

*Н.С. Зубарев*

## **БИОИНФОРМАТИКА И ЕЁ ПРИМЕНЕНИЕ В МЕДИЦИНЕ**

*Научный руководитель: канд. мед. наук, доц. Л.М. Сычик*

*Кафедра биологии*

*Белорусский государственный медицинский университет, г. Минск*

*N.S. Zubarev*

## **BIOINFORMATICS AND ITS APPLICATION IN MEDICINE**

*Tutor: PhD, associate professor L.M. Sychic*

*Department of Biology*

*Belarusian State Medical University, Minsk*

**Резюме.** В современных реалиях новейшие исследования в области биологии и медицины не обходятся без использования биоинформатики. В данной статье рассматриваются такие области применения, как аннотация геномов, молекулярный докинг и вычислительная биология рака.

**Ключевые слова:** Биоинформатика, докинг, вычислительная биология, рак.

**Resume.** Nowadays the latest research in the fields of biology and medicine can't exist without the use of bioinformatics. This article discusses such applications as genome annotation, molecular docking, and cancer computational biology.

**Keywords:** Bioinformatics, docking, computational biology, cancer.

**Актуальность.** В современном научном мире в ходе практических и теоретических исследований в области биологии и медицины учёным необходимо работать с огромным объёмом информации, которую человеку невозможно быстро проанализировать. С такими пластами данных справляются с помощью методов биоинформатики, поэтому данная междисциплинарная область сейчас активно развивается.

**Цель:** При изучении и анализе современной научной литературы рассмотреть основные области применения биоинформатики в биологии, охарактеризовать подходы и методы, используемые в медицине.

**Задачи:**

1. Изучить и проанализировать вклад в вычислительную биологию: геномику, фармакологию, биологию рака, как приоритетного направления современной онкологии.

2. Рассмотреть области применения аннотации геномов и проблемы, с которыми сталкиваются в ходе этого процесса.

3. Изучить процесс молекулярного докинга и его применение в медицине в области разработки лекарственных средств.

**Материалы и методы.** Изучение литературы, моделирование, анализ полученных данных и синтез материала на их основе.

**Результаты и их обсуждение.** Биоинформатика — это междисциплинарная область, объединяющая общую биологию, молекулярную биологию, кибернетику, генетику, химию, компьютерные науки, математику и статистику. Есть множество таких масштабных биологических проблем, которые требуют анализа больших объёмов данных и использования компьютерных моделей, как аннотация геномов, молекулярный докинг и вычислительная биология рака.

Анализ генетических последовательностей позволяет анализировать сходство генов в рамках одного вида или разных, производить поиск генов и регуляторных последовательностей в геноме, составлять филогенетические деревья.

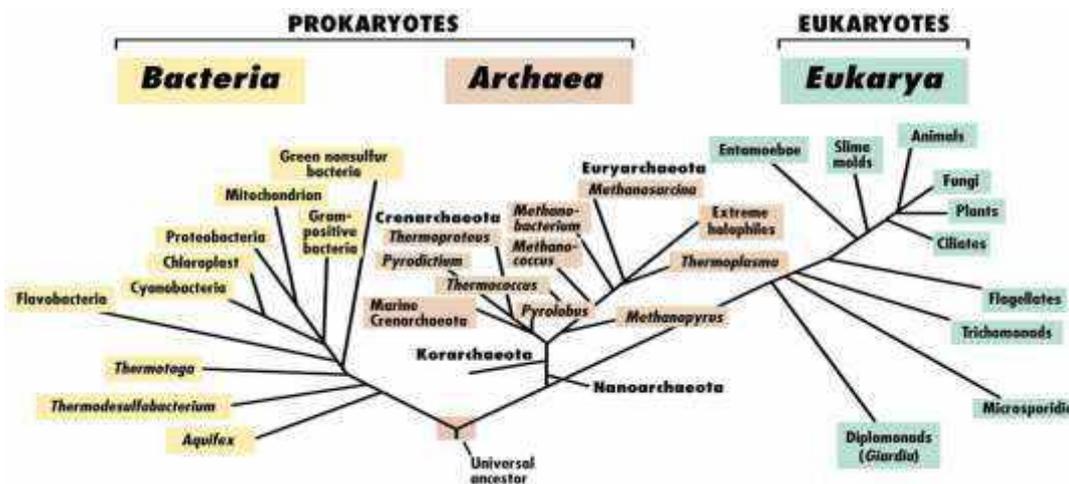


Рис. 1 — Пример филогенетическое дерева

Вычислительная биология, включающая многие аспекты биоинформатики, - это наука об использовании биологических данных для разработки алгоритмов или модели для понимания биологических систем и взаимосвязей. Вычислительная геномика изучает геномы клеток и организмов. Если вы работаете с геномом бактерий, то компьютеры могут напрямую вычислить расположение генов, поскольку они уложены очень простыми блоками. Но у эукариот гены не устроены такими блоками — они разбиты на фрагменты: экзоны и интроны. Специалисты разработали компьютерные программы, которые могут определять расположение генов, но результаты такого анализа до сих пор не очень надёжны. По этой причине исследователи в области вычислительной геномики работают над пониманием функций некодирующих областей человеческого генома путём разработки вычислительных и статистических методов. После полной реализации это может позволить врачам анализировать геном отдельного пациента. Это открывает возможность персонализированной медицины, прописывая лечение на основе ранее существовавших генетических паттернов человека.

Вычислительная фармакология позволяет на основе химических и геномных данных идёт поиск взаимосвязи с эффективностью действия лекарства, однако необходим, опять же, анализ огромных массивов данных, так что идёт разработка методов и инструментов, тут мы и возвращаемся к проблеме вычислительной геномики.

Рассматривая вычислительную биологию рака, мы прежде всего говорим о том, что ДНК раковых клеток расшифровываются по всему миру, и сейчас уже есть открытые базы, из которых можно скачивать геномные данные опухолей, сравнивать их. Основное внимание уделяется:

1. Характеристике опухолей.
2. Анализу экспрессии генов, транскриптома в раковых клетках.
3. Построению атласов, карт сигнальных путей, вовлечённых в канцерогенез.

Также необходимо для понимания изложить основные характеристики раковой клетки:

1. Поддержание высокого уровня сигналов к размножению.
2. Подавление супрессоров роста.
3. Неограниченное деление.
4. Избегание иммунного надзора.
5. Устойчивость к апоптозу.
6. Нестабильность генома.
7. Стимуляция ангиогенеза.
8. Метастазирование.

Различные учёные и кампании предпринимают попытки изобразить всё в виде карты сигнальных путей, на основании взаимосвязи сигналов, получаемых клеткой и внутренних процессов.

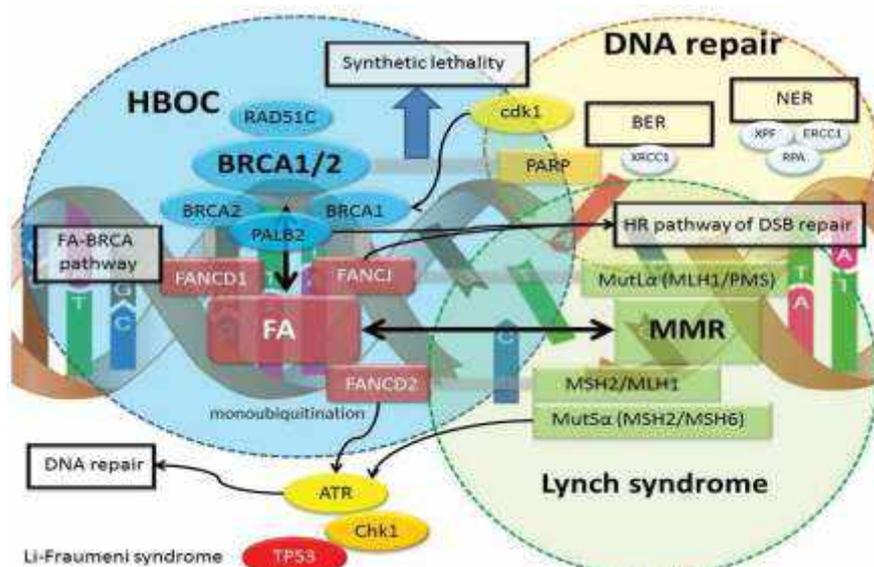


Рис. 2 — Атлас генома рака

Атласы на данный момент разрабатываются во множестве стран, это крайне перспективное направление, которое стремится помочь в изучении частных видов рака. Суть: есть несколько основных блоков, в которых анализируется гены и выявляются закономерности, приводящие к развитию рака.

Пример острой миелоидной лейкемии: В 99,5 % случаев идентифицируется по меньшей мере одна несинонимическая мутация в каком-то гене из следующего списка: гены опухолевого супрессора (15,5 %), гены, связанные с метилированием ДНК (43,5 %), сигнальные гены (59%), гены, модифицирующие хроматин (30,5 %), гены миелоидного транскрипционного фактора (22 %), гены когезинового комплекса (13%) и сплайсосомные гены (13,5%).

Этапы исследования:

1. Анализируются большие объёмы данных транскриптомов.
2. Составляются графики экспрессии генов в раковых клетках.
3. Производится деление опухолей на молекулярные подтипы/кластеры.

На основе этих данных уже может организовываться персонализированный подход к лечению больных.

Молекулярный докинг — метод молекулярного моделирования, позволяющий предсказать наиболее выгодную для образования устойчивого комплекса ориентацию и конформацию одной молекулы (лиганда) в сайте связывания другой (рецептора). Данные о положении и конформации партнёров используются для предсказания силы взаимодействия посредством так называемых оценочных функций. В случае, если лиганд является макромолекулой, докинг называют макромолекулярным.

В основе данного метода лежит концепция “ключ-замок”, соответственно главная задача - найти оптимальную конформацию. Данное положение достигается в случае, когда свободная энергия связывания минимальна.

Для этого нам в первую очередь необходима структура мишени. Для этого можно использовать: рентгеноструктурный или ядерномагнитный анализ. Также есть структуры, которые можно скачать из общей базы данных (Например, RCSB PDB). Далее мы находим структуру лиганда и программное обеспечение (Различные молекулярные редакторы (Avogadro, ChemBio3d и т.д.), программы для самого докинга).

Опишем наши пошаговые действия:

1. Скачаем интересующую структуру, например, протеина из PDB.
2. Подготовим лиганд. При скачивании структуры она обычно подаётся без атомов водорода, которые нам необходимы для дальнейшей работы. Также различные заместители в молекуле могут быть протонированными
3. Используем молекулярные редакторы для корректировки структуры лиганда и протеина (воспользуемся, например, программой Avogadro). Из файла с протеином удаляются низкомолекулярные соединения, молекулы воды (хотя они влияют на взаимодействие белка с лигандом)
4. Программа определяет активный центр белка и строит сетку потенциалов для докинга:

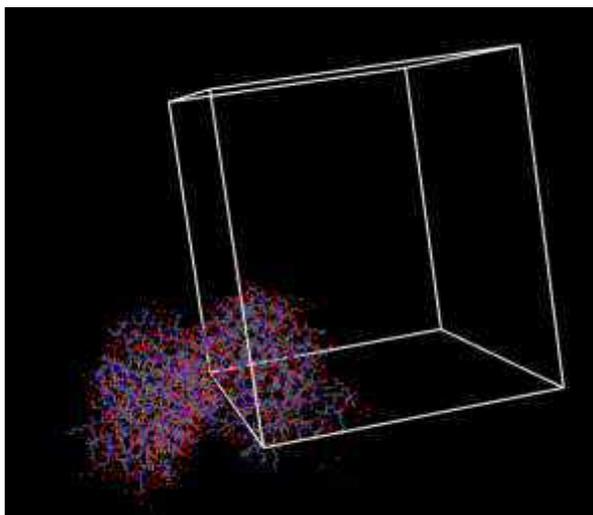


Рис. 3 — Сетка потенциалов для докинга

5. Осуществляем сам докинг в определённых программах и обрабатываем ре-

зультат. На выходе получается комплекс показателей, описывающих средство с данным белком. Это всё — параметры скоринг-функции.

Также необходимо описать сами методы докинга. Жёсткий докинг. В данном случае считаются неизменными:

1. Длины связей.
2. Углы.
3. Торсионные углы.

Важно понимать, что при взаимодействии с лигандом белок претерпевает конформационные изменения, поэтому жёсткий докинг некорректно описывает схему взаимодействия.

Выходом из ситуации является гибкий докинг. Здесь допускаются конформационные изменения, что даёт наиболее приближенные результаты. Однако это значительно замедляет процесс, так что нужно учитывать лишь небольшое количество изменений.

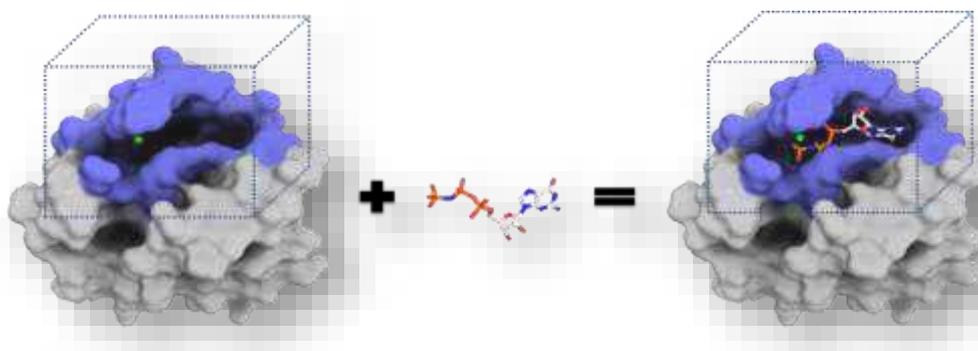


Рис. 4 — Иллюстрация молекулярного докинга.

Таким образом, мы можем предсказать аффинность и активность молекулы-лекарства по отношению к белку-мишени, а из библиотеки определить наиболее подходящее лекарство по степени взаимодействия. Эти данные используются в дальнейшем для разработки лекарственных средств.

**Выводы:** применение в медицине биоинформатики как науки, сочетающей в себе такие дисциплины, как химия, молекулярная и общая биология, компьютерные науки и статистика, позволяет решить ряд вопросов по оценке рисков, ранней диагностике, подбору оптимального лечения и дальнейшему прогнозу заболевания.

#### Литература

1. Баженова О., О'Брайен [Электронный ресурс]. Применение биоинформатики в медицинских исследованиях // Здоровье – основа человеческого потенциала: проблемы и пути их решения. 2014. №1. Режим доступа: <https://cyberleninka.ru/article/n/primenenie-bioinformatiki-vmeditsinskihissledovaniyah>.
2. Гуреев М.А., Кадочников В.В., Порозов Ю.Б. Молекулярный докинг и его верификация в контексте виртуального скрининга. – СПб: Университет ИТМО, 2018 – 50 с.
3. D. H. Huson and D. Bryant, Application of Phylogenetic Networks in Evolutionary Studies Mol. Biol. Evol., 23(2):254-267, 2006.