «Биоинформатика» обучающийся должен:

знать:

- фундаментальные понятия, законы, теории постгеномной биологии;
- задачи биоинформатического анализа и его связь с другими науками;
- принципы работы современных баз данных по структуре геномов, белков и другой биологической информации.

уметь:

- абстрагироваться от несущественного при моделировании реальных биологических процессов;
- пользоваться своими знаниями для решения фундаментальных и прикладных задач постгеномной биологии;
- создавать компьютерные программы, используемые в биоинженерии и биоинформатике, и самостоятельно осваивать новые ресурсы (базы данных и программы) и экспериментальные методы;
- делать качественные выводы при переходе к предельным условиям в изучаемых проблемах;
- осваивать новые предметные области, теоретические подходы и экспериментальные методики;
- определять актуальность целей и задач и практическую значимость исследования;
- проводить анализ результатов и методического опыта исследования применительно к общей фундаментальной проблеме в избранной области;
- работать на современном, в том числе и уникальном вычислительном оборудовании;
- эффективно использовать информационные технологии и компьютерную технику для достижения необходимых теоретических и прикладных результатов.

владеть:

- навыками освоения большого объема информации;
- навыками самостоятельной работы в Интернете;
- культурой моделирования биологических задач;
- навыками грамотной обработки результатов опыта и сопоставления с теоретическими данными;
- практикой исследования и решения теоретических и прикладных задач молекулярной медицины;
- навыками теоретического анализа задач геномики, транскриптомики, протеомики и метаболомики, связанных с изучением свойств биологических систем на молекулярном и субклеточном уровнях структурной организации.

2. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Во время текущего контроля студент должен уметь ответить на следующие вопросы:

- 1) История развития компьютерной обработки биологических данных. Определение биоинформатики.
- 2) Методы определения последовательности ДНК.
- 3) Форматы результатов секвенирования.
- 4) Однонуклеотидные полиморфизмы и методы их детекции.
- 5) Программы для сборки геномов de novo.
- 6) Поиск геномных транслокаций.
- 7) Понятие алгоритма. Вычислительная сложность алгоритмов. Методы сравнения алгоритмов.
- 8) Задача сравнения генетических и белковых последовательностей. Методы выравнивания: парное и множественное, локальное и глобальное. Алгоритм глобального выравнивания Нидльмана-Вунша (Needleman-Wunsh). Алгоритм локального выравнивания Смита-Уотермана (Smith-Waterman). Gibbs sampling.
- 9) Пакет Blast. Назначение и основные возможности. Алгоритм.
- 10) FASTA. Назначение и основные возможности. Алгоритм.
- 11) Поиск повторов, комплементарностей и симметрий в последовательностях.
- 12) Основы методов анализа данных. Регрессионный анализ. Дискриминантный анализ. Методы кластеризации. Факторный анализ.

- 13) Распознавание структурно-функциональных мотивов в генетических текстах. Понятие консенсуса, весовой матрицы. Оценка точности распознавания.
- 14) Методы распознавания промоторов.
- 15) Понятие модели. Приемы и способы моделирования. Основные этапы построения математических моделей.
- 16) Статистический анализ геномных данных.
- 17) Визуализация данных биологического эксперимента.

Во время занятий могут проходить интерактивные обсуждения в чатах курса, что будет являться домашним заданием. Возможно выполнение патентного поиска в качестве самостоятельной задачи. Успешное выполнение всех заданий по курсу и выполнение контрольных срезов знаний дает преимущество на дифференцированном зачете.

3. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

Вопросы для подготовки к дифференцированному зачету:

- 1) История развития компьютерной обработки биологических данных. Определение биоинформатики.
- 2) Методы определения последовательности ДНК.
- 3) Форматы результатов секвенирования.
- 4) Однонуклеотидные полиморфизмы и методы их детекции.
- 5) Программы для сборки геномов de novo.
- 6) Поиск геномных транслокаций.
- 7) Понятие алгоритма. Вычислительная сложность алгоритмов. Методы сравнения алгоритмов.
- 8) Задача сравнения генетических и белковых последовательностей. Методы выравнивания: парное и множественное, локальное и глобальное. Алгоритм глобального выравнивания Нидльмана-Вунша (Needleman-Wunsh). Алгоритм локального выравнивания Смита-Уотермана (Smith-Waterman). Gibbs sampling.
- 9) Пакет Blast. Назначение и основные возможности. Алгоритм.
- 10) FASTA. Назначение и основные возможности. Алгоритм.
- 11) Поиск повторов, комплементарностей и симметрий в последовательностях.
- 12) Основы методов анализа данных. Регрессионный анализ. Дискриминантный анализ. Методы кластеризации. Факторный анализ.
- 13) Распознавание структурно-функциональных мотивов в генетических текстах. Понятие консенсуса, весовой матрицы. Оценка точности распознавания.
- 14) Методы распознавания промоторов.
- 15) Понятие модели. Приемы и способы моделирования. Основные этапы построения математических моделей.
- 16) Статистический анализ геномных данных.
- 17) Визуализация данных биологического эксперимента.

Критерии оценивания

Оценка отлично (10 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (9 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (8 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо (7 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо (6 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо (5 баллов) - выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно (4 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно (3 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно (2 балла) - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно (1 балл) - выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

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

При проведении устного дифференцированного зачета обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету на устном дифференцированном зачете не должен превышать одного астрономического часа.