

Валідація панелі розширеного генетичного скринінгу CarrierSeq

Виявляючи патогенні варіанти в сотнях генів, розширений скринінг носійства CarrierSeq дозволяє майбутнім батькам оцінити ризик передачі аутосомно-рецесивного або Х-зчепленого захворювання. Відсоток виявлення пар, що перебувають у групі ризику, залежить від кількості аналізованих генів, поширеності відповідних захворювань у певній популяції та аналітичної чутливості методу для виявлення варіантів, що викликають захворювання.

Володимир Рушак,

канд. біол. наук,
науковий директор
генетичної лабораторії
«Ультрагеном»

Наталія Ольхович,

докт. біол. наук,
завідувачка лабораторії
медичної генетики
спеціалізованого
медико-генетичного
центру «Охматдиту»

Вікторія Бадюк,

канд. мед. наук,
доцент, завідувачка
клініко-діагностичної
лабораторії клініки
генетики репродукції
«Вікторія»

У даній статті наведено результати аналітичної валідації панелі CarrierSeq на основі секвенування 420 генів з повним покриттям по кодуючих регіонах. Для цього були використані 30 анонімних зразків ДНК пацієнтів з раніше визначеним носійством аутосомно-рецесивних та Х-зчеплених захворювань або підтвердженим діагнозом спадкового синдрому. Ідентифікація одонуклеотидних варіантів, коротких вставок/делецій та числа копій для всіх генів, які входять до складу панелі CarrierSeq, здійснювалась за допомогою капілярного електрофорезу. Після цього здійснювали аналіз зразків методом секвенування наступного покоління (NGS) та біоінформатичних алгоритмів, включаючи спеціалізовані алгоритми для технічно складних генів та індивідуального аналізу, та проводили валідацію згідно з визначеними критеріями оцінки задовільності результатів.

Валідація панелі CarrierSeq з використанням 30 зразків людської ДНК, показала > 96 % аналітичної чутливості методу та високу конкордантність результатів з такими, що отримані за допомогою капілярного електрофорезу. Продемонстрована здатність методу виявляти варіанти норми та здійснювати аналіз числа копій гену.

Перевірена нами аналітична чутливість панелі CarrierSeq як методу ідентифікації різних типів генетичних варіантів дозволяє рекомендувати її для використання у якості інструменту генетичного скринінгу подружніх пар, що мають репродуктивні плани, особливо для захворювань зі складним молекулярними патогенезом, що значно підвищує ймовірність виявлення родин з групи високого ризику народження дітей з аутосомно-рецесивними та Х-зчепленими синдромами.

Актуальність

Скринінг носійства з метою виявлення спадкових генетичних дефектів був важливим елементом стратегії репродуктивної медицини впродовж останніх 50 років. Дослідження показують, що більшість людей мають принаймні один генетичний варіант, пов'язаний з важким або навіть летальним розладом, який може проявитися в наступних поколіннях (*Chong JX et al., Gao Z et al.*). Однак до недавніх пір ця практика була обмежена певною кількістю моногенних тестів, які переважно проводились для осіб або популяцій груп підвищеного ризику на основі раси, етнічної належності чи походження. Хоча для певних субпопуляцій рекомендується проходити спеціалізований цільовий скринінг носійства, зосереджуючи увагу на захворюваннях із високою поширеністю, зростаюча глобальна мобільність та взаємозв'язок призвели до значного зростання генетичної різноманітності населення в усьому світі.

Як наслідок, генетичні розлади, які раніше були притаманні певним етнічним групам, тепер можуть виникати з більш високою частотою у широких верств населення. Внаслідок цього підходи до генетичної діагностики та роль цього аналізу у репродуктивній медицині швидко змінюються.

Існує понад 1000 патологічних станів, обумовлених рецесивними мутаціями в одному гені, які відрізняються за віком початку та важкістю перебігу (*Bell CJ et al.*). Частота кожного з цих станів серед загальної популяції є дуже невисокою, проте в сукупності ці захворювання є причиною приблизно 20 % дитячої смертності та 10 % госпіталізацій немовлят (*Costa T. et al., Kumar P. et al.*). Скринінг на носійство таких мутацій перед вагітністю або в пренатальний період дає змогу поінформувати пари про ризик народження дитини з серйозною патологією та доступні варіанти репродуктивного планування.

Через зростаючу якість та зменшення вартості геномних технологій, станом на сьогодні можливо проводити загальнонаціональний скринінг носійства одночасно на широкий спектр патологій, так званий розширений скринінг носійства – Expanded carrier screening (ECS). Очікується, що універсальна панель ECS здатна виявити більше пар із ризиком серйозних генетичних захворювань у нащадків, ніж панелі, розроблені на основі етнічної належності, що охоплюють значно меншу кількість генів (*Haque IS et al.*) Американський коледж акушерів-гінекологів (ACMG) нещодавно визнав ECS як прийнятну стратегію для планування вагітності та пренатального скринінгу носійства (*Committee opinion no. 690 summary: carrier screening in the age of genomic medicine*). Нещодавніми дослідженнями було показано, що приблизно 80 % пар, у яких

було виявлено носійство небезпечних мутацій, використовують альтернативні репродуктивні варіанти для зменшення ризиків народження хворої дитини (*Ghioffi CE et al., Archibald AD et al.*).

Обов'язковою умовою для використання ECS є висока достовірність ідентифікації кожної мутації — як для ідентифікації ризику у пар (де обидва партнери є носіями аутосомно-рецесивної патології або жінка-партнер є носієм патології, зчепленої з хромосомою X) так і для мінімізації ризиків у пари, коли лише у одного партнера виявлено носійство патології.

Статус носія для деяких патологій важко ідентифікувати за допомогою стандартних підходів секвенування наступного покоління та біоінформаційного аналізу через особливості послідовності генів захворювання. Тобто ці умови вимагають особливої уваги та спеціального аналізу. Послідовності з тринуклеотидними повторами (наприклад, повтори CGG у гені FMR1 — синдром ламкої X-хромосоми) та високогомологічні гени (наприклад, гени SMN1 та SMN2 — спінальна м'язова атрофія) ускладнюють ідентифікацію варіантів, однак ці складні гени беруть участь у розвитку серйозних захворювань. Наприклад, в США одна з таких патологій, як синдром ламкої X-хромосоми, спінальна м'язова атрофія, вроджена гіперплазія надниркових залоз (CAH), дефіцит 21-гідроксилази та α -таласемія зустрічається в 54 з 100 000 вагітностей (*Beauchamp KA et al.*).

В даній роботі ми охарактеризували ефективність панелі ECS, яка складається з 420 генів, використовуючи метод NGS для ідентифікації однонуклеотидних варіантів (SNV), невеликих вставок та делецій (*indels*), включаючи такі складні гени, як CFTR (трансмембранний регулятор муковісцидозу) та DMD (м'язова дистрофія, тип Дюшенна та Беккера). Дотримуючись існуючих клінічних рекомендацій (*Aziz N et al., Rehm HL et al.*), ми здійснювали аналітичне підтвердження чутливості тесту та його придатність для використання в якості скринінгового методу генетичної діагностики населення.

Матеріали та методи

Дизайн панелі

Панель Ion AmpliSeq CarrierSeq™ складається з 14 044 ампліконів, які покривають 99,94 % кодуючих послідовностей генів, а також фланкуючих регіонів розміром 25 bp, що знаходяться на межі інтрону/екзону всіх 420 генів (повний список генів можна знайти на сайті [thermofisher.com/carrierseq](https://www.thermofisher.com/carrierseq)). Дизайн панелі передбачає високі показники покриття, в результаті чого лише 33 гени мають покриття кодуючих/фланкуючих регіонів менше 100 % та всього 3.

Цільові регіони охоплюють більш ніж 36000 однонуклеотидних замін (SNV) та коротких вставок/делецій, описаних в базі даних ClinVar та ряду непублічних баз генетичних варіантів. Близько 2000 з них (5,5 %) цільових варіантів розташовані в областях на межі кодуючих регіонів, що висуває високі вимоги до якості покриття в областях меж інтрону / екзону.

Крім того, певні міжгенні та інтронні ділянки з відомими некодуючими патологічними варіантами також були включені в склад панелі. Також в склад

панелі входять ряд ідентифікаційних ампліконів, які дозволяють визначити стать та ідентифікувати зразок після проведення аналізу в разі непередбачених обставин.

Хоча аналіз варіативності числа копій гену проводиться для всіх 420 генів, для 33 генів, таких як CFTR (трансмембранний регулятор провідності муковісцидозу) та DMD (М'язова дистрофія Дюшенна) застосовується окремий алгоритм секвенування та аналізу зі збільшеним покриттям для максимізації чутливості щодо складних та небезпечних патологій.

Аналітичні зразки та оцінка результативності

Всього в роботі було використано 30 анонімних зразків ДНК пацієнтів з підтвердженим діагнозом спадкових захворювань, наданими Лабораторією медичної генетики СМГЦ НДСЛ «ОХМАТДИТ» МОЗ України та ТОВ «Клініка генетики репродукції «Вікторія».

Для оцінки задовільності результатів досліджень застосовували бальну шкалу, де:

2 бали – нараховувались при абсолютному збігу результатів досліджень;

1 бал – нараховувався при частковому збігу результатів досліджень;

0 балів – при відсутності збігу результатів досліджень.

Оцінка ступеню точності результатів досліджень являла собою відношення кількості набраних балів до максимально можливої кількості балів, помножене на 100 %.

Підготовка та контроль якості бібліотек

Підготовку бібліотек з отриманих зразків ДНК проводили за допомогою набору реагентів Ion Ampliseq™ CarrierSeq™ ECS Panel (Thermo Fisher Scientific, США). ДНК була фрагментована до ~200 п.н. і трансформована в бібліотеки шляхом зшивки ампліконів зі специфічними адаптерами. Отримані бібліотеки ампліфікували за допомогою кодованих мультиплексних праймерів допомогою ПЛР та. Після декількох етапів очистки бібліотек, включаючи очистку магнітними частками, проводили їх кількісне вимірювання за допомогою Qubit™ Fluorometer.

Подальша підготовка бібліотек здійснювалась з використанням станції автоматичної пробопідготовки Ion Chef (Thermo Fisher Scientific, США) та за допомогою набору реагентів та пластику Ion 530™ Kit-Chef. 100 % отриманих бібліотек відповідали критеріям прийнятності для якості бібліотек. 100 % зразків продемонстрували однорідність > 97 %. Середня глибина прочитання становила 390х.

Секвенування проводили за допомогою Ion 530™ Chip Kit на приладі Ion GeneStudio™ S5. Одночасно аналізувались максимум 4 зразки ДНК на чіпі Ion 530™ (Thermo Fisher Scientific, США, кат № A27764). Аналіз отриманих даних проводився за допомогою Ion Torrent Suite™ Browser версії 5.14 та Ion Reporter™ версії 5.16.0.2. Оцінку варіантів згідно з критеріями патогенності та розрахунок передбачення патогенності виявлених варіантів проводили за допомогою програмного середовища Igenity™ 2.0.7.

Результати

Усі визначені варіанти автоматично класифікувались згідно зі стандартами та рекомендаціями ACMG щодо інтерпретації генетичних варіантів. Для забезпечення точності інтерпретації, усі виявлені патогенні, ймовірно патогенні та варіанти з конфліктуєчими даними переглядались лабораторним персоналом особисто. Після звірки з рядом додаткових публічних баз даних приймалось остаточне рішення щодо включення виявленого варіанту у звіт. Доброякісні варіанти та варіанти з невизначеною значимістю не підлягали перегляду та не включались у звіт.

Отримані результати наведені в *Таблиці 1*. У результаті процедури паралельного дослідження з розділеними пробами анонімних зразків пацієнтів з встановленим діагнозом спадкових захворювань в двох лабораторіях було показано, що точність виявлення мутацій в генах спадкових захворювань людини за допомогою методу секвенування нового покоління та панелі CarrierSeq становить 96,6 %.

Зразок 11 розглядався як підтвердження специфічності методу, оскільки виявлена обома методами мутація є варіантом норми і була правильно інтерпретована як доброякісна. Виявлена в зразках 12А та 23А втрата ділянки гену була ідентифікована CarrierSeq та інтерпретована як втрата копійності гену, що дозволяє говорити про чутливість панелі відносно такого складного для аналізу явища, як варіативність числа копій гену.

У двох випадках, 2Б та 15Б, було ідентифіковано один з двох варіантів, діагностованих за допомогою капілярного електрофорезу. Ряд генів, асоційованих з клінічними станами, мають паралоги або псевдогени, що значно ускладнює їх молекулярний аналіз, зокрема існування множинних поліморфних алелів у гені IDUA вимагає високої аналітичної точності для диференціювання патогенних варіантів від нейтральних поліморфізмів (*Li P et al.*). Ген GBA має розташований поруч псевдоген GBAP1, у якого в певних екзонах спостерігаються послідовності з високою ідентичністю (*Straniero L et al.*). Серед цих генів та їх гомологів часто відбуваються рекомбінації та конверсія генів, що може призвести до зміни кількості копій та хибнопозитивних результатів.



ТАБЛИЦЯ 1

Порівняльні результати діагностики генетичних варіантів методами прямого автоматичного секвенування за Сенгером та методом NGS

Результати дослідження, отримані лабораторією медичної генетики СМГЦ НДСЛ «ОХМАТДИТ» МОЗ України (метод прямого автоматичного секвенування за Сенгером)		Результати дослідження, отримані ТОВ «Ультрагеном» (метод NGS)		Бали
Номер зразка	Виявлені мутації в генах спадкових захворювань	Номер зразка	Виявлені мутації в генах спадкових захворювань	
1	2	3	4	5
1А	IDUA c.1598C>T (p.Pro533Leu) IDUA c.208C>T (p.Gln70Ter)	1Б	IDUA c.1598C>T (p.Pro533Leu); IDUA c.208C>T (p.Gln70Ter)	2
2А	IDUA c.1163C>A (p.Thr388Lys) IDUA c.19_20insC	2Б	IDUA c.1163C>A (p.Thr388Lys);	1
3А	GALT c.855G>T p. Lys285Asn	3Б	GALT c.855G>T (p.Lys285Asn); GALT c.-1039_753del3162 LOSS	2
4А	ARSA c.465+1G>A	4Б	ARSA c.465+1G>A	2
5А	SGSH c.220C>T (p.Arg74Cys); SGSH c.1105G>A (p.Glu369Lys)	5Б	SGSH c.220C>T (p.Arg74Cys); SGSH c.1105G>A (p.Glu369Lys)	2
6А	GLB1 c.397-2A>G	6Б	GLB1 c.397-2A>G	2
7А	GLB1 c.841C>T p.His281Tyr	7Б	GLB1 c.985C>T (p.His329Tyr)	2
8А	GLB1c.699delG (p.Gln234ArgfsTer20); GLB1 c.1175 T>G (p.Leu392Arg)	8Б	GLB1 c.699delG (p.Gln234ArgfsTer20); GLB1 c.1175 T>G (p.Leu392Arg)	2
9А	GLB1 c.841C>T (p.His281Tyr) GLB1 c.438_440delTCT	9Б	GLB1 c.985C>T (p.His329Tyr); GLB1 c.438_440del	2
10А	GLB1 c.841C>T (p.His281Tyr)	10Б	GLB1 c.985C>T (p.His329Tyr)	2
11А	GBA - норма (-/-)	11Б	GBA c.1225-34C>A (Benign)	2
12А	GALT делеція всієї ділянки гену в одному алелі	12Б	GALT g.34646494-34650897 LOSS (втрата копійності гену)	2
13А	GBA c.680A>G (p.Asn227Ser); GBA c.732C>A (p.Tyr244Ter)	13Б	GBA c.680A>G (p.Asn227Ser); GBA c.732C>A (p.Tyr244Ter)	2
14А	GALT c.855G>T (p.Lys285Asn) GALT c.425T>A (p.Met142Lys)	14Б	GALT c.855G>T (p.Lys285Asn) GALT c.425T>A (p.Met142Lys)	2
15А	GBA c.901delC (p.Arg301ValfsTer3) GBA p.R87W	15Б	GBA c.901del (p.Arg301ValfsTer3)	1

Стандарти

Продовження таблиці 1

1	2	3	4	5
16A	IDS c.596_599delAACA	16Б	IDS c.596_599delAACA	2
17A	TPP1 c.622C>(p.Arg208Ter); TPP1 c.622C>T (p.Tyr406Ter)	17Б	TPP1 c.622C>T (p.Arg208Ter); TPP1 c.1218T>A (p.Tyr406Ter)	2
18A	TPP1 c.1677_1678 delTC	18Б	TPP1 g.6614559_6614560AG[1]	2
19A	TPP1 c.622C>T (p.Arg208Ter) TPP1 c.1436_1437delCA (p.Pro479ArgfsTer11);	19Б	TPP1 c.622C>T (p.Arg208Ter) TPP1 c.1436_1437del (p.Pro479ArgfsTer11);	2
Результати дослідження, отримані ТОВ «Клініка генетики репродукції «Вікторія» (метод прямого автоматичного секвенування за Сенгером)		Результати дослідження, отримані ТОВ «Ультрагеном» (метод NGS)		
20A	SGSH c.1105G>A (p.Glu369Lys)	20Б	SGSH c.1105G>A (p.Glu369Lys)	2
21A	SGSH c.220C>T (p.Arg74Cys)	21Б	SGSH c.220C>T (p.Arg74Cys)	2
22A	CFTR c.2374C>T (p.Arg792Ter); ZFYVE26 c.2756-1G>A	22Б	CFTR c.2374C>T (p.Arg792Ter); ZFYVE26 c.2756-1G>A	2
23A	CFTR del2.3(21kb)(c.54-5940_273+10250del); CFTR ΔF508del(-CTT)(c.1521_1523delCTT)	23Б	CFTR g.117504038-117509495 LOSS (втрата копійності гену) CFTR c.1521_1523delCTT	2
24A	CFTR c.3485G>T (p.Arg1162Leu)	24Б	CFTR c.3485G>T (p.Arg1162Leu)	2
25A	LIFR c.3288C>A (p.Asn1096Lys); TTN c.30450C>A (p.Tyr10150Ter)	25Б	LIFR c.3288C>A (p.Asn1096Lys); TTN c.30450C>A (p.Tyr10150Ter)	2
26A	TTN c.30560-1G>C	26Б	TTN c.30560-1G>C	2
27A	LIFR c.3288C>A (p.Asn1096Lys) TTN c.30450C>A (p.Tyr10150Ter); TTN c.30560-1G>C	27Б	LIFR c.3288C>A (p.Asn1096Lys) TTN c.30450C>A (p.Tyr10150Ter); TTN c.30560-1G>C	2
28A	NEB c.16929G>A	28Б	NEB c.16929G>A	2
29A	NEB c.9473C>G	29Б	NEB c.9473C>G	2
30A	CAPN3 c.549delA (p.Thr184Argfs)	30Б	CAPN3 c.549delA (p.Thr184Argfs)	2

- Максимально можлива кількість балів — 60
- Кількість балів за результатами оцінки — 58

- Кількість ідентичних результатів — 96.6 %
- Кількість неспівпадінь — 3.3 %

Висновки

Панель CarrierSeq забезпечує надійну та доступну оцінку ризику для багатьох серйозних рецесивних та Х-зчеплених захворювань одночасно. Отримані дані свідчать на користь можливості використання CarrierSeq як скринінгового методу для діагностики широкого спектра рідкісних рецесивних спадкових захворювань з високою чутливістю та специфічністю.

Варто зазначити, що хоча частота кожного окремого захворювання є низькою, загальна частота 420 включених до складу панелі захворювань перевищує частоту синдрому Дауна (1 на 800 живонароджених), відносно якого здійснюється плановий скринінг.

Включення в протокол алгоритмів аналізу числа копій гену (CNV) стосовно всіх генів панелі значно збільшує шанси ідентифікувати пари, у яких є ризик народження дитини з важким станом. CNV можна ідентифікувати в звичайних клінічних умовах за допомогою загальноприйнятих технологій (наприклад, MLPA), проте тестування MLPA недоступне для одночасного аналізу сотень генів і вимагає додаткових етапів лабораторної обробки, які можуть супроводжуватись помилкою лаборанта. Використовуючи позитивний зразок з ідентифікованою втратою алелі гену, ми продемонстрували можливості високої чутливості NGS для ідентифікації змін копійності гену. Аналіз впливу змін копійності гену на частоту виникнення різних патологічних станів є предметом подальшого вивчення і становить значний науковий інтерес.

Метою даної роботи було визначення можливості виявлення різних типів варіантів (SNV, indels, CNV) та розробити протокол ідентифікації та інтерпретації отриманих результатів. Результати роботи показали, що панель CarrierSeq з високою точністю виявляє генетичні варіанти, ідентифіковані таким еталонним аналітичним методом, як капілярний електрофорез. Як і в інших аналітичних дослідженнях скринінгових тестів на основі NGS, призначених для виявлення нових генетичних варіантів, обмеження нашого аналізу полягає в тому, що аналітичну чутливість неможливо встановити для кожного ідентифікованого варіанту. Одночасному аналізу підлягають біля тисячі мегабаз геномної послідовності, а частота варіацій на багатьох ділянках генів не дозволяє отримати валідаційні зразки для їх включення в аналітичне порівняльне дослідження. Аналітична чутливість, однак, може бути знижена при порівняно великих інсерціях/делеціях, а клінічна чутливість може бути меншою, ніж аналітична чутливість через раніше не описані патогенні варіанти в інтронах або патогенні CNV, які не ідентифікуються під час аналізу.

Враховуючи надзвичайне генетичне розмаїття українців, що склалось історично, широке використання панелі CarrierSeq в якості скринінгового тесту та статистичний аналіз отриманих результатів дозволить отримати цінні дані для ряду патологій відносно частот розповсюдженості їх носійства та захворюваності серед населення України, що в подальшому може складати значний інтерес з точки зору розробки національної стратегії попередження та лікування рідкісних захворювань. ■