

**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**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Системная биология
<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Медицинская физика и биоинформатика Физтех-школа Биологической и Медицинской Физики кафедра молекулярной и трансляционной медицины
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Экзамен

Аудиторных часов: 60 всего, в том числе:

лекции: 0 час.

семинары: 60 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 45 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Программу составил: Г.П. Арапиди, канд. биол. наук, профессор

Программа обсуждена на заседании кафедры молекулярной и трансляционной медицины 21.02.2025

## Аннотация

Курс посвящен изучению основ статистика и биоинформатики, теорий, методов обработки данных масс-спектрометрического анализа и результатов секвенирования последовательностей РНК и ДНК. Студенты овладеют навыками проведения вычислительного эксперимента и компьютерной обработки результатов секвенирования РНК, ДНК и определения последовательностей белков и пептидов с использованием методов математической статистики и ресурсов общедоступных биоинформатических баз данных.

### 1. Цели и задачи

#### Цель дисциплины

- изучение основ статистика и биоинформатики, теорий, методов обработки данных масс-спектрометрического анализа и результатов секвенирования последовательностей РНК и ДНК. Формирование научного мировоззрения. Овладение навыками проведения вычислительного эксперимента и компьютерной обработки результатов секвенирования РНК, ДНК и определения последовательностей белков и пептидов с использованием методов математической статистики и ресурсов общедоступных биоинформатических баз данных.

#### Задачи дисциплины

- формирование у студентов научного стиля мышления, умения ориентироваться в потоке научной и технической информации;
- формирование навыков проведения вычислительного эксперимента и анализа его результатов;
- освоение приемов и методов решения конкретных задач из различных областей системной биологии: протеомики, геномики, транскриптомики, эпигеномики;
- приобретение базовых знаний о задачах биоинформатического анализа и его связи с другими науками, способствующих успешному освоению различных курсов.

### 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.1 Способен планировать и проводить научные исследования самостоятельно или в составе научного коллектива
	ПК-2.2 Способен проводить апробацию результатов научно-исследовательской работы посредством публикации научных статей и участия в конференциях
ПК-3 Способен профессионально работать с исследовательским и испытательным оборудованием (приборами и установками, специализированными пакетами прикладных программ) в избранной предметной области	ПК-3.1 Понимает принципы работы используемого оборудования (специализированных пакетов прикладных программ)
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)
	ПК-3.3 Способен оценивать точность полученных экспериментальных (численных) результатов

### 3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основные методы обработки данных масс-спектрометрического анализа и результатов секвенирования последовательностей РНК и ДНК, применение этих методов в важнейших практических приложениях.

уметь:

- проводить вычислительные эксперименты и компьютерную обработку результатов секвенирования РНК, ДНК и определения последовательностей белков и пептидов с использованием методов математической статистики и ресурсов общедоступных биоинформатических баз данных.

владеть:

- приемами и методами решения конкретных задач из различных областей системной биологии: протеомики, геномики, транскриптомики, эпигеномики.

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

#### 4.1. 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	Введение в системную биологию. Основы биостатистики и методы визуализации в биоинформатике. Работа в среде программирования R		12		9
2	NGS секвенирование. Геномика.		12		9
3	Транскриптомика		12		9
4	Эпигеномика.		12		9
5	Протеомика.		12		9
Итого часов			60		45
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

#### 4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 2 (Весенний)

1. Введение в системную биологию. Основы биостатистики и методы визуализации в биоинформатике. Работа в среде программирования R

Введение в системную биологию. История развития компьютерной обработки биологических данных. Определение биоинформатики. Общее представление о задачах биоинформатического анализа и его связи с другими науками. Области применения.

Теория вероятностей и описательная статистика. Генеральная совокупность и выборка. Статистическая значимость; нулевая и альтернативная гипотезы. Распределения, статистики и параметры. Параметрические и непараметрические статистические методы и критерии. Метод главных компонент.

Сравнение выборок и визуализация. Сравнение выборок по Стьюденту. Использование критерия Фишера для сравнения выборок. Непараметрические аналоги параметрических методов. Поправки на множественные сравнения. Визуализация данных биологического эксперимента.

Работа в командной строке и работа в среде программирования R. Подключение к удаленному серверу. Работа в командной строке Linux. Основы работы в среде программирования R. Логические конструкции и циклы в среде программирования R. Скачивание и чтение файлов. Работа с data.frame и data.table. The Base Plotting System. The Lattice and the ggplot2 Plotting Systems.

## 2. NGS секвенирование. Геномика.

Определение геномики. Структурная и функциональная геномика. Распространенные технологии секвенирования. Методы определения последовательности ДНК. Подходы к высокопроизводительному секвенированию ДНК. Секвенирование по Сенгеру. Секвенаторы нового поколения. Секвенаторы третьего поколения.

Контроль качества NGS данных. Сборка геномов de novo. Форматы данных секвенирования. Анализ качества секвенированных последовательностей нуклеотидов. Сборка геномов de novo. Подходы и программы-ассемблеры. Различия алгоритмов сборки коротких и длинных ДНК-прочтений. Оценка качества сборки. Поиск и аннотация генов.

Методы выравнивания биологических последовательностей. 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).

Выравнивание коротких фрагментов на геном. Поиск геномных вариаций. Ресеквенирование ДНК. Референсный геном. Алгоритмы и программы картирования коротких фрагментов. Однонуклеотидные полиморфизмы, структурные вариации и методы их детекции. Функциональная классификация полиморфизмов.

## 3. Транскриптомика

РНК и методы их исследования. Транскриптомный анализ. Типы РНК в клетке и методы их секвенирования. Анализ данных miRNA-seq. Особенности анализа данных секвенирования lncRNA. Гибридизация при помощи чипов. Секвенирование мРНК. Определение дифференциально экспрессирующихся генов. Интерпретация данных секвенирования. Анализ альтернативного сплайсинга.

Секвенирование единичных клеток. Сравнение RNA-seq и scRNA-seq. Применение данных scRNA-seq. Существующие платформы для изоляции отдельных клеток (Drop-seq, 10Xgenomics, Smart-seq2). Основная схема обработки scRNA-Seq. Оценка качества данных. Способы кластеризации данных. Дополнительные стадии обработки scRNA-Seq: CellPhoneDB, RNA velocity, определений траекторий и прочее.

## 4. Эпигеномика.

Метилирование ДНК: роль, функции и основные методы анализа. Молекулярно-биологические основы метилирования ДНК. Роль метилирования ДНК при эмбриональном развитии и онкогенезе. Методы определения уровня метилирования ДНК. Прикладные задачи метилирования ДНК.

Гистоновый код и открытый хроматин. Гистоновые модификации: роль и функции. Иммунопреципитация хроматина и ее приложения. Методы определения регионов открытого хроматина (DNase-seq, ATAC-seq, FAIRE-seq). Приложения DNase-seq и ATAC-seq.

Пространственная структура хроматина. Роль пространственной структуры хроматина в регуляции экспрессии. Методы определения пространственной структуры хроматина. Базы данных по 3D-геномике.

## 5. Протеомика.

Идентификация белков и пептидов по данным масс-спектрометрии. Протеомика. Анализ белков и пептидов методами масс-спектрометрии. Идентификация белков по масс-спектру. Идентификация пептидов методом de novo. Идентификация пептидных последовательностей с использованием «баз данных» белковых последовательностей. Программные пакеты для протеомного анализа.

Валидация результатов идентификации и оценка ошибки. Методы валидации идентифицированных пептидов. Преобразование Score в вероятность. "Target-Decoy" подход для валидации идентифицированных пептидов. Идентификация белков на основе результатов идентификации пептидов.

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

- учебные аудитории для проведения занятий лекционного / семинарского типа;
- аудитории, оснащенные компьютерной техникой с подключением к сети «Интернет»;
- компьютер и мультимедийное оборудование (проектор, звуковая система),
- индивидуальные вычислительные средства студентов (персональные компьютеры) для выполнения домашних заданий.

## **6. Перечень рекомендуемой литературы**

### Основная литература

Основная литература:

Литература предоставляется базовой кафедрой:

1. Практическая статистика для специалистов Data Science [2022] Питер Брюс, Питер Гедек, Эндрю Брюс.
2. Kappelmann-Fenzl M (2021) Next generation sequencing and data analysis. Springer, Cham
3. Guo-Cheng Yuan (2019) Computational Methods for Single-Cell Data Analysis. Springer
4. Renato Paro, Ueli Grossniklaus, Raffaella Santoro, Anton Wutz (2021) Introduction to Epigenetics. Springer Cham. 215c
5. Silvio Bicciato, Francesco Ferrari (2022) Hi-C Data Analysis. Methods and Protocols. Springer. 354c

### Дополнительная литература

Дополнительная литература:

Литература предоставляется базовой кафедрой:

1. Масс-спектрометрия в органической химии Издание второе, переработанное и дополненное Москва: ТЕХНОСФЕРА, 2015. - 704с.
2. A decade's perspective on DNA sequencing technology. Elaine R. Mardis. Nature volume 470, pages 198–203. doi: 10.1038/nature09796
3. Neus Visa, Antonio Jordán-Pla (2018) Chromatin Immunoprecipitation. Springer 243c

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

<http://www.ncbi.nlm.nih.gov>

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

PubMed

## **9. Методические указания для обучающихся по освоению дисциплины (модуля)**

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения дисциплины, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- напряжённой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к преподавателю.

Возможен промежуточный контроль знаний студентов в виде решения задач в соответствии с тематикой занятий.

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

**по направлению:** Прикладные математика и физика  
**профиль подготовки:** Медицинская физика и биоинформатика  
Физтех-школа Биологической и Медицинской Физики  
кафедра молекулярной и трансляционной медицины  
**курс:** 1  
**квалификация:** магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Экзамен

**Разработчик:** Г.П. Арапиди, канд. биол. наук, профессор

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.1 Способен планировать и проводить научные исследования самостоятельно или в составе научного коллектива
	ПК-2.2 Способен проводить апробацию результатов научно-исследовательской работы посредством публикации научных статей и участия в конференциях
ПК-3 Способен профессионально работать с исследовательским и испытательным оборудованием (приборами и установками, специализированными пакетами прикладных программ) в избранной предметной области	ПК-3.1 Понимает принципы работы используемого оборудования (специализированных пакетов прикладных программ)
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)
	ПК-3.3 Способен оценивать точность полученных экспериментальных (численных) результатов

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Системная биология» обучающийся должен:

### знать:

- основные методы обработки данных масс-спектрометрического анализа и результатов секвенирования последовательностей РНК и ДНК, применение этих методов в важнейших практических приложениях.

### уметь:

- проводить вычислительные эксперименты и компьютерную обработку результатов секвенирования РНК, ДНК и определения последовательностей белков и пептидов с использованием методов математической статистики и ресурсов общедоступных биоинформатических баз данных.

### владеть:

- приемами и методами решения конкретных задач из различных областей системной биологии: протеомики, геномики, транскриптомики, эпигеномики.

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

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

- 1) Генетика и секвенирование генома. Что изучают генетика и геномика? Каким образом генетическая информация закодирована в последовательности ДНК? Проект генома человека завершен – зачем пересеквенировать геномы людей? Задачи геномного секвенирования. Перечислите виды мутации (классификация мутации по значимости). Какие технологии секвенирования Вам известны? В чем отличия секвенаторов разного поколения? Расскажите про секвенирование первого поколения?



2) Секвенаторы «нового поколения». Какие технологии секвенирования Вам известны? В чем отличия секвенаторов разного поколения? Расскажите про секвенаторы второго поколения? Расскажите про секвенаторы третьего поколения? Сравните между собой секвенаторы разных поколений. Для каких задач какие секвенаторы более применимы? Почему до сих пор для решения задач крупных проектов широко используются именно секвенаторы второго поколения, а не третьего?

3) Анализ качества секвенированных последовательностей нуклеотидов. На примере платформы Illumina расскажите об ошибках, которые возникают при секвенировании ДНК. Какой алгоритм устранения ошибок секвенирования и какие программные продукты для этого подходят? Что такое fastq формат файла и шкала Phred? Какие параметры отслеживает программа FastQC? Приведите примеры нормальных значений этих параметров и отклонений? Как бороться с такими ошибками секвенирования? Что такое тримминг и какие действия по триммингу возможны?

4) Сборка генома de novo. Какие подходы по сборке генома существуют? Расскажите про алгоритм Overlap-Layout-Consensus. В чем его недостаток? Для каких задач используется? Расскажите про De Bruijn граф. В чем заключается решение задачи De Bruijn графа? Как решение задачи зависит от значения параметра k? Какие существуют трудности при сборке генома и как они влияют на решение задачи De Bruijn графа? Что нужно знать о данных, из которых вы собираетесь делать сборку генома? Как эти параметры могут повлиять на результат сборки? На какие параметры следует обращать внимание при оценке качества сборки и как валидировать сборку?

5) Методы выравнивания биологических последовательностей. В чем заключается задача выравнивания, и какие существуют методы выравнивания? Расскажите о способах представления выравниваний: текстовое, графическое, матрица сходства. Алгоритм Нидлмана-Вунша, приведите пример его использования для глобального выравнивания двух последовательностей. В чем принципиальное отличие локального выравнивания от глобального, алгоритм Смита-Ватермана.

6) Поиск гомологичных последовательностей. Программа BLAST: опишите основной алгоритм поиска и какие существуют модификации. Какие существуют форматы выдачи программы BLAST? Расскажите про задачу множественного выравнивания, почему стандартные подходы с использованием динамического программирования непригодны в этот случае? Какие есть альтернативы?

7) Выравнивание коротких фрагментов на геном. Что такое референсный геном? Откуда появился референсный геном человека? Почему референсный геном человека до сих пор продолжают улучшать? Какие задачи можно решить, выравнивая последовательности на референсный геном? Почему задача выравнивания последовательностей на референсный геном считается вычислительно сложной, какие существуют подходы для решения этой задачи? Как выбрать оптимальный картировщик? Приведите пример наиболее популярных алгоритмов и в общих словах опишите принцип их работы. Расскажите о форматах BAM и SAM.

8) Поиск геномных вариаций. Что такое геномная вариация, какие бывают виды геномных вариаций? Идентификация вариантов по данным NGS-эксперимента, геномные браузеры. Расскажите о формате данных VCF. Какие бывают виды мутаций и как оценить их значимость? Что значит аннотировать вариант, какие инструменты и базы данные используют для этого?

9) Основные понятия математической статистики. Что изучает математическая статистика? Чем отличаются описательная и индуктивная статистика? Что такое выборка и генеральная совокупность? Какие бывают типы переменных? Какие бывают числовые характеристики выборки? Что такое статистическая гипотеза? Дайте определение основной и альтернативной гипотезы? Чем отличаются ошибки первого и второго рода? Что такое уровень значимости статистической гипотезы?

10) Что такое статистическая гипотеза? Дайте определение основной и альтернативной гипотезы? Чем отличаются ошибки первого и второго рода? Что такое уровень значимости статистической гипотезы? Опишите последовательность действий при построении и проверке статистической гипотезы. Приведите примеры статистических критериев для проверки гипотезы. Чем отличаются непараметрические и параметрические критерии в статистике?

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 экзамене.

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

- 1) Перечислите основные типы графиков, используемых для визуализации биологических данных. Расскажите для каких данных используется каждый график, кратко опишите алгоритм его построения. Дайте постановку задачи, для решения которой применяется критерий Стьюдента. При каких условиях применяется критерий Стьюдента?
- 2) Откуда возникает проблема множественных сравнений при проверке статистических гипотез? Что такое поправка Бонферрони и FDR? Что такое точный критерий Фишера? Приведите пример любой биологической задачи, для которой уместно использование точного критерия Фишера.
- 3) Транскриптомика. Что изучает транскриптомика? Что такое транскрипт? Опишите типы РНК клетки? Что такое микро РНК и длинная белок некодирующая РНК? Какие функции в клетке они выполняют? Чем отличаются протоколы секвенирования микро РНК и белок некодирующей РНК от протокола секвенирования белок кодирующей РНК?
- 4) Анализ транскриптомных данных. Для чего изучают мРНК в клетке? Как узнать последовательность мРНК? Сравните достоинства и недостатки двух методов: гибридизация при помощи чипов и RNA-seq. Опишите в общих чертах схему анализа данных RNA-seq. Что такое метод главных компонент (PCA)? Что такое геновая онтология и зачем она нужна?
- 5) Секвенирование единичной клетки. Опишите отличия тотального секвенирования от секвенирования единичной клетки. Какие задачи помогает решить scRNA-seq? Опишите подход для секвенирования РНК единичных клеток? Опишите основные трудности при проведении секвенирования единичных клеток? Зачем бороться с ПЦР дубликатами при анализе данных секвенирования единичной клетки? Опишите изоляцию клеток при помощи капель (Drop-seq). Опишите принципы баркодирования для создания библиотеки секвенирования единичной клетки.
- 6) Секвенирование единичной клетки. Опишите отличия тотального секвенирования от секвенирования единичной клетки. Какие задачи помогает решить scRNA-seq? Опишите схему анализа данных scRNA-seq. На какие параметры качества данных стоит обращать внимание в отчете CellRanger? Какие этапы анализа реализованы в пакете Seurat?
- 7) Что такое метилирование ДНК? В каких контекстах оно происходит у млекопитающих/растений / бактерий? Приведите примеры организмов, в которых метилирования нет совсем. Функции метилирования ДНК у млекопитающих? Infinium MethylationEPIC: опишите эксперимент и расскажите об обработке данных.
- 8) Роль и функции метилирования ДНК? Белки, которые «пишут», «снимают» и «читают» метилирование ДНК. Основные методы определения метилирования ДНК? WGBS/RRBS: опишите эксперимент и расскажите об обработке данных.
- 9) Что такое гистоны и хроматин? Как химическое изменение гистонов регулирует активность хроматина? Белки, которые «пишут», «снимают» и «читают» гистоновые метки. Ассоциированные с этими белками заболевания. Методы определения гистоновых модификаций. ChIP-seq. Алгоритм обработки данных ChIP-seq. Базы данных с ChIP-seq.
- 20) Хроматин и модификации хроматина: роль и функции? Методы определения открытого хроматина? Какие задачи можно решать при помощи этих методов? Базы данных с информацией об открытых регионах хроматина?
- 10) Пространственная структура хроматина. Зачем она нужна, что регулирует? Методы определения пространственной структуры. DamID – суть метода. Что такое LAD? Как можно использовать ChIP-seq для определения регионов генома, формирующих определенные ядерные компартменты?
- 11) Что такое TAD? 3C, Hi-C и ChiA-PET: различие между методами и области применения? Привести пример заболевания, которое возникает из-за изменения 3D-структуры хроматина.
- 12) Анализ белков и пептидов методами масс-спектрометрии. Что такое протеомика и какие у неё задачи? Как устроен масс-спектрометр? Расскажите на примерах об устройстве источника ионов, масс-анализатора и детектора. Что такое тандемный масс-спектрометр? Приведите примеры. Как работает масс-спектрометр для получения MS/MS спектров? Что такое изотопы и что они дают для масс-спектрометрии белковых молекул? Можно ли идентифицировать белок по масс-спектру, без использования спектров фрагментации?

13) Идентификация пептидов по масс-спектру. Опишите стандартную схему LC-MS/MS эксперимента. В чем особенности Bottom-up протеомики: преимущества и недостатки? Опишите схему MS/MS анализа? Классификация пептидных ионов-фрагментов. Как происходит идентификация пептидов по базам данных белковых структур? Какие базы данных белков существуют? Опишите примерный алгоритм идентификации пептидов методом de novo. Какие параметры важно знать для правильной идентификации пептида? Можно ли проанализировать все возможные модификации за время одного анализа?

14) Методы валидации идентифицированных пептидов. Опишите алгоритм определения верной идентификации в SEQUEST и X! Tandem. Что дает объединение результатов идентификации разных поисковых алгоритмов? Как его можно реализовать? Опишите Target-Decoy подход для валидации идентифицированных пептидов?

15) Как можно идентифицировать белки на основе результатов идентификации пептидов? Как оценить достоверность идентификации белка? Какие возможны сложности и как их преодолеть? В чем разница между Top-down и Bottom-up протеомикой? Опишите преимущества и недостатки каждого из методов.

#### Примеры билетов на экзамене

##### Билет №1

1) Что такое и TAD? 3C, Hi-C и ChiA-PET: различие между методами и области применения? Привести пример заболевания, которое возникает из-за изменения 3D-структуры хроматина.

2) Анализ белков и пептидов методами масс-спектрометрии. Что такое протеомика и какие у неё задачи? Как устроен масс-спектрометр? Расскажите на примерах об устройстве источника ионов, масс-анализатора и детектора. Что такое тандемный масс-спектрометр? Приведите примеры. Как работает масс-спектрометр для получения MS/MS спектров? Что такое изотопы и что они дают для масс-спектрометрии белковых молекул? Можно ли идентифицировать белок по масс-спектру, без использования спектров фрагментации?

##### Билет №2

1) Перечислите основные типы графиков, используемых для визуализации биологических данных. Расскажите для каких данных используется каждый график, кратко опишите алгоритм его построения. Дайте постановку задачи, для решения которой применяется критерий Стьюдента. При каких условиях применяется критерий Стьюдента?

2) Откуда возникает проблема множественных сравнений при проверки статистических гипотез? Что такое поправка Бонферрони и FDR? Что такое точный критерий Фишера? Приведите пример любой биологической задачи, для которой уместно использование точного критерия Фишера.

#### 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.

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

При проведении устного экзамена обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету на устном экзамене не должен превышать одного астрономического часа.