

**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.ОД3**

**«Прикладная биоинформатика»**

**Направление подготовки: 06.06.01 Биологические науки**

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

**Квалификация выпускника: Исследователь. Преподаватель-исследователь.**

**Разработчик:**

д.ф-м.н. В.Ю. Макеев

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 «Прикладная биоинформатика» является третьей обязательной учебной дисциплиной модуля вариативной части Блока 1 образовательной программы по направленности (профилю) 03.02.07 Генетика.

Объём курса составляет 4 зачетных единиц или 144 академических часа, из них 30 академических часов лекций, 42 практических занятий (семинары), 72 академических часов самостоятельной внеаудиторной работы аспирантов, включая подготовку к зачету.

## **1. Цели и задачи**

### **Цель дисциплины**

дать аспирантам наиболее важные представления об основных подходах биоинформатического анализа применяемых для изучения структуры и функционирования геномного локуса.

### **Задачи дисциплины**

- формирование базовых знаний о различных ресурсах и их организации содержащих информацию и программы для анализа биологических данных;
- практическое освоение аспирантами методов биоинформатического анализа на примере одного из участков генома человека;
- формирование у аспирантов основных навыков и приобретение ими практического опыта, необходимого для проведения самостоятельных научных исследований в области биоинформатического анализа.

## **2. Место дисциплины (модуля) в структуре образовательной программы**

Дисциплина «Прикладная биоинформатика» включает в себя разделы, которые могут быть отнесены к обязательной учебной дисциплине модуля вариативной части Блока 1 образовательной программы.

Дисциплина «Прикладная биоинформатика» базируется на дисциплинах:

Генетика;

Биостатистика;

Дополнительные главы биоинформатики;

Анализ данных высокопроизводительного секвенирования.

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

Освоение дисциплины направлено на формирование следующих общепрофессиональных, универсальных и профессиональных компетенций аспиранта:

### ***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);

- Готовность использовать современные методы и технологии научной коммуникации на государственном и иностранном языках (УК-4);

- 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 направленности 03.02.07 Генетика (ПК-1);

- Обладание представлениями о фундаментальных основах биологических процессов, 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);

- Обладание опытом профессионального участия в научных дискуссиях, умение представлять полученные в исследованиях результаты в виде отчетов и научных публикаций (стендовые доклады, рефераты и статьи в периодической научной печати) (ПК-4);

- Владение методами отбора материала, преподавания и основами управления процессом обучения фундаментальной биологии (ПК-5).

В результате освоения дисциплины, обучающиеся должны

**знать:**

- основные геномные элементы;
- базы данных содержащих информацию о геномных элементах;
- инструменты для анализа геномных элементов;
- программы для представления данных по геномным элементам;
- различные форматы данных;
- методы, с помощью которых получают биологические данные;
- методы исследования генома.

**уметь:**

- проводить биоинформатический анализ участка генома;
- вычленять информацию о функционировании участка генома;
- уметь собирать всю имеющуюся информацию из различных источников для анализа;
- проводить комплексный анализ различных видов данных;
- выстраивать гипотезы и планировать эксперименты для подтверждения наблюдаемых феноменов.

**владеть:**

- навыками освоения большого объёма информации;
- культурой постановки и моделирования биологических задач.

**4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий**

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Лекции	Практич. занятия (Семинары)	Самост. работа
1	Функциональные элементы генома	2	2	4
2	Основные ресурсы для анализа функциональных элементов генома	4	4	8
3	Геномные баузеры	4	6	8
4	Анализ структуры локуса генома человека	4	6	12
5	Экспрессионный анализ	2	6	8
6	Изучение регуляции локуса	2	4	4
7	Исследование кодирующего потенциала гена	2	2	4
8	Анализ функции белка	2	4	6
9	Сравнительно-геномный анализ локуса	2	2	4
10	Особенности анализа экспрессии генов	2	2	4
11	Анализ мутации в генах	2	2	4
12	Теория дизайна лабораторного эксперимента и представления данных	2	2	4
<b>Итого часов: 142 ак.ч.</b>		<b>30</b>	<b>42</b>	<b>70</b>
<b>Зачеты – 2 ак.ч.</b>				
<b>Общая трудоёмкость – 144 ак.ч.</b>				

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

### **Тема 1. Функциональные элементы генома**

Геном. Доля транскрибируемой ДНК. Мусорная ДНК. Информационная емкость. Реализация ДНК как генетического материала. Анализ первичной структуры ДНК и её функции. Проект ENCODE. Гены человека. Псевдогены, их классификация. Процессированные псевдогены. Механизмы функционального действия процессированных псевдогенов. Регуляторные участки в геноме: промотор, ТАТА-бокс, энхансер, сайленсер, инсулятор. Повторяющиеся последовательности в ДНК. Тандемные повторы: микросателлиты, минисателлиты и сателлиты. Болезни экспансии тринуклеотидных повторов. Диспергированные повторы: транспозоны и ретротранспозоны. Открытие мобильных элементов. Полиморфизм ДНК. Функционирование вторичной структуры ДНК. G-квадруплексы в промоторах и теломерах.

### **Тема 2. Основные ресурсы для анализа функциональных элементов генома**

Сервер NCBI. Основные базы данных расположенные на нём: GenBank, Genome, Gene, dbSNP, dbEST, UniGene, PubMed, OMIM, Protein, и другие. Запросы онлайн. Перекрёстный анализ. Скачивание данных для локального анализа.

### **Тема 3. Геномные баузеры**

Геномные баузеры: MapView, USCS, ensemble. Устройство и пользование баузером USCS. Создание индивидуальных сессий. Обмен информацией в сессиях. Сборки геномов. Трэки, управление и настройки трэков. Трэк гены. Сравнение различных аннотаций генов человека. Трэки мРНК и EST. Совместный анализ трэков гены, мРНК и EST.

### **Тема 4. Анализ структуры локуса генома человека.**

Кодирующие и некодирующие РНК. lncRNA, miRNA. Псевдогены. Повторяющиеся последовательности. Семейство Alu повторов.

### **Тема 5. Экспрессионный анализ**

Анализ данных микрочипов Affymetrix. Устройство пробсета. База данных BioGPS. Анализ данных RNAseq разных проектов. Количественный и качественный анализ. Выбор вида представления данных. Экспрессионный анализ по клеточным культурам, по локализации транскриптов, различным условиям обработки клеточных культур, фракции полиА<sup>+</sup> или полиА<sup>-</sup>. Сравнение структуры генов, РНК и EST. Данные Fantom5.

### **Тема 6. Изучение регуляции локуса**

Регуляция с помощью miRNA. CLASH data. Антисмысловые взаимодействия. Транскрипционные факторы. Гистоновые модификации. Метилирование. Чувствительность к ДНКазам.

#### **Тема 7. Исследование кодирующего потенциала гена**

Поиск открытой рамки считывания. Определение белковых изоформ продукта. Анализ доменной структуры белка. Онлайн инструменты: CDS, CDART. База данных Pfam.

#### **Тема 8. Анализ функции белка**

Выявление белковых партнёров. Пост-трансляционные модификации белка. Предсказание функции белка на основе его взаимодействий с белками-партнерами. Анализ данных базы «IntAct Molecular Interaction Database». Анализ клеточных путей. База данных neXtProt. Данные проекта «The human protein atlas». PeptideAtlas. Uniprot

#### **Тема 9. Сравнительно-геномный анализ локуса**

Консервативность гена, эволюция локуса. Ортологи, паралоги, гомологи. Анализ генов паралогов в геноме человека.

#### **Тема 10. Особенности анализа экспрессии генов.**

Особенности анализа экспрессии генов в опухолях и при исследовании хронических заболеваний человека База данных GEO. Датасеты, их устройство. Депонирование данных. Инструменты для анализа данных. GEO2R.

#### **Тема 11. Анализ мутации в генах**

Мутации, их классификация. Синонимичные и несинонимичные замены. Полиморфизм в гене. вариации дозы гена. Функциональный анализ мутации. Предсказательные подходы: консервативность, программы Mutation taster, PolyPhen-2, SIFT+Provean. Экспериментальные методы.

#### **Тема 12. Теория дизайна лабораторного эксперимента и представления данных**

Выбор объекта исследования. Постановка задач и целей. Выбор методического решения поставленной задачи. Определение сроков эксперимента и объёма работы. Расходные материалы. Выбор журнала для публикации работы. Импакт фактор. Анализ основных разделов статьи. Логика формирования материалов для разных разделов публикации. Черновой вариант. Ответы рецензенту. Представление результатов в виде устных и постерных сообщений. Оформление работы.

## **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), доска для рисования, маркеры.

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

### Основная литература

1) Computational Biology and Applied Bioinformatics. Edited by Heitor Silverio Lopes and Leonardo Magalhães Cruz, ISBN 978-953-307-629-4, 456 pages, Publisher: InTech, Chapters published September 02, 2011 under CC BY-NC-SA 3.0 license. DOI: 10.5772/772

2) Applied Bioinformatics: An Introduction. Selzer, Paul Maria, Marhöfer, Richard, Rohwer, Andreas. 2008, XIV, 287 p.

### Дополнительная литература

1) Bioinformatics for High Throughput Sequencing. Rodríguez-Ezpeleta, Naiara, Hackenberg, Michael, Aransay, Ana M. (Eds.). 2012, XI, 255 p.

2) Bioinformatics. Tools and Applications. Edwards, David, Stajich, Jason, Hansen, David (Eds.). 2009, XII, 451 p.

3) Systems and Synthetic Biology. Editors-in-Chief: R. Weiss; P.K. Dhar.

## **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:**

д.ф-м.н. В.Ю. Макеев

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ  
ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ  
ПО ДИСЦИПЛИНЕ  
«Прикладная биоинформатика»**

**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);
- Готовность использовать современные методы и технологии научной коммуникации на государственном и иностранном языках (УК-4);
- 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 направленности 03.02.07 Генетика (ПК-1);

- Обладание представлениями о фундаментальных основах биологических процессов, 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);

- Обладание опытом профессионального участия в научных дискуссиях, умение представлять полученные в исследованиях результаты в виде отчетов и научных публикаций (стендовые доклады, рефераты и статьи в периодической научной печати) (ПК-4);

- Владение методами отбора материала, преподавания и основами управления процессом обучения фундаментальной биологии (ПК-5).

## **2. Показатели оценивания компетенций**

В результате изучения дисциплины «Прикладная биоинформатика» обучающийся должен:

### **знать:**

- основные геномные элементы;
- базы данных содержащих информацию о геномных элементах;
- инструменты для анализа геномных элементов;
- программы для представления данных по геномным элементам;
- различные форматы данных;
- методы с помощью которых получают биологические данные;
- методы исследования генома.

### **уметь:**

- проводить биоинформатический анализ участка генома
- вычленять информацию о функционировании участка генома;
- уметь собирать всю имеющуюся информацию из различных источников для анализа;
- проводить комплексный анализ различных видов данных;
- выстраивать гипотезы и планировать эксперименты для подтверждения наблюдаемых феноменов.

**владеть:**

- навыками освоения большого объёма информации;
- культурой постановки и моделирования биологических задач.

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

Промежуточная аттестация по дисциплине «Прикладная биоинформатика» осуществляется в форме зачета. Зачет проводится в устной форме.

1. Функциональные элементы генома
2. Анализ первичной структуры ДНК и её функции. Проект ENCODE.
3. Гены человека. Различные базы данных по генам.
4. Псевдогены, их классификация. Процессированные псевдогены. Механизмы функционального действия процессированных псевдогенов.
5. Регуляторные участки в геноме: промотор, ТАТА-бокс, энхансер, сайленсер, инсулятор.
6. Повторяющиеся последовательности в ДНК. Тандемные повторы: микросателлиты, минисателлиты и сателлиты. Болезни экспансии тринуклеотидных повторов.
7. Диспергированные повторы: транспозоны и ретротранспозоны. Открытие мобильных элементов.
8. Полиморфизм ДНК.
9. Функционирование вторичной структуры ДНК. G-квадруплексы в промоторах и теломерах.
10. Основные ресурсы для анализа функциональных элементов генома
11. Сервер NCBI. Основные базы данных расположенные на нём: GenBank, Genome, Gene, dbSNP,
12. Базы данных dbEST, UniGene, PubMed, OMIM, Protein, и другие. Запросы онлайн.
13. Геномные баузеры: MapView, USCS, ensemble.
14. Устройство и пользование баузером USCS. Создание индивидуальных сессий. Обмен информацией в сессиях. Сборки геномов.
15. Трэки, управление и настройки трэков. Трэк гены. Сравнение различных аннотаций генов человека.
16. Трэки мРНК и EST. Совместный анализ трэков гены, мРНК и EST.
17. Анализ структуры локуса генома человека.
18. Кодирующие и некодирующие РНК. lncRNA, miRNA.
19. Псевдогены. Повторяющиеся последовательности. Семейство Alu

19. Псевдогены. Повторяющиеся последовательности. Семейство Alu
20. Экспрессионный анализ. Анализ данных микрочипов Affymetrix. Устройство пробсета. База данных BioGPS.
21. Анализ данных RNAseq разных проектов. Количественный и качественный анализ. Выбор вида представления данных.
22. Данные проекта Fantom5.
23. Регуляция с помощью miRNA. CLASH data.
24. Антисмысловые взаимодействия.
25. Транскрипционные факторы.
26. Гистоновые модификации. Метилирование. Чувствительность к ДНКазам.
27. Поиск открытой рамки считывания. Определение белковых изоформ продукта. Анализ доменной структуры белка.
28. Онлайн инструменты: CDS, CDART. База данных Pfam.
29. Выявление белковых партнёров. Пост-трансляционные модификации белка.
30. Предсказание функции белка на основе его взаимодействий с белками-партнерами. Анализ данных базы «IntAct Molecular Interaction Database».
31. Анализ клеточных путей. База данных neXtProt. Данные проекта «The human protein atlas». PeptideAtlas. Uniprot
32. Сравнительно-геномный анализ локуса.
33. Консервативность гена, эволюция локуса. Ортологи, паралоги, гомологи. Анализ генов паралогов в геноме человека.
34. Особенности анализа экспрессии генов в опухолях и при исследовании хронических заболеваний человека База данных GEO. Датасеты, их устройство. Депонирование данных. Инструменты для анализа данных. GEO2R.

### Тестирование

1. Наиболее часто встречающаяся замена нуклеотида в геноме человека
  - а)  $T \rightarrow G$
  - б)  $C \rightarrow T$
  - в)  $A \rightarrow C$
2. Один из двух генов, возникших в результате дупликации исходного гена организма при хромосомной мутации это
  - а) Паралог
  - б) Ортолог
  - в) Синоним
3. Сайты связывания транскрипционных факторов с ДНК исследуют с помощью
  - а) RNA-Seq
  - б) BS-Seq
  - в) ChIP-Seq

#### **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.