

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**Федеральное государственное бюджетное учреждение науки
Институт общей генетики им. Н.И. Вавилова
Российской академии наук
(ИОГен РАН)**

«ПРИНЯТО»

«УТВЕРЖДАЮ»

На заседании Ученого совета ИОГен РАН

Директор ИОГен РАН

Протокол № 1 от «19» февраля 2019 г. д.б.н.  А.М. Кудрявцев



**РАБОЧАЯ ПРОГРАММА
ПО ДИСЦИПЛИНЕ**

Б1.В.ВД2

«Анализ данных высокопроизводительного секвенирования»

Направление подготовки: 06.06.01 Биологические науки

**Уровень образования: высшее образование – подготовка кадров высшей
квалификации**

Квалификация выпускника: Исследователь. Преподаватель-исследователь.

Разработчик:

к.ф-м.н. А.С. Касьянов

Москва, 2019 г.

Рабочая программа составлена на основании федерального государственного образовательного стандарта, разработанного для реализации основных профессиональных образовательных программ высшего образования - программ подготовки научно-педагогических кадров в аспирантуре по направлению подготовки кадров высшей квалификации 06.06.01 «Биологические науки».

Согласно Федеральному государственному образовательному стандарту высшего образования по направлению подготовки 06.06.01 Биологические науки (уровень подготовки кадров высшей квалификации), утвержденному приказом Минобрнауки РФ № 871 от 30 июля 2014 г., и учебному плану аспирантов, разработанного на основе этого стандарта, дисциплина «Анализ данных высокопроизводительного секвенирования» является второй обязательной учебной дисциплиной модуля вариативной части Блока 1 образовательной программы по направленности (профилю) 03.02.07 Генетика.

Объём курса составляет 3 зачетные единицы или 108 академических часа, из них 20 академических часов лекций, 32 академических часов семинаров и 56 академических часов самостоятельной внеаудиторной работы аспирантов, включая подготовку к зачету.

1. Цели и задачи

Цель дисциплины:

Дисциплина «Анализ данных высокопроизводительного секвенирования» представляет обучающимся возможность освоить основные методы, применяющиеся при обработке данных высокопроизводительного секвенирования. Аспиранты получают широкий обзор основных видов данных, генерируемых платформами высокопроизводительного секвенирования, с упором на данные полногеномного и полнотранскриптомного секвенирования.

Цель изучения дисциплины: знакомство аспирантов с известными на данный момент способами обработки данных, получаемых в результате высокопроизводительного секвенирования.

Задачи дисциплины:

- формирование базовых знаний об особенностях данных, получаемых с помощью платформ высокопроизводительного секвенирования;
- практическое освоение аспирантами методов для анализа биологических данных, полученных с помощью высокопроизводительного секвенирования;
- формирование у аспирантов основных навыков разработки методов для анализа данных и приобретение ими практического опыта, необходимого для проведения самостоятельных научных исследований в области вычислительной обработки биологических данных, полученных с помощью технологий высокопроизводительного секвенирования.

2. Место дисциплины (модуля) в структуре образовательной программы аспирантуры

Дисциплина «Анализ данных высокопроизводительного секвенирования» включает в себя разделы, которые могут быть отнесены к вариативной части цикла и базируется на дисциплинах: Генетика, Биостатистика и Прикладная биоинформатика.

3. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Универсальные компетенции

- способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении

исследовательских и практических задач, в том числе в междисциплинарных областях (УК-1);

- способность проектировать и осуществлять комплексные исследования, в том числе междисциплинарные, на основе целостного системного научного мировоззрения с использованием знаний в области истории и философии науки (УК-2);

- готовность участвовать в работе российских и международных исследовательских коллективов по решению научных и научно-образовательных задач (УК-3);

- способность планировать и решать задачи собственного профессионального и личностного развития (УК-5).

Общепрофессиональные компетенции

- способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий (ОПК-1);

- готовность к преподавательской деятельности по основным образовательным программам высшего образования (ОПК-2).

Профессиональные компетенции

- способность к самостоятельному проведению научно-исследовательской работы и получению научных результатов, удовлетворяющих установленным требованиям к содержанию диссертаций на соискание ученой степени кандидата наук по направленной специальности (ПК-1);

- обладание представлениями о системе фундаментальных понятий и методологических аспектов биологии, форм и методов научного познания (ПК-2);

- способность приобретать новые знания с использованием современных научных методов и владение ими на уровне, необходимом для решения задач, возникающих при выполнении профессиональных функций (ПК-3).

В результате освоения дисциплины, обучающиеся должны

знать:

- основные физические принципы, лежащие в основе технологий высокопроизводительного секвенирования;

- основные алгоритмы и структуры данных, применяемые при сборке de novo геномов и транскриптомов, структурной аннотации геномных последовательностей, картировании чтений;

- статистические методы, применяющиеся при анализе данных, полученных с помощью высокопроизводительного секвенирования;

- вычислительные задачи, возникающие при обработке данных, полученных с использованием высокопроизводительного секвенирования;

уметь:

- применять основные программные средства, предназначенные для обработки данных, полученных с использованием высокопроизводительного секвенирования;

- применять основные алгоритмические идеи для разработки новых методов и алгоритмов для обработки данных, полученных с использованием высокопроизводительного секвенирования;

владеть:

- навыками освоения большого объема информации;
- культурой постановки и моделирования вычислительных задач обработки биологических данных, полученных с использованием технологий высокопроизводительного секвенирования.

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

Всего часов	Аудиторные занятия (час):	Практические занятия (Семинары) (час)	Самостоятельная работа (час)	Подготовка к зачету
108	20	32	54	2

№ п/п	Тема (раздел) дисциплины	Виды учебных занятий, включая самостоятельную работу				
		Лекции	Практические занятия Семинары	Практики	Научные исслед.	Самост. работа
1	Технологии высокопроизводительного секвенирования	2	2			6
2	Основы работы с командной строкой Linux	3	6			6
3	Предобработка результатов секвенирования	2	4			6
4	De novo сборка геномов и транскриптомв	3	4			6
5	Аннотация геномных последовательностей	2	4			6
6	Ресеквенирование	2	2			6
7	RNA-seq	2	4			6
8	Метагеномика	2	4			6
9	CHiP-seq	2	2			6
Итого часов		20	32			54
Подготовка к экзамену/зачету		2 часа				
Общая трудоёмкость		108 час., 3 зач.ед.				

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

Тема 1. Технологии высокопроизводительного секвенирования

Физические принципы и технологические решения, используемые в технологиях высокопроизводительного секвенирования. Характеристики основных платформ высокопроизводительного секвенирования.

Тема 2. Основы работы с командной строкой Linux

Командная оболочка Bash. Устройство файловой системы в операционных системах семейства Linux. Команды cd, ls, pwd, cp, mv, rm, more, head, tail, grep. Редактор vi.

Тема 3. Предобработка результатов секвенирования

Основные типы ошибок, свойственные технологиям высокопроизводительного секвенирования. Основные форматы данных. Оценка качества чтений. Тримминг.

Тема 4. *de novo* сборка геномов и транскриптомов

Алгоритмы *de novo* сборки, основанные на графа де Брейна и графах перекрытий. Особенности геномных последовательностей, затрудняющих сборку. Оценка качества сборки. Практические аспекты больших геномных проектов. Особенности сборки транскриптомов *de novo*.

Тема 5. Аннотация геномных последовательностей

Основные принципы построения алгоритмов аннотации. Оценка качества аннотации. Практические аспекты применения алгоритмов аннотации для эукариотических геномов.

Тема 6. Ресеквенирование

Картирование чтений на референсный геном. Преобразование Барроуза-Уилера для картирования ридов при секвенировании ДНК. Оценка качества картирования. SNP calling. Особенности, возникающие при детекции соматических мутаций.

Тема 7. RNA-seq

Особенности картирования чтений, полученных в результате RNA-seq эксперимента на референсный геном. Методы нормализации и анализ экспрессии генов.

Тема 8. Метагеномика

Таргетное секвенирование 16S рРНК. Таксономический анализ и анализ биоразнообразия. Полнометагеномное секвенирование. De novo сборка и аннотация генов.

Тема 9. ChIP-seq

Взаимодействие ДНК и белка. Методы для изучения ДНК-белкового взаимодействия, применяющиеся до появления высокпроизводительного секвенирования. ChIP – seq протокол. Основные методы анализа ChIP-seq данных.

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

Оборудование, необходимое для лекций и семинаров: компьютер, Windows, MS Office и мультимедийное оборудование (проектор, звуковая система),

6. Перечень основной и дополнительной литературы, необходимой для освоения дисциплины (модуля)

Основная литература:

1. Phillip Compeau, Pavel Pevzner, Bioinformatics Algorithms: An Active Learning Approach 2014 Book
2. Xinkun Wang Next-Generation Sequencing Data Analysis 2016 Book
3. Ion Mandoiu, Alexander Zelikovsky. Computational Methods for Next Generation Sequencing Data Analysis 2016 Book
4. Eija Korpelainen, Jarno Tuimala, Panu Somervuo , Mikael Huss, Garry Wong RNA-seq Data Analysis: A Practical Approach. 2014 Book.

Дополнительная литература: -

7. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

-

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

Научно-библиографические и патентные базы данных в области биологии, доступные по сети Интернет в бесплатном режиме - Science Citation Index (Web of Science), Medline (PubMed), Научная электронная библиотека (НЭБ), Российская патентная БД ФГУ ФИПС и американская патентная БД USPAFULL; электронные адреса крупных научных издательств, предоставляющих доступ к полным текстам текущих и архивным выпускам этих журналов.

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

Доступ в Интернет. Дополнительно программное обеспечение не требуется.

10. Методические указания для обучающихся по освоению дисциплины

Аспирант, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике.

В результате изучения дисциплины аспирант должен знать основные определения дисциплины, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

– посещения всех занятий, предусмотренных учебным планом по дисциплине;

– ведения конспекта занятий;

– напряжённой самостоятельной работы аспиранта.

Самостоятельная работа включает в себя:

– чтение рекомендованной литературы;

– проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;

– решение задач, предлагаемых аспирантам на занятиях;

– подготовку к выполнению заданий текущей и промежуточной аттестации.

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

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

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

11. Фонд оценочных средств для проведения промежуточной аттестации по итогам обучения

Приложение А

12. Составители программы:

к.ф-м.н. А.С. Касьянов

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ
ПО ДИСЦИПЛИНЕ
«Анализ данных высокопроизводительного секвенирования»**

1. Компетенции, формируемые в процессе изучения дисциплины

Освоение дисциплины направлено на формирование у обучающегося следующих универсальных (УК), общепрофессиональных (ОПК) и профессиональных (ПК) компетенций:

Универсальные компетенции

- способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях (УК-1);

- способность проектировать и осуществлять комплексные исследования, в том числе междисциплинарные, на основе целостного системного научного мировоззрения с использованием знаний в области истории и философии науки (УК-2);

- готовность участвовать в работе российских и международных исследовательских коллективов по решению научных и научно-образовательных задач (УК-3);

- способность планировать и решать задачи собственного профессионального и личностного развития (УК-5).

Общепрофессиональные компетенции

- способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий (ОПК-1);

- готовность к преподавательской деятельности по основным образовательным программам высшего образования (ОПК-2).

Профессиональные компетенции

- способность к самостоятельному проведению научно-

исследовательской работы и получению научных результатов, удовлетворяющих установленным требованиям к содержанию диссертаций на соискание ученой степени кандидата наук по направленной специальности (ПК-1);

- обладание представлениями о системе фундаментальных понятий и методологических аспектов биологии, форм и методов научного познания (ПК-2);

- способность приобретать новые знания с использованием современных научных методов и владение ими на уровне, необходимом для решения задач, возникающих при выполнении профессиональных функций (ПК-3).

2. Показатели оценивания компетенций

В результате изучения дисциплины «Анализ данных высокопроизводительного секвенирования» обучающийся должен:

знать:

- основные физические принципы, лежащие в основе технологий высокопроизводительного секвенирования;

- основные алгоритмы и структуры данных, применяемые при сборке *de novo* геномов и транскриптомов, структурной аннотации геномных последовательностей, картировании чтений;

- статистические методы, применяющиеся при анализе данных, полученных с помощью высокопроизводительного секвенирования;

- вычислительные задачи, возникающие при обработке данных, полученных с использованием высокопроизводительного секвенирования;

уметь:

- применять основные программные средства, предназначенные для обработки данных, полученных с использованием высокопроизводительного секвенирования;

- применять основные алгоритмические идеи для разработки новых методов и алгоритмов для обработки данных, полученных с использованием высокопроизводительного секвенирования;

владеть:

- навыками освоения большого объема информации;
- культурой постановки и моделирования вычислительных задач обработки биологических данных, полученных с использованием технологий высокопроизводительного секвенирования.

3. Перечень типовых контрольных заданий, используемых для оценки знаний, умений, навыков

Промежуточная аттестация по дисциплине «Анализ данных высокопроизводительного секвенирования» осуществляется в форме зачета. Зачет проводится в устной форме.

1. Основные физические принципы, лежащие в основе технологий высокопроизводительного секвенирования
2. Поколения технологий секвенирования. Основные принципиальные отличия технологий секвенирования второго поколения от первого.
3. Основные ошибки в данных, возникающие при использовании различных платформ высокопроизводительного секвенирования
4. Алгоритмы сборки de novo геномных последовательностей.
5. Особенности геномных последовательностей, приводящие к трудностям при сборке de novo.
6. Оценка качества геномных сборок
7. Особенности сборки транскриптомов de novo.
8. Оценка качества транскриптомной сборки.
9. Основные методы, используемые при аннотации геномных последовательностей.
10. Оценка качества аннотации.
11. Картирование чтений на референсный геном. Преобразование Барроуза-Уилера.
12. SNP calling.
13. Особенности детекции соматических мутаций на основе данных высокопроизводительного секвенирования.
14. Дизайн RNA-seq эксперимента.
15. Основные способы нормализации экспрессионных данных.
16. Анализ диф. экспрессии.
17. Таргентное секвенирование 16s РНК в метагеномике.
18. Полнометагеномное секвенирование
19. Таксономический анализ и анализ биоразнообразия.
20. De novo сборка и аннотация данных, полученных в результате полнометагеномного секвенирования
21. Дизайн ChIP – seq эксперимента.
22. Основные элементы вычислительного конвейера, используемого для обработки данных, полученных в результате ChIP-seq эксперимента.

Тестирование:

1. Что такое N50
 - а) Средняя длина контигов
 - б) Медианная длина контигов
 - в) Взвешанная по длине медианная длина контигов

2. Какой минимум биологических повторностей необходим для RNA-Seq экспериментов
 - а) 2
 - б) 3
 - в) 5

3. Чем отличается секвенирование 3-го поколения от секвенирования 2-го поколения
 - а) Производительностью
 - б) Качеством прочтений
 - в) Отсутствием амплификации

4. Критерии оценивания

Оценка отлично (5 баллов) - выставляется аспиранту, 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 хорошо (4 балла) - выставляется аспиранту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка удовлетворительно (3 балла) - выставляется аспиранту, 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 неудовлетворительно (2 балла) - выставляется аспиранту, 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.

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

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