

Национальный исследовательский университет

Высшая школа экономики

Концепция
Магистерской программы
«Анализ данных в биологии и медицине»

НАПРАВЛЕНИЕ 01.04.02

«Прикладная математика и информатика»

Степень (квалификация): магистр

Москва 2018

1. Общая характеристика

Форма обучения - очная.

Основной языки преподавания – русский и английский

Нормативный срок освоения – 2 года при очной форме обучения, трудоемкостью 120 зачетных единиц.

- Руководитель программы - д.б.н., профессор Михаил Сергеевич Гельфанд (gelfand@iitp.ru), заместитель директора по научной работе ИППИ РАН, руководитель учебно-научного центра «Биоинформатика» ИППИ РАН, профессор факультета биоинженерии и биоинформатики МГУ им. М.В. Ломоносова.

2. Цель программы

Целью программы является подготовка будущих лидеров биоинформатических исследований, способных разрабатывать и применять на практике вычислительные методы для решения задач в различных областях биологии и медицины. Основным преимуществом таких специалистов станет междисциплинарное образование и полноценное понимание как математического аппарата, так и биологических систем. Программа представляет собой конкурентный эталон магистерского образования и квалификации биоинформатиков, благодаря наличию которого в НИУ ВШЭ усилится интеграция информационных областей знания и смежных им дисциплин.

Данная программа является важным этапом развития факультета компьютерных наук. Благодаря программе и на её базе будут созданы различные курсы по биоинформатике, доступные для всех студентов ФКН, а также других факультетов через систему майноров. По мере появления квалифицированных кадров возникнет и возможность развития новых научных групп и научно-учебных лабораторий в области анализа данных в биологии в НИУ ВШЭ.

3. Целевая аудитория

Программа рассчитана на два контингента обучающихся:

- лица с подготовкой в области прикладной математики и/или анализа данных на уровне бакалавриата ведущих вузов, в которых ведется подготовка по прикладной математике, анализу данных и компьютерным наукам (НИУ ВШЭ, МГУ, МФТИ и др.), желающие войти в новую для себя предметную и профессиональную область наук о жизни;

- лица с хорошей биологической/медицинской и достаточной математической подготовкой, работающие в области наук о жизни и желающие приобрести или усовершенствовать навыки математического моделирования и анализа медико-биологических данных

4. Актуальность и значимость

В связи с масштабным развитием технологий в области наук о живых системах, современная биология ежегодно накапливает огромные массивы данных, для обработки которых необходимо грамотное владение компьютерными методами и специальными алгоритмами. Ввиду этого ежегодно возрастает спрос на высококвалифицированных специалистов в области вычислительной биологии и биоинформатики, способных применять математический аппарат для решения биологических проблем.

Данная программа строится на базе опыта, наработанного в рамках трех взаимосвязанных образовательных инициатив, в которых участвовал академический руководитель программы М.С. Гельфанд:

- факультет биоинженерии и биоинформатики МГУ;
- учебно-научный центр «Биоинформатика» Института проблем передачи информации им. А.А. Харкевича РАН;
- Московская школа биоинформатики (бывш. отделение биоинформатики Школы анализа данных Яндекса)

Интеграция ключевых центров по обучению и изучению биоинформатики является основным конкурентным преимуществом существующей программы.

5. Обоснование потребности в магистрах

Опыт существующих российских программ по подготовке биоинформатиков (Московская школа биоинформатики, Москва; Институт биоинформатики, Санкт-Петербург) указывает одновременно на интерес к дополнительному образованию и профессиональной переподготовке среди абитуриентов и на высокий спрос на специалистов с полным высшим образованием среди работодателей – биотехнологических компаний. Конкурс на подобные программы каждый год составляет около 5 человек на место. Кроме того, получение степени

магистра в области анализа данных в биологии позволит выпускникам успешно продолжать научную карьеру в аспирантурах России и за рубежом.

Последние несколько лет на базе магистерской программы «Науки о данных» факультета компьютерных наук реализуется вариативная дисциплина по медицинской информатике, год от года интерес к которой среди студентов устойчиво возрастает. Таким образом, внешняя по отношению к факультету потребность в развитии направления биоинформатики и вычислительной биологии дополняется благоприятной средой, уже формирующейся на самом факультете и создающей возможность конструктивного взаимодействия, обмена курсами между существующей и новой программами и т.п.

Магистерские программы по биоинформатике и вычислительной биологии широко распространены в ведущих мировых университетах:

- Гарвардский университет, Бостон, США, программа по вычислительной биологии;
- Стенфордский университет, Пало-Альто, США, программа по биоинформатике и анализу данных в медицине;
- Технический университет Мюнхена, программа по статистическим методам в биологии и медицине и др.;

В России возможность получить степень магистра в области биоинформатики предоставляют:

- факультет биоинженерии и биоинформатики МГУ, Москва
- Кафедра биоинформатики, факультет биологической и медицинской физики МФТИ, Москва
- магистерская программа «Биомедицинские науки и технологии», Сколковский институт науки и технологий, Москва;
- кафедра биоинформатики и медицинской кибернетики, институт фундаментальной медицины и биологии КФУ, Казань;
- кафедра биоинформатики факультета ВМиК ННГУ им. Н.И. Лобачевского, Нижний Новгород;
- кафедра биоинженерии и биоинформатики, институт приоритетных технологий ВолГУ, Волгоград;

- специальность «Биоинженерия и биоинформатика», биологический факультет СГУ, Саратов;
- кафедра информационной биологии, факультет естественных наук НГУ, Новосибирск;

6. Требования к выпускникам магистерской программы

Система подготовки магистров будет соответствовать оригинальному образовательному стандарту НИУ ВШЭ для подготовки магистров по прикладной математике и информатике. В частности, все требования оригинального образовательного стандарта применимы к выпускникам предлагаемой магистерской программы. В дополнении к этим требованиям, магистр в области прикладной математики должен понимать современные тенденции развития естественных наук и основные актуальные задачи естествознания, должен быть способен к совместной научной деятельности с представителями информатики, экономики, физики, биологии и других наук, должен уметь строить математические модели естественных систем. Магистр должен быть подготовлен к дальнейшему обучению в аспирантуре, а также к исследовательской деятельности.

В дополнение к основным перечисленным требованиям направления, выпускники данной программы по результатам обучения будут:

- Разбираться в актуальных задачах медицины и биологии
- Ориентироваться на рынке медицинской информатики и вычислительной биологии
- Способны вести научно-исследовательскую работу в области анализа омиксных данных
- Уметь работать с современными базами данных генетической и медико-биологической информации (OMIM, Cosmic, ClinVar, HGMD, dbSNP, 1000 Genomes и NHLBI-ESP 6500 exomes и др.)
- Понимать взаимосвязь между анализируемыми объектами и процессами, происходящими в организме человека
- Уметь определять оптимальные для решаемой задачи методы анализа медико-биологических данных
- Способны вести самостоятельную исследовательскую работу по анализу и моделированию биологических процессов
- Понимать, как эволюционные процессы связаны с молекулярными механизмами, происходящими в клетке и организме

- Способны вести аналитическую работу со структурами молекул, моделировать лекарственные препараты и анализировать их воздействие
- Способны выдвигать медико-биологические гипотезы и доказывать их с помощью компьютерных и математических методов

В соответствии с оригинальным образовательным стандартом НИУ ВШЭ, программа предусматривает формирование у студента следующих компетенций:

Системные (СК):

СК2.СК-М2 Способен создавать новые теории, изобретать новые способы и инструменты профессиональной деятельности.

СК5.СК-М5 Способен принимать управленческие решения и готов нести за них ответственность.

СК7.СК-М7 Способен организовать многостороннюю коммуникацию и управлять ею.

Профессиональные (ПК):

Социально-личностные (СЛК):

ПК1.СЛК-М1 Способен транслировать нормы здорового образа жизни, увлекать своим примером.

ПК 2.СЛК-М2 Способен задавать, транслировать правовые и этические нормы в профессиональной и социальной деятельности.

ПК 3.СЛК-М3 Способен использовать социальные и мультикультурные различия для решения проблем в профессиональной и социальной деятельности.

ПК 4.СЛК-М4 Способен определять, транслировать общие цели в профессиональной и социальной деятельности.

ПК 5.СЛК-М5 Способен к осознанному выбору стратегий межличностного взаимодействия.

ПК 6.СЛК-М6 Способен анализировать и разрешать для себя мировоззренческие, социально и личностно значимые проблемы.

ПК 7.СЛК-М7 Способен строить профессиональную деятельность, бизнес и делать выбор, руководствуясь принципами социальной ответственности.

ПК 8.СЛК-М8 Способен порождать принципиально новые идеи и продукты, обладает креативностью, инициативностью.

Б) Инструментальные (ИК):

ПК 9.ИК-М1.2н Способен организовать научно-исследовательскую деятельность.

ПК10.ИК-М2.1нми Способен анализировать и воспроизводить смысл междисциплинарных текстов с использованием языка и аппарата прикладной математики.

ПК11.ИК-М2.2нми Способен создавать междисциплинарные тексты с использованием языка и аппарата прикладной математики.

ПК12.ИК-М2.5 Способен публично представлять результаты профессиональной деятельности (в том числе с использованием информационных технологий).

ПК13.ИК-М4.1 Способен осуществлять целенаправленный многокритериальный поиск информации о новейших научных и технологических достижениях в сети Интернет и в других источниках.

ПК14.ИК-М5.1нми Способен описывать проблемы и ситуации профессиональной деятельности, используя язык и аппарат прикладной математики при решении междисциплинарных проблем.

ПК16.ИК-М7.1нми Способен использовать в профессиональной деятельности знания в области естественных наук, математики и информатики, понимание основных фактов, концепций, принципов теорий, связанных с прикладной математикой и информатикой.

ПК17.ИК-М7.2нми Способен строить и решать математические модели в соответствии с направлением подготовки и специализацией.

ПК18.ИК-М7.3нми Способен понимать и применять в исследовательской и прикладной деятельности современный математический аппарат.

ПК19.ИК-М7.4нми Способен в составе научно-исследовательского и производственного коллектива решать задачи профессиональной деятельности в соответствии с профилем подготовки, общаться с экспертами в других предметных областях.

ПК20.ИК-М7.5нми Способен применять в исследовательской и прикладной деятельности современные языки программирования и языки манипулирования данными, операционные системы, электронные библиотеки и пакеты программ, сетевые технологии и т.п.

По видам деятельности:

- научно-исследовательская: СК-2, ПК1-ПК-8, ПК-9, ПК-10, ПК-11, ПК-12, ПК-13, ПК-14, ПК-16, ПК-17, ПК-18, ПК-19, ПК-20;

- *проектная*: СК- 5, СК- 7, ПК1-ПК-8, ПК-12, ПК-15, ПК-19,

- *информационно-аналитическая*: СК- 6, ПК1-ПК-8, ПК-13, ПК-14, ПК-17, ПК-20;

В качестве итоговой государственной аттестации выпускник магистратуры должен подготовить и защитить магистерскую диссертацию. Результаты защиты могут учитываться как результаты вступительных экзаменов в аспирантуру по соответствующей специальности.

7. Структура учебного плана

Адаптационные курсы:

1. Комбинаторика и теория вероятности
2. Математика

Базовые курсы направления:

3. Современные методы анализа данных
4. Современные методы принятия решений: Алгоритмы в биоинформатике

Обязательные курсы программы:

5. Прикладная статистика
6. Молекулярная эволюция
7. Программирование на языке R
8. Анализ данных в биологии и медицине
9. Введение в молекулярную биологию
10. Практическая биоинформатика
11. Сравнительная геномика
12. Биоинформатика для высокопроизводительного секвенирования
13. Структурная биоинформатика и моделирование лекарств
14. Системная биология и персонализированная медицина
15. Медицинская информатика (совместный курс с программой «Науки о данных»)

8. Концепция научно-исследовательского (проектного) семинара по исследованию медико-биологических данных

В рамках проектного семинара студентам будет предложено выполнить научно-исследовательский проект. Основным объектом научно-исследовательской работы станут медико-биологические данные, например, геномные данные онкобольных или больных, страдающих нейродегенеративными заболеваниями. Партнерами семинара станут крупные медико-биологические центры, такие как ФНКЦ ДГОИ им. Дмитрия Рогачева и РГМУ им. Н.И. Пирогова.

Проектная работа включает в себя анализ литературных данных, выбор объекта исследования, постановку научной гипотезы, подготовку, реализацию и тестирование методов проверки гипотезы, анализ и визуализацию полученных результатов, представление проделанной работы в формате письменного отчёта (на русском языке) и презентации (на английском языке). Проектная работа может быть продолжена на летней практике или в рамках магистерской диссертации.

Примеры проектных работ: «Кластеризация раковых транскриптов», «Предсказание эпитопов Т-клеток», «Автоматический анализ регенеративной способности клеток кожи человека».

9. Трудоустройство выпускников

Междисциплинарность программы позволяет выпускникам выбирать дальнейшее направление профессионального развития среди множества вариантов. Одним из них может стать развитие научной карьеры, как в России, так и за рубежом. Во многих городах России, особенно в Москве и в Петербурге, в научных и научно-технических лабораториях, исследовательских группах и медико-биологических учреждениях существует постоянный спрос на специалистов вычислительной биологии и биоинформатики. Кроме того, после получения степени магистра выпускники программы получают возможность обучаться на PhD программам за границей и работать в зарубежных лабораториях. Другим направлением развития может стать работа в индустрии. Биотехнологические компании, которых с каждым годом становится всё больше и больше, постоянно находятся в поиске специалистов по вычислительной биологии. Например, на данный момент среди коммерческих компаний и лабораторий существует большой спрос на специалистов в области обработки данных высокопроизводительного секвенирования и микрочипов.

10. Обеспеченность профессорско-преподавательским составом

Команда преподавателей:

- д.б.н., проф. М.С.Гельфанд – зам. директора по научным вопросам и заведующий УНЦ «Биоинформатика ИППИ РАН, профессор ФББ МГУ (руководитель)
- д.б.н., проф. А.А.Миронов – профессор ФББ МГУ

- к.ф.-м.н. С.А.Спирин – с.н.с. НИИ ФХБ им. Белозёрского
- к.б.н. Г.А.Базыкин – зав. сектором ИППИ РАН
- к.ф.-м.н. А.Р.Рубинов – с.н.с. ИППИ РАН
- д.х.н. А.В.Головин – ст. преп. ФББ МГУ
- к.ф.-м.н. М.С.Попцова – доцент ДБДИП НИУ ВШЭ
- PhD О.С.Пьяных – профессор ДАДИИ НИУ ВШЭ
- к.ф.-м.н. П.В.Мазин – доцент кафедры ТМСС НИУ ВШЭ
- PhD Кертес-Фаркаш Аттила – доцент ДАДИИ НИУ ВШЭ
- Сапунов Г.В. – приглашенный преподаватель ТМСС
- к.ф.-м.н. Пятницкий М.А. – доцент ТМСС

Большинство членов команды преподают биоинформатику в течение многих лет на ФББ МГУ и в Московской школе биоинформатики. Михаил Сергеевич Гельфанд, Андрей Александрович Миронов, Сергей Александрович Спирин формировали программу обучения специалистов по биоинформатике, которая используется на факультете биоинженерии и биоинформатики МГУ.

11. Партнеры программы

В качестве партнёров программы могут выступить:

- НИИ Физико-химической биологии им. А.Н. Белозерского МГУ
- Институт общей генетики им. Н.А.Вавилова РАН
- Институт биоорганической химии им. М.М. Шемякина и Ю.А. Овчинникова РАН
- НИИ биомедицинской химии им. В.Н. Ореховича РАМН
- Институт проблем передачи информации им. А.А. Харкевича РАН
- Институт микробиологии им. С.Н.Виноградского РАН
- Институт молекулярной биологии им. В.А.Энгельгардта РАН
- НИИ физико-химической медицины ФМБА РФ
- Сколковский институт науки и технологий
- Московская школа биоинформатики
- Институт биоинформатики
- Лейденский Университет (Нидерланды)

А также компании:

- Литех (www.lytech.ru)

- Knomics (www.knomics.ru)
- Atlas Biomed Group (www.atlas.ru)
- iBinom (www.ibinom.com)

Все преподаватели в команде активно занимаются наукой, регулярно публикуют научные статьи и являются научными руководителями многих студентов и аспирантов.

Библиометрия

По Web of Science (http://expertcorps.ru/science/whoiswho/by_branch/bio-all)

- М.С.Гельфанд: 208 публикаций, 6680 цитирований, Хирш=46
- А.А.Миронов: 119 публикаций, 4538 цитирований, Хирш=34
- Г.А.Базыкин: 25 публикаций, 220 цитирований, Хирш=8 (первая публикация 2004 г.)
- А.В.Головин: 38 публикаций, 183 цитирований, Хирш=8

12. Участие ведущих зарубежных ученых

К участию в лекционных мероприятиях планируется привлечь следующих зарубежных учёных:

- Кондрашов Алексей Симонович, Мичиганский университет в Энн-Арбор, индекс Хирша 37
- Новичков Павел Сергеевич, Национальная лаборатория Беркли, индекс Хирша 14
- Равчеев Дмитрий Андреевич, Люксембургский центр системной биомедицины, индекс Хирша 7

13. Дальнейшие перспективы развития программы

Помимо основной программы АДБМ планируется активно развивать факультативные инициативы по изучению и анализу «омиксных» данных и их взаимосвязи с медицинскими показателями. Подобные программы будут одинаково востребованы среди студентов, выбирающих как академическую карьеру, так и карьеру в индустрии. Кроме того, активное развитие биологических дисциплин прогнозирует высокую востребованность квалифицированных кадров в области анализа данных, имеющих узкоспециальную биологическую специализацию. Одним из перспективных направлений видится создание

дополнительной программы по подготовке специалистов в области компьютерной нейробиологии.

14. Поступление на программу

Поступление на программу осуществляется по результатам конкурса портфолио и собеседованием с экзаменационной комиссией.