

МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное автономное образовательное учреждение
высшего образования «Российский национальный исследовательский медицинский
университет имени Н.И. Пирогова»**

**Министерства здравоохранения Российской Федерации
ФГАОУ ВО РНИМУ им Н.И.Пирогова Минздрава России (Пироговский Университет)**

Институт биомедицины (МБФ)

УТВЕРЖДАЮ

Директор Института

Прохорчук Егор Борисович

Доктор биологических наук,

Член-корреспондент

Российской академии наук

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

**Б.2.О.У.01 Практика по направлению профессиональной деятельности (анализ данных
высокопроизводительного секвенирования)**

**для образовательной программы высшего образования - программы Магистратуры
по направлению подготовки (специальности)**

06.04.01 Биология

направленность (профиль)

Медицинская биоинформатика

Настоящая рабочая программа дисциплины Б.2.О.У.01 Практика по направлению профессиональной деятельности (анализ данных высокопроизводительного секвенирования) (далее – рабочая программа дисциплины) является частью программы Магистратуры по направлению подготовки (специальности) 06.04.01 Биология. Направленность (профиль) образовательной программы: Медицинская биоинформатика.

Форма обучения: очная

Составители:

№	Фамилия, Имя, Отчество	Учёная степень, звание	Должность	Место работы	Подпись
---	---------------------------	------------------------------	-----------	--------------	---------

Рабочая программа дисциплины рассмотрена и одобрена на заседании кафедры (протокол № _____ от «__» _____ 20__).

Рабочая программа дисциплины рекомендована к утверждению рецензентами:

№	Фамилия, Имя, Отчество	Учёная степень, звание	Должность	Место работы	Подпись
---	---------------------------	------------------------------	-----------	--------------	---------

Рабочая программа дисциплины рассмотрена и одобрена советом института Институт биомедицины (МБФ) (протокол № _____ от «__» _____ 20__).

Нормативно-правовые основы разработки и реализации рабочей программы дисциплины:

1. Федеральный государственный образовательный стандарт высшего образования – магистратура по специальности 06.04.01 Биология, утвержденный приказом Министерства науки и высшего образования Российской Федерации от «11» августа 2020 г. No 934 рук (Далее – ФГОС ВО);
2. Общая характеристика образовательной программы;
3. Учебный план образовательной программы;
4. Устав и локальные акты Университета.

© Федеральное государственное автономное образовательное учреждение высшего образования «Российский национальный исследовательский медицинский университет имени Н.И. Пирогова» Министерства здравоохранения Российской Федерации.

1. Общие положения

1.1. Цель и задачи освоения дисциплины

1.1.1. Цель.

получение обучающимися профессионально-практической подготовки, направленной на формирование умений в области анализа данных высокопроизводительного секвенирования, а также на формирование у студентов целостного представления о содержании, видах и формах профессиональной деятельности.

1.1.2. Задачи, решаемые в ходе освоения программы дисциплины:

- Сформировать у студентов комплекс знаний и навыков, необходимых для проведения анализ данных высокопроизводительного секвенирования, выполнение индивидуальных заданий, предлагаемых руководителем практики.

1.2. Место дисциплины в структуре образовательной программы

Дисциплина «Практика по направлению профессиональной деятельности (анализ данных высокопроизводительного секвенирования)» изучается в 2, 3 семестре (ах) и относится к обязательной части блока Б.2 практика. Является Учебная практика.

Общая трудоемкость дисциплины составляет 8.0 з.е.

Для успешного освоения настоящей дисциплины обучающиеся должны освоить следующие дисциплины: R, биостатистика; Биоинформатика; Молекулярная биология; Практика по профилю профессиональной деятельности (лаборантская практика).

Знания, умения и опыт практической деятельности, приобретенные при освоении настоящей дисциплины, необходимы для успешного освоения дисциплин: Системная биология; Алгоритмы биоинформатики.

Знания, умения и опыт практической деятельности, приобретенные при освоении настоящей дисциплины, необходимы для успешного прохождения практик: Преддипломная практика, в том числе научно-исследовательская работа.

1.3. Планируемые результаты освоения дисциплины

Семестр 2

Код и наименование компетенции	
Код и наименование индикатора достижения компетенции	Планируемые результаты освоения дисциплины (модуля)
ПК-4 Способен планировать и реализовывать проведение научных исследований в области медицинской биоинформатики и смежных областях	
ПК-4.ИД1 Распределяет задачи в рамках исследовательского проекта формирует план научного эксперимента.	Знать: Основные форматы файлов, применяемых в NGS. Принципы использования данных, полученных в исследованиях с использованием NGS, а также основные алгоритмы и программы, используемые для работы с данными секвенирования.
	Уметь: Формировать постановку задачи анализа NGS данных, обосновывает актуальность, значимость, ожидаемые результаты и возможные сферы их применения.
	Владеть практическим опытом (трудовыми действиями): Использования программ для работы с данными секвенирования.

Код и наименование компетенции	
Код и наименование индикатора достижения компетенции	Планируемые результаты освоения дисциплины (модуля)
ПК-4 Способен планировать и реализовывать проведение научных исследований в области медицинской биоинформатики и смежных областях	
ПК-4.ИД1 Распределяет задачи в рамках исследовательского проекта формирует план научного эксперимента.	Знать: Основные форматы файлов, применяемых в NGS. Принципы использования данных, полученных в исследованиях с использованием NGS, а также основные алгоритмы и программы, используемые для работы с данными секвенирования.
	Уметь: Формировать постановку задачи анализа NGS данных
	Владеть практическим опытом (трудовыми действиями): Основные форматы файлов, применяемых в NGS. Принципы использования данных, полученных в исследованиях с использованием NGS, а также основные алгоритмы и программы, используемые для работы с данными секвенирования.

2. Формы работы обучающихся, виды учебных занятий и их трудоёмкость

Формы работы обучающихся / Виды учебных занятий / Формы промежуточной аттестации	Всего часов	Распределение часов по семестрам		
		2	3	
Учебные занятия				
Контактная работа обучающихся с преподавателем в семестре (КР), в т.ч.:	105	51	54	
Семинарское занятие (СЗ)	105	51	54	
Самостоятельная работа обучающихся в семестре (СРО), в т.ч.:	144	90	54	
Подготовка курсовой работы	30	0	30	
Иные виды самостоятельной работы (в т.ч. выполнение практических заданий проектного, творческого и др. типов)	114	90	24	
Промежуточная аттестация (КРПА), в т.ч.:	12	3	9	
Экзамен (Э)	9	0	9	
Зачет (З)	3	3	0	
Подготовка к экзамену (СРПА)	27	0	27	
Общая трудоёмкость дисциплины (ОТД)	в часах: ОТД = КР+СРО+КРПА+СРПА	288	144	144
	в зачетных единицах: ОТД (в часах)/36	8.00	4.00	4.00

3. Содержание дисциплины

3.1. Содержание разделов, тем дисциплины

2 семестр

№ п/п	Шифр компетенции	Наименование раздела (модуля), темы дисциплины	Содержание раздела и темы в дидактических единицах
Раздел 1. Работа с операционной системой Linux (на основе Ubuntu). Анализ результатов высокопроизводительного секвенирования. Поиск, аннотация и интерпретация генетических вариантов.			
1	ПК-4.ИД1	Тема 1. Анализ результатов высокопроизводительного секвенирования. Поиск, аннотация и интерпретация генетических вариантов.	Изучение подходов к анализу данных высокопроизводительного секвенирования. Методы сборки генома. Формат fastq, оценка качества секвенирования. Анализ качества данных, фильтрация и очистка данных. Сборка на основе прохождения по графу перекрытий. Гамильтоновы пути и универсальные строки. de Bruijn граф и проход по графу de Bruijn-a (Эйлеров путь, построение de Bruijn графа из k-меров, мосты Калининграда, Теорема Эйлера). Сборка генома из пар прочтений. Форматы данных, которые используются при сборке геномов (SAM/BAM, VCF/BCF), сортировка данных. Выравнивание на референтный геном (Bowtie2, BWA, SAMTools). Идентификация и аннотация вариантов (varscan2: с использованием данных dbSNP, 1000 Genomes). Визуализация результатов анализа (IGV, BEDTools). Получение навыков работы с наборами программ Genome Analysis Toolkit (GATK), VarScan при анализе данных высокопроизводительного секвенирования.
Раздел 2. Методы изучения генной экспрессии, эпигенетических маркёров и организации хроматина.			
1	ПК-4.ИД1	Тема 1. Сборка генома	Сборка и аннотация прокариотических геномов.

2	ПК-4.ИД1	Тема 2. Методы изучения генной экспрессии, эпигенетических маркёров и организации хроматина.	Картирование транскриптомных данных. Поиск дифференциально экспрессированных генов. Функциональный анализ транскриптомных данных. . Поиск дифференциально метилированных генов. Поиск места посадки транскрипционных факторов. Функциональная аннотация эпигеномных данных.
---	----------	--	---

3.2. Перечень разделов, тем дисциплины для самостоятельного изучения обучающимися

Разделы и темы дисциплины для самостоятельного изучения обучающимися в программе не предусмотрены.

4. Тематический план дисциплины.

4.1. Тематический план контактной работы обучающихся с преподавателем.

№ п/п	Виды учебных занятий / форма пром. аттестации	Период обучения (семестр) Порядковые номера и наименование разделов. Порядковые номера и наименование тем разделов. Темы учебных занятий.	Количество часов контактной работы	Виды контроля успеваемости	Формы контроля успеваемости и промежуточной аттестации	
					КП	РЗ
1	2	3	4	5	6	7
2 семестр						
Раздел 1. Работа с операционной системой Linux (на основе Ubuntu). Анализ результатов высокопроизводительного секвенирования. Поиск, аннотация и интерпретация генетических вариантов.						
Тема 1. Анализ результатов высокопроизводительного секвенирования. Поиск, аннотация и интерпретация генетических вариантов.						
1	СЗ	Общие этапы анализа NGS данных.	3	Т	1	1
2	СЗ	Менеджер пакетов conda. Работа с окружениями. База данных SRA. Форматы данных секвенирования, оценка качества секвенирования. Пакеты SRA-toolkit, FastQC и MultiQC.	3	Т	1	1
3	СЗ	Картирование. Индексирование генома. Выравнивание на референсный геном. Картировщик Bowtie2	3	Т	1	1
4	СЗ	Форматы SAM/BAM. Анализ качества картирования. Визуализация результатов картирования (IGV)	3	Т	1	1

5	СЗ	Возможности Genome Analysis Toolkit (GATK). Предобработка данных картирования.	3	Т	1	1
6	СЗ	Выявление коротких вариантов (SNPs и Indels) в индивидуальных образцах	3	Т	1	1
7	СЗ	Формат VCF. Фильтрация VCF. Аннотация генетических вариантов человека.	3	Т	1	1
8	СЗ	Variant calling на примере HaploTypeCaller.	3	Т	1	1
9	СЗ	Выявление и аннотация митохондриальных генетических вариантов. Определение гаплогруппы.	3	Т	1	1

Раздел 2. Методы изучения генной экспрессии, эпигенетических маркёров и организации хроматина.

Тема 1. Сборка генома

1	СЗ	Сборка генома на основе графа де Брюина. Сборка генома на основе прохождения по графу перекрытий.	3	Т	1	1
2	СЗ	Аннотация геномной сборки.	3	Т	1	1

Тема 2. Методы изучения генной экспрессии, эпигенетических маркёров и организации хроматина.

1	СЗ	Транскриптомика. Методы изучения транскриптома.	3	Т	1	1
2	СЗ	Предобработка данных Bulk RNA-seq. Нормализация. Анализ самосогласованности образцов.	3	Т	1	1
3	СЗ	Анализ самосогласованности образцов. Поиск дифференциально-экспрессированных генов.	3	Т	1	1

4	СЗ	Статистический анализ данных Bulk RNA-seq. Понятие дифференциально экспрессированных генов.	3	Т	1	1
5	СЗ	Функциональный анализ Bulk RNA-Seq.	3	Т	1	1
6	СЗ	Аннотирование результатов анализа Bulk RNA-seq. Анализ обогащения по функциональной принадлежности.	3	Т	1	1

3 семестр

Раздел 1. Методы изучения генной экспрессии, эпигенетических маркёров и организации хроматина.

Тема 1. Методы изучения генной экспрессии, эпигенетических маркёров и организации хроматина.

1	СЗ	Эпигенетика. Подходы к изучению метилирования ДНК.	3	Т	1	1
2	СЗ	Анализ данных бисульфитного секвенирования.	3	Т	1	1
3	СЗ	Предобработка данных бисульфитного секвенирования.	3	Т	1	1
4	СЗ	Поиск дифференциально метилированных регионов.	3	Т	1	1
5	СЗ	Функциональный анализ дифференциально метилированных регионов.	3	Т	1	1
6	СЗ	Функциональный анализ дифференциально метилированных регионов. Пакет HOMER.	3	Т	1	1

7	СЗ	Объединение анализа дифференциальной экспрессии с дифференциальным метилированием.	3	Т	1	1
8	СЗ	Технология ChIP-Seq.	3	Т	1	1
9	СЗ	Предобработка данных ChIP-Seq.	3	Т	1	1
10	СЗ	Пакеты Bioconductor для работы с данными ChIP-Seq исследований.	3	Т	1	1
11	СЗ	Анализ данных ChIP-Seq эксперимента.	3	Т	1	1
12	СЗ	Аннотирование результатов анализа ChIP-Seq.	3	Т	1	1
13	СЗ	Технология секвенирования Singlecell.	3	Т	1	1
14	СЗ	Анализ самосогласованности данных секвенирования Single cell RNA.	3	Т	1	1
15	СЗ	Предобработка данных Single cell RNA.	3	Т	1	1
16	СЗ	Анализ данных секвенирования Singlecell RNA.	3	Т	1	1
17	СЗ	Аннотирование результатов анализа Singlecell RNA.	3	Т	1	1
18	СЗ	Зачётное занятие по практике.	3	Т	1	1

Текущий контроль успеваемости обучающегося в семестре осуществляется в формах, предусмотренных тематическим планом настоящей рабочей программы дисциплины.

Формы проведения контроля успеваемости и промежуточной аттестации обучающихся /виды работы обучающихся

№ п/п	Формы проведения текущего контроля успеваемости и промежуточной аттестации обучающихся (ФТКУ)	Виды работы обучающихся (ВРО)
1	Контроль присутствия (КП)	Присутствие

2	Решение практической (ситуационной) задачи (РЗ)	Решение практической (ситуационной) задачи
---	---	--

4.2. Формы проведения промежуточной аттестации

2 семестр

1) Форма промежуточной аттестации - Зачет

2) Форма организации промежуточной аттестации -Контроль присутствия, Опрос комбинированный

3 семестр

1) Форма промежуточной аттестации - Экзамен

2) Форма организации промежуточной аттестации -Контроль присутствия, Опрос комбинированный

5. Структура рейтинга по дисциплине

5.1. Критерии, показатели проведения текущего контроля успеваемости с использованием балльно-рейтинговой системы.

Рейтинг по дисциплине рассчитывается по результатам текущей успеваемости обучающегося. Тип контроля по всем формам контроля дифференцированный, выставляются оценки по шкале: "неудовлетворительно", "удовлетворительно", "хорошо", "отлично". Исходя из соотношения и количества контролей, рассчитываются рейтинговые баллы, соответствующие системе дифференцированного контроля.

2 семестр

Виды занятий	Формы текущего контроля успеваемости /виды работы	Кол-во контролей	Макс. кол-во баллов	Соответствие оценок рейтинговым баллам ***					
				ТК	ВТК	Отл.	Хор.	Удовл.	
Семинарское занятие	СЗ Решение практической (ситуационной) задачи	РЗ	17	493	В	Т	29	20	10
Сумма баллов за семестр				493					

3 семестр

Виды занятий	Формы текущего контроля успеваемости /виды работы	Кол-во контролей	Макс. кол-во баллов	Соответствие оценок рейтинговым баллам ***					
				ТК	ВТК	Отл.	Хор.	Удовл.	
Семинарское занятие	СЗ Решение практической (ситуационной) задачи	РЗ	18	522	В	Т	29	20	10
Сумма баллов за семестр				522					

5.2. Критерии, показатели и порядок промежуточной аттестации обучающихся с использованием балльно-рейтинговой системы. Порядок перевода рейтинговой оценки обучающегося в традиционную систему оценок

Порядок промежуточной аттестации обучающегося по дисциплине (модулю) в форме зачёта

По итогам расчета рейтинга по дисциплине в 2 семестре, обучающийся может быть аттестован по дисциплине без посещения процедуры зачёта, при условии:

Оценка	Рейтинговый балл
Зачтено	291

Порядок промежуточной аттестации обучающегося по дисциплине (модулю) в форме экзамена

По итогам расчета рейтинга по дисциплине в 3 семестре, обучающийся может быть аттестован с оценками «отлично» (при условии достижения не менее 90% баллов из возможных), «хорошо» (при условии достижения не менее 75% баллов из возможных), «удовлетворительно» (при условии достижения не менее 60% баллов из возможных) и сданных на оценку не ниже «удовлетворительно» всех запланированных в текущем семестре рубежных контролей без посещения процедуры экзамена. В случае, если обучающийся не согласен с оценкой, рассчитанной по результатам итогового рейтинга по дисциплине, он обязан пройти промежуточную аттестацию по дисциплине в семестре в форме экзамена в порядке, предусмотренном рабочей программой дисциплины и в сроки, установленные расписанием экзаменов в рамках экзаменационной сессии в текущем семестре. Обучающийся заявляет о своем желании пройти промежуточную аттестацию по дисциплине в форме экзамена не позднее первого дня экзаменационной сессии, сделав соответствующую отметку в личном кабинете по соответствующей дисциплине. В таком случае, рейтинг, рассчитанный по дисциплине не учитывается при процедуре промежуточной аттестации. По итогам аттестации обучающийся может получить любую оценку из используемых в учебном процессе: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Оценка	Рейтинговый балл
Отлично	900
Хорошо	750
Удовлетворительно	600

6. Фонд оценочных средств по дисциплине (модулю) для проведения текущего контроля и промежуточной аттестации

2 семестр

Перечень практических умений и навыков для подготовки к промежуточной аттестации в форме зачёта

1. Обобщенная схема анализа NGS данных. Преаналитический этап анализа NGS: выделение биоматериала, фрагментация и обогащение. Влияние преаналитического этапа на дальнейший анализ.
2. Технология секвенирования Illumina. Вариации строения ампликона Illumina.
3. Base-calling процедура. Структура fastq файла. Шкала Phred. Базы данных NGS.
4. Анализ качества данных секвенирования. Отчёт fastqc. Программа Multiqc.
5. Факторы, влияющие на качество NGS данных.
6. Картирование и выравнивание. Общая схема картирования. Картировщики: принцип работы и настраиваемые параметры.
7. Понятие референсного генома. Референсный геном человека. Различия в версиях референсного генома человека.
8. Подходы к индексированию. Значение индексирования. Проблемы индексирования. Конструкции индекса: Суффиксный бор, суффиксное дерево, суффиксный массив, хеш-таблицы. Примеры использования в картировщиках.
9. Индексирование референса. Прямое и обратное преобразование Барроуза-Уилера (BWT). Примеры использования в картировщиках.
10. Индексирование референса. FM-индекс. Сравнение подходов к индексированию.
11. Картирование. Подходы к поиску seeds, примеры использования в картировщиках.
12. Выравнивание биологических последовательностей. Алгоритмы глобального и локального выравнивания. Примеры использования в картировщиках.
13. Классификация выравнивания прочтения. Взаиморасположение прочтений. Особенности картирования с учётом интронов, примеры использования в картировщиках.

14. Форматы SAM и BAM. Средства работы с sam/bam файлами. Возможности пакета samtools.
15. Анатомия формата SAM. Содержимое и значение заголовка SAM. Понятия mapQ, размер вставки. Строение строки CIGAR.
16. Анализ качества картирования. Понятие покрытия. Факторы, влияющие на качество картирования.
17. Визуализация и интерпретация показателей качества картирования на примере samtools stats. Integrative Genome Viewer. Qualimap.
18. Классификации генетических вариантов. Общий алгоритм поиска генетических вариантов. GATK4. Предварительная обработка bam файлов.
19. Понятие дубликатов в данных секвенирования. Поиск дубликатов на примере MarkDuplicates.
20. Программы Variant Calling. Принцип работы на примере HaplotypeCaller. Поиск активных регионов. Значение локальной сборки. Оценка гаплотипа.
21. Оценка вероятности генотипа, теорема Байеса. 22. Variant Calling Format. Анатомия vcf-файла. Формат gVCF.
23. Средства работы с vcf. Возможности bcftools. Подходы к фильтрации vcf-файлов.
24. Аннотация vcf-файлов. Базы данных для аннотирования. Типичные отличия индивидуального генома от референса.
25. Клиническая классификация генетических вариантов: назначение, группы доказательности. Присвоение классификации. 2
6. Клиническая классификация генетических вариантов: популяционные данные
27. Клиническая классификация генетических вариантов: информация о состоянии аллелей, структуре и функции белка. Фазирование генотипа. Связь варианта с потерей функции. Компьютерные методы оценки патогенности.
28. Митохондриальная ДНК. Понятие гетероплазмии. Митохондриальный референс. Анализ гаплогрупп митохондриальной ДНК.
29. Основные этапы в сборке генома de novo. Понятие покрытия в сборке генома, причины неравномерности покрытия. Сложности в сборке генома de novo.

30. Технологии секвенирования длинных прочтений. Принципы работы, преимущества и недостатки. Особенности оценки качества.
31. Оценка размера генома. Редактирование ошибок в прочтениях.
32. Подходы к сборке контигов. Параметры сборки контигов. Скаффолдинг.
33. Overlap Layout Consensus. Принцип построения и обхода. Сборщики, основанные на OLC.
34. Графы де Брёйна. Принцип построения и обхода. Сборщики, основанные на графе де Брёйна.
35. Сравнение подходов к сборке генома. Сборщики, использующие гибридный подход.
36. Понятие транскриптома. Особенности его сборки. Программы для сборки транскриптома.
37. Оценка качества геномной сборки: программы и метрики.
38. Аннотация геномной сборки. Принцип, форматы, базы данных, программные средства. Различия в подходах к аннотации прокариотических и эукариотических геномов.
39. Транскриптомика. Подходы к изучению транскриптомики. Общая схема анализа RNA-seq. Основные методы подготовки РНК библиотек. Понятие реплик. Консорциум ENCODE.
40. Базы данных в транскриptomике. Стратегии анализа транскриptomных данных.
41. Анализ качества РНК библиотеки.
42. Подходу к выравниванию транскриptomных данных. Картирование прочтений на экзон-экзонные границы. Примеры картировщиков.
43. Анализ качества картирования транскриptomных данных. Программы, метрики.
44. Картирование и подсчёт экспрессии на примере STAR. Входные и выходные данные, принцип работы.
45. Альтернативный сплайсинг и оценка геной экспрессии. Подходы к учёту экспрессии изоформ.
46. Псевдокартирование на примере kallisto. Преимущества и недостатки псевдокартировщиков. Выходные данные kallisto quant.
47. Нормализация каунтов. Причины необходимости нормализации каунтов и способы.

48. Проверка самосогласованности данных RNA-seq. Входные данные, описательная статистика, корреляционный анализ и PCA.

49. Понятие дифференциальной экспрессии. Пакеты для анализа дифференциальной экспрессии: входные и выходные данные, принцип работы, настраиваемые параметры.

50. Функциональный анализ. Архитектура Gene Ontology. Подходы к анализу GO enrichment.

3 семестр

Перечень вопросов для подготовки к промежуточной аттестации в форме экзамена

1. Обобщенная схема анализа NGS данных. Преаналитический этап анализа NGS: выделение биоматериала, фрагментация и обогащение. Влияние преаналитического этапа на дальнейший анализ.

2. Технология секвенирования Illumina. Вариации строения ампликона Illumina.

3. Base-calling процедура. Структура fastq файла. Шкала Phred. Базы данных NGS.

4. Анализ качества данных секвенирования. Отчёт fastqc. Программа Multiqc.

5. Факторы, влияющие на качество NGS данных.

6. Картирование и выравнивание. Общая схема картирования. Картировщики: принцип работы и настраиваемые параметры.

7. Понятие референсного генома. Референсный геном человека. Различия в версиях референсного генома человека.

8. Подходы к индексированию. Значение индексирования. Проблемы индексирования. Конструкции индекса: Суффиксный бор, суффиксное дерево, суффиксный массив, хеш-таблицы. Примеры использования в картировщиках.

9. Индексирование референса. Прямое и обратное преобразование Барроуза-Уилера (BWT). Примеры использования в картировщиках.

10. Индексирование референса. FM-индекс. Сравнение подходов к индексированию.
11. Картирование. Подходы к поиску seeds, примеры использования в картировщиках.
12. Выравнивание биологических последовательностей. Алгоритмы глобального и локального выравнивания. Примеры использования в картировщиках.
13. Классификация выравнивания прочтения. Взаиморасположение прочтений. Особенности картирования с учётом интронов, примеры использования в картировщиках.
14. Форматы SAM и BAM. Средства работы с sam/bam файлами. Возможности пакета samtools.
15. Анатомия формата SAM. Содержимое и значение заголовка SAM. Понятия mapQ, размер вставки. Строение строки CIGAR.
16. Анализ качества картирования. Понятие покрытия. Факторы, влияющие на качество картирования.
17. Визуализация и интерпретация показателей качества картирования на примере samtools stats. Integrative Genome Viewer. Qualimap.
18. Классификации генетических вариантов. Общий алгоритм поиска генетических вариантов. GATK4. Предварительная обработка bam файлов.
19. Понятие дубликатов в данных секвенирования. Поиск дубликатов на примере MarkDuplicates.
20. Программы Variant Calling. Принцип работы на примере HaplotypeCaller. Поиск активных регионов. Значение локальной сборки. Оценка гаплотипа.
21. Оценка вероятности генотипа, теорема Байеса. 22. Variant Calling Format. Анатомия vcf-файла. Формат gVCF.

23. Средства работы с vcf. Возможности bcftools. Подходы к фильтрации vcf-файлов.

24. Аннотация vcf-файлов. Базы данных для аннотирования. Типичные отличия индивидуального генома от референса.

25. Клиническая классификация генетических вариантов: назначение, группы доказательности. Присвоение классификации. 2

6. Клиническая классификация генетических вариантов: популяционные данные

27. Клиническая классификация генетических вариантов: информация о состоянии аллелей, структуре и функции белка. Фазирование генотипа. Связь варианта с потерей функции. Компьютерные методы оценки патогенности.

28. Митохондриальная ДНК. Понятие гетероплазмии. Митохондриальный референс. Анализ гаплогрупп митохондриальной ДНК.

29. Основные этапы в сборке генома de novo. Понятие покрытия в сборке генома, причины неравномерности покрытия. Сложности в сборке генома de novo.

30. Технологии секвенирования длинных прочтений. Принципы работы, преимущества и недостатки. Особенности оценки качества.

31. Оценка размера генома. Редактирование ошибок в прочтениях.

32. Подходы к сборке контигов. Параметры сборки контигов. Скаффолдинг.

33. Overlap Layout Consensus. Принцип построения и обхода. Сборщики, основанные на OLC.

34. Графы де Брёйна. Принцип построения и обхода. Сборщики, основанные на графе де Брёйна.

35. Сравнение подходов к сборке генома. Сборщики, использующие гибридный подход.

36. Понятие транскриптома. Особенности его сборки. Программы для сборки транскриптома.
37. Оценка качества геномной сборки: программы и метрики.
38. Аннотация геномной сборки. Принцип, форматы, базы данных, программные средства. Различия в подходах к аннотации прокариотических и эукариотических геномов.
39. Транскриптомика. Подходы к изучению транскриптомики. Общая схема анализа RNA-seq. Основные методы подготовки РНК библиотек. Понятие реплик. Консорциум ENCODE.
40. Базы данных в транскриptomике. Стратегии анализа транскриptomных данных.
41. Анализ качества РНК библиотеки.
42. Подходу к выравниванию транскриptomных данных. Картирование прочтений на экзон-экзонные границы. Примеры картировщиков.
43. Анализ качества картирования транскриptomных данных. Программы, метрики.
44. Картирование и подсчёт экспрессии на примере STAR. Входные и выходные данные, принцип работы.
45. Альтернативный сплайсинг и оценка геной экспрессии. Подходы к учёту экспрессии изоформ.
46. Псевдокартирование на примере kallisto. Преимущества и недостатки псевдокартировщиков. Выходные данные kallisto quant.
47. Нормализация каунтов. Причины необходимости нормализации каунтов и способы.
48. Проверка самосогласованности данных RNA-seq. Входные данные, описательная статистика, корреляционный анализ и PCA.

49. Понятие дифференциальной экспрессии. Пакеты для анализа дифференциальной экспрессии: входные и выходные данные, принцип работы, настраиваемые параметры.
50. Функциональный анализ. Архитектура Gene Ontology. Подходы к анализу GO enrichment.
51. Эпигенетика. Методы изучения метилирования современной ДНК. Бисульфитная конверсия: принцип, подходы к анализу, примеры программ.
52. Оценки качества данных бисульфитного секвенирования. Требования к дизайну исследования метилирования ДНК.
53. Особенности выравнивания бисульфитного секвенирования. Программы, входные и выходные данные. Оценка качества выравнивания RRBS данных. M-bias plot.
54. Принцип определения метилированных участков. Расчёт уровня метилирования. Средний уровень метилирования генома.
55. Подходы к поиску дифференциально метилированных регионов. Типы дифференциально метилированных регионов. Функциональный анализ дифференциально метилированных регионов.
56. Технология ChIP-Seq. Возможности и принцип работы. Контроли в ChIP-Seq. Требования к дизайну исследования. Оценка качества данных ChIP-Seq эксперимента.
57. Peak-calling в анализе ChIP-Seq. Программы, пакеты, входные и выходные данные. Типы получаемых пиков и их интерпретация.
58. Аннотация и функциональный анализ данных ChIP-Seq.
59. Технология Single Cell RNA-Seq.

Экзаменационный билет для проведения экзамена

--

Федеральное государственное автономное образовательное учреждение
высшего образования «Российский национальный исследовательский медицинский
университет
имени Н.И. Пирогова» Министерства здравоохранения Российской Федерации
ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России (Пироговский Университет)
Экзаменационный билет № _____

для проведения экзамена по дисциплине Б.2.О.У.01 Практика по направлению
профессиональной деятельности (анализ данных высокопроизводительного
секвенирования)
по программе Магистратуры
по направлению подготовки (специальности) 06.04.01 Биология
направленность (профиль) Медицинская биоинформатика

Федеральное государственное автономное образовательное учреждение
высшего образования «Российский национальный исследовательский медицинский
университет
имени Н.И. Пирогова» Министерства здравоохранения Российской Федерации
(ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России)

Кафедра биоинформатики

Экзаменационный билет № 1

для проведения экзамена по дисциплине

«Алгоритмы биоинформатики»

по направлению подготовки 06.04.01 Биология

1. Картирование. Подходы к поиску seeds, примеры использования в картировщиках
2. Нормализация каунтов. Причины необходимости нормализации каунтов и способы.

Заведующий кафедрой _____ *Фамилия, Инициалы*

Заведующий Лагунин Алексей Александрович
Кафедра биоинформатики МБФ

7. Методические указания обучающимся по освоению дисциплины

Для подготовки к занятиям семинарского типа обучающийся должен

1. Ознакомиться с теоретическим материалом, изложенным в учебниках, методических пособиях и в личном кабинете студента.
2. Выполнить письменные домашние задания для подготовки к занятию.
3. Ознакомиться и оформить лабораторную работу по теме занятия.
4. Подготовить конспект к занятию по изучаемой теме.

При подготовке к экзамену необходимо

1. Ознакомиться с примером билета, выносимого на промежуточную аттестацию в форме зачета;
2. Ознакомиться с тестовыми заданиями в личном кабинете студента;
3. Проанализировать материал и наметить последовательность его повторения;
4. Определить наиболее простые и сложные темы и (или) разделы дисциплины;
5. Повторить материал по наиболее значимым/сложным темам и (или) разделам дисциплины по конспектам лекций и учебной литературе, а также электронным образовательным ресурсам.

Самостоятельная работа студентов (СРС) включает в себя

Закрепление и углубление полученных знаний, умений и навыков, поиск и приобретение новых знаний, выполнение учебных заданий, подготовку к предстоящим занятиям, текущему контролю успеваемости и промежуточной аттестации.

Выполнение домашних заданий осуществляется в форме:

- работы с учебной, учебно-методической и научной литературой, электронными образовательными ресурсами (например, просмотр видеолекций), конспектами обучающегося: чтение, изучение, анализ, сбор и обобщение информации, её конспектирование; подготовки ответов на вопросы.
- решения задач, выполнения письменных заданий и упражнений.

8. Учебно-методическое, информационное и материально-техническое обеспечение дисциплины

8.1. Перечень литературы по дисциплине:

№ п /п	Наименование, автор, год и место издания	Используется при изучении разделов	Количество экземпляров в библиотеке	Электронный адрес ресурсов
1	2	3	4	5
1	NGS. Высокопроизводительное секвенирование, Ребриков Д. В., 2020	Методы изучения геной экспрессии, эпигенетических маркёров и организации хроматина. Работа с операционной системой Linux (на основе Ubuntu). Анализ результатов высокопроизводительного секвенирования. Поиск, аннотация и интерпретация генетических вариантов.	0	https://rsmu.informsystema.ru/uploader/fileUpload?name=86bn.pdf&show=dcatalogues/1/5053/86bn.pdf&view=true

8.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», в том числе профессиональных баз данных, необходимых для освоения дисциплины (модуля)

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

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

1. Автоматизированный информационный комплекс «Цифровая административно-образовательная среда РНИМУ им. Н.И. Пирогова»
2. Система управления обучением
3. ОС Ubuntu
4. Anaconda
5. R Studio
6. IGV: Integrative Genomics Viewer

8.4. Материально-техническое обеспечение дисциплины (модуля)

Каждый обучающийся в течение всего периода обучения обеспечен индивидуальным неограниченным доступом к электронной информационно-образовательной среде университета из любой точки, в которой имеется доступ к информационно-телекоммуникационной сети «Интернет» (далее - сеть «Интернет»), как на территории Университета, так и вне ее.

Электронная информационно-образовательная среда университета обеспечивает:

- доступ к учебному плану, рабочей программе дисциплины, электронным учебным изданиям и электронным образовательным ресурсам, указанным в рабочей программе дисциплины;

- формирование электронного портфолио обучающегося, в том числе сохранение его работ и оценок за эти работы.

Университет располагает следующими видами помещений и оборудования для материально-технического обеспечения образовательной деятельности для реализации образовательной программы дисциплины (модуля):

№ п /п	Наименование оборудованных учебных аудиторий	Перечень специализированной мебели, технических средств обучения
1	Аудитории для проведения занятий лекционного типа, занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оборудованная мультимедийными и иными средствами обучения	
2	Аудитории, оборудованные мультимедийными и иными средствами обучения	Компьютерная техника с возможностью подключения к сети “Интернет”, Стулья, Столы, Компьютер персональный, Проектор мультимедийный, Компьютеры для обучающихся, Экран для проектора, Электрические розетки и подсетка на каждом столе
3	Учебные аудитории для проведения промежуточной аттестации	Компьютерная техника с возможностью подключения к сети “Интернет”, Стулья, Столы, Компьютер персональный, Проектор мультимедийный, Электрические розетки и подсетка на каждом столе, Компьютеры для обучающихся, Экран для проектора

4	Помещения для самостоятельной работы обучающихся, оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду организации	учебная мебель (столы, стулья), компьютерная техника с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду
---	--	--

Университет обеспечен необходимым комплектом лицензионного и свободно распространяемого программного обеспечения (состав определяется в рабочей программе дисциплины и подлежит обновлению при необходимости). Библиотечный фонд укомплектован печатными изданиями из расчета не менее 0,25 экземпляра каждого из изданий, указанных в рабочей программе дисциплины, на одного обучающегося из числа лиц, одновременно осваивающих соответствующую дисциплину.

Обучающимся обеспечен доступ (удаленный доступ), в том числе в случае применения электронного обучения, дистанционных образовательных технологий, к современным профессиональным базам данных и информационным справочным системам, состав которых определяется в рабочей программе дисциплины и подлежит обновлению (при необходимости).

Обучающиеся из числа инвалидов обеспечены печатными и (или) электронными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья.

Приложение 1
к рабочей программе
дисциплины (модуля)

Сведения об изменениях в рабочей программе дисциплины (модуля)

для образовательной программы высшего образования – программы бакалавриата/специалитета /магистратуры (оставить нужное) по направлению подготовки (специальности) (оставить нужное) _____ (код и наименование направления подготовки (специальности)) направленность (профиль) « _____ » на _____ учебный год.

Рабочая программа дисциплины с изменениями рассмотрена и одобрена на заседании кафедры _____ (Протокол № _____ от « ____ » _____ 20 ____).

Заведующий _____ кафедрой _____ (подпись)
_____ (Инициалы и фамилия)

Приложение 2
к рабочей программе
дисциплины (модуля)

Формы проведения текущего контроля успеваемости и промежуточной аттестации

Формы проведения текущего контроля успеваемости и промежуточной аттестации	Сокращённое наименование	
Контроль присутствия	Присутствие	КП
Решение практической (ситуационной) задачи	Практическая задача	РЗ

Виды учебных занятий и формы промежуточной аттестации

Формы проведения текущего контроля успеваемости и промежуточной аттестации	Сокращённое наименование	
Семинарское занятие	Семинар	СЗ
Экзамен	Экзамен	Э
Зачет	Зачет	З

Виды контроля успеваемости

Формы проведения текущего контроля успеваемости и промежуточной аттестации	Сокращённое наименование	
Текущий тематический контроль	Тематический	Т
Промежуточная аттестация	Промежуточная аттестация	ПА