

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

© Коллектив авторов

СПЛАЙС-ВАРИАНТЫ мРНК ГЕНОВ ЦИТОХРОМОВ P450: АНАЛИЗ МЕТОДОМ НАНОПОРОВОГО СЕКВЕНИРОВАНИЯ В ТКАНИ ПЕЧЕНИ ЧЕЛОВЕКА И КЛЕТКАХ ЛИНИИ HepG2

К.А. Дейниченко¹, К.Г. Птицын¹, С.П. Радько^{1}, Л.К. Курбатов¹, И.В. Вахрушев¹,
И.В. Буромский², С.С. Маркин¹, А.И. Арчаков¹, А.В. Лисица¹, Е.А. Пономаренко¹*

¹Научно-исследовательский институт биомедицинской химии им. В.Н. Ореховича,
119121, Москва, ул. Погодинская, 10; *эл. почта: radkos@yandex.ru

²Российский национальный исследовательский медицинский университет им. Н.И. Пирогова, Москва

Методом нанопорового секвенирования проведён анализ транскриптов цитохромов P450 в образцах ткани печени трёх доноров и клетках линии HepG2 (гепатобластома человека). Показано, что прямое секвенирование мРНК с использованием нанопорового секвенатора MinION (Oxford Nanopore Technologies) позволяет получать количественные профили транскриптов (и их сплайс-вариантов) для генов суперсемейства цитохрома P450, кодирующих изоформы, при участии которых метаболизируется большая часть (~80%) лекарственных веществ. Профили сплайс-вариантов существенно различаются для разных доноров. Экспрессия генов цитохромов P450 на уровне транскриптов в клетках линии HepG2 значительно слабее в сравнении с их экспрессией в нормальной ткани печени, что затрудняет исследование альтернативного сплайсинга транскриптов цитохромов P450 в клетках HepG2 методом прямого нанопорового секвенирования мРНК. Профили экспрессии генов суперсемейства цитохрома P450 на уровне транскриптов в ткани печени человека и клетках HepG2 значительно различаются как количественно, так и качественно.

Ключевые слова: сплайс-варианты; нанопоровое секвенирование; цитохромы P450; ткань печени человека; клетки линии HepG2

DOI: 10.18097/PBMC20226802117

ВВЕДЕНИЕ

Альтернативный сплайсинг — формирование нескольких вариантов зрелой мРНК в ходе процессинга одного предшественника мРНК (пре-мРНК) за счёт регулируемого вырезания и соединения экзонов в различных комбинациях (при этом возможно сохранение части интронов и/или потеря части экзонов) — приводит к существенному увеличению разнообразия белков у высших эукариот, позволяя одному гену генерировать несколько различных белковых продуктов [1]. В случае генов человека пре-мРНК около 95% мультиэкзонных генов подвергается альтернативному сплайсингу для кодирования белков с различными функциями [1, 2]. Известно, что ~15% наследственных и раковых заболеваний человека связаны с альтернативным сплайсингом [2]. Альтернативный сплайсинг пре-мРНК играет важную роль в развитии и дифференцировке клеток, а также в регулировании различных метаболических процессов, лежащих в основе их жизнедеятельности [1, 2]. Широкое распространение технологий секвенирования следующего поколения (next generation sequencing, NGS) имело революционное воздействие на исследование альтернативного сплайсинга [3]. Появление в последнее десятилетие секвенаторов третьего поколения, в том числе нанопоровых секвенаторов компании “Oxford Nanopore Technologies” (“ONT”), которые позволяют проводить прямое секвенирование молекул РНК и кДНК с длиной, сопоставимой с длиной полноразмерных транскриптов (секвенирование “длинными прочтениями”), упростило анализ сплайс-вариантов мРНК и существенно повысило надёжность их детекции [4].

В целом существует хорошая корреляция между значениями экспрессии белок-кодирующих генов, полученными с помощью секвенирования “длинными прочтениями” и “короткими прочтениями”: например, коэффициент корреляции между ранжированными значениями экспрессии генов при анализе одного биологического образца платформами ONT MinION и Illumina NovaSeq 6000 составил 0,914 [5]. Тем не менее, секвенирование с помощью “коротких прочтений” часто не способно правильно идентифицировать сплайс-варианты транскриптов одного гена при их реконструкции из “коротких прочтений” из-за схожести изоформ мРНК — недостаток, которого лишено нанопоровое секвенирование, основанное на “длинных прочтениях” [6].

Цитохромы P450 (CYP) образуют суперсемейство ферментов, катализирующих метаболизм большого числа эндогенных и экзогенных соединений. Они участвуют в биотрансформации ксенобиотиков, катализируя, в том числе, значительную долю (около 80%) реакций, связанных с метаболизмом лекарственных веществ [7, 8]. У человека известно 57 генов, кодирующих CYP, которые представлены 18 семействами и 44 подсемействами [9, 10]. Несмотря на значительное число изоформ цитохромов P450 человека, всего 10 изоформ, а именно CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4 и CYP3A5, ответственны за метаболизм более чем 80% лекарственных препаратов, используемых в клинической практике; при этом метаболизм половины из них катализируется всего двумя изоформами — CYP3A4 и CYP2D6 [8].

Большая часть биотрансформации лекарственных веществ происходит в печени, при этом печень также является основной мишенью для образующихся токсичных метаболитов [11, 12]. Несколько клеточных линий, происходящих из злокачественных опухолей печени человека (в первую очередь такие линии как HepG2, Huh7 и HepaRG), используют в настоящее время как модели для проведения исследований механизмов биотрансформации, а также для оценки потенциальной гепатотоксичности существующих и перспективных лекарственных веществ (например, [12-16]). Среди этих линий, HepG2 остаётся одной из широко используемых несмотря на то, что она характеризуется сниженной активностью цитохромов P450 по сравнению с первичными гепатоцитами [16]. Следует отметить, что клеточная линия Huh7 также показывает сниженную активность цитохромов P450 [17]. Другая клеточная линия, HepaRG, хотя и имеет ряд характеристик, близких к характеристикам первичных гепатоцитов (в том числе в отношении экспрессии ряда цитохромов P450 [18]), требует поддержания её в дифференцированном состоянии (линия HepaRG фактически представляет терминально дифференцированные клетки-предшественники гепатоцитов человека) [19]. В отношении линии HepG2 можно также отметить, что она используется как модель в исследованиях процессов, связанных с злокачественной трансформацией гепатоцитов (например, [20, 21]).

Альтернативный сплайсинг пре-мРНК генов суперсемейства цитохрома P450 человека может приводить к формированию почти тысячи уникальных последовательностей мРНК, включая варианты, связанные с удержанием интронов (retained introns) и с образованием преждевременных стоп-кодонов (premature termination codons), которые эффективно удаляются через механизм “нóнсенс-опосредованного распада мРНК” (nonsense-mediated mRNA decay) [10]. Тем не менее, разнообразие сплайс-вариантов мРНК цитохромов P450 может потенциально привести к формированию около пятисот функционально активных белковых продуктов [10]. В рамках данного исследования мы провели анализ сплайс-вариантов мРНК цитохромов P450 в ткани печени человека и клетках линии HepG2 методом нанопорового секвенирования. Ранее подобный анализ не проводился. Установлено, что профили экспрессии сплайс-вариантов транскриптов генов суперсемейства цитохрома P450 характеризуются межиндивидуальной вариабельностью. Исследование альтернативного сплайсинга мРНК генов цитохромов P450 в клетках линии HepG2 методом прямого нанопорового секвенирования мРНК ограничено низким уровнем их экспрессии.

МЕТОДИКА

Биологические образцы и выделение РНК

Для проведения исследования, образцы печени доноров мужского пола (возраст — 65, 38 и 54 лет, далее донор 1, 2 и 3) собирали *post mortem*

при вскрытии. У доноров не было известных заболеваний печени (причиной смерти были острая сердечная недостаточность или несовместимая с жизнью травма). Временной интервал между временем смерти и забором материала не превышал 8 ч. Образцы немедленно помещали в раствор RNeasy RNA Stabilization Reagent (“Thermo Fisher Scientific”, США) и хранили при температуре -20°C до использования. Клетки линии HepG2 (“Merck”, США) культивировали в среде DMEM/F12 с добавлением 10% эмбриональной телячьей сыворотки (“ПанЭко”, Россия), 100 ед/мл пенициллина и 100 мкг/мл стрептомицина (“Thermo Fisher Scientific”). По достижении 80% конfluenceности клетки трипсинизировали добавлением 2 мл 0,25% раствора трипсина (“ПанЭко”) в смеси с 2 мл раствора Версена (“ПанЭко”) и отмывали центрифугированием в фосфатно-солевом растворе (“ПанЭко”). Осадок клеток немедленно использовали для выделения РНК.

Для экстракции суммарной РНК использовали набор RNeasy Mini Kit (“Qiagen”, США) согласно инструкции производителя. Качество РНК оценивали на биоанализаторе Agilent Bioanalyser 2100 (“Agilent Technologies”, США), используя набор Agilent RNA 6000 Nano Kit (“Agilent Technologies”). Значения RIN (RNA Integrity Number) для препаратов суммарной РНК были 7,5 или выше. Выделение мРНК из препаратов суммарной РНК проводили с использованием набора Dynabeads™ mRNA Purification Kit (“Thermo Fisher Scientific”), позволяющим изолировать поли(A)+ РНК. Концентрацию РНК определяли на спектрофотометре NanoDrop 1000 (“Thermo Fisher Scientific”).

Нанопоровое секвенирование

Нанопоровое секвенирование проводили на секвенаторе MinION (“ONT”, Великобритания), используя проточную ячейку FLO-MIN106 (“ONT”). Подготовку ячейки к секвенированию и приготовление библиотек для секвенирования выполняли с использованием набора Direct RNA Sequencing Kit SQK-RNA002 (“ONT”), предназначенного для прямого секвенирования РНК, следуя протоколу производителя. Количество мРНК, использованной для приготовления библиотек, варьировало от 500 до 750 нг. Обратную транскрипцию проводили с помощью ревертазы SuperScript III Reverse Transcriptase (“Thermo Fisher Scientific”), достройку концов и лигирование адаптеров — с помощью набора NEBNext Quick Ligation Module (“New England Laboratories”, Великобритания). Очистку нуклеиновых кислот при приготовлении библиотеки для секвенирования выполняли с помощью магнитных микросфер Agencourt RNAClean XP (“Beckman Coulter”, США). Время секвенирования составляло от 48 ч до 72 ч, объём секвенирования — 0,8 млрд. оснований для клеток HepG2 и варьировал от 0,6 млрд. до 1,5 млрд. оснований для образцов ткани печени человека.

Процедуру перевода сигналов нанопорового секвенатора в нуклеотидные последовательности (base-calling) выполняли с использованием программы guppy_basecaller 3.1.5 (“ONT”) [22] на платформе

облачных вычислений Amazon Web Services Elastic Cloud 2 [23]. Исходные данные секвенирования в формате файлов fast5, генерируемые секвенатором MinION, обрабатывали с использованием виртуальной машины p3.2xlarge (процессоры 8×2,7 ГГц, графический процессор Nvidia V100). При обработке данные фильтровали программой guppy_basecaller с параметром оценки качества >7,0. Последующий контроль качества “прочтений” проводили с помощью скрипта MinIONqc.R. Картирование выполняли, используя программу minimap2 v.2.17 [24] и референсный транскриптом, загруженный с <https://www.encodegenes.org/human> (Release 32 (GRCh38.p13)). При картировании учитывались “прочтения”, покрывающие сплайс-вариант транскрипта как полностью, так и частично. Количественную оценку представленности транскриптов проводили с помощью программы Salmon 0.12/1.1.0 [25], используя инструмент salmon quant программного пакета SAMtools. Метрику TPM (transcripts per million) использовали для количественной оценки представленности транскриптов.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Анализ транскриптов, кодируемых генами цитохромов P450, в ткани печени человека

Количество “прочтений” при нанопоровом секвенировании транскриптома ткани печени составляло около 1,8 млн., 0,7 млн. и 1,0 млн. для доноров 1, 2 и 3 соответственно. В каждом случае более трёх четвертей “прочтений” было успешно картировано на референсный транскриптом. Число транскриптов с ненулевым значением TPM составило 30104 (донор 1), 21882 (донор 2) и 24324 (донор 3). Для донора 1 последующий анализ выявил транскрипты 41 гена цитохромов P450, а для доноров 2 и 3 — 37 генов у каждого (дополнительные материалы, табл. S1). Более высокое число генов цитохромов P450 с детектированной экспрессией в случае донора 1 может быть в определенной степени связано с большим объемом (и, соответственно, глубиной) секвенирования по сравнению с донорами 2 и 3. Действительно, для трёх генов, чьи транскрипты обнаружены в ткани печени донора 1, но не 2 и 3 (*CYP11A1*, *CYP21A2*

и *CYP46A1*), уровень представленности транскриптов низок (TPM = 0,439, 0,385 и 0,197 соответственно; дополнительные материалы, табл. S1). В то же время экспрессия гена *CYP17A1* не обнаружена для донора 3, но обнаружена для доноров 1 и 2, причём уровень его экспрессии в ткани печени донора 2 почти в 10 раз превышает уровень для донора 1, несмотря на больший объём секвенирования (дополнительные материалы, табл. S1). Для гена *CYP2A13*, экспрессия детектировалась на относительно заметном уровне для донора 2 (с наименьшим объёмом секвенирования), но не детектировалась для доноров 1 и 3. Таким образом, число генов цитохромов P450, чьи транскрипты обнаруживаются методом нанопорового секвенирования, определяется как чувствительностью детекции (объёмом или глубиной секвенирования), так и уровнями их экспрессии, которые могут значительно различаться у доноров. Следует отметить, что хотя в рассмотрение принимались все транскрипты со значениями TPM больше нуля, среди транскриптом генов цитохромов P450 (включая сплайс-варианты) наименьшее значение TPM составило 0,145 (дополнительные материалы, табл. S1). Это значение выше 0,1 — значения, часто принимаемого в качестве порога отсека при нанопоровом секвенировании [5].

Для 34 генов суперсемейства цитохрома P450, экспрессия на уровне транскриптов была обнаружена в ткани печени всех доноров (дополнительные материалы, табл. S2). Парное сравнение лог-трансформированных значений представленности (в TPM) транскриптов этих генов (рис. 1) показало относительно высокую степень корреляции между донорами 1 и 2 и донорами 1 и 3 (коэффициент корреляции Пирсона r_p равен 0,897 и 0,887 соответственно), но заметно меньшую корреляцию между донорами 2 и 3 ($r_p = 0,700$). Среди пятёрки генов, характеризующихся наиболее высокой представленностью транскриптов у данного донора (дополнительные материалы, табл. S2), только 3 гена совпадают у всех доноров (*CYP2E1*, *CYP2C9*, *CYP4A11*), при этом два гена совпадают для доноров 1 и 2 (*CYP2A6*) и 1 и 3 (*CYP3A4*), а ещё два гена (*CYP1A2* и *CYP2A7*) входят в соответствующие пятёрки у доноров 3 и 2.

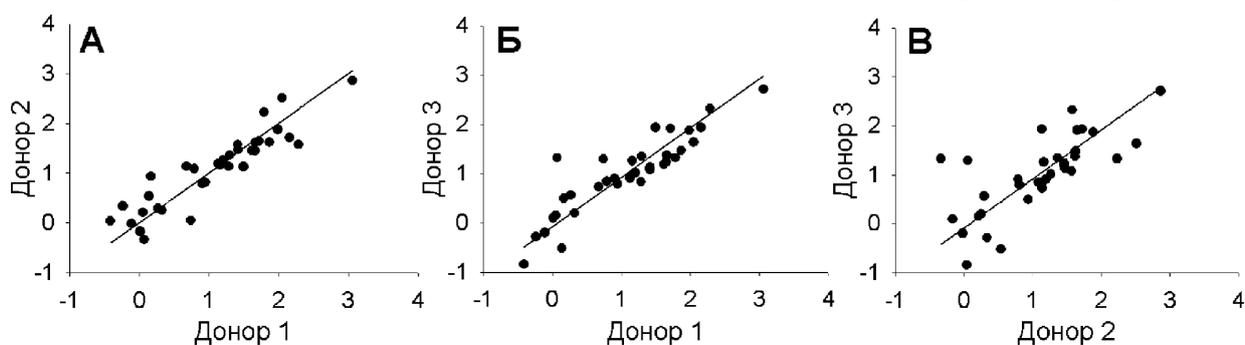


Рисунок 1. Корреляции между лог-трансформированными (по основанию 10) значениями представленности (в TPM) транскриптов генов суперсемейства цитохрома P450, обнаруженных методом нанопорового секвенирования во всех образцах ткани печени человека. Парное сравнение: **А** — доноры 1 и 2, **Б** — доноры 1 и 3, **В** — доноры 2 и 3. Значения коэффициента корреляции Пирсона: 0,897 (**А**), 0,887 (**Б**), 0,700 (**В**).

Таким образом, количественно экспрессия генов цитохромов P450 действительно значительно различается у доноров. Вероятно, эти различия отражают индивидуальную реакцию на воздействие неизвестных нам, но несомненно присутствующих разнообразных факторов, способных индуцировать или супрессировать экспрессию различных генов суперсемейства цитохрома P450.

Известно, что лекарственные вещества метаболизируются в печени с участием 17 изоформ цитохромов P450 — CYP1A1/2, CYP1B1, CYP2A6, CYP2B6, CYP2C8/9/18/19, CYP2D6, CYP2E1, CYP2J2, CYP3A4/5/7/43 и CYP4F2 [26, 27], средняя представленность которых варьирует в широком диапазоне — от 93 пмоль/мг ткани (CYP3A4) до 0,4 пмоль/мг ткани (CYP2C18) [27]. При этом масс-спектрометрический анализ микросом, полученных из образцов ткани печени человека, позволил выявить 8 из этих 17 изоформ (CYP1B1, CYP2A6, CYP2C8/9, CYP2D6, CYP2E1, CYP3A4 и CYP4F2), а также дополнительно изоформу CYP4A11 [28]. Метод нанопорового секвенирования идентифицировал транскрипты, кодирующие все упомянутые выше 18 изоформ, по меньшей мере в образцах ткани печени двух доноров одновременно, при этом транскрипты, кодирующие изоформы CYP3A43 и CYP2J2, не обнаружены лишь у донора 2 и донора 1 соответственно (дополнительные материалы, табл. S1 и S2).

Альтернативный сплайсинг мРНК цитохромов P450 в ткани печени человека

На рисунке 2 представлено распределение числа генов цитохромов P450, чьи транскрипты были идентифицированы в ткани печени человека при обработке данных нанопорового секвенирования, и соответствующего суммарного количества сплайс-вариантов транскриптов для трех доноров. Для донора 1 количество детектированных сплайс-вариантов транскриптов варьировало от 1 до 5, а для доноров 2 и 3 — от 1 до 4 (дополнительные материалы, табл. S1). Среднее число сплайс-вариантов на транскрипт варьировало от 1,7 для донора 2 до 1,9 для донора 1 (рис. 2). Более высокое число сплайс-вариантов, выявляемое для донора 1, вероятней всего, связано с большим объемом (глубиной) секвенирования, что позволяет детектировать низкопредставленные сплайс-варианты транскриптов. В среднем по донорам на каждый транскрипт, кодирующий отдельную изоформу цитохрома P450, приходится 1,8 сплайс-варианта, что сопоставимо со средним числом сплайс-вариантов транскриптов для белок-кодирующих генов, выявленным в рамках данного исследования, которое лежало в пределах от 1,5 (донор 3) до 1,7 (донор 1). Это указывает на то, что в ткани печени человека разнообразие сплайс-вариантов транскриптов генов суперсемейства цитохрома P450 в результате альтернативного сплайсинга пре-мРНК сопоставимо с общим разнообразием сплайс-вариантов транскриптов белок-кодирующих генов (или, возможно, незначительно превышает его).

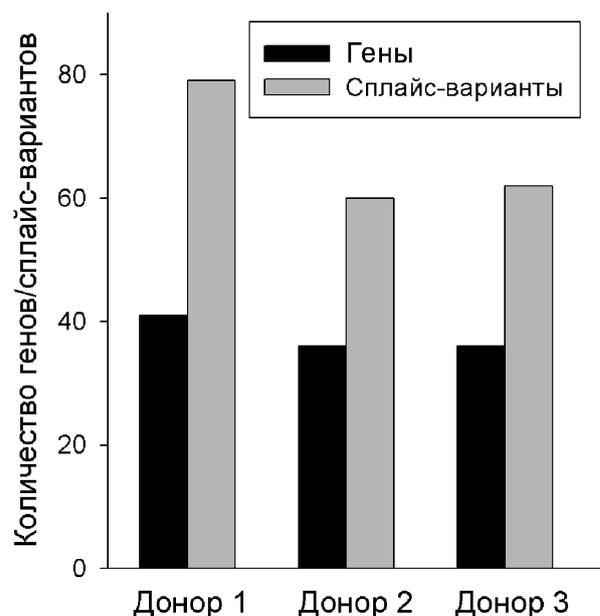


Рисунок 2. Гистограмма распределения числа генов цитохромов P450 (черные столбцы) и суммарного (по всем транскриптам) числа соответствующих сплайс-вариантов, чья экспрессия была детектирована методом нанопорового секвенирования (серые столбцы), по донорам.

Распределение числа сплайс-вариантов транскрипта по 34 генам, чья экспрессия была детектирована у всех доноров, представлено на рисунке 3 для доноров 1, 2 и 3. Для удобства сравнения гены ранжированы по возрастанию количества сплайс-вариантов транскрипта у донора 1. Следует отметить, что в случае донора 1 суммарное количество сплайс-вариантов, выявленных для данной группы из 34 генов, было заметно больше, чем в случае доноров 2 и 3 (72, 57 и 59 соответственно, дополнительные материалы, табл. S2), что, как уже упоминалось выше, может быть связано с существенно большим объемом секвенирования. Это затрудняет прямое сравнение распределения для донора 1 с соответствующими распределениями для доноров 2 и 3 (рис. 3). Для доноров 2 и 3 с сопоставимым числом выявленных сплайс-вариантов профили распределения сплайс-вариантов по генам различаются. Например, несмотря на меньший объем секвенирования для донора 2, для генов *CYP8B1*, *CYP4F12*, *CYP4F3*, *CYP4A11* и *CYP2D7* число сплайс-вариантов соответствующих транскриптов выше, чем у донора 3 (рис. 3). Таким образом, существует различие между донорами не только по количественной экспрессии тех или иных генов суперсемейства цитохрома P450, но и в качественном отношении — по представленности разных сплайс-вариантов. Это подтверждает важную роль альтернативного сплайсинга в обеспечении индивидуального разнообразия ферментов, ответственных за биотрансформацию ксенобиотиков, через разнообразие вариантов транскриптов, каждый из которых может кодировать соответствующий вариант изоформы цитохрома P450.

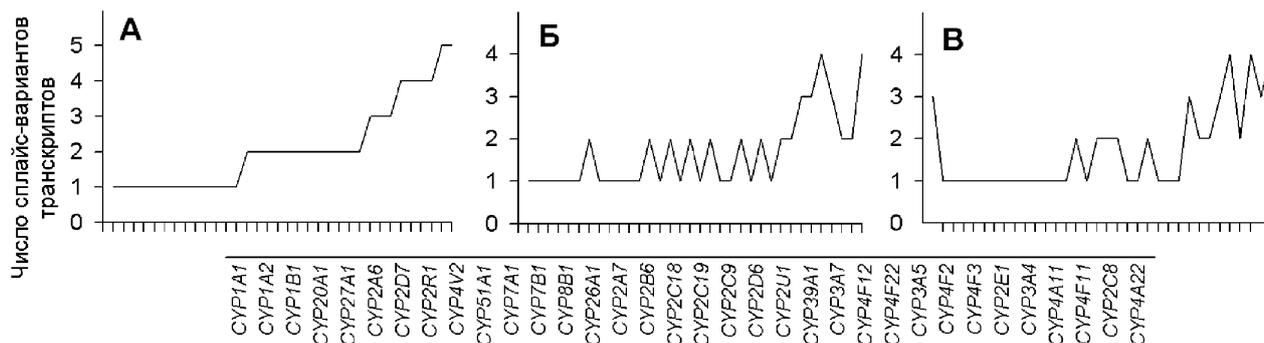


Рисунок 3. Профили альтернативного сплайсинга (количество сплайс-вариантов на транскрипт) для образцов ткани печени донора 1 (панель А), донора 2 (панель Б) и донора 3 (панель В). Представлены транскрипты цитохромов P450, детектированные у всех доноров. Расположение генов цитохромов P450 на осях абсцисс панелей указано под горизонтальной чертой. Гены в ряду ранжированы по возрастанию уровня их экспрессии (в TPM) для донора 1.

Как упоминалось во введении, около половины лекарственных препаратов метаболизируется двумя изоформами, CYP3A4 и CYP2D6 [7], которые входят в число изоформ цитохромов P450, представленных в печени человека [26, 27]. Транскрипты, кодирующие эти изоформы, были обнаружены методом нанопорового секвенирования в образцах ткани печени всех доноров (дополнительные материалы, табл. S2). Для гена *CYP3A4* анализ выявил у всех доноров присутствие четырёх сплайс-вариантов мРНК (CYP3A4-201, -202, -206 и -208, таблица 1) из шести известных (www.proteinatlas.org/ENSG00000160868-CYP3A4), образованных в результате *цис*-сплайсинга, а также присутствие у одного из доноров сплайс-варианта CYP3A43-204 (донор 3) или CYP3A43-205 (донор 1), являющихся продуктами транс-сплайсинга пре-мРНК генов *CYP3A4* и *CYP3A43* (www.proteinatlas.org/ENSG0000021461-CYP3A43). Тогда как сплайс-вариант CYP3A43-205 кодирует предсказанный белок, который до настоящего времени не обнаружен в тканях и/или клетках человека (имеет статус PE3 — “Protein inferred by homology”), остальные сплайс-варианты кодируют белковые продукты, которые имеют статус “Experimental evidence at protein level” (PE1). Обнаруженные сплайс-варианты транскриптов гена *CYP3A4* могут транслироваться в варианты изоформы CYP3A4 с различной биотрансформирующей активностью. Как видно из таблицы 1, количественный профиль данных сплайс-вариантов различается у доноров, что, наряду с интегральной экспрессией данной изоформы, потенциально может предопределять индивидуальный ответ человеческого организма на приём лекарств.

Для гена *CYP2D6* анализ образцов ткани печени человека методом нанопорового секвенирования выявил два сплайс-варианта транскриптов из четырёх известных (www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000100197;r=22:42126499-42130865) — CYP2D6-201 и CYP2D6-203, оба из которых обнаружены у донора 1, при этом у доноров 2 и 3 встречается только по одному из этих сплайс-вариантов (табл. 1). Сплайс-вариант CYP2D6-201 кодирует белок, который имеет статус PE1 (www.proteinatlas.org/ENSG00000100197-CYP2D6),

в то время как другой обнаруженный сплайс-вариант CYP2D6-203 аннотирован в базе данных Ensembl как “retained intron”, (www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000100197;r=22:42126499-42130865). Хотя сплайс-варианты транскриптов, имеющие в своем составе последовательности интронов (биотип “retained intron”), могут в принципе транслироваться в функционально активные белки [29], данный сплайс-вариант аннотирован как не дающий белкового продукта. Оставшиеся два известных сплайс-варианта транскриптов, CYP2D6-202 и CYP2D6-204, которые не были детектированы в нашем исследовании, представляют белок-кодирующий транскрипт (CYP2D6-204) и транскрипт, образованный преждевременным стоп-кодоном (premature termination codon). В последнем случае мРНК быстро подвергается деградации при помощи механизма, известного как “nonsense-mediated mRNA decay” [10], и её эффективная детекция может быть затруднена. Сплайс-вариант CYP2D6-202 либо экспрессирован на уровне, недостаточном для его детекции в нашем исследовании, либо его экспрессия является тканезависимой (*CYP2D6* также экспрессирован у человека в тонком кишечнике и двенадцатипёрстной кишке (www.proteinatlas.org/ENSG00000100197-CYP2D6/tissue), хотя и на значительно более низком уровне, чем в печени). Дополнительно к сплайс-вариантам, указанным в таблице 1, нанопоровое секвенирование позволило также детектировать сплайс-вариант CYP2D6-205, являющийся продуктом псевдогена *CYP2D6* (www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000272532;r=CHR_HSCR22_2_CTG1:42126481-42130888), который был обнаружен у всех доноров (значения TPM равнялись 2,225, 15,375 и 1,762 для доноров 1, 2 и 3 соответственно).

Анализ транскриптов, кодируемых генами цитохромов P450, в клетках линии HepG2

В результате нанопорового секвенирования транскриптома клеток линии HepG2 (гепатобластома человека [16]) было получено около 0,8 млн “прочтений”, почти 80% которых было успешно картировано на референсный транскриптом.

АНАЛИЗ СПЛАЙС-ВАРИАНТОВ мРНК ГЕНОВ ЦИТОХРОМОВ P450

Таблица 1. Сплайс-варианты транскриптов, кодируемых генами *CYP3A4* и *CYP2D6*, а также сплайс-варианты транскриптов гена *CYP3A43*, которые имеют в своём составе копии экзонов гена *CYP3A4* как результат транс-сплайсинга

Сплайс-вариант	Название гена	Длина (нт)	Биотип	TPM*		
				Донор 1	Донор 2	Донор 3
CYP3A4-201	<i>CYP3A4</i>	2165	Белок-кодирующий (protein coding)	18,223	5,184	25,407
CYP3A4-202		2342		50,931	8,891	89,633
CYP3A4-206		2781		44,435	8,778	71,893
CYP3A4-208		1982		80,365	15,586	23,743
CYP3A43-204	<i>CYP3A43</i>	1709		–	–	1,483
CYP3A43-205		990		0,519	–	–
CYP2D6-201	<i>CYP2D6</i>	1433		–	13,194	–
CYP2D6-203		1714		Retained intron	30,717	–

Примечание: * “–” — сплайс-вариант не найден.

Число транскриптов с ненулевым значением TPM составило 28386, среди них было выявлено 16 транскриптов, кодируемых генами суперсемейства цитохрома P450 (дополнительные материалы, табл. S1). Уровень экспрессии был существенно ниже, чем в случае образцов ткани печени человека. Если в ткани печени максимальные значения TPM составляли 1147, 731 и 515 для доноров 1, 2 и 3 соответственно (продукт гена *CYP2E1*), то в клетках HepG2 максимальное значение было менее 20 (продукт гена *CYP24A1*, дополнительные материалы, табл. S1). Большая часть транскриптов, экспрессированных в HepG2 (11 из 16), была детектирована в ткани печени всех доноров, один транскрипт (продукт гена *CYP27B1*) был обнаружен у донора 1. Для четырёх генов семейства цитохрома P450, а именно *CYP24A1*, *CYP2W1*, *CYP2S1* и *CYP19A1*, экспрессия на уровне транскриптов была детектирована только в клетках HepG2, но не в ткани печени, причём по уровню экспрессии эти гены занимают 1, 2, 5 и 8 места соответственно в ряду генов цитохромов P450, экспрессированных в клетках HepG2 (дополнительные материалы, табл. S1).

Таким образом, как в количественном, так и в качественном отношении экспрессия генов цитохрома P450, выявляемая методом нанопорового секвенирования в ткани печени и клетках HepG2, существенно различается. Как и в случае сравнения доноров, качественные различия в профиле экспрессированных генов могут отражать соотношение между уровнем их количественной экспрессии и чувствительностью детекции транскриптов при анализе конкретного образца методом нанопорового секвенирования. Тем не менее, количество “прочтений” (и, соответственно, чувствительность детекции низкопредставленных транскриптов) была в случае секвенирования клеток HepG2 выше, чем в случае образца ткани печени донора 2 (0,8 млн. vs. 0,7 млн.), но различия как по количественным, так и качественным характеристикам профилей экспрессии генов суперсемейства цитохрома P450 между тканью печени донора 2 и клетками HepG2 принципиально были такими же, как и в случае сравнения с донорами 1 и 3. Ни транскрипты гена *CYP3A4*, ни *CYP2D6*,

чья белковые продукты ответственны у человека за метаболизм значительной части лекарственных веществ, в клетках HepG2 на данном уровне чувствительности не детектировались.

Наибольшее число сплайс-вариантов транскрипта (а именно три) было обнаружено в клетках HepG2 для гена *CYP24A1*, который у них также является наиболее высоко экспрессированным (при этом его экспрессия не была обнаружена ни у одного донора, дополнительные материалы, табл. S1). Экспрессия одиннадцати генов, *CYP1A1*, *CYP3A5*, *CYP3A7*, *CYP4F2*, *CYP4F3*, *CYP4F11*, *CYP4F12*, *CYP8B1*, *CYP20A1*, *CY27A1*, *CYP51A1*, была выявлена как в образцах печени, так и в клетках линии HepG2 (табл. 2). При этом для семи генов из одиннадцати, *CYP3A7*, *CYP4F2*, *CYP4F3*, *CYP4F12*, *CYP8B1*, *CY27A1*, *CYP51A1*, в клетках HepG2 детектировался сплайс-вариант транскрипта, который был наиболее представленным в ткани печени (за исключением генов *CY27A1* и *CYP51A1*, для которых был обнаружен только один сплайс-вариант в обоих случаях). Нельзя исключить, что в этих случаях наблюдаемое отсутствие других сплайс-вариантов связано просто с тем, что их представленность лежит ниже уровня чувствительности детекции в силу более низкой экспрессии генов цитохромов P450 в целом в клетках HepG2 по сравнению с тканью печени. Для трёх генов, *CYP1A1*, *CYP3A5*, *CYP4F11*, как количественная, так и качественная представленность сплайс-вариантов *per se* заметно варьировала между донорами (табл. 2), что затрудняет оценку различий в характере их экспрессии в ткани печени и клетках HepG2. Только для одного гена из одиннадцати, *CYP20A1*, наблюдалось очевидное различие в экспрессии в отношении сплайс-вариантов — тогда как в ткани печени экспрессирован только вариант *CYP20A1-210* у всех доноров, в клетках HepG2 экспрессирован другой вариант, *CYP20A1-203*, причём уровень его экспрессии выше уровней для варианта *CYP20A1-210* в ткани печени (табл. 2). Таким образом, для генов цитохромов P450, чья экспрессия детектируется как в клетках HepG2, так и в ткани печени, состав сплайс-вариантов в большинстве случаев существенно не различается.

Таблица 2. Представленность (в TPM*) сплайс-вариантов транскриптов генов суперсемейства цитохрома P450, обнаруженных как в ткани печени человека, так и в клетках линии HepG2

Сплайс-вариант	Идентификатор (Ensembl)	Длина (нт)	Донор 1	Донор 2	Донор 3	HepG2
CYP1A1-201	ENST00000379727.8	2353	–	0,464	–	–
CYP1A1-202	ENST00000395048.6	2358	–	–	6,071	–
CYP1A1-209	ENST00000612821.4	2152	–	–	10,778	3,186
CYP1A1-210	ENST00000617691.4	2271	1,185	–	4,146	1,162
CYP3A5-201	ENST00000222982.8	1470	0,636	2,540	3,654	0,587
CYP3A5-202	ENST00000339843.6	3070	1,516	–	2,803	–
CYP3A5-203	ENST00000439761.3	4223	5,847	3,622	1,550	–
CYP3A7-201	ENST00000336374.4	1829	1,795	1,792	1,579	0,472
CYP3A7-205	ENST00000620220.4	1454	0,318	–	–	–
CYP4F2-201	ENST00000011989.11	1817	1,124	4,453	0,755	–
CYP4F2-202	ENST00000221700.11	2111	15,687	32,653	11,156	0,818
CYP4F2-206	ENST00000589654.2	249	9,270	–	–	–
CYP4F3-201	ENST00000221307.13	4803	12,749	15,413	9,435	0,539
CYP4F3-203	ENST00000586182.6	2088	2,634	1,860	–	–
CYP4F3-206	ENST00000591058.5	2733	0,637	1,262	0,911	–
CYP4F11-201	ENST00000248041.12	2727	–	–	3,560	1,308
CYP4F11-202	ENST00000326742.12	2627	2,672	–	1,994	2,257
CYP4F11-203	ENST00000402119.8	2911	0,826	4,596	0,823	–
CYP4F11-204	ENST00000591841.1	2281	2,431	–	–	–
CYP4F11-205	ENST00000620614.4	2658	7,530	10,995	1,694	–
CYP4F12-201	ENST00000324632.10	1437	2,511	10,056	–	0,742
CYP4F12-215	ENST00000550308.6	1464	2,222	3,737	5,386	–
CYP8B1-201	ENST00000316161.6	3438	41,259	27,962	15,319	0,251
CYP8B1-202	ENST00000437102.1	1646	–	0,664	–	–
CYP20A1-203	ENST00000429815.6	1562	–	–	–	3,869
CYP20A1-210	ENST00000613925.4	1736	1,033	0,688	1,247	–
CYP27A1-201	ENST00000258415.9	1645	19,836	23,266	22,178	6,402
CYP51A1-201	ENST00000003100.13	2905	1,475	8,649	3,131	0,594

Примечание: * “–” — сплайс-вариант не найден.

ЗАКЛЮЧЕНИЕ

Анализ транскриптома ткани печени человека методом нанопорового секвенирования позволяет обнаружить транскрипты изоформ суперсемейства цитохрома P450, при участии которых метаболизируется большая часть (~80%) лекарственных веществ в печени, а также исследовать межиндивидуальные различия в качественном и количественном профиле сплайс-вариантов транскриптов цитохромов P450. Профиль сплайс-вариантов отличается для разных доноров, демонстрируя межиндивидуальную вариабельность. Экспрессия генов цитохромов P450 на уровне транскриптов в клетках линии HepG2 значительно ниже в сравнении с их экспрессией в нормальной ткани печени. Это лимитирует исследование альтернативного сплайсинга транскриптов

цитохромов P450 в клетках HepG2 только их наиболее представленными сплайс-вариантами. Исследование всего разнообразия сплайс-вариантов транскриптов цитохромов P450 в данной клеточной линии требует существенного увеличения объема (глубины) секвенирования, которое может быть достигнуто в рамках использованной технологии секвенирования переходом от прямого секвенирования молекул мРНК (данное исследование) к секвенированию кДНК с её предварительной ПЦР-амплификацией.

ФИНАНСИРОВАНИЕ

Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021-2030 годы) (№122030100168-2).

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Сбор образцов ткани печени человека был одобрен Этическим комитетом Российского государственного медицинского университета им. Н.И. Пирогова (протокол № 3 от 15.03.2018 г.) и проводился при информированном согласии представителей доноров.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

Дополнительные материалы доступны в электронной версии статьи на сайте журнала (pbmc.ibmc.msk.ru).

ЛИТЕРАТУРА

- Nilsen T.W., Graveley B.R. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature*, **463**(7280), 457-463. DOI: 10.1038/nature08909
- Jiang W., Chen L. (2021) Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. *Comput. Struct. Biotechnol. J.*, **19**, 183-195. DOI: 10.1016/j.csbj.2020.12.009
- de Klerk E., Hoen P.A. (2015) Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends Genet.*, **31**(3), 128-139. DOI: 10.1016/j.tig.2015.01.001
- Радько С.П., Курбатов Л.К., Птицын К.Г., Киселева Я.Ю., Пономаренко Е.А., Лисица А.В., Арчаков А.И. (2018) Перспективы использования секвенаторов третьего поколения для количественного профилирования транскриптома. *Biomed. Chem.: Res. Meth.*, **1**(4), e00086. [Radko S.P., Kurbatov L.K., Ptitsyn K.G., Kiseleva Y.Y., Ponomarenko E.A., Lisitsa A.V., Archakov A.I. (2018) Prospects for the use of third generation sequencers for quantitative profiling of transcriptome. *Biomed. Chem.: Res. Meth.*, **1**(4), e00086.] DOI: 10.18097/bmcrm00086
- Pyatnitskiy M.A., Arzumanian V.A., Radko S.P., Ptitsyn K.G., Vakhrushev I.V., Poverennaya E.V., Ponomarenko E.A. (2021) Oxford nanopore MinION Direct RNA-Seq for systems biology. *Biology (Basel)*, **10**(11), 1131. DOI: 10.3390/biology10111131
- de Paoli-Iseppi R., Gleeson J., Clark M.B. (2021) Isoform age — splice isoform profiling using long-read technologies. *Front. Mol. Biosci.*, **8**, 711-733. DOI: 10.3389/fmolb.2021.711733
- Danielson P.B. (2002) The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Curr. Drug Metab.*, **3**(6), 561-597. DOI: 10.2174/1389200023337054
- Zhao M., Ma J., Li M., Zhang Y., Jiang B., Zhao X., Huai C., Shen L., Zhang N., He L., Qin S. (2021) Cytochrome P450 enzymes and drug metabolism in humans. *Int. J. Mol. Sci.*, **22**(23), 12808. DOI: 10.3390/ijms222312808
- Sim S.C., Ingelman-Sundberg M. (2013) Update on allele nomenclature for human cytochromes P450 and the human cytochrome P450 allele (CYP-allele) nomenclature database. *Methods Mol. Biol.*, **987**, 251-259. DOI: 10.1007/978-1-62703-321-3_21
- Annalora A.J., Marcus C.B., Iversen P.L. (2017) Alternative splicing in the cytochrome P450 superfamily expands protein diversity to augment gene function and redirect human drug metabolism. *Drug Metab. Dispos.*, **45**(4), 375-389. DOI: 10.1124/dmd.116.073254
- Almazroo O.A., Miah M.K., Venkataramanan R. (2017) Drug metabolism in the liver. *Clin. Liver Dis.*, **21**(1), 1-20. DOI: 10.1016/j.cld.2016.08.001
- Gupta R., Schrooders Y., Hauser D., van Herwijnen M., Albrecht W., Ter Braak B., Brecklinghaus T., Castell J.V., Elenschneider L., Escher S., Guye P., Hengstler J.G., Ghallab A., Hansen T., Leist M., Maclennan R., Moritz W., Tolosa L., Tricot T., Verfaillie C., Walker P., van de Water B., Kleinjans J., Caiment F. (2021) Comparing *in vitro* human liver models to *in vivo* human liver using RNA-Seq. *Arch. Toxicol.*, **95**(2), 573-589. DOI: 10.1007/s00204-020-02937-6
- Broekman M.M., Roelofs H.M., Wong D.R., Kerstholt M., Leijten A., Hoentjen F., Peters W.H., Wanten G.J., de Jong D.J. (2015) Allopurinol and 5-aminosalicylic acid influence thiopurine-induced hepatotoxicity *in vitro*. *Cell. Biol. Toxicol.*, **31**(3), 161-171. DOI: 10.1007/s10565-015-9301-1
- Waldherr M., Misik M., Ferk F., Tomc J., Zegura B., Filipic M., Mikulits W., Mai S., Haas O., Huber W.W., Haslinger E., Knasmuller S. (2018) *Arch. Toxicol.*, **92**(2), 921-934. DOI: 10.1007/s00204-017-2109-4
- Romualdo G.R., Leroy K., Costa C.J.S., Prata G.B., Vanderborgh B., da Silva T.C., Barbisan L.F., Andraus W., Devisscher L., Câmara N.O.S., Vinken M., Cogliati B. (2021) Use of HuH6 and other human-derived hepatoma lines for the detection of genotoxins: a new hope for laboratory animals? *Cancers (Basel)*, **13**(21), 5583. DOI: 10.3390/cancers13215583
- Arzumanian V.A., Kiseleva O.I., Poverennaya E.V. (2021) The curious case of the HepG2 cell line: 40 years of expertise. *Int. J. Mol. Sci.*, **22**(23), 13135. DOI: 10.3390/ijms222313135
- Choi S., Sainz B. Jr., Corcoran P., Uprichard S., Jeong H. (2009) Characterization of increased drug metabolism activity in dimethyl sulfoxide (DMSO)-treated Huh7 hepatoma cells. *Xenobiotica*, **39**(3), 205-217. DOI: 10.1080/00498250802613620
- Kanebratt K.P., Andersson T.B. (2008) Evaluation of HepaRG cells as an *in vitro* model for human drug metabolism studies. *Drug Metab Dispos.*, **36**(7), 1444-1452. DOI: 10.1124/dmd.107.020016
- Gripon P., Rumin S., Urban S., Le Seyec J., Glaise D., Cannie I., Guyomard C., Lucas J., Trepo C., Guguen-Guillouzo C. (2002) Infection of a human hepatoma cell line by hepatitis B virus. *Proc. Natl. Acad. Sci. USA*, **99**(24), 15655-15660. DOI: 10.1073/pnas.232137699
- Ren Y., Wang Y., Hao S., Yang Y., Xiong W., Qiu L., Tao J., Tang A. (2020) BMP2 secretion from hepatocellular carcinoma cell HepG2 enhances angiogenesis and tumor growth in endothelial cells via activation of the MAPK/p38 signaling pathway. *J. Cancer.*, **11**(23), 6939-6949. DOI: 10.7150/jca.48100
- Chinnadurai R., Porter A.P., Patel M., Lipat A.J., Forsberg M.H., Rajan D., Hematti P., Capitini C.M., Bruker C. (2021) Hepatocellular carcinoma cells are protected from immunolysis by mesenchymal stromal cells through indoleamine 2,3 dioxygenase. *Front. Cell. Dev. Biol.*, **9**, 715905. DOI: 10.3389/fcell.2021.715905
- Wick R.R., Judd L.M., Holt K.E. (2019) Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.*, **20**(1), 129. DOI: 10.1186/s13059-019-1727-y

23. Шаповалова В.В., Радько С.П., Птицын К.Г., Краснов Г.С., Наход К.В., Конаш О.С., Виноградина М.А., Пономаренко Е.А., Дружиловский Д.С., Лисица А.В. (2020) Обработка длинных чтений транскриптомного секвенирования на облачной вычислительной платформе amazon web services. *Biomed. Chem.: Res. Meth.*, **3**(4), e00131. [Shapovalova V.V., Radko S.P., Ptitsyn K.G., Krasnov G.S., Nakhod K.V., Konash O.S., Vinogradina M.A., Ponomarenko E.A., Druzhilovskiy D.S., Lisitsa A.V. (2020) Processing oxford nanopore long reads using amazon web services. *Biomed. Chem.: Res. Meth.*, **3**(4), e00131.] DOI: 10.18097/bmcrm00131
24. Li H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**(18), 3094-3100. DOI: 10.1093/bioinformatics/bty191.
25. Patro R., Duggal G., Love M.I., Irizarry R.A., Kingsford C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*, **14**(4), 417-419. DOI: 10.1038/nmeth.4197
26. Anzenbacher P., Anzenbacherova E. (2001) Cytochromes P450 and metabolism of xenobiotics. *Cell. Mol. Life Sci.*, **58**(5-6), 737-747. DOI: 10.1007/pl00000897
27. Achour B., Barber J., Rostami-Hodjegan A. (2014) Expression of hepatic drug-metabolizing cytochrome P450 enzymes and their inter-correlations: a meta-analysis. *Drug Metab. Dispos.*, **42**(8), 1349-1356. DOI: 10.1124/dmd.114.058834
28. Petushkova N.A., Kanaeva I.P., Lisitsa A.V., Sheremetyeva G.F., Zgoda V.G., Samenkova N.F., Karuzina I.I., Archakov A.I. (2006) Characterization of human liver cytochromes P450 by combining the biochemical and proteomic approaches. *Toxicol. In Vitro*, **20**(6), 966-974. DOI: 10.1016/j.tiv.2006.01.005
29. Jacob A.G., Smith C.W.J. (2017) Intron retention as a component of regulated gene expression programs. *Hum. Genet.*, **136**(9), 1043-1057. DOI: 10.1007/s00439-017-1791-x

Поступила в редакцию: 20. 02. 2022.
После доработки: 16. 04. 2022.
Принята к печати: 18. 04. 2022.

SPLICE VARIANTS OF mRNA OF CYTOCHROME P450 GENES: ANALYSIS BY THE NANOPORE SEQUENCING METHOD IN HUMAN LIVER TISSUE AND HepG2 CELL LINE

K.A. Deynichenko¹, K.G. Ptitsyn¹, S.P. Radko^{1*}, L.K. Kurbatov¹, I.V. Vakhrushev¹, I.V. Buromski², S.S. Markin¹, A.I. Archakov¹, A.V. Lisitsa¹, E.A. Ponomarenko¹

¹Institute of Biomedical Chemistry,
10 Pogodinskaya str., Moscow 119121, Russia; *e-mail: radkos@yandex.ru
²Pirogov Russian National Research Medical University, Moscow, Russia

The analysis of cytochrome P450 transcripts was carried out by the nanopore sequencing in liver tissue samples of three donors and HepG2 line cells. It has been demonstrated that direct mRNA sequencing with a MinION nanopore sequencer (Oxford Nanopore Technologies) allows one to obtain quantitative profiles for transcripts (and their splice variants) of cytochrome P450 superfamily genes encoding isoforms involved in metabolism of the large (~80%) part of drugs. The splice variant profiles substantially differ for donors. The cytochrome P450 gene expression at the transcript level is significantly weaker in cells of the HepG2 line compared with that in the normal liver tissue. This limits the capability of the direct mRNA nanopore sequencing for studying alternative splicing of cytochrome P450 transcripts in HepG2 cells. Both quantitative and qualitative profiles of the cytochrome P450 gene expression at the transcript level are notably differ in human liver tissue and HepG2 cells.

Key words: splice variants; nanopore sequencing; cytochromes P450; human liver tissue; HepG2 cell line

Funding. The work was performed within the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021-2030) (no. 122030100168-2).

Received 20.02.2022; revised 16.04.2022; accepted 18.04.2022.