

Возможности и ограничения высокопроизводительного секвенирования в диагностике моногенных заболеваний

Л.И. Шагам, В.Ю. Воинова

ОСП Научно-исследовательский клинический институт педиатрии им. академика Ю.Е. Вельтищева ГБОУ ВПО РНИМУ им. Н.И. Пирогова Минздрава РФ, Москва

High-performance sequencing in the diagnosis of monogenic diseases: Possibilities and limitations

L.I. Shagam, V.Yu. Voinova

Academician Yu.E. Veltishchev Research Clinical Institute of Pediatrics, N.I. Pirogov Russian National Research Medical University, Ministry of Health of the Russian Federation, Moscow

Статья посвящена актуальным вопросам применения высокопроизводительного секвенирования в диагностике моногенных заболеваний. *De facto* данный метод уже сегодня широко используется в клинической практике. Бурное развитие высокопроизводительного секвенирования требует от врачей-генетиков и педиатров, специализирующихся на наследственных болезнях, понимания его возможностей и ограничений, выбора стратегии диагностики для каждого конкретного пациента и грамотной интерпретации полученного результата.

Ключевые слова: дети, моногенные заболевания, высокопроизводительное секвенирование, NGS, генетическая диагностика.

The paper deals with the topical issues of using high-performance sequencing to diagnose monogenic diseases. The *de facto* method is widely used in clinical practice. The rapid development of high-performance sequencing requires that medical geneticists and pediatricians specializing in hereditary diseases should understand its possibilities and limitations, choose a strategy for diagnosis in each specific patient, and competently interpret the obtained results.

Key words: children, monogenic diseases, high-performance sequencing, NGS, genetic diagnosis.

Высокопроизводительное секвенирование (англ. next generation sequencing, NGS – секвенирование следующего/нового поколения) – совокупность методов высокопроизводительного считывания нуклеотидной последовательности ДНК и РНК, широко используемых в научных исследованиях начиная с середины нулевых годов XXI столетия [1]. Метод активно применяется в фундаментальных и прикладных исследованиях, а также в диагностике моногенных заболеваний – обширной группы наследственных болезней, вызываемых мутациями в одном гене. На данный момент описано 4645 моногенных наследственных болезней и признаков, молекулярная основа которых уже определена [2]. Многие из этих заболеваний генетически гетерогенны, т.е. могут возникать вследствие мутаций одного из многих генов (до нескольких десятков). Так, синдром Барде–Бидля, характеризующийся специфическим легко распознаваемым симптомокомплексом (пигментный ретинит, ожирение, полидактилия, аномалии почек, гипогонадизм и психические нарушения), в настоящее время связывают с 19 различными генами (*BBS1*,

BBS2, ... *BBS19*), локализованными в разных хромосомах.

Технические возможности высокопроизводительного секвенирования позволяют считывать последовательность одновременно большого числа генов и выявлять патогенные мутации, ответственные за возникновение генетически гетерогенных заболеваний [3]. Кроме того, данная технология способствует открытию генетической этиологии недифференцированных форм патологии. Так, молекулярная основа многих моногенных синдромов была открыта благодаря секвенированию экзонов у пациентов с идентичными клиническими признаками. В качестве примера можно привести открытие мутаций гена *SMARCA2* как причины синдрома Николаидеса–Барайцера [4]. Наконец, у практически здоровых индивидуумов благодаря высокопроизводительному секвенированию становится возможным выяснение генетической предрасположенности к ряду заболеваний [5].

Генетическое обследование пациентов с подозрением на моногенные болезни применяется для уточнения диагноза, прогнозирования тяжести состояния пациента и динамики течения его заболевания, а также медико-генетического консультирования семей, имеющих больного ребенка, с целью оценки риска для следующих беременностей и выработки рекомендаций по пренатальной диагностике. Генетическое тестирование для ряда заболеваний и их групп (например, таких как нейрональный цероидный липофуциноз или X-сцепленные формы умственной

© Шагам Л.И., Воинова В.Ю., 2016

Ros Vestn Perinatol Pediat 2016; 2:105–109

Адрес для корреспонденции: Шагам Лев Иосифович – н.с. НИЛ общей патологии НИКИ педиатрии им. академика Ю.Е. Вельтищева
Воинова Виктория Юрьевна – д.м.н., вед.н.сотр. отдела психоневрологии и наследственных заболеваний с нарушением психики того же учреждения
125412 Москва, ул. Талдомская, д. 2

отсталости) уже является основным методом диагностики [6]. В то же время применение подавляющего большинства высокопроизводительных секвенаторов в целях диагностики не разрешено на территории Российской Федерации, что связано с отсутствием регистрационных удостоверений для осуществления диагностической деятельности. Исключением является лишь пиросеквенатор GS Junior компании Roche diagnostics (Швейцария), однако его возможности ограничены (см. ниже). В связи с этим подавляющее большинство подобных исследований на территории нашей страны в настоящее время проводится в статусе научной работы.

Стратегия планирования обследования пациента с помощью NGS

Для эффективной диагностики генетически детерминированного состояния важно придерживаться принятой в профессиональном сообществе последовательности шагов по выявлению определенного заболевания — диагностического алгоритма, важнейшее место в котором занимают как клиничко-генеалогическое, так и генетическое исследования. Алгоритм генетического обследования зависит от мутаций, характерных для конкретного заболевания.

Например, для муковисцидоза характерны горячие точки мутаций, изменения в которых выявляются у более чем 75% пациентов в российской популяции [7]. В связи с этим у больных с таким направляющим диагнозом целесообразно, в первую очередь, провести детекцию мутаций в горячих точках и только в случае отсутствия последних осуществить полное секвенирование гена. Другим примером может служить энцефаломиопатия Ли (исключительно генетически гетерогенное митохондриальное заболевание), молекулярно-генетическую диагностику которого начинают с анализа частых мутаций гена *SURF1*. При их отсутствии проводят секвенирование данного гена, секвенирование митохондриальной ДНК, секвенирование ядерного экзома [8]. Возможности высокопроизводительного секвенирования, как правило, ограничены при выявлении делеций и вставок генетического материала, превышающих по размеру 10–50 пар нуклеотидов (п.н.), в частности делеций нескольких экзонов. В связи с этим для заболеваний, которым свойственны подобные мутации, рекомендуется проводить их выявление методом MLPA (*англ.* мультиплексная лигаза-зависимая амплификация зонда) или ПЦР (полимеразная цепная реакция) в реальном времени. Для других заболеваний, таких как туберозный склероз или синдром Альпорта, необходимо сразу проводить полное секвенирование генов, поскольку горячие точки мутаций для них или не характерны, или очень редки, а делеции и вставки размером более 10 п.н. встречаются лишь в небольшом проценте случаев.

Если число генов, связанных с наследственным

заболеванием, превышает один, высокопроизводительное секвенирование, как правило, является наиболее экономически обоснованным методом исследования с разрешением до одного нуклеотида (в сравнении с секвенированием по Сэнгеру). Применяются следующие альтернативные стратегии NGS: секвенирование генома, секвенирование экзома и генные панели, специально разработанные под заболевание или группу болезней.

Полногеномное секвенирование является наиболее подробным из возможных вариантов исследования ДНК, однако ввиду высокой стоимости и трудоемкости интерпретации используется в основном в научных исследованиях. В то же время лишь 1% генома составляют кодирующие участки, в которых, согласно современным оценкам, локализуется подавляющее большинство патогенных мутаций — около 85% [9]. Что же касается уже описанных на сегодняшний день в научной литературе точечных мутаций, связанных с моногенными заболеваниями, 97% из них располагаются именно в кодирующих экзонах генов или смежных участках интронов протяженностью 5–10 п.н. При этом патогенные мутации в 5'-и 3'-нетранслируемых областях, промоторах генов, интронах (исключая экзон-интронную границу) и межгенных областях встречаются весьма редко [10, 11]. В связи с этим секвенирование экзома или отдельных групп генов в клинической практике более целесообразно.

Секвенирование экзома (полноэкзомное секвенирование) — стандартизированное исследование, в результате которого происходит считывание кодирующей последовательности всех генов человека (иногда помимо кодирующих экзонов и смежных участков интронов считываются 5'-и 3'-нетранслируемые области, промоторы генов). Затем, в зависимости от направляющего диагноза, проводится биоинформатический анализ того или иного набора генов. Отметим, что вопрос о выдаче в заключении информации по мутациям генов, не связанных с предполагаемым заболеванием, является дискуссионным. Согласно популярным рекомендациям [12] следует сообщать пациенту только заведомо патогенные мутации с высокой пенетрантностью в 56 генах, связанных с заболеваниями, которые можно предотвратить. Однако с данным мнением можно не соглашаться, поскольку информация о генетической предрасположенности к заболеваниям, которые не актуальны на данный момент для пациента, может в дальнейшем повлиять на процесс их профилактики, выбора методов терапии и др.

Генная панель разрабатывается под наследственное заболевание или группу болезней. Она позволяет секвенировать определенные гены, при этом не происходит чтение остальных участков экзома. Плюсами данного подхода, как правило, является лучшее покрытие исследуемых генов и более низкая цена анализа в сравнении с секвенированием экзома

при том же качестве. Минусы – необходимость разработки и апробации генной панели под каждое заболевание или группу. Частным случаем генной панели является «клинический экзом» – данное исследование подразумевает секвенирование всех генов, связанных с известными моногенными болезнями.

Некоторые технологические особенности высокопроизводительного секвенирования

Принцип высокопроизводительного секвенирования предполагает многократное прочтение (*покрытие*) каждого секвенируемого нуклеотида. Средним покрытием называют среднее число прочтений каждого нуклеотида из секвенируемой области. Минимальным покрытием, делающим возможным получение достоверной информации о последовательности ДНК, считается не менее 10–20-кратного (10–20X). Это объясняется законами теории вероятности. Так, предположим, что покрытие участка ДНК, несущего мутацию в гетерозиготном состоянии (аллели АВ), составляет 10X. В предположении равной вероятности чтения каждого из аллелей вероятность того, что ни разу не будет считан аллель А, составляет $0,5^{10} = 0,001$. Аналогично, вероятность того, что ни разу не будет прочитан аллель В, составляет 0,001. Итого, суммарная вероятность ложного результата составляет 0,002, что является приемлемым. При более низком покрытии такая вероятность значительно увеличивается.

Однако задача по 10–20-кратному покрытию всех нуклеотидов из целевых областей является весьма сложной как с технической, так и с финансовой точки зрения. Так, всегда участки из секвенируемых областей характеризуются неравномерным покрытием, и при среднем покрытии 100X, 200X и т.д. находятся области, покрываемые менее 10 раз. Для максимизации процента изучаемых участков генома, покрываемых не менее 10X, 15X или 20X, среднее покрытие увеличивают до сотен раз в пересчете на нуклеотид. Такой процент называют *полнотой покрытия*. При этом увеличение полноты покрытия для экзома является задачей весьма дорогостоящей. Так, исследование экзома с полнотой покрытия 98% может стоить до 5 000 долларов США, при том что анализ экзома со стандартной полнотой покрытия 90–95% стоит от 400 долларов США. Исследование клинического экзома позволяет снизить стоимость анализа в сравнении с полноэкзомным исследованием, а также более безболезненно увеличивать среднее покрытие и, как следствие, полноту покрытия генов. Однако если ген, связанный с заболеванием пробанда, еще не идентифицирован или идентифицирован недавно и не входит в версию клинического экзома, применяемую лабораторией, то с помощью панели «клинический экзом» мутация, лежащая в основе патологического фенотипа, не будет обнаружена.

Секвенирование полного, клинического экзома

или генной панели может быть проведено при помощи различных технологий. На сегодняшний день наиболее популярными методиками высокопроизводительного секвенирования являются Illumina/Solexa (Illumina, США) и Ion Torrent (Thermo Fisher Scientific, США). Кроме того, анализы иногда проводят на платформе 454 (пиросеквенаторы компании Roche Diagnostics, Швейцария), разработка которой официально прекращена в 2013 г., а поддержка будет осуществляться лишь до конца 2016 г. Отметим, что на последней платформе может быть выполнено только исследование генной панели, но не всего экзома.

С точки зрения эффективности детекции мутаций, технологии Illumina, Ion Torrent и 454 обладают схожими характеристиками, позволяя выявлять однонуклеотидные замены и небольшие делеции/вставки, при этом выявление делеций/вставок размером от 10–50 п.н. затруднено. Технология Illumina является несколько более точной в детекции небольших вставок и делеций генетического материала. Ряд биоинформатических методов позволяет выявлять также вариации копийности размером от нескольких экзонов, однако эти методики зачастую характеризуются большим количеством ложноположительных результатов и в настоящее время применяются в диагностических задачах весьма редко.

Анализ результатов высокопроизводительного секвенирования пока автоматизирован не полностью и требует работы высококвалифицированного биоинформатика. Важнейшей в данном процессе является задача по маркировке артефактов, выявляемых в процессе высокопроизводительного секвенирования. Для полного исключения артефактов рекомендуется альтернативным способом верифицировать все патогенные мутации, найденные методом высокопроизводительного секвенирования. Альтернативными способами верификации служат такие методы, как: секвенирование по Сэнгеру, ПЦР в реальном времени, MLPA и т.п.

Клиническая интерпретация результата секвенирования

Ввиду большого числа исследуемых при NGS генов неудивительно, что может быть идентифицировано множество вариантов последовательности ДНК. Полногеномное секвенирование обычно генерирует 3–4 млн генетических вариантов, которые отличаются от человеческой эталонной последовательности, в то время как полноэкзомное секвенирование генерирует 15–20 тыс. вариантов в пределах кодирующих участков [13]. Подавляющее большинство из этих вариантов являются непатогенными полиморфизмами, распространенными в популяции человека, и только 1–2 варианта фактически связаны с фенотипом, из-за которого и было предпринято тестирование.

Для клинической интерпретации результата NGS

необходима его корректная аннотация (описание свойств генетических вариантов), которая обычно включает для каждого изменения ДНК:

- ген, в котором зафиксирован данный вариант;
- заболевания, ассоциированные с данным геном с номерами по каталогу OMIM;
- характеристику выявленных изменений на уровне генома и соответствующих им предполагаемых изменений на уровне РНК и белка (молчащая замена, миссенс-, нонсенс-замена, мутация сдвига рамки считывания, мутация сайта сплайсинга и т.п.);
- популяционную частоту данного генетического варианта;
- оценку патогенности мутации согласно базам данных HGMD [14] и ClinVar [15], а также специализированным базам данных различных заболеваний (например, LOVD [16] и многие другие, специфические для каждой группы нозологий)[17]. Такие базы данных построены на информации из научных работ, а также неопубликованных данных научных и клинических лабораторий по всему миру;
- информацию о консервативности региона, содержащего мутацию, у разных видов;
- оценку эффекта мутации, полученную путем использования моделей предсказания *in silico*.

На основании вышеупомянутой информации принимается решение о клинической значимости генетического варианта. Следует отметить, что на сегодняшний день основная ценность большинства баз данных генетических вариантов заключается в возможности отбора потенциально патогенных изменений ДНК и ссылок на научную литературу, в которой был описан каждый из выявленных вариантов. Грубой ошибкой является классификация генетического варианта как патогенного на основании, например, только соответствующей записи в базе данных HGMD или LOVD, без детального ознакомления с первоисточниками. Кроме того, важную роль в процессе клинической интерпретации играет наличие полной информации о фенотипе больного.

И все же в значительной доле случаев имеющиеся, в том числе клинические, данные не позволяют отнести выявленный генетический вариант к категории однозначно патогенных или непатогенных. Тем не менее обычно удается отобрать потенциально патогенные варианты, которые могут потребовать более внимательного рассмотрения. При этом большое значение имеет анализ сегрегации предполагаемого патогенного варианта в семье пробанда. Так, при аутосомно-доминантных заболеваниях определяют, передается ли данный вариант в родословной вместе с патологическим фенотипом, а при аутосомно-рецессивной патологии — получил ли пробанд по одному рецессивному аллелю от каждого из родителей (транс-положение мутаций, которое подтверждает клиническую значимость) либо унаследовал обе мутации в аллеле, полученном от одного родителя

(цис-положение, отвергающее клиническую значимость). Косегрегация генетического варианта с заболеванием свидетельствует в пользу патогенности данного варианта, хотя и не доказывает ее; в то же время отсутствие сегрегации является доказательством непатогенности варианта. Помимо этого, весьма значимым доказательством патогенности генетического варианта у пробанда является его полное отсутствие у обоих родителей, что указывает на возникновение данной мутации *de novo* (при этом нужно исключить ложное отцовство) [18].

Поскольку вышеупомянутые подходы зачастую являются решающими в клинической интерпретации, врачу, направляющему пациента на высокопроизводительное секвенирование, желательно заранее позаботиться о доступности для исследования образцов крови родителей. Кроме того, в случае полноэкзомных анализов по ряду нозологий показано, что изучение семейного трио повышает диагностический выход в среднем с 22 до 31%, т.е. в 1,4 раза [19].

В настоящее время используется следующая классификация генетических вариантов по признаку их патогенности [18]:

- патогенные;
- вероятно патогенные;
- варианты с неясной клинической значимостью;
- вероятно непатогенные;
- непатогенные.

Общепринятой стратегии классификации генетических вариантов по вышеперечисленным группам предполагаемой патогенности пока нет, каждая лаборатория, как правило, использует свои собственные критерии. Один из вариантов критериев содержится в рекомендациях Американской Ассоциации Медицинских Генетиков от 2015 г. [18].

Выявление в генетическом материале пациента патогенных или вероятно патогенных мутаций в генах, связанных с изучаемым заболеванием, в соответствующем аллельном состоянии и количестве свидетельствует в пользу направляющего диагноза. При этом важно понимать, что наличие у пациента генетического варианта с неясной клинической значимостью в гене, соответствующем предполагаемому диагнозу, вовсе не означает, что обнаруженный вариант как-либо связан с имеющимся у пробанда заболеванием. Кроме того, отрицательный результат не означает, что направляющий диагноз полностью исключен: нужно иметь в виду ограничения метода и чувствительность выполняемого теста.

Ограничения применения NGS в диагностике наследственной патологии

Практика показывает, что сегодня даже очень опытные врачи-генетики, понимая достоинства и незаменимость высокопроизводительного секвенирования, связанные с возможностью детекции

мутаций с разрешением до одного нуклеотида на обширных областях генома, не всегда трезво оценивают слепые зоны метода и знают, на какие генетические обследования следует направлять пациента с отрицательным результатом секвенирования экзома или геномной панели. Ограничения же заключаются в том, что в ходе высокопроизводительного секвенирования не могут быть исключены следующие виды мутаций:

- находящиеся в областях, покрываемых менее 10 раз;
- делеции и вставки генетического материала размером более 10–50 п.н. в гетерозиготном состоянии, например, делеция экзона на аутосоме;
- сбалансированные транслокации, инверсии;
- мутации в участках, не покрываемых геномной панелью или экзомом, например, в регуляторных областях;

- варианты, находящиеся в областях ДНК «низкой сложности», в том числе изменения числа триплетных повторов.

Поскольку технология NGS и связанный с ней биоинформатический анализ продолжают развиваться, многие из этих ограничений, возможно, будут сведены к минимуму в не слишком отдаленном будущем.

Высокопроизводительное секвенирование является популярнейшим, мощнейшим и незаменимым методом генетического тестирования. Однако на сегодняшний день его использование в диагностике по всему миру связано с рядом формальных и методических ограничений. В сложившейся ситуации крайне важно для врача, направляющего пациента на генетическое исследование, понимать как возможности, так и ограничения высокопроизводительного секвенирования.

ЛИТЕРАТУРА (REFERENCES)

1. Liu L., Li Y., Li S. et al. Comparison of Next-Generation Sequencing Systems. *J Biomed Biotechnol* 2012; 2012: 1–11.
2. <http://www.omim.org/statistics/entry>
3. Korf B.R., Rehm H.L. New approaches to molecular diagnosis. *JAMA* 2013; 309: 1511–1521.
4. Van Houdt J.K.J., Nowakowska B.A., Sousa S.B. et al. Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nat Genet* 2012; 44: 445–449, S1.
5. Johnston J.J., Rubinstein W.S., Facio F.M. et al. Secondary variants in individuals undergoing exome sequencing: screening of 572 individuals identifies high-penetrance mutations in cancer-susceptibility genes. *Am J Hum Genet* 2012; 91: 97–108.
6. Riva D., Bulgheroni S., Pantaleoni C. Mental retardation. D.Riva, S.Bulgheroni, C. Pantaleoni (eds.). Montrouge: John Libbey Eurotext, 2007; 232.
7. Петрова Н.В. Молекулярно-генетические и клинико-генотипические особенности муковисцидоза в российских популяциях. Автореф. дисс. ... докт. биол. наук. М 2009; 21. (Petrova N.V. Molecular genetic and clinical characteristics of cystic fibrosis in Russian population. Avtoref. diss. ... doct.biol.n. Moscow 2009; 21.)
8. Николаева Е.А., Новиков П.В. Проблема диагностики и дифференциальной диагностики митохондриальных заболеваний у детей. *Педиатрия* 2014; 6: 75–83. (Nikolaeva E.A., Novikov P.V. Problem of diagnostics and differential diagnosis of mitochondrial diseases at children. *Pediatriya* 2014; 6: 75–83.)
9. Majewski J., Schwartzenuber J., Lalonde E. et al. What can exome sequencing do for you? *J Med Genet* 2011; 48: 580–589.
10. <https://portal.biobase-international.com/hgmd/pro/stats.php>
11. Ma M., Ru Y., Chuang L.-S. et al. Disease-associated variants in different categories of disease located in distinct regulatory elements. *BMC Genomics* 2015; 16: Suppl 8: S3.
12. Green R.C., Berg J.S., Grody W.W. et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 2013; 15: 565–574.
13. Bamshad M.J., Ng S.B., Bigham A.W. et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011; 12: 745–755.
14. <http://www.hgmd.cf.ac.uk/ac/index.php>
15. <http://www.ncbi.nlm.nih.gov/clinvar>
16. <http://www.lovd.nl/3.0/home>
17. Facio F.M., Lee K., O'Daniel J.M. A genetic counselor's guide to using next-generation sequencing in clinical practice. *J Genet Couns* 2014; 23: 455–462.
18. Richards S., Aziz N., Bale S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; 17: 405–423.
19. Lee H., Deignan J.L., Dorrani N. et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 2014; 312: 1880–1887. doi:10.1001/jama.2014.14604.

Поступила 15.02.16