

## ЛЕКЦИЯ

### БИОИНФОРМАТИКА, ГЕНОМИКА И ПРОТЕОМИКА - НАУКИ О ЖИЗНИ XXI СТОЛЕТИЯ.

АРЧАКОВ А.И.

Биоинформатика, геномика и протеомика - это науки о жизни, которые будут определять начало XXI века, так же как молекулярная биология, иммунология и биотехнология определяли конец XX века.

Геномика возникла раньше двух других наук этой триады и полностью вошла в наш век. Геномика определяет развитие современных наук о жизни. Это наука, изучающая структуру и функцию генов, наука, которая инвентаризирует гены, создавая, таким образом, геномные карты живых существ. 20 лет назад никто бы не поверил в возможность инвентаризации всего генома человека. Сейчас это реальность. Прежде всего, инвентаризируются гены, связанные с заболеваниями. Можно назвать точную дату возникновения геномики - октябрь 1990 года, время подписания проекта "Геном человека". Полное прочтение генома человека планировалось в 2005 году. Наука идет с опережением, уже сделан непрерывный сиквенс нескольких хромосом. Теперь можно с уверенностью сказать, что полное чтение генома человека закончится в 2001 году, когда будет получен сиквенс 23 хромосом человека, это 140 тыс. генов, содержащих 3 млрд оснований с точностью до 0,01%. На проведение этих работ планировалось 3 млрд. долларов, но обойдется дешевле. Сомнений в необходимости проведения подобных работ нет. Работы в этом направлении будут проводиться еще быстрее, т.к. геном человека - это некий ориентир в развитии техники сиквенирования. Самое важное для человечества в ближайшем будущем - определение геномов микроорганизмов. Будут прочитаны законы возбудителей наиболее страшных заболеваний и на основе этого созданы новые лекарства для лечения этих заболеваний. Если до 1998 года исследования геномов микроорганизмов развивались за счет государственного финансирования, то сейчас в эти исследования вкладывают деньги все ведущие фармацевтические фирмы. Знание генома человека очень важно для медицины. Раньше считали, что генетика - это только наследственные заболевания. Сейчас ясно, что наследственные заболевания составляют всего 2 % от всего числа заболеваний, т.е. не являются социально значимыми. До недавнего времени 98% заболеваний не связывалось с генетикой, теперь стало совершенно очевидно, что "все болезни от генов". При наследственных болезнях наблюдается дефект гена, а в других случаях - нарушение регуляции экспрессии гена. Для каждого патологического процесса можно выявить нарушение регуляции гена, в результате чего развивается этот патологический процесс. Все это социально значимые заболевания. Наблюдается

тесная связь: геномика → дефектные гены → генная терапия. В настоящее время актуальны исследования по созданию новых лекарственных препаратов с использованием антисмысловых олигонуклеотидов, блокирующих экспрессию того или иного гена. Это то, что уже дала геномика. Основной недостаток геномики: мы знаем весь геном, но не знаем, что реально экспрессируется и что реально работает в клетке.

Из геномики выросла новая наука - протеомика. Протеомика занимается инвентаризацией белков, т.е., реально работающих молекулярных машин, в клетке. Задача протеомики на несколько порядков сложнее, чем у геномики. Всего 10 лет назад мы начали инвентаризировать гены, сейчас мы можем инвентаризировать белки. В настоящее время возможно не только считывать последовательности, но и читать и анализировать все модифицированные белки: фосфорилированные, гликозилированные, процессированные и многие другие. Самое главное то, что мы можем прямо в патологически измененных тканях видеть диспропорцию между белками. Протеомный анализ ведут следующим образом. Первый этап - двумерный электрофорез, который дает разделение по молекулярной массе и изоэлектрической точке. Второй этап - анализ. Показано, что в патологически измененных тканях содержание некоторых белков увеличено (это имеет значение для онкологии), других - уменьшено. Третий этап - масс-спектрометрия, с помощью которой можно прочесть последовательности этих белков с точностью до атомов. Можно выходить на атомарное разрешение и определять остатки глюкозы, фосфорной кислоты в молекуле белка, можно читать смеси белков, не разделяя их. Важно знать, что вся эта информация имеется в современных банках данных. Там можно найти до 4 млн. геномных последовательностей, но только 10 тыс. известных трехмерных структур для 2 тыс. отдельных белков (данные на октябрь 1999 года). Поэтому о белках, реально работающих молекулярных машинах клетки, мы, фактически, ничего не знаем. Задачей протеомики является проанализировать белок, установить его последовательность, соотнести с банком данных, сделать рентген и установить структуру. Стоимость проведения рентгеноструктурного анализа отдельного белка составляет 200 тыс. долларов, реально довести стоимость анализа до 20 тыс. долларов, но, так или иначе, эта наука не для нас, - слишком дорого. Самый простой комплекс, который позволяет работать в этой области, стоит 500 - 700 тыс. долларов - тоже не для нас. Экономически наиболее доступная для нас наука - биоинформатика.

Биоинформатика - это область науки, которая занимается примерно тем, чем занимались классическая биохимия, молекулярная биология и биотехнология, но не в пробирке, а с помощью вычислительной, компьютерной техники. Это, так называемая, "сухая" биохимия, не из области технократических наук, где все решает техника, а интеллектуальная наука, где есть место разуму, мысли и образованию. Биоинформатика - это путь от гена к лекарству через структуру макромолекулы. Все, что раньше мы делали с помощью, теперь недоступных для нас, экспериментов, включая ЯМР, рентгеноструктурный анализ, сейчас можно сделать с помощью вычислений. В это мало кто верит, но факт остается фактом. Если есть геном, его можно разметить и найти границы гена не при помощи клонирования отдельных генов, а с помощью определенных компьютерных программ. Если есть последовательность белка, можно перейти к пространственной структуре и функции. Это не сказка, не фантазия, а реальность. Кроме того, на основании этих пространственных моделей можно

сконструировать определенные лекарства. Не все фармакологи принимают биоинформатику восторженно, но эта наука бурно развивается за рубежом, и несмотря ни на что и у нас в стране, революционным является следующее: вместо нескольких сотен мишеней, на которые направлено действие лекарств сегодня, эта наука готова дать десятки тысяч новых мишеней, предсказать функцию генов, дешифровать последовательность белков. Все до удивления просто. Если получена новая последовательность ДНК→белок, ее можно сравнить с существующими последовательностями в банках данных. Если существуют гомологи - ответ ясен, можно, с определенной степенью точности, предсказать структуру, функцию белка. Даже в случае, если нет гомолога, то можно также что-то сделать, но с гораздо меньшей точностью. Наука далеко продвинулась в геномике, анализе генетического материала. Что мы делаем с геном? Мы размечаем геном, функционально классифицируем гены, предсказываем функцию генов и, затем, это вторая часть анализа генома, мы выбираем гены-мишени, на которые направлены наши лекарства. Что значит экспериментально установить функцию белка? Это огромный, трудоемкий процесс, требующий огромных затрат. Если вы работаете в области информатики и дешифруете последовательность белка, вы работаете так, как работали криптографы во все времена - из последовательности отдельных букв читается весь биологический текст, и, на основании этого, выясняется, что это за белок. Наиболее впечатляющие результаты были доложены этим летом проф. Айзенбергом в Австрии. Используя дешифровальные методы анализа 6 тыс. белков, не проводя ни одного эксперимента, он предсказал функцию сразу 2,5 тыс. белков. Приведенные им результаты производят громадное впечатление. Не так давно биохимики тратили всю жизнь, чтобы определить структуру одного белка и выявить его функцию. С помощью новых подходов можно одновременно предсказать функцию тысячи белков. Это новое слово в науке. Самым интересным остается вопрос, как в настоящее время делают новые лекарства. Геном человека будет расшифрован в начале следующего года. Определяются геномы микроорганизмов. Затем ведут поиск пересечения между ними, выясняют, какие гены важны для функции тех или иных микроорганизмов, накладывается геном эукариотических клеток, исключается некоторое количество генов, существующих у людей, чтобы их не повредить, и на область пересечения направляется лекарство. Одна из первых работ в этой области появилась в 1999 году. В нашем институте ведутся подобные исследования А. Дубановым. Наше программное обеспечение позволяет, анализируя геномные данные и применяя балльную систему, дать каждому гену определенную оценку. Например, если исходить из 10-балльной системы, нескольким тысячам генов можно поставить 10, т.е., если мы подберем лекарство, которое найдет эту мишень, успех обеспечен. Если оценка меньше 5, то и стараться нечего, на такие мишени нечего направлять лекарство. С помощью биоинформатики Е.Ф. Колесанова и сотр. занимаются созданием вакцин против вируса гепатита С. Проводили компьютерный анализ белков Е1 и Е2, вируса гепатита С. Была создана база данных, в этой базе данных были найдены участки, которые, по сути дела, инвариантны для всех 827 последовательностей оболочечных белков. Известно, что вирусная изменчивость - основное препятствие для создания вакцин, а найденные пептиды консервативны. Если будет получена антипептидная вакцина к какому-то набору из этих пептидов, если эта вакцина будет обладать вирус-нейтрализующими свойствами, то проблема изменчивости не будет играть

существенной роли. Самое главное - зная консервативные последовательности, можно провести синтез антигенных детерминант и создавать вакцину. Консервативные участки прямо или опосредованным путем участвуют во взаимодействии вируса с рецептором. Это очень важно. Если мы сделаем такой пептидомиметик, то есть шанс сделать лекарство, которое блокирует это взаимодействие. Консервативные участки - это участки нуклеации, т.е. участки, которые играют наиболее важную роль, когда из линейной структуры молекула превращается в глобулу. Это может быть новый класс лекарств, который будет блокировать фолдинг белков. Это может стать основным путем, который позволит бороться с вирусными инфекциями. Это нечто совершенно новое! Консервативные участки существуют у всех без исключения вирусных вариантов. У разных вирусов они разные, но у одного типа вируса они одинаковы. Как перейти от последовательности к пространственной структуре? Переход может быть сделан расчетным путем. Всю глобулу, ядро белка и активный центр можно смоделировать, рассчитать по гомологии, если идентичность порядка 40%, с разрешением до 0,2 нм, т.е. разрешением рентгеноструктурного анализа. Сложнее рассчитать другие части молекулы: поверхность, мембранные участки белка. Разрешение в этом случае составляет 0,4 - 0,6 нм, это уже очень большие атомарные расстояния. Очень быстро происходит внедрение биоинформатики в фармакологию. Классическая фармакология не очень хорошо принимает эти новые концепции, однако биоинформатика уже получила мировое признание, работы в этом направлении ведутся во всем мире. Фармакологи должны знать, куда направлять лекарство. Пока фармакология проводит ковровые бомбардировки, биоинформатика стремится к тому, чтобы иметь конкретные цели для конкретных участков, знать, куда направлять то или иное лекарство. Чтобы проводить такого рода исследования быстро, нужны расчеты. Путь нового лекарства до рынка 12-15 лет, до рынка доходит 1 соединение из многих тысяч, стоимость составляет 500 - 700 млн. долларов, из них клиника - 80%. Новые науки: биоинформатика, фармакогеномика существенно сокращают и сроки (6 - 7 лет), и стоимость (300 - 400 млн. долларов) создания новых лекарств. Новые науки уменьшают затраты на первую стадию разработки. Ничего не надо синтезировать, все, что нужно современной фармакологии, уже есть в базах данных, это более 6 млн. низкомолекулярных соединений. Мы должны найти мишень, рецептор, оценить, как то или иное соединение из базы данных взаимодействует с рецептором. Затем высокоэффективный скрининг, используемый во всем мире в фармакологических целях, который позволяет одновременно анализировать несколько тысяч различных соединений в параллельном режиме. Найденные таким образом базовые структуры уже на следующей стадии путем их химической модификации превращаются в конечное лекарство, таким образом в триаде - геномика-протеомика-биоинформатика - по сути дела заложены все новые подходы к созданию принципиально новой медицины будущего: новые лекарства, новые методы диагностики.