

Е.Ш. Кулигина¹, А.П. Соколенко¹, И.В. Бизин¹, А.А. Романько¹, М.О. Анисимова¹,
А.О. Иванцов¹, Е.Н. Имянитов^{1,2}

Поиск новых генетических детерминант наследственного рака молочной железы с помощью полноэкзомного секвенирования: способы оценки патогенности и алгоритмы селекции кандидатных вариантов

¹ФГБУ «НМИЦ онкологии имени Н. Н. Петрова» Министерства здравоохранения Российской Федерации,
²ФГБОУ ВО «СЗГМУ им. И.И. Мечникова» Министерства здравоохранения Российской Федерации,
Санкт-Петербург

Примерно каждая десятая карцинома молочной железы (РМЖ) развивается в результате наследования патогенной мутации в гене опухолевого супрессора. Более половины РМЖ с признаками наследственной этиологии до сих пор не нашли генетического объяснения. Использование высокопроизводительного секвенирования имеет большой потенциал в плане открытия новых предрасполагающих мутаций. Основная проблема полноэкзомных исследований (WES) состоит в дифференциации истинно патогенных каузативных мутаций и нейтральных генных полиморфизмов. Авторами выполнен экзомный анализ 32 образцов ДНК от больных наследственным РМЖ неизвестной генетической этиологии. Разработан оригинальный алгоритм селекции кандидатных мутаций с учётом их частоты у онкологических больных и здоровых лиц, потенциальной патогенности мутации, функций гена, молекул-партнёров. В результате был сформирован список из 245 потенциально патогенных наследственных мутаций. В ходе валидирующего молекулярно-эпидемиологического исследования по схеме «случай-контроль» получены данные, которые свидетельствуют о вероятном вкладе в формирование наследственного риска РМЖ нескольких ранее не изучавшихся в этой связи мутаций. Например, мутация *BRCA1* p.Gln356Arg в гомозиготном состоянии встречалась у 0.9% случайных больных РМЖ и была ассоциирована с 7-кратным увеличением риска РМЖ. Это подтверждает существование предрасполагающих к РМЖ генов, действующих по рецессивному механизму.

Ключевые слова: рак молочной железы, наследственный

Введение

Примерно каждый десятый случай рака молочной железы (РМЖ) связан с наследованием патогенной мутации в гене опухолевого супрес-

сора [16, 32, 40]. На долю известных предрасполагающих («каузативных») мутаций в классических и «новых» генах наследственного рака (*BRCA1*, *BRCA2*, *CHEK2*, *NBN*, *PALB2*, *ATM*, *BLM*, *PTEN*, *BRIPI*, *TP53*, *NF1*, *STK11*, *RAD51D*, *CDH1* etc) приходится менее половины таких случаев РМЖ [7; 45; 10; 28]. Еще 8-12% случаев наследственного РМЖ можно объяснить с позиций полигенной модели, в рамках которой за причину развития рака принимают сочетанное носительство нескольких неблагоприятных средне- и низко-пенетрантных аллелей [11, 13, 30]. Так или иначе, значительная доля РМЖ с признаками наследственной этиологии до сих пор не имеет генетического объяснения.

Использование высокопроизводительного секвенирования для анализа геномов пациентов с наследственными раковыми синдромами обещало дать скорое разъяснение многим случаям РМЖ с неизвестной наследственной этиологией и пополнить список генов «наследственного рака» новыми кандидатами [29]. Однако, как и в случае других онкопатологий со сложным мультифакторным патогенезом, несмотря на всеобщие усилия, в этом направлении достигнуты весьма скромные успехи. В качестве новых генетических детерминант РМЖ были предложены редкие мутации в генах репарации *XRCC1* [15, 36], *FANCM* [12, 19], *RECQL* [9], *MCPH1* [25], однако ни одна из них не смогла сравниться по клинической значимости с «каноническими» мутациями в генах *BRCA1/2*.

Возможно, относительно низкая информативность экзомного анализа связана с гетерогенностью исследуемого заболевания и существованием множества разнообразных фенокопий. С другой стороны, каузативные мутации могли быть упущены из-за неоптимального алгоритма селекции кандидатных вариантов.

В данной работе с помощью глубокого секвенирования исследовали полную последовательность 32 экзомов *BRCA*-негативных пациенток с признаками наследственного РМЖ. Был разработан оптимальный для данных условий алгоритм

селекции потенциально патогенных вариантов, сформирован список кандидатных мутаций, выполнены валидирующие эксперименты.

Материалы и методы

Пациентки

Были исследованы 32 образца ДНК, полученные из периферической крови от больных раком молочной железы с ярко выраженными клиническими признаками наследственного заболевания. Эти пациентки характеризовались необычно ранним возрастом возникновения заболевания (<45 лет), билатеральностью поражения молочных желёз и/или семейным анамнезом РМЖ, что дало основания предполагать у них наличие высокопенетрантного генетического фактора риска РМЖ. Отбор образцов на полноэкзомное секвенирование подразумевал также отсутствие характерных для популяции Северо-запада России рекуррентных мутаций в известных генах наследственного РМЖ — *BRCA1/2*, *CHEK2* и *NBS1*.

Полноэкзомное секвенирование и биоинформатический процессинг данных

Выполнено полноэкзомное секвенирование (WES) на платформе Illumina NextSeq. Для приготовления библиотек использовали набор для экзомного обогащения Nextera Rapid Capture Exome (Illumina, USA). Для обработки «сырых» данных применяли программный комплекс на основе Illumina Best Practice. Полученные с секвенатора файлы объединялись прикладной утилитой *zcat* и подавались на вход утилиты BWA. В качестве референсного генома была

выбрана версия hg19. После выравнивания прочтений на референс Sequence Alignment/Map (SAM) формат конвертировали в бинарный сжатый формат BAM программой Samtools; Дальнейшая обработка до этапа получения файлов формата Variant Call Format (VCF) осуществлялась при помощи команд программного комплекса Genome Analysis Toolkit (GATK) версии 3.6. Для улучшения работы алгоритма коррекции данные были локально выровнены вокруг часто встречающихся инсерций и делеций, полученных из базы данных The 1000 Genomes Project. Для определения наследственных полиморфизмов и их аннотации использовался инструмент HaplotypeCaller и snpEff версии 4.3. Схема эксперимента представлена на рис. 1.

Молекулярно-эпидемиологическая валидация кандидатных вариантов

Частоты мутантных аллелей были исследованы с помощью ПЦР в группах пациенток с признаками наследственного рака (n=640) и в группе онкологически здоровых доноров пожилого и среднего возраста (45-89 лет) (n=805). Наиболее перспективные варианты были направлены на расширенное исследование по схеме «случай — контроль» с привлечением 1200 случайных «последовательных» больных РМЖ и 750 здоровых доноров. Для детекции гетерозиготных мутаций применяли анализ кривых плавления высокой точности (HRM) с последующим секвенированием по Сэнгеру аномальных фрагментов. Для анализа гомозиготных мутаций использовали аллель-специфическую ПЦР (AS-PCR). Величина риска, ассоциированного с аллелем, оценивалась с помощью показателя OR (odds ratio), который представляет собой отношение шансов заболеть среди носителей патогенного генотипа и носителей «дикого типа» [35]. Схема молекулярно-эпидемиологического исследования представлена на рис. 2.

Результаты и обсуждение

Проблема отбора кандидатных мутаций

В ходе полного секвенирования кодирующей части генома (whole exome sequencing, WES) каждой пациентки с РМЖ обнаруживается приблизительно 11000-13000 несинонимичных альтернативных вариантов разной степени патогенности. Список генных вариантов, обнаруженных нами с помощью WES у 32 больных РМЖ «высокого риска», насчитывал более 52000 позиций, так как большинство полиморфных вариантов повторяются в популяции. Представляется крайне затруднительным идентифицировать единичные причинно-значимые для каждой пациентки события внутри такого огромного массива информации. Таким образом, основной научной проблемой подобных исследований является не столько сложность лабораторных экспериментов (приготовление библиотек, секвенирование и т.д.), сколько оптимизация биоинформатического процессинга и разработка алгоритма фильтрации данных. Суть работы в данном случае состоит в том, чтобы отделить истинно патогенные мутации, явившиеся генетической причиной наследственной патологии, от нейтральных по отношению к данному признаку полиморфных вариантов. Чтобы сформулировать оптимальный

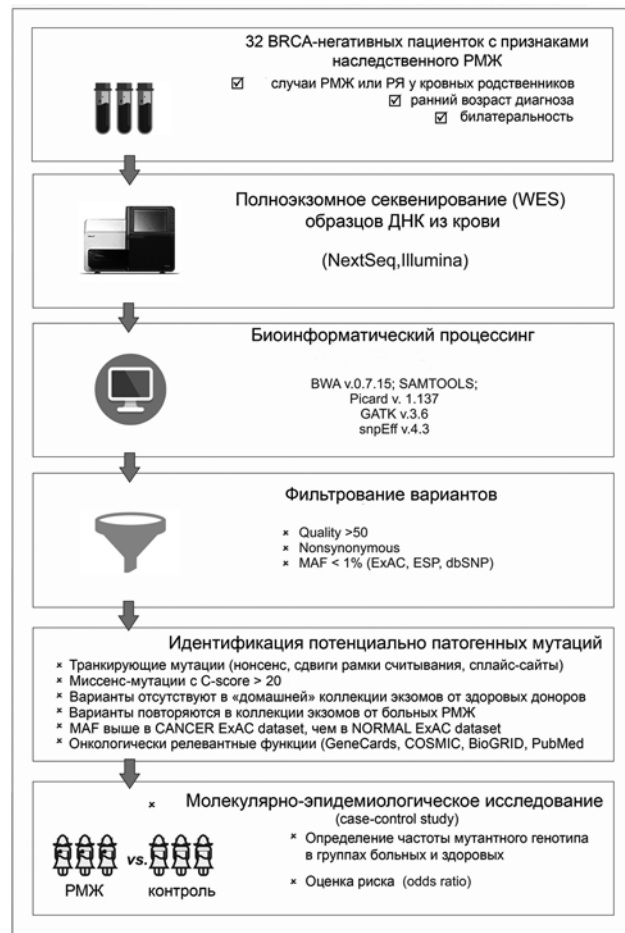


Рис. 1. Схема поиска новых генов наследственного РМЖ с помощью полноэкзомного секвенирования

Таблица 1. Частоты некоторых «канонических» патогенных наследственных мутаций РМЖ и показатели ассоциированного с ними риска

Ген, Мутация	ID, dbSNP	MAF, ExAC, %	MAF, ExAC, Non_Cancer, %	MAF, ExAC, Cancer, %	OR, 95% CI*
BRCA1 5382 insC	rs397507247	0,02-0.24	0,01	0,07	8.2 (4.49-14.86) (не Ашкенази)
CHEK2 1100delC	rs555607708	0,18–1.00	0,18	0,17	2.6 (2.00-3.27)
CHEK2 ivs2+1G>A	rs121908698	0,00-0.01	0,01	0,05	4.0 (0.5-30.8)
PALB2 c.509_510delGA	rs515726124	0,00-0.01	0,00	0,03	15.4 (0.90-263.55)
BLM p.Gln548*	rs200389141	0,00-0.02	0,01	0,06	2.5 (1.29-4.78)

* данные из работ Noskowicz et al., 2014, Sokolenko et al., 2012; Kuligina et al., 2011, Buslov et al., 2005; Satagopan et al., 2001

Таблица 2. Результаты молекулярно-эпидемиологического тестирования кандидатных вариантов

Тип мутации	C-score	Ген	Мутация	Генотип: гетеро-/гомозигота	Метод анализа	РМЖ «высокого риска» (wt/wt - mut/wt - mut/mut)	«Пожилой» контроль (wt/wt - mut/wt - mut/mut)	ВЫВОД
миссенс	26,5	ING1	p.Pro319Leu	гетеро	HRM	N=588 (586-2-0)	N=450 (450-0-0)	рекуррентная
нонсенс	40	EMSY	p.Gly934*	гетеро	HRM	N=562 (562-0-0)	N=441 (441-0-0)	приватная
миссенс	26	HELLS	p.Arg53Cys	гетеро	HRM	N = 517 (515-1-0)	N = 338 (338-0-0)	рекуррентная
миссенс	25,2	TP53INP1	p.Glu27Asp	гетеро	HRM	N=550 (498-1-0)	N=556 (556-0-0)	рекуррентная
нонсенс	39	SFXN2	p.Gln267*	гетеро	HRM	N=285 (284-1-0)	N=278 (277-1-0)	полиморфизм
нонсенс	32	PZP	p.Arg680*	гетеро	HRM	N= 1260 (1253-7-0)#	N=1100 (1100-0-0)	рекуррентная
миссенс	34	MLH3	p.Cys1393Phe	гетеро	HRM	N=180 (180-0-0)	N=180 (180-0-0)	приватная
миссенс	34	MMS19	p.Arg1012Cys	гетеро	HRM	N= 583 (583-0-0)	N= 441 (441-0-0)	приватная
миссенс	26,6	TP63	p.Ala139Val	гетеро	HRM	N=410 (410-0-0)	N=442 (442-0-0)	приватная
миссенс	35	WT1	p.Arg370His	гетеро	HRM	N=586 (586 -0-0)	N= 361 (361-0-0)	приватная
миссенс	27,5	WRNIP1	p.Gly466Trp	гетеро	HRM	N= 637 (637-0-0)	N=494 (494-0-0)	приватная
миссенс	18,29	BRCA1	p.Gln356Arg	гомо	AS-PCR	N=812 (710-95-7)#	N=775 (675-99-1)	рекуррентная
нонсенс	42	ATRIP	p.Arg760*	гетеро	HRM	N=562 (562-0-0)	N=441 (441-0-0)	приватная
миссенс	22	EXO1	p.Gly759Glu	гомо	AS-PCR	N=128 (128-0-0)	N= 282 (282-0-0)	приватная

Примечание: # - частота мутации в группе пациентов существенно выше, чем в контроле (χ^2 , $p = 0.07$)

для наших условий алгоритм отбора кандидатных мутаций, мы проанализировали подходы, которые применяются другими научными группами для решения этой задачи.

Стандартные подходы к классификации генных полиморфизмов

Общепринятым «золотым стандартом» для оценки патогенности вариантов в генах, ассоциированных с моногенными наследственными заболеваниями, являются рекомендации ACMG-AMP (American College of Medical Genetics and Genomics/Association for Molecular Pathology) [3, 23, 39]. В данный протокол включены 28 критериев, которые адресованы разным аспектам проявления патогенности мутации: популяционные данные о частоте аллеля, эпидемиологические исследования (case-control study), функциональные эксперименты, результаты *in silico* предикции, анализ сегрегации в семьях и др. Каждый критерий имеет свой код и вводится с

определенным весом, соответствующим степени убедительности данного признака. В результате суммарной оценки всех параметров, вариант переводится в одну из категорий патогенности: однозначно патогенный (pathogenic, P), вероятно патогенный (likely pathogenic, LP), вариант с неясным функциональным значением (variant of uncertain significance, VUS), возможно нейтральный (likely benign, LB), нейтральный (benign, B). Важно, что, несмотря на то, что эта процедура весьма стандартизована, в отдельных случаях допускаются произвольные модификации «веса» критерия со стороны куратора, и конечное решение, таким образом, во многом зависит от личного мнения эксперта.

Наш опыт применения данного «калькулятора патогенности» показал, что только очень незначительное количество обнаруженных вариантов попадает в категорию «патогенный» и «вероятно патогенный». Из 15437 вариантов с популяционной частотой <1% только 64 были отфильтрованы как патогенные, 2529 оказались



Рис. 2. Эпидемиологическая верификация кандидатных мутаций по схеме «случай-контроль»

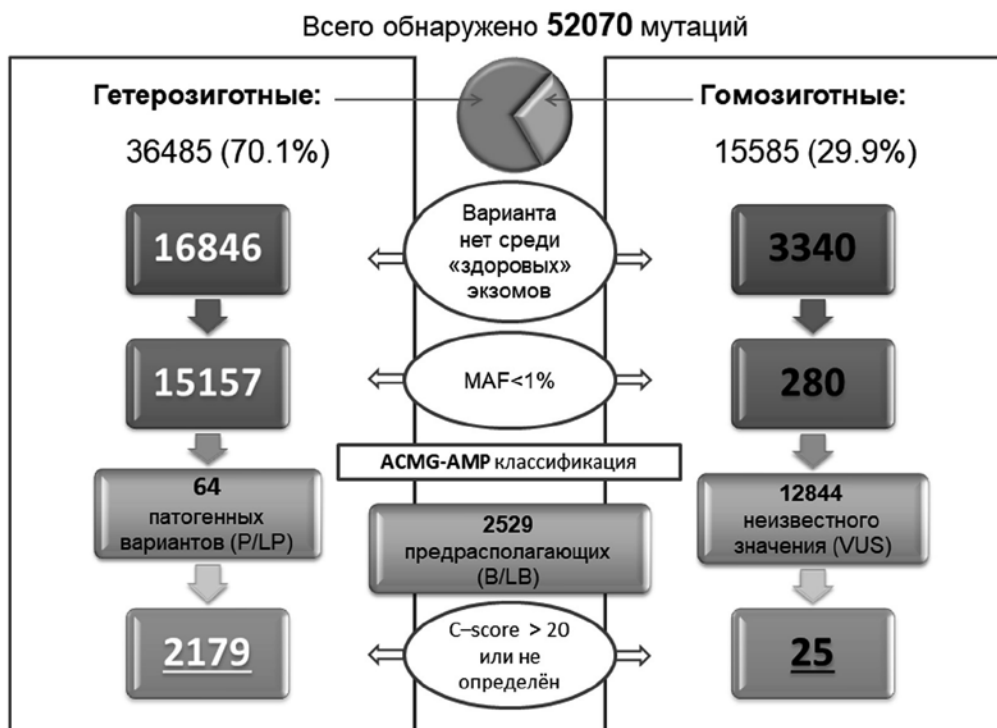


Рис. 3. Статистика фильтрации альтернативных вариантов, обнаруженных в результате экзомного анализа 32 пациенток с наследственным РМЖ

слабо предрасполагающими, а большая часть (12844) так и осталась вариантами с неясным значением, VUS (рис. 3). По-видимому, низкая эффективность применения стандартного протокола ACMG-AMP в наших условиях связана с отсутствием информации по ряду важных

параметров: куратор обычно не располагает данными ни о функциональных исследованиях мутантного белка, ни о ко-сегрегации аллельного варианта и какого-либо фенотипического признака в родословной. Фактически, алгоритм селекции в нашем случае сводится к комбина-

ции нескольких фильтров (MAF, *in silico* патогенность, встречаемость в «домашних» коллекциях экзомов от пациентов и здоровых доноров) и анализу баз данных, посвященных функциям продукта (GeneCard, BioGrid, UniProt), соматическим раковым мутациям (COSMIC, My Cancer Genome, cBioPortal), наследственным мутациям (HGMD, LOVD, ClinVar), поиску релевантных публикаций в базе PubMed.

Частота альтернативного аллеля в популяции

Первым и наиболее объективным показателем патогенности мутации является популяционная частота вариантного аллеля (minor allele frequency, MAF): чем она ниже, тем больше вероятность того, что мутация приводит к серьезным последствиям. В таблице 1 приведены результаты нашего анализа частот наиболее известных патогенных мутаций в генах наследственного рака, которые встречаются в российской популяции, а также показатели ассоциированного с ними риска OR (odds ratio) [4, 6, 21, 34, 46; ExAC Browser, <http://exac.broadinstitute.org/>]. Как следует из этих наблюдений, частота патогенной мутации в популяции обычно не превышает 0.1%, за исключением делеции *CHEK2 1100delC*, которая представляет собой довольно частое событие в некоторых популяциях (MAF ~1%). Повидимому, этот частотный лимит (MAF <1%) и следует выбрать в качестве порогового значения при селекции потенциально патогенных кандидатных мутаций (рис. 3).

В качестве источника информации о частотах полиморфных аллелей принято пользоваться одной из трех гигантских общедоступных баз данных: 1000 Genomes Project Consortium [1]; Exome Sequencing Project (ESP) [49] и ExAC Browser [18]. Последняя является, пожалуй, наиболее значимой: она агрегирует данные полноэкзомного секвенирования 60706 индивидуумов, сгруппированных по ряду атрибутов, в том числе по этнической и континентальной принадлежности, и, что особенно важно, по онкологическому анамнезу. Таким образом, представляется возможность сравнить частоту интересующего варианта в двух подмножествах базы данных: dataset CANCER (n=7601) vs. dataset NORMAL (n=53105). Если обнаруживается существенное увеличение встречаемости какого-либо варианта среди больных раком по сравнению со здоровой популяцией, то это может косвенно указывать на его предрасполагающую онкогенную роль и повышать приоритет мутации в списке кандидатов на валидирующие исследования. Существенно, что каждый альтернативный вариант, представленный в этой базе, сопровождается ме-

таданнами — ссылками на базы аллельных полиморфизмов dbSNP и патогенных клинически-важных мутаций UCSC и ClinVar.

In silico предикторы патогенности мутаций

Мутации принято подразделять на транкрирующие варианты (преждевременные стоп-кодоны, сдвиги рамки считывания и нарушения сайтов сплайсинга) и варианты с неясным значением (variants with uncertain significance, VUS), куда входят, главным образом, аминокислотные замены (миссенс-мутации), а также делеции без сдвига рамки считывания, структурные вариации, утраты стоп-кодонов и др. Транкрирующие мутации, бесспорно, приводят к нарушению функций соответствующего белкового продукта, и весь вопрос состоит в том, действительно ли этот белок может играть существенную роль в проонкогенных процессах. Вторая группа вариантов имеет неочевидные последствия и требует специального анализа.

Существующие инструменты, разработанные для предсказания патогенности аллельных вариантов, могут базироваться на анализе отдельных предиктивных параметров и их комбинаций (SIFT, PolyPhen, Panther, MutationTaster) или использовать нейросетевой интегральный подход и машинное обучение (CADD, FATHMM). Например, SIFT [<http://sift.jcvi.org/>], в котором локализована мутация [33] PolyPhen-2 [<http://genetics.bwh.harvard.edu/pph2/>], объединяет 8 различных параметров, в том числе, влияние замены на вторичную структуру белка, аминокислотный контекст и консервативность фрагмента [2]. MutationTaster [<http://www.mutationtaster.org/>] апробирован в плане классификации вариантов с неясным значением [42]. Неутешительна специфичность и чувствительность детекции активирующих патогенных мутаций — так, например, знаменитая онкогенная мутация *BRAF V600E* по заключению различных аналитических алгоритмов является малофункциональной [23]. Перечисленные инструменты имеют весьма умеренную информативность в отношении сплайсинговых мутаций. Для такого рода анализа рекомендуется использовать специальные «сплайсинговые» предикторы Human Splicing Finder [8]; GeneSplicer [37].

К сожалению, мутации в некодирующих областях (интронах, межгенных спейсерах, регуляторных элементах) также находятся за пределами компетенции перечисленных стандартных *in silico* предикторов [27]. Между тем, отсутствие заметного прогресса в поисках новых генетических детерминант рака в рамках традиционного анализа экзомов заставляет исследователей обратить большее внимание на некодирующие, в

первую очередь, регуляторные области генома. Для решения этой задачи в последнее время стали появляться актуальные биоинформатические инструменты. Например, недавно обнаруженный продукт CScare [41] с 70%-ой аккуратностью распознает функционально значимые некодирующие мутации. FATHMM-indel специально разработан для анализа малых инсерций/делеций за пределами экзона [14]; OncodriveFML был создан для селекции драйверных соматических мутаций в опухолях, он хорошо анализирует промоторы и сайты связывания регуляторных РНК [31]. Учитывая несовершенство каждого отдельно взятого предиктора, рекомендуется одновременно привлекать к анализу несколько разных инструментов.

В данном исследовании мы пользовались одним из наиболее популярных на сегодняшний день биоинформатических инструментов оценки потенциальной патогенности — Combined Annotation Dependent Depletion (CADD), интегральным методом определения последствий мутаций [20]. База данных CADD включает более 8.6 миллиардов записей, которые содержат все возможные варианты единичных замен нуклеотидов в человеческом геноме. Для каждого варианта при помощи нейросетевого подхода на основе результатов распространенных программ-предикторов патогенности рассчитана интегральная оценка (C-score) их влияния на синтезируемый белок. Фильтруя варианты по пороговому значению C-score, можно выделить для дальнейшего рассмотрения заранее заданную долю наиболее потенциально опасных вариантов. Данный выбор является вполне удовлетворительным, особенно, если учесть, что CADD — предиктор лучше «разбирается» в аминокислотных заменах, чем в транкирующих мутациях. Независимое апробирование данного предиктора на тестовых коллекциях генных вариантов продемонстрировало специфичность выявления патогенных мутаций 79%, а чувствительность — 74% [26].

В нашей работе в дополнение к двум описанным фильтрам (частотному, MAF <1%) и «патогенному» (C-score > 20), мы также использовали третий — отсутствие данного мутантного варианта в «домашней» коллекции экзотов от онкологически-здоровых индивидуумов (n=17). Результаты фильтрации представлены на рис. 3. Все варианты, удовлетворившие перечисленным условиям, были вручную проверены с помощью секвенирования и процессинга данных. Взвешенное решение о включении варианта в список кандидатов на дальнейшую валидацию принимал куратор после изучения подробной информации о данном гене, паттерне его экспрессии, соматических мутациях в опухолях и функциях

соответствующего белка с помощью ресурсов GeneCard (<http://www.genecards.org>); COSMIC (<http://cancer.sanger.ac.uk/cosmic>), Pubmed (<http://www.ncbi.nlm.nih.gov/pubmed>), BioGrid (<http://thebiogrid.org/>).

Способы валидации каузативной роли кандидатной мутации

Наиболее убедительным признаком патогенности служит выявление мутации у больных членов семьи при её отсутствии у здоровых родственников (сегрегация). Однако в распоряжении специалистов редко имеются многочисленные родословные, в которых можно проследить вертикальную передачу мутации и фенотипа. К тому же, рак молочной железы — заболевание частое и гетерогенное по патогенезу — что означает, что в пределах семьи могут комбинироваться как наследственные, так и спорадические случаи, что затрудняет анализ сегрегации.

Для нашей работы наиболее пригодным способом валидации оказалось молекулярно-эпидемиологическое тестирование кандидатных мутаций. Оно базируется на том положении, что истинно предрасполагающие к РМЖ варианты должны достоверно чаще встречаться в группе онкологических пациенток по сравнению со здоровыми женщинами, причем это преобладание будет наиболее выражено в «генетически-обогащенной» группе больных (семейные случаи, ранние и билатеральные случаи) [21, 22, 38, 47]. Молекулярно-эпидемиологическое исследование по схеме «случай-контроль» требует генотипирования больших когорт пациентов и здоровых контролей (рис. 2).

Результаты подобного молекулярно-генетического тестирования можно условно подразделить на три варианта: 1) кандидатная мутация не обнаруживается ни в одной из когорт; 2) мутация встречается только в группе больных или её частота в группе больных значительно выше, чем в группе здоровых контролей; 3) мутация встречается в обеих выборках, причем её частота в группе больных и здоровых не различается или выше среди здоровых «контролей». В первом случае вариант относили к редким «приватным» мутациям, во втором — к «потенциально предрасполагающим рекуррентным», в третьем — к «нейтральным полиморфизмам». Результаты молекулярно-эпидемиологической валидации наиболее интересных вариантов представлены в табл. 2. Нами было обнаружено пять потенциально предрасполагающих рекуррентных мутаций (PZP p.Arg680*, BRCA1 p.Gln356Arg, ING1 p.Pro319Leu, HELLS p.Arg53Cys, TP53INP1 p.Glu27Asp) и восемь уникальных «приватных» мутаций (EMSY p.Gly934*, MLH3 p.Cys1393Phe,

MMS19 p.Arg1012Cys, TP63 p.Ala139Val, WT1 p.Arg370His, WRNIP1 p.Gly466Trp, ATRIP p.Arg760, EXO1 p.Gly759Glu.*

Ген *PZP* (pregnancy-zone protein) до сих пор не был описан как имеющий отношение к каким-либо онкологическими заболеваниями. Он кодирует плазменный белок-ингибитор протеиназ и предположительно участвует в процессах клеточной адгезии. Его концентрация в плазме крови резко возрастает во время беременности, а также у некоторых пациентов с раком яичников [48]. В одной из работ была зафиксирована его подавленная экспрессия в клеточных линиях рака молочной железы [24]. Несмотря на неоднозначную функцию данного гена, для обнаруженной нами нонсенс-мутации *PZP p.Arg680** были получены весьма убедительные данные эпидемиологической верификации. Среди 1260 случаев РМЖ было обнаружено 7 носителей (0.56%), тогда как среди 1100 здоровых женщин не найдено ни одного. Показатель риска OR составил 13.2 (95 % CI: 0.75 — 230.85), *p*-value = 0.078. Среди семи носительниц мутации, пять женщин имели признаки наследственной передачи РМЖ: трое были весьма молоды на момент постановки диагноза (младше 45 лет), одна больная имела билатеральную форму рака и в одном случае наблюдался РМЖ у матери.

Миссенс-мутация *BRCA1 p.Gln356Arg* (rs1799950; MAF = 4,21%; C-score = 18.3) представляла для нас особенный интерес, поскольку полногеномный анализ выявил её в гомозиготном состоянии. Эта мутация уже привлекала к себе внимание [5,17]. В результате выполненных исследований типа «случай-контроль» были получены стандартные показатели риска: $OR_{\text{per allele}}$ (в пересчёте на аллель) и OR для носителей мутации; они оказались весьма небольшими, в диапазоне от 0.94 до 1.7. Мы предлагаем рассматривать эту мутацию в рамках рецессивной модели наследования и, следовательно, считать патогенным только гомозиготный генотип. Нами было обнаружено 7 (0.9%) гомозигот в группе РМЖ с высоким риском (n=812), 6 (0.8%) гомозигот в группе случайных РМЖ (n=720), в то время как среди здоровых женщин (n=775) этот генотип встретился всего один раз (0.1%). Показатель риска $OR_{\text{per allele}}$, как и описано ранее, был незначителен — 1.03 (95 % CI: 0.79 — 1.38; *p* = 0.878); и в то же время, для гомозигот OR он выглядел весьма впечатляюще — 6.7 (95 % CI: 0.83 — 54.83; *p* = 0.075). Примечательно, что в когорте больных было зафиксировано достоверное нарушение равновесия Харди-Вайнберга за счет выраженного избытка гомозигот (*p* = 0.02). Ни одна из носительниц гомозиготного генотипа не имела случаев заболевания РМЖ у матери, у одной пациентки РМЖ был диагностирован у

сестры. Эти обстоятельства вполне согласуются с представлением о механизме трансмиссии рецессивного заболевания. Таким образом, были впервые получены данные, свидетельствующие о существовании предрасполагающих к РМЖ генов, действующих по рецессивному механизму. Интересно, что такой мутацией в данном случае оказалась мутация самого знаменитого «ракового» гена — *BRCA1*, который до сих пор являлся общепризнанным примером доминантного наследования.

Кроме исследованных выше генетических вариантов, к числу потенциально предрасполагающих были отнесены весьма привлекательные редкие мутации в генах опухолевых супрессоров *ING1* (inhibitor of growth) и *TP53INP1* (tumor protein p53 inducible nuclear protein 1) и в гене фермента хеликазы *HELLS*. Во всех этих случаях были найдены дополнительные носительницы среди больных РМЖ с высоким риском и не обнаружено ни одного здорового носителя.

Достоверное преобладание мутантного генотипа в группе пациенток по сравнению со здоровым контролем служит косвенным свидетельством причастности соответствующей мутации к формированию повышенного онкологического риска. Однако окончательным доказательством функциональной значимости мутации могут считаться только результаты структурного и функционального анализа мутантного белка *in vitro* и различных тестов *in vivo*.

Таким образом, в представленном исследовании оригинальный алгоритм отбора мутаций наследственной предрасположенности к РМЖ был применен к анализу данных экзомного секвенирования 32 пациенток с признаками наследственной патологии. В результате валидирующего молекулярно-эпидемиологического исследования получены данные, которые свидетельствуют о вероятном вкладе в формирование наследственного риска РМЖ нескольких ранее не изучавшихся в этой связи мутаций — *PZP p.Arg680**, *BRCA1 p.Gln356Arg*, *ING1 p.Pro319Leu*, *HELLS p.Arg53Cys*, *TP53INP1 p.Glu27Asp*.

Работа выполнена при поддержке гранта РНФ № 16-45-02011

ЛИТЕРАТУРА

1. 1000 Genomes Project Consortium, Auton A., Brooks L.D. et al. A global reference for human genetic variation // *Nature*. — 2015. — Vol. 526. — P. 68-74.
2. Adzhubei I.A., Schmidt S., Peshkin L. et al. A method and server for predicting damaging missense mutations // *Nat. Methods*. — 2010. — Vol. 7. — P. 248-249.
3. Amendola L.M., Jarvik G.P., Leo M.C. et al. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory

- Research Consortium // *Am J. Hum Genet.* — 2016. — Vol. 98. — P. 1067-1076.
4. Bogdanova N., Enssen-Dubrowskaja N., Feshchenko S. et al. Association of two mutations in the CHEK2 gene with breast cancer // *Int. J. Cancer.* — 2005. — Vol. 116. — P. 263-266.
 5. Baynes C., Healey C.S., Pooley K.A. et al. Common variants in the ATM, BRCA1, BRCA2, CHEK2 and TP53 cancer susceptibility genes are unlikely to increase breast cancer risk // *Breast Cancer Res.* — 2007. — Vol. 9. — P. R27.
 6. Buslov K.G., Iyevleva A.G., Chekmariova E.V. et al. NBS1 657del5 mutation may contribute only to a limited fraction of breast cancer cases in Russia // *Int. J. Cancer.* — 2005. — Vol. 114. — P. 585-589.
 7. Couch F.J., Shimelis H., Hu C. et al. Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer // *JAMA Oncol.* — 2017. — Vol. 3. — P. 1190-1196.
 8. Desmet F.O., Hamroun D., Lalande M. et al. Human Splicing Finder: an online bioinformatics tool to predict splicing signals // *Nucleic. Acids Res.* — 2009. — Vol. 37. — P. e67.
 9. Cybulski C., Carrot-Zhang J., Kluźniak W. et al. Germline RECQL mutations are associated with breast cancer susceptibility // *Nat. Genet.* — 2015. — Vol. 47. — P. 643-646.
 10. Easton D.F., Pharoah P.D., Antoniou A.C. et al. Gene-panel sequencing and the prediction of breast-cancer risk // *N. Engl. J. Med.* — 2015. — Vol. 372. — P. 2243-2257.
 11. Gorodnova T.V., Kuligina E.Sh., Yanus G.A. et al. Distribution of FGFR2, TNRC9, MAP3K1, LSP1, and 8q24 alleles in genetically enriched breast cancer patients versus elderly tumor-free women // *Cancer Genet. Cytogenet.* — 2010. — Vol. 199. — P. 69-72.
 12. Gracia-Aznarez F.J., Fernandez V., Pita G. et al. Whole exome sequencing suggests much of non-BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles // *PLoS One.* — 2013. — Vol. 8. — P. e55681.
 13. Fachal L., Dunning A.M. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer // *Curr. Opin. Genet. Dev.* — 2015. — Vol. 30. — P. 32-41.
 14. Ferlaino M., Rogers M.F., Shihab H.A. et al. An integrative approach to predicting the functional effects of small indels in non-coding regions of the human genome // *BMC Bioinformatics.* — 2017. — Vol. 18. — P. 442.
 15. Hilbers F.S., Wijnen J.T., Hoogerbrugge N. et al. Rare variants in XRCC2 as breast cancer susceptibility alleles // *J. Med. Genet.* — 2012. — Vol. 49. — P. 618-620.
 16. Hoang LN, Gilks BC. Hereditary Breast and Ovarian Cancer Syndrome: Moving Beyond BRCA1 and BRCA2 // *Adv. Anat. Pathol.* — 2017. — Sep 13.
 17. Johnson N., Fletcher O., Palles C. et al. Counting potentially functional variants in BRCA1, BRCA2 and ATM predicts breast cancer susceptibility // *Hum Mol. Genet.* — 2007. — Vol. 16. — P. 1051-1057.
 18. Karczewski KJ, Weisburd B, Thomas B, et al. The ExAC browser: displaying reference data information from over 60 000 exomes // *Nucleic. Acids Res.* — 2017. — Vol. 45. — P. D840-D845.
 19. Kiiski J.I., Tervasmäki A., Pelttari L.M. et al. FANCM mutation c.5791C>T is a risk factor for triple-negative breast cancer in the Finnish population // *Breast Cancer Res Treat.* — 2017. — Vol. 166. — P. 217-226.
 20. Kircher M., Witten D.M., Jain P. et al. A general framework for estimating the relative pathogenicity of human genetic variants // *Nat. Genet.* — 2014. — Vol. 46. — P. 310-315.
 21. Kuligina E., Reiner A., Imyanitov E.N., Begg C. Evaluating cancer epidemiologic risk factors using multiple primary malignancies // *Epidemiology.* — 2010. — Vol. 21. — P. 366-372.
 22. Kuligina E.Sh., Sokolenko A.P., Mitiushkina N.V. et al. Value of bilateral breast cancer for identification of rare recessive at-risk alleles: evidence for the role of homozygous GEN1 c.2515_2519delAAGTT mutation // *Fam Cancer.* — 2013. — Vol. 12. — P. 129-132.
 23. Li M.M., Datto M., Duncavage E.J. et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists // *J. Mol. Diagn.* — 2017. — Vol. 19. — P. 4-23.
 24. Liang X., Huuskonen J., Hajivandi M. et al. Identification and quantification of proteins differentially secreted by a pair of normal and malignant breast-cancer cell lines // *Proteomics.* — 2009. — Vol. 9. — P. 182-193.
 25. Mantere T., Winqvist R., Kauppila S., Grip M. et al. Targeted Next-Generation Sequencing Identifies a Recurrent Mutation in MCPH1 Associating with Hereditary Breast Cancer Susceptibility // *PLoS Genet.* — 2016. — Vol. 12. — P. e1005816.
 26. Masica D.L., Karchin R. Towards Increasing the Clinical Relevance of In Silico Methods to Predict Pathogenic Missense Variants // *PLoS Comput Biol.* — 2016. — Vol. 12. — P. e1004725.
 27. Mather C.A., Mooney S.D., Salipante S.J. et al. CADD score has limited clinical validity for the identification of pathogenic variants in noncoding regions in a hereditary cancer panel // *Genet. Med.* — 2016. — Vol. 18. — P. 1269-1275.
 28. Mavaddat N., Antoniou A.C., Easton D.F., Garcia-Closas M. Genetic susceptibility to breast cancer // *Mol. Oncol.* — 2010. — Vol. 4. — P. 174-191.
 29. Melloni GEM, Mazzarella L., Bernard L. et al. A knowledge-based framework for the discovery of cancer-predisposing variants using large-scale sequencing breast cancer data // *Breast Cancer Res.* — 2017. — Vol. 19. — P. 63.
 30. Michailidou K., Hall P., Gonzalez-Neira A. et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk // *Nat. Genet.* — 2013. — Vol. 45. — P. 353-361.
 31. Mularoni L., Sabarinathan R., Deu-Pons J. et al. OncoPrint: a general framework to identify coding and non-coding regions with cancer driver mutations // *Genome Biol.* — 2016. — Vol. 17. — P. 128.
 32. NCCN Guidelines genetic/familial high-risk assessment: breast and ovarian (2016) V.2.
 33. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function // *Nucleic. Acids Res.* — 2003. — Vol. 31. — P. 3812-3814.
 34. Noskowitz M., Bogdanova N., Bermisheva M. et al. Prevalence of PALB2 mutation c.509_510delGA in unselected breast cancer patients from Central and Eastern Europe // *Fam Cancer.* — 2014. — Vol. 13. — P. 137-142.
 35. Pagano M., Gauvreau K. (2000) Principles of biostatistics. 2nd ed. Belmont, CA: Brooks/Cole.

36. Pelttari L.M., Kiiski J.I., Ranta S. et al. RAD51, XRCC3, and XRCC2 mutation screening in Finnish breast cancer families // Springerplus. — 2015. — Vol. 4. — P. 92
37. Perteau M., Lin X., Salzberg S.L. GeneSplicer: a new computational method for splice site prediction // Nucleic. Acids Res. — 2001. — Vol. 29. — P. 1185e1190.
38. Rahman N., Seal S., Thompson D. et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene // Nat. Genet. — 2007. — Vol. 39. — P. 165-167.
39. Richards S., Aziz N., Bale S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology // Genet. Med. — 2015. — Vol. 17. — P. 405-424.
40. Ripperger T., Gadzicki D., Meindl A., Schlegelberger B. Breast cancer susceptibility: current knowledge and implications for genetic counseling // Eur. J. Hum Genet. — 2009. — Vol. 17. — P. 722-731.
41. Rogers M.F., Shihab H.A., Gaunt T.R., Campbell C. CS-cape: a tool for predicting oncogenic single-point mutations in the cancer genome // Sci. Rep. — 2017. — Vol. 7. — P. 11597.
42. Sadowski C.E., Kohlstedt D., Meisel C. et al. BRCA1/2 missense mutations and the value of in-silico analyses // Eur. J. Med. Genet — 2017. — Vol. 60. — P. 572-577.
43. Satagopan J.M., Offit K., Foulkes W. et al. The lifetime risks of breast cancer in Ashkenazi Jewish carriers of BRCA1 and BRCA2 mutations // Cancer Epidemiol. Biomarkers Prev. — 2001. — Vol. 10. — P. 467-473.
44. Schwarz J.M., Rödelsperger C., Schuelke M., Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations // Nat. Methods. — 2010. — Vol. 7. — P. 575-576.
45. Slavin T.P., Maxwell K.N., Lilyquist J. et al. The contribution of pathogenic variants in breast cancer susceptibility genes to familial breast cancer risk // NPJ Breast Cancer. — 2017. — Vol. 3. — P. 22.
46. Sokolenko A.P., Iyevleva A.G., Preobrazhenskaya E.V. et al. High prevalence and breast cancer predisposing role of the BLM c.1642 C>T (Q548X) mutation in Russia // Int. J. Cancer. — 2012. — Vol. 130. — P. 2867-2873.
47. Sokolenko A.P., Mitiushkina N.V., Buslov K.G. et al. High frequency of BRCA1 5382insC mutation in Russian breast cancer patients // Eur. J. Cancer. — 2006. — Vol. 42. — P. 1380-1384.
48. Teng H., Zhang W.Y., Zhu F.Q. A study on the serum pregnancy zone protein levels in pregnant women and patients with gynecological tumors // Chin Med J. (Engl). — 1994. — Vol. 107. — P. 910-914.
49. Tennessen J.A., Bigham A.W., O'Connor T.D. et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes // Science. — 2012. — Vol. 337. — P. 64-69.

*E.S. Kuligina¹, A.P. Sokolenko¹, I.V. Bizin¹,
A.A. Romanko¹, M.O. Anisimova¹, A.O. Ivantsov¹,
E.N. Imyanitov^{1,2}*

Searching for missing heritability for breast cancer by whole exome sequencing: evaluation of variant pathogenicity and selection of candidate mutations

¹N.N. Petrov National Medical Research Center of Oncology
²I.I. Mechnikov North-West State Medical University
St. Petersburg

Approximately 10% of breast cancer (BC) incidence is attributed to inherited germ-line mutations in tumor suppressor genes, and about a half of all hereditary BC cases remain unexplained by known genetic determinants. Whole exome sequencing (WES) offers great promise for the identification of novel hereditary cancer genes, however the interpretation of disease-causing significance for newly identified variants presents a challenge. Here we present the results of WES analysis of 32 hereditary BC patients with unknown causative mutations. In total, 52070 germ-line genetic variants were identified. We designed a special selection algorithm, which uses a sequence of filtering steps for highlighting probably pathogenic mutations. The pipeline is capable to compute minor allele frequencies (MAF) based on public databases; to estimate the potential pathogenicity of mutation according to CADD-database; to summarize the data regarding the gene functions, its pattern of expression and protein interactions networks. As a result we assembled the list of 245 potentially deleterious germ-line mutations probably associated with increased risk of BC. Some of these variants were validated in a case-control study. For example, homozygous missense substitution *BRCA1 p.Gln356Arg* was detected in 0.9% of consecutive BC cases; it was associated with 7-fold increased BC risk (95 % CI: 0.83 — 54.83). This finding supports the existence of recessive inheritance of breast cancer predisposition.

Key words: whole exome sequencing, candidate mutations, breast cancer, *BRCA1 p.Gln356Arg* mutation

Поступила в редакцию 04.12.2018 г.