

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Нижегородский государственный
университет им. Н.И. Лобачевского»
Институт биологии и биомедицины

УТВЕРЖДЕНО
решением ученого совета ННГУ
протокол от
«16» июня 2021 г. №8

Рабочая программа дисциплины

Базы данных

(наименование дисциплины (модуля))

Уровень высшего образования

Специалитет

Направление подготовки / специальность

30.05.03 Медицинская кибернетика

Направленность образовательной программы

(указывается профиль / магистерская программа / специализация)

Форма обучения

Очная

Нижегород
2021 год

1. Место и цели дисциплины в структуре ОПОП

Дисциплина «Бзы данных» к базовым дисциплинам цикла Б.1 профессиональный) ОПОП по специальности 30.05.03 Медицинская кибернетика. Дисциплина обязательна для освоения в 8 семестре.

Студенты к моменту освоения дисциплины «базы данных » демонстрируют базовые представления о строении и функциях ДНК и РНК, методиках секвенирования и анализирования геномов, полученные при изучении таких дисциплин специалитета как «Биология», «Цитология», «Анализ данных», «Научное программирование». Студенты имеют навыки работы с языком программирования «R», имеют опыт работы поиска достоверной научной информации в международных базах данных и составления презентаций и сообщений.

Целями освоения дисциплины «базы данных» являются:

1. Ознакомление с основными базами данных и методами работы с ними.
2. Изучение основных принципов биоинформатики.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников)

Формируемые компетенции (код компетенции, уровень освоения – при наличии в карте компетенции)	Планируемые результаты обучения по дисциплине (модулю), характеризующие этапы формирования компетенций
<p><i>(ПК-1)</i> способен творчески использовать в научной и производственно-технологической деятельности знания фундаментальных и прикладных разделов дисциплин (модулей), определяющих направленность (профиль) программы магистратуры <i>(Этап формирования - начальный)</i></p>	<p><i>З (ПК-1) – знать</i> механизмы транскрипции генов; методы секвенирования; способы обработки данных; полученных при секвенировании; основные базы данных; современные представления о строении и функционировании хромосом, свойства генетического кода.</p> <p><i>У (ПК-1) уметь</i> работать в основных базах данных, применять методы анализа данных.</p> <p><i>В (ПК-1) – владеть</i> методами анализа данных РНК-секвенирования.</p>

3. Структура и содержание дисциплины (модуля)

Объем дисциплины (модуля) составляет 2 зачетных единицы, всего 72 часа, из которых 65 часов составляет контактная работа обучающегося с преподавателем (32 часа занятия лекционного типа, 32 часа практические занятия, 1 час мероприятия промежуточной аттестации), 7 часов составляет самостоятельная работа обучающегося (включая 36 часов подготовки к экзамену).

Наименование и краткое содержание разделов и тем дисциплины (модуля), форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	в том числе	
		контактная работа (работа во взаимодействии с преподавателем), часы из них	Самостоятельная работа

		Занятия лекционного типа	Занятия практического типа	Всего	
Тема 1. Направления и задачи биоинформатики. Центральная догма молекулярной биологии	9	4	4	8	1
Тема 2. Секвенирование	9	4	4	8	1
Тема 3. Алгоритм Смита-Ватермана. Матрицы штрафов. Эвристические подходы и оптимизации	9	4	4	8	1
Тема 4. Первичная обработка RNA-seq. Матрица экспрессии генов.	9	4	4	8	1
Тема 5. Поиск дифференциально экспрессированных генов в двух группах (непарный и парный). Множественное тестирование и методы коррекции	9	4	4	8	1
Тема 6. Корреляционный анализ. Графовое представление взаимодействий.	13	6	6	12	1
Тема 7. Матрица экспрессии генов. Классификация. Линейная регрессия	13	6	6	12	1
<i>В т.ч. текущий контроль</i>	1				
Промежуточная аттестация – зачет (36 часов)					

Текущий контроль успеваемости реализуется в рамках практических занятий. Промежуточный контроль осуществляется при проведении экзамена.

4. Образовательные технологии

В соответствии с рабочей программой и тематическим планом изучение дисциплины проходит в виде аудиторной и самостоятельной работы студентов. Учебный процесс в аудитории осуществляется в форме лекционных и практических занятий.

Для достижения планируемых результатов обучения в дисциплине «базы данных и основы биоинформатики» используются следующие образовательные технологии:

1. *Традиционные технологии:* информационные лекции, семинары.
2. *Информационно-коммуникационные технологии:* лекции-визуализации, беседы, доклады, презентации, различные формы самостоятельной работы студентов (самостоятельное изучение литературы, составление опорных конспектов, подготовка докладов).
3. *Интерактивные технологии:* семинары-дискуссии
4. *Технологии проблемного обучения:* лекции с проблемным изложением материала, учебные дискуссии.

5. Учебно-методическое обеспечение самостоятельной работы обучающихся

Основные виды самостоятельной работы студентов в рамках освоения дисциплины:

- изучение понятийного аппарата дисциплины;
- проработка тем дисциплины;
- работа с основной и дополнительной литературой;
- самоподготовка к занятиям семинарского типа;
- работа в библиотеке;
- изучение сайтов по темам дисциплины в сети Интернет;
- подготовка к экзамену.

5.1 Методические указания по подготовке студентов к текущему и промежуточному контролю по дисциплине «базы данных и основы биоинформатики»

Подготовка к экзамену.

Промежуточная аттестация по итогам освоения дисциплины проходит в форме экзамена.

Подготовка к экзамену является концентрированной систематизацией всех полученных знаний по дисциплине «базы данных и основы биоинформатики».

В начале семестра рекомендуется внимательно изучить перечень вопросов к экзамену по данной дисциплине, а также использовать в процессе обучения программу, другие методические материалы, разработанные кафедрой по данной дисциплине. Это позволит в процессе изучения тем сформировать более правильное и обобщенное видение студентом существа того или иного вопроса за счет:

- а) уточняющих вопросов преподавателю;
- б) подготовки докладов по отдельным темам;
- в) самостоятельного уточнения вопросов на смежных дисциплинах;
- г) углубленного изучения вопросов темы по учебным пособиям.

После изучения соответствующей тематики рекомендуется проверить наличие и формулировки вопроса по этой теме в перечне вопросов к экзамену, а также попытаться изложить ответ на этот вопрос. Если возникают сложности при раскрытии материала, следует вновь обратиться к лекционному материалу, материалам практических занятий, уточнить терминологический аппарат темы, а также проконсультироваться с преподавателем.

Вопросы к экзамену приведены в п. 6.4. данной рабочей программы.

6. Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю), включающий:

6.1 Перечень компетенций выпускников образовательной программы с указанием результатов обучения (знаний, умений, владений), характеризующих этапы их формирования, описание показателей и критериев оценивания компетенций на различных этапах их формирования.

(ПК-1): способность творчески использовать в научной и производственно-технологической деятельности знания фундаментальных и прикладных разделов дисциплин (модулей), определяющих направленность (профиль) программы магистратуры. Начальный этап формирования

Индикаторы компетенции	Критерии оценивания						
	«плохо»	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«очень хорошо»	«отлично»	«превосходно»

Знать: механизмы транскрипции генов; методы секвенирования; способы обработки данных; полученных при секвенировании; основные базы данных; современные представления о строении и функционировании и хромосом, свойства генетического кода	Отсутствие знаний материала	Наличие грубых ошибок в основном материале	Знание основного материала при наличии ошибок	Знание основного материала с заметными погрешностями	Знание основного материала с незначительными погрешностями	Знание основного материала без ошибок	Знание основного материала и дополнительного материала без ошибок
Уметь: работать в основных базах данных, применять методы анализа данных	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продемонстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания но не в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения, решены все основные задачи с отдельными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
Владеть: методами анализа данных РНК-секвенирования.	Отсутствие минимальных владений. Невозможность оценить владение вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы владение материалом. Имели место грубые ошибки	Продемонстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания но не в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения, решены все основные задачи с отдельными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
Шкала оценок по проценту правильно выполненных заданий	0-20%	21-50%	51-70%	71-80%	81-90%	91-99%	100%

(ПК-2): способность планировать и реализовывать профессиональные мероприятия (в соответствии с направленностью (профилем) программы магистратуры).

Начальный этап формирования

Индикаторы компетенции	Критерии оценивания						
	«плохо»	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«очень хорошо»	«отлично»	«превосходно»
Знать: основные алгоритмы в биоинформатике; методы сборки геномов; строение и функции биополимеров, их компонентов и комплексов.	Отсутствие знаний материала	Наличие грубых ошибок в основном материале	Знание основного материала при наличии ошибок	Знание основного материала с заметными погрешностями	Знание основного материала с незначительными погрешностями	Знание основного материала без ошибок	Знание основного и дополнительного материала без ошибок
Уметь: выполнять неточный поиск подпоследовательностей в больших текстах; составлять алгоритмы для решения задач биоинформатики; применять общие схемы разработки программ для реализаций собственных алгоритмов	Отсутствие минимальных владений. Невозможность оценить владение вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы владение материалом. Имели место грубые ошибки	Продемонстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения, решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
Владеть: основами алгоритмических методов биоинформатики и разработки их программных реализаций.	Отсутствие минимальных владений. Невозможность оценить владение вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы владение материалом. Имели место грубые ошибки	Продемонстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения, решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
Шкала оценок по проценту правильно выполненных заданий	0-20%	21-50%	51-70%	71-80%	81-90%	91-99%	100%

6.2 Описание шкал оценивания результатов обучения по дисциплине

Промежуточный контроль качества усвоения студентами содержания дисциплины проводится в виде экзамена, на котором определяется:

- уровень усвоения студентами основного учебного материала по дисциплине;
- уровень понимания студентами изученного материала;
- способности студентов использовать полученные знания для решения

конкретных задач.

Экзамен проводится в устной форме в виде ответа обучающегося на теоретические вопросы курса (с предварительной подготовкой), а также выполнения практических заданий. Собеседование проводится в форме вопросов, на которые студент должен дать

краткий ответ. Практическая часть экзамена предусматривает беседу по подготовленному проекту. Итоговая оценка за экзамен получается как средняя арифметическая из теоретического ответа по билету и оценки за выполнение практических заданий в течении семестра.

Шкала оценивания устного ответа на экзамене

Оценка	Уровень подготовки
Превосходно	Высокий уровень подготовки, безупречное владение теоретическим материалом, студент демонстрирует творческий подход к решению нестандартных ситуаций. Студент дал полный и развернутый ответ на все теоретические вопросы билета. Студент активно работал на практических занятиях, чему подтверждением является высокий средний балл за текущую успеваемость.
Отлично	Высокий уровень подготовки с незначительными ошибками. Студент дал полный и развернутый ответ на все теоретические вопросы билета. Студент активно работал на практических занятиях, чему подтверждением является высокий средний балл за текущую успеваемость.
Очень хорошо	Хорошая подготовка. Студент дал полный ответ на все теоретические вопросы билета, но допустил небольшие неточности в определениях понятий, процессов и т.п. Студент активно работал на практических занятиях, имеет высокие средний балл за текущую успеваемость.
Хорошо	В целом хорошая подготовка с заметными ошибками или недочетами. Студент дал ответ на все теоретические вопросы билета, но допустил неточности в определениях понятий, процессов и т.п. Имеются ошибки при ответах на дополнительные и уточняющие вопросы экзаменатора. Студент работал на практических занятиях, имеет хорошие средний балл за текущую успеваемость.
Удовлетворительно	Минимально достаточный уровень подготовки. Студент показал минимальный уровень теоретических знаний, сделал существенные ошибки при ответе на экзаменационный вопрос, но при ответах на наводящие вопросы, смог правильно сориентироваться и в общих чертах дать правильный ответ. Студент посещал практические занятия, но имеет низкие средний балл за текущую успеваемость.
Неудовлетворительно	Подготовка недостаточная и требует дополнительного изучения материала. Студент дал ошибочные ответы, как на теоретические вопросы билета, так и на наводящие и дополнительные вопросы экзаменатора. Студент посещал практические занятия, но имеет очень низкий средний балл за текущую успеваемость.

Плохо	Студент отказался отвечать на экзаменационный билет.
-------	--

*информация предоставляется преподавателем, ведущим практические занятия.

Критерии оценивания практического задания

Предел длительности контроля	10-15 минут
Предлагаемое количество заданий	1-2
Критерии оценки:	
«отлично»	Ответ на вопросы задания изложен четко и логично, является достоверным; содержит элементы анализа. Практические действия выполнены верно

«хорошо»	Ответ на вопросы задания, в целом, изложен четко и логично, является достоверным; не содержит элементы анализа и носит описательный характер. Практические действия выполнены верно
«удовлетворительно»	Ответ на вопросы задания неполный и поверхностный. В практических действиях имеются ошибки
«неудовлетворительно»	Ответ на вопросы задания неполный и поверхностный. Практические действия не выполнены

6.3 Критерии и процедуры оценивания результатов обучения по дисциплине (модулю), характеризующих сформированность компетенций

Для оценивания результатов обучения в виде знаний используются следующие процедуры и технологии:

- собеседование по вопросам на экзамене.

Для оценивания результатов обучения в виде умений и владений используются следующие процедуры и технологии:

- практические задания

6.4 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения, характеризующих сформированность компетенций и (или) для итогового контроля сформированности компетенции.

6.4.1 Для оценки сформированности знаний компетенции ПК-1 и ПК-2 используются:

***1. Собеседование (устный опрос)** по вопросам, выносимым на практические занятия и экзамен.*

Устный опрос используется для контроля знаний студентов в качестве проверки результатов освоения основных категорий, принципов и закономерностей по дисциплине «базы данных и основы биоинформатики».

Вопросы для подготовки к экзамену.

1. Центральная догма молекулярной биологии
2. Секвенирование. ДНК-секвенирование. РНК-секвенирование
3. Сборка геномов de novo, с использованием референсных геномов
4. Аннотация и анализ генов
5. Мутации, их ассоциации с болезнями
6. Экспрессия РНК и медицинские приложения
7. Предсказание взаимодействий белков, фолдинг, drug discovery
8. Задача о парном выравнивании. Глобальное и локальное выравнивание. Алгоритм Нилдмана-Вунша. Алгоритм Смита-Ватермана.
9. Основные базы данных
10. BLAST
11. Форматы fasta
12. Первичная обработка РНК-секвенирования
13. Матрица экспрессии генов
14. edgeR и DESeq2
15. Поиск дифференциально экспрессированных генов в двух группах
16. Статистическая обработка данных. Множественное тестирование и методы коррекции p-value
17. Графовое представление взаимодействий

18. Корреляционный анализ
19. Визуализация: PCA, t-SNE
20. Иерархическая кластеризация. Основные проблемы классификации: проклятие размерности, overfitting.
21. Простейшие идеи классификации: kNN, SVM

6.4.2 Для оценки сформированности умений компетенции ПК-1 и ПК-2 используются:

Практические задания (ППЗ)

Простые ПКЗ предполагают решение одной или нескольких задач в одно или два действия. К ним можно отнести: простые ситуационные задачи с коротким ответом или простым действием; несложные задания по выполнению конкретных действий. Простые ПКЗ содержат четкую инструкцию по выполнению или алгоритм действий.

Примеры практических заданий для оценки умений ПК1-1 и ПК-2:

Задание 1. Вручную сделать глобальное выравнивание нуклеотидных последовательностей ATGAGTCTCT и CTGTCTCCTG, используя матрицу штрафов DNAfull и линейный штраф за гэп равный 10. Проверить результат с помощью EMBOSS Needle.

Задание 2. Построить глобальное выравнивание нуклеотидных последовательностей человеческого и мышиного гена, кодирующего бета-актин. Построить выравнивание аминокислотных последовательностей соответствующих белков. Объяснить получившуюся разницу.

Задание 3. Написать функцию, принимающую на вход нуклеотидную последовательность и возвращающую ее обратно-комплементарную.

Задание 4. Используя Python, построить хеш-таблицу (dict) для позиций k -мер'ов в геноме SARS-CoV-2 (k –переменная).

Задание 5. Посчитать экзонные длины генов GNG4, SPRR4.

Задание 6. Провести вычисления метода median-of-ratios для матрицы TCGA (без использования DESeq2), убедиться, что результат совпал с DESeq2.

Задание 7. Посчитать нормировочные множители для матрицы TCGA с помощью библиотеки edgeR; выяснить, сильно ли отличаются коэффициенты от DESeq2.

Задание 8. Вывести топ-50 наиболее дифференциально экспрессированных генов по t -критерию Стьюдента (после нормализации FPKM на sizeFactors и логарифмирования).

Задание 9. Главный комплекс гистосовместимости человека кодируется генами HLA-A, HLA-B и HLA-C (класс I) и HLA-DRB1, HLA-DQB1 и HLA-DPB1 (класс II). С использованием корреляционного анализа убедиться, что фактор транскрипции СИТА активирует данные гены.

Задание 10. Даны гены, значимо изменившие свою экспрессию в клеточной линии HT-29 при некотором воздействии. Проанализировав функциональную принадлежность данных генов, предположить, что это было за воздействие.

6.5 Методические материалы, определяющие процедуры оценивания.

Положение «О проведении текущего контроля успеваемости и промежуточной аттестации обучающихся в ННГУ», утвержденное приказом ректора ННГУ от 13.02.2014 г. №55-ОД.

Положение о фонде оценочных средств, утвержденное приказом ректора ННГУ от 10.06.2015 №247-ОД.

7. Учебно-методическое и информационное обеспечение дисциплины (модуля)

а) основная литература:

1. An Introduction to Bioinformatics Algorithms. Neil C. Jones, Pavel A. Pevzner. MIT Press. 2004.

2. Computational Molecular Biology: An Algorithmic Approach. Pavel A. Pevzner. MIT Press. 2000.
3. Алексеев В.Е., Таланов В.А. Графы. Модели вычислений. Структуры данных: Учебник. – Нижний Новгород: Изд-во ННГУ, 2005. 307 с.
4. С., Основы биоинформатики. - М.-Ижевск: НИЦ "Регуляторная и хаотическая динамика", Институт компьютерных исследований, 2007. - 320 стр.

б) дополнительная литература

1. Johnes N., Pevzner P., An introduction to bioinformatics algorithms. – MIT Press, 2004. – 454 p.
2. Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology. Daniel M. Gusfield. Cambridge University Press, 1997.
3. Томас Х. Кормен, Чарльз И. Лейзерсон, Рональд Л. Ривест, Клиффорд Штайн. Алгоритмы: построение и анализ, 3-е издание // Introduction to Algorithms, Third Edition. — М.: «Вильямс», 2013. — 1328 с.

в) Интернет ресурсы

1. <http://bioinformatics.ru/>
2. <https://www.lektorium.tv/course/22933>

г) программное обеспечение

Для подбора источников информации используются электронные ресурсы ННГУ (электронный каталог, электронные библиотеки). При выполнении практических заданий используются программы Microsoft Windows Professional 7 Russian, Microsoft Office Word, RStudio, Jupyter Notebook, языки программирования R и Python.

8. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения занятий лекционного и практического типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации, укомплектованные учебной мебелью, а также помещения для хранения и профилактического обслуживания учебного оборудования. Для проведения занятий имеются демонстрационное оборудование (доска, переносное мультимедийное оборудование (проектор, ноутбук, экран)).

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду организации.

Программа составлена в соответствии с требованиями ФГОС ВО с учетом рекомендаций ОПОП ВО по направлению 06.04.01. Биология, магистерская программа «Биоинформатика».

Автор(ы) _____ м.н.с., Нерсиян С.А.

Рецензент _____ д.б.н., доц Ведунова М.В.

Директор ИББМ, зав. кафедрой

общей и медицинской генетики _____ д.б.н., доц. Ведунова М.В.

Программа одобрена на заседании методической комиссии ИББМ от 24 февраля 2021 г., протокол № 4. _____