

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

УТВЕРЖДЕНО
исполнительный директор

М.А. Смирнова

	Рабочая программа дисциплины (модуля)
по дисциплине:	Инструментальные основы системной биологии и биостатистика
по направлению:	Биотехнология
профиль подготовки:	Системная и синтетическая биология Физтех-школа Биологической и Медицинской Физики кафедра системной и синтетической биологии
курс:	1
квалификация:	магистр

Семестры, формы промежуточной аттестации:

1 (осенний) - Зачет
2 (весенний) - Экзамен

Аудиторных часов: 75 всего, в том числе:

лекции: 30 час.
семинары: 45 час.
лабораторные занятия: 0 час.

Самостоятельная работа: 165 час.

Подготовка к экзамену: 30 час.

Всего часов: 270, всего зач. ед.: 6

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

Программа обсуждена на заседании кафедры системной и синтетической биологии 11.04.2024

Аннотация

Дисциплина инструментальные основы системной биологии и биостатистика даст студентам целостное представление о современных методах экспериментальной работы с биологическими системами. Применение мультидисциплинарных фундаментальных знаний в приложении к решению аналитических задач в персонализированной биомедицине.

1. Цели и задачи

Цель дисциплины

- получение студентами целостного представления о трансляции фундаментальных знаний в получение и аналитику экспериментальных результатов. Курс представляет собой практико-ориентированную аналитическую методологию. Методы анализа и оценки экспериментальных данных для развития биомедицинских технологий.

Задачи дисциплины

- научиться рассчитывать и анализировать сложные процессы в живых системах;
- овладеть экспериментальными методиками в мультиомиксных технологиях.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области физико-математических наук и информационно-коммуникационных технологий	ОПК-4.1 Способен применять знания и навыки по использованию информационно-коммуникационных технологий для поиска и изучения научной литературы, применения прикладных программных продуктов
	ОПК-4.2 Способен применять знания в области физико-математических наук для решения поставленной задачи, формулирования выводов и оценки полученных результатов
	ОПК-4.3 Способен аргументировано выбирать способ проведения научного исследования
ОПК-5 Способен и готов к повышению квалификации, профессиональному росту и руководству коллективом в сфере своей профессиональной деятельности, толерантно воспринимая социальные, этнические, конфессиональные и культурные различия	ОПК-5.1 Способен работать в коллективе, толерантно воспринимая социальные, этнические, конфессиональные и культурные различия
	ОПК-5.2 Владеет навыком руководства малым коллективом в сфере своей профессиональной деятельности
	ОПК-5.3 Стремится к получению новых знаний, профессиональному и личностному росту

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- технологии геномного, транскриптомного, протеомного и метаболомного анализов;
- фундаментальные основы экспериментальных методов в молекулярной и системной биологии;
- основы статистики для интерпретации результатов мультиомиксных исследований.

уметь:

- ставить цели и формулировать задачи, понимать поставленные цели и задачи при создании аналитических систем в медико-биологических исследованиях;
- использовать свои знания для решения задач и проведения экспериментальных исследований;
- оценивать корректность постановок задач и строить алгоритмы достижения их оптимального решения в различных (в том числе меняющихся) условиях.

владеть:

- полученные знания и умения должны быть использованы для создания биомедицинских продуктов и технологий.

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	Редукционизм и холизм в биологии XIX–XX века.		4		8
2	Молекулярный состав живой системы и методы его изучения. Технологические революции в молекулярной биологии- изменение наших знаний о составе и свойствах клетки.		4		7
3	Примеры и основные результаты системных исследований прокариот. Построение математических моделей. Принцип неопределенности. Синтетическая биология. Этап развития системного анализа.		4		8
4	Современные методы определения первичной структуры нуклеиновых кислот, мономолекулярное секвенирование, секвенирование единичных клеток, метагеномика, методы анализа паттерна метилирования ДНК.		3		7
5	Контроль качества NGS данных. Сборка геномов de novo. Методы выравнивания биологических последовательностей.		4		8
6	Типы масс-спектрометров. Виды масс-спектрометрического анализа.		3		7
7	Масс-спектрометрия в биологии. Top–down и bottom-up подходы в протеомном анализе.		4		7
8	Введение в системную и синтетическую биологию.		4		8
9	Идентификация белков и пептидов по данным масс-спектрометрии. Валидация результатов идентификации и оценка ошибки.	5	2		17
10	Протеогеномика. Области применения протеогеномных технологий для коррекции аннотации геномов, альтернативного сплайсинга и процессинга белков. Профилирование и таргетное определение метаболитов. Флаксомика.	5	2		18

11	Методы количественного анализа в протеомике. Методы масс-спектрометрического анализа для определения комплексов белков. Интерактомика. Методы определения посттрансляционных модификаций (ПТМ).	5	3		17
12	Основы биостатистики и методы визуализации в биоинформатике.	5	3		18
13	Методы кросс-валидации результатов мультиомиксного анализа. Технология планирования экспериментов для кросс-валидации.	5	3		18
14	Базы данных. Понятие о форматах. Базы знаний по омиксным данным.	5	2		17
Итого часов		30	45		165
Подготовка к экзамену		30 час.			
Общая трудоёмкость		270 час., 6 зач.ед.			

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

Семестр: 1 (Осенний)

1. Редукционизм и холизм в биологии XIX–XX века.

Ключевые аналогии, использовавшиеся для понимания жизни. Редукционизм vs холизм. Представления о феномене жизни со стороны всех разновидностей редукционизма и холизма (по В.И. Моисееву). Редукция или/и поиск структуры. Две стратегии единой науки. Объяснить сложное через элементарные свойства простого. Наука до XVII века. Иррациональный холизм. XVII век. Сильный редукционизм. XVIII век. Атеистический механицизм. XIX век. Возрождение холизма. XIX век. Неовитализм. Эмерджентный подход. Порядок из хаоса. Теория аутопоэза Матураны и Варелы («теория Сантьяго»).

2. Молекулярный состав живой системы и методы его изучения. Технологические революции в молекулярной биологии- изменение наших знаний о составе и свойствах клетки.

Секвенирование ДНК: исторический экскурс. Секвенирование по Сэнгеру: «плюс–минус» метод. Основные этапы метода секвенирования по Сэнгеру. Секвенирование ДНК по Сэнгеру: метод «терминаторов». Автоматическое секвенирование по Сэнгеру. «Фабрики» секвенирования ДНК по Сэнгеру. Секвенирование ДНК методом химической дегградации. Секвенирование ДНК по максаму–гилберту. Развитие методов секвенирования ДНК первого поколения. Успехи применения секвенирования ДНК первого поколения. George Church разработал colony sequencing. Масс-спектрометрия. Протеомика. «Омиксные» подходы системной биологии.

3. Примеры и основные результаты системных исследований прокариот. Построение математических моделей. Принцип неопределенности. Синтетическая биология. Этап развития системного анализа.

Микоплазмы – объект для системной биологии. Микоплазмы – модель минимальной клетки. Сборка генома *Mycoplasma gallisepticum*. Метод точного картирования и измерения активности промоторов. Создание высокоэффективного вектора для трансформации микоплазм. Результаты мультиомиксного подхода. Неспецифическая регуляция: проскальзывание через терминатор. Белковый ответ *M. gallisepticum* при стрессе ничтожен по сравнению с изменениями на уровне мРНК. Абсолютная квантификация РНК в клетке. Как отделить шумовой ответ от адаптивного? Высокопроизводительная реконструкция регуляторных сетей молликут.

4. Современные методы определения первичной структуры нуклеиновых кислот, мономолекулярное секвенирование, секвенирование единичных клеток, метагеномика, методы анализа паттерна метилирования ДНК.

Генетика и геномика. Геном человека. Нуклеотид. Строение ДНК/РНК. Виды наследственной изменчивости. Точечные генные мутации. Хромосомные мутации. Геномное секвенирование. Технологии секвенирования. Развитие технологий секвенирования. Полимеразная цепная реакция (ПЦР). Общая схема работы NGS: от исходной ДНК до «букв» на экране. Общие принципы работы NGS. Roche 454: пиросеквенирование на микрочипе. Roche 454: пиросеквенирование на микрочипе. SoLID – секвенирование с помощью лигирования. Ионное полупроводниковое секвенирование IonTorrent. Секвенирование с помощью синтеза на чипе (Illumina). Применение MALDI для анализа ДНК. Секвенирование на платформе PacBio. Нанопоровое секвенирование (Oxford Nanopore). Менее известные технологии NGS. Секвенирование единичных молекул. Сводная таблица для некоторых платформ секвенирования. Неизученные виды (сборка геномов *de novo*). Полногеномное ресеквенирование. Секвенирование экзона. ДНК-белковые взаимодействия (ChIP-Seq). Подготовка библиотек кДНК для высокопроизводительного секвенирования. Как секвенировать транскриптом? Метагеномика – подход, основанный на секвенировании.

5. Контроль качества NGS данных. Сборка геномов *de novo*. Методы выравнивания биологических последовательностей.

Секвенирование путем синтеза (Illumina). Источники ошибок, связанные с адаптерами. Источники ошибок, связанные с фазировкой. Падение интенсивности кластеров. Анализ качества (Quality Check). Значение качества, шкала Phred. FastQC: Общая статистика. Качество прочтения нуклеотидов. Качество прочтений внутри ячейки. Суммарное качество прочтений. Нуклеотидный состав прочтений. GC-состав прочтений. Количество неопределенных нуклеотидов. Длины прочтений. Уровень дубликаций прочтений. Перепредставленные последовательности. Наличие адаптеров. Высокая степень фрагментации генома. Тримминг – удаление ошибок секвенирования. Сборка генома *de novo*. Алгоритм сборки Overlap-Layout-Consensus. Граф De Bruijn. К-меры. Проблемы при сборке геномов. Влияние повтора на сборку генома. Объединение прочтений в контиги. От контигов к скаффолдам. «Paired-End» и «Mate Pair» библиотеки. Оценка качества сборки. QUAST. Валидация сборки. Финализация сборки. The NCBI Eukaryotic/Prokaryotic Genome Annotation Pipeline. Поиск открытых рамок считывания (ORFs). Методы выравнивания. Точечная матрица сходства (dot plot). Алгоритм Нидлмана-Вунша. Штрафы и матрицы замен. Алгоритм Смита-Ватермана. Множественное выравнивание. Правила при множественном выравнивании. Методы множественного выравнивания. Обратная задача – выравнивание выравниваний. The Basic Local Alignment Search Tool (BLAST). BLAST – поиск гомологичных последовательностей.

6. Типы масс-спектрометров. Виды масс-спектрометрического анализа.

Масс-спектрометрия. Масс-спектрометр Астона. Ионы и вакуум. Времяпролетный масс-анализатор. Калибровка времяпролетного масс-анализатора. Разрешающая способность. Изотопы. Изотопное распределение. Ортогональный времяпролетный масс-анализатор. Квадруполь. Квадруполь – диаграмма стабильности. Квадруполь – сканирование. Режимы работы квадруполь. Квадруполь – основной транспортный элемент. Орбитальная ионная ловушка. Виды масс-анализаторов.

7. Масс-спектрометрия в биологии. Top-down и bottom-up подходы в протеомном анализе.

Молекулы как предмет научного интереса. Как измерить уровень глюкозы? Аналитическая химия. Методы аналитической химии. Молекулярный состав клетки. Элементный состав клетки. Аналитическая химия для системной биологии. Особенности масс-спектрометрии биомолекул. Электроспрейная ионизация. MALDI - Matrix Assisted Laser Desorption/Ionization. MALDI – матрица. Другие методы ионизации. Протеомика. Top-down и bottom-up подходы в протеомном анализе. «Трипсинолиз». Идентификация – гипотезы. Tandemная масс-спектрометрия. CID – Collision-Induced Dissociation. Фрагментация ионов - газозная реакция. Библиотеки спектров фрагментации. Номенклатура дочерних ионов пептидов. ETD – Electron Transfer Dissociation. Фрагментация пептидов. De novo секвенирование пептидов. Способы фрагментации. Тройной квадруполь. Квадруполь-времяпролётный масс-спектрометр. Современные масс-спектрометры.

8. Введение в системную и синтетическую биологию.

Теория систем и ее роль в биологии. Роль научной методологии. Ранние примеры системной методологии. Методы биологической науки. Стадии развития биологии. Клеточная теория жизни. ДНК и наследственная информация. 1930-е годы – научная революция и начало молекулярной биологии как самостоятельной науки. Один ген – один фермент. Жизнь с точки зрения физики. Изучение отдельных молекул. Синтетическая биология. Вызовы синтетической биологии.

Семестр: 2 (Весенний)

9. Идентификация белков и пептидов по данным масс-спектрометрии. Валидация результатов идентификации и оценка ошибки.

Принцип идентификации. Гипотезы-приманки (decoy). Идентификация белков. Данные-зависимый анализ (DDA). Неполнота данных. Типы эксперимента. Прицельный ВЭЖХ-МС анализ. MRM - Выбор протеотипических пептидов, ионов, фрагментов. Воспроизводимость.

10. Протеогеномика. Области применения протеогеномных технологий для коррекции аннотации геномов, альтернативного сплайсинга и процессинга белков. Профилирование и таргетное определение метаболитов. Флаксомика.

Протеогеномика. Изоформы белков. N-терминомика. Библиотеки спектров фрагментации. Метаболомный анализ. Флаксомика.

11. Методы количественного анализа в протеомике. Методы масс-спектрометрического анализа для определения комплексов белков. Интерактомика. Методы определения посттрансляционных модификаций (ПТМ).

Хроматография. ВЭЖХ-МС. Количественный анализ. Виды количественного анализа. Абсолютная квантификация. Относительная квантификация. SILAC (метаболическое введение метки). Изобарные метки. Изотопы ^{13}C и ^{15}N . Белковые комплексы, кросс-сшивающие агенты. Аффинное выделение комплексов AP-MS, TAP-MS. Интерактом – карта белок-белковых взаимодействий. Обогащение фосфопептидов.

12. Основы биостатистики и методы визуализации в биоинформатике.

Математическая статистика. Наблюдение, генеральная совокупность, выборка. Способы формирования репрезентативной выборки. Частотное распределение переменной. Среднее значение vs медиана vs мода. Квартили. Box plot vs форма распределения. Scatterplots/Stripcharts. Меры изменчивости. Меры отклонения формы распределения. Статистическая гипотеза. Тестирование гипотез в статистике. Ошибки первого и второго рода. Z-оценка (стандартизованная оценка). Область принятия гипотезы vs критическая область. Статистика - критерий проверки гипотезы. Одновыборочные критерии. Распределение Стьюдента. Сравнение двух средних нормальных совокупностей. Статистический критерий Стьюдента. Принятие решения. Непараметрические (ранговые) критерии. Критерий Вилкоксона. Точный критерий Фишера. Анализ обогащения по функциональной принадлежности. Gene Set Enrichment Analysis, GSEA. Множественные сравнения. Поправка Бонферрони. Поправка Бенджамини-Хохберга (FDR). Тепловые карты (heatmap). Снижение размерности пространства. Principal Components Analysis (PCA). Кривые выживаемости Каплана-Мейера. Диаграмма рассеяния (scatterplot). Volcano plot. Диаграмма Венна. UpSet plot.

13. Методы кросс-валидации результатов мультиомиксного анализа. Технология планирования экспериментов для кросс-валидации.

Протеогеномная коррекция аннотации. Проверка транскриптомных данных с помощью ПЦР в реальном времени. Интеграция омиксных данных. Протеогеномная коррекция аннотации. Проверка транскриптомных данных с помощью ПЦР в реальном времени. Интеграция омиксных данных.

14. Базы данных. Понятие о форматах. Базы знаний по омиксным данным.

Форматы FASTA и FASTQ. GenBank. Базы данных и репозитории. Основные БД белковых последовательностей. Universal Protein Resource (UniProt). GeneCards. Human Proteome Project. Human Protein Atlas. Blood Atlas. STRING-DB.

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

Учебные аудитории, оснащенные мультимедийным оборудованием (экран, проектор, аудио и видеоаппаратура, ноутбук с подключением к сети «Интернет», микрофоны).

Лабораторное оборудование:

- tandemный масс-спектрометр высокого разрешения Exploris 480 в комплексе с нанохроматографической системой Ultimate 3000;
- tandemный масс-спектрометр QTrap 6500+ с системой ионной подвижности в комплексе с системой УВЭЖХ ExionLC AD.

6. Перечень рекомендуемой литературы

Основная литература

Основная литература предоставлена базовой кафедрой

1. Кулделл Н., Берштейн Р., Ингрэм К., Харт К.М. На пути к синтетической биологии.
2. Аналитическая химия. Методы идентификации и определения веществ: учебное пособие. - 4-е издание

Дополнительная литература

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

Стентон Гланц Медико-биологическая статистика, электронная книга
<https://medstatistic.ru/articles/glantz.pdf>

8. 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 Яндекс телемост. Яндекс диск для доступа к материалам курса. Приветствуется наличие во время занятий смартфонов/ноутбуков для участия в интерактивных упражнениях.

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

Успешное освоение курса требует напряжённой самостоятельной работы студента. В программе курса приведено минимально необходимое время для работы студента над темой.

Самостоятельная работа включает в себя:

- проработку учебного материала (по конспектам лекций, учебной и научной литературе);
- подготовку ответов на вопросы, предназначенных для самостоятельного изучения, доказательство отдельных утверждений, свойств.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению: Биотехнология
профиль подготовки: Системная и синтетическая биология
Физтех-школа Биологической и Медицинской Физики
кафедра системной и синтетической биологии
курс: 1
квалификация: магистр

Семестры, формы промежуточной аттестации:

1 (осенний) - Зачет
2 (весенний) - Экзамен

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

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области физико-математических наук и информационно-коммуникационных технологий	ОПК-4.1 Способен применять знания и навыки по использованию информационно-коммуникационных технологий для поиска и изучения научной литературы, применения прикладных программных продуктов
	ОПК-4.2 Способен применять знания в области физико-математических наук для решения поставленной задачи, формулирования выводов и оценки полученных результатов
	ОПК-4.3 Способен аргументировано выбирать способ проведения научного исследования
ОПК-5 Способен и готов к повышению квалификации, профессиональному росту и руководству коллективом в сфере своей профессиональной деятельности, толерантно воспринимая социальные, этнические, конфессиональные и культурные различия	ОПК-5.1 Способен работать в коллективе, толерантно воспринимая социальные, этнические, конфессиональные и культурные различия
	ОПК-5.2 Владеет навыком руководства малым коллективом в сфере своей профессиональной деятельности
	ОПК-5.3 Стремится к получению новых знаний, профессиональному и личностному росту

2. Показатели оценивания компетенций

В результате изучения дисциплины «Инструментальные основы системной биологии и биостатистика» обучающийся должен:

знать:

- технологии геномного, транскриптомного, протеомного и метаболомного анализов;
- фундаментальные основы экспериментальных методов в молекулярной и системной биологии;
- основы статистики для интерпретации результатов мультиомиксных исследований.

уметь:

- ставить цели и формулировать задачи, понимать поставленные цели и задачи при создании аналитических систем в медико-биологических исследованиях;
- использовать свои знания для решения задач и проведения экспериментальных исследований;
- оценивать корректность постановок задач и строить алгоритмы достижения их оптимального решения в различных (в том числе меняющихся) условиях.

владеть:

- полученные знания и умения должны быть использованы для создания биомедицинских продуктов и технологий.

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

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

1. Редукционизм и холизм в биологии XIX–XX века.
2. Молекулярный состав живой системы и методы его изучения
3. Состав и свойства клетки
4. Принципы статистики в живых системах
5. Принцип неопределенности
6. Понятие синтетической биологии

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

Типовые вопросы для зачета

1. Примеры и основные результаты системных исследований прокариот.
2. Построение математических моделей для описания функции живых систем.
3. Основы системного анализа.
4. Принципы работы масс-спектрометра.
5. Методы количественного анализа в протеомике.
6. Понятие протеогеномики.
7. «Редукционизм» и «холизм» в биологии XIX–XX века.
8. Молекулярный состав живой системы и методы его изучения
9. Построение математических моделей. Принцип неопределенности.
10. Масс-спектрометрия в биологии. Top–down и bottom–up протеомный анализ.

Типовые вопросы для экзамена

1) 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? Опишите последовательность действий при построении и проверке статистической гипотезы. Приведите примеры статистических критериев для проверки гипотезы. Чем отличаются непараметрические и параметрические критерии в статистике?

2) 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? 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) 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?

4) 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?

5) 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?

6) 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?

7) 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. Поиск гомологичных последовательностей, программа BLAST и его варианты. Расскажите про задачу множественного выравнивания, почему стандартные подходы с использованием динамического программирования непригодны

в

8) Введение в транскриптомику. 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? 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 геномная онтология и зачем она нужна?

9) Введение в секвенирование мРНК единичной клетки. 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 одноклеточного секвенирования? 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. Опишите в общих чертах схему анализа данных scRNA-seq.

10) Масс-спектрометрия. Блок-схема масс-спектрометра. Ионы и вакуум. Расчёт отношение массы к заряду. Многозарядные ионы. Изотопы. Изотопное распределение. Разрешающая способность масс-анализатора.

11) Времяпролётный масс-анализатор. Блок-схема масс-спектрометра. Ионы и вакуум. Расчёт отношение массы к заряду. Многозарядные ионы. Принцип работы времяпролётного масс-анализатора. Калибровка масс-анализатора.

12) Квадрупольный масс-анализатор. Блок-схема масс-спектрометра. Ионы и вакуум. Расчёт отношение массы к заряду. Многозарядные ионы. Принцип работы квадрупольного масс-анализатора. Диаграмма стабильности.

13) Tandemная масс-спектрометрия. Блок-схема масс-спектрометра. Ионы и вакуум. Расчёт отношение массы к заряду. Многозарядные ионы. Блок-схема tandemного масс-спектрометра. Родительские и дочерние ионы, спектры фрагментации и их библиотеки. Номенклатура дочерних ионов пептидов.

14) Масс-спектрометрия биомолекул, ESI и ETD. Принцип работы электроспрея. Принцип ETD (диссоциация, индуцированная переносом электрона). Номенклатура дочерних ионов пептидов.

15) Масс-спектрометрия биомолекул, MALDI и CID. Принцип работы MALDI. Принцип CID (диссоциация, индуцированная столкновением). Номенклатура дочерних ионов пептидов.

16) Абсолютный количественный анализ с помощью масс-спектрометрии. Виды количественного анализа. Калибровочная кривая. Способы калибровки с помощью стандартов. Эффект матрицы.

17) Протеомный анализ. SILAC. Общая схема пробоподготовки для протеомного анализа «снизу-вверх». Трипсин. Способы внесения метки в протеомном анализе. Схема SILAC.

18) Протеомный анализ. Изобарные метки. Общая схема пробоподготовки для протеомного анализа «снизу-вверх». Трипсин. Способы внесения метки в протеомном анализе. Принцип работы изобарных меток.

19) Режимы работы tandemного масс-спектрометра. Прицельный анализ методом MRM. Сравнение DDA и DIA (данные-зависимый и данные-независимый анализ).

20) Протеомный анализ. Принцип работы поисковых машин. Оценка FDR. Protein inference, от пептидов к белкам. Изоформы белков.

Примеры билетов:

Билет 1

1. Типы масс-спектрометров. Виды масс-спектрометрического анализа. Масс-спектрометрия в биологии. Top-down и bottom-up протеомный анализ.

2. Интерактомика. Методы определения посттрансляционных модификаций (ПТМ).

Билет 2

1. Байесовский подход в анализе данных. Методы кросс-валидации результатов мультиомиксного анализа. Технология планирования экспериментов для кросс-валидации.
2. Современные методы определения первичной структуры нуклеиновых кислот, мономолекулярное секвенирование, секвенирование единичных клеток, метагеномика, методы анализа паттерна метилирования ДНК.

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.

Оценка «зачтено» выставляется студенту, если он показал 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 «не зачтено» выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

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

При проведении зачета и дифференцированного зачета обучающемуся предоставляется 45 минут на подготовку. Опрос обучающегося не должен превышать одного астрономического часа.