УСПЕХИ СОВРЕМЕННОЙ БИОЛОГИИ, 2023, том 143, № 2, с. 107–113

УДК 577.1:57.088

НЕСИНОНИМИЧНЫЕ ОДНОНУКЛЕОТИДНЫЕ ЗАМЕНЫ И ИНДЕЛЫ: ВКЛАД В МОЛЕКУЛЯРНЫЙ ПОСТГЕНОМНЫЙ ПОРТРЕТ КЛЕТОЧНОЙ ЛИНИИ HepG2

© 2023 г. Е. В. Поверенная[1], *, О. И. Киселева[1], В. А. Арзуманян[1], М. А. Пятницкий[1], И. В. Вахрушев[1], Е. А. Пономаренко[1]

[1]*Научно-исследовательский институт биомедицинской химии им. В.Н. Ореховича, Москва, Россия*

**e-mail: k.poverennaya@gmail.com*

Поступила в редакцию 09.12.2022 г.
После доработки 09.12.2022 г.
Принята к публикации 29.12.2022 г.

В геноцентричном режиме проведен сравнительный анализ результатов геномного, транскриптомного и протеомного профилирования образцов клеточной линии HepG2. Показана прослеживаемость на транскриптомном и протеомном уровнях изменений, связанных с наличием несинонимичных однонуклеотидных замен и инделов в геноме. На транскриптомном уровне регистрируется большинство молекулярных событий, вызванных аберрациями на геномном уровне. В то же время лишь единичные протеоформы могут быть детектированы из числа кодируемых выбранными мутантными генами. Вероятно, это связано с методическими ограничениями протеомных методов, не позволяющих регистрировать протеоформы, присутствующие в образце в низких концентрациях. Результаты исследования согласуются с ранее полученными данными других научных групп и описывают принципиальные методические решения, требуемые для расшифровки молекулярного постгеномного портрета биологических образцов с разрешением на уровне аберрантных молекул.

Ключевые слова: мутации, геном, транскриптом, протеом, протеоформы, молекулярное профилирование

DOI: 10.31857/S0042132423020096, **EDN:** KMZWKK

ВВЕДЕНИЕ

Реализация генетической информации в каждом органе, ткани и клетке живых организмов зависит от конкретной роли, направленной на обеспечение функционирования целостного организма. Для человека известно около 20 тыс. белок-кодирующих генов (Ensembl release 94), что в несколько раз меньше первоначальных оценок, которые были сделаны на старте международного проекта “Геном человека” (Venter et al., 2001; Smith, Kelleher, 2013). В каждом типе ткани экспрессируется в среднем 50–60% всего генома (GTEx Consortium, 2015), при этом количество вариантов белковых продуктов (протеоформ), кодируемых одним геном, может достигать 100 (Ponomarenko et al., 2016; Aebersold et al., 2018). Многообразие протеоформ обусловливает протеомную гетерогенность.

Основные источники протеомной гетерогенности – несинонимичные однонуклеотидные замены, альтернативный сплайсинг и посттрансляционные модификации. Каждый вариант белка (протеоформа) может иметь отличную от других функцию, участвуя в различных биологических процессах (Poverennaya et al., 2020). Показано, что из 65 тыс. несинонимичных полиморфизмов, найденных в геноме человека, более 30% являются причиной возникновения белков с измененными функциями (Yip et al., 2008). Следовательно, изменение или нарушение функции только одного белка может приводить к перестройке взаимосвязей между различными уровнями (транскриптомным, протеомным, метаболомным) реализации генетической информации, меняя молекулярный постгеномный портрет.

Развитие технологий молекулярного анализа в направлении высокопроизводительного секвенирования и масс-спектрометрии привело к накоплению существенных массивов данных, что, в свою очередь, способствовало углублению знаний о возможностях реализации генома. Полученные в ходе таких исследований наборы данных используют для системного изучения и последующего мультиомного моделирования (Vitrinel et al., 2019) взаимосвязей между различными омикс-слоями – геномным, транскриптомным, трансляторным, протеомным. Выявленные механизмы эпигенетической регуляции экспрессии, механизмы регуляции экспрессии гена на уровне трансля-

ции, связи между уровнем транскрипта и белка (Edfors et al., 2016) – сложные, не позволяющие описать процессы реализации генетической информации линейными закономерностями, например простой зависимостью между количеством мРНК и соответствующего белка в клетке (Tenzer et al., 2016; Poverennaya et al., 2017). Ключевой фактор для корректного моделирования биологических процессов – коэффициент переноса/потерь информации между различными слоями реализации генетической информации и, в итоге, конечным фенотипическим проявлением.

В данной работе изучен вопрос о последовательной реализации мутаций (несинонимичных однонуклеотидных замен и инделов) на примере клеточной линии HepG2 (Arzumanian et al., 2021).

МАТЕРИАЛЫ И МЕТОДЫ

Культивирование клеточной линии HepG2

Клеточная линия HepG2 (гепатобластома человека, SCC249) была приобретена у Merck KGaA, Deutschland. После оттаивания клетки культивировали в среде DMEM/F12 с добавлением 10% эмбриональной телячьей сыворотки и 100 ед./мл пенициллина/стрептомицина (все от Gibco, USA) в CO_2-инкубаторе в стандартных условиях (5% CO_2, 37°C, влажность 100%). Для подготовки образцов клетки отделяли 0.25%-ным раствором трипсин-ЭДТА (ПанЭко, Россия), трижды промывали PBS и подсчитывали в камере Горяева. Все используемые операции и реагенты были максимально унифицированы для каждого образца нестационарной клеточной линии, чтобы свести к минимуму возможные технические ошибки.

Полногеномное секвенирование

ДНК выделяли с помощью набора PureLink Genomic DNA Mini Kit (Thermo FS, USA), согласно протоколу производителя. В финальной концентрации ДНК 100 нг в объеме 50 мкл была подготовлена библиотека с использованием TruSeq Nano DNA Library Prep Kit (Illumina, USA). Контроль концентрации полученной библиотеки осуществляли с помощью набора Qubit dsDNA HS Assay Kit (Thermo FS, USA) на флуориметре Qubit 2.0 (Invitrogen, USA). Секвенирование было произведено на приборе NovaSeq 6000 (Illumina, USA) с длиной прочтения 2 × 150 п.о. Качество геномных данных проверяли с помощью программы FastQC (Trivedi et al., 2014). Для обработки результатов секвенирования использовали фреймворк bcbio-nextgen (https://github.com/bcbio/bcbio-nextgen). Анализ данных выполнялся в соответствии с рекомендациями GATK best practices от Broad Inst., USA.

Поиск несинонимичных однонуклеотидных замен и инделов проводили относительно референсного генома (Ensembl v103) в GATK HaplotypeCaller (van der Auwera et al., 2013), как для геномных данных. Полученные генетические варианты аннотировались с помощью программы Jannovar; их также искали в каталоге Cancer Gene Census (https://cancer.sanger.ac.uk/census).

Собранные данные депонированы в SRA NCBI, номер проекта PRJNA765908.

Транскриптомное секвенирование

РНК выделяли с помощью набора PureLink RNA Mini Kit (Thermo FS, USA), согласно протоколу производителя. В финальной концентрации РНК 10 нг в объеме 50 мкл была подготовлена кДНК-библиотека с использованием TruSeq Stranded mRNA Library Prep Kit (Illumina, USA). Контроль концентрации полученной библиотеки осуществляли с помощью набора Qubit dsDNA HS Assay Kit (Thermo FS, USA) на флуориметре Qubit 2.0 (Invitrogen, USA). Секвенирование было произведено на приборе NovaSeq 6000 (Illumina, USA) с длиной прочтения 2 × 100 п.о. Качество геномных данных проверяли с помощью программы FastQC. Для обработки результатов секвенирования использовали фреймворк bcbio-nextgen (https://github.com/bcbio/bcbio-nextgen). Анализ данных выполнялся в соответствии с рекомендациями GATK best practices от Broad Inst., USA.

Поиск несинонимичных однонуклеотидных замен и инделов проводили относительно референсного генома (Ensembl v103) в GATK HaplotypeCaller (van der Auwera et al., 2013), как для геномных данных. Полученные генетические варианты аннотировались с помощью программы Jannovar; их также искали в каталоге Cancer Gene Census (https://cancer.sanger.ac.uk/census).

Собранные данные депонированы в SRA NCBI, номер проекта PRJNA765908. Детально описание опубликовано в статье (Pyatnitskiy et al., 2021).

Масс-спектрометрия

Белковая фракция из предварительно алкилированного и восстановленного лизата клеток линии HepG2 была гидролизирована на пептиды модифицированной протеазой трипсина (Thermo FS, USA). Далее был выполнен хромато-масс-спектрометрический анализ в панорамном режиме в трех технических повторениях. Хроматографическое разделение пептидов проводили с помощью хроматографа Ultimate 3000 RSLC Nano system (Thermo FS, USA). Масс-спектрометрический анализ выполняли с помощью масс-спектрометра сверхвысокого разрешения Orbitrap Fusion (Thermo FS, USA). Полученные файлы масс-спектров были проанализировали с помощью пакета SearchGUI (версия 4.0.5) с использованием трех поисковых алгоритмов (X!Tandem, MS-GF+ и OMSSA)

(Kiseleva et al., 2018). В качестве референсной базы аминокислотных последовательностей белков, по которой производили биоинформатический поиск, использовали библиотеку, созданную на основе белковых сиквенсов человека из репозитория SwissProt (v.2022_01) с добавлением оригинальных последовательностей с аминокислотными заменами, идентифицированными на транскриптомном уровне, а также белков-контаминантов из базы cRAP (Mellacheruvu et al., 2013). Белок считали идентифицированным, если удавалось детектировать не менее двух протеотипических пептидов длиной 9–26 аминокислотных остатков. Отсечение по ложноположительным идентификациям как для пептидов, так и для белков и для совпадающих пар теоретического и экспериментального спектров установлено на уровне 1%.

Полученные протеомные данные доступны в Mendeley Data (https://data.mendeley.com/).

Опубликованные данные

Данные полногеномного секвенирования с аналогичными параметрами секвенирования исследуемого образца клеток HepG2 (технология Illumina, парноконцевое прочтение с длиной рида 100 п.о., покрытие не менее 90X) выгружены из SRA NCBI (идентификаторы SRR14832808, SRR5296491).

Данные транскриптомного секвенирования с аналогичными параметрами секвенирования исследуемого образца клеток HepG2 (технология Illumina, парноконцевое прочтение с длиной рида 150 п.о., с суммарным количеством ридов не менее 30 млн ридов) выгружены из SRA NCBI SRR17514594 (PRJNA795717), SRR10011494 (PRJNA561411), SRR12132985 (PRJNA643657).

Биоинформатическая обработка полученных данных проводилась в соответствии с анализом данных секвенирования исследуемого образца.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Прогресс развития высокопроизводительных омикс-технологий способствовал расширению знаний о возможностях реализации генома. Метаанализ данных зачастую приводит к противоречивым результатам, которые связаны не только с различием технологических платформ, на которых были выполнены исследования, но и с изначальной гетерогенностью образцов. В 2019 г. на примере широкомасштабного профилирования образцов клеточной линии HeLa, предоставленных 13 лабораториями, было показано наличие нескольких молекулярных и фенотипических профилей этих клеток (Liu et al., 2019), что изменило представление о клеточных линиях как о стабильных конструктах. Следовательно, для оценки стабильности путей реализации генетической информации необходим комплексный анализ результатов мультиомного (геномного, транскриптомного и протеомного) профилирования конкретного образца.

На примере клеток HepG2 (аликвоты получены в максимально унифицированных условиях одновременно из одного образца клеточной линии) проведен полногеномный, транскриптомный и протеомный анализ для оценки гетерогенности соответствующих молекулярных профилей с учетом реализации мутаций (несинонимичных однонуклеотидных замен (nsSNP) и инделов). Для создания молекулярного портрета гетерогенности омикс-слоев полученные результаты сопоставлены с опубликованными ранее результатами профилирования клеточной линии HepG2.

Поиск генетических вариантов по данным полногеномного секвенирования

В результате полногеномного секвенирования (WGS, whole genome sequencing) клеток HepG2 найдено 11674 генетических варианта, из которых 11186 вариантов относились к классу несинонимичных однонуклеотидных замен и 488 вариантов являлись инделами. Найденные nsSNP были распределены по 5858 генам. Из них в 1961 гене встречались только гомозиготные генетические варианты (92 индела, 2694 однонуклеотидные замены), в 3155 генах встречались только гетерозиготные генетические варианты (237 инделов, 4841 однонуклеотидная замена). Из 5858 генов для 187 генов известна их роль в онкогенезе, согласно каталогу Cancer Gene Census. При этом в 55 генах встречались только гомозиготные генетические варианты, в 103 генах – только гетерозиготные.

Поиск генетических вариантов по данным РНК-секвенирования

По данным РНК-секвенирования (RNASeq) клеточной линии HepG2 выявлено 13 тыс. экспрессирующихся генов, для которых детектированы 1781 мутация, включая 1633 однонуклеотидные замены и 148 инделов.

Найденные несинонимичные генетические варианты распределены по 1296 генам. Из них в 340 генах встречались только гомозиготные генетические варианты (12 инделов, 394 однонуклеотидные замены), в 880 генах встречались только гетерозиготные генетические варианты (119 инделов, 1018 однонуклеотидных замен). Из 1296 генов для 56 генов известна их роль в развитии онкогенеза, согласно каталогу Cancer Gene Census. В 14 генах встречались только гомозиготные, а в 36 генах – только гетерозиготные генетические варианты.

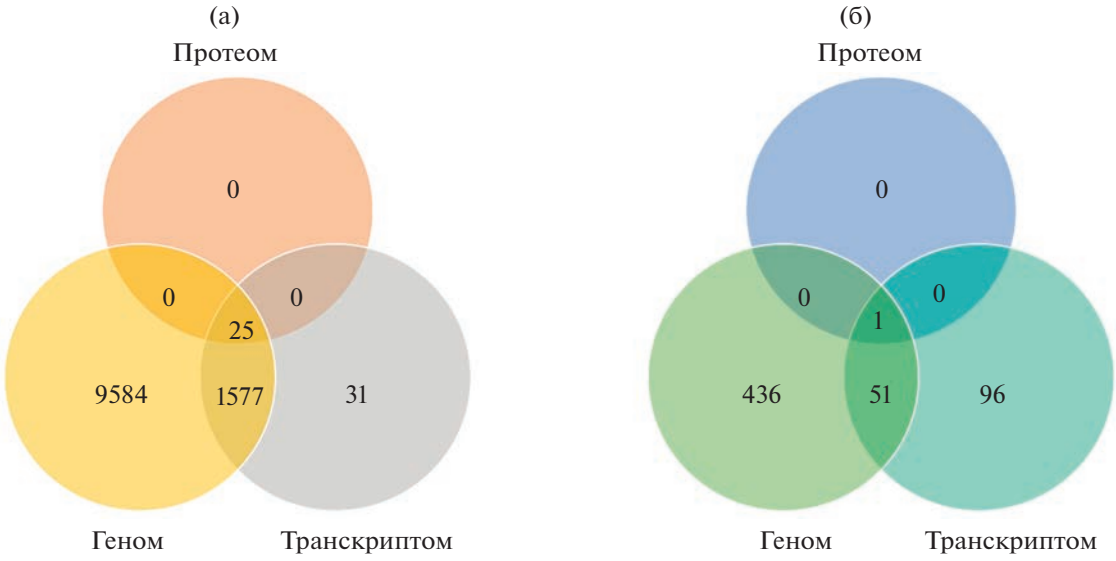

Рис. 1. Выявленные несинонимичные однонуклеотидные замены (а) и инделы (б) по результатам полногеномного, транскриптомного и протеомного профилирования опухолевой клеточной линии HepG2.

Пересечение результатов поиска генетических вариантов согласно полногеномным и транскриптомным данным

Мы сопоставили 11 674 генетических варианта, обнаруженных в результате полногеномного секвенирования, и 1781 генетический вариант, найденный в результате секвенирования транскриптома. Сопоставление проводили для каждого типа отдельно – для однонуклеотидных замен и для инделов.

На рис. 1 представлена диаграмма Венна для nsSNP. Как следует из данных, 98% вариантов, найденных по данным RNASeq, подтверждены по данным WGS. Из общих 1602 вариантов 1128 являются гетерозиготами и 474 представляют собой гомозиготы. Из 31 варианта, детектированного только через RNASeq, 23 – условные гетерозиготы и 8 – условные гомозиготы. Также из этих 31 варианта наиболее распространены замены C→T, G→A и T→C (каждая замена в 6 вариантах). Интересно отметить, что первые два типа замен могут быть объяснены феноменом РНК-редактирования. Остальные замены встречались только 4 раза (G→C) и реже.

Далее мы анализировали пересечение инделов. Как и следовало ожидать, сплайсирование РНК существенно осложняет работу алгоритмов поиска инделов, делая качество поиска последних существенно ниже, по сравнению с исследованием однонуклеотидных замен. Из всех 148 инделов, найденных с помощью RNASeq, только 35% подтверждены данными WGS.

Пересечение результатов поиска генетических вариантов согласно полногеномным, транскриптомным и протеомным данным

В результате протеомного масс-спектрометрического профилирования клеток HepG2 при сопоставлении со стандартной библиотекой канонических и сплайс-опосредованных белковых последовательностей (UniProt) надежно детектировано 1236 белковых продуктов, кодируемых 1206 генами человека. Помимо канонических белковых продуктов, удалось обнаружить пять сплайс-опосредованных вариантов (P35613-2, P17096-2, Q00325-2, P29692-3, O94925-3, Q70UQ0-4). При использовании кастомизированной библиотеки, созданной на основании результатов транскриптомного секвенирования (TPM > 0.01), из 25 141 последовательности, кодируемой 13 681 геном, выявлены 1407 белковых последовательностей, кодируемых 1363 генами.

На протеомном уровне идентифицировано 25 несинонимичных однонуклеотидных замен, из которых на геномном уровне 18 представлены гомозиготами, а 7 – гетерогизотами, а также один индел. Надо отметить, что, несмотря на выполнение критериев достоверности при панорамном масс-спектрометрическом анализе, обнаруженные мутации требуют дополнительной валидации в таргетном режиме.

Анализ опубликованных результатов молекулярного профилирования

Для оценки стабильности генома клеточной линии HepG2 полученные данные о разнообразии мутации и экспрессии транскриптов сопоставлены с опубликованными результатами дру-

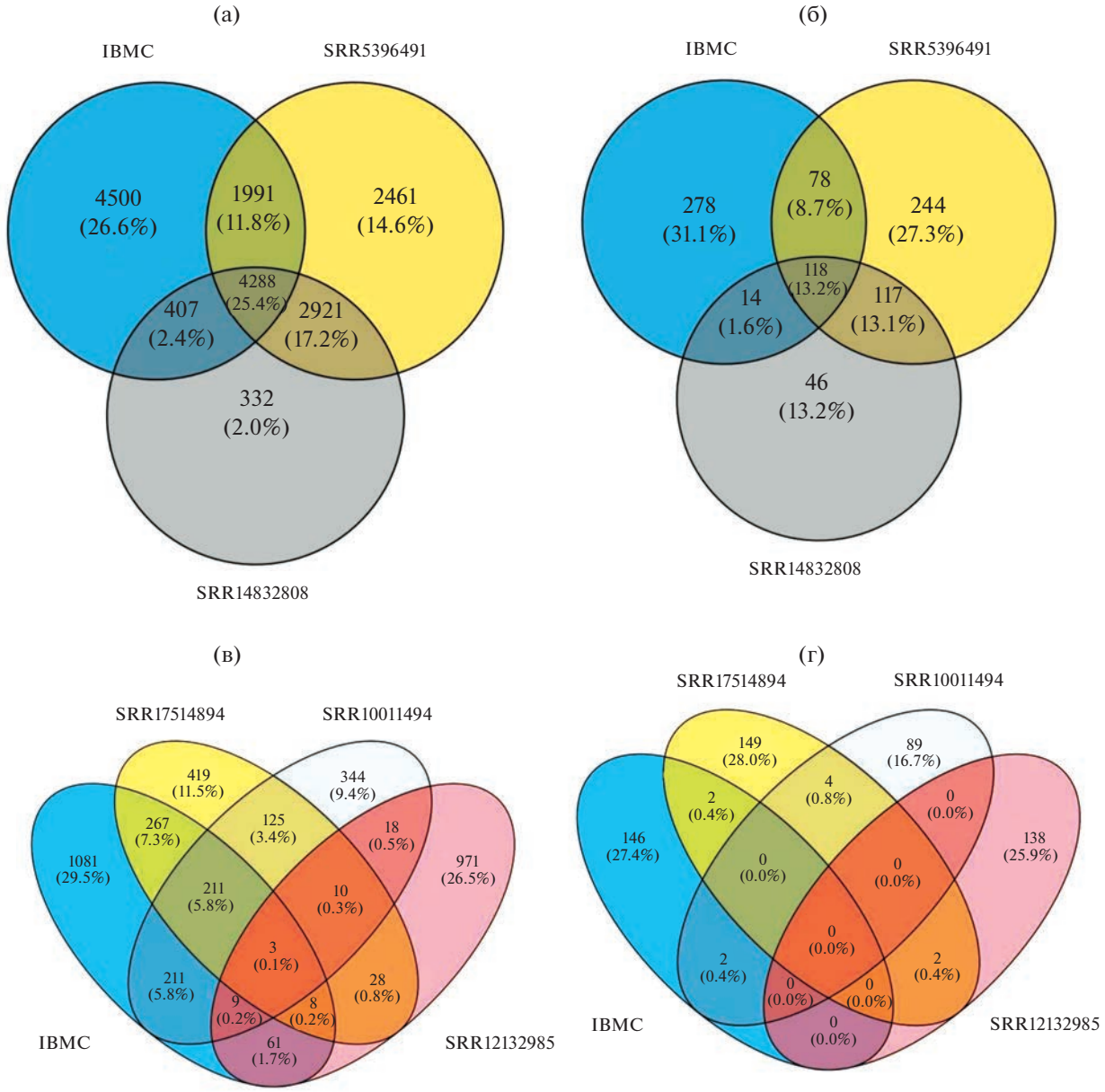

Рис. 2. Диаграмма Венна, отражающая пересечение между наборами данных секвенирования нуклеотидных последовательностей клеточной линии HepG2 (публичные источники). (а) – геномный уровень, точечные нуклеотидные замены; (б) – геномный уровень, инделы; (в) – транскриптомный уровень, точечные нуклеотидные замены; (г) – транскриптомный уровень, инделы. IBMC – данные, полученные в настоящем исследовании. SRR... – данные, полученные в результате экспериментов других научных групп, депонированные в базе SRA NCBI.

гих научных групп. Результаты протеомного анализа позволяют получить существенно меньшее покрытие белок-кодирующих генов, в сравнении с результатами секвенирования. Это объясняется, с одной стороны, ограничениями по чувствительности протеомных аналитических методов (Vavilov et al., 2022), а с другой – тем, что не все протеотипические пептиды могут быть детектированы вследствие физико-химических особенностей, и стандартные подходы для выявления белков с мутациями не всегда являются информативными.

В результате анализа полногеномных данных выявлено суммарно 17805 мутаций, представленных 16900 точечными нуклеотидными заменами (рис. 2а) и 895 инделами (рис. 2б). Среди полученных данных только 38% nsSNP обнаружены во всех образцах, 48% встречаются хотя бы в двух из трех проанализированных образцов, при этом 21% встречается как в исследуемом образце, так и в опубликованных данных. В случае транскриптомного анализа данных характер распределения результатов был похож (рис. 2в), а 24% инделов, об-

наруженных в исследуемом образце, также встречаются и в опубликованных данных (рис. 2г).

Сопоставление транскриптомных данных показало, что общими (для собственных и опубликованных данных) являются 4900 мутаций, что составляет 29% от общего количества на геномном уровне. Во всех четырех образцах выявлены только три несинонимичные однонуклеотидные замены, для исследуемого образца клеток HepG2 62% случаев выявленных замен являются уникальными событиями. Анализ встречаемости инделов показал, что практически все случаи уникальны для конкретного образца.

Необходимость изучения тонких процессов реализации генетической информации в рамках одного образца подтверждается наблюдением, что при исследовании одного образца на транскриптомном уровне реализуется 14% выявленных на геномном уровне мутаций, тогда как при сопоставлении опубликованных данных этот показатель равен 1%. Незначительная доля выявляемых мутаций обусловлена как биологическими причинами (экспрессируется около 65% генов, при этом транскрипты могут не содержать экзонов с мутациями), так и техническими (например, из-за менее точного анализа мутаций на транскриптомном уровне, по сравнению с исследованием генома).

ЗАКЛЮЧЕНИЕ

Отсутствие полной расшифровки генома (то есть путей реализации генома в фенотип) – не только фундаментальная, но и прикладная проблема. Развитие молекулярной биологии и медицины невозможно без целостного представления о функционировании организма, органов и тканей, клеток, без детального описания взаимосвязей генов, кодируемых ими продуктов и протекающих процессов.

На примере данного исследования показана необходимость изучения тонких процессов реализации генетической информации. Полученные сведения наглядно демонстрируют, что разные образцы одной и той же клеточной линии могут иметь разный молекулярный профиль, специфичный по выявляемым мутациям на разных омикс-уровнях. Изучение тонкостей процессов реализации генетической информации требует проведения последовательного комплексного анализа и на этапе становления основ системной биологии должно быть выполнено мультиомное профилирование для каждого исследуемого образца.

БЛАГОДАРНОСТИ

Секвенирование выполнено по заказу компанией Центр генетики и репродуктивной медицины “ГЕНЕТИКО”. Масс-спектрометрический анализ белков, а также биоинформатический анализ выполнен с использованием оборудования и ресурсов Центра коллективного пользования “Протеом человека” при ИБМХ.

ФИНАНСИРОВАНИЕ

Работа была выполнена при поддержке гранта Российского научного фонда № 20-14-00328.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют, что у них нет конфликта интересов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая статья не содержит каких-либо исследований с использованием людей или животных в качестве объектов. Исследуемые клетки HepG2 были получены из Sigma-Aldrich (Merck KGaA, Deutschland).

СПИСОК ЛИТЕРАТУРЫ

Aebersold R., Agar J.N., Amster I.J. et al. How many human proteoforms are there? // Nat. Chem. Biol. 2018. V. 14. № 3. P. 206–214.

Arzumanian V.A., Kiseleva O.I., Poverennaya E.V. The curious case of the HepG2 cell line: 40 years of expertise // Int. J. Mol. Sci. 2021. V. 22. № 23. P. 13135.

Edfors F., Danielsson F., Hallström B.B.M. et al. Gene-specific correlation of RNA and protein levels in human cells and tissues // Mol. Syst. Biol. 2016. V. 12. № 10. P. 883.

GTEx Consortium. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans // Science. 2015. V. 348. № 6235. P. 648–660.

Kiseleva O., Poverennaya E., Shargunov A., Lisitsa A. Proteomic Cinderella: customized analysis of bulky MS/MS data in one night // J. Bioinform. Comput. Biol. 2018. V. 16. № 1. P. 1740011.

Liu Y., Mi Y., Mueller T. et al. Multi-omic measurements of heterogeneity in HeLa cells across laboratories // Nat. Biotechnol. 2019. V. 37. № 3. P. 314–322.

Mellacheruvu D., Wright Z., Couzens A.L. et al. The CRAPome: a contaminant repository for affinity purification – mass spectrometry data // Nat. Methods. 2013. V. 10. № 8. C. 730–736.

Pyatnitskiy M.A., Arzumanian V.A., Radko S.P. et al. Oxford nanopore minION direct RNA-Seq for systems biology // Biology. 2021. V. 10. № 11. P. 1131.

Ponomarenko E.A., Poverennaya E.V., Ilgisonis E.V. et al. The size of the human proteome: the width and depth // Int. J. Anal. Chem. 2016. V. 2016. P. 7436849.

Poverennaya E.V., Ilgisonis E.V., Ponomarenko E.A. et al. Why are the correlations between mRNA and protein levels so low among the 275 predicted protein-coding genes on human chromosome 18? // J. Proteome Res. 2017. V. 16. № 12. P. 4311–4318.

Poverennaya E., Kiseleva O., Romanova A., Pyatnitskiy M. Predicting functions of uncharacterized human proteins: from canonical to proteoforms // Genes. 2020. V. 11. № 6. P. 677.

Smith L.M., Kelleher N.L. Proteoform: a single term describing protein complexity // Nat. Methods. 2013. V. 10. № 3. P. 186–187.

Tenzer S., Leidinger P., Backes C. et al. Integrated quantitative proteomic and transcriptomic analysis of lung tumor and control tissue: a lung cancer showcase // Oncotarget. 2016. V. 7. № 12. P. 14857–14870.

Trivedi U.H., Cézard T., Bridgett S., Montazam A. et al. Quality control of next-generation sequencing data without a reference // Front. Genet. 2014. V. 5. P. 111.

van der Auwera G.A., Carneiro M.O., Hartl C. et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline // Curr. Protoc. Bioinformatics. 2013. V. 43. № 1110. P. 11.10.1–11.10.33.

Vavilov N., Ilgisonis E., Lisitsa A. et al. Number of detected proteins as the function of the sensitivity of proteomic technology in human liver cells // Curr. Protein Pept. Sci. 2022. V. 23. № 4. P. 290–298.

Venter J.C., Adams M.D., Myers E.W. et al. The sequence of the human genome // Science. 2001. V. 291. № 5507. P. 1304–1351.

Vitrinel B., Koh H.W.L., Kar F.M. et al. Exploiting interdata relationships in next-generation proteomics analysis // Mol. Cell. Proteomics. 2019. V. 18. № 8. Suppl. 1. P. S5–S14.

Yip Y.L., Famiglietti M., Gos A. et al. Annotating single amino acid polymorphisms in the UniProt/Swiss-Prot knowledgebase // Hum. Mutat. 2008. V. 29. № 3. P. 361–366.

Non-Synonymous Single-Nucleotide Mutations and Indels: Contribution to the Molecular Postgenome Portrait of the HepG2 Cell Line

E. V. Poverennaya[a], *, O. I. Kiseleva[a], V. A. Arzumanian[a], M. V. Pyatnitskiy[a], I. V. Vakhrushev[a], and E. A. Ponomarenko[a]

[a] *Orekhovich Institute of Biomedical Chemistry, Moscow, Russia*

**e-mail: k.poverennaya@gmail.com*

A comparative analysis of the results of genomic, transcriptomic, and proteomic profiling of HepG2 cell line was carried out in the gene-centric mode. The traceability at the transcriptomic and proteomic levels of changes associated with nonsynonymous single nucleotide substitutions and indels in the genome was shown. Most of the molecular events caused by aberrations at the genomic level are recorded at the transcriptomic level. Only single proteoforms encoded by the selected mutant genes can be reliably detected due to the methodological limitations of proteomic methods, which do not allow the registration of proteoforms present in the sample at low concentrations. The results are consistent with the previously obtained data of other scientific groups and describe the principal methodological solutions required for deciphering the molecular postgenomic portrait of biological samples with a resolution at the level of aberrant molecules.

Keywords: mutations, genome, transcriptome, proteome, proteoforms, molecular profiling