ГЕНЕТИКА, 2022, том 58, № 10, с. 1155–1163

ОБЗОРНЫЕ И ТЕОРЕТИЧЕСКИЕ СТАТЬИ

УДК 575.224.2

РОЛЬ СПЛАЙСИНГА В ПАТОГЕНЕЗЕ МОНОГЕННЫХ БОЛЕЗНЕЙ

© 2022 г. Н. А. Скрябин[1, *], Д. И. Жигалина[1], В. А. Степанов[1]

[1]Научно-исследовательский институт медицинской генетики, Томский национальный исследовательский медицинский центр Российской академии наук, Томск, 634050 Россия

*e-mail: nikolay.skryabin@medgenetics.ru

Поступила в редакцию 04.04.2022 г.
После доработки 23.05.2022 г.
Принята к публикации 31.05.2022 г.

Несмотря на развитие технологий экзомного и полногеномного секвенирования и их рутинное использование при диагностике наследственных заболеваний, эффективность выявления патогенных генетических вариантов для методов на основе анализа ДНК составляет менее 50%. Одной из основных причин может являться неэффективность данных подходов при поиске генетических вариантов, ответственных за нарушение сплайсинга пре-мРНК. В этом обзоре рассматриваются результаты работ по поиску аномалий сплайсинга при наследственных орфанных заболеваниях с помощью РНК-секвенирования и возможности клинического применения данного метода.

Ключевые слова: РНК, орфанные болезни, мутации, сплайсинг, экспрессия.

DOI: 10.31857/S0016675822100083

Отличительной особенностью редких (орфанных) болезней является небольшая частота встречаемости в популяции. Однако число самих нозологий чрезвычайно велико и с каждым годом увеличивается. В мире, по экспертным оценкам, известно около 6172 редких заболеваний [1], большинство из них (71.9%) имеют в своей основе генетическую природу [2]. В онлайн базе данных менделевских наследственных заболеваний человека описано 5918 фенотипов различных наследственных нозологий с известными молекулярными механизмами развития патологического состояния, и еще более 3000 фенотипов заболеваний без выявленных молекулярных механизмов [3].

На настоящий момент общее количество больных орфанными болезнями в мире составляет примерно 3.5–5.9% от общего населения Земли, что соответствует от 263 до 446 млн человек [2]. При этом таких данных по Российской Федерации на данный момент нет. На основе данных среднеевропейских и среднемировых показателей распространенности орфанных заболеваний, а также численности населения РФ (146.9 млн человек по состоянию на 2018 г.), можно предположить, что потенциальное количество пациентов с редкими заболеваниями в РФ может составлять от 5.1 до 8.6 млн человек.

Другой особенностью данных заболеваний является чрезвычайно высокая гетерогенность как в отношении систем и органов, которые они поражают, так и в степени клинического проявления. Все это в совокупности приводит к значительным сложностям диагностики орфанных болезней. Для пациента установление правильного диагноза в течение нескольких лет может быть сопряжено с прохождением множества диагностических тестов, а также посещением большого количества специалистов. В результате некоторые пациенты не доживают до постановки диагноза. С учетом гетерогенности наследственных орфанных болезней одним из условий в улучшении диагностики является расширение спектра используемых технологий, направленных на поиск известных наследственных мутаций и на картирование новых генов и генетических вариантов, связанных с развитием наследственных заболеваний.

Выявление этиологически значимых мутаций позволяет уточнить молекулярный диагноз у пациентов. Это необходимо для медико-генетического консультирования, поскольку позволяет оценить наследование патогенетически значимых мутаций, дает возможность выявлять носителей мутаций и семейные формы заболеваний.

Постановка точного молекулярного диагноза существенно улучшает возможность проведения профилактических мероприятий. Так, ранняя диагностика наследственных болезней обмена позволяет использовать заместительную терапию, что в свою очередь приводит к нормализации функций или снижению выраженности патологического процесса. Для некоторых наследственных заболеваний возможно внутриутробное лечение

(например, при некоторых ацидуриях, галактоземии). Развитие заболевания в настоящее время можно предотвратить путем коррекции (лечения) после рождения больного. Типичными примерами таких болезней могут быть галактоземия, фенилкетонурия, гипотиреоз и др.

Точная идентификация мутации при наследственных заболеваниях необходима для их пренатальной и преимплантационной диагностики. В результате пренатальной диагностики, в случае выявления мутаций у плода, проводится прерывание беременности, тем самым исключается рождение больных детей. Постановка диагноза и выявление носительства мутаций у родителей позволяет предложить пациентам процедуры преимплантационной генетической диагностики при планировании беременности с целью исключения патогенетически значимых наследственных дефектов, и, соответственно, снизить риск повторного рождения больных детей в семьях, что должно способствовать снижению груза наследственных заболеваний, затрат на лечение и реабилитацию пациентов.

ПОИСК МУТАЦИЙ В ДНК

Достижения в области секвенирования сделали возможным анализ всего экзома и генома, практически данные подходы стали рутинными инструментами врача-генетика. Эти достижения привели к значительному улучшению эффективности диагностики и увеличению числа идентифицируемых генов, лежащих в основе редких заболеваний. Одной из первых работ, в которой экзомное секвенирование было использовано для идентификации причинной мутации при наследственном заболевании является работа международной группы авторов под руководством американских исследователей из Медицинского института Говарда Хьюза [4]. С помощью экзомного секвенирования ими был установлен молекулярный диагноз у ребенка с подозрением на синдром Барттера. По результатам анализа была идентифицирована гомозиготная миссенс-мутация в гене *SLC26A3*, ответственном за развитие врожденной хлоридной диареи. Таким образом, был выставлен диагноз, отличный от направительного, что стало возможным за счет секвенирования всей кодирующей последовательности генома [4]. С тех пор экзомное секвенирование в клинической практике используется все чаще и значительно повысило эффективность молекулярной диагностики, сокращая “диагностическую одиссею” пациентов с моногенными заболеваниями.

В настоящее время несмотря на существенный прогресс в области развития технологий молекулярно-генетической диагностики орфанных заболеваний, остается много нерешенных проблем. Эффективность выявления патогенетически значимых мутаций с использованием передовых технологий, базирующихся на анализе ДНК, таких как экзомное и геномное секвенирование, составляет по разным оценкам от 30 до 50% [5, 6]. Так, группой американских ученых была проведена оценка диагностической ценности экзомного секвенирования у детей с моногенными заболеваниями. При анализе 40 клинических случаев генетические дефекты были выявлены у 12 (30%) пациентов, среди которых 47% мутаций ранее не упоминались в литературе. Кроме того, 36 пациентам был проведен анализ вторичных находок по отношению к основному диагнозу (“случайные” находки). В результате у трех пациентов (8%) были идентифицированы генетические варианты, приводящие к нарушениям, которые требуют медицинского вмешательства [5].

Схожая работа была проведена группой австралийских исследователей, в которой была оценена диагностическая ценность экзомного секвенирования для детей с наследственными болезнями, при этом молекулярный диагноз был поставлен в 52% случаев [6]. Кроме того, у 35% пациентов были выявлены диагнозы, отличающиеся от направительного, а у 26% была скорректирована тактика клинического ведения больного. Авторы также провели экономический анализ различных траекторий диагностики пациентов и установили, что экзомный анализ, выполненный при первичном обращении, мог привести к дополнительной экономии затрат в размере 9020 австралийских долларов по сравнению со стандартными подходами к диагностике моногенных болезней.

В одном из наиболее масштабных исследований по оценке эффективности экзомного секвенирования был проведен анализ 3040 пациентов. В результате общая диагностическая ценность экзомного секвенирования составила 28.8%. Стоит отметить, что при анализе только пробандов, диагностическая ценность составила 23.6%, а при анализе трех членов семьи – 31% [7]. Таким образом, анализ “трио” позволяет улучшить эффективность идентификации причинной мутации за счет выявления генетических вариантов *de novo*, что упрощает классификацию новых вариантов. Стоит отметить, что при этом стоимость исследования возрастает в три раза.

Использование полногеномного секвенирования, вопреки ожиданиям о значительном увеличении диагностической эффективности, существенно не улучшает ситуацию. В результате полногеномного секвенирования 103 пациентов из Канады с наследственными заболеваниями, молекулярный диагноз был выставлен у 41% пациентов [8]. Тем не менее использование такого подхода позволило идентифицировать у 18 пациентов мутации, расположенные в некодирующей последовательности ДНК. В метаанализе, проведенном исследователями из Оксфорда, было проведено сравнение

стоимости и эффективности экзомного и геномного секвенирования. Всего в работе были проанализированы данные 27 исследований с использованием экзомного секвенирования и три исследования секвенирования полного генома. В результате было показано, что в среднем эффективность секвенирования экзома составила 35%, а секвенирования генома – 49% [9].

В случаях, когда не удается идентифицировать патогенные генетические варианты, возможен повторный анализ данных экзомного или геномного секвенирования через определенное время. В некоторых случаях это позволяет выявить варианты, которые не были обнаружены при первом анализе. Повышение диагностической ценности в данном случае обусловлено несколькими факторами [10]:

• открытие новых генов/вариантов, ассоциированных с заболеваниями;

• изменение классификации ранее выявленных вариантов вследствие расширения баз данных, проведения функциональных исследований;

• улучшение референсных геномов;

• развитие биоинформатических алгоритмов поиска вариантов, в том числе с использованием методов машинного обучения;

• анализ новых типов вариантов;

• сбор более подробных сведений о пациенте и/или возрастные изменения клиники пациента.

Таким образом, возможно повышение диагностической ценности экзомного/геномного секвенирования примерно на 10–20% [10, 11]. Несмотря на то, что в некоторых случаях такой подход не позволяет идентифицировать новые патогенетически значимые варианты [12].

Эффективность секвенирования как экзома, так и генома имеет свои ограничения. В большей степени это обусловлено тем, что методики, основанные на анализе ДНК, не позволяют идентифицировать мутации, которые не влияют на аминокислотный состав белка. Большинство патогенетически значимых генетических вариантов, идентифицированных в настоящее время, относятся к миссенс- и нонсенс-мутациям (84%), поскольку для их поиска использовалось секвенирование ДНК. Кроме того, на эффективность секвенирования ДНК влияют проблемы, связанные с повторяющимися последовательностями, GC-богатыми регионами, неполным охватом зондами кодирующей последовательности и сложностями с выравниванием коротких последовательностей, что приводит к пропуску вариантов внутри регионов с плохим охватом.

Одним из механизмов, приводящих к возникновению наследственных заболеваний и не детектируемых с помощью секвенирования ДНК, как изменяющих аминокислотную последовательность белка, является нарушение сплайсинга. На настоящий момент при интерпретации геномных данных учитываются только варианты, затрагивающие канонические сайты сплайсинга. Доля таких генетических вариантов оценивается примерно в 8.7% [13]. При этом, группой исследователей из Великобритании и Испании с помощью математического моделирования было спрогнозировано, что 62% из всех патогенетически значимых генетических вариантов могут приводить к аномалиям сплайсинга РНК [14].

СПЛАЙСИНГ В НОРМЕ И ПАТОЛОГИИ

На настоящий момент известно около 20000 генов, кодирующих белки человека и около 150000 изоформ транскриптов. Следовательно, в среднем каждый ген человека имеет около семи различных изоформ [15]. Альтернативный сплайсинг характерен для 90% интронсодержащих генов человека [16]. Основным эффектором реакции сплайсинга РНК является сплайсосома, комплекс из сотен взаимодействующих белков и малых ядерных РНК (**мяРНК**), включая малые ядерные рибонуклеопротеины (**мяРНП**) U1, U2, U4, U5 и U6. Каждый интрон пре-мРНК фланкирован 5'-экзоном и 3'-экзоном и содержит различные консервативные сигналы сплайсинга, распознаваемые сплайсосомой: 5'-сайт сплайсинга, последовательность точки ветвления, 3'-сайт сплайсинга и полипиримидиновый тракт, расположенный на 5–40 пн выше 3'-конца интрона. Поскольку этих сигналов сплайсинга недостаточно для регуляции сплайсинга, точность сплайсинга пре-мРНК зависит от взаимодействий между транс-действующими факторами (белками и рибонуклеопротеинами) и цис-действующими элементами (последовательностями пре-мРНК), включая экзонный энхансер сплайсинга, экзонный сайленсер сплайсинга, интронный энхансер сплайсинга и интронный сайленсер сплайсинга. Все эти элементы оказывают свое воздействие, модулируя связывание факторов сплайсинга, которые, в свою очередь, положительно или отрицательно регулируют включение определенного экзона в состав зрелой мРНК [17].

Сплайсинг пре-мРНК играет важную роль в формировании разнообразия белков и функционировании различных клеток и тканей организма, что сказывается на роли нарушения нормальных паттернов сплайсинга в дисфункции генов и развитии заболеваний. Описаны заболевания, в основе которых лежат мутации, затрагивающие сплайсосомы. Так, мутации в гене *SNRPB*, кодирующем полипептиды B и B1 мяРНП, приводят к развитию церебро-косто-мандибулярного синдрома; мутации в гене *EFTUD2*, приводят к развитию одного из типов нижнечелюстно-лицевого дизостоза; мутации в гене *SF3B4* (Splicing Factor 3b Subunit 4) идентифицированы при синдроме Нагера и др.

[15]. Среди многих генов, ответственных за развитие пигментного ретинита, описаны шесть генов, участвующих в процессинге пре-мРНК (*PRPF3*, *PRPF4*, *PRPF6*, *PRPF8*, *PRPF31* и *SNRNP200*) [18]. Кроме того, показана роль альтернативного сплайсинга в развитии солидных злокачественных новообразований [19–21]. Патогенные генетические варианты в генах *SF3B1*, *U2AF1* и *U2AF2* приводят к развитию некоторых типов миелоидных новообразований [18].

Тем не менее основная роль сплайсинга в развитии различных патологий обусловлена не нарушением механизмов процессинга пре-мРНК, а изменениями в регуляторных последовательностях самих генов. Для поиска таких генетических вариантов одним из удобных и доступных инструментов является секвенирование РНК.

ПОИСК МУТАЦИЙ С ИСПОЛЬЗОВАНИЕМ РНК-СЕКВЕНИРОВАНИЯ

Однонуклеотидные варианты в некодирующих частях генов могут быть ответственны за значительную часть наблюдаемых вариаций фенотипов [22]. Точный сплайсинг пре-мРНК, необходимый для соответствующей трансляции белка, зависит от наличия консенсусных последовательностей, которые определяют границы между экзонами и интронами, и регуляторных последовательностей, распознаваемых механизмом сплайсинга. Точечные мутации в этих консенсусных последовательностях могут вызвать неправильное распознавание экзона и интрона и привести к образованию аберрантного транскрипта гена. Мутация сплайсинга может происходить как в интронах, так и в экзонах и нарушать существующие сайты сплайсинга или регуляторные последовательности сплайсинга, создавать новые или активировать криптические сайты сплайсинга. Обычно такие мутации приводят к ошибкам в процессе сплайсинга и могут привести к неправильному удалению интрона, пропуску или появлению дополнительного экзона [23].

На настоящий момент известно 23868 мутаций, приводящих к нарушениям сплайсинга, которые ответственны за наследственные заболевания человека. Частота таких генетических нарушений составляет 8.7% от всех мутаций, вызывающих наследственные болезни [13]. Вероятно, это число занижено, поскольку большинство описанных мутаций были идентифицированы с помощью секвенирования геномной ДНК без учета влияния мутаций на сплайсинг. Недавние исследования указывают на высокую частоту и важную роль мутаций сплайсинга в этиологии наследственных заболеваний, включая миодистрофию Дюшенна [24], муковисцидоз [25], болезнь Элерса–Данло [26], наследственные болезни сетчатки [27] и другие моногенные патологии. При анализе геномной ДНК их можно легко не заметить и ошибочно классифицировать как синонимичные изменения или доброкачественные аминокислотные замены. Однако анализ РНК ясно показывает, что такие мутации оказывают значительное влияние на сплайсинг пре-мРНК. Предполагалось, что более крупные гены с длинными интронами более склонны к дефектам сплайсинга, но теперь стало очевидно, что значительное число мутаций в меньших генах также вызывает аномальный сплайсинг мРНК [28]. Кроме того, многие из выявленных сплайсинговых мутаций находятся вне канонических сайтов сплайсинга и могут быть легко пропущены при анализе геномной ДНК. Появляется все больше свидетельств того, что неправильная классификация мутаций является распространенной ошибкой, и общее число дефектов сплайсинга, вероятно, недооценивается [29].

В последние годы начинают накапливаться данные об использовании РНК-секвенирования для поиска патогенетически значимых мутаций у пациентов с моногенными заболеваниями (табл. 1). По данным некоторых исследователей, диагностическая ценность РНК-секвенирования находится в пределах 10–35% для разных групп пациентов [30]. Кроме того, было показано что использование РНК-секвенирования в качестве диагностического метода может помочь расширить наши знания о патогенетической значимости вариантов с неизвестной клинической значимостью (variants of unknown significance, VUS), идентифицированных с помощью секвенирования ДНК [17]. В табл. 1 представлена информация об исследованиях по поиску патогенных генетических вариантов с использованием РНК-секвенирования.

Так, группа исследователей из Массачусетса проанализировали 63 пациента с подозрением на моногенные мышечные заболевания (миопатии и мышечные дистрофии) и 184 контрольных образца из проекта The Genotype-Tissue Expression (GTEx). При этом у 13 пациентов были диагностированы патогенные варианты, влияющие на транскриптом (нонсенс-мутации и мутации в канонических сайтах сплайсинга), которые использовались в качестве положительного контроля. У 16 пациентов без установленного диагноза с помощью экзомного секвенирования были идентифицированы спрогнозированные варианты, влияющие на сплайсинг ($n = 4$), либо сильные гены-кандидаты ($n = 12$). У 34 пациентов не было выявлено ни того, ни другого. В качестве материала для исследования были использованы образцы биопсии мышц. По результатам проведенной работы в 35% случаев удалось выявить патогенные варианты. Наиболее высокая частота обнаружения патогенных генетических вариантов, пропущенных при экзомном и полногеномном секвенировании, была в группе пациентов со спрогнозированными кандидатными вариантами (50%) и

Таблица 1. Эффективность РНК-секвенирования при моногенных заболеваниях

Авторы	Технология пробоподготовки	Источник РНК	Количество прочтений на образец	Количество обследованных пациентов	Эффективность выявления патогенных генетических вариантов
[30]	Набор реактивов для обогащения полиА-транскриптами (Illumina TruSeq)	Скелетные мышцы	50–100 млн прочтений на образец. Один образец был секвенирован с глубиной 500 млн чтений	50 пациентов с генетически невыявленными редкими мышечными заболеваниями	35% (17/50)
[36]	Набор реактивов TrueSeq Stranded mRNA Sample Prep Kit (Illumina)	Кровь	Примерно 50 млн чтений на образец	94 пациента с подозрением на различные орфанные заболевания и невыясненным диагнозом	7.5% (6/80)
[31]	Набор реактивов для обогащения полиА-транскриптами (TruSeq Illumina)	Фибробласты кожи	Примерно 50–100 млн чтений на образец	25 семей с невыясненным генетическим диагнозом	36% (9/25)
[33]	Набор реактивов Illumina TruSeq Stranded mRNA	Кровь, культивированные фибробласты кожи и клетки почечного эпителия	–	155 случаев с менделевскими наследственнымизаболеваниями с невыясненным генетическим диагнозом по результатам полногеномного секвенирования	13.5% (21/155)
[34]	–	Кровь, фибробласты кожи, мышцы или костный мозг	Примерно 50–100 млн чтений на образец	48 семей (91 образец), для которых ранее было проведено геномное секвенирование	15% (7/48)
[35]	Набор реактивов NEBNext Ultra Directional RNA Library Prep Kit для Illumina (New England Biolabs).	Кровь	–	Пациенты с 257 генетическими вариантами с неясной клинической значимостью	33% (58/257)
[32]	Набор реактивов для обогащения полиА-транскриптами (Illumina)	49 (42%) – фибробласты, 18 (16%) – кровь, 48 (42%) – две ткани	Примерно 30–50 млн чтений на образец	115 пациентов: для 72 (63%) пациентов было проведено экзомноесеквенирование,для 29 (25%) – геномное секвенирование, для 14 (12%) экзомное и геномное секвенирование	12% (14/115)

с сильным геном-кандидатом (66%). При этом даже в группе пациентов без гена-кандидата или кандидатного варианта в 21% случаев удалось детектировать мутации [30].

В аналогичной работе канадских исследователей на выборке пациентов с нейромышечными заболеваниями были продемонстрированы более значительные результаты по оценке вклада РНК-секвенирования в поиск новых патогенных мутаций. Было показано, что в 36% случаев (9/25) секвенирование РНК позволяет выявлять причинную мутацию у пациентов без установленного диагноза после экзомного секвенирования [31].

Исследователи из Медицинского Колледжа Бейлора (Хьюстон, США) провели транскриптомный анализ для 182 пациентов с наследственными заболеваниями без установленного диагноза после экзомного секвенирования и хромосомного микроматричного анализа, в результате в 17% случаев были идентифицированы патогенные генетические варианты [32]. Были выявлены патогенные мутации разных типов: мутации в канонических сайтах сплайсинга (7%), синонимичные мутации в экзонах (7%), мутации в интронах (43%), мутации в промоторах генов (7%) и вариации числа копий ДНК (36%).

Аналогичная работа была проведена S. Maddirevula с соавт., в которой был проведен полнотранскриптомный анализ 155 пациентов без идентифицированной мутации с помощью экзомного секвенирования. Генетические варианты, приводящие к потере транскриптов (transcript-deleterious variants, TDV), были найдены в 13.5% случаев. Кроме того, в данной работе был проведен анализ тканеспецифичной экспрессии генов с TDV. В частности были проанализированы образцы РНК, полученные из крови, фибробластов кожи и клеток почечного эпителия, выделенных из мочи. Было обнаружено, что 84.1% (195 из 232) проанализированных генов экспрессируются в РНК клеток крови, 85.8% (199 из 232) в РНК фибробластов и 90% (209 из 232) в РНК клеток почечного эпителия. Большинство генов экспрессировались во всех трех источниках РНК (75.5%), и только 2.6% (6 из 232 генов) не экспрессировались ни в одном из них [33].

Комбинирование методов оценки ДНК и РНК приводит к повышению диагностической значимости массового параллельного секвенирования, что было показано в работе американских исследователей из Калифорнийского университета. Они проанализировали 234 образца от пациентов без установленного диагноза с помощью экзомного, полногеномного и транскриптомного секвенирования. В результате диагностическая ценность методов анализа ДНК составила 31%, в то время как добавление РНК-секвенирования добавило еще 7%, расширив общую диагностическую ценность до 38%. Кроме того, с помощью РНК-секвенирования была установлена патогенетическая значимость для 18% генетических вариантов, найденных с помощью секвенирования ДНК [34]. H. Wai с соавт. с помощью количественного ПЦР в режиме реального времени и РНК-секвенирования проанализировали функциональную значимость 257 генетических вариантов с неясной клинической значимостью. В результате проведенной работы было установлено, что 58 вариантов (33%) были ассоциированы с аномалиями сплайсинга, т.е. была установлена патогенетическая значимость вариантов с неясной клинической значимостью [35].

Одной из основных сложностей при использовании РНК-секвенирования является тканеспецифичная экспрессия многих генов и мало- или недоступность для анализа многих целевых тканей. В то время как для анализа с помощью секвенирования ДНК используется наиболее доступный материал – периферическая кровь, для анализа транскриптома данный биологический материал может быть малоинформативным. Группа исследователей из Стэнфорда провела секвенирование РНК из цельной крови для 94 пациентов с подозрением на различные орфанные заболевания (неврологические, скелетно-мышечные и ортопедические, гематологические и офтальмологические), но с невыясненным диагнозом. Экспрессия в клетках крови была показана для 76% из 284 генов, связанных с неврологическими расстройствами, и 66% всех генов, чувствительных к потере функции (loss-of-function intolerance). В 7.5% случаев был выставлен диагноз, при этом гены-кандидаты были обозначены для еще 16.7%. Авторы идентифицировали кандидатные гены с использованием анализа уровня экспрессии генов, аллель-специфической экспрессии и прогнозирования аномалий сплайсинга. Данная работа показала широкую применимость технологии РНК-секвенирования, даже для пациентов, у которых целевая ткань является труднодоступной [36].

Другим подходом для поиска аномалий сплайсинга является использование компьютерных программ и алгоритмов. Исследователями из лаборатории искусственного интеллекта компании “Illumina” была разработана компьютерная программа “SpliceAI” для предсказания аномалий сплайсинга *in silico* на основе данных секвенирования ДНК [37]. “SpliceAI” представляет собой остаточную нейронную сеть, имеющую сетевую архитектуру, состоящую из 32 расширенных сверточных слоев, которые могут распознавать детерминанты последовательности, охватывающие большие регионы генома. Для обучения нейронной сети авторы использовали аннотированные последовательности транскриптов пре-мРНК в GENCODE. При этом точность предска-

зания событий сплайсинга для транскриптов пре-мРНК в тестовом наборе данных составила 95%. Даже гены размером более 100 тпн, такие как *CFTR*, часто реконструируются с идеальной точностью [37]. Дальнейшее улучшение подходов для поиска аномалий сплайсинга с помощью моделирования *in silico* может повысить эффективность диагностики наследственных заболеваний. Кроме того, это позволит расширить наши представления о механизмах регуляции такого сложного процесса как сплайсинг пре-мРНК.

Вне зависимости от подходов и методов, которые используются для поиска аномалий сплайсинга, идентифицированные генетические варианты требуют верификации с использованием методов функционального анализа. Наиболее удобным инструментом для этого является использование системы минигенов. Минигенные конструкции представляют собой участки генов, содержащие экзон и фланкирующие интронные области с регуляторными элементами. Использование данной модельной системы позволяет определить патогенность различных генетических вариантов посредством оценки их влияния на эффективность сплайсинга, а также провести поиск экзонных и интронных энхансеров и сайленсеров сплайсинга. Кроме того, с помощью минигенов можно оценить роль сайтов сплайсинга в установлении базового уровня распознавания экзонов и для установления роли различных транс-регуляторов на отдельные события сплайсинга [38].

РНК-секвенирование имеет хорошие перспективы в качестве одного из инструментов для диагностики наследственных орфанных заболеваний. Тем не менее существует ряд вопросов, без ответов на которые могут возникнуть сложности с анализом и интерпретацией получаемых результатов. К таким вопросам можно отнести тканеспецифичную экспрессию генов и выбор тканей для анализа при различных нозологиях, оптимизацию и стандартизацию методики РНК-секвенирования и методов биоинформатической обработки данных, понимание необходимого уровня аберрантного сплайсинга для возникновения патологического фенотипа, влияние внутригенного и межгенного контекста на развитие аномалий сплайсинга. Для ответа на все данные вопросы необходимо более детальное изучение и глубокое понимание фундаментальных основ сплайсинга пре-мРНК. Знание этих механизмов позволит не только улучшить диагностику, но и значительно продвинуться в лечении орфанных болезней с помощью препаратов, модулирующих сплайсинг.

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (Федеральная научно-техническая программа развития генетических технологий на 2019–2027 гг., соглашение № 075-15-2021-1061, РФ 193021X0029).

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Orphanet (Электронный ресурс). URL: https://www.orpha.net/
2. *Nguengang Wakap S., Lambert D.M., Olry A. et al.* Estimating cumulative point prevalence of rare diseases: Analysis of the Orphanet database // Eur. J. Hum. Genet. 2020. V. 28. № 2. P. 165–173. https://doi.org/10.1038/s41431-019-0508-0
3. OMIM (Электронныйресурс). P. Online Mendelian Inheritance in Man. URL: https://www.omim.org/
4. *Choi M., Scholl U.I., Ji W. et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing // Proc. Natl Acad. Sci. 2009. V. 106. № 45. P. 19096–19101. https://doi.org/10.1073/pnas.0910672106
5. *Valencia C.A., Husami A., Holle J. et al.* Clinical impact and cost-effectiveness of whole exome sequencing as a diagnostic tool: A pediatric center's experience // Front. Pediatr. 2015. V. 3. https://doi.org/10.3389/fped.2015.00067
6. *Tan T.Y., Dillon O.J., Stark Z. et al.* Diagnostic impact and cost-effectiveness of whole-exome sequencing for ambulant children with suspected monogenic conditions // JAMA Pediatr. 2017. V. 171. № 9. P. 855. https://doi.org/10.1001/jamapediatrics.2017.1755
7. *Retterer K., Juusola J., Cho M.T. et al.* Clinical application of whole-exome sequencing across clinical indications // Genet. Med. 2016. V. 18. № 7. P. 696–704. https://doi.org/10.1038/gim.2015.148
8. *Lionel A.C., Costain G., Monfared N. et al.* Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test // Genet. Med. 2018. V. 20. № 4. P. 435–443. https://doi.org/10.1038/gim.2017.119
9. *Schwarze K., Buchanan J., Taylor J.C. et al.* Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature // Genet. Med. 2018. V. 20. № 10. P. 1122–1130. https://doi.org/10.1038/gim.2017.247
10. *Robertson A.J., Tan N.B., Spurdle A.B. et al.* Reanalysis of genomic data: An overview of the mechanisms and complexities of clinical adoption // Genet. Med. 2022. P. 1–13. https://doi.org/10.1016/j.gim.2021.12.011
11. *Liu P., Meng L., Normand E.A. et al.* Reanalysis of clinical exome sequencing data // N. Engl. J. Med. 2019. V. 380. № 25. P. 2478–2480. https://doi.org/10.1056/NEJMc1812033

12. *Tan N.B., Stapleton R., Stark Z. et al.* Evaluating systematic reanalysis of clinical genomic data in rare disease from single center experience and literature review // Mol. Genet. Genomic Med. 2020. V. 8. № 11. P. 1–19. https://doi.org/10.1002/mgg3.1508
13. *Stenson P.D., Mort M., Ball E.V. et al.* The Human Gene Mutation Database: Towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies // Hum. Genet. 2017. V. 136. № 6. P. 665–677. https://doi.org/10.1007/s00439-017-1779-6
14. *López-Bigas N., Audit B., Ouzounis C. et al.* Are splicing mutations the most frequent cause of hereditary disease? // FEBS Lett. 2005. V. 579. № 9. P. 1900–1903. https://doi.org/10.1016/j.febslet.2005.02.047
15. *Jiang W., Chen L.* Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing // Comput. Struct. Biotechnol. J. 2021. V. 19. P. 183–195. https://doi.org/10.1016/j.csbj.2020.12.009
16. *Kalsotra A., Cooper T.A.* Functional consequences of developmentally regulated alternative splicing // Nat. Rev. Genet. 2011. V. 12. № 10. P. 715–729. https://doi.org/10.1038/nrg3052
17. *Marco-Puche G., Lois S., Benítez J. et al.* RNA-seq perspectives to improve clinical diagnosis // Front. Genet. 2019. V. 10. https://doi.org/10.3389/fgene.2019.01152
18. *Scotti M.M., Swanson M.S.* RNA mis-splicing in disease // Nat. Rev. Genet. 2016. V. 17. № 1. P. 19–32. https://doi.org/10.1038/nrg.2015.3
19. *Wu Z.-H., Tang Y., Zhou Y.* Alternative splicing events implicated in carcinogenesis and prognosis of thyroid gland cancer // Sci. Rep. 2021. V. 11. № 1. P. 4841. https://doi.org/10.1038/s41598-021-84403-6
20. *Marin J.J.G., Reviejo M., Soto M. et al.* Impact of alternative splicing variants on liver cancer biology // Cancers (Basel). 2022. V. 14. № 1. P. 18. https://doi.org/10.3390/cancers14010018
21. *Kim B.-H., Woo T.-G., Kang S.-M. et al.* Splicing variants, protein-protein interactions, and drug targeting in hutchinson-gilford progeria syndrome and small cell lung cancer // Genes (Basel). 2022. V. 13. № 2. P. 165. https://doi.org/10.3390/genes13020165
22. *Wachs A.S., Bohne J.* Two sides of the same medal: Noncoding mutations reveal new pathological mechanisms and insights into the regulation of gene expression // WIREs RNA. 2021. V. 12. № 1. P. 1–21. https://doi.org/10.1002/wrna.1616
23. *Anna A., Monika G.* Splicing mutations in human genetic disorders: Examples, detection, and confirmation // J. Appl. Genet. 2018. V. 59. № 3. P. 253–268. https://doi.org/10.1007/s13353-018-0444-7
24. *Habara Y., Takeshima Y., Awano H. et al. In vitro* splicing analysis showed that availability of a cryptic splice site is not a determinant for alternative splicing patterns caused by +1G>A mutations in introns of the dystrophin gene // J. Med. Genet. 2009. V. 46. № 8. P. 542–547. https://doi.org/10.1136/jmg.2008.061259
25. *Sanz D.J., Hollywood J.A., Scallan M.F. et al.* Cas9/gRNA targeted excision of cystic fibrosis-causing deep-intronic splicing mutations restores normal splicing of CFTR mRNA // PLoS One. 2017. V. 12. № 9. P. e0184009. https://doi.org/10.1371/journal.pone.0184009
26. *Symoens S., Malfait F., Vlummens P. et al.* A novel splice variant in the n-propeptide of COL5A1 causes an eds phenotype with severe kyphoscoliosis and eye involvement // PLoS One. 2011. V. 6. № 5. P. e20121. https://doi.org/10.1371/journal.pone.0020121
27. *Weisschuh N., Buena-Atienza E., Wissinger B.* Splicing mutations in inherited retinal diseases // Prog. Retin. Eye Res. 2021. V. 80. P. 100874. https://doi.org/10.1016/j.preteyeres.2020.100874
28. *Chen M., Manley J.L.* Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches // Nat. Rev. Mol. Cell Biol. 2009. V. 10. № 11. P. 741–754. https://doi.org/10.1038/nrm2777
29. *Xiong H.Y., Alipanahi B., Lee L.J. et al.* The human splicing code reveals new insights into the genetic determinants of disease // Science. 2015. V. 347. № 6218. https://doi.org/10.1126/science.1254806
30. *Cummings B.B., Marshall J.L., Tukiainen T. et al.* Improving genetic diagnosis in Mendelian disease with transcriptome sequencing Genotype-Tissue Expression Consortium // Sci. Transl. Med. 2017. V. 9. № 386. https://doi.org/10.1126/scitranslmed.aal5209
31. *Gonorazky H.D., Naumenko S., Ramani A.K. et al.* Expanding the boundaries of RNA sequencing as a diagnostic tool for rare mendelian disease // Am. J. Hum. Genet. 2019. V. 104. № 3. P. 466–483. https://doi.org/10.1016/j.ajhg.2019.01.012
32. *Murdock D.R., Dai H., Burrage L.C. et al.* Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing // J. Clin. Invest. 2021. V. 131. № 1. https://doi.org/10.1172/jci141500
33. *Maddirevula S., Kuwahara H., Ewida N. et al.* Analysis of transcript-deleterious variants in Mendelian disorders: Implications for RNA-based diagnostics // Genome Biol. Genome Biology. 2020. V. 21. № 1. P. 145. https://doi.org/10.1186/s13059-020-02053-9
34. *Lee H., Huang A.Y., Wang L. et al.* Diagnostic utility of transcriptome sequencing for rare Mendelian diseases // Genet. Med. 2020. V. 22. № 3. P. 490–499. https://doi.org/10.1038/s41436-019-0672-1
35. *Wai H.A., Lord J., Lyon M. et al.* Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance // Genet. Med. 2020. V. 22. № 6. P. 1005–1014. https://doi.org/10.1038/s41436-020-0766-9
36. *Frésard L., Smail C., Ferraro N.M. et al.* Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts // Nat. Med. 2019. V. 25. № 6. P. 911–919. https://doi.org/10.1038/s41591-019-0457-8
37. *Jaganathan K., Kyriazopoulou Panagiotopoulou S., McRae J.F. et al.* Predicting splicing from primary sequence with deep learning // Cell. 2019. V. 176. № 3. P. 535–548. e24. https://doi.org/10.1016/j.cell.2018.12.015
38. *Cooper T.A.* Use of minigene systems to dissect alternative splicing elements // Methods. 2005. V. 37. № 4. P. 331–340. https://doi.org/10.1016/j.ymeth.2005.07.015

The Role of Splicing in the Pathogenesis of Monogenic Diseases

N. A. Skryabin[a], *, D. I. Zhigalina[a], and V. A. Stepanov[a]

[a] *Research Institute of Medical Genetics, Tomsk National Research Medical Center, Russian Academy of Sciences, Tomsk, 634050 Russia*

**e-mail: nikolay.skryabin@medgenetics.ru*

Despite the development of exome and whole genome sequencing technologies and their routine use in the diagnosis of hereditary diseases, the efficiency of detection of pathogenic genetic variants for methods based on DNA analysis is less than 50%. One of the main reasons may be the inefficiency of these approaches in the search for genetic variants responsible for impaired pre-mRNA splicing. This review discusses the results of work on the search for splicing abnormalities in hereditary orphan diseases using RNA sequencing and the possibility of clinical application of this method.

Keywords: RNA, orphan diseases, mutations, splicing, expression.