**МИНОБРНАУКИ РОССИИ**

**Федеральное государственное автономное образовательное учреждение высшего образования**

**«ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

**Институт высоких технологий и пьезотехники**

**Кафедра прикладной информатики и инноватики**

**Направление 09.03.03 - Прикладная информатика**

**Реферат**

**Дисциплина: «Введение в проектную деятельность»**

**Тема: «Биоинформатика»**

Выполнил:

студент 1 курса 6 группы

Лапич В. Р.  
\_\_\_\_\_\_\_\_\_\_\_\_

Проверил:

Кандидат физико-математических наук, доцент

Бычков А. А.  
\_\_\_\_\_\_\_\_\_\_\_\_

**Ростов-на-Дону – 2021**

**Содержание**

[**Введение** 3](#_Toc87910107)

[**1 История развития биоинформатики** 4](#_Toc87910108)

[**2 Цели и задачи биоинформатики** 5](#_Toc87910109)

[**3 Области применения биоинформатики** 5](#_Toc87910110)

[**4 Структурная биоинформатика** 7](#_Toc87910111)

[**5 Биоинформатика последовательностей** 7](#_Toc87910112)

[**6 Компьютерная геномика** 8](#_Toc87910113)

[**Заключение** 10](#_Toc87910114)

[**Список источников** 11](#_Toc87910115)

# **Введение**

Что такое биоинформатика? Смысл этого понятия связан с применением компьютерных технологий для решения проблем биологии и связанных с ней наук. Итак, биоинформатика — это междисциплинарная область, которая разрабатывает методы и средства для понимания биологических данных. В качестве междисциплинарной области науки биоинформатика сочетает компьютерный анализ, статистику, теорию вероятностей и набор инструментов для анализа и интерпретации биологических данных. Сбор данных, их хранение и интеллектуальный анализ данных стали основными проблемами для биотехнологов и ученых биологов из-за резкого роста количественных данных в биологии, таких как полные геномы биологических видов, включая геном человека, последовательности белков, трехмерные структуры белков, базы данных метаболических путей, клеточные линии, информация, связанная с биоразнообразием. Биоинформатика, используя современные методы, была призвана исправить эти проблемы. В наше время развитие информационных технологий, в частности интернета, используется для сбора, доступа и распространения постоянно растущей информации в области биологии и биотехнологии. Биоинформатика — это очень широкая область, и она охватывает такие вопросы, как картирование, секвенирование, сравнение последовательностей, идентификация генов, моделирование белков, сетевые базы данных, визуализация и этика. Это междисциплинарный предмет, который, с одной стороны, требует создания биологической информационной инфраструктуры, а с другой - требует биологических исследований, основанных на вычислениях. Все это зависит от большого объема экспериментальных и полученных данных.

Все это приводит к развитию ресурсов данного сообщества и с этого начинается разработка приложений биоинформатики для анализа генетических данных.

# **1 История развития биоинформатики**

Информационные технологии начали использоваться для хранения и анализа данных молекулярной биологии с 1960 года. Это было вызвано стремительным увеличением информации о строении белков, играющих в клетке регуляторную, ферментативную и строительную роль и нуклеиновых кислот, обеспечивающих хранение и передачу наследственной информации***.*** В это время впервые были опубликованы нуклеотидные последовательности малых РНК и разработаны алгоритмы предсказаний их вторичной структуры. Термин “Биоинформатика” был первоначально введен Беном Хеспером и Полиной Хогевег в 1970 году и определен как изучение процессов информатики в биотических системах. [2] Стандартным подходом к решению ряда биологических проблем являлся метод сравнения объектов с эталоном. К 1981 году было открыто 579 человеческих генов. Марвин Каррутерс и Лири Худ совершили огромный прорыв в биоинформатике, когда изобрели метод автоматического секвенирования ДНК. В 1988 году была основана международная организация генома человека. Это международная организация ученых, участвующих в проекте "Геном человека". В 1990 году был начат проект "Геном человека". К 1991 году было изучено в общей сложности 1879 человеческих генов. В 1993 году Генетон, исследовательский центр генома человека во Франции, создал физиологическую карту генома человека. Три года спустя Генетон опубликовал окончательную версию генетической карты человека. [3]

# **2 Цели и задачи биоинформатики**

Основными задачами биоинформатики являются:

1. Разработка алгоритмов анализа биостатистики для определения связей между элементами больших объёмов данных;
2. Формирование баз данных, способных реализовывать хранение наборов биологических данных и управление ими;
3. Применение этих средств для анализа и объяснения биологических показателей различного типа.

К первостепенным целям биоинформатики можно отнести:

1. Организацию данных наиболее удобным способом для экспериментаторов с целью внесения корректировок, по мере поступления новых сведений;
2. Разработка методов и программным средств для распознавания биологических данных, в частности, когда наборы данных большие и сложноустроенные.

# **3 Области применения биоинформатики**

Современную биоинформатику можно разделить на структурную биоинформатику, биоинформатику последовательностей и компьютерную геномику. Инфекционные заболевания в настоящее время являются основной причиной смерти детей и молодежи в мире. По данным ВОЗ, ежегодно от инфекционных заболеваний умирает более 13 миллионов человек. В развивающихся странах регистрируется наибольшее число смертей от инфекционных заболеваний, и этому способствовало отсутствие лекарств и высокая стоимость лекарств, если таковые имеются. Одной из основных возникающих проблем является разработка дешевых и эффективных лекарств, которая может быть решена путем рационального проектирования лекарств с использованием биоинформатики. Кроме того, фармацевтическая промышленность перешла от процесса проб и ошибок при поиске лекарств к рациональному и структурному дизайну лекарств. Благодаря успешному и надежному процессу разработки лекарственных средств можно сократить время и затраты на разработку эффективных фармакологических средств. Процесс идентификации целевого лекарственного средства и поиска кандидатов на лекарственное средство может быть ускорен, и на основе молекулярного моделирования и моделирования могут быть разработаны более безопасные или эффективные лекарственные средства. [7] Персонализированная медицина — это модель здравоохранения, которая адаптирована к уникальному генетическому коду каждого человека. Генетический профиль пациента может помочь врачу предсказать предрасположенность к определенным заболеваниям, обеспечить надлежащее лечение и соответствующую дозу для уменьшения побочных эффектов. Он применяется в лечении рака, заболеваний, связанных с диабетом, и ВИЧ. Биоинформатика используется в персонализированной медицине для анализа данных секвенирования генома или анализа экспрессии генов микрочипов в поисках мутаций или вариантов генов, которые могут повлиять на реакцию пациента на конкретный препарат или изменить прогноз заболевания. [5] Профилактическая медицина фокусируется на здоровье отдельных людей, сообществ и определенных групп населения. Она использует различные методы исследования, включая биостатистику, биоинформатику и эпидемиологию, для понимания закономерностей и причин болезней, а также для преобразования такой информации в программы, предназначенные для предотвращения заболеваний, инвалидности и смерти. Примером профилактической медицины является скрининг новорожденных сразу после рождения на наличие нарушений здоровья, таких как генетические заболевания или нарушения обмена веществ, которые поддаются лечению, но клинически не проявляются в период новорожденности. Для разработки таких скрининговых тестов для выявления заболевания на ранней стадии исследователи используют биоинформатические инструменты для анализа данных геномики на предмет возможных биомаркеров заболевания. [9]

# **4 Структурная биоинформатика**

Структурная биоинформатика направлена на решение биологических проблем путем анализа последовательностей биологических молекул, таких как ДНК и белок, с использованием вычислительных алгоритмов, средств информатики и программного обеспечения для оценки молекулярных данных. Некоторые из применений этой новой области приведены ниже: Предсказание структуры белка: Понимание корреляции между аминокислотной последовательностью и трехмерной структурой белка может быть полезно для определения структуры белка по его аминокислотной последовательности. Многочисленные инструменты биоинформатики могут быть использованы для прогнозирования структуры и функций белка, включая предсказание вторичной структуры, моделирование гомологии, пронизывание белка, предсказание мотива, домена, трансмембранной спирали, сигнального пептида и т. д. [10]

# **5 Биоинформатика последовательностей**

Этот раздел биоинформатики занимается анализом нуклеотидных и белковых последовательностей. Анализ последовательности — это термин, который представляет собой вычислительный анализ последовательности ДНК, РНК или пептида для извлечения знаний о ее свойствах, биологической функции, структуре и эволюции. Для эффективного проведения анализа последовательности важно сначала понять источник данных, т. е. различные экспериментальные методы, используемые для определения биологической последовательности. Затем следуя аналитическим стратегиям, выяснить, является ли последовательность геномной, транскриптомной или протеомной. Базы данных, в которых в настоящее время хранятся огромные данные об этих биомолекулах, необходимо будет сначала проверить на наличие аналогичных последовательностей, которые могли бы направлять экспериментальные анализы для функциональных исследований. Для проведения биоинформатического анализа часто используются программные средства и веб-сервисы. Небольшие органические молекулы или метаболиты, которые необходимы организмам для жизни и роста необходимо изучать в контексте их взаимодействия с генами и белками через метаболические пути. [1]

# **6 Компьютерная геномика**

На сегодняшний день определены почти полные последовательности геномов многих организмов. Прочтение полной нуклеотидной последовательности того или иного генома не является главной задачей. Различные проекты по секвенированию генома предоставляют множество информации о последовательности белка, но не содержат информации о структуре или функциях белка. Наиболее эффективным методом отсеивания полезных белков из этих геномных баз данных является компьютерное предсказание функции белка. Однако современные методы, которые в основном основаны на последовательности, ограничены степенью сходства между последовательностями неизвестной и известной функции; они все чаще терпят неудачу, поскольку идентичность последовательности расходится в сумеречную зону идентичности последовательности и выходит за ее пределы. На практике от 30 до 60 процентов всех белков могут быть функционально идентифицированы с помощью современного программного обеспечения, основанного на последовательности. Чтобы расширить уровень аннотации молекулярных функций на более широкий класс белковых последовательностей, необходимо будет разработать методы идентификации функции белка, основанные непосредственно на парадигме "последовательность-структура-функция". Представлен один из таких подходов. Идея состоит в том, чтобы сначала предсказать нативную структуру, используя методы сворачивания или нарезания нитей, а затем определить ее молекулярную или биохимическую функцию, сопоставив активный участок в предсказанной структуре белка с участком в белке известной функции. Основываясь на этих предварительных результатах, в ближайшие 5–10 лет, вероятно, будут разработаны вычислительные инструменты, которые позволят прогнозировать третичную структуру однодоменных белков со средним разрешением, более надежную идентификацию белковых лигандов, методы прогнозирования белков, имеющих специфические четвертичные взаимодействия, и начало восходящего подхода для идентификации важных белков в метаболических путях и путях передачи сигналов. [11]

# **Заключение**

Компьютерный анализ биологических последовательностей все больше оказывает влияние на биологию. Вычислительный анализ биопоследовательностей и инструменты поиска в базах данных в настоящее время являются неотъемлемой частью этой науки и привели к многочисленным важным научным открытиям за последнее время. Большинство из них были получены в результате поиска в базе данных, который выявил неожиданные сходства между молекулами, о которых ранее не было известно. Главной задачей на будущее является создание компьютерных методов, которые могут интерпретировать биопоследовательности, используя еще более глубокую интеграцию биологических знаний и статистических методов. Это позволит биологам работать на более высоком уровне, где их труд может принести максимальную пользу.

# **Список источников**

1. “Биоинформатика: метод во главе угла.” **URL**: <https://scfh.ru/papers/bioinformatika-metod-vo-glave-ugla/> (дата обращения: 13.11.2021).
2. “История развития биоинформатики” **URL**: <http://vladggu.ru/piter/72100-istoriya-razvitiya-bioinformatiki.html> (дата обращения: 13.11.2021).
3. “Проект «Геном человека».” **URL**: [https://ru.wikipedia.org/wiki/Проект\_«Геном\_человека»](https://ru.wikipedia.org/wiki/Проект_) (дата обращения: 13.11.2021).
4. “Биоинформатика-Bioinformatics.” **URL**: <https://ru.wikichi.ru/wiki/Bioinformatics> (дата обращения: 13.11.2021).
5. Базлов, Д. А. КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ МАКРОМОЛЕКУЛЯРНЫХ ОБЪЕКТОВ И БИОЛОГИЧЕСКИХ СИСТЕМ / Д. А. Базлов, А. В. Цивов. — 2. — Ярославль: ЯрГУ, 2013. — 48 c. — Текст: непосредственный.
6. “Что такое биоинформатика?” **URL**: <https://studfile.net/preview/1713924/> (дата обращения:14.11.2021).
7. “Структурная биоинформатика-все самое интересное на ПостНауке.”  **URL**: <https://postnauka.ru/video/42434> (дата обращения: 14.11.2021).
8. “Структурная биология, биоинформатика и молекулярная динамика” **URL**: <http://conf.bionet.nsc.ru/sbb2016/section_2/> (дата обращения: 14.11.2021).
9. “Данные внутри нас: чем занимаются биоинформатики?” **URL**: <https://habr.com/ru/company/skillfactory/blog/522444/> (дата обращения: 14.11.2021).
10. “Сравнение биологических последовательностей.” **URL**: <https://habr.com/ru/post/221813/> (дата обращения: 15.11.2021).
11. “Компьютерная геномика-Биоинформатика и проект генома” **URL**: <https://studwood.ru/1582224/estestvoznanie/kompyuternaya_genomika> (дата обращения: 15.11.2021).