

НОВЕЙШИЕ ДОСТИЖЕНИЯ МОЛЕКУЛЯРНОЙ БИОЛОГИИ И ГЕНЕТИКИ В БОРЬБЕ С СЕРДЕЧНО-СОСУДИСТЫМИ ПАТОЛОГИЯМИ

ГЕНЕРАЛЬНЫЙ ДИРЕКТОР
ФГБУ «СЕВЕРО-
ЗАПАДНЫЙ ФЕДЕРАЛЬНЫЙ
МЕДИЦИНСКИЙ
ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
ИМЕНИ В.А. АЛМАЗОВА»
МИНЗДРАВА РОССИИ
Евгений Владимирович
Шляхто



Развитие методологии молекулярной медицины и генетики позволило сегодня иначе взглянуть на происхождение целого ряда заболеваний, которые ранее считались идиопатическими. Новые методы диагностики и лечения стали применяться при кардиомиопатии, врожденных пороках сердца, сосудистых аномалиях, некоторых формах нарушений ритма и синдрома внезапной смерти. Что же касается всей действующей системы диагностики и лечения, то она сегодня носит эмпирический характер и не обеспечивает ни раннего выявления патологии в группах риска (например, у близких родственников), ни надежных методов профилактики. Внедрение молекулярно-генетических технологий является важнейшим элементом в развитии так называемой персонализированной медицины, позволяющей проводить раннюю диагностику и профилактику осложнений и основанной на точном установлении молекулярно-генетических поломок. Этот абсолютно новый для медицины подход создает условия не только для существенного снижения смертности, но и для улучшения отдаленных результатов высокотехнологичных операций на сердце и сосудах.

Молекулярно-генетическая диагностика является дорогостоящей и трудозатратной областью лабораторной диагностики. Несмотря на существенный прогресс в медико-биологических технологиях, данная область пока не может стать высокоавтоматизированной, и ра-

бота над одним клиническим случаем занимает несколько недель и даже месяцев. В настоящее время в клиническую практику во многих странах активно внедряются методы секвенирования нового поколения, что, несомненно, в ближайшем будущем сделает генетическую диагностику более доступной, в том числе и в нашей стране.

Основой для создания и внедрения технологии секвенирования нового поколения явилась публикация полного генома человека в 2001 году, что положило начало новой эпохе в биологии, медицине и множестве смежных отраслей. Геном человека был прочтен при помощи сэнгеровского секвенирования, старого метода, отличающегося малой производительностью. Необходимо было приготовить тысячи плазмидных векторов, содержащих в своем составе те или иные части генома. Это был очень объемный и сложный проект, который длился 10 лет; до того как вышла первая публикация, работа в его рамках шла с участием множества научных институтов по всему миру. Ресурсоемкость, скорость и дороговизна сэнгеровского способа секвенирования не позволяли использовать данный метод в качестве обычного клинического секвенирования больших областей генома (совокупность последовательностей генов, содержащихся в наборе хромосом одного человека) или экзона (совокупность последовательностей частей генов, кодирующих белок). Однако благодаря своей точности и стандартизованности метод до сих пор актуален и активно используется для получения последовательностей отдельных коротких генов или их участков, а также для верификации результатов секвенирования нового поколения.

Ситуация в корне поменялась после разработки методов высокопроцессивного параллельного секвенирования (или секвенирования нового поколения, next generation sequencing, NGS). Данная технология заключается в том, что секвенируются одновременно миллиарды последовательностей. Однако по сравнению с методом Сэнгера (600–1000 пар оснований) длина индивидуального рида (одна секвенированная последовательность) при таком способе секвенирования невелика (35–300 пар

оснований). За счет огромного количества перекрытий отдельных коротких ридов друг с другом такой способ секвенирования позволяет прочесть полную последовательность генома человека менее чем за неделю. Но это требует большого количества вычислительных ресурсов. По сути, секвенирование нового поколения представляет собой не один, а целую группу методов, основанных на принципе параллельного секвенирования, но отличающихся особенностями секвенируемых последовательностей. Среди них можно выделить:

- секвенирование генома;
- секвенирование целевых участков (экзома или отдельных групп генов);
- секвенирование транскриптома РНК;
- Chip-Seq, или секвенирование участков ДНК, взаимодействующих с различными белками;
- бисульфидное секвенирование метилома (эпигенома).

Наиболее востребованным и экономически целесообразным в клинической практике является способ секвенирования целевых участков генома. В упрощенном варианте этот способ состоит из следующих этапов:

1. Выделение и характеристика ДНК пациента.
2. Подготовка библиотеки (совокупность целевых участков ДНК пациента, подготовленных соответствующим образом для массового параллельного секвенирования).
3. Секвенирование и обработка сырых данных, получение ридов.
4. Выравнивание (картирование) ридов на целевые участки референсного генома.
5. Фильтрация и аннотирование вариантов, верификация вариантов, интерпретация вариантов.

Каждый этап процесса имеет свои особенности и сложности, с которыми приходится сталкиваться техникам, лаборантам, врачам и генетикам.

ВЫДЕЛЕНИЕ И ХАРАКТЕРИСТИКА ДНК ПАЦИЕНТА

Методика секвенирования нового поколения предъявляет повышенные требования к качеству исходных нуклеиновых кислот. ДНК выделяется из крови пациента способами, основанными на переосаждении этиловым спиртом. Эти способы, в отличие от фенольной экстракции, безвредны для персонала и требуют меньше времени. Экстрагированную ДНК проверяют несколькими методами, среди которых классический электрофорез, спектрометрия и флюорометрия. Важной метрикой является концентрация: она должна быть определена с максимальной точностью, переизбыток или недостаток ДНК в пределах 5 нг может привести к серьезному нарушению всей процедуры пробоподготовки. Обычно протоколы требуют небольших концентраций ДНК, но в таких количествах ДНК быстро подвергается деградации, поэтому приведение ДНК к рабочему объему осуществляется непосредственно перед началом процедуры подготовки библиотеки.

ПОДГОТОВКА БИБЛИОТЕКИ

Первым этапом подготовки библиотеки с гибридным обогащением является короткое фрагментирование тотальной ДНК пациента. В зависимости от метода пробоподготовки длина может варьироваться (от 30 до 600 пар оснований). Методы фрагментирования могут быть физическими (например, ультразвук) или ферментативными (при помощи ферментов, разрезающих ДНК по определенным или случайным сайтам). Эффективность фрагментирования оценивают при помощи микрофлюидного электрофореза. После этого этапа производят отбор необходимых целевых последовательностей ДНК пациента: для этого смешивают фрагментированную ДНК с раствором, содержащим зонды. Зонды представляют собой искусственно синтезированные фрагменты ДНК, содержащие последовательности, комплементарные тем участкам целевой ДНК пациента, которые необходимо секвенировать. Вслед за этим следует этап гибридизации, при котором происходит комплементарное связывание одноцепочечной ДНК-последовательности пациента и комплементарной ей последовательности зонда. На следующем этапе производят очистку фрагментов ДНК пациента, гибридованных с зондами, при помощи магнитных частиц. Этот процесс возможен благодаря тому, что один из нуклеотидов, содержащийся в зонде, модифицирован таким образом, что содержит в своем составе белок биотин. Гибридизационную смесь смешивают с магнитными частицами, которые содержат на своей поверхности белок стрептавидин, способный связаться комплементарно с биотином. После этой операции магнитные частицы со стрептавидином, прочно связанные с биотином зондов, которые, в свою очередь, гибридованы с целевыми фрагментами ДНК пациента, несколько раз отмывают в буферных растворах, чтобы избавиться от ДНК без последовательности интереса, не связанной с зондами. Дальнейший этап – амплификационное обогащение, и проводится оно при помощи ПЦР-амплификации. Этот этап необходим, чтобы библиотека содержала достаточное для секвенса количество последовательностей. Далее библиотеку очищают от компонентов ПЦР-смеси и остатков зондов при помощи связывающих ДНК магнитных частиц и оценивают эффективность пробоподготовки и качественный состав библиотек. Для этого используют флюорометрию, микрофлюидный электрофорез и ПЦР в реальном времени. Особенностью массового параллельного секвенирования является то, что на этапе подготовки библиотек от разных пациентов в каждой целевой последовательности пациента добавляется последовательность специфического бар-кода.

СЕКВИРОВАНИЕ И ОБРАБОТКА СЫРЫХ ДАННЫХ, ПОЛУЧЕНИЕ РИДОВ

Наиболее распространенной платформой секвенирования нового поколения является платформа Illumina. Фирма Illumina выпускает несколько приборов, отличающихся количеством одновременно секвенируемых пос-



1



А



Б



В



Г

СЕКВЕНАТОР ILLUMINA MISEQ (ИЗОБРАЖЕНИЯ С САЙТА ILLUMINA.COM): А – ОБЩИЙ ВИД ПРИБОРА; Б – НАБОР РЕАКТИВОВ ДЛЯ ОДНОГО ЗАПУСКА ПРИБОРА; В – УСТАНОВКА В ПРИБОР ПРОТОЧНОЙ ЯЧЕЙКИ И КАРТРИДЖА С РЕАКТИВАМИ; Г – ИНТЕРФЕЙС ПРИБОРА ВО ВРЕМЯ СЕКВЕНИРОВАНИЯ

ледовательностей и объемами информации, получаемой на выходе. Существуют приборы (HiSeq), способные за один запуск прочитать одновременно несколько человеческих геномов с огромным покрытием. Запуск такого прибора занимает несколько недель и стоит 20–30 тыс. долларов. Альтернативой является прибор NextSeq, способный к прочтению одного генома с небольшим покрытием или экзона с большим. Наиболее удобен для клинического использования младший представитель семейства Illumina – прибор MiSeq (рис. 1). Он используется для прочтения генома или экзона человека, но его производительности вполне хватает для получения сиквенса с целевым обогащением. Для того чтобы получить объем информации, достаточный для выявления вариаций, библиотеки смешивают и секвенируют одновременно за одну процедуру запуска прибора, после чего биоинформатическими методами отделяют риды одного пациента от ридов другого, используя известные последовательности бар-кодов. В результате одного запуска на жестком диске прибора генерируется набор первично обработанных данных в виде пары файлов (прямое и обратное чтение) с расширением .fastq (стандартный файл .fasta содержит строку информации и строку последовательности, а файл .fastq дополнительно включает информацию о качестве чтения для каждого нуклеотида).

ВЫРАВНИВАНИЕ (КАРТИРОВАНИЕ) РИДОВ НА ЦЕЛЕВЫЕ УЧАСТКИ РЕФЕРЕНСНОГО ГЕНОМА, ДЕТЕКТИРОВАНИЕ ВАРИАНТОВ

Выравниванием называют вычислительный (программный, компьютерный) процесс соотнесения или картирования каждого короткого рида по отношению к соответствующей позиции референсного генома (например, текущей версии собранного и аннотированного генома человека). Выравнивание представляет собой первый шаг в определении и разделении типов вариаций, присутствующих в изучаемом геноме. Оно осуществляется при помощи специализированного программного обеспечения. Можно выделить программы BWA и Bowtie2, которые принимают на входе референсный геном и файлы, содержащие короткие риды (.fastq), а на выходе генерируют файл .bam (содержит полные последовательности ридов, к которым добавляется хромосомная координата по соответствующей сборке) или .bed (содержит имена коротких ридов, к которым добавляется хромосомная координата по соответствующей сборке). Последний из них используется для расчета различных



метрик качества проведенного запуска, а файл .bam – на дальнейших этапах получения информации. Для облегчения выравнивания используемая программа должна быть толерантна к небольшим различиям коротких ридов с референсным геномом. В зависимости от размера выравниваемой области генома и числа генерированных ридов выравнивание требует разных вычислительных ресурсов.

Количество выравненных или картированных ридов на каждый нуклеотид референсного генома или целевой области генома называется покрытием. Данный параметр является одним из основных показателей качества проведенного секвенирования и определяет информативность и достоверность полученных данных. После этапа выравнивания следует этап определения вариантов. Для него также существуют специальные программные пакеты. На входе их принимают файл .bam и референсный геном, а на выходе генерируется файл .vcf, представляющий таблицу с координатой и характеристиками обнаруженных полиморфизмов. Чем больше покрытие в данной позиции, тем больше уверенность в том, что детектированный здесь генетический вариант действительно представлен в исследуемом геноме. В настоящее время алгоритмы определения однобуквенных вариантов (SNP, снип, SNV) при помощи технологии секвенирования нового поколения достаточно надежны. Однако другие типы вариаций – малые вставки (1–100 бп) или делеции-инделлы (вырезание от 1 до 100 нуклеотидов) – более сложны в детекции посредством данного метода, так как требуют обратной связи программы, определяющей варианты, с программой, осуществляющей выравнивание. Однако масштабные вариации генома по-прежнему подразумевают разработку более совершенных алгоритмов или модернизацию технологии секвенирования.

ФИЛЬТРАЦИЯ И АННОТИРОВАНИЕ ВАРИАНТОВ, ВЕРИФИКАЦИЯ ВАРИАНТОВ, ИНТЕРПРЕТАЦИЯ ВАРИАНТОВ

Файл .vcf содержит информацию о хромосомной позиции варианта, текстовое обозначение референсного и замененного нуклеотида, а также техническую информацию о качестве детектирования в данной позиции. Чтобы соотнести эту информацию с генами, кодирующими последовательностями или регуляторными областями, используются аннотирующие программы. Эти программы на входе принимают файл .vcf, а на выходе генерируют таблицу (в стандартном формате Excel) с подробной и актуальной информацией о генетическом варианте. Информация включает:

- частоту обнаруженного генетического варианта в популяции;
- название, аббревиатуру и номер гена (по соответствующей системе);
- влияние на кодирующую последовательность (ведет ли вариант к замене одной аминокислоты на другую);

– наличие публикаций о клинически значимых состояниях в связи с данным генетическим вариантом.

Результатом работы аннотирующей программы является документ, предоставляемый врачу – клиническому генетику группой секвенирования для дальнейшей работы.

Упрощение и масштабирование технологии секвенирования привело к резкому увеличению количества генетических вариантов, обнаруживаемых у индивидуума. Это повлекло сложности в трактовке полученных генетических результатов и потребовало новой системы оценки патогенного значения каждого выявленного генетического варианта. Прежде, в эру секвенирования по Сэнгеру, патологическим считался генетический вариант, который отсутствовал как минимум у 200 здоровых индивидуумов (400 наборов хромосом). При этом все новые генетические варианты, которые приводили к замене аминокислоты и не обнаруживались в подобной группе контроля, автоматически считались патологическими. Благодаря современным технологиям массивного параллельного секвенирования в геноме человека выявлено множество вариантов, встречающихся в популяции с частотой 1:1000, 1:10 000, 1:20 000 и т.д. В связи с этим возникли вопросы: могут ли данные варианты иметь патогенное значение? как трактовать наличие подобных замен? является ли факт обнаружения подобного генетического варианта ранее у пациента со сходной патологией достаточным показателем его патогенности или же является случайным совпадением? является ли обнаружение подобного редкого генетического варианта прогностически значимым или его фенотипическая реализация зависит от множества комбинаций с другими генетическими вариантами? что является определяющим в реализации патогенного эффекта таких редких популяционных генетических вариантов – факторы внешней среды или генетический фон пациента? Все эти вопросы осложняют работу врача – клинического генетика в трактовке вновь обнаруженных генетических вариантов и требуют разработки новых критериев оценки клинической значимости патогенных вариантов. В настоящее время предложено несколько классификаций для оценки клинической значимости обнаруженных вариантов. Одна из наиболее популярных классификаций предложена частной компанией, обеспечивающей услуги по массивному параллельному секвенированию в клинических целях. Такая классификация учитывает частоту обнаружения данного генетического варианта в популяции, информацию о структурном моделировании обнаруженной замены и предсказание функционального эффекта мутации, а также ранее описанные случаи клинически значимой патологии у пациентов с такой генетической заменой. В соответствии с предложенной классификацией все обнаруживаемые варианты можно разделить:

- на патологические;
- вероятно патологические;
- с неопределенной значимостью;
- вероятно доброкачественные;
- доброкачественные.



Именно эта классификация сейчас наиболее часто применяется в оценке вариантов при сердечно-сосудистой патологии, в частности при трактовке генетических причин кардиомиопатий и нарушений сердечного ритма.

Подводя итоги первых лет работы с использованием методики массивного параллельного секвенирования, еще раз подчеркнем, что современные технологии молекулярной биологии и генетики дают мощные возможности для изучения патогенеза многих заболева-

ний и позволяют заглянуть внутрь молекулярных событий, лежащих в их основе. Однако никакие даже самые современные и информативные технологии не смогут заменить клинический опыт, профессионализм и глубокие знания в области медицины и биологии. Только сочетание этих компонентов дает возможность реализовать в клинической медицине персонализированный подход, который, кроме всего прочего, должен учитывать уникальность каждой клинической ситуации.

СТАТЬЯ ПОДГОТОВЛЕНА ПРИ УЧАСТИИ
ДИРЕКТОРА ИНСТИТУТА МОЛЕКУЛЯРНОЙ БИОЛОГИИ И ГЕНЕТИКИ
ФГБУ «СЗФМИЦ ИМЕНИ В.А. АЛМАЗОВА» МИНЗДРАВА РОССИИ
А.А. Костаревой
СТАРШЕГО НАУЧНОГО СОТРУДНИКА ИНСТИТУТА МОЛЕКУЛЯРНОЙ
БИОЛОГИИ И ГЕНЕТИКИ ФГБУ «СЗФМИЦ ИМЕНИ В.А. АЛМАЗОВА»
МИНЗДРАВА РОССИИ
А.М. Киселёва