

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»
(МФТИ, Физтех)

УТВЕРЖДАЮ

Ректор МФТИ

д-р физ.-мат. наук, профессор

Д. В. Ливанов



2025 г.

**Дополнительная профессиональная
программа повышения квалификации
«Прикладные аспекты анализа данных высокопроизводительного
секвенирования»**

УГСН 09.00.00 Информатика и вычислительная техника

Направление подготовки 09.03.01 «Информатика и вычислительная техника»

ОКВЭД 63.11 Деятельность по обработке данных, предоставление услуг по размещению информации и связанная с этим деятельность

Москва 2025

1. Общая характеристика программы

1.1.Цель реализации программы

Целью реализации дополнительной профессиональной программы повышения квалификации «Прикладные аспекты анализа данных высокопроизводительного секвенирования» является формирование у слушателей основных навыков разработки методов для анализа данных и приобретение ими практического опыта, необходимого для проведения самостоятельных научных исследований области вычислительной обработки биологических данных, полученных с помощью технологий высокопроизводительного секвенирования.

Лицам, успешно освоившим программу и прошедшим итоговую аттестацию, выдается удостоверение о повышении квалификации установленного образца.

1.2.Совершенствуемые и/или приобретаемые компетенции

Компетенции, формируемые и совершенствуемые в результате обучения, представлены в таблицах 1 и 2.

Таблица 1

№	Компетенции в соответствии с профессиональным стандартом 06.042 Специалист по большим данным	Код компетенции
1	Способен к созданию информационных технологий нового поколения, обеспечивающих экономически эффективное извлечение полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, и применение этих технологий в информационно-аналитической деятельности, в системах управления и принятия решений, а также для разработки на их основе новых продуктов и услуг	ПК-1

Таблица 2

№	Компетенции в соответствии с направлением подготовки 09.03.01 Информатика и вычислительная техника	Код компетенции
1.	Способен применять естественнонаучные и общетеchnические знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ОПК-1

1.3.Планируемые результаты обучения

Планируемые результаты обучения представлены в таблице 3.

Таблица 3

№	Уметь - знать	Направление подготовки 09.03.01 Информатика и вычислительная техника
		Квалификация: бакалавр
		Код компетенции
1	Знать:	ОПК-1

	<ul style="list-style-type: none"> – основы молекулярной биологии и механизмы, лежащие в основе методов исследования в генетике, геномике, молекулярной биологии – физические принципы и технологические решения, используемые в технологиях высокопроизводительного секвенирования; характеристики основных платформ высокопроизводительного секвенирования – основы операционной системы Linux, принципы построения компьютерных сетей, основные сетевые протоколы, файловую систему Linux, текстовые редакторы, права доступа Linux, принципы работы процессов Linux, оболочку Bash, основы администрирования файловой системы Linux, основы управления пакетами программ – основные типы ошибок, свойственные технологиям высокопроизводительного секвенирования, основные форматы данных, оценки качества чтений, тримминг – этапы сборки транскриптома De novo, используя стандартные протоколы подготовки образцов – основные принципы построения алгоритмов аннотации – принципы секвенирования фрагментов ДНК, обобщенная последовательность которых уже известна, с целью обнаружения индивидуальных отличий конкретного образца – основные принципы метода RNA-seq – принципы оценивания таксономического разнообразия метагенома на основе использования 16S рРНК последовательностей – область применения и сущность метода ChIP-seq, его характеристики, принципиальное отличие от метода ChIP-on-chip <p>Уметь:</p> <ul style="list-style-type: none"> – применять навыки информационного поиска для исследований в области молекулярной биологии, генетики – выбирать тот или иной метод секвенирования в зависимости от объекта исследования и постановки задачи – работать с командной строкой, настраивать графический интерфейс, управлять учетными записями пользователей, работать с файлами и каталогами, работать с 	
--	--	--

	<p>текстовыми редакторами, конфигурировать разделы файловой системы, конфигурировать дисковые квоты, управлять правами доступа к файлам и каталогам</p> <ul style="list-style-type: none"> – оценивать качество ридов, представляя их в виде графиков; определять источники ошибок в ридов; использовать программы для тримминга (например, Trimmomatic и т.п.) – применять различные алгоритмы сборки, в зависимости от длины изучаемых фрагментов; – анализировать качество сборки, используя специальные программы метрики оценки качества – правильно производить оценку качества аннотации – корректно применять различные подходы для картирования чтений – проводить анализ экспрессии генов – оценивать состав микробиоты с использованием полнометагеномного секвенирования, аннотировать и анализировать метаболические пути – обрабатывать и анализировать ChIP-Seq — данные и решать вычислительные задачи, возникающие при обработке данных, полученных с использованием высокопроизводительного секвенирования 	
		Профессиональный стандарт 06.042 Специалист по большим данным
		Код компетенции
2	<p>Знать: предметную область анализа, теоретические и прикладные основы анализа больших данных, типы анализа больших данных, виды аналитики, статистические методы, применяющиеся при анализе данных; методы и инструментальные средства управления аналитическими проектами по исследованию больших данных, современный опыт использования анализа больших данных.</p> <p>Уметь: разрабатывать и оценивать модели больших данных, осуществлять интеграцию и преобразование данных в ходе работ по анализу больших данных, адаптировать и развертывать модели в предметной среде; проводить аналитические работы с использованием технологий больших данных, как индивидуально, так и, осуществляя руководство малыми аналитическими группами.</p>	ПК-1

1.4. Категория обучающихся

Программа повышения квалификации предназначена для специалистов в области биоинформатики и (или) информационных технологий, имеющих высшее или среднее профессиональное образование, а также лиц, получающих высшее или среднее профессиональное образование.

1.5. Форма обучения

Очная с применением дистанционных образовательных.

Программа может быть реализована в сетевой форме.

1.6. Объем программы

72 академических часа.

1.7. Режим обучения

2 месяца, 9 ак. часов в неделю

2. Содержание программы

2.1. Учебный (тематический) план

Учебный (тематический) план программы представлен в таблице 4.

Таблица 4

№ п/п	Наименование разделов (модулей) и тем	Всего, час.	Лекции	Практич. работа	Самост. работа	Форма контроля
УЧЕБНЫЙ ПЛАН						
1.	Командная строка Linux	8	3	1	4	
2.	Основы языка R	8	4	2	2	
3.	Технологии высокопроизводительного секвенирования	6	2		4	
4.	Проверка качества секвенирования. Trimming.	6	1	1	4	
5.	Обработка результатов секвенирования. Ресеквенирование	10	4	2	4	
6.	RNA-seq	4	2	2		
7.	ChIP-seq	6	3	3		
8.	Полнометагеномное секвенирование	4	2	2		

№ п/п	Наименование разделов (модулей) и тем	Всего, час.	Лекции	Практич. работа	Самост. работа	Форма контроля
9.	De novo сборка геномов и транскриптомов	10	4	2	4	
10.	Аннотация геномных последовательностей	10	4	2	4	
	Итого:	72	30	16	26	Зачет по совокупности выполнения практических заданий

2.2.Календарный учебный график

Календарный учебный план составляется при сформированной группе с учетом уровня их подготовки.

Календарный учебный график отражает периоды теоретических занятий, практик, процедур промежуточной и итоговой аттестаций и т.д.

КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

Учебные недели	1	2	3	4	5	6	7	8
Учебные занятия (Т)	Т	Т	Т	Т	Т	Т	Т	Т
Практические занятия (П)	П	П	П	П	П	П	П	П
Самостоятельная работа (СР)	С Р	С Р	С Р	С Р	С Р	С Р	С Р	С Р
Стажировка (С)								
Контроль Зачет, экзамен (З, Э)								
Итоговая аттестация (А)								

2.3.Рабочая программа

Содержание учебной программы приведено в таблице 5.

Таблица 5

№ п/п	Наименование модуля, разделов и тем	Содержание обучения, наименование и тематика практических занятий (семинаров), самостоятельной работы	Объем, ак. час.
1.	Командная строка Linux	<p>Лекция. Командная оболочка Bash. Устройство файловой системы в операционных системах семейства Linux. Команды cd, ls, pwd, cp, mv, rm, more, head, tail, grep. Редактор vi.</p> <p>Практическая работа Выполнение заданий по теме лекции.</p> <p>Самостоятельная работа Самостоятельное выполнение заданий по теме лекции.</p>	8
2.	Основы языка R	<p>Лекция. Базовые понятия. R studio. Основы написания кода. Программные конструкции. Пакеты.</p> <p>Практическая работа Выполнение заданий по теме лекции.</p> <p>Самостоятельная работа Самостоятельное выполнение заданий по теме лекций.</p>	8
3.	Технологии высокопроизводительного секвенирования	<p>Лекция. Физические принципы и технологические решения, использующиеся в технологиях высокопроизводительного секвенирования. Характеристики основных платформ высокопроизводительного секвенирования.</p> <p>Самостоятельная работа Самостоятельное выполнение заданий по теме лекции.</p>	6
4.	Проверка качества секвенирования. Trimming.	<p>Лекция. Проверка качества секвенирования. Подбор условий тримминга ридов. Тримминг с помощью инструментов на выбор (trimmomatic/seqtk).</p> <p>Практическая работа Выполнение заданий по теме лекции.</p>	6

№ п/п	Наименование модуля, разделов и тем	Содержание обучения, наименование и тематика практических занятий (семинаров), самостоятельной работы	Объем, ак.час.
		<i>Самостоятельная работа</i> Самостоятельное выполнение заданий по теме лекций.	
5.	Одоброботка результатов секвенирования. Ресеквенирование.	<i>Лекция.</i> Картирование чтений на референсный геном. Преобразование Барроуза-Уилера для картирования ридов при секвенировании ДНК. Оценка качества картирования. SPN calling. Особенности, возникающие при детекции соматических мутаций. <i>Практическая работа</i> Выполнение заданий по теме лекции. <i>Самостоятельная работа</i> Самостоятельное выполнение заданий по теме лекции.	10
6.	RNA-seq	<i>Лекция.</i> Особенности картирования чтений, полученных в результате RNA- seq эксперимента на референсный геном. Методы нормализации и анализ экспрессии генов. <i>Практическая работа</i> Выполнение заданий по теме лекции.	4
7.	ChIP-seq	<i>Лекция.</i> Взаимодействие ДНК и белка. Методы для изучения ДНК-белкового взаимодействия, применяющего до появления высокопроизводительного секвенирования. ChIP-seq протокол. Основные методы анализа ChIP-seq данных. <i>Практическая работа</i> Выполнение заданий по теме лекции.	6
8.	Полнометагеномное секвенирование	<i>Лекция.</i> Таргетное секвенирование 16S рРНК. Таксономический анализ и анализ биоразнообразия. Полнометагеномное секвенирование. De novo сборка и аннотация генов. <i>Практическая работа</i> Выполнение заданий по теме лекции.	4

№ п/п	Наименование модуля, разделов и тем	Содержание обучения, наименование и тематика практических занятий (семинаров), самостоятельной работы	Объем, ак.час.
9.	De novo сборка геномов и транскриптомов	Лекция. Алгоритмы de novo сборки, основанные на графе де Брейна и графах перекрытий. Особенность геномных последовательностей, затрудняющих сборку. Оценка качества сборки. Практические аспекты больших геномных проектов. Особенности сборки транскриптомов de novo. Практическая работа Выполнение заданий по теме лекции. Самостоятельная работа Самостоятельное выполнение заданий по теме лекций.	10
10.	Аннотация геномных последовательностей	Лекция. Основные принципы построения алгоритмов аннотации. Оценка качества аннотации. Практические аспекты применения алгоритмов аннотации для эукариотических геномов. Практическая работа Выполнение заданий по теме лекции. Самостоятельная работа Самостоятельное выполнение заданий по теме лекций.	10
	Итого:		72

3. Формы аттестации и оценочные материалы

3.1. Формы аттестации

Итоговая аттестация по программе является обязательной,

Форма итоговой аттестации: совокупность выполненных практических заданий.

Критерии оценивания практических заданий:

- выполнение / не выполнение практических заданий.

В ведомость итоговой аттестации оценка выставляется в соответствии с нижеприведенной таблицей 6.

Таблица 6

Наименование модуля, разделов и тем	Основные показатели оценки	Формы и методы контроля и оценки
Прикладные аспекты анализа данных высокопроизводительного	ОПК-1, ПК-1	Выполнение практических заданий

секвенирования		
----------------	--	--

Оценивание проводится преподавателем на основе представленных ниже критериев и шкалы оценки.

3.1.1. Критерии оценивания компетенций (результатов)

Максимальная оценка 100%, в том числе:

- успешно выполнено 1 практическое задание — от 60% до 74%;
- успешно выполнено 2 практических задания — от 75% до 89%;
- успешно выполнено 3 практических задания — от 90% до 100%.

3.1.2. Шкала соответствия оценок

таблица 7

Оценка по пятибалльной шкале	Рейтинговая оценка, %
«Отлично» (5)	90 – 100%
«Хорошо» (4)	75 – 89%
«Удовлетворительно» (3)	60 – 74%
«Неудовлетворительно» (2)	Менее 60%

3.1.3. Описание шкалы оценивания

Итоговым результатом считается оценка по результатам выполнения практических заданий, проставленная преподавателем в зачетной-экзаменационной ведомости.

- «зачтено» с рейтинговой оценкой «отлично» выставляется слушателю, если сумма баллов по результатам выполнения практических заданий составляет 90% и более;

- «зачтено» с рейтинговой оценкой «хорошо» выставляется слушателю, если сумма баллов по результатам выполнения практических заданий от 75% до 89%;

- «зачтено» с рейтинговой оценкой «удовлетворительно» выставляется слушателю, если сумма баллов по результатам выполнения практических заданий от 60% до 74%;

- «не зачтено» выставляется слушателю, если сумма баллов по результатам выполнения практических заданий менее 60%.

3.2.Оценочные материалы

Таблица 8

Наименование модуля, разделов и тем	Основные показатели оценки	Формы и методы контроля и оценки	Вес задания, %
Темы 1-2	ОПК – 1, ПК - 1	Практические задания по темам и лекции	10
Тема 3-5	ОПК – 1, ПК - 1	Практические задания по темам и лекции	10

Наименование модуля, разделов и тем	Основные показатели оценки	Формы и методы контроля и оценки	Вес задания, %
Темы 6-10	ОПК – 1, ПК - 1	Практические задания по темам и лекции	30
Итоговая аттестация	ОПК – 1, ПК - 1	Выполнение практических заданий	50

Пример практического задания:

De novo сборка. Задачи практического задания:

1. Проведите сборку чтений образца SRR21730522 из предыдущего задания (геном *Escherichia coli*). Проведите оценку параметров сборки: длину генома, количество контигов, N50, L50, глубину и ширину покрытия. Сравните длину полученной сборки с известным значением для длины генома *Escherichia coli*.

2. Проведите сборку чтений образца SRR21016598 из предыдущего задания (транскриптом *Arabidopsis thaliana*). Для сборки используйте программу gnaSPAdes. Проведите оценку параметров сборки: количество контигов, долю чтений, которые картируются на сборку.

Геном *A. thaliana* содержит более 27 тысяч белок-кодирующих генов. Сравните полученное путем сборки количество контигов с этим значением, насколько сильно они отличаются? Предположите, с чем могут быть связаны различия.

Теперь попробуйте сделать сборку этих же данных с помощью стандартного геномного протокола SPAdes. Опишите, как изменятся параметры сборки.

4. Организационно-педагогические условия реализации программы

4.1. Учебно-методическое обеспечение и информационное обеспечение программы

4.1.1. Список литературы

Основная литература

1. Ребриков Д.В., Коростин Д.О., Шубина Е.С.: NGS. Высокопроизводительное секвенирование. Лаборатория знаний, - 2021.
2. Xinkun Wang Next-Generation Sequencing Data Analysis, - 2016.
3. Ion Mandoiu, Alexander Zelikovsky. Computational Methods for Next Generation Sequencing Data Analysis, - 2016.

Дополнительная литература

1. Phillip Corpean, Pavel Pevzner: Bioinformatics Algorithms: An Active Learning Approach, - 2014.
2. B.Alberts, et al. Ed. Garland Sc.: Molecular Biology of the Cell. 6th Edition, 2014.
Eija korpelainen, Jaro Tuimala, Pasi Somervuo, Mikael Huss, Garry Wong. RNA-seq Data Analysis: A Practical Approach, - 2014.

4.1.2. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся:

1. Коничев, А. С. Молекулярная биология: учебник для вузов / А. С. Коничев,
2. Г. А. Севастьянова, И. Л. Цветков. - 5-е изд., - 2021.
3. Скворцова Н.Н. Основы молекулярной биологии: Учебное пособие.
4. СПб.: Университет ИТМО; ИХиБт, - 2015.
5. Курячий Г.В., Маслинский К.А. Операционная система Linux. Курс лекций. Учебное пособие, - 2019.
6. Хамидуллин Р.Я. Теория вероятностей и математическая статистика, - 2020.

4.1.3. Ресурсы информационно-телекоммуникационной сети "Интернет"

1. Тестер регулярных выражений - <https://www.regextester.com>
2. Интерактивный учебник по SQL - <http://www.sql-tutorial.ru/ru/content.html>

4.2. Описание материально-технической базы, необходимой для осуществления образовательного процесса по программе

Таблица 9

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Система дистанционного обучения	Лекция	Персональные компьютеры/ноутбуки, доступ в Интернет, программное обеспечение Microsoft Visual Studio 2010 или выше
Система дистанционного обучения / аудитория	Практическая работа	Персональные компьютеры/ноутбуки, доступ в Интернет, программное обеспечение Microsoft Visual Studio 2010 или выше
-	Самостоятельная работа	Персональные компьютеры/ноутбуки, доступ в Интернет

5. Организация образовательного процесса

В таблице 10 описаны образовательные технологии.

Таблица 10


№ п/п	Вид занятия	Форма проведения занятий	Цель
1	Лекция	Ознакомление с теоретическими основами анализа данных	актуализация и систематизация теоретических знаний по дисциплине
2	Практическая работа	Выполнение практических заданий	осознание связей между теорией и практикой, повышение степени понимания материала

3	Самостоятельная работа	Самостоятельное изучение дополнительных материалов и литературы.	получение дополнительных теоретических знаний
4	Итоговая аттестация	По совокупности выполнения практических заданий	контроль освоения программы

6. Составители программы

Бахута Полина Александровна, ФБМФ, аналитик

Согласовано,
Эксперт ОСОП

 Ж. И. Зубцова

Согласовано
Директор ФБМФ

 Д. В. Кузьмин

Пояснительная записка
к разработке и реализации программы повышения квалификации
«Прикладные аспекты анализа данных высокопроизводительного
секвенирования»

№	Информация о программе и организаторе курса	Данные
1	Планируемое название дополнительной профессиональной или общеобразовательной программы (далее - программы)	Прикладные аспекты анализа данных высокопроизводительного секвенирования
2	Вид программы	Повышение квалификации
3	Выдаваемый документ	Удостоверение повышения квалификации установленного образца.
4	Форма обучения	Очно с применением дистанционных образовательных технологий
5	Режим обучения	Смешанные занятия
6	Объем, в ак. ч.	72 ак. часа
7	Подразделение	ФБМФ
8	Контактное лицо	Бахута Полина Александровна, ФБМФ, аналитик Тел.: +7 902 812 38 86 Эл.почта: bakhuta.pa@mipt.ru
9	Цель и задачи программы	Целью реализации дополнительной профессиональной программы повышения квалификации «Прикладные аспекты анализа данных высокопроизводительного секвенирования» является формирование у слушателей 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.
10	Краткое содержание программы	1. Командная строка Linux 2. Основы языка R 3. Технологии высокопроизводительного секвенирования 4. Проверка качества секвенирования. Trimming. 5. Обработка результатов секвенирования. Ресеквенирование 6. RNA-seq 7. ChIP-seq 8. Полнометагеномное секвенирование 9. De novo сборка геномов и транскриптомов 10. Аннотация геномных последовательностей
11	Целевая аудитория программы	Специалисты, имеющие высшее или среднее профессиональное образование, а также лица, получающие высшее или среднее профессиональное образование, успешно прошедшие вступительные испытания
12	Продолжительность	9 недель
13	Сроки (период) обучения	21 июля 2025 – 21 сентября 2025
14	Количество слушателей	30 человек

15	Источник финансирования	ФЛС
16	Стоимость обучения, р.	Тариф «Базовый» - 49 500 руб. Тариф «Эксперт» - 66 000 руб.
17	Условие запуска курса	Запись необходимого количества слушателей (25 человек)
18	Оборудование	Персональные компьютеры/ноутбуки, доступ в Интернет, программное обеспечение Microsoft Visual Studio 2010 или выше.
19	Состав преподавателей	Медведева Ю.А. – к.б.н., старший научный ФИЦ Биотехнологии РАН Ковтун А.С. – к.б.н., старший научный сотрудник лаборатории генетики микроорганизмов ИОГен РАН
20	Теги по программе	#анализданныхNGS

Директор ФБМФ

 Д.В. Кузьмин

«ПРИКЛАДНЫЕ АСПЕКТЫ АНАЛИЗА ДАННЫХ ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СЕКВЕНИРОВАНИЯ»

Директор ФБМФ

[illegible]

Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»
(МФТИ, Физтех)

ВЫПИСКА ИЗ ПРОТОКОЛА № 10

заседания учебно-методического совета от 23 июня 2025 года.

ПОВЕСТКА:

Рассмотрение дополнительных общеобразовательных и профессиональных программ.

Проректор по учебной работе А. А. Воронов.

СЛУШАЛИ: заместителя директора (Центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск") А. И. Рыбакову о представлении дополнительных общеобразовательных и профессиональных программ (Центр «Пуск», МФТИ).

ПОСТАНОВИЛИ:

Рекомендовать к утверждению в установленном порядке дополнительную профессиональную программу повышения квалификации
«Прикладные аспекты анализа данных высокопроизводительного секвенирования».

Решение принято единогласно.

Форма проведения заседания: заочная

Председатель УМС МФТИ

Ученый секретарь УМС МФТИ



А.А. Воронов

М.В. Березникова