

Санкт-Петербургский государственный университет

**РАБОЧАЯ ПРОГРАММА
УЧЕБНОЙ ДИСЦИПЛИНЫ**

**Многомерный анализ биологических данных
Multiscale Analysis of Biological Data**

Язык(и) обучения:

русский

Трудоемкость (границы трудоемкости) в зачетных единицах: 3

Регистрационный номер рабочей программы:

Санкт-Петербург

2018

Раздел 1. Характеристики учебных занятий

1.1. Цели и задачи учебных занятий

Курс охватывает круг вопросов, связанных с классификацией, обработкой и анализом больших массивов экспериментальных данных в биологии. Задачи курса - дать студентам навыки биологической интерпретации многомерных данных, их статистического анализа, познакомить с классическими методами анализа больших данных в молекулярной биологии и биоинформатике, а также основными программными реализациями.

1.2. Требования к подготовленности обучающегося к освоению содержания учебных занятий (пререквизиты)

Курсы биологической статистики и молекулярной биологии

1.3. Перечень результатов обучения (learning outcomes)

ОКМ-1, ОКМ-2, ОКМ-3, ОКМ-4, ОКМ-5, ПКМ-3, ПКМ-6, ПКМ-8

В результате освоения дисциплины обучающиеся получают представления о типах данных и подходах к их анализу, о базовых алгоритмах понижения размерности и кластеризации, и их применении в анализе омиксных проектов. Также студенты получают практические навыки работы с реальными данными большой размерности, полученными в последние годы в больших международных проектах (таких как «1000 genomes»).

1.4. Перечень активных и интерактивных форм учебных занятий

1. Лекции
2. Интерактивные формы – практические занятия с реальными данными секвенирования ДНК и РНК.
3. Самостоятельная работа — освоение операционной системы LINUX и базовых статистических концепций с использованием онлайн программ «Введение в LINUX» и «Основы статистики», самостоятельный анализ данных реальных проектов.

Раздел 2. Организация, структура и содержание учебных занятий

2.1. Организация учебных занятий

Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся															
Период обучения (модуль)	Контактная работа обучающихся с преподавателем										Самостоятельная			Объ	Тр
	Лекции	Семинары	Консультации	Практические занятия	Лабораторные работы	Контрольные работы	Коллоквиумы	Текущий контроль	Промежуточная аттестация	Итоговая аттестация	Под руководством преподавателя	В присутствии преподавателя	Сам. Раб. с использованием		
ОСНОВНАЯ ТРАЕКТОРИЯ															
очная форма обучения															

	4	12	2	6					2			34	36		12		24	3
	50-120	50-120	50-120	10-24					50-120			10-24	10-24		50-120			
ИТОГО:	4	12	2	6					2			34	36		12		24	3

Формы текущего контроля успеваемости, виды промежуточной и итоговой аттестации			
Период обучения (модуль)	Формы текущего контроля успеваемости	Виды промежуточной аттестации	Виды итоговой аттестации (только для программ итоговой аттестации и дополнительных образовательных программ)
ОСНОВНАЯ ТРАЕКТОРИЯ			
очная форма обучения			
Семестр 1		зачет	

2.2. Структура и содержание учебных занятий

Основной курс Основная траектория Очная форма обучения

Период обучения (модуль): С01

№ п/п	Наименование тем (раздела, части)	Вид учебных занятий	Количество часов
1	Введение	Лекции	2
		Семинары	0
		Сам. раб. с использованием методических материалов	0
2	Дисперсионный анализ	Лекции	0
		Семинары	2
		Сам. раб. с использованием методических материалов	1
3	Регрессионный анализ	Лекции	0
		Семинары	2
		Сам. раб. с использованием методических материалов	1
4	Многомерная регрессия	Лекции	0
		Семинары	2
		Сам. раб. с использованием методических материалов	1
5	Кластерный анализ	Лекции	0
		Семинары	2
		Сам. раб. с использованием методических материалов	1
6	Метод главных компонент	Лекции	0
		Семинары	2
		Сам. раб. с использованием методических материалов	1
7	Подходы к анализу реальных	Лекции	2
		Семинары	2

	многомерных данных	Консультации	2
		Сам. раб. с использованием методических материалов	1

1. Введение.

Основные подходы к анализу многомерных данных в биологии, типы и форматы данных. Основные базы данных и основы работы с ними.

2. Дисперсионный анализ

Однофакторный дисперсионный анализ, множественные сравнения в ANOVA, многофакторный ANOVA.

3. Регрессионный анализ

Корреляция и регрессия, линейная регрессия с одним предиктором.

4. Многомерная регрессия

Регрессионный анализ с несколькими переменными. Логистическая регрессия.

5. Кластерный анализ.

Метод к-средних. Иерархическая кластеризация.

6. Метод главных компонент.

Методы понижения размерности. Интерпретация, использование PCA в приложении к реальным данным..

7. Подходы к анализу реальных многомерных данных.

Анализ наборов данных из крупных международных баз данных и биоинформатических проектов (JGI IMG/M, «100 Genomes»).

Раздел 3. Обеспечение учебных занятий

3.1. Методическое обеспечение

3.1.1 Методические указания по освоению дисциплины

Для освоения дисциплины необходимо отработать всю программу практических занятий, выполнить проектное задание и сдать зачет. Выполнение каждого практического занятия фиксируется преподавателем подписью в альбоме обучающегося при проверке правильности выполнения программы практического занятия.

3.1.2 Методическое обеспечение самостоятельной работы

Самостоятельная работа обеспечивается онлайн курсами «Введение в LINUX» (авторы курса А. Гуревич и А. Пржибельский) и «Основы статистики» (автор курса А. Карпов)

3.1.3 Методика проведения текущего контроля успеваемости и промежуточной аттестации и критерии оценивания

Аттестация по дисциплине фиксируется выставлением положительной оценки в рамках промежуточной аттестации, проводимой в форме зачета по окончании 1-го семестра.

Обучающиеся должны представить лектору сертификаты платформы Stepic, свидетельствующий об успешном освоении онлайн курсов «Введение в Linux» и «Основы статистики».

Сдача зачета осуществляется в устной форме и включает в себя краткую презентацию результатов выполнения проекта.

Зачет не может быть поставлен, если обучающийся не освоил базовые основы LINUX и не ознакомился с основными концепциями анализа данных; не выполнил и/или не сдал проектное задание.

Во всех остальных случаях ставится зачет.

3.1.4 Методические материалы для проведения текущего контроля успеваемости и промежуточной аттестации (контрольно-измерительные материалы, оценочные средства)

Примеры заданий для проведения промежуточной аттестации:

1. Кластеризация данных экспрессии генов *S.cerevisiae*
2. Анализ методом главных компонент результатов проекта «1000 Genomes».

3.1.5 Методические материалы для оценки обучающимися содержания и качества учебного процесса

Используются контрольно-измерительные материалы (анкеты), разработанные на факультете для оценки содержания и качества учебного процесса.

3.2. Кадровое обеспечение

3.2.1 Образование и (или) квалификация преподавателей и иных лиц, допущенных к проведению учебных занятий

Для проведения лекционной части курса и консультаций необходимо высшее образование по специальности или опыт работы по направлению «Биоинформатика», наличие ученой степени кандидата или доктора биологических наук и опыта чтения дисциплин в области биоинформатики.

Для проведения практических занятий и приема зачета необходимо высшее образование по специальности или опыт работы по направлению «Биоинформатика», опыт проведения практических занятий по Биоинформатике. Требования к степени и званию не предъявляются.

3.2.2 Обеспечение учебно-вспомогательным и (или) иным персоналом

Для обеспечения реализации курса необходим 1 ИТ специалист.

Помимо этого, необходимо обеспечить по 1 преподавателю на группу практических занятий из 24 человек. Необходимо обеспечить ИТ сопровождение практических занятий, включающее:

1. подготовку занятий, в том числе установку необходимого пакета программ на каждый персональный компьютер в аудитории (список программ приводится ниже);
2. обеспечить бесперебойный доступ к скоростному интернету, ИТ-специалист должен быть доступным в случае возникновения нештатных ситуаций.

3.3. Материально-техническое обеспечение

3.3.1 Характеристики аудиторий (помещений, мест) для проведения занятий

Для проведения лекционных занятий необходима стандартно оборудованная лекционная аудитория с мультимедийным оборудованием.

Аудитория для практических занятий (компьютерный класс) должна содержать мультимедийное оборудование, 12 персональных компьютеров и 24 стула для обучающихся (2 обучающихся на один компьютер), а также стол, стул и компьютер для преподавателя.

3.3.2 Характеристики аудиторного оборудования, в том числе неспециализированного компьютерного оборудования и программного обеспечения общего пользования

12 компьютеров с установленной операционной системой LINUX, Microsoft Power Point для показа слайдов и экран для их демонстрации во время практических занятий.

3.3.3 Характеристики специализированного оборудования

Не требуется

3.3.4 Характеристики специализированного программного обеспечения

Операционная система LINUX

3.3.5 Перечень и объемы требуемых расходных материалов

Мел белый, мел цветной (на каждое лекционное и практические занятия)

3.4. Информационное обеспечение

3.4.1 Список обязательной литературы

См. Раздел 3.4.3.

3.4.2 Список дополнительной литературы

Не требуется

3.4.3 Перечень иных информационных источников

1. Основы статистики <https://stepik.org/course/76/syllabus>
2. Введение в Linux <https://stepik.org/course/Введение-в-Linux-73/syllabus>

Раздел 4. Разработчики программы

Райко Михаил Петрович, с.н.с.

mike.rayko@gmail.com

+79213302127

Санкт-Петербургский государственный университет

**РАБОЧАЯ ПРОГРАММА
УЧЕБНОЙ ДИСЦИПЛИНЫ**

Биологические базы данных: назначение, создание, использование

Biological Databases: Application, Construction, Specificity

Язык(и) обучения:

русский

Трудоемкость (границы трудоемкости) в зачетных единицах: 2

Регистрационный номер рабочей программы: 057264

Санкт-Петербург

2018

Раздел 1. Характеристики учебных занятий

1.5. Цели и задачи учебных занятий

Цель курса - знакомство обучающихся с основными базами данных, используемыми в биоинформатике (NCBI Entrez, ENA и т. д.), обучение составлению запросов и анализу больших массивов, работе с API наиболее важных баз и созданию собственных баз данных.

1.6. Требования к подготовленности обучающегося к освоению содержания учебных занятий (пререквизиты)

Курсы молекулярной биологии и программирования на Python.

1.7. Перечень результатов обучения (learning outcomes)

ОКМ-1, ОКМ-2, ОКМ-3, ОКМ-4, ОКМ-5, ПКМ-3, ПКМ-5, ПКМ-8

В результате освоения дисциплины обучающиеся получают представления об организации больших массивов данных и доступе к ним. Они узнают о различных типах данных и способах их представления, и о том, как автоматически обрабатывать большие запросы, получать необходимые данные для сравнения полученных результатов с уже известными.

1.8. Перечень активных и интерактивных форм учебных занятий

Активные формы – самостоятельная работа по освоению REST API систем NCBI Entrez, ENA и UniProt.

Интерактивные формы – практические занятия по поиску реальных данных в базах, составлению автоматических запросов, обработке полученных данных и проверке гипотез.

Самостоятельная работа в присутствии преподавателя с использованием методических материалов над реальными задачами с последующей презентацией результатов.

Раздел 2. Организация, структура и содержание учебных занятий

2.1. Организация учебных занятий

Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся																	
Период обучения (модуль)	Контактная работа обучающихся с преподавателем											Самостоятельная			Объём	Итого	
	Лекции	Семинары	Консультации	Практические занятия	Лабораторные работы	Контрольные работы	Коллоквиумы	Текущий контроль	Промежуточная аттестация	Итоговая аттестация	Под руководством преподавателя	В присутствии преподавателя	Сам. Раб. с использованием	Текущий контроль (сам. раб.)			Промежуточная аттестация (сам. раб.)
ОСНОВНАЯ ТРАЕКТОРИЯ																	
очная форма обучения																	
	4	16	2	4				2			4	28		12		34	2
	50-120	10-24	50-120					50-120				10-24		50-120			

ИТОГО:	4	16	2	4					2			4	28		12		34	2
--------	---	----	---	---	--	--	--	--	---	--	--	---	----	--	----	--	----	---

Формы текущего контроля успеваемости, виды промежуточной и итоговой аттестации			
Период обучения (модуль)	Формы текущего контроля успеваемости	Виды промежуточной аттестации	Виды итоговой аттестации (только для программ итоговой аттестации и дополнительных образовательных программ)
ОСНОВНАЯ ТРАЕКТОРИЯ			
очная форма обучения			
Семестр 1		зачет	

2.2. Структура и содержание учебных занятий

Основной курс Основная траектория Очная форма обучения

Период обучения (модуль): С01

№ п/п	Наименование тем (раздела, части)	Вид учебных занятий	Количество часов
1	Введение	Лекции	2
		Практические занятия	0
		Сам. раб. с использованием методических материалов	4
2	Нуклеотидные базы данных. GenBank, ENA, DDBJ	Лекции	2
		Практические занятия	2
		Сам. раб. с использованием методических материалов	4
3	Белковые базы данных. NCBI Protein, UniProt, ExPASy	Лекции	0
		Практические занятия	2
		Сам. раб. с использованием методических материалов	4
4	Анализ эволюции, базы ортологов. EggNOG, Pfam, TreeFam.	Лекции	0
		Практические занятия	4
		Сам. раб. с использованием методических материалов	4
5	Метаинформация, метаболические пути. Microbes Online, MetaCyc, KEGG	Лекции	0
		Практические занятия	4
		Сам. раб. с использованием методических материалов	4
6	Entrez E-utils, REST API	Лекции	0
		Практические занятия	4
		Сам. раб. с использованием методических материалов	4
7	Самостоятельное создание баз данных.	Лекции	0
		Практические занятия	4
		Сам. раб. с использованием методических материалов	4

8. Введение.

Роль баз данных в современных биомедицинских исследованиях. Основные принципы, задачи и подходы к организации и хранению геномных, протеомных и других данных, получаемых современными методами молекулярной биологии.

9. Нуклеотидные базы данных. GenBank, ENA, DDBJ

Базы последовательностей ДНК — результатов секвенирования, собранных геномов, проаннотированных генов. Форматы данных, метаданные, способы представления и запросы.

10. Белковые базы данных. NCBI Protein, UniProt, ExPASy.

Базы последовательностей белков. Способы представления аннотации и доменной организации. Связь с базами трехмерной структуры белков.

11. Анализ эволюции, базы ортологов. EggNOG, Pfam, TreeFam.

Биоинформатические подходы к анализу эволюции доменной структуры белков. Предсказание доменов в различных белках, построение НММ-моделей. Запросы через командную строку и поиск ортологических наборов.

12. Метаинформация, метаболические пути. Microbes Online, MetaCyc, KEGG.

Поиск метаданных, фильтрация по различным параметрам. Связь различных баз между собой, утилита eLink. Анализ метаболических путей в различных геномах.

13. Entrez E-utils, REST API.

Запрос к различным базам данных через командную строку, написание скриптов для автоматизации запросов и последующего анализа. Особенности доступа к различным базам.

14. Самостоятельное создание баз данных.

Основы SQL, создание небольших баз на основе собственных результатов, публикация данных.

Раздел 3. Обеспечение учебных занятий

3.1. Методическое обеспечение

3.1.1 Методические указания по освоению дисциплины

Для освоения дисциплины необходимо отработать всю программу практических занятий, выполнить проектное задание и сдать зачет. Выполнение каждого практического занятия фиксируется преподавателем подписью в альбоме обучающегося при проверке правильности выполнения программы практического занятия.

3.1.2 Методическое обеспечение самостоятельной работы

Самостоятельная работа обеспечивается онлайн курсом СПбГУ «Базы данных» на открытой платформе Stepic и руководством NCBI Entrez Programming Utilities Help.

3.1.3 Методика проведения текущего контроля успеваемости и промежуточной аттестации и критерии оценивания

Аттестация по дисциплине фиксируется выставлением положительной оценки в рамках промежуточной аттестации, проводимой в форме зачета по окончании курса.

Обучающиеся должны представить лектору сертификат платформы Stepic, свидетельствующий об успешном освоении онлайн курса «Базы данных».

Сдача зачета осуществляется в устной форме и включает в себя краткую презентацию результатов выполнения проекта.

Зачет не может быть поставлен, если обучающийся не освоил базовые основы работы с NCBI API, не способен найти и скачать требуемые данные из изучаемых баз, не выполнил и/или не сдал проектное задание.

Во всех остальных случаях ставится зачет.

3.1.4 Методические материалы для проведения текущего контроля успеваемости и промежуточной аттестации (контрольно-измерительные материалы, оценочные средства)

Примеры заданий для проведения промежуточной аттестации:

3. Анализ таксономической представленности известных метагеномов из горячих источников (по данным NCBI Assembly, MG-RAST).
4. Эволюция опсинов у позвоночных (по данным UniProt и EggNOG).

3.1.5 Методические материалы для оценки обучающимися содержания и качества учебного процесса

Используются контрольно-измерительные материалы (анкеты), разработанные на факультете для оценки содержания и качества учебного процесса.

3.2. Кадровое обеспечение

3.2.1 Образование и (или) квалификация преподавателей и иных лиц, допущенных к проведению учебных занятий

Для проведения лекционной части курса и консультаций необходимо высшее образование по специальности или опыт работы по направлению «Биоинформатика», наличие ученой степени кандидата или доктора биологических наук и опыта чтения дисциплин в области биоинформатики.

Для проведения практических занятий и приема зачета необходимо высшее образование по специальности или опыт работы по направлению «Биоинформатика», опыт проведения практических занятий по Биоинформатике. Требования к степени и званию не предъявляются.

3.2.2 Обеспечение учебно-вспомогательным и (или) иным персоналом

Для обеспечения реализации курса необходим 1 ИТ специалист.

Помимо этого, необходимо обеспечить по 1 преподавателю на группу практических занятий из 24 человек. Необходимо обеспечить ИТ сопровождение практических занятий, включающее:

3. подготовку занятий, в том числе установку необходимого пакета программ на каждый персональный компьютер в аудитории (список программ приводится ниже);
4. обеспечить бесперебойный доступ к скоростному интернету, ИТ-специалист должен быть доступным в случае возникновения нештатных ситуаций.

3.3. Материально-техническое обеспечение

3.3.1 Характеристики аудиторий (помещений, мест) для проведения занятий

Для проведения лекционных занятий необходима стандартно оборудованная лекционная аудитория с мультимедийным оборудованием.

Аудитория для практических занятий (компьютерный класс) должна содержать мультимедийное оборудование, 12 персональных компьютеров и 24 стула для обучающихся (2 обучающихся на один компьютер), а также стол, стул и компьютер для преподавателя.

3.3.2 Характеристики аудиторного оборудования, в том числе неспециализированного компьютерного оборудования и программного обеспечения общего пользования

12 компьютеров с установленной операционной системой LINUX, Microsoft Power Point для показа слайдов и экран для их демонстрации во время практических занятий.

3.3.3 Характеристики специализированного оборудования

Не требуется

3.3.4 Характеристики специализированного программного обеспечения

Операционная система LINUX

3.3.5 Перечень и объемы требуемых расходных материалов

Мел белый, мел цветной (на каждое лекционное и практические занятия)

3.4. Информационное обеспечение

3.4.1 Список обязательной литературы

См. Раздел 3.4.3.

3.4.2 Список дополнительной литературы

Не требуется

3.4.3 Перечень иных информационных источников

3. Онлайн курс СПбГУ «Базы данных» <https://stepik.org/course/2614/>
4. Entrez Programming Utilities Help (<https://www.ncbi.nlm.nih.gov/books/NBK25501/>)

Раздел 4. Разработчики программы

Райко Михаил Петрович, с.н.с.

mike.rayko@gmail.com

+79213302127