

Министерство науки и высшего образования РФ  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»  
**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

**Б1.В.03 Основы биоинформатики**

---

наименование дисциплины (модуля) в соответствии с учебным планом

Направление подготовки / специальность

06.04.01 Биология

---

Направленность (профиль)

06.04.01.06 Геномика и биоинформатика

---

Форма обучения

очная

---

Год набора

2023

---

Красноярск 2023

## **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

Программу составили \_\_\_\_\_

д.ф-м.н., Профессор, Садовский Михаил Георгиевич

должность, инициалы, фамилия

## 1 Цели и задачи изучения дисциплины

### 1.1 Цель преподавания дисциплины

Целью преподавания дисциплины является формирование у магистров базовых знаний об этапах геномного проекта в части обработки данных секвенирования, способностей к решению вопросов геномики и биоинформатики как в области науки, так и в области медицины, сельского хозяйства, лесного хозяйства, животноводства.

### 1.2 Задачи изучения дисциплины

В задачи курса входит:

- знакомство с методами оценки качества результатов секвенирования,
- знакомство с алгоритмами попарного и множественного выравнивания нуклеотидных последовательностей,
- знакомство с статистическими методами обработки транскриптомных данных
- изучение основных баз данных и форматов хранения данных в биоинформатике.

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

Код и наименование индикатора достижения компетенции	Запланированные результаты обучения по дисциплине
<b>ПК-1: Способен осуществлять выбор форм и методов научно-исследовательской деятельности в соответствии с профилем научного исследования</b>	
ПК-1.2: Способен: - решать задачи, связанные с проведением исследований с использованием современных методических подходов и специализированного оборудования	
<b>ПК-3: Способен выполнять работы, связанные с исследованием и анализом генома и протеома живых организмов в т. ч. в областях здравоохранения, лесного хозяйства и охраны природы.</b>	
ПК-3.1: Умеет: - в полном объеме планировать и реализовывать проведение лабораторных молекулярно-генетических исследований живых организмов; - планировать и реализовывать проведение работ с биоинформационными ресурсами.	

<p>ПК-3.2: Владеет:</p> <ul style="list-style-type: none"> <li>- современными методами обработки и интерпретации генетической информации при проведении научных исследований;</li> <li>- методами обработки данных геномного секвенирования, полученных с разных платформ; способностью извлекать необходимые данные из банков генетических данных;</li> <li>- знаниями для обработки полученных результатов, анализа и осмысливания их с учетом имеющихся литературных данных.</li> </ul>	
<p>ПК-3.3: Способен:</p> <ul style="list-style-type: none"> <li>- использовать знания геномики и биоинформатики для объяснения важнейших биохимических процессов, протекающих в живых организмах, как в норме, так и при возникновении патологий;</li> <li>- ориентироваться в вопросах, связанных с анализом нуклеиновых кислот и белков;</li> </ul>	

#### **1.4 Особенности реализации дисциплины**

Язык реализации дисциплины: Русский.

Дисциплина (модуль) реализуется с применением ЭО и ДОТ

URL-адрес и название электронного обучающего курса: <https://e.sfu-kras.ru/course/view.php?id=12485>.

## 2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	е
		1
<b>Контактная работа с преподавателем:</b>	<b>0,83 (30)</b>	
занятия лекционного типа	0,28 (10)	
практические занятия	0,56 (20)	
<b>Самостоятельная работа обучающихся:</b>	<b>2,17 (78)</b>	
курсовое проектирование (КП)	Нет	
курсовая работа (КР)	Нет	
<b>Промежуточная аттестация (Экзамен)</b>	<b>1 (36)</b>	

### 3 Содержание дисциплины (модуля)

#### 3.1 Разделы дисциплины и виды занятий (тематический план занятий)

		Контактная работа, ак. час.							
№ п/п	Модули, темы (разделы) дисциплины	Занятия лекционного типа		Занятия семинарского типа				Самостоятельная работа, ак. час.	
				Семинары и/или Практические занятия		Лабораторные работы и/или Практикумы			
		Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС
<b>1.</b>									
	1. Методы секвенирования. Сравнительный анализ методов секвенирования и областей их применения. Секвенирование по Сэнгеру. Оценка качества секвенирования.	2							
	2. Введение в работу с сервером. Базы данных. Геномные браузеры. Форматы данных.	2							
	3. Алгоритмы выравнивания последовательностей и сжатия данных. Преобразование Барроуза -Уилера и Алгоритм Смита - Ватермана.	2							
	4. Статистические методы в биоинформатике. Методы анализа дифференциальной экспрессии генов.	2							

5. Предсказание структуры белка. Введение в структуру белка. Структурные свойства белков. Базы данных для классификации структуры: SCOP & CATH. Фолдинг белка: теория и эксперимент. Прогнозирование вторичной структуры, прогнозирование структуры ab initio, распознавание фолдинга (threading), сравнительное моделирование (основанное на гомологии).	2								
<b>2.</b>									
1. Методы секвенирования. Оценка качества результатов секвенирования по Сэнгеру. Редактирование хроматограмм.			2						
2. Введение в работу с сервером. Базы данных. Геномные браузеры. Форматы данных. Преобразование форматов файлов. SNP calling.			6						
3. Алгоритмы выравнивания. Парное и множественное выравнивание. Оценка качества и редактирование выравниваний.			4						
4. Статистические методы в биоинформатике. P-value, z-score, resampling. Дифференциальная экспрессия генов.			6						
5. Предсказание структуры белка. Базы данных для классификации структуры белков. Прогнозирование структуры ab initio, распознавание фолдинга (threading), сравнительное моделирование (основанное на гомологии). Практикум по работе на веб-сервере SWISS-MODEL и с другим ПО.			2						
<b>3.</b>									
1. Методы секвенирования.								16	

2. Введение в работу с сервером.							16	
3. Алгоритмы выравнивания последовательностей и сжатия данных.							16	
4. Статистические методы в биоинформатике.							16	
5. Предсказание структуры белка.							14	
Всего	10		20				78	



## **4 Учебно-методическое обеспечение дисциплины**

### **4.1 Печатные и электронные издания:**

1. Попов В. В. Геномика с молекулярно-генетическими основами(Москва: URSS).
2. Кребс Д., Голдштейн Э., Килпатрик С. Гены по Льюису: научное издание(Москва: Издательство "Лаборатория знаний").
3. Сетубал Ж., Мейданис Ж., Миронов А. А. Введение в вычислительную молекулярную биологию: перевод с английского(Ижевск: Институт компьютерных исследований).
4. Игнасимуту С. Основы биоинформатики: перевод с английского (МоскваМосква: [R&C Dynamics] Регулярная и хаотическая динамика [РХД]).
5. Глик Б., Пастернак Д., Янковский Н. К. Молекулярная биотехнология: принципы и применение: перевод с английского(Москва: Мир).
6. Козлов Н. Н. Математический анализ генетического кода: монография (Москва: БИНОМ, Лаборатория знаний).
7. Хаубольд Б., Вие Т., Чудов С. В., Артамонова И. И. Введение в вычислительную биологию. Эволюционный подход(Москва: Регулярная и хаотическая динамика).
8. Каменская М. А. Информационная биология: учебное пособие для вузов по направлению подготовки бакалавров и магистров 020200 "Биология" и биологическим специальностям(Москва: Academia (Академия)).
9. Леск А., Миронов А. А., Швядас В. К. Введение в биоинформатику: учеб. пособие: пер. с англ.(Москва: БИНОМ, Лаборатория знаний).
10. Титова Н. М., Савченко А. А., Замай Т. Н., Боровкова Г. И., Субботина Т. Н., Инжеваткин Е. В. Биохимия и молекулярная биология: учебно-методический комплекс дисциплины(Красноярск: ИПК СФУ).
11. Титова Н. М., Савченко А. А., Замай Т. Н., Боровкова Г. И., Субботина Т. Н., Инжеваткин Е. В. Биохимия и молекулярная биология: электронный учебно-методический комплекс по дисциплине (№ 175-2007) (Красноярск: СФУ).
12. Субботина Т. Н. Молекулярная биология: учеб.-метод. пособие для самостоят. работы [для студентов спец. 020208.65 «Биохимия»] (Красноярск: СФУ).

**4.2 Лицензионное и свободно распространяемое программное обеспечение, в том числе отечественного производства (программное обеспечение, на которое университет имеет лицензию, а также свободно распространяемое программное обеспечение):**

1. Современные биоинформатические исследования требуют умения решать поставленные задачи с использованием самого разнообразного программного обеспечения, от пользовательских скриптов, размещенных в репозиториях, до дорогостоящего проприетарного ПО, такого как CLC Genomics Workbench. Философия современного биоинформатического сообщества заключается в том, что любую задачу можно решить несколькими способами: с использованием бесплатно распространяемого ПО, при помощи онлайн-сервисов (пайплайнов) и проприетарного ПО, или самостоятельно создать новый программный продукт для решения конкретной пользовательской задачи. В рамках данного курса используется только свободно распространяемое ПО: BLAST, UGENE, MEGA, BioEdit, SeaView, PhyML, RaxML, MrBayes, PartitionFinder, jModelTest, IQTREE, FigTree, signalP, TargetP, TopPred2.

#### **4.3 Интернет-ресурсы, включая профессиональные базы данных и информационные справочные системы:**

1. Биоинформатика – та область знаний, в которой ресурсы Интернет используются практически для решения любой задачи. Вся биоинформатика основана на создании баз данных, наполнении их результатами научных работ исследователями со всего мира, открытости доступа к этим данным и сравнении новых результатов с уже опубликованными.
2. В рамках освоения дисциплины используется одна из крупнейших информационных систем в области биологии медицины, биофизики Национального центра биотехнологической информации (National Center for Biotechnology Information (NCBI)), США ([www.NCBI.nlm.nih.gov](http://www.NCBI.nlm.nih.gov)).
3. БД NCBI являются достаточно сложным инструментарием с разнообразным функционалом.
4. Ниже приведено краткое описание основных БД NCBI, которые могут быть полезны при освоении тем дисциплины.
5. БД Nucleotide (<http://www.NCBI.nlm.nih.gov/sites/Entrez?db=nucleotide>) объединяет данные последовательностей нуклеиновых кислот из нескольких исходных БД, в том числе GenBank, RefSeq и др. Данные могут быть найдены по регистрационному номеру, имени автора, наименованию организма, генома/белка, а также ряду других параметров.
6. БД Protein (<http://www.NCBI.nlm.nih.gov/sites/Entrez?db=protein>) является коллекцией аминокислотных последовательностей из нескольких источников, в том числе из GenBank, RefSeq и TPA, а также SwissProt, PIR, PRF и PDB.
7. БД Structure (<http://www.NCBI.nlm.nih.gov/Structure/index.shtml>) организуют доступ к результатам молекулярного моделирования макромолекул и связанным с ними БД: трехмерных биомолекулярных структур полученных с помощью рентгеновской кристаллографии и ЯМР-спектроскопии; БД химических структур небольших органических молекул; к информации об их биологической активности и т. д.

8. БД Gene (<http://www.NCBI.nlm.nih.gov/sites/Entrez?db=gene>) представляет собой инструмент для просмотра данных из широкого спектра геномов. Каждая запись – это один из генов определенного организма. Минимальный набор данных в гене запись включает уникальный идентификатор, т. н. Gene-ID.
9. БД dbMHC (<http://www.NCBI.nlm.nih.gov/gv/mhc/main.cgi?cmd=init>) предоставляет открытую платформу, где научное сообщество может размещать, просматривать и редактировать данные MajorHistocompatibilityComplex (МНС) для человека. БД dbMHC полностью интегрирована с другими ресурсами NCBI, а также с Международной рабочей группой гистосовместимости (IHWG).
10. DbSNP (<http://www.NCBI.nlm.nih.gov/SNP/>) – БД одиночных нуклеотидных полиморфизмов, полиморфных повторяющихся элементов, включающая как гибридные данные, так и полученные только экспериментальным путем.
11. БД ReferenceSequence (RefSeq) (<http://www.NCBI.nlm.nih.gov/RefSeq/>), содержащая последовательности, в том числе геномных ДНК, белков и т. д., является основой для проведения функциональных исследований, геной идентификации, сравнительного анализа и т. п. В частности, релиз от 11.07.2012 включал в себя описания 16 393 342 белков и 17 605 организмов.
12. БД Genomic Biology представляет собой объединение нескольких ресурсов и инструментов геномной биологии, в том числе геномных карт для Fruitfly, Human, Malariaparasite, Mouse, Rat, Retroviruses, Zebrafish и т. д., которые дополнительно содержат ссылки на интернет-ресурсы и БД, касающиеся рассматриваемых видов.
13. В БД UniGene (<http://www.NCBI.nlm.nih.gov/unigene/>) полноразмерные mRNA последовательности организованы в уникальные кластеры, представляющие известные или предполагаемые гены. Для кластеров доступна информация по картированию, экспрессии и другие ресурсы.
14. HomoloGene (<http://www.NCBI.nlm.nih.gov/homologene>) – инструмент для автоматизированного выявления гомологов среди аннотированных генов, который сравнивает нуклеотидные последовательности между парами организмов в целях выявления предполагаемых ортологов.
15. Basic Local Alignment Search Tool (<http://www.NCBI.nlm.nih.gov/BLAST/>) - основной метод поиска гомологичных последовательностей на основе локального выравнивания.
16. Public repository Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) - публичная электронная библиотека данных экспрессии генов «Омнибус Экспрессии Генов»

17. GenBank (<http://www.NCBI.nlm.nih.gov/genbank/index.html>) – БД, содержащая доступные последовательности нуклеотидов для более чем 260 000 организмов, вся информация в генетическом банке данных сопровождается библиографическими ссылками и биологическими аннотациями. GenBank автоматически интегрирует информацию о геноме и БД белковых последовательностей для изучения, учитывая таксономию, геном, белковую структуру и другую информацию.
18. Для представления последовательностей в GenBank предложено два инструмента:
19. ● BankIt – интернет-представление одной или нескольких последовательностей;
20. ● Sequin – интернет-представление для длинных последовательностей, полных геномов, результатов популяционных и филогенетических исследований.
21. Объединяющим фактором и при этом крайне удобным инструментом поиска в NCBI является поисковая система Search NCBI databases (<http://www.NCBI.nlm.nih.gov/sites/gquery>). Она обеспечивает одновременный доступ как к нуклеотидным и белковым последовательностям (GenBank, EMBL, DDBJ, PIR-International, PRF, Swiss-Prot и PDB, GenPept, RPF), 3-мерным структурам и популяционным данным, так и к библиографическим БД (PubMed, PubMed Central и т. д.). Доступ к поисковой системе Search NCBI databases может быть легко получен с помощью прямого интернет-адреса (<http://www.NCBI.nlm.nih.gov/gquery/>) либо посредством использования стартовой страницы NCBI (<http://www.NCBI.nlm.nih.gov/>). На этой странице приведен полный перечень инструментария и БД NCBI и существует возможность получить доступ к любой из перечисленных БД.
22. Крайне полезным инструментом, который сохраняет информацию о пользователе, используется для более точной настройки поисковых запросов в NCBI (<http://www.NCBI.nlm.nih.gov/index.html>) и т. д., является сервис «My NCBI» (<http://www.NCBI.nlm.nih.gov/sites/MyNCBI/>). Этот инструмент позволяет сохранять результаты поиска, выбирать форматы отображения, фильтрации, настраивать автоматический поиск и отправлять его результаты по электронной почте. Пользователи «My NCBI» могут сохранять свои БД, построенные на основе поисковых запросов в NCBI, и управлять политикой общественного доступа.

## **5 Фонд оценочных средств**

Оценочные средства находятся в приложении к рабочим программам дисциплин.

## **6 Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю)**

Аудиторный класс, наличие проектора для демонстрации наглядных пособий и экрана. Компьютерный класс, Internet, свободно распространяемое ПО: BLAST, UGENE, MEGA, BioEdit, SeaView, PhyML, RaxML, MrBayes, PartitionFinder, jModelTest, IQTREE, FigTree, signalP, TargetP, TopPred2.