

РОССИЙСКАЯ АКАДЕМИЯ НАУК
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ ХИМИЧЕСКОЙ БИОЛОГИИ И ФУНДАМЕНТАЛЬНОЙ
МЕДИЦИНЫ

На правах рукописи

Кечин Андрей Андреевич

**Разработка и применение метода определения
мутаций в генах *BRCA1* и *BRCA2* у больных раком
молочной железы и раком яичников**

03.01.03 – молекулярная биология

Диссертация на соискание ученой степени
кандидата биологических наук

Научный руководитель:
к.б.н. Филипенко М.Л.

Новосибирск – 2017

Оглавление

Оглавление	2
Список сокращений.....	4
Введение	6
Глава 1. Обзор литературы.....	11
Рак молочной железы и рак яичников	11
Рак молочной железы	11
Рак яичников	12
История открытия генов <i>BRCA1</i> и <i>BRCA2</i>	13
Строение и функции белков, кодируемых генами <i>BRCA1/2</i>	15
Ген <i>BRCA1</i>	15
Ген <i>BRCA2</i>	17
Методы выявления мутаций в генах <i>BRCA1/2</i>	18
Методы приготовления библиотек для таргетного секвенирования генов <i>BRCA1/2</i>	19
Технологии секвенирования NGS	22
Методы анализа получаемых после секвенирования прочтений.....	24
Статистика мутаций в генах <i>BRCA1/2</i>	27
Пенетрантность мутаций в генах <i>BRCA1/2</i>	28
Соматические мутации	30
«hotspot»-мутации в генах <i>BRCA1/2</i>	32
Мутации в генах <i>BRCA1/2</i> в выборках пациентов из России.....	32
Заключение.....	34
Глава 2. Материалы и методы.....	35
Материалы.....	35
Пациенты.....	35
Реактивы.....	36
Растворы и буферы	36
Готовые коммерческие наборы.....	37
Ферменты	37
Олигонуклеотиды	37
Методы.....	37
Выделение ДНК.....	37
Определение мутаций с повышенной частотой	38
Приготовление библиотеки	39
Электрофорез ДНК в ПААГ	48
Переосаждение ДНК изопропанолом.....	48
Секвенирование библиотеки	48
Анализ данных.....	48
Статистическая обработка результатов	52
Секвенирование экзонов генов <i>BRCA1/2</i> по Сэнгеру	53
Глава 3. Результаты и их обсуждение	54
Методы приготовления библиотеки.....	54
Метод приготовления библиотеки для образцов ДНК, выделенных не из гистологических блоков.....	54
Мультиплексный метод приготовления библиотеки.....	55
Анализ данных.....	59
Оптимизация алгоритма и подбор параметров.....	59
Создание единого пакета анализа NGS-данных экзонов генов <i>BRCA1/2</i>	69
Исследование CNV и потери гетерозиготности в генах <i>BRCA1/2</i>	79
Потеря гетерозиготности и мутации, ассоциированные с развитием РМЖ/РЯ.....	81

Механизм возникновения регионов с ЛОН: CNV или геновая конверсия?.....	84
Выводы	89
Список литературы.....	91
Приложение 1.....	107
Приложение 2.....	110

Список сокращений

а.о. – аминокислотных остатков

гДНК – геномная ДНК

ВП – верно положительные

ГР – гомологичная рекомбинация

ДНК – дезоксирибонуклеиновая кислота

е.а. – единицы активности

ИХБФМ СО РАН – Институт химической биологии и фундаментальной медицины
Сибирского отделения Российской академии наук

КРРМЖ – кластерные районы РМЖ

КРРЯ – кластерные районы РЯ

ЛО – ложно отрицательные

ЛП – ложно положительные

ЛПАА – линейный полиакриламид

НГСК – негомологичное соединение концов

п.о. – пар оснований

ПАА – полиакриламид

ПААГ – полиакриламидный гель

ПГБ – парафинированный гистологический блок

ПО – программное обеспечение

ПСА – персульфат аммония

ПЦР – полимеразная цепная реакция

РМЖ – рак молочной железы

РФ – Российская Федерация

РЯ – рак яичников

СПЛ – секвенирование с помощью лигирования

СПС – секвенирование с помощью синтеза

ТЕМЭД – тетраметилэтилендиамин

Трис – трисгидроксиметиламинометан

ФГБУ – федеральное государственное бюджетное учреждение

ЭДТА – этилендиаминтетрауксусная кислота

ВАМ – карта выравнивания последовательностей в бинарном виде (binary alignment map)

ВІС – база данных мутаций в генах *BRCA1/2* (breast cancer information core)

BWA – программа для картирования прочтений (burrows-wheeler alignment)

CNV – вариация числа копий (copy number variation)

dNTP – дезоксинуклеозидтрифосфат (deoxy nucleoside triphosphate)

GATK – сокращенное название программы GenomeAnalysisToolkit

LOH – потеря гетерозиготности (loss of heterozygosity)

MLPA – мультиплексная амплификация лигированных зондов (multiplex ligation-dependent amplification)

MPS – массовое параллельное секвенирование (massive parallel sequencing)

NES – последовательность экспорта из ядра (nuclear export sequence)

NGS – секвенирование нового поколения (next-generation sequencing)

NLS – последовательность ядерной локализации (nuclear localization sequence)

PARP – поли(АДФ-рибоза)-полимераза (poly (ADP-ribose) polymerase)

SAM – карта выравнивания последовательностей (sequence alignment map)

SDS – додецилсульфат натрия (sodium dodecyl sulfate)

SNV – однонуклеотидная вариация (single nucleotide variation)

Введение

Ежегодно в России регистрируется около 66 тысяч новых случаев рака молочной железы (РМЖ) и около 14 тысяч новых случаев рака яичников (РЯ) (по данным на 2015 год) [1]. Одни из основных факторов, предрасполагающих к развитию данных заболеваний – это мутации в генах *BRCA1* и *BRCA2*, повышающие риск возникновения опухоли до 80 лет до 72% для РМЖ и до 44% для РЯ [2]. Определение таких мутаций имеет значение не только для выявления пациентов группы риска, но и для назначения пациентам с состоявшимся раком таргетной терапии препаратами-ингибиторами системы репарации (олапариб, рупапариб) или препаратами, эффективность применения которых зависит от активности системы репарации клетки (например, цисплатин). Золотым стандартом тестирования мутаций в *BRCA1* и *BRCA2* является секвенирование по Сэнгеру. Однако, большой размер генов (размеры кодирующей области - 5592 п.о. и 10257 п.о., соответственно) и небольшое число мутаций с повышенной частотой в популяции («hotspot»-мутации) делает эту процедуру слишком дорогой.

В настоящий момент большое распространение получила технология массового параллельного секвенирования (MPS), или секвенирования нового поколения (NGS). Благодаря ей стало возможным быстро и с меньшими финансовыми затратами одновременно анализировать множество участков генома и даже целые геномы. Одним из применений данной технологии является определение последовательностей выбранных фрагментов ДНК, в том числе и экзонов генов *BRCA1/2* – так называемое, таргетное секвенирование. При этом процедура исследования включает в себя три наиболее важных этапа: приготовление библиотеки, секвенирование и анализ полученных прочтений.

В литературе уже описаны как методы приготовления библиотек для таргетного NGS, так и методы обработки получаемых прочтений. Большинство описанных методов предполагают использование коммерческих наборов, позволяющих с высокой чувствительностью и специфичностью выявлять мутации в кодирующих последовательностях *BRCA1/2*, однако в то же время повышающие себестоимость такого исследования для пациента в десятки раз. Кроме того, на момент начала работы над диссертацией не было опубликовано работ, в которых использовалась платформа MiSeq (Illumina), что связано с тем, что компания не имела клинической аккредитации. Поэтому большинство исследователей концентрировали свое внимание на технологиях секвенирования компаний Roche и ThermoFisher Scientific – 454 и Ion Torrent. Однако использование этих технологий для анализа генов *BRCA1/2* до сих пор остается под вопросом, поскольку большинство мутаций в этих генах находятся в гомополимерных

районах, в которых повышена частота ошибок секвенирования для технологий 454 и Ion Torrent. Поэтому разработка собственного метода приготовления библиотеки для платформы MiSeq (Illumina) является актуальной задачей.

Завершающим этапом исследования генов *BRCA1/2* с помощью технологий NGS является обработка и анализ полученных после секвенирования прочтений. При этом могут быть использованы несколько разных подходов. Первым подходом является последовательный запуск всех программ через командную строку с выбором всех необходимых параметров их работы. Минус такого подхода – трудность его адаптации в других лабораториях. Вторым подходом является использование таких систем составления автоматических протоколов, как Galaxy (<https://galaxyproject.org>) и GeneXplain (<http://genexplain.com>), позволяющие с помощью компьютерной мыши легко составлять протоколы из тех программ, которые уже установлены на сервере. Однако при необходимости установки новых инструментов или модификации старых, возникает потребность в создании скриптов, интерфейсов и прочих настройках. Кроме того, подобные системы требуют постоянный доступ в интернет. Третьим подходом является использование готовых протоколов, в которых уже готовые программы объединены в единую оболочку с помощью BASH, Python, Perl и других скриптов. Такие протоколы легко могут быть модифицированы и не требуют доступа в интернет. В то же время, все параметры в них должны быть заранее подобраны и адаптированы к конкретной задаче на репрезентативной выборке результатов секвенирования, чтобы конечному пользователю не требовалось проводить подбор их значений параметров. На сегодняшний день доступны только коммерческие автоматические протоколы обработки данных. Поэтому актуальным является разработка свободно доступного протокола обработки NGS данных генов *BRCA1/2* и его проверка на представительной выборке образцов. Помимо объединения готовых программ в единый протокол, существует и необходимость разработки программы для удаления последовательностей праймеров из полученных прочтений. Это обусловлено тем, что созданные ранее программы адаптированы только для удаления последовательностей адаптеров, поскольку таргетное секвенирование, основанное на амплификации (amplicon-based targeted NGS) получило широкое распространение только в последнее время. Кроме того, нет универсальных программ для выявления CNV по данным NGS генов *BRCA1/2* на множестве пациентов без применения референсных образцов, при котором выявление инсерций и делеций осуществляется после нормализации. В то же время, по некоторым данным, крупные перестройки могут составлять от 2 до 8% наследуемых мутаций в генах *BRCA1/2* [3, 4]. Кроме того, неисследованной остаётся распространённость крупных перестроек в генах *BRCA1/2* в

опухолевой ткани.

Другая слабо исследованная область – это распространение мутаций в генах *BRCA1/2* среди больных РМЖ и РЯ жителей России. Несмотря на то, что с помощью технологии NGS обнаружено большое число вариаций в генах *BRCA1/2* (более 12000 в базе данных ClinVar), российские популяции не были охарактеризованы по представленности мутаций во всей кодирующей последовательности. Были проведены работы по определению частот мутаций, имеющих более высокую частоту, чем другие (5382insC, 185delAG, 4154delA, C61G, 3819delGTAAA, 3875delGTCT, 2080delA, 6174delA) [5], однако это не позволяет сделать вывод о распространенности мутаций в генах *BRCA1/2* среди больных РМЖ или РЯ ввиду неисследованности остальных районов данных генов. Кроме того, неясным остается природа так называемой потери гетерозиготности (LOH) в генах *BRCA1/2*: происходит ли это из-за CNV или же в результате геномной конверсии [6]. Связано это с проблемами исследования CNV в парафинированных гистологических блоках (ПГБ).

Цель и задачи исследования. Целью настоящей работы являлась разработка и применение метода выявления герминальных и соматических мутаций в генах *BRCA1* и *BRCA2* у пациенток больных раком молочной железы и раком яичников, проживающих в Российской Федерации. Для достижения поставленной цели выполнялись следующие задачи:

1. Разработка метода приготовления библиотеки кодирующих последовательностей генов *BRCA1/2*, основанный на амплификации выбранных фрагментов и предназначенный для последующего секвенирования на MiSeq Illumina.
2. Разработка программ для автоматической обработки получаемых после секвенирования прочтений, включая программу для удаления последовательностей праймеров из прочтений, а также выявления CNV в генах *BRCA1/2*.
3. Выявление герминальных и соматических вариаций (SNV, короткие инсерции и делеции, CNV) в генах *BRCA1/2* в выборке больных раком молочной железы и раком яичников, проживающих на территории РФ, с использованием разработанных методов и программ.
4. Сравнение выявляемых герминальных и соматических вариаций (SNV, короткие инсерции и делеции, CNV), а также оценка представленности феномена потери гетерозиготности в генах *BRCA1/2* среди исследуемых пациентов в ДНК, выделенной из крови и из парафинированных гистологических блоков.

Научная новизна и практическая ценность работы. В рамках работы был разработан первый в РФ полный отечественный метод приготовления библиотеки для

секвенирования генов *BRCA1/2*. Также разработаны новые программы, позволяющие проводить обработку и анализ прочтений, получаемых при секвенировании генов *BRCA1/2* с помощью технологии NGS. В том числе, разработана новая программа, позволяющая выявлять CNV в образцах ДНК, выделенных как из крови, так и из ткани опухоли в гистологических блоках, заключенных в парафин. Впервые проведено сравнение CNV, выявляемых в ДНК, выделенной из крови и из гистологических блоков. Результаты работы имеют высокую практическую ценность для генетики, молекулярной биологии и медицины. Применение разработанных методов позволяет снизить стоимость генетического исследования пациентов больных РМЖ и/или РЯ как для выявления пациентов с высоким риском развития опухоли, так и для выбора тактики лечения больных. Разработанный протокол эффективен для анализа мутаций *BRCA1/2* в ДНК, выделенных из ткани опухоли в гистологических блоках, заключенных в парафин, что увеличивает процент выявляемых мутаций и расширяет группу пациенток, для которых может быть эффективна таргетная терапия.

Апробация работы и публикации. По материалам работы опубликованы 3 статьи, получен 1 патент, результаты работы были представлены на трех конференциях:

- Ермоленко Н. А., Боярских У. А., Кечин А. А., Лазарев Л. Ф., Петрова В. Д., Мазитова А. М., Кушлинский Н. Е., Филипенко М. Л. Опыт клинического использования платформы MiSeq Illumina для диагностики мутаций *BRCA1* и *BRCA2* // Технологии живых систем. – 2015. – Т. 12. – №1. – С. 11–23.
- Ermolenko N. A., Boyarskikh U. A., Kechin A. A., Mazitova A. M., Khrapov E. A., Petrova V. D., Lazarev A. F., Kushlinskii N. E., Filipenko M. L. Massive parallel sequencing for diagnostic genetic testing of BRCA genes - a single center experience // Asian Pac. J. Cancer Prev. – 2015. – V. 16. – No. 17. – С. 7935–7941.
- Kechin A., Boyarskikh U., Kel A., Filipenko M. cutPrimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing // J. Comput. Biol. – 2017. – V. 24. – №11. – С. 1138–1143.
- Кечин А. А., Боярских У. А., Ермоленко Н. А., Храпов Е. А., Филипенко М. Л. Способ определения нуклеотидных последовательностей экзонов генов *BRCA1* и *BRCA2* // Патент РФ № 2612894 от 13.03.2017 г., заявка № 2015153232 от 11.12.2015 г.
- Кечин А. А., Боярских У. А., Филипенко М. Л. Разработка комплекса программ для исследования кодирующих последовательностей генов и их участков, ассоциированных с заболеваниями. Международный биотехнологический конгресс (Москва, 20–22 февраля 2017 г.). – С. 374–375.

- Кечин А. А., Боярских У. А., Ермоленко Н. А., Храпов Е. А., Тюляндина А. С., Лазарева Д. Г., Филипенко М. Л. Разработка метода выявления мутаций в генах *BRCA1* и *BRCA2* с помощью секвенирования следующего поколения (NGS) // IX Всероссийская научно-практическая конференция с международным участием «Молекулярная диагностика 2017» (Москва, 18–20 апреля 2017 г.). – Т. 2. – С. 290–291.
- Кечин А. А., Боярских У. А., Ермоленко Н. А., Храпов Е. А., Тюляндина А. С., Лазарева Д. Г., Филипенко М. Л. Разработка метода выявления клинически значимых структурных вариантов в генах *BRCA1* и *BRCA2* с использованием массового параллельного секвенирования в образцах ДНК, выделенной из крови и гистологических блоков // Вторая международная научно-практическая конференция «NGS в медицинской генетике» (Суздаль, 26–28 апреля 2017 г.). – С. 32.

Вклад автора. Основные результаты получены автором самостоятельно. Процедуры по разработке и оптимизации метода приготовления библиотек были выполнены совместно с к.б.н. Боярских У.А., Ермоленко Н.А. Образцы ДНК из гистологических блоков были выделены совместно с Троменшлегер И.Н. Образцы ДНК из крови были выделены Задорожным А.В. Разработка программ, анализ данных NGS и статистический анализ полученных данных сделаны лично автором.

Объем и структура диссертации. Диссертация состоит из введения, обзора литературы, материалов и методов исследования, результатов собственных исследований и их обсуждения, выводов и списка цитируемой литературы. Текст изложен на 118 страницах, иллюстрирован 23 рисунками, включает 11 таблиц, список литературы содержит 199 библиографических источников.

Глава 1. Обзор литературы

Рак молочной железы и рак яичников

Рак молочной железы – наиболее распространенный тип рака среди женщин. Ежегодно в России им заболевают около 66 тысяч женщин (по данным на 2015 год) и более 20 тысяч ежегодно от него умирают [1]. В то же время рак яичников считается самым смертоносным видом опухоли для женщин: ежегодно от него умирают более 7 тысяч человек, а заболевают – 14 тысяч [1]. Далее будут рассмотрены основные характеристики РМЖ и РЯ и пути их лечения, а затем история выявления ассоциации между РМЖ, РЯ и нарушениями в генах *BRCA1/2*.

Рак молочной железы

РМЖ представляет собой злокачественную опухоль железистой ткани молочной железы. Основными причинами развития заболевания считается нарушение генетического и гормонального контроля над ростом клеток в молочной железе [7], что ведет к их активной пролиферации и разрастанию ткани, в том числе в протоки и прилегающие ткани. Выявление и определение типов таких нарушений является одним из подходов для классификации РМЖ, что необходимо для выбора тактики лечения больных.

Согласно рекомендациям Российского общества клинической онкологии (<http://www.rosoncweb.ru/>) первый метод, чаще всего применяемый при выявлении опухолей – хирургическое вмешательство. Исключения составляют случаи, когда такое лечение невозможно. Следующим этапом, в случае недостаточности хирургического вмешательства или прогрессирования опухоли, является использование адъювантной химиотерапии (таксаны и антрациклины, блокирующие пролиферацию клеток). При этом последовательность типов химиопрепаратов четко прописана и регламентирована (рисунок 1 *Рисунок 1*).

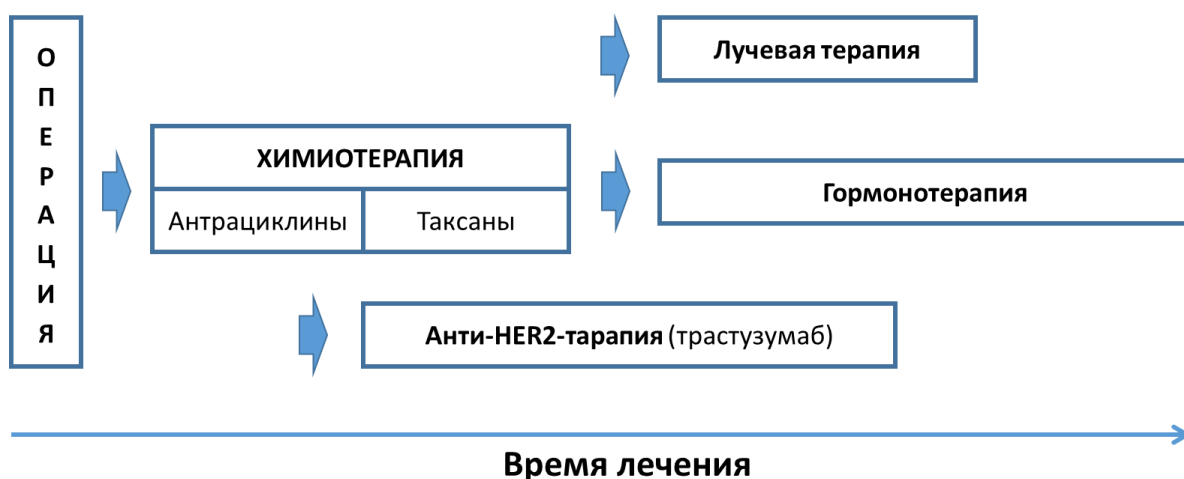


Рисунок 1. Рекомендуемая последовательность применения видов терапии при раке молочной железы. Пересечение прямоугольников по оси абсцисс означает совместное использование типов лечения. Антрациклины и таксаны применяются последовательно. Адаптировано из [8].

Кроме того, регламентирована тактика лечения при наличии или отсутствии на клетках опухоли рецепторов HER2, рецепторов к эстрогену и прогестерону, а также наследуемых мутаций в генах *BRCA1/2*. При наличии последних пациентам рекомендуется назначать препараты на основе платины (карбоплатин, цисплатин) [8], поскольку такие препараты вносят одно- и двунитевые разрывы в ДНК клеток, а мутации в этих генах снижают уровень работы системы репарации клетки. Одновременно с этим, в скором времени ожидается внедрение в практику препаратов-ингибиторов PARP (поли(АФД-рибоза)-полимеразы, poly (ADP-ribose)-polymerase) (например, олапариба) для лечения РМЖ. Для олапариба сейчас заканчиваются клинические испытания [9], после чего станет актуальным выявление у больных РМЖ не только герминальных, но и соматических мутаций, что значительно увеличит нагрузку и повысит требования для медико-генетических лабораторий.

Рак яичников

РЯ представляет собой гетерогенное злокачественное новообразование внутри или на поверхности яичника(-ов). Основными причинами высокой смертности больных этим видом рака являются возможность выявления заболевания в основном на поздних стадиях, а также высокая миграционная активность малигнизированных клеток. Большинство случаев (до 90 %) относятся к эпителиальному типу [10], который может иметь различное гистологическое строение [11]:

- Серозная карцинома (подразделяется на карциномы с низкой и высокой

злокачественностью) – до 71 % [10];

- Эндометриоидная карцинома – 10 % [10];
- Муцинозная карцинома – 3 % [10];
- Светлоклеточная карцинома – 5 % [10];
- Злокачественная опухоль Бреннера;
- Серозно-муцинозная карцинома;
- Недифференцированная карцинома;
- Смешанная эпителиальная карцинома.

На последние четыре типа приходится оставшиеся 11 % всех случаев. От гистологического типа зависит набор дополнительных обследований, которые необходимо провести пациенту. Для всех больных с серозной и эндометриоидной карциномой рекомендуется проводить генетическое консультирование и определение мутаций в генах *BRCA1/2* [11].

Лечение РЯ так же, как и РМЖ, рекомендуется начинать с полной или оптимальной циторедуктивной операции, то есть операции, нацеленной на удаление опухолевых масс. После этого, в зависимости от типа опухоли и стадии опухолевого процесса, пациенту может быть назначена химиотерапия такими препаратами, как паклитаксел, доцетаксел (относятся к группе таксанов), карбоплатин, цисплатин (относятся к группе препаратов на основе платины), циклофосфан [11]. В случае рецидива (более 70% всех случаев [12]), чувствительности опухоли к препаратам платины и наличия мутаций в генах *BRCA1/2*, больным рекомендуют назначать PARP-ингибитор олапариб [11].

Помимо олапариба, на сегодняшний день показали высокую эффективность против клеток рака яичников и другие препараты-ингибиторы PARP, проходящие (нирапариб, велипариб, талазопариб [13]) или уже прошедшие клинические испытания и получившие одобрение FDA (например, препарат рукапариб [14]). Высокая эффективность этих препаратов, а также в скором времени возможная их доступность (из-за большого их разнообразия на рынке и, как следствие, снижения стоимости) дадут возможность большему числу пациентов получать такую терапию. Данное обстоятельство увеличит необходимость в исследованиях генов *BRCA1/2*.

Таким образом, выявление мутаций в генах *BRCA1/2* у больных РМЖ и РЯ с использованием эффективных и максимально недорогих методов, является актуальной задачей.

История открытия генов *BRCA1* и *BRCA2*

Сегодня большинству исследователей и врачей, связанных с работой в области РМЖ

и РЯ, хорошо известно об ассоциации РМЖ, РЯ и генов *BRCA1/2*. Однако данная связь долгое время оставалась не выявленной ввиду относительно низкой пенетрантности мутаций (по сравнению со многими другими заболеваниями, не включая опухоли), а также сложной структуры генов. И лишь с появлением методов молекулярной биологии, позволяющих с помощью маркерных последовательностей картировать ассоциированные участки, эти гены были найдены.

Впервые распространенность РМЖ среди членов одной семьи была замечена еще в 1948 году на данных о 459 семьях [15]. Тогда было замечено, что пациенты, у которых в семье были больные РМЖ умирали чаще от РМЖ, чем пациенты из общей популяции. В то же время, исследователями не было найдено ассоциации между семейной формой и более ранним развитием заболевания, как это известно сейчас. Скорее всего, это было связано с невозможностью авторами понять, когда заболевание возникло у исследуемых пациентов. После этого исследования еще долгое время не было новых прорывных данных о наследуемости РМЖ или РЯ. Однако был сделан прорыв в области биотехнологий, когда в 1983 году путем исследования ассоциации полиморфных локусов впервые был выявлен район, ассоциированный с болезнью Хантингтона [16]. И в 1990 году ассоциация РМЖ с районом 17й хромосомы 17q21 была показана путем типирования 183 полиморфных локусов у 329 пациентов из 23 семей со 146 случаями РМЖ (в среднем, по 6 случаев на семью) [17]. Приблизительно в то же время исследователи стали связывать между собой РМЖ и РЯ как заболевания, которые ассоциированы с одним и тем же предрасполагающим генетическим фактором [18]. В 1993 году профессор Даг Истон, изучив 214 семей, у которых РМЖ был ассоциирован с районом 17q21, сузил искомый район 17й хромосомы до 8,3 сМ для мужчин и 18,0 сМ для женщин (разница для мужчин и женщин связана с разницей в частоте рекомбинаций гомо- и гетерогаметного полов) [19]. Кроме того, в этой же работе было высказано предположение о существовании и других *BRCA*-генов. В мае 1994 под руководством профессора Брюса Пондера район поиска был уменьшен до менее, чем 1 сМ [20]. А в августе того же года этой же командой была опубликована карта района 17й хромосомы, на которой были размечены 20 генов-кандидатов для *BRCA1* [21]. И, наконец, в октябре 1994 года первыми ген *BRCA1* идентифицировали исследователи под руководством Марка Скольника [22] – будущего основателя скандально известной компании Myriad Genetics, которая получила патент на последовательность генов *BRCA1* и *BRCA2*. Район, содержащий второй ген ассоциации с РМЖ (*BRCA2*), был найден в сентябре 1994 года [23], однако на точную идентификацию этого гена ушло еще больше года, и в декабре 1995 в журнале *Science* вышла статья, описывающая первые 6 наследственных мутаций в гене *BRCA2*, которые были

ассоциированы с семейным РМЖ [24]. Таким образом, были выявлены гены, мутации в которых имеют одну из самых высоких пенетрантностей среди всех опухолевых синдромов [25] и ведут к так называемому синдрому наследственного рака молочной железы и яичников [26].

С 1995 года по 2017 год в базе данных ClinVar стало уже описано более 12000 вариаций в генах *BRCA1* и *BRCA2*, из которых около 5000 считаются мутациями, повышающими риск развития РМЖ и/или РЯ. При этом они часто имеют различное влияние на строение и функции кодируемых белков, что связано с их сложной структурой и большим количеством выполняемых функций.

Строение и функции белков, кодируемых генами *BRCA1/2*

Ген *BRCA1*

Ген *BRCA1* (синонимы: *BRCA1*, *BRCC1*, *BROVCA1*, *FANCS*, *IRIS*, *PNCA4*, *PPP1R53*, *PSCP*, *RNF53*) располагается на длинном плече 17й хромосомы в позиции 17q21, имея координаты 41196312 – 41277500 пар оснований (п.о.) (сборка GRCh37). Данный район содержит большое количество повторенных последовательностей, что считается одним из основных механизмов возникновения крупных перестроек в этом гене [27]. Ген включает в себя 24 экзона, 22 из которых кодирующие, и еще один экзон является кодирующим лишь в одной из изоформ мРНК. Кодируемый геном белок *BRCA1* состоит из 1863 аминокислот и имеет вес 220 кДа.

Мутации в гене *BRCA1* с высокой вероятностью (до 72%) в течение жизни пациента приводят к развитию РМЖ и РЯ [2], а также вносят вклад в риск развития рака фаллопиевых труб [28, 29], поджелудочной железы и простаты [30]. Несмотря на некоторую тканеспецифичность в развитии опухолей, обусловленных мутациями в гене *BRCA1*, данный ген участвует во многих процессах, происходящих во всех клетках организма (таблица 1, рисунок 2). Так, ген имеет строгую зависимость экспрессии от клеточного цикла [31–34]: в G_0 и G_1 фазах наблюдается низкий уровень экспрессии, а затем он возрастает от G_1 к G_2/M . Также показано, что *BRCA1* вовлечен в прохождение клеткой контрольных точек между G_1-S и G_2-M [31, 35–41]. Основной функцией *BRCA1* считается участие, по крайней мере, в двух механизмах ДНК-репарации двуцепочечных разрывов: негомологичное соединение концов (НГСК) и гомологичная рекомбинация (ГР) [27, 42, 43]. Сегодня считается, что *BRCA1* регулирует оба механизма НГСК: классический и альтернативный – выбирая между ними в зависимости от стадии клеточного цикла [44]. В других исследованиях было показано, что *BRCA1* участвует в

регуляции транскрипции [45, 46], перестройке хроматина [47, 48] и, возможно, даже входит в холофермент РНК-полимеразы II [49, 50]. Однако до сих пор не известно ни механизмов участия BRCA1 в этих процессах, ни достоверных данных, а лишь косвенные эксперименты, чаще всего проведенные на отдельных доменах белка. Кроме того, неизвестной остается третичная структура большей части белка. Таким образом, потенциально продукт гена *BRCA1* вовлечен во многие процессы на клеточном уровне (рисунок 2).

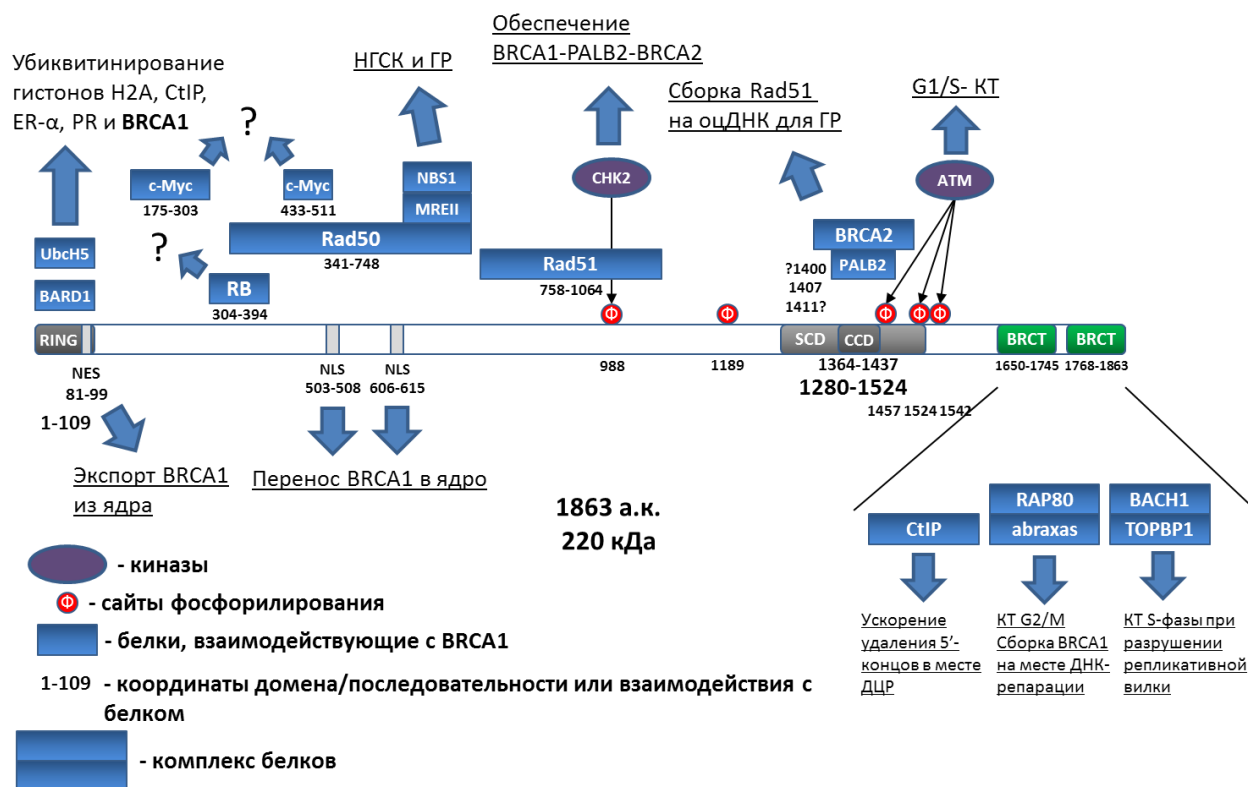


Рисунок 2. Схема строения и функционирования белка BRCA1. На схеме представлены домены белка RING, SCD и BRCT, все белки и комплексы белков, которые с ними взаимодействуют, а также для чего данное взаимодействие необходимо. Для белков, для которых известны координаты взаимодействия с BRCA1, эти координаты указаны под соответствующим прямоугольником. Масштаб соблюден. Основа для схемы была взята из работы Clark [51] и была дополнена из литературы и приведена к правильному масштабу. НГСК – негомологичное соединение концов; ГР – гомологичная рекомбинация; оцДНК – одноцепочечная ДНК; КТ – контрольная точка; ДЦР – двуцепочечный разрыв; ER – α-эстрогеновый рецептор; PR – прогестероновый рецептор.

Таблица 1. Функции и активности белка BRCA1

Функция/Активность	Домен/Последоват.	Прямое взаимодей.	Непрямое взаимодей.	Ссылки
Е3-убиквитин-лигазная активность	RING	BARD1, UbcH5	Гистоны H2A	[52, 53]

Перемещение в ядро	NLS	импортин -α	-	[54]
Перемещение из ядра	NES	CRM1	-	[55–57]
Гомологичная рекомбинация	CCD и S988	PALB2	BRCA2	[58–61]
ДНК-репарация при репликации	BRCT	BACH1	TOPBP1	[62]
Прохождение контрольной точки S-фазы	SCD (S1387)	ATM	комплекс MRN	[41]
	BRCT	BACH1	TOPBP1	[63]
Прохождение контрольной точки G2/M	BRCT	Abraxas	RAP80	[64, 65]
	BRCT	CtIP	комплекс MRN	[66]
	SCD (S1423, S1524)	ATM	комплекс MRN	[67]
Рестрикция концов ДНК при репарации	BRCT и RING	CtIP	комплекс MRN	[66, 68]
Сборка комплекса на месте повреждения ДНК	BRCT	Abraxas	RAP80	[19, 20]
Регуляция транскрипции	BRCT	?	?	[45, 46]
Перестройка хроматина	260–553	BRG1	комплекс SWI/SNF	[47, 48]

Ген *BRCA2*

Ген *BRCA2* (синонимы: *FANCD1*, *FAD*, *BROVCA2*, *BRCC2*, *FAD1*, *FANCB*, *FANCD*, *GLM3*, *PNCA2*) цитогенетически располагается на длинном плече 13й хромосомы в позиции 13q12.3, абсолютные координаты на хромосоме 32889617 – 32973809 п.о. (сборка GRCh37). Так же, как и для гена *BRCA1*, данный район содержит большое число повторенных последовательностей, что предположительно является одной из причин возникновения здесь крупных перестроек. Ген включает в себя 27 экзонов, 26 из которых кодирующие, и относится к классу генов-супрессоров опухоли. Кодируемый геном белок *BRCA2* состоит из 3418 аминокислот и имеет вес 384 кДа. Так же, как и в случае гена *BRCA1*, мутации в гене *BRCA2* в течение жизни приводят с высокой вероятностью (до 80% [69]) к РМЖ или РЯ, а также увеличивают риск развития рака груди у мужчин, рака поджелудочной железы, простаты и других тканей [70]. Основной функцией белка *BRCA2* является участие в ДНК-репарации двуцепочечных разрывов с помощью механизма ГР [43, 71], в котором участвует и белок-рекомбиназа Rad51. Также во многих работах было показано, что *BRCA2* участвует в контроле правильности прохождения митоза [72, 73], а его отсутствие приводит к нарушениям числа хромосом. На сегодняшний день до конца структура белка неизвестна. Однако некоторые домены и функциональные блоки были описаны (рисунок 3).

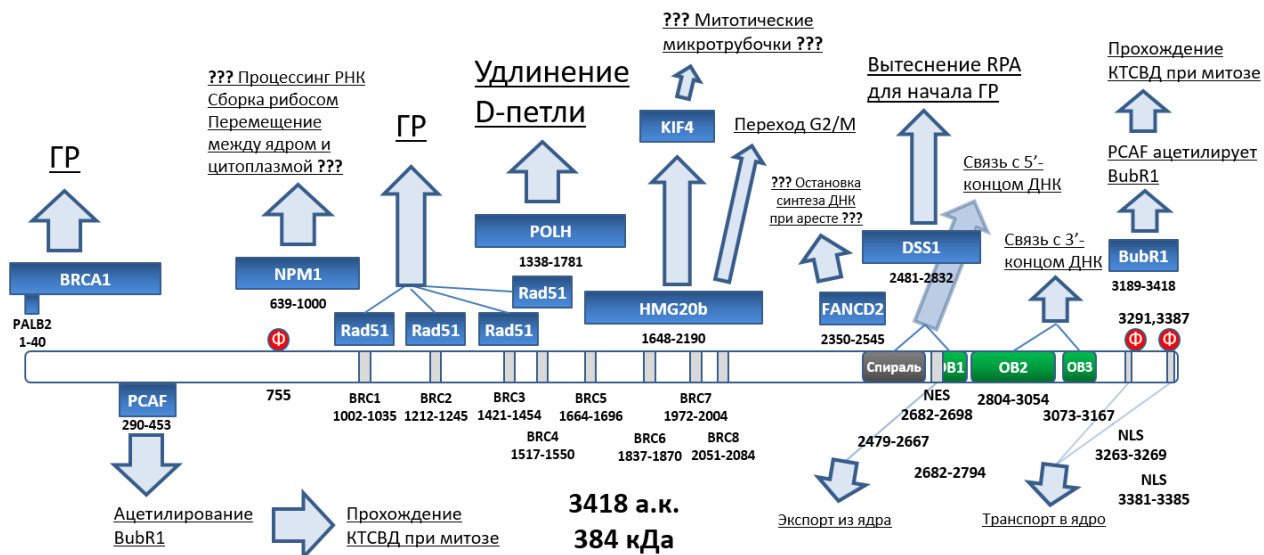


Рисунок 3. Схема строения белка BRCA2. На схеме представлены домены белка BRC1-BRC8, α-спиральный домен, OB1-OB3, а также последовательности NES (экспорта из ядра) и NLS (локализации в ядре); все белки, которые с ними взаимодействуют, а также для чего данное взаимодействие необходимо. Для белков, для которых известны координаты взаимодействия с BRCA2, эти координаты указаны под соответствующим прямоугольником. Масштаб соблюден. За основу была взята схема из [74]. Были добавлены данные из литературы и GenBank, а также произведено расположение элементов согласно масштабу. ГР – гомологичная рекомбинация; КТСВД – контрольная точка сборки веретена деления.

Методы выявления мутаций в генах BRCA1/2

Первым доступным коммерческим тестом для анализа генов BRCA1/2 стал BRCAAnalysis®, минимальная стоимость которого составляла порядка 3000–4000 долларов США на одного пациента. Сам метод был основан на секвенировании по Сэнгеру и включал в себя исследование всей кодирующей последовательности генов BRCA1 и BRCA2 [75]. При этом для пациента альтернативных вариантов не было, так как компания имела 5 патентов, покрывающих как последовательности самих генов, так и их использование, в том числе в некоммерческих организациях, включая научные институты [76]. Действие патентов было остановлено судом в 2010 году, признав их некорректными. Другими проблемами для использования секвенирования по Сэнгеру для генов BRCA1/2 являются большое число экзонов и большая протяженность генов. Поэтому для скрининговых исследований применялись такие методы проверки наличия наиболее распространенных мутаций, как олигонуклеотидная гибридизация, аллель-специфичная ПЦР, гетеродуплексный анализ, одноцепочечный конформационный полиморфизм [77],

тест на укорочение белка [78] и другие. С появлением NGS стали появляться и методы по анализу всей кодирующей последовательности генов *BRCA1/2*.

Методы приготовления библиотек для таргетного секвенирования генов *BRCA1/2*

Одна из первых работ описана в статье под руководством профессора Кинг в 2010 году [79]. Тогда был разработан метод приготовления библиотеки путем фрагментации гДНК, лигирования адаптеров и отбора целевых фрагментов с помощью гибридизации для дальнейшего секвенирования на приборе GenomeAnalyzer (Illumina). Авторам удалось показать работоспособность такого подхода на 20 пациентах, выявив все выявленные ранее 15 точечных и коротких мутаций, а также 6 крупных перестроек в генах *BRCA1*, *BRCA2*, *BRIP1*, *CDH1*, *CHEK2*, *MLH1*, *MSH2*, *p53*, *PALB2* и *STK11*. Все эти мутации были ранее выявлены с помощью секвенирования по Сэнгеру всей кодирующей последовательности генов, а также мультиплексной амплификации лигированных проб (MLPA).

Позже появилась работа, в которой использовалась технология NGS на основе пиросеквенирования, однако в ней же были показаны и ограничения такого подхода, связанные с пропуском делеций или инсерций в гомополимерах длиной более 6 нуклеотидов [80]. Данное ограничение является критическим для генов *BRCA1/2*, поскольку многие распространенные мутации связаны именно с длинными гомополимерами (5382insC, 2080delA). Также появились работы, в которых использовались протяженная ПЦР [81], платформа Ion Torrent и различные варианты приготовления библиотек (**Таблица 2**). Однако основными ключевыми моментами развития методов стало появление системы баркодирования, которое позволило анализировать одновременно большое число образцов за один запуск прибора, а также возможность использования метода приготовления библиотеки на ДНК, выделенной из ПГБ. Важным условием применимости метода на практике является его тестирование на репрезентативной выборке образцов ДНК, в том числе различного качества и выделенных из различных источников. Суммируя работы, в которых были использованы одинаковые коммерческие наборы, такое условие может выполняться для коммерческих наборов (BRCA Master Kit), однако и для них в некоторых работах показаны встречающиеся проблемы с чувствительностью и/или специфичностью [82].

Таблица 2. Примеры работ, описывающих свой опыт по выявлению мутаций в генах *BRCA1* и *BRCA2*, с помощью технологии NGS. Для каждой из работ указан метод приготовления библиотеки и платформа для ее секвенирования. Собственный набор

означает, что авторы использовали собственные реагенты вместо готового коммерческого продукта.

№	Метод приготовления библиотеки	Платформа для NGS	Сомат/гермин	Число образц.	Ссылка, год
1	Фрагментация гДНК, лигирование адаптеров и отбор целевых фрагментов с помощью гибридизации	Genome Analyzer GAIIX (Illumina)	Герм.	20	[79], 2010
2	Paired End Flow Cell v4 (коммерческий набор)	Genome Analyzer GAIIX (Illumina)	Герм.	20	[79], 2010
3	Мультиплексная амплификация экзонов (собственный набор *)	454 GS-FLX (Roche)	Герм.	30	[79], 2011
4	Мультиплексная амплификация экзонов (BRCA MASTR kit)	GS Junior (Roche)	Герм.	6	[83], 2012
5	Мультиплексная амплификация экзонов (BRCA MASTR assay v1.2)	454 GS-FLX (Roche)	Герм.	63	[84], 2012
6	Амплификация длинных фрагментов с последующей рестрикцией (собственный набор)	GS Junior (Roche)	Герм.	5	[85], 2012
7	Амплификация длинных фрагментов с последующей рестрикцией (собственный набор)	Genome Analyzer II (Illumina)	Герм.	12	[81], 2012
8	Мультиплексная амплификация экзонов (BRCA MASTR kit)	GS Junior (Roche)	Герм.	28	[86], 2013
9	Мультиплексная амплификация экзонов (BRCA MASTR assay v2.0)	Ion Torrent PGM (Life Technologies)	Герм.	62	[87], 2014
10	Мультиплексная амплификация экзонов (BRCA MASTR assay v2.1)	454 GS-FLX (Roche)	Герм.	70	[88], 2015
11	Мультиплексная амплификация экзонов (собственный набор)	Ion Torrent PGM (Life Technologies)	Герм.	26	[82], 2015
	Мультиплексная амплификация экзонов (Ion AmpliSeq BRCA1/2)	Ion Torrent PGM (Life Technologies)	Герм.	32	
12	Мультиплексная амплификация экзонов (собственный набор)	GS Junior (Roche)	Герм.	10	[89], 2015
13	Мультиплексная амплификация экзонов (GeneRead v1 и v2, Ion AmpliSeq BRCA1/2)	MiSeq (Illumina)	Сомат.	68	[90], 2015
14	BRCA Tumor MASTR Plus assay	MiSeq (Illumina)	Сомат. и герм.	97	[91], 2016

В целом, все методы приготовления библиотек можно разделить на три основные группы (**рисунок 4**):

- Двухстадийная амплификация (**рисунок 4А**). На первой стадии амплифицируются целевые фрагменты; на второй – путем амплификации к полученным фрагментам добавляются индексирующие последовательности и адаптеры для секвенирования [82, 83, 89]. Данный подход используется в коммерческих наборах BRCA MASTR kit, Ion Ampliseq BRCA1/2, а также был использован в нашей работе.
- Амплификация целевых фрагментов, а затем лигирование адаптеров и индексирующих последовательностей (**рисунок 4Б**). Данный подход используется в коммерческих наборах Qiagen GeneRead BRCA1/2 panel.
- Прочие подходы, которые пока не получили широкого распространения: фрагментация гДНК, пришивание адаптеров путем лигирования и отбор целевых фрагментов с помощью гибридизации [79]; протяженная амплификация фрагментов с последующей рестрикцией (**рисунок 4В**) [81, 85]; амплификация перекрывающихся фрагментов с формированием петли [92]; а также обогащение целевыми фрагментами с помощью одномолекулярных молекулярных перевернутых зондов (single-molecule molecular inversion probes) (**рисунок 4Г**) [93]. Последний подход в настоящий момент используется компанией Roche в продуктах HEAT-Seq Enrichment Systems.

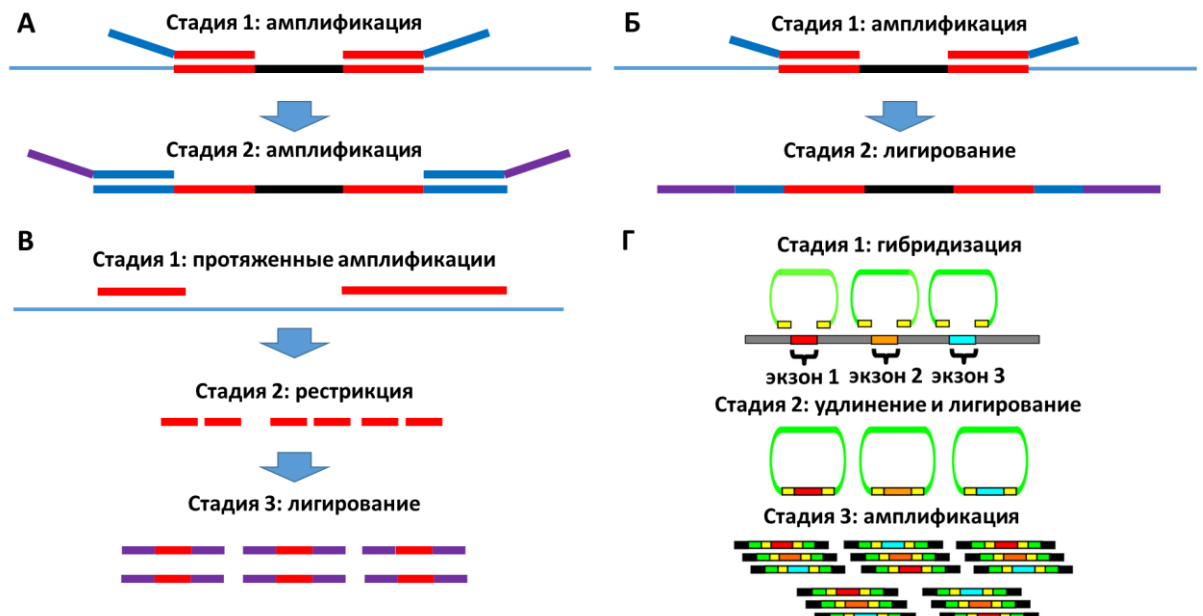


Рисунок 4. Основные методы приготовления библиотек для таргетного NGS. А. Двухстадийная амплификация. На первом этапе происходит наработка ген-специфичных последовательностей с праймеров, содержащих на 5'-конце последовательности

комплементарные 3'-концам олигонуклеотидов, добавляемых на втором этапе. На втором этапе в ампликоны вводятся индексирующие последовательности, а также адаптеры для секвенирования. **Б.** Амплификация + лигирование. На первом этапе амплифицируются целевые фрагменты, после чего с помощью лигирования к ним пришиваются адаптеры и индексирующие последовательности. **В.** Амплификация с помощью протяженной ПЦР. На первом этапе происходит наработка длинных фрагментов, которые затем разрезаются эндонуклеазами рестрикции. К полученным фрагментам с помощью лигирования пришиваются адаптеры и индексирующие последовательности. **Г.** Обогащение целевыми фрагментами с помощью одномолекулярных молекулярных перевернутых зондов. На первом этапе происходит гибридизация олигонуклеотидов, связанных между собой линкером и специфичных к целевой последовательности. Далее осуществляется удлинение одного из олигонуклеотидов и лигирование с другим. После этого полученные фрагменты амплифицируются с праймеров, комплементарных участкам линкера. Адаптировано из [93].

Каждый из подходов имеет свои недостатки и преимущества. Так, при двухстадийной амплификации возможно большее смещение покрытия между ампликонами. Однако такой подход позволяет избежать дополнительных процедур, связанных с очисткой после каждого этапа, что сокращает вероятность ошибок. Амплификация с лигированием снижает число амплификаций, так как индексирующие последовательности и адаптеры вводятся лигированием, однако после каждого этапа (амплификация, отбор фрагментов одинаковой длины и лигирование) необходимо проведение очистки фрагментов на магнитных частицах. В то же время методы, основанные на протяженной амплификации, совершенно не подходят для образцов ДНК, выделенных из ПГБ, поскольку такие образцы имеют сильную фрагментацию [94]. В свою очередь, метод с одномолекулярными перевернутыми молекулярными зондами на данный момент недостаточно изучен.

Разработка, адаптация и улучшение новых методов приготовления библиотек для выявления мутаций в генах *BRCA1/2* происходит и сегодня [92, 95, 96]. Однако до этой работы в литературе не было описано методов, разработанных, адаптированных и проверенных в России.

Технологии секвенирования NGS

Все технологии NGS можно разделить на две категории: для получения коротких прочтений (35–700 п.о.) и для получения длинных прочтений (более 700 п.о.) [97]. Длинные прочтения в большинстве случаев нужны для изучения ранее не

секвенированных геномов организмов и имеют несколько существенных недостатков, не позволяющих пока применять такие технологии в клинических исследованиях. Во-первых, стоимость в пересчете на 1 миллион прочитанных нуклеотидов гораздо выше (в 10–100 раз) по сравнению с технологиями для коротких прочтений [97]. Во-вторых, на текущий момент ни одна из технологий с большой протяженностью прочтений не может конкурировать по точности с короткими прочтениями. И, наконец, такие технологии имеют значительно меньшую пропускную способность. В связи с этим, далее будут рассматриваться только технологии с получением коротких прочтений генома.

Все «короткочитающие» NGS технологии можно разделить на два класса: секвенирование с помощью лигирования (СПЛ) и секвенирование с помощью синтеза (СПС) [97]. В первом случае зонды, меченные флуорофором, гибридизуются и лигируются с ДНК-фрагментом, после чего по флуорофору считывается, какое(-ие) основание(-я) было(-и) лигировано(-ы). Во втором случае используется полимеразы и по сигналу от флуорофора или по изменению рН определяется, какой нуклеотид был включен. В обоих подходах для усиления сигнала прибор сначала производит клонирование загруженных фрагментов библиотеки на подложке либо на частицах, после чего осуществляется само секвенирование. Технологии СПЛ имеют важный недостаток по сравнению с СПС ввиду сложного кодирования двухбуквенных зондов, и в случае ошибки в какой-либо позиции теряется информация обо всех последующих позициях (подробнее см. [97]). Однако в идеальных условиях возможно получение данных достаточного качества и с этой технологией, хотя такие условия редки в случае клинических образцов.

Все технологии СПС можно также разделить на две категории: технологии с циклическим обратимым терминированием (Illumina, Qiagen) и технологии с добавлением одного нуклеотида (454, Ion Torrent) [97]. В первом случае при каждом добавлении смеси всех терминирующих нуклеотидов (А, Т, Г и Ц), меченных флуорофорами (мечены все нуклеотиды – в случае Illumina, и только часть – в случае Qiagen, что и является их главным отличием) синтез останавливается и происходит считывание добавленных нуклеотидов в каждой точке подложки. Во втором случае в каждом цикле добавляется только один вид нуклеотидов (А, Т, Г или Ц), после чего в каждой точке оценивается количество выделившегося пирофосфата (в случае 454) или протонов (в случае Ion Torrent). Обе последние технологии имеют весомую проблему с гомополимерными последовательностями длиной более 6 нуклеотидов из-за отсутствия терминации [97]. Данный факт является критичным для исследования генов *BRCA1/2*, поскольку многие мутации располагаются именно в гомополимерах. Кроме того, в 2016 году технология 454 прекратила свое развитие [97].

В свою очередь, технология Illumina имеет на сегодняшний день наиболее низкий уровень ошибок, а также широкую представленность на рынке [97], что ведет к ее активному тестированию пользователями и большей надежности. Поэтому для исследования генов *BRCA1/2* с перспективой на применение в клинической практике нами была выбрана платформа на основе технологии Illumina – MiSeq (**рисунок 5**).

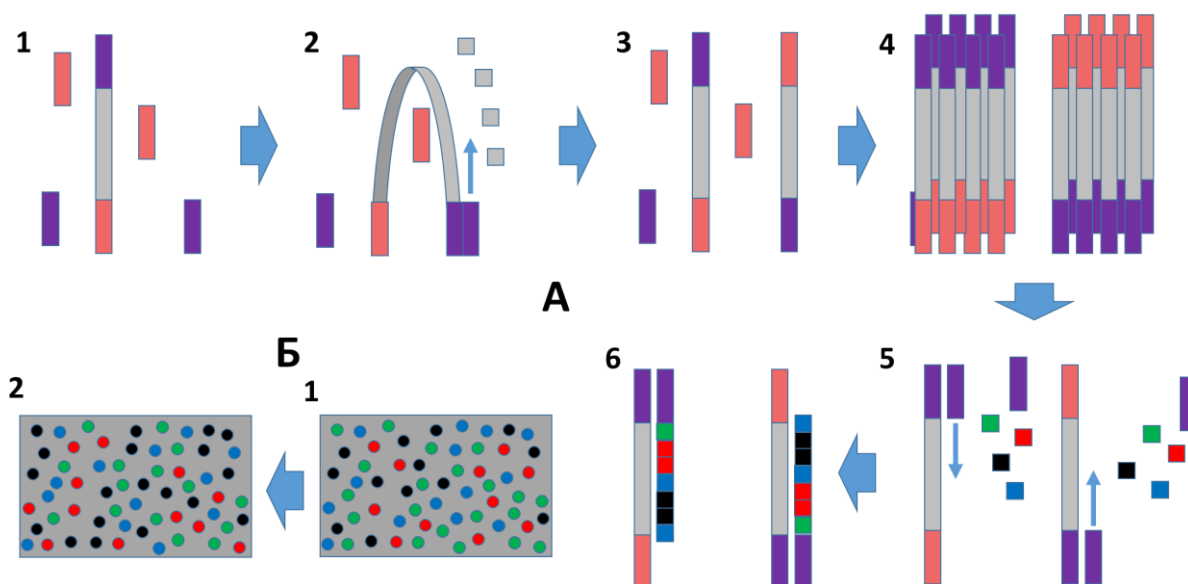


Рисунок 5. Технология секвенирования Illumina. **А.** Последовательность процессов, происходящих при секвенировании библиотеки. **1.** Загруженные одноцепочечные фрагменты, содержащие на концах адаптеры (обозначены розовым и фиолетовым цветами), прикрепляются к подложке, на которой содержатся олигонуклеотиды, комплементарные адаптерам на концах фрагментов. **2.** Далее каждый фрагмент свободным концом гибридизуется с прикрепленным к подложке праймером и начинается так называемая мостиковая ПЦР (bridge PCR). **3.** По окончании амплификации цепи расплавляются. **4.** Процесс многократно повторяется, что приводит к накоплению кластеров клонов загруженных фрагментов. **5.** Ко всем кластерам добавляется смесь меченных терминирующих нуклеотидов, а также праймеров комплементарных одному из адаптеров. **6.** Такой процесс повторяется заданное используемым набором реактивов число раз. После этого так же циклично добавляются и терминирующие меченные нуклеотиды для другого адаптера, и цепь читается с другой стороны. **Б.** Схематический пример изображений, фиксируемых оптической системой прибора и преобразуемые в последовательности нуклеотидов в каждом кластере.

Методы анализа получаемых после секвенирования прочтений

Формат файлов, содержащих прочтения NGS

Программы для обработки и анализа данных NGS

Наиболее простым алгоритмом обработки данных NGS (получаемых прочтений) является их картирование (определение позиций в геноме) и выравнивание относительно референсного генома, а затем выявление несовпадающих позиций. Именно такой подход чаще всего и используется стандартными протоколами обработки данных Illumina [98, 99]. Однако он будет работать только в случае идеальных данных и на небольшом числе образцов, и даже в этом случае возможны ложноположительные и/или ложноотрицательные мутации [99]. Поэтому далее будет представлено краткое описание некоторых программ, рекомендованных разработчиками GenomeAnalysisToolkit и позволяющих улучшить результаты обработки данных [100].

Наиболее часто используемой программой для картирования прочтений является программа BWA [101], позволяющая быстро обрабатывать миллионы прочтений благодаря разработанным ими преобразованиям последовательностей прочтений и генома. При этом программа может картировать не только последовательность прочтения целиком, но и лишь небольшую ее часть, что, с одной стороны, позволяет автоматически избавляться от последовательностей адаптеров на концах прочтений, а, с другой стороны, может картировать и прочтения праймер-димеров.

Следующая программа, используемая при обработке прочтений – программа samtools [102] – позволяет производить различного вида преобразования с файлами формата SAM (sequence alignment map – карта выравнивания последовательностей) и BAM (бинарный вид SAM) – основного формата файлов, получаемых после картирования прочтений. Во-первых, данная программа необходима для конвертации между форматами SAM и BAM. Во-вторых, samtools позволяет индексировать («составлять план») BAM-файла, что ускоряет работу с ним других программ и что требуется всеми современными программами, взаимодействующими с BAM-файлами.

Еще одна важная программа, работающая с SAM- и BAM-файлами – Picard (<https://broadinstitute.github.io/picard/>). Она также позволяет быстро анализировать и преобразовывать такого типа файлы. Так, она часто используется исследователями для выявления и удаления ПЦР-дубликатов из BAM-файлов [103]. Однако в случае, если библиотека готовится с помощью амплификации, такая процедура может только увеличить число ложноотрицательных результатов.

Программа GenomeAnalysisToolkit [104] представляет собой целый пакет программ. Наиболее полезные из них UnifiedGenotyper (для выявления вариаций), RealignerTargetCreator и IndelRealigner (для перевыравнивания прочтений вокруг инсерций

и делеций для повышения точности их идентификации), BaseRecalibrator (для пересчитывания качества каждого нуклеотида в прочтениях после их выравнивания). Каждая из них направлена на улучшение получаемых после обработки данных результатов.

Программа snpEff служит для аннотации выявляемых вариаций [105]. Так как программы по выявлению вариаций приводят только их координаты, референсные и альтернативные аллели, качество и различные технические параметры, требуется дополнительная аннотация того, в какие гены попадает вариация (и попадает ли), в какие экзоны или интроны, меняет ли аминокислоту в данной позиции кодируемого белка и т.д. Все это может быть аннотировано с помощью snpEff, при этом программа аннотирует по всем образуемым мРНК, и пользователь может выбрать нужную.

Наконец, последняя программа, которая необходима для исследования вариаций – ANNOVAR – проводит аннотацию выявленных вариаций по различным базам данных и программам предсказания функционального значения [106], включая 1000Genomes [107], dbSNP [108], ClinVar [109], ExAc, SIFT [110], PolyPhen [111] и другие. ANNOVAR также может аннотировать вариации по влиянию и расположению в гене, однако менее удобна.

Существующие на сегодняшний день программы позволяют проводить обработку данных NGS, однако при их применении по-отдельности пользователю необходимо самому подобрать программы, а также параметры к ним, проанализировав 100–200 образцов с известными в них мутациями. Поэтому разработка автоматического протокола обработки данных NGS генов *BRCA1/2*, в котором все параметры уже будут подобраны является актуальной задачей.

Статистика мутаций в генах *BRCA1/2*

Мутации, нарушающие структуру и важнейшие функции белка, в гене *BRCA1* в среднем увеличивают риск развития рака молочных желез до 80 лет до 72%, а рака яичников до 44% [2]. Для гена *BRCA2* эти же значения составляют 69% и 17%, соответственно [2]. В то же время, безотносительно возраста 46% носителей мутаций в гене *BRCA1* имеют РМЖ, 12% – РЯ, а 5% – и РМЖ, и РЯ; 52% носителей мутаций в *BRCA2* имеют РМЖ, 6% – РЯ, 2% – РМЖ и РЯ одновременно [112]. На сегодняшний день известно более 12 тысяч вариаций, записанных в базе данных ClinVar, около 5 тысяч из которых считаются мутациями, ассоциированными с развитием РМЖ и/или РЯ. Большая часть известных мутаций была найдена благодаря появлению методов таргетного NGS [79]. С тех пор были проведены скрининги с помощью NGS на носительство мутаций среди больных РМЖ и РЯ в различных странах: Германии [113], Китае [114–116],

Бразилии [117], Чили [118], Колумбии [119] и небольшое исследование было сделано даже в Ливане [120]. В то же время, в России до этой работы были проведены только исследования представленности «hotspot»-мутаций [121] и CNV [4] у больных РМЖ или РЯ [122], а также была исследована кодирующая последовательность гена *BRCA1* у больных РМЖ на наличие коротких мутаций и CNV [123].

Пенетрантность мутаций в генах *BRCA1/2*

Пенетрантность мутации в медицинской генетике – это доля людей, имеющих мутацию и ассоциированное с ней заболевание одновременно среди всех людей, несущих данную мутацию. Для генов *BRCA1/2* чаще рассчитывается общая пенетрантность для всех мутаций, которые нарушают их функции. Пенетрантность может быть оценена в течение всей жизни человека (так называемый риск развития заболевания в течение жизни), до определенного возраста или безотносительно возраста. Значения пенетрантности относительно и безотносительно возраста были приведены выше, в то же время, риск развития в течение жизни близок к риску развития заболевания до 80 лет. Поэтому далее будут рассмотрены факторы, от которых может зависеть пенетрантность мутаций, нарушающих функции белков *BRCA1* и *BRCA2*, за исключением фактора возраста.

Зависимость от позиции в гене

Показано, что мутации в определенных районах генов *BRCA1* и *BRCA2* могут вести к более высокому риску развития РМЖ и более низкому – РЯ – или, наоборот, более высокому риску РЯ и более низкому – РМЖ [112], выделяя так называемые кластерные районы РМЖ (КРРМЖ) и кластерные районы РЯ (КРРЯ). Так, в *BRCA1* выделяют три КРРМЖ: с.179 – с.505, с.4328 – с.4945 и с. 5261 – с.5563; и один КРРЯ: с.1380 – с.4062. В свою очередь, в *BRCA2* выделяют три КРРМЖ: с.1 – с.596, с.772 – с.1806, с.7394 – с.8904; и три КРРЯ: с.3249 – с.5681, с.5946delТ и с.6645 – с.7471 [112]. Специфичность мутаций в разных районах *BRCA*-генов к РМЖ или РЯ связывают с различными функциями, которые выполняют определенные участки кодируемых белков.

Зависимость от исследуемой популяции

Можно предположить, что пенетрантность мутаций в *BRCA1/2* может зависеть от анализируемой популяции. Однако результаты последней крупной работы на данных о почти 10 тысячах женщинах-носителей мутаций в генах *BRCA1/2* из Великобритании, Нидерландов и Франции не показали разницы в пенетрантности между странами [2]. В то

же время, в более ранних работах до появления NGS можно встретить результаты для небольших популяций, где пенетрантность для носителей отличалась от общей статистики [124].

Зависимость от фактора родителя

Другой интересный фактор, который может влиять на пенетрантность исследуемых мутаций – фактор родителя, от которого человек наследовал мутацию. Известно, что в процессе развития геном преимплантационного эмбриона претерпевает ряд деметилирований и метилирований, и аллели от отца и от матери могут иметь различные уровни экспрессии. И в недавней работе было показано, что копия гена *BRCA1*, полученная от отца, имеет более высокий уровень экспрессии по сравнению с копией, полученной от матери [125]. Кроме того, авторы показали, что эмбрионы, несущие отцовскую копию с мутацией, развивались медленней, чем те, что несли материнскую копию с мутацией. В том числе, в другой работе было показано, что риск развития РМЖ при наследовании мутации от отца достоверно выше, чем при наследовании от матери [126]. Однако при учете смещения результатов из-за проведения опросов, данная разница была нивелирована, что может быть связано с тем, что после фильтрации выборка стала нерепрезентативной. Поэтому необходимы дальнейшие исследования данной области.

Генетические модификаторы

Многие исследования были направлены на выявление генетических маркеров-модификаторов риска развития РМЖ и/или РЯ у носителей мутаций в *BRCA1/2*. В 2015 году была проведена работа по поиску генов-кандидатов, однако ни один из генов не показал достоверных результатов [127]. С другой стороны, ранее была показана ассоциация районов 1q32 и 4q32.3 с развитием РМЖ и РЯ у носителей мутаций в *BRCA1*, соответственно, а также 17q21.31 – с развитием РЯ у носителей мутаций в *BRCA1* или *BRCA2* [128]. Кроме того, в недавней работе было разработано полигенное значение риска, рассчитываемое исходя из 94 SNP, выявленных ранее [129]. Интересным оказался результат работы, в которой было показано, что протяженная делеция в районе CYP2A7 (19q13.2) снижала риск развития РЯ у носителей мутаций в гене *BRCA1* [130]. Другое направление по поиску генов-модификаторов риска – поиск ассоциации с SNP, связанных с дифференциальной экспрессией. В таких работах были выявлены районы 4q21 [131] и 11q22.3 [132], SNP которых влияли на риск развития РМЖ у носителей мутаций в *BRCA1/2*. На данный момент, все полученные результаты по генам- и SNP-модификаторам требуют дополнительной проверки и изучения, хотя и представляют огромный интерес

как для исследователей, так и для клинической практики.

Гомозиготы, биаллельные мутации, трансгетерозиготы

Еще одной важной особенностью генов *BRCA1/2* является отсутствие мутаций, нарушающих структуру и важнейшие функции белков и находящиеся у пациента в гомозиготном состоянии. На мышах было показано, что животные гомозиготные по определенной мутации в *BRCA1* или *BRCA2*, гибнут на эмбриональной стадии [133–136]. Также, редко встречаются и биаллельные мутации, находящиеся в транс-положении. В литературе было описано четыре таких случая [137], один из которых позже был распознан как артефакт анализа [138]. Во втором случае, второй мутацией была миссенс-мутация, ассоциация которой с развитием заболевания до сих пор остается под вопросом [139]. И, наконец, в третьей работе было показано, что у пациента, содержащего делецию в 11м экзоне и мутацию, выщепляющую 10й экзон со сдвигом рамки считывания, в транс-положении, была повышена экспрессия изоформы $\Delta 9/10$ мРНК *BRCA1* [137], приводящая к образованию функционального белка [140]. В еще одном случае, описанном в литературе, пациент так же имел одну делецию и одну миссенс-мутацию [141], что привело к развитию множественных врожденных аномалий. Похожие нарушения (в том числе, анемию Фанкони) можно наблюдать и у людей, имеющих две мутации в гене *BRCA2* [142]. Другой, более частый тип случаев – трансгетерозиготные пациенты, которые несут по одной мутации в генах *BRCA1* и *BRCA2*. В недавно опубликованной работе среди более, чем 32 тысяч женщин-носительниц мутаций в *BRCA1/2*, нашли 93 таких пациента (0,3%) [143]. Такие пациенты достоверно чаще заболели РМЖ, чем пациенты, несущие одну из тех же мутаций. Однако такой же эффект для РЯ наблюдался только для гена *BRCA1*. Таким образом, до сегодняшнего дня не найдено случаев мутаций, находящихся в гомозиготном состоянии или в транс-положении, для которых не было найдено дополнительных объяснений (как в случае с повышенной экспрессией одной из изоформ мРНК). Это же подтверждает и исследование семьи из матери, отца и шести сестер, несущих две различные мутации, но при этом эти мутации ни разу не встретились в транс-положении в одном геноме [144].

Соматические мутации

Изучение соматических мутаций у больных РМЖ и РЯ представляет большой интерес как с фундаментальной, так и с практической точек зрения. С одной стороны, данные о соматических мутациях позволяют лучше понимать механизмы онкогенеза. С другой стороны, выявленные соматические мутации могут быть поводом для назначения

более эффективных препаратов: на основе платины или ингибиторов PARP. Наиболее актуальным из литературных обзоров по соматическим мутациям в РЯ является обзор М. Moschetta и соавторов, в котором было показано, что в среднем соматические мутации в генах *BRCA1/2* наблюдаются в 5–7% всех случаев РЯ [145]. Похожие результаты были получены и в более поздних работах [91, 146]. В то же время, для больных РМЖ соматические мутации встречаются примерно в 3% случаев [147, 148]. Кроме того, во всех или большинстве случаев у пациентов с мутацией в опухолевой ткани наблюдался феномен «потеря гетерозиготности» (LOH – loss of heterozygosity). Потеря гетерозиготности означает потерю в определенном участке генома гетерозиготных аллелей, что может быть вызвано одним из двух событий: делеция участка генома в одной из копий хромосом либо замена одной из родительских копий копией от другого родителя в процессе геномной конверсии. В результате последнего варианта такой человек несет две копии гена или его части от одного родителя.

Потерю гетерозиготности связывают с механизмом развития опухоли, в частности с так называемой двухударной теорией онкогенеза Нудсона [149], которая утверждает, что первым толчком (или «ударом») к развитию опухоли является точечная мутация или же изменение в уровне экспрессии гена, однако само по себе оно не приводит к заболеванию. И только в случае второго «удара», в виде потери аллеля дикого типа (что и называется потерей гетерозиготности) клетка малигнизируется. На сегодняшний день, считается, что потеря гетерозиготности вызвана, в первую очередь, именно геномной конверсией, а не крупными перестройками типа CNV [6], однако этот вопрос требует дальнейшего изучения и может быть исследован благодаря развитию методов выявления CNV в образцах ДНК из опухолевой ткани.

Одним из ограничений анализа соматических мутаций является небольшая длина фрагментов ДНК, выделяемых из ПГБ. В большинстве случаев длина выделяемых фрагментов ДНК составляет около 200 п.о., и только в случае хорошо приготовленных гистологических блоков и подходящем методе выделения – до 600 п.о. [94]. Еще одно не менее важное требование к такому исследованию – это высокое покрытие всех исследуемых областей гена: для обеспечения возможности идентифицировать мутации с низкой частотой в исследуемой ткани. Существующие на сегодня приборы и реагенты для NGS позволяют получить любое необходимое покрытие, однако, чем выше требуемое покрытие – тем меньше образцов ДНК может быть взято в один анализ.

Таким образом, область исследования соматических мутаций, в частности CNV, в генах *BRCA1/2* является слабоизученной.

«hotspot»-мутации в генах *BRCA1/2*

Одной из проблем анализа генов *BRCA1/2* является отсутствие или слабая представленность мутаций с повышенной частотой в популяции («hotspot»-мутаций). Согласно исследованиям, частота какой-либо одной мутации среди всех носителей мутаций в *BRCA1/2* в популяции чаще всего составляет не более 11–16% [150, 151]. Однако стоит обратить внимание на тот факт, что в некоторых исследованиях, которые используют небольшое число пациентов или проводят их отбор, например, по признаку родства, или анализируют небольшие практически замкнутые популяции, частота некоторых мутаций может оказаться высокой. Примером такой выборки является популяция евреев Ашкенази, в которой мутации «основателя» (185delAG – *BRCA1*, с.66_67delAG; 5382insC – *BRCA1*, с.5263_5264insC; и 6174delT – *BRCA2*, с.5946_5946delT) составляют до 90% всех мутаций в генах *BRCA1/2* [152]. Для других популяций многие «hotspot»-мутации пересекаются с перечисленными выше, однако их частота гораздо ниже [153].

Мутации в генах *BRCA1/2* в выборках пациентов из России

Одной из первых работ по поиску мутаций в генах *BRCA1/2* на выборке пациентов из России стала работа под руководством В.А.Ponder в 1997 году [154]. В данной работе были проанализированы 19 пациентов из 19 семей, в которых было зарегистрировано по несколько случаев РМЖ и/или РЯ. При этом «hotspot»-мутации 5382insC и 4154delA (или 4153delA – старая версия обозначения данной мутации) составили 86% всех выявленных мутаций. На тот момент для исследователей были доступны три основных метода: одноцепочечный конформационный полиморфизм, анализ гетеродуплексов и тест на укорочение белка. В другой работе в 2002 году были исследованы 25 семей из России с РМЖ/РЯ, для которых было найдено четыре мутации (16%), три из которых были 5382insC [155]. Такой же результат был получен и для 52 больных РМЖ, для которых мутации были выявлены в 16% случаев [156]. До некоторого времени размеры выборки всех проводимых в России исследований распространенности мутаций в генах *BRCA1/2* среди больных РМЖ/РЯ не превышали 100 образцов [157, 158]. Впервые 177 образцов больных РЯ на наличие мутации 4154delA были исследованы в работе Крылова Н.И. и соавторы [159]. Данная мутация была найдена у 1% пациентов. В другой работе также под руководством Имянитова Е.Н. с помощью ПЦР были исследованы 302 образца больных РМЖ, для 15% из которых были выявлены мутации, ассоциированных с развитием заболевания [160]. В 2009 году был разработан метод детекции мутации 5382insC на основе технологии «дуплекс-скорпион», использующей флуоресцентно меченные

олигонуклеотиды, что расширило разнообразие методов для выявления мутаций [161]. В 2012 году на выборке из 7920 контрольных образцов и 564 образцов больных РМЖ из Сибирского региона было подтверждено, что мутация 5382insC ассоциирована с развитием РМЖ [162]. В 2014 году в России был предложен метод выявления мутаций в генах *BRCA1/2* с использованием биочипов [163]. В 2017 году вышла статья, в которой исследовалась почти полная кодирующая последовательность гена *BRCA1* (не все экзоны были взяты в анализ) с помощью секвенирования по Сэнгеру на выборке больных РМЖ из Республики Бурятия [164]. При этом были найдены шесть новых вариантов с неизвестным клиническим значением.

Таким образом, для русской популяции так же, как и для многих других, характерно наличие нескольких «hotspot»-мутаций, которые рекомендованы к скрининговым исследованиям [5]:

- 5382insC (*BRCA1*, NM_007294.3: c.5266dupC);
- 4154delA (*BRCA1*, NM_007294.3: c.4035_4035delA);
- 185delAG (*BRCA1*, NM_007294.3: c.66_67delAG);
- Cys61Gly (*BRCA1*, NM_007294.3: c.181T>G);
- 3819delGTAAA (*BRCA1*, NM_007294.3: c.3700_3704delGTAAA);
- 3875delGTCT (*BRCA1*, NM_007294.3: c.3756_3759delGTCT);
- 2080delA (*BRCA1*, NM_007294.3: c.1961delA);
- 6174delT (*BRCA2*, NM_000059.3: c.5946_5946delT).

Однако, в связи с отсутствием до этой работы недорогих и эффективных методов исследования всей кодирующей последовательности генов *BRCA1/2* в России, кроме вышеперечисленных мутаций, практически никакие более редкие мутации так и не были изучены. Исключение составляет работа, в которой с помощью цифровой ПЦР и MLPA были исследованы CNV у больных РМЖ и частично – РЯ (только одна из мутаций была проверена на выборке из 184 пациентов) [4]. Из результатов следует, что CNV обнаруживаются примерно у 3% больных РМЖ с подозрением на наследственную форму. Также была проведена небольшая работа по исследованию части кодирующей последовательности гена *BRCA1* у больных РМЖ с одним и более риском развития заболевания (наличие больных РМЖ родственников, раннее развитие заболевания и др.) на наличие коротких мутаций и CNV с помощью ПЦР в режиме реального времени и секвенирования по Сэнгеру [123]. Поскольку не была исследована вся кодирующая последовательность гена *BRCA1* и не был исследован ген *BRCA2*, необходимо дальнейшее изучение распространения мутаций в генах *BRCA1/2* в русских популяциях.

Заключение

Анализ литературы показал, что РМЖ и РЯ – наиболее частые и смертоносные виды рака у женщин, наследственная природа которых предполагалась еще с 1948 года. Сегодня уже известно, что именно гены *BRCA1* и *BRCA2*, в первую очередь, отвечают за наследственную составляющую данных заболеваний и разработаны методы для исследования их кодирующей последовательности с помощью технологии NGS. Несмотря на то, что методы таргетного NGS генов *BRCA1/2* применяются во всем мире еще с 2010 года, в России не было разработано ни одного метода приготовления библиотек или же программ для автоматической обработки получаемых данных. Также в литературе не описано ни одного метода исследования генов *BRCA1/2* с помощью NGS, который бы позволял с высокой точностью выявлять SNV, короткие инсерции и делеции, а также CNV, как в образцах ДНК, выделенных из крови, так и выделенных из ПГБ. Развитие таких методов даст толчок к исследованиям процессов, происходящим до и во время развития опухоли. Другим важным последствием появления таких методов, в том числе в России, станет повышение продуктивности исследования генов *BRCA1/2* у пациентов с уже диагностированным РМЖ и/или РЯ.

Во многих странах уже проведены скрининговые исследования всей кодирующей последовательности генов *BRCA1/2* с помощью NGS, что дает возможность сказать о распространении мутаций в популяции, в том числе о представленности «hotspot»-мутаций. В то же время, в России было изучено только распространение «hotspot»-мутаций и CNV только для больных РМЖ и, в основном, в гене *BRCA1*. Поэтому необходимо более полное изучение данной проблемы.

Таким образом, разработка собственного метода по выявлению мутаций в генах *BRCA1/2* с помощью NGS и его применение на клинических образцах ДНК, выделенных из крови и гистологических блоков, с целью изучения представленности мутаций в русской популяции имеет не только научно-фундаментальное, но и прикладное значение, поскольку имеет непосредственное применение в клинической практике.

Глава 2. Материалы и методы

Материалы

Пациенты

Всего в исследовании было проанализировано 1112 образцов ДНК, выделенных либо только из крови (592), либо только из ПГБ (64), либо из обоих материалов (228) от 884 пациентов. Большинство образцов (608 образцов для 426 пациентов) были использованы для отработки методов, и выявленные в них вариации и мутации представлены не будут, так как данная часть работы проводилась по закрытому исследованию или выявленные вариации будут использованы в другой работе.

В исследование вошли следующие выборки пациентов:

А). 96 больных РМЖ, имеющих указания на наследственный характер заболевания (два случая РМЖ до 50 лет в одной семье, три и более случая РМЖ в одной семье, билатеральный РМЖ, РМЖ и семейная ассоциация с другими типами рака, РМЖ в молодом возрасте: до 40 лет), но не несущих «hotspot»-мутаций. Все пациенты наблюдались в Алтайском краевом онкологическом диспансере в период с 2006 по 2008 г. и проживали на территории Алтайского края РФ. Все 96 пациентов были взяты из выборки из 337 пациентов, для каждого из которых ранее был проведен прескрининг на носительство наиболее частых мутаций в генах *BRCA1* (5382insC, C61G, 4154delA, 185delAG, 2080delA, 3819del5, 3875del4) и *BRCA2* (6174delT, 9318del4, 1528del4). При этом было найдено 38 носителей (11,3%), ни один из которых не был взят в текущую выборку. 75 образцов ДНК были выделены из клеток периферической крови, 14 – из капли крови, высушенной на бумаге и 7 – из буккального эпителия.

Б). 144 больных раком предстательной железы, пациенты Барнаульского филиала ФГБУ «Российский онкологический научный центр им. Н.Н. Блохина» и КГБУЗ «Красноярский краевой клинический онкологический диспансер им. А.И. Крыжановского». Все образцы ДНК (144) были выделены только из крови и были использованы только для отработки методов, результаты выявления мутаций будут опубликованы в другой работе.

В). 107 больных раком яичников, пациенты ФГБУ «Российский онкологический научный центр им. Н.Н. Блохина». Для 24 пациентов ДНК была выделена из крови и ПГБ, для 81 – только из крови, для 2 – только из ПГБ. Из анализа NGS были исключены образцы крови двух пациентов, несущих «hotspot»-мутации 5382insC, определенных с

помощью количественного ПЦР.

Г). 103 больных раком яичников, пациенты Барнаульского филиала ФГБУ «Российский онкологический научный центр им. Н.Н. Блохина». Для 20 пациентов ДНК была выделена и из крови, и из ПГБ, для 80 – только из крови, для трех – только из ПГБ. Из анализа NGS были исключены образцы крови 12 пациентов, несущих «hotspot»-мутации 5382insC и 4154del4, определенных с помощью количественного ПЦР.

Д). 7 больных билатеральным раком молочных желез, пациенты ФГБУ «НИИ онкологии им. Н.Н. Петрова». Все образцы ДНК (7) были выделены из крови.

Е). 145 пациентов – больные раком яичников, пациенты различных онкологических центров Сибири и Дальнего востока. Большая часть образцов (84) – пациенты Новосибирских онкологических центров. Для двух пациентов ДНК была выделена и из крови, и из ПГБ, для 111 – только из крови, для 32 – только из гистологического блока. Из анализа NGS были исключены образцы крови 42 пациентов, несущих «hotspot»-мутации 5382insC, 4154del4, 2080delAA, 3819del5, 300A/C, определенных с помощью количественного ПЦР.

Ж). 282 пациента – больные раком яичников, пациенты различных онкологических центров Сибири и Дальнего востока. Для 182 пациентов ДНК была выделена и из крови, и из ПГБ, для 73 – только из крови, для 27 – только из гистологического блока. Все образцы были получены при выполнении закрытой программы и использованы только для отработки методов, результаты выявления мутаций по договоренности не будут представлены в этой работе.

Все пациенты выборок В, Г, Е и Ж имели позитивный ответ на лечение цисплатином.

Реактивы

Акриламид, N,N'- метилен-бисакриламид, N,N,N',N' – тетраметилэтилендиамин (ТЕМЭД) («ApplyChem», Германия), Tween – 20 («Serva», США), трисгидроксиметиламинометан (Трис), этилендиаминтетрауксусная кислота (ЭДТА), додецилсульфат натрия (SDS) («ICN Biochemicals», США), минеральное масло (ЗАО «ВЕКТОН», Россия), дезоксинуклеозидтрифосфаты (dNTP), перегнаный этиловый спирт 96%; изопропанол («Химмед», Россия).

Все остальные реактивы: HCl, NaCl, NaAc, MgCl₂ хлороформ, изопропанол, изоамиловый спирт, персульфат аммония – отечественного производства и имели категорию не ниже химически чистого.

Растворы и буферы

0,5× TBE: 0,045 М Трис-борная кислота; 0,001 М ЭДТА (pH 8,0).

Буфер нанесения образцов ДНК в полиакриламидный гель: 0,5% бромфеноловый синий, 0,5% ксиленцианол, 50% глицерин в воде.

Раствор бромистого этидия: сток – 10мг/мл, рабочий раствор – 0,5мкг/мл в H₂O.

30% раствор акриламида (30 г акриламида и 0,8 г бисакриламида на 100 мл воды).

TE буфер: 10мМ Трис-НСl (pH=8,0), pH 7,4, 1мМ ЭДТА.

Фенол, уравновешенный буфером TE, pH 8,0.

Раствор А для депарафинизации (NaOH 0,1М, Tween-20 0,5%).

Буфер Б (10мМ Трис-НСl pH 8,0, 100мМ NaCl, 10мМ ЭДТА)

Готовые коммерческие наборы

QIAamp® DNA Blood Mini Kit (50) (Qiagen), GeneRead BRCA panel v2 (Qiagen).

Ферменты

В работе использовали ДНК-полимеразу AmpliTaq Gold (Life Technologies), ДНК-полимеразу Taq-Klen (ЗАО «Биосан», Россия), протеиназу К («Serva», США).

Олигонуклеотиды

Дезоксирибонуклеотидные праймеры синтезированы в Институте химической биологии и фундаментальной медицины (ИХБФМ СО РАН).

Методы

Выделение ДНК

724 образца ДНК были выделены из венозной крови с использованием стандартной процедуры, включающей выделение и лизис мононуклеарных клеток периферической крови, гидролиз белков протеиназой К, очистку ДНК экстракцией примесей фенол-хлороформом и осаждение ДНК в этаноле [165].

21 образец ДНК были выделены при помощи набора реагентов QIAamp® DNA Blood Mini Kit (50) (Qiagen) согласно инструкции производителя (14 из капли крови, высушенной на бумаге и 7 – буккальный эпителий). Концентрацию ДНК определяли при помощи Fluorescent DNA Quantitation Kit (Bio-Rad).

292 образца ДНК были выделены из ПГБ по следующей методике. Сначала нарезали парафиновые блоки на микротоме (толщина среза 5–10 мкм). Получив 5–7 срезов, помещали их в центрифужную пробирку Eppendorf емкостью 1,5 мл и проводили депарафинизацию. Для этого добавляли 400 мкл свежеприготовленного раствора А. Затем

нагревали полученную смесь при температуре 99°C в течение 10 мин. Центрифугировали 10 мин при 10'000g. Содержимое пробирки разделялось на 3 фазы: 1 – парафин, 2 – водная фаза, 3 – осадок, содержащий ткань. ДНК содержится, как в водной фазе, так и в 3 фазе. Для последующего выделения использовался только осадок. Его переносили в чистую пробирку. Добавляли 300мкл буфера Б, 40 мкл 10% SDS и 10 мкл протеиназы К (10мг/мл) и минеральное масло. Инкубировали полученную смесь при 62°C 12–24 ч. Затем отбирали водную фазу и переносили в чистые пробирки. Далее добавляли 150 мкл фенола, уравновешенного ТЕ, 150 мкл хлороформа, тщательно перемешивали до образования эмульсии. Центрифугировали 10 мин при 10'000g. Верхнюю фазу переносили в чистые пробирки, добавляли 5 мкл 0,5% ЛПАА, 30мкл 3М ацетата натрия (рН=5,0), 300мкл изопропанола. Инкубировали 15–20 мин при комнатной температуре. Затем центрифугировали 10 мин при 10'000g. Супернатант осторожно отбирали, осадок промывали 75% этиловым спиртом. Пробирки с осадком подсушивали при 37°C не менее 15 мин. После чего добавляли 100мкл буфера ТЕ (10мМ Tris-HCl рН 8,0). Хранили полученный раствор ДНК при -20°C.

Определение мутаций с повышенной частотой

Для всех образцов ДНК, выделенных из крови, проводили скринирование на «hotspot»-мутации методом асимметричной ПЦР в режиме реального времени с использованием флуоресцентно меченного зонда, комплементарного исследуемому участку ДНК. Для каждого образца проводили ПЦР с использованием пары праймеров и зонда. Структуры праймеров и зондов для типирования приведены в **таблице 3**. Общий объем реакционной смеси составлял 18 мкл, смесь содержала 40–100 нг ДНК; 1000 нМ праймера 1; 100 нМ праймера 2; 100 нМ TaqMan-зонда, конъюгированного с FAM, R6G или ROX; 200 мкМ dNTP, амплификационный буфер, термостабильную Taq-Klen-полимеразу – 0,5 е.а./реакцию. ПЦР проводили в следующих условиях: начальная денатурация 3 мин при 96 °C; затем 54 цикла, включающих денатурацию при 96°C – 3 сек. отжиг праймеров при 62°C – 3 сек. и последующую элонгацию при 72°C – 6 сек. После ПЦР регистрировали кривые плавления продуктов ПЦР в течение 100 циклов, повышая температуру на 0,4°C в каждом цикле от начальной температуры 40°C, регистрацию флуоресцентного сигнала производили на каждом цикле. ПЦР проводили в амплификаторе CFX384 Touch™ Real-Time PCR Detection System по 384 образца.

Таблица 3. Последовательности праймеров и зондов, которые использовались для выявления «hotspot»-мутаций.

	Праймер 1	Праймер 2	Зонд
5382InsC	5'- GCTTCTCTTTCTCTT ATCCTGATG-3'	5'- TGAGGGAGGGAGCT TTACC-3'	5'-R6G- CCTGGGATTCCTTG CTCGC-BHQ-3'
300AC	5'- TGCTGAAACTTCTC AACCAG-3'	5'- CTTCCAACCTAGCA TCATTACC-3'	5'-FAM- CATAAAGGACCCTG TGAAGG-BHQ-3'
185delAG	5'- TTCTTCCCTAGTATG TAAGGTCA-3'	5'- GAAATGGATTTATCT GCTCTTC-3'	5'-ROX- AAAATCTTAGAGTG TCCCATCTG-BHQ-3'
4153delA A	5'- CTTCTTGATTATTTT CTTCCAAGC-3'	5'- GGAGTTGGTCTGAG TGACAAG-3'	5'-FAM- CAGATGATGAAGAA AGAGGAACG-BHQ- 3'
3819del	5'- CCTGCTTCCAACAC TTGTT-3'	5'- GAGATGCCTTTGCC AATATTACCTG-3'	5'-R6G- AGAAGGTATATTGTT TACTTTACCA-BHQ- 3'

Приготовление библиотеки

Метод приготовления библиотеки для образцов ДНК, выделенных не из гистологических блоков

Первоначально, использовали метод приготовления библиотеки, использующий праймеры, применяемые ранее для секвенирования кодирующих последовательностей генов *BRCA1/2* по Сэнгеру, часть из которых заменили в процессе оптимизации протокола. Метод удобен для образцов с небольшим количеством ДНК, выделяемой не из гистологических блоков, поскольку в ПГБ фрагменты ДНК имеют небольшую длину. В данном методе образуемые ампликоны имеют максимальную длину около 500 п.о. (рисунок 7).

индивидуальных пациент-специфичных библиотек и включение в состав ампликонов библиотек индексирующих (баркодов) и адаптерных нуклеотидных последовательностей. Второй раунд ПЦР выполняли с праймерами-адаптерами, содержащими на 3'-конце 20-ти нуклеотидную последовательность аналогичную таковой на 5'-конце ампликон-специфичных праймеров; уникальную 8-ми нуклеотидную последовательность (i) для индексирования индивидуальных библиотек и адаптерные последовательности на 5'-конце (таблицы 3 и 4 Приложения 1). Все последовательности праймеров-адаптеров были взяты из протоколов, рекомендуемых компанией Illumina [166]. Амплификация пациент-специфичных библиотек осуществляли в тех же условиях, что и первый раунд ПЦР. Протокол амплификации включал этапы: инкубация при 94°C в течение 12 минут; 8 циклов, состоящих из денатурации при 94°C (6 сек), отжига праймеров при 58°C (10 сек), элонгации при 72°C (50 сек); заключительная элонгация при 72°C (2 мин). ПЦР проводили на приборе CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad).

Все полученные в результате второго раунда индивидуальные библиотеки ампликонов нормализовали и объединяли в эквимольном количестве. Нормализацию количества продуктов ПЦР в индивидуальных библиотеках осуществляли на основании значений максимальной флуоресценции амплификационной смеси при плавлении продуктов. Пулированный образец очищали на магнитных частицах AMPure XP. Концентрацию ДНК в образце оценивали при помощи количественной ПЦР с праймерами 5'-AATGATACGGCGACCCACCGA-3', 5'-CAAGCAGAAGACGGCATAACGA-3' и TaqMan зондом 5'-FAM-TCCCTACACGACGCTCTCCG-FQ-3'. Для построения калибровочной кривой использовали образцы, приготовленные при помощи серийного разведения стандартной библиотеки PhiX. На основании полученных результатов концентрацию пулированного образца довели до 10 нМ.

Всего данным методом были проанализированы 96 образцов ДНК больных РМЖ (выборка А).

Мультиплексный метод приготовления библиотеки

Первый метод не адаптирован для приготовления библиотеки из образцов ДНК, выделенных из ПГБ, ввиду высокой фрагментированности (выделяемые фрагменты обычно длиной около 200 п.о.), а также он не подходит для выявления CNV, поскольку все ампликоны различной длины и большинство экзонов покрыты лишь одним ампликоном. Поэтому нами был применен и оптимизирован метод амплификации кодирующей последовательности генов *BRCA1/2* фрагментами по 150 п.о. Это позволяет обойти вышеуказанные ограничения.

Для получения равномерного покрытия между разными образцами ДНК секвенируемых в одной реакции, проводили нормализацию концентрации ДНК с помощью количественной ПЦР гена IL-17. Для этого использовали следующие олигонуклеотиды: 5'-СТТGATGCTCTCGCTCTTCG -3' (прямой праймер), 5'-TGTAGCCCTGGTCAGACTG -3' (обратный праймер), 5'-FAM-STGCCGCTGCTCCTCCTCG-BHQ-3' (зонд). Программу для амплификации использовали следующую: 96°C – 3 мин, (96°C – 10 сек., 60°C – 2 мин) × 40 циклов. Для построения калибровочной кривой использовали собственные плазмидные стандарты с концентрациями 5, 100 и 10000 копий/мкл, что соответствует приблизительно 17, 333 и 33000 пг/мкл.

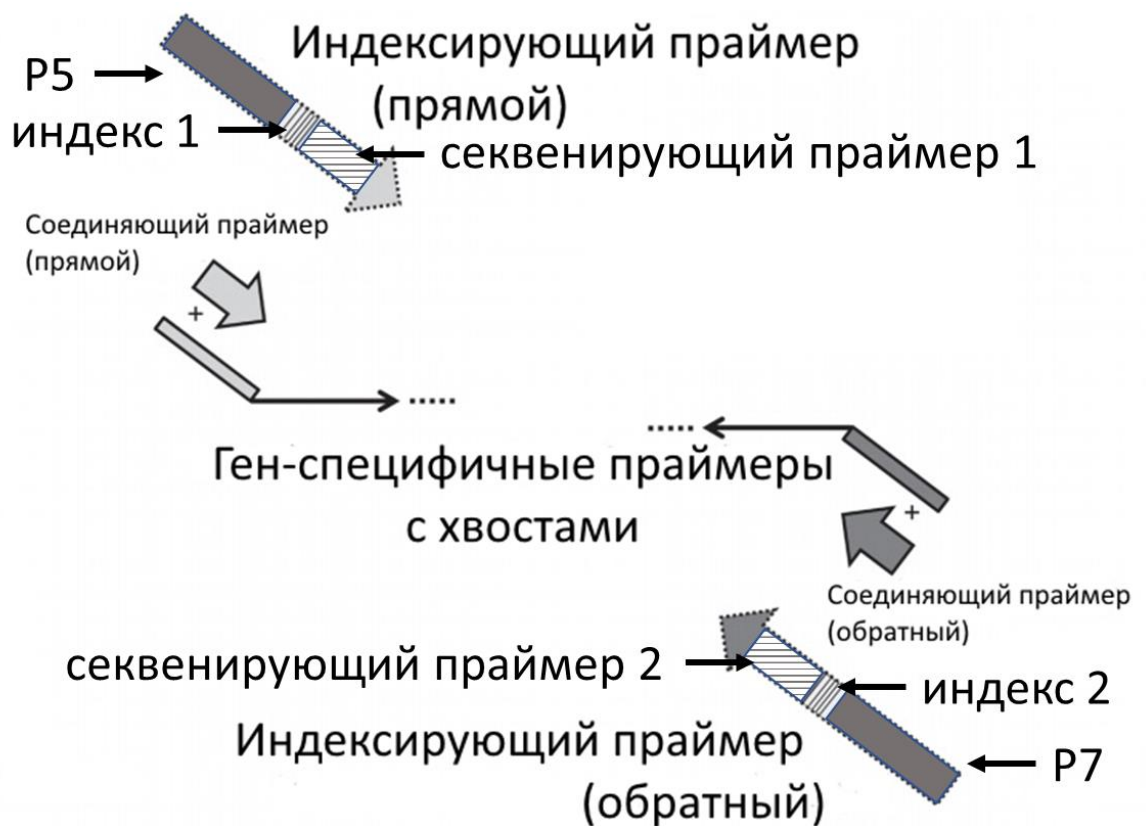


Рисунок 8. Схема приготовления библиотеки. Сначала амплификация проводится без индексированных праймеров, с низкой концентрацией ген-специфичных праймеров и высокой – соединяющих праймеров, поскольку с них амплификация идет интенсивнее, а ген-специфичные праймеры необходимы только для создания первоначальной матрицы. После наработки нужного фрагмента в смесь добавляются индексированные праймеры, содержащие баркоды, по которым различаются пациенты, последовательности P5 и P7, с которых осуществляется амплификация фрагментов в приборе MiSeq, и последовательности, обратные комплементарные секвенирующим праймерам, с которых происходит «чтение» последовательности фрагмента.

Общая схема метода приготовления библиотеки представлена на **рисунке 8** *Рисунок 8*. Последовательности ген-специфичных праймеров были любезно предоставлены Даниэлем Парком, разработчиком программы hi-plex [167] и приведены в **таблице 1 Приложения 2**. Всего в данном методе используется 190 пар ген-специфичных праймеров, расположенных друг за другом (**рисунок 9**). Для каждого образца ДНК проводили четыре мультиплексных амплификации (M1, M2, M3 и M4), по 46–48 пар праймеров в каждой. Первоначально, праймеры мультиплексировали последовательно 1-2-3-4-1-2-3-4 и т.д. (по порядку расположения комплементарных им фрагментов в гене). ПЦР проводили в конечном объеме 20 мкл, содержащем 65 мМ Трис-НСl (рН 8,9), 24 мМ сульфат аммония; 3,0 мМ MgCl₂; 0,01% Tween-20; 0,2 мМ dNTP; по 8 нМ ген-специфичных праймеров M1, M2, M3 или M4, по 0,5 мкМ соединяющих праймеров, 3–20 нг ДНК и 1 е.а. «hot-start» Taq-полимеразы («Биосинтек», Россия). В таком составе ПЦР проводили по следующей программе: 96 °С – 12 минут; (96 °С 20 с, 55 °С – 1 минута, 60 °С – 1 минута, 65 °С – 1 минута, 70 °С – 1 минута) x 18 циклов. После этого, смесь нагревали до 94 °С, амплификацию останавливали и добавляли индексирующие праймеры в конечной концентрации 1 мкМ каждый. После чего проводили еще 4 цикла (96 °С – 30с, 68 °С – 1 минута, 70 °С – 1 минута) и 68 °С 20 минут. Далее объединяли полученные фрагменты по мультиплексам и разделяли в 6% ПААГ. При вышеуказанном мультиплексировании праймеров, в геле были обнаружены неспецифичные фрагменты ДНК длиной около 190 п.о. (**рисунок 10**).

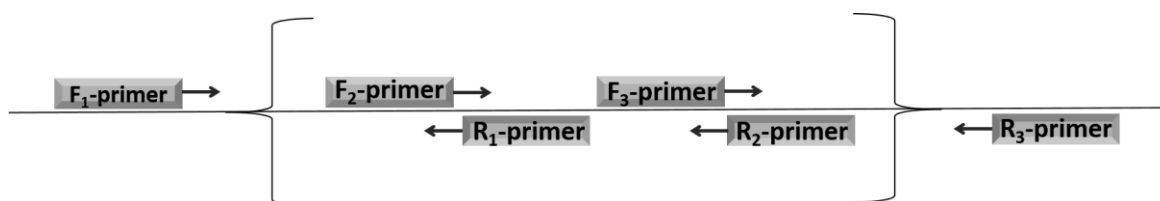


Рисунок 9. Позиции отжига праймеров. Фигурные скобки отражают границы последовательности, которую необходимо секвенировать (например, экзон гена). Ампликоны расположены таким образом, чтобы покрыть всю интересующую нас последовательность

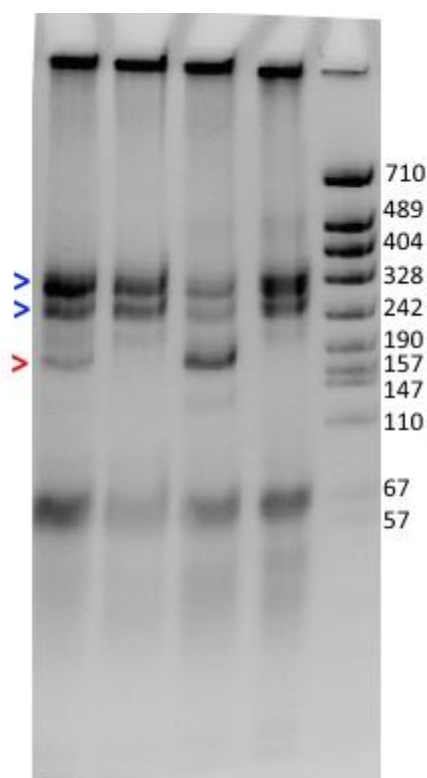


Рисунок 10. Электрофореграмма фрагментов ДНК, полученных после мультиплексной амплификации экзонов генов *BRCA1/2*. В качестве маркера длин использовался рBlueScript MSP I, длины его фрагментов указаны справа. Слева синими стрелками указаны нужные фрагменты (280 п.о. – фрагменты с двумя включенными индексирующими праймерами и 240 п.о. – фрагменты с одним включенным индексирующим праймером), красной – неспецифичный фрагмент амплификации (190 п.о.), возникший из-за определенного сочетания мультиплексированных праймеров в М1 и М3.

Далее провели поиск пар праймеров, которые давали неспецифичную амплификацию. Для этого последовательно каждый мультиплекс (М1 и М3) делили на два, пока не были найдены такие пары. В результате, пара праймеров *BRCA1_1* была перенесена из первого в четвертый мультиплекс, *BRCA1_4* – из первого во второй, *BRCA1_15* – из третьего в первый. После этого, все реакции проходили специфично. Далее вырезали нужную полосу из геля и элюировали из нее ДНК. Для этого растирали вырезанный гель в пробирке, добавляли 500мкл ТЕ и 200 мкл 5М NaCl, перемешивали, оставляли на ночь при 65°C. После короткого центрифугирования, отбирали супернатант (около 600 мкл), осаждали ДНК в изопропанол и растворяли осадок в 12 мкл буфера ТЕ, после чего смеси четырех мультиплексов смешивали в одну пробирку.

Затем проводили реамплификацию полученных фрагментов в общем объеме 25 мкл, содержащем 65 мМ Трис-НСl (рН 8,9), 24 мМ сульфат аммония; 3,0 мМ MgCl₂; 0,01%

Tween-20; 0,2 мМ dNTP, 0,3 мкМ праймеров Is5 и Is6, 10 мкл элюированной из геля ДНК и 1 е.а. «hot-start» Taq-полимеразы («Биосинтек», Россия). Программу для амплификации использовали следующую: 96 °С – 12 минут, (96 °С – 20с, 60 °С – 30с, 72 °С – 1 минута) × 5 циклов. Продукты амплификации переосаждали в изопропанол, растворяя осадки в 20 мкл Трис-НСl 10мМ.

Далее концентрацию библиотеки определяли при помощи ПЦР в режиме реального времени относительно контрольного образца (18 нг/мкл, 100 нМ). Для этого готовили следующие разведения контрольного образца: 1/5000, 1/25000, 1/125000, 1/625000, 1/3125000, а также анализируемых образцов: 1/1000, 1/10000, 1/100000. Каждая амплификационная смесь содержала: 65 мМ Трис-НСl (рН 8,9), 24 мМ сульфат аммония; 3,0 мМ MgCl₂; 0,01% Tween-20; 0,2 мМ dNTP, 0,3 мкМ праймеров Is5, Is6, 0,1 мкМ зонда (5'-HEX-CCATCTCATCCCTGCGTGTCTC-BHQ-3'), образцы ДНК и 1 е.а. «hot-start» Taq-полимеразы («Биосинтек», Россия). Программу для амплификации использовали следующую: 96 °С – 12 минут, (96 °С – 10с, 60 °С – 40с) × 36 циклов.

Данный метод был проверен на 83 образцах ДНК, выделенных из крови (часть пациентов из выборок Б-Е), после чего была получена статистика по покрытию ампликонов, используя samtools mpileup. На основании этого скорректировали концентрации праймеров. Вторая и последующая версии мультиплексного метода в дальнейшем отличались только соотношением концентраций праймеров. В дальнейшем, метод был использован для приготовления библиотек еще для двух (вторая версия) и 835 образцов (7 образцов были повторно секвенированы с первого применения данного метода): как выделенных из крови, так и из ПГБ.

Коммерческий набор GeneRead BRCA panel v2 (Qiagen)

Библиотеку готовили в соответствии с инструкцией фирмы-изготовителя. Для приготовления библиотеки с помощью набора Qiagen использовали только образцы с концентрацией ДНК не менее 2,5 нг/мкл для образцов, выделенных из крови, и не менее 10 нг/мкл – для образцов ДНК из ПГБ (согласно инструкции производителя). Предварительно все образцы ДНК развели до концентрации 2,5 нг/мкл в 10 мМ Трис-НСl (рН=8,0). Для каждого образца ДНК четыре аликвоты по 4 мкл внесли в четыре лунки 96-луночного планшета (всего по 16 мкл на образец). Далее на льду приготовили четыре смеси оставшихся реагентов, необходимых для проведения ПЦР: по 4,4 мкл GeneRead DNAseq Panel PCR Buffer (5×), 11 мкл Primer mix pool (№ 1, 2, 3 или 4) (2×), 1,5 мкл GeneRead HotStarTaq DNA Polymerase (6 U/мкл) и 0,7 мкл DNase-free water. К каждой аликвоте образца ДНК в планшете добавили по 16 мкл приготовленных ПЦР-смесей с

соответствующим номером пула праймеров (1, 2, 3 или 4), перемешивая содержимое пипетированием. Планшет(-ы) заклеили пленкой и провели амплификацию при следующих условиях на приборе CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad). 95 °C – 15 минут, (95 °C – 15с, 60 °C – 4 мин.) × 22 цикла.

По окончании программы амплификации планшет помещали на лед и проводили пулирование и очистку фрагментов на частицах AMPure XP. Для этого для каждого образца объединяли четыре использованных пула (всего 80 мкл) и переносили в новый планшет и перемешивали. Далее из полученной смеси отбирали 40 мкл и переносили в 96-луночный планшет для очистки. Добавляли 36 мкл AMPure XP и перемешивали пипетированием. Инкубировали в течение 5 минут при комнатной температуре. Устанавливали планшет на магнитный штатив для отделения частиц от супернатанта. Когда раствор становился прозрачным (порядка 5 минут), аккуратно переносили по 70 мкл супернатанта в новый 96-луночный планшет. Добавляли 64 мкл новых AMPure XP к супернатанту, перемешивая содержимой пипетированием, и инкубировали в течение 5 минут при комнатной температуре. Устанавливали планшет на магнитный штатив для отделения частиц от супернатанта. Когда раствор становился прозрачным (порядка 5 минут), аккуратно удаляли супернатант. Далее промывали дважды оставшиеся частицы в 200 мкл 80% этанола, перемещая планшет из одного положения в другое на магнитном штативе так, чтобы частицы перемещались с одного края на другой. Остатки этанола удаляли и подсушивали частицы в течение 15 минут на магнитном штативе. Для элюции ДНК с магнитных частиц добавляли 28 мкл воды, не содержащей нуклеаз, и перемешивали пипетированием. Устанавливали планшет на магнитный штатив и, когда раствор становился прозрачным, переносили 25 мкл супернатанта в новый 96-луночный планшет.

Далее для каждого пулированного образца проводили восстановление концов ДНК, для чего отбирали в новый 96-луночный планшет по 20,5 мкл полученного после ПЦР-обогащения супернатанта и добавляли к нему по 2,5 мкл End-Repair Buffer (10×) и 2 мкл End-Repair Enzyme Mix, размещая планшет на льду. Перемешивали компоненты многократным пипетированием и инкубировали в амплификаторе при 25 °C в течение 30 минут, а затем при 75 °C – 20 минут. Коротко центрифугировали и возвращали планшет на лед. Далее для каждого образца проводили аденилирование. Для этого, в каждую лунку планшета добавляли по 3 мкл A-addition buffer (10×) и 3 мкл Klenow Fragment (3'→5' exo⁻), перемешивали содержимое пипетированием и инкубировали в амплификаторе при 37 °C 30 минут, а затем при 75 °C – 10 минут. Далее для каждого образца проводили лигирование адаптеров. Для этого, в каждую лунку добавляли по 45 мкл Ligation Buffer

(2×), 1 мкл NEBNext® Multiplex Oligos for Illumina® (E7335S – для 24 образцов или E7335L – для 96 образцов ДНК), 4 мкл T4 DNA Ligase и 9 мкл DNase-free water, перемешивали пипетированием и инкубировали в амплификаторе при 25 °С 10 минут.

Далее снова проводили очистку на частицах AMPure XP. Для этого к каждому пулированному образцу ДНК добавляли по 108 мкл AMPure XP beads, перемешивали пипетированием и инкубировали 5 минут при комнатной температуре. Помещали планшет на магнитный штатив и, когда раствор становился прозрачным, аккуратно удаляли супернатант. После этого, дважды промывали частицы в 200 мкл 80% этанола, перемещая планшет из одного положения в другое на магнитном штативе так, чтобы частицы перемещались с одного края на другой. Остатки этанола удаляли и подсушивали частицы в течение 10 минут на магнитном штативе. Для элюции ДНК с магнитных частиц добавляли 19 мкл воды, не содержащей нуклеаз, и перемешивали пипетированием. Устанавливали планшет на магнитный штатив и, когда раствор становился прозрачным, переносили 17 мкл супернатанта в новый 96-луночный планшет.

Для введения индексирующих последовательностей использовали индексирующие праймеры, используемые в собственном методе. Для этого проводили ПЦР в конечном объеме 20 мкл, содержащем 65 мМ Трис-НСl (рН 8,9), 24 мМ сульфат аммония; 3,0 мМ MgCl₂; 0,01% Tween-20; 0,2 мМ dNTP; 1 мкМ индексирующих праймеров, 2 мкл приготовленной библиотеки и 1 е.а. «hot-start» Taq-полимеразы («Биосинтек», Россия). В таком составе ПЦР проводили по следующей программе: 96 °С – 12 минут; (96 °С – 30с, 68 °С – 1 минута, 70 °С – 1 минута) × 10 циклов и 68 °С – 20 минут.

Далее проводили очистку полученной библиотеки на магнитных частицах AMPure XP. Для этого добавляли к ПЦР-смесям по 16 мкл частиц AMPure XP, перемешивали и инкубировали 5 минут при комнатной температуре. Помещали планшет на магнитный штатив и, когда раствор становился прозрачным, аккуратно отбирали по 32 мкл супернатанта в новый 96-луночный планшет. После этого, добавляли по 8 мкл частиц AMPure XP к супернатанту, перемешивали пипетированием и инкубировали 5 минут при комнатной температуре. Далее снова помещали планшет на магнитный штатив и, когда раствор становился прозрачным, аккуратно удаляли супернатант. После этого, дважды промывали частицы в 200 мкл 80% этанола, перемещая планшет из одного положения в другое на магнитном штативе так, чтобы частицы перемещались с одного края на другой. Остатки этанола удаляли и подсушивали частицы в течение 10 минут на магнитном штативе. Для элюции ДНК с магнитных частиц добавляли 30 мкл воды, не содержащей нуклеаз, и перемешивали пипетированием. Устанавливали планшет на магнитный штатив и, когда раствор становился прозрачным, переносили 28 мкл супернатанта в новый 96-

луночный планшет. Концентрацию библиотеки определяли так же, как и для библиотеки, приготовленной собственным мультиплексным методом.

Данный метод был применен для 105 образцов ДНК: 82 образца из крови и 23 – из ПГБ (часть пациентов из выборки Ж).

Электрофорез ДНК в ПААГ

Для получения 6% ПААГ к 6%-ному раствору ПАА в $0,5 \times$ ТВЕ добавляли 1/100 по объему ПСА и 1/1000 по объему ТЕМЭД. Гель заливали между двумя стеклами, расположенными горизонтально и соединенными зажимами, вставляли гребенку. После полимеризации гель использовали для проведения электрофореза.

Вертикальный электрофорез проводили в буфере $0,5 \times$ ТВЕ. Раствор анализируемой ДНК смешивали с буфером для нанесения (1:10) и помещали в лунки геля. После прохождения электрофореза гель помещали на 5–10 мин в раствор бромистого этидия (10 мкг/мл) для окрашивания ДНК. Электрофореграмму документировали при помощи транс-иллюминатора и CCD-камеры Gel Doc («Bio-Rad», США).

Переосаждение ДНК изопропанолом

К раствору ДНК добавляли 1/10 объема 3М NaAc (pH=5,0), 5 мкл 0,5% ЛПАА, перемешивали и добавляли равный объем изопропанола. Инкубировали при -20°C 30 минут. Далее центрифугировали 15 минут при $10'000g$, убирали супернатант, дважды промывали осадок 70% спиртом, подсушивали осадок при 37°C в течение 5–10 минут и растворяли осадок в Трис-HCl 10 мМ или буфере TE.

Секвенирование библиотеки

Секвенирование библиотеки (10 нМ) выполняли при использовании наборов реагентов MiSeq Reagent Kit v2 (300 циклов) на платформе MiSeq (Illumina) согласно инструкциям фирмы производителя. Секвенирование проводили в ЦКП «Геномика» ИХБФМ СО РАН и в Лаборатории геномики ИМКБ СО РАН.

Анализ данных

Полученные FASTQ-файлы с прочтениями анализировали при помощи программы FastQC [168]. Первоначально, последующий анализ данных проводили с использованием разработанного нами алгоритма, не оформленного в единую программу. Картирование прочтений каждого пациента проводили на последовательности генов *BRCA1* (NC_000017.10 регион 41184133–41289677) и *BRCA2* (NC_000013.10 регион 32876987–

32986437), а также в одном из вариантов анализа на последовательности 13й и 17й хромосом (версия сборки человеческого генома GRCh37). Перед картированием каждую референсную последовательность индексировали программы BWA [101]. Картирование осуществляли функцией `mem` программы BWA. Далее полученный SAM-файл конвертировали в BAM-файл функцией `view` программы `samtools` [102], и последний сортировали и индексировали функциями `sort` и `index` `samtools` (все параметры использовали по умолчанию). На данном этапе в разных вариантах анализа проводили дополнительное выравнивание инсерций-делеций функциями `RealignerTargetCreator` и `IndelRealigner` программы `GenomeAnalysisToolkit` (GATK) [104]. Выявление SNV и инсерций-делеций проводили тремя программами: `Freebayes` [169], `samtools` и GATK. Для `Freebayes` в качестве параметров были указаны следующие: минимальное число прочтений, содержащих альтернативный аллель – 5, а минимальный процент прочтений, содержащих альтернативный аллель – 7%, остальные параметры использовали по умолчанию. Для `samtools` и GATK использовали стандартные параметры. Затем все найденные SNV и инсерции-делеции аннотировали с помощью `ANNOVAR` [106]. Дальнейшие фильтры накладывали в программе `Excel` (Microsoft). Аннотацию клинической значимости вариаций проводили с помощью разработанных нами `python`-скриптов на основании баз данных BIC (<http://research.nhgri.nih.gov/bic/>), ClinVar [109] и dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>). Кроме этого, проверяли наличие и частоту выявленных структурных вариантов в базе данных 1000Genomes [107].

В дальнейшем вся процедура анализа данных была оптимизирована (в том числе был добавлен этап вырезания последовательностей праймеров из прочтений) и оформлена в единый пакет по анализу данных NGS на MiSeq Illumina – «BRCA-analyzer», включающий в себя программу по удалению последовательностей праймеров из прочтений и основную программу по анализу прочтений. Сравнение производительности и результатов секвенирования проводили со стандартным WEB-анализатором Qiagen GeneRead Targeted Exon Enrichment Panel Data Analysis (<http://ngsdataanalysis.sabiosciences.com/NGS2/>).

cutPrimers – программа для удаления последовательностей праймеров из прочтений

Для вырезания последовательностей праймеров из прочтений нами была разработана программа на языке программирования `Python`. В качестве входных данных программа использует четыре FASTA-файла с последовательностями праймеров, расположенных на 5'- и 3'-концах R1 и R2-прочтений и FASTQ-файлы с прочтениями. Кроме того, программа может обрезать последовательности праймеров только на 5'-конце (например, при длине

ампликона 150 оснований, а длине прочтения – 75). Для чтения и записи FASTQ- и FASTA-файлов программа использует класс SeqIO из пакета Biopython [170]. Для поиска последовательностей праймеров в прочтениях cutPrimers использует два модуля Python: regex, который ищет последовательности с помощью регулярных выражений, и multiprocessing – модуль, позволяющий проводить все операции в несколько потоков. В процессе обработки прочтений программа может сохранять статистику ошибок в них. Если ошибка присутствует только в одном из прочтений, такая ошибка считается ошибкой секвенирования, если в обоих – ошибкой синтеза праймеров. Такая дополнительная информация полезна для оценки качества синтеза праймеров. Кроме того, cutPrimers может детектировать праймер-димеры, которые могут образовываться в процессе амплификации. Алгоритм cutPrimers имеет более высокую точность идентификации благодаря одновременному поиску последовательностей праймеров в обоих прочтениях. cutPrimers находится в свободном доступе по адресу: <https://github.com/aakechin/cutprimers>. Программа работает в операционных системах Linux, Windows и Mac OS.

BRCA-analyzer – программа для автоматической многопоточковой обработки прочтений

Конечный протокол обработки данных NGS генов BRCA1 и BRCA2 включает в себя следующие этапы. Обрезание последовательностей адаптеров с помощью Trimmomatic [171]; вырезание последовательностей праймеров с помощью cutPrimers; обрезание концов прочтений по качеству с помощью Trimmomatic; картирование прочтений с помощью BWA [101]; конвертация SAM-файла в BAM-файл с помощью samtools [102]; сортировка картированных прочтений с помощью samtools; добавление тэгов групп прочтений, используя picard; индексация BAM-файла с помощью samtools; перевыравнивание картированных прочтений вокруг районов с инсерциями/делециями с помощью GATK [104]; пересчет качества оснований с помощью GATK; выявление вариаций с помощью GATK UnifiedGenotyper; аннотация вариаций с помощью snpEff; конвертация VCF-файла во входной файл ANNOVAR; пересчет соотношения аллелей; объединение вариаций разных пациентов в один файл; аннотация вариаций с помощью ANNOVAR [106]; добавление позиций вариаций относительно старой референсной последовательности генов BRCA1/2 (для клиницистов); аннотация вариаций с помощью информации из базы данных VIC [172]; конвертация файла с результатами в формат Excel и разделение вариаций по их клиническому значению и релевантности. С помощью ANNOVAR программа аннотирует вариации по информации из следующих баз данных: 1000Genomes [107], ExAc [173], Esp6500 [174], Kaviar [175], ClinVar [176] and Cosmic

[177]. Кроме того, используя ANNOVAR, для миссенс-вариаций программа предсказывает влияние замены аминокислоты на функции белка по программам SIFT [110], PolyPhen [111], LRT [178], MutationTaster [179], MutationAccessor [180], FATHMM [181], RadialSVM [182] и LR. Конечный Excel-файл содержит следующие листы: «All» – все выявленные вариации, «All filtered» – все выявленные и прошедшие фильтры, «Pathogenic» – мутации, ассоциированные или вероятно ассоциированные с развитием РМЖ/РЯ (клинически значимые), «Pathogenic filtered» – мутации, ассоциированные или вероятно ассоциированные с развитием РМЖ/РЯ и прошедшие фильтры, «Predicted pathogenic filtered» – вариации, предсказанные *in silico* как нарушающие функции белка. В качестве фильтров вариаций использовались следующие параметры: значение качества, вычисленное программой по выявлению вариаций, – не менее 500 – и процент прочтений с альтернативным аллелем – не менее 14%. Клиническое значение вариаций определялось в соответствии с рекомендациями ACMG [183] с некоторыми адаптациями к анализу генов *BRCA1/2*. В качестве клинически значимых мы рассматривали нуль-варианты (нонсенс, сдвиг рамки считывания, нарушение донорного или акцепторного сайта сплайсинга, потеря иницирующего кодона), если для них не было показано, что они нейтральные (например, как это показано для K3326X или c.9976A>T в гене *BRCA2*). Также, клинически значимыми считались те вариации, для которых ранее было показано, что они имеют ассоциацию с развитием РМЖ/РЯ, и это записано в базах данных ClinVar и/или BIC (например, C61G или c.181T>G). В лист предсказанных клинически значимыми записывались миссенс-вариации, которые были предсказаны всеми программами как клинически значимые (с высокой или средней вероятностью нарушающими структуру и функции белка), аннотированы по 1000Genomes как редкие (частота менее 0,1%) и не показано, что она являются нейтральными.

Все этапы, для которых программа не указана, выполняются с помощью собственных Python-скриптов. Весь протокол от картирования упакован в единый Python-скрипт, названный BRCA-analyzer. BRCA-analyzer находится в свободном доступе по адресу: <https://github.com/aakechin/BRCA-analyzer>. Программа использует дополнительные модули Python: argparse (для чтения входящих аргументов) и multiprocessing (для многопоточности). В репозитории также расположен скрипт для оценки покрытия всех ампликонов.

Сравнение BRCA-analyzer с аналогичными программами по чувствительности выполняли на данных NGS библиотеки, полученной коммерческим набором GeneRead BRCA panel v2 (Qiagen). Всего таких образцов было 105: 82 образца из крови и 23 – из ПГБ. Для подтверждения всех найденных клинически значимых вариантов (и одним, и

другим протоколом) было использовано секвенирование по Сэнгеру. Сравнение с другими модифицируемыми автоматическими протоколами не проводили (например, bcbio-nextgen [184]), поскольку они используют те же программы, что и BRCA-analyzer, и результат зависит от выбранных программ и параметров.

Выявление CNV по данным NGS

CNV выявляли, используя собственную программу. Программа оценивает покрытие для каждого фрагмента, соответствующего каждому ампликону, после чего проводит нормализацию полученных значений относительно покрытия остальных ампликонов данного пациента и покрытия данного ампликона для всех пациентов. В конечном итоге, проводится статистическая обработка полученных нормализованных значений, выдавая в качестве результата для каждого образца список найденных делеций/дупликаций и соответствующие им значения статистики (p-value). Также для каждого образца программа строит график нормализованных значений покрытия

Определение потери гетерозиготности

Потеря гетерозиготности определялась путем оценки количества прочтений с помощью samtools mpileup [102] для всех вариантов, выявленных в образцах ДНК из крови и/или парафинированного гистологического блока, как это сделано в программе VarScan 2 [185]. Для отбора наиболее вероятных LOH использовали два критерия: p-value для соматического события (соматическая мутация или LOH) – менее 0,05 (односторонний точный тест Фишера; критерий использовался в [186]); разница между процентом прочтений, несущих альтернативный аллель (здесь и далее – аллель, отличный от референсной последовательности генома человека), в образце ДНК из крови и из гистологического блока – не менее 20% (критерий использовался в [187]). Для определения статистической значимости увеличения или уменьшения процента прочтений, содержащих альтернативный аллель, также использовался односторонний точный тест Фишера.

Статистическая обработка результатов

Статистическую обработку результатов проводили в программах Statistics 7.0, Microsoft Excel и статистическими пакетами Python (Scipy, statistics). Нормальность распределения значений покрытия ампликонов для разных пациентов оценивали с помощью теста Шапиро-Уилка.

Секвенирование экзонов генов *BRCA1/2* по Сэнгеру

Аmplификацию экзонов генов *BRCA1/2* осуществляли с использованием праймеров, приведенных в **таблицах 1–2 Приложения 1**. Амплификацию образцов ДНК выполняли в 16 мкл реакционной смеси, содержащей 10 мМ Трис-НCl (pH 8,9), 2,5 мМ MgCl₂, 55 мМ KCl, dNTP (каждый в концентрации 0,2 мМ), 0,5 U AmpliTaq Gold (Life Technologies), прямой и обратный праймеры (каждый в концентрации 300 нМ) и ДНК (10–25 нг). Протокол амплификации включал этапы: инкубация при 94°C в течение 12 минут; 38 циклов, состоящих из денатурации при 94°C (6 сек), отжига праймеров при 58°C (10 сек), элонгации при 72°C (50 сек); заключительная элонгация при 72°C (2 мин).

Секвенирование фрагментов ДНК по методу Сэнгера проводили на автоматическом секвенаторе ABI 3130XL Genetic Analyzer («Applied Biosystems», США) с использованием набора BigDye Terminator Cycle Sequencing Ready Reaction Kit («Perkin Elmer», США) согласно рекомендациям производителя. Реакционная смесь (объем 10 мкл) содержала 1–1,5 пмоль ДНК, 1 пмоль праймера и 4 мкл смеси из набора BigDye Terminator Cycle Sequencing Ready Reaction Kit. Реакцию Сэнгера осуществляли в амплификаторе Eppendorf в следующем температурном режиме:

- а) 1 цикл – 96°C 20 сек;
- б) 2 цикла – 96°C 10 сек, 64°C 4 мин;
- в) 4 цикла – 96°C 10 сек, 60°C 4 мин;
- г) 12 циклов – 96°C 10 сек, 50°C 7 сек, 60°C 4 мин.

После прохождения реакции избыток терминаторов удаляли переосаждением в изопропанол.

Глава 3. Результаты и их обсуждение

Методы выявления клинически значимых вариантов в генах *BRCA1/2* с использованием NGS включают в себя несколько основных стадий: (1) приготовление библиотеки; (2) секвенирование библиотеки; (3) анализ данных. В данной работе секвенирование библиотек проводили сотрудники ЦКП «Геномика» ИХБФМ СО РАН и Лаборатории геномики ИМКБ СО РАН, поэтому обсуждения данной части работы представлено не будет. В то же время, было проведено сравнение разработанных методов приготовления библиотек и подходов к анализу получаемых прочтений с уже существующими. Для этого были оценены общепринятые параметры.

Методы приготовления библиотеки

Методы приготовления библиотек для таргетного NGS оценивают по трем основным параметрам: (1) комплексность (в английском варианте «complexity») – отношение уникальных и повторяющихся прочтений полученной библиотеки; (2) процент целевых прочтений (в английском варианте «on target») – процент прочтений, комплементарных выбранным для анализа районам, среди всех прочтений; (3) – равномерность покрытия (в английском варианте «uniformity») – равномерность распределение прочтений по выбранным для анализа районам [188]. Первый параметр не используется для библиотек, обогащение целевых фрагментов для которых проводилось с помощью амплификации, поскольку данный показатель будет заведомо занижен по сравнению с остальными подходами. Оставшиеся два параметра были оценены для собственного мультиплексного метода и коммерческого набора реагентов с помощью программы samtools mpileup и собственных Python-скриптов. Кроме того, было проведено сравнение разработанного метода с коммерческим Qiagen по числу процедур, которые необходимо провести, и по стоимости использования для лаборатории-разработчика.

Метод приготовления библиотеки для образцов ДНК, выделенных не из гистологических блоков

Первый метод приготовления библиотек, основанный на праймерах для секвенирования по Сэнгеру, был разработан в качестве первой пробной версии (**рисунок 11**). Медианное значение покрытия каждого нуклеотида по ампликонам (86 шт.) составило 927,0 (от 182,0 до 2269,9). Медианное значение покрытия по пациентам составило 925,9 (от 240,0 до 4562,8) со стандартным отклонением 614,9. Полученные значения покрытия оказались достаточно высокими и проходят по требованиям для выявления герминальных

вариаций, которые предъявляются к лабораториям Американской коллегии медицинских генетиков (более 10–20×) [189]. Однако, поскольку данный метод не подходит для образцов ДНК, выделенных из гистологических блоков, и трудозатратен, был разработан новый мультиплексный подход.

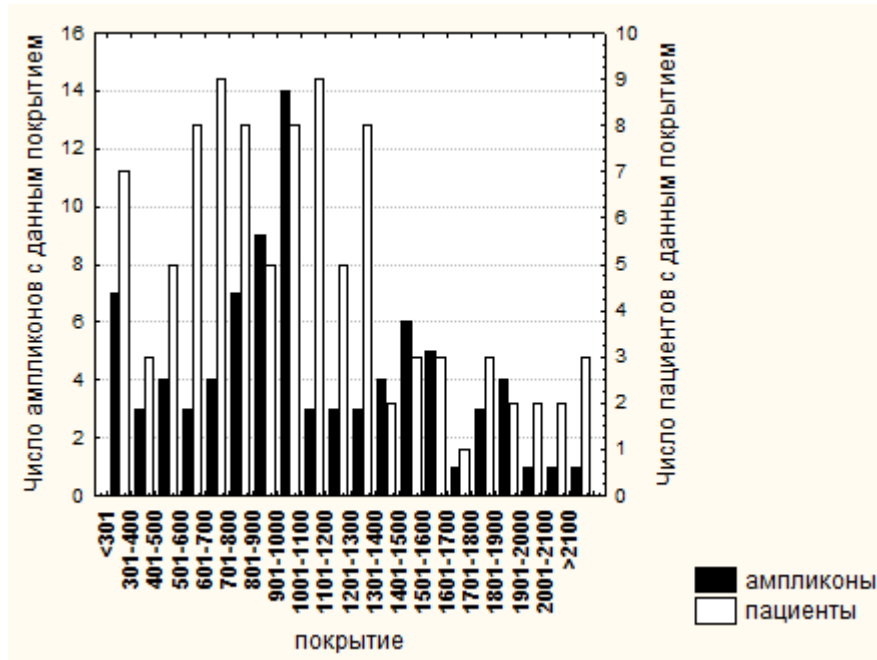


Рисунок 11. Распределение покрытия для ампликонов и пациентов.

Мультиплексный метод приготовления библиотеки

Показатели покрытия

Первая версия мультиплексного метода приготовления библиотеки («krlex1») была оценена на 83 образцах ДНК. Покрытие оценивалось таким же образом, как и для предыдущего метода. Медианное значение покрытия каждого нуклеотида составило 85,8 прочтений (от 0 до 1387,0). Число ампликонов, медианное значение покрытия которых было ниже 20 прочтений, составило 27 штук. Вероятно, это было вызвано различной эффективностью наработки целевых фрагментов для разных ампликонов. Поэтому концентрации используемых праймеров в мультиплексах были изменены по следующему принципу: если медианное значение покрытия было менее 20 прочтений – концентрации такой пары праймеров увеличивались в 4 раза; если от 20 до 30 – в 2 раза; если более 300 – концентрации праймеров уменьшались в 2 раза. Так как в первой версии большое число ампликонов не было покрыто минимум 20 прочтениями, результаты этого запуска не использовались для выявления вариаций.

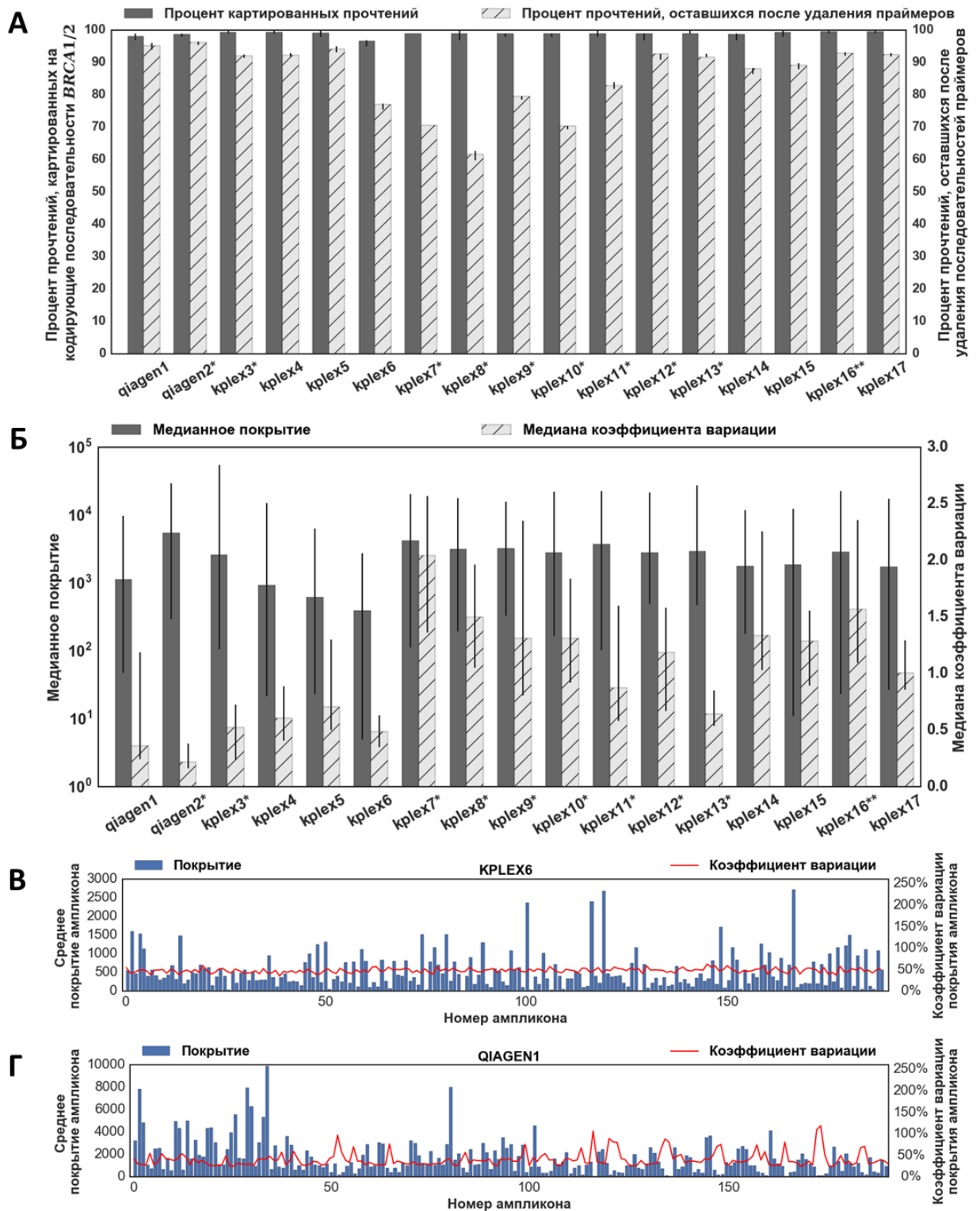


Рисунок 12. Сравнение показателей библиотек, приготовленных с помощью коммерческого набора реагентов Qiagen и собственного разработанного метода (обозначены на рисунках как «kplex» с номером запуска). **А.** Сравнение по проценту прочтений, остающихся после удаления последовательностей праймеров (правая вертикальная ось, светлые штрихованные полосы гистограммы) и по проценту прочтений, картированных на кодирующие последовательности генов *BRCA1/2* + 100 нуклеотидов

рядом (левая вертикальная ось, темные полосы гистограммы). **Б.** Сравнение по медиане средних значений покрытия ампликонов (левая вертикальная ось, темные полосы гистограммы) и медиане коэффициента вариации покрытия среди пациентов (правая вертикальная ось, светлые штрихованные полосы) для каждого из ампликонов. **В** и **Г.** Примеры разброса среднего покрытия и коэффициента вариации для KPLEX6 и QIAGEN1. Разброс значений представляет собой максимальное и минимальное значения. * - библиотеки, в которых присутствовали только образцы ДНК из ПГБ; ** - библиотеки, в которых присутствовали как образцы ДНК из ПГБ, так и из крови; в оставшихся библиотеках – только образцы ДНК из крови.

Вторая версия мультиплексного метода («kplex2») оценивалась на 2 образцах ДНК и только с целью оценки распределения покрытия по ампликонам. Медианное покрытие каждого нуклеотида составило 835,0 прочтений (от 63,1 до 4107,3). Таким образом, нам удалось добиться, чтобы все ампликоны были покрыты минимум 20 прочтениями. Для последующих запусков концентрации праймеров изменялись незначительно (не более, чем в 1,5 раза). Показатели покрытия были оценены и сравнены со значениями для коммерческого набора Qiagen (**рисунок 12**). Медианное значение покрытия для коммерческого набора составило 801,7 (от 27,4 до 2634,2).

По проценту прочтений, относящихся к целевым последовательностям, разработанный нами метод приготовления библиотеки не уступал коммерчески доступному (98,5% – для разработанного метода, против 99,0% – для Qiagen) (**рисунок 12А**). Однако для некоторых из библиотек, приготовленных с использованием собственных реагентов, наблюдалось снижение процента прочтений, в которых правильным образом идентифицировались последовательности праймеров (медиана для разработанного метода – 89,3%, для Qiagen – 95,7%) (KPLEX6–11 на **рисунке 12А**). Возможно, это связано с различиями в качестве синтеза праймеров разных партий. По медиане и разбросу значений покрытия различных ампликонов разработанный метод имел значения близкие к значениям для коммерческого набора реагентов (1337 прочтений для библиотек из образцов ДНК, выделенных из крови, против 1143 прочтений – для Qiagen). Следует отметить, что для библиотек с образцами ДНК, выделенных из ПГБ, медианное значение покрытия выше, чем для библиотек с образцами ДНК из крови. Для разработанного метода коэффициент вариации изменялся между запусками больше, чем для коммерческого набора. При этом для многих запусков равномерность покрытия была сопоставима с равномерностью для библиотек, приготовленных с помощью реагентов Qiagen (**рисунок 12В** и **12Г**).

Таким образом, разработанные методы приготовления библиотек позволяют получить последовательности всех ампликонов с высоким значением покрытия.

Число процедур при приготовлении библиотеки

Для разработанного мультиплексного метода приготовления библиотеки и коммерческого набора реагентов были оценены число процедур и время, необходимое для их проведения (**рисунок 13**). Для оценки числа процедур использовалось допущение, что минимальная процедура – это добавление какого-либо реагента в одну лунку планшета или пробирку и его перемешивание; удаление супернатанта или перенос из одной лунки в другую. Также допускалось, что продолжительность рабочего дня составляет 8 часов.

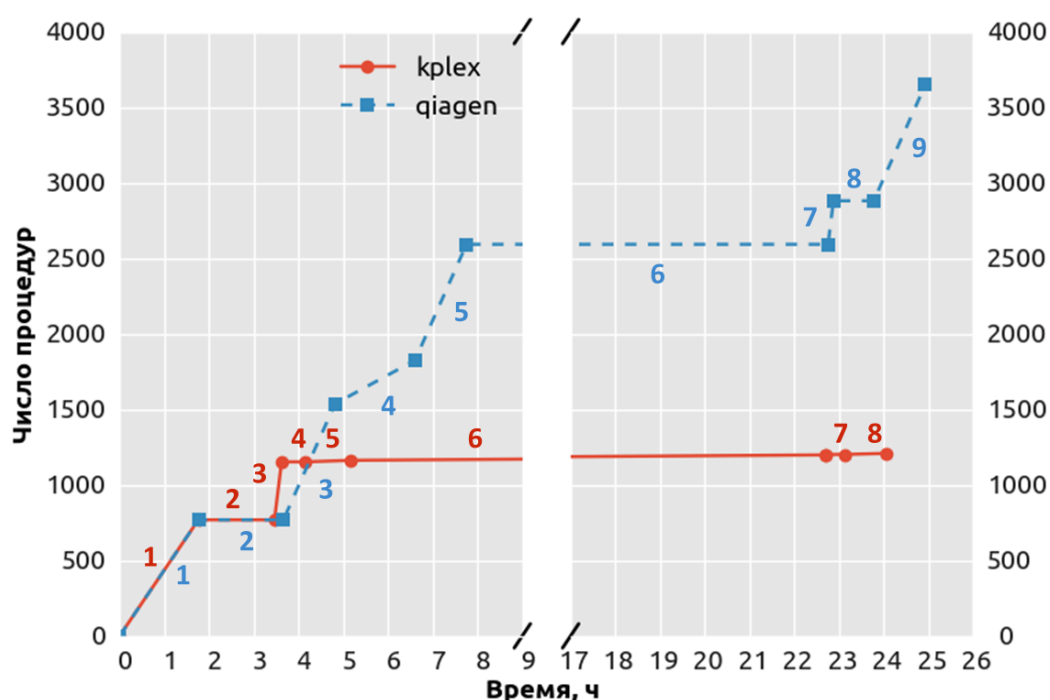


Рисунок 13. Сравнение по числу процедур (вертикальная ось) и затрачиваемому времени (горизонтальная ось) на приготовление библиотеки для 96 образцов с помощью коммерческого набора реагентов (qiagen) и разработанного мультиплексного метода (kplex). Числа рядом с прямыми линиями обозначают этап приготовления библиотеки.

Процедура приготовления библиотеки для Qiagen была разбита на следующие основные этапы (подробнее см. Материалы и методы): (1) раскапывание всех реагентов, необходимых для амплификации целевых фрагментов (для каждого из 96 образцов по четыре мультиплекса, по два реагента в каждую лунку); (2) ПЦР; (3) очистка полученных ампликонов на частицах AMPure (по 8 процедур для 96 образцов); (4) подготовка фрагментов и лигирование NEB-адаптера (по 3 процедуры для 96 образцов); (5) очистка полученных фрагментов на частицах AMPure (по 8 процедур для 96 образцов); (6)

окончание рабочего дня; (7) раскапывание всех реагентов, необходимых для включения индексирующих последовательностей (для каждого из 96 образцов по три реагента в каждую лунку); (8) ПЦР; (9) очистка на частицах AMPure (по 8 процедур для 96 образцов). Суммарно, для приготовления библиотеки с помощью Qiagen проводится 3648 процедур для 96 образцов. Время приготовления библиотеки составляет около 26 часов, с учетом прерывания рабочего дня.

Для разработанного мультиплексного метода были выделены следующие основные этапы: (1) раскапывание всех реагентов, необходимых для амплификации целевых фрагментов (для каждого из 96 образцов по четыре мультиплекса, по два реагента в каждую лунку); (2) ПЦР; (3) добавление индексирующих праймеров с помощью ПЦР (по одной процедуре на каждый из четырех мультиплексов 96 образцов); (4) ПЦР; (5) объединение образцов по четырем мультиплексам и разделение полученных фрагментов в ПААГ (суммарно, около 10 процедур для четырех мультиплексов); (6) элюция ДНК из ПААГ (около 9 процедур для каждого из мультиплексов, проходит в течение ночи); (7) раскапывание всех реагентов, необходимых для амплификации полученной библиотеки (все мультиплексы объединены в одну смесь, поэтому три процедуры) и ПЦР; (8) пересаживание продуктов амплификации в изопропанол (семь процедур). Суммарно, для приготовления библиотеки с помощью разработанного метода проводится 1208 процедур для 96 образцов. Время приготовления библиотеки составляет около 24 часов, с учетом прерывания рабочего дня.

Таким образом, разработанный метод позволяет значительно (в три раза) сократить число процедур при приготовлении библиотеки, что снижает риск контаминации образцов другими образцами. В том числе, риск контаминации снижается за счет того, что процедура внесения индексирующих последовательностей производится на первых этапах приготовления библиотеки (на **рисунке 13** – это процедура №3). В то же время, для Qiagen индексирующие последовательности вводятся только на этапе №7.

Анализ данных

Следующим этапом работы необходимо было провести анализ получаемых после секвенирования библиотеки данных, а именно прочтений загруженных в прибор фрагментов. Для этого нами использовались как уже существующие свободно доступные программы, так и разработанные в процессе выполнения работы.

Оптимизация алгоритма и подбор параметров

Оптимизация работы алгоритма анализа данных NGS генов *BRCA1/2* проводилась на

протяжении всех 20 запусков NGS (1 библиотека на основе праймеров для секвенирования по Сэнгеру, 2 библиотеки Qiagen и 17 библиотек, приготовленных мультиплексным методом). Ниже приведено описание первоначального алгоритма обработки данных, применявшегося для первого запуска, а затем алгоритма и программ, которые были применены ко всем запускам, начиная со второго.

Сортировка прочтений в соответствии с принадлежностью пациентом

Секвенирование первой библиотеки, приготовленной на основе праймеров для секвенирования по Сэнгеру, было выполнено на более старой версии программного обеспечения MiSeq Illumina, которое не разделяло прочтения по баркодам. Поэтому прочтения были разделены вручную.

В результате секвенирования пулированной библиотеки были получены файлы формата FASTQ, содержащие прочитанные последовательности. Общее число пар прочтений составило 8'282'483, длина прочтений варьировала от 249 до 251 нуклеотидов, доля GC пар составила 36 %, качество прочтений (quality score) варьировало от 32 до 39. В зависимости от сочетания нуклеотидных последовательностей баркодов (индексов) файлы FASTQ были сортированы по принадлежности к пациентам. При сравнении прочитанных индексов с соответствующими референсными последовательностями было выявлено, что 70,68% (5854421) пар прочтений содержали баркоды без замен в нуклеотидной последовательности. Остальная же часть индексов содержали одну (13,22 %), две (3,04 %) или более замен, во многих случаях равновероятно относясь в двум или более референсным индексам (13,06 %). Для последующего анализа были отобраны только пары прочтений, несущие баркоды без замен.

Картирование прочтений библиотеки. Анализ глубины секвенирования

Картирование пар прочтений было выполнено при помощи функции map программы BWA [101]. В качестве референсных последовательностей использовались последовательности генов *BRCA1* (регион 41184133–41289677 хромосомы 17 сборки GRCh37) и *BRCA2* (регион 32876987–32986437 хромосомы 13 сборки GRCh37). Доля прочтений, картированных на последовательность *BRCA1* в среднем составила $35,9 \pm 4,5\%$ (от 27,0 до 50,1% в зависимости от пациента). Доля прочтений, картированных на последовательность *BRCA2* составила $63,9 \pm 4,7\%$ (от 50,0 % до 72,8 %). Оставшиеся прочтения не имели преимущественного картирования на какой-либо район хромосом

человека.

Число прочтений (покрытие) для каждого из экзонов генов *BRCA1* и *BRCA2* составило в среднем $1610,2 \pm 693,6$ (216,9–3156,8) прочтений (**рисунок 14**). При этом для 18-ого экзона гена *BRCA2* покрытие было значительно ниже среднего (216,9 прочтений).



Рисунок 14. Распределение покрытия генов *BRCA1* и *BRCA2* по экзонам и образцам ДНК. По горизонтальной оси обозначены интервалы, содержащие различное число прочтений. По вертикальным – число экзонов (левая ось) и число образцов (правая ось) с соответствующим покрытием.

Дополнительно мы проанализировали покрытие генов *BRCA1* и *BRCA2* для каждого пациента. Мы нашли, что для 86 пациентов (90%) покрытие в среднем составило 1459 ± 513 прочтений (766–3384 прочтений, в зависимости от пациента). Для оставшихся 10 пациентов среднее покрытие было меньше 750 прочтений (см. **рисунок 14**).

Выявление коротких SNV, инсерций и делеций для 16 контрольных образцов больных РМЖ

Из 96-ти образцов случайным образом были выбраны 16 образцов, из которых была сформирована контрольная группа. Последовательности экзонов гена *BRCA1*, а также прилегающие к ним сайты сплайсинга (не менее 15 нуклеотидов) были секвенированы методом Сэнгера. При этом было определено 69 полиморфных вариантов: 11 уникальных SNV и одна делеция.

Анализ данных NGS контрольных образцов (так, как это было описано выше, но только для гена *BRCA1*) был выполнен при помощи инструмента GATK. При использовании параметров по умолчанию в последовательностях 16 образцов было определено 103 полиморфных варианта гена *BRCA1* (92 SNV и 11 делеций). То есть 32,9% полиморфных вариантов, определенных при помощи NGS секвенирования являлись ложноположительными результатами. Для того, чтобы исключить ошибочные полиморфные варианты, мы использовали несколько дополнительных фильтров, среди них: число прочтений, картированных на данную позицию (покрытие) и процент прочтений, несущих альтернативный аллель, среди общего числа прочтений в данной позиции (процент альтернативных прочтений).

Дополнительно мы сравнили три программы по выявлению SNV и инсерций/делеций: Freebayes, samtools и GATK UnifiedGenotyper. Результаты сравнения программного обеспечения и использования различных фильтров при определении полиморфных вариантов приведены в **таблице 4**. При этом сравнение проводилось по следующим значениям: число правильно определенных мутаций (верно положительные – ВП), число ошибочно определенных мутаций (ложноположительные – ЛП) и число ошибочно не выявленных мутаций (ложноотрицательные – ЛО). Наилучшие результаты достигались при использовании программы GATK с дальнейшим исключением полиморфных вариантов с показателями ниже следующих:

- покрытие менее 20 прочтений
- доля прочтений, несущих альтернативный аллель, менее 14 %

При выборе критериев отсекающего предпочтения отдавалось тем, при которых достигалась максимальная чувствительность теста ВП/(ВП+ЛО), из которых далее выбираются критерии, дающие максимальную специфичность. При использовании указанных выше параметров анализа данных NGS секвенирования чувствительность теста составляла 98,6%, специфичность – 94,4%. Специфичность определялась как ВП/(ЛП+ВП).

Таблица 4. Результаты анализа данных NGS 16 контрольных образцов при сравнении программ Freebayes, samtools и GATK с использованием следующих фильтров: минимальное покрытие и минимальный процент прочтений, несущих альтернативный аллель. Звездочкой отмечены оптимальные параметры анализа данных NGS. ВП – число верно положительных результатов; ЛП – число ложноположительных результатов; ЛО – число ложноотрицательных результатов. * - лучшие параметры для анализа данных

Программа картирования	Программа выявления мутаций	Пороговое покрытие позиции	Мин. % прочтений с альт. аллелем	ВП	ЛП	ЛО
BWA, IndelRealignment	Freebayes	100	10	61	32	8
			30	57	3	12
		20	10	65	36	4
			20	63	22	6
			30	61	3	8
	40		60	3	9	
	samtools	100	10	55	3	14
			30	55	3	14
		20	10	59	9	10
			20	59	3	10
			30	59	3	10
	GATK*	100	10	64	6	5
			30	56	3	13
		20*	10	68	7	1
			14*	68*	4*	1*
			20	65	3	4
			30	60	3	9
			40	60	3	9

Выявление коротких SNV, инсерций и делеций для 96 образцов больных РМЖ

Критерии поиска полиморфных вариантов, определенные при анализе 16 контрольных образцов, были использованы нами для определения мутаций в последовательностях генов *BRCA1* и *BRCA2* всех 96 образцов. Все SNV и инсерции/делеции, обнаруженные в результате анализа данных NGS секвенирования 96 образцов, приведены в **таблицах 5 и 6**.

Всего были выявлены 41 уникальных SNV и 20 инсерций/делеций. Все найденные полиморфные варианты были проверены при помощи секвенирования методом Сэнгера. При этом 6 найденных SNV и большая часть (18) инсерций/делеций не подтвердились.

Нами были проанализированы параметры ошибочно определенных полиморфных вариантов. Было обнаружено, что два из ложноположительных SNV располагались в

области праймеров, с которых осуществлялась амплификация экзонов генов *BRCA1* и *BRCA2*. Их наличие, вероятно, связано с ошибочностью синтеза праймеров. Оставшиеся 4 ложноположительных SNV располагались в позиции, непосредственно прилегающей к 3'-концу одного из праймеров. Покрывание для этих SNV в среднем составило 239 ± 27 прочтений (44–250), процент прочтений, несущих альтернативный аллель, составил $24 \pm 6\%$ (14–43%). Качество их секвенирования (quality или Q-score – качество, вычисленное программой для выявления вариаций) существенно варьировало и в среднем составило 1243 ± 740 (50–3376). Все четыре позиции имели прочтения как с одного, так и с другого конца, однако альтернативный аллель определялся преимущественно в той группе прочтений, которая соответствовала праймеру, к которому непосредственно прилегал SNV. В связи с этим, для этих SNV нами был проанализирован параметр FS VCF-файлов, отвечающий за вероятность ошибочного определения вариации из-за присутствия его только на прочтениях исключительно одного направления. Было выявлено, что для ошибочно определенных SNV данный показатель всегда больше 1, в то время как для правильно определенных он всегда меньше 0,001. Поэтому одним из решений такой проблемы может стать добавление в разработанный алгоритм этапа исключения таких SNV. В результате применения данного фильтра нам удалось избавиться от неправильно определенных SNV.

Из 17 ошибочно определенных делеций 11 также располагались в области праймеров. Такое число ложноположительных полиморфных вариантов в последовательностях комплементарных праймерам, использовавшихся для амплификации, по-видимому, говорит о недостаточно высоком качестве синтеза олигонуклеотидов. Одной из наиболее очевидных причин может быть неполное кэпирование 5'-концевого нуклеотида синтезируемой цепи олигонуклеотида, что приводит к увеличению фракции олигонуклеотида длиной $n-1$. В случаях если последовательность праймера содержит несколько повторяющихся нуклеотидов, то ошибка синтеза в этой области будет накапливаться, симулируя наличие однобуквенной делеции в последовательности праймера. Например, в последовательности TAAAT аденин может быть не включен в любой из трех позиций, однако при выравнивании все три ошибки будут представлены как одна.

Другим решением проблемы нахождения ложноположительных вариаций в части прочтения, соответствующей праймеру, является вырезание последовательностей праймеров из прочтений, что и было применено в дальнейшем.

Таблица 5. SNV, обнаруженные в результате анализа данных NGS 96 образцов. (#) данные приведены только для пациентов, несущих редкие (по данным 1000 Genomes Project частота аллеля <0,1%) мутации.

Подтв. секв. по Сэнгеру	dbSNP_ID	Позиция в гене <i>BRCA1/2</i>	Позиция в белке	Реф. аллель	Альт. аллель	Частота альт. аллеля в 1000Genomes	Число пациентов-носителей #	Примеч.
<i>BRCA2</i> (NC_000013.10)								
ДА	rs28897700	16358	p.Q66Q	A	G	0,0005	1	
	rs28897701	16383	p.A75P	G	C	-	1	
	rs766173	29494	p.N289H	A	C	0,0591		
	rs28897706	29607	p.S326R	C	A	0,0005	1	
	rs144848	29743	p.N372H	A	C	0,2444		
		29796	p.P389P	G	A	-	1	
	rs1801439	29994	p.S455S	A	G	0,0591		
	rs28897708	30143	p.I505T	T	C	0,0009	1	
	rs1801499	33735	p.H743H	T	C	0,0587		
	rs1799944	34477	p.N991D	A	G	0,0633		
	rs1801406	34902	p.K1132K	A	G	0,2835		
	rs80358600	35021	p.S1172L	C	T	0,0023		
	rs1799952	35022	p.S1172S	G	A	0,0028		
	rs543304	35313	p.V1269V	T	C	0,1820		
	rs28897734	36705	p.S1733S	C	T	0,0023		
		37165	p.T1887A	A	G	-	1	
	rs4987117	37250	p.T1915M	C	T	0,0116		
	rs1799955	52246	p.S2414S	A	G	0,2495		
rs28897744	53687	p.T2515I	C	T	-	2		
rs11571833	95640	p.K3326*	A	T	0,0056			
НЕТ	rs431825281	30050	p.L474P	T	C	-		только на одной цепи
	rs431825335	37290	p.E1928E	A	G	-		
		34699	p.I1065V	A	G	-		
		36008	p.N1501S	A	G	-		в области праймера
		36317	p.L1604P	T	C	-		

BRCA1 (NC_000017.10)								
ДА	rs1799967	38843	p.M1652I	C	T	0,0093		
	rs4986854	38916	p.M1628T	A	G	0,0042		
	rs1799966	38962	p.S1613G	T	C	0,3329		
	rs1800744	42356	p.S1512I	C	A	0,0005	1	
	rs1060915	50338	p.S1436S	A	G	0,3087		
	rs16942	59868	p.K1183R	T	C	0,3296		
	rs4986852	60297	p.S1040N	C	T	0,0121		
	rs16941	60303	p.E1038G	T	C	0,3082		
	rs799917	60804	p.P871L	G	A	0,4912		
	rs16940	61105	p.L771L	A	G	0,3068		
	rs1799949	61334	p.S694S	G	A	0,3296		
	rs4986850	61339	p.D693N	C	T	0,0405		
		61733	p.S561S	A	G	-	1	
	rs80357197	62165	p.N417K	A	C	-	1	
	62349	p.Q356R	T	C	0,0284			
НЕТ		61742	p.K558K	T	C	-		SNV области праймера ⁶

Таблица 6. Инсерции/делеции, выявленные в результате анализа данных NGS секвенирования 96-ти образцов. (*) Данные приведены только для пациентов, несущих редкие (по данным 1000 Genomes Project частота аллеля <0,001) мутации. **Кач.** – Качество определения мутации; **Покр.** – покрытие; **Число пац.-нос.** – число пациентов-носителей

Подтв. секв. по Сэнгеру	Тип мутации	Позиция в гене BRCA1/2	Реф. аллель	Альт. аллель	Есть в ВИС	Кач.	Покр.	Доля прочтений с альт. аллелем	Число пац.-нос. *	Примеч.
BRCA2 (NC_000013.10)										
ДА	Дел	52256	AG	-	+	5931	249	0,40	1	
НЕТ	Дел	29699	A	-	-	962	250	0,14		Делеция в области праймера
	Дел	29701	GT	-	-	196	250	0,15		
	Дел	29701	G	-	-	1038	250	0,16		
	Дел	30294	A	-	-	208	90	0,23		

	Дел	35651	A	-	-	126	250	0,61		
	Дел	36326	T	-	-	674	250	0,14		
	Дел	36329	A	-	-	870	250	0,14		
	Дел	37976	A	-	-	586	168	0,15		
	Дел	37978	T	-	-	1084	250	0,14		
	Дел	52212	C	-	-	1174	248	0,15		
	Дел	30435	A	-	+	270	183	0,15		Делеция в поли-А последовательности
	Инс	77036	-	A	+	443	250	0,16		
	Дел	77037	A	-	+	346	250	0,14		
	Дел	95907	A	-	+	171	250	0,18		
<i>BRCA1 (NC_000017.10)</i>										
ДА	Дел	72104	GG	-	-	2483	250	0,24	2	
НЕТ	Дел	60087	T	-	+	171	250	0,16		Делеция в поли-А последовательности
	Дел	61455	T	-	+	583	241	0,15		Делеция в поли-А последовательности
	Дел	62448	C	-	-	457	250	0,18		Делеция в области праймера

Оставшиеся шесть ошибочно определенных делеционных вариантов располагались в составе гомополимеров, в частности в поли-А-трактах. Предположительно, источником таких ошибок может быть неправильное включение нуклеотидов Taq-полимеразой на этапах амплификации. Этапы приготовления библиотеки содержали два раунда ПЦР: на стадии амплификации фрагментов ДНК и на стадии присоединения к полученным ампликонам адаптерных последовательностей. Так как в своей работе мы использовали Taq-полимеразу, не обладающую 3'→5' редактирующей активностью, то вероятность случайных ошибок довольно высока, в особенности при амплификации гомополимерных последовательностей.

Например, делеция 2080delA (с.1961delA) в гене *BRCA1* (экзон 11) была определена в 15% прочтений у одного из пациентов. Секвенирование методом Сэнгера соответствующего фрагмента ДНК, амплифицированного при помощи AmpliTaq Gold (Life Technologies) полимеразы показало наличие делеции 2080delA в анализируемом образце. В то же время секвенирование ампликона, полученного при помощи Pfu-полимеразы, показало наличие только неизмененного аллеля. Эти данные указывают на возможную роль полимеразы в ошибках секвенирования. Стоит отметить и тот факт, что большинство из ошибочно определенных нами инсерций/делеций, расположенных в поли-А-трактах, с различной частотой присутствуют в базе данных BIC (в частности, описанная выше делеция 2080delA). Можно предположить, что в некоторых случаях аннотированная в BIC делеция также была определена неверно.

Анализ сайтов сплайсинга

Для анализа SNV и инсерций/делеций в сайтах сплайсинга был использован список полиморфных вариантов, полученный с помощью программы GATK после картирования на 13-ю и 17-ю хромосомы одновременно. Аннотация SNV и инсерций-делеций проводилась программой ANNOVAR. В качестве района, потенциально влияющего на сплайсинг, был выбран район -40 п.о. от границы «экзон-интрон». После этого все вариации были проаннотированы по клинической значимости, на основе баз данных BIC (<http://research.nhgri.nih.gov/bic/>) и ClinVar (<http://www.ncbi.nlm.nih.gov/clinvar/>). Аннотация проводилась с помощью собственного Python-скрипта.

После фильтрации полученного списка на основании подобранных нами параметров (минимальное покрытие – 20 прочтений, минимальный процент прочтений с альтернативным аллелем – 14%) в районе «-40 п.о.» генов *BRCA1* и *BRCA2* были выявлены вариации у 43 и 95 пациентов, соответственно. Всего было выявлено 77 вариаций для гена *BRCA1* и 348 – для *BRCA2*. Из них уникальных 5 и 10, соответственно. Для гена *BRCA1* 3 из 5 были найдены в базах данных BIC и ClinVar. Для *BRCA2* 4 из 13

были найдены в BIC и 5 – в ClinVar. Ни одна из указанных вариаций по этим базам данных не обладала клинически значимым эффектом. Затем для всех SNV были определены частоты встречаемости в 1000Genomes. Для гена *BRCA1* был найден только 1 SNV, для гена *BRCA2* – 3 (таблица 7).

Таблица 7. Интронные вариации, найденные в районе «-40» от границы экзон-интрон. **Реф. ал.** – референсный аллель; **Альт. ал.** – альтернативный аллель.

Ген	ID dbSNP	Позиция в CDS	Реф. ал.	Альт. ал.	Частота в 1000Genomes	Число пациентов
<i>BRCA1</i>	rs799923	с.442-34C>T	G	A	0,098	39
<i>BRCA1</i>	rs273898672	с.80+1243C>T	G	A	-	1
<i>BRCA1</i>	rs273898667	с.80+118->AT	-	AT	-	2
<i>BRCA1</i>	-	с.80+119AT>-	AT	-	-	1
<i>BRCA1</i>	-	с.4676-25A>G	T	C	-	1
<i>BRCA2</i>	rs9534262	с.7806-14T>C	T	C	0,453	73
<i>BRCA2</i>	-	с.9648+22G>A	G	A	-	1
<i>BRCA2</i>	rs81002830	с.68-7T>A	T	A	0,0023	2
<i>BRCA2</i>	rs11571818	с.9257-16T>C	T	C	0,0046	2

Создание единого пакета анализа NGS-данных экзонов генов

BRCA1/2

Используя результаты использования первоначального алгоритма анализа данных NGS, а также литературу и учитывая новое появившееся программное обеспечение, алгоритм анализа был установлен таким, как это описано в разделе «BRCA-analyzer – программа для автоматической многопоточковой обработки прочтений» раздела «Материалы и методы».

По сравнению с алгоритмом, описанном выше, были сделаны следующие изменения:

- Введен этап удаления последовательностей праймеров из прочтений;
- Добавлен этап дополнительного выравнивания прочтений вблизи инсерций и делеций (данная процедура рекомендована создателями программы GenomeAnalysisToolkit);
- Добавлен этап пересчета значений качества нуклеотидов в прочтениях после их картирования и выравнивания;
- Аннотация по расположению в генах и влиянию на аминокислотную

последовательность кодируемого белка проводится с помощью snpEff;

- Проводится пересчет соотношения числа прочтений с референсным и альтернативными аллелями;
- Аннотация по базам данных и программа предсказания значения миссенс-мутаций проводится с помощью ANNOVAR.

Сравнение *cutPrimers* с аналогичными программами

Разработанная программа для удаления последовательностей праймеров из прочтений была сравнена со следующими аналогичными программами: cutadapt, BBDuk и AlienTrimmer. Сравнение с Trimmomatic – одной из наиболее известных программ для удаления технических последовательностей из прочтений – приведено не будет, поскольку получаемые с ее помощью результаты аналогичны результатам AlienTrimmer. Сравнение трех программ проводили по следующим параметрам: 1) процент прочтений с длиной не менее 90 нуклеотидов (все ампликоны имели длину 100 или 94), успешно обработанных программой и записанных в выходной файл; 2) время обработки всех прочтений; 3) число ампликонов, покрытых минимум 30 прочтениями (для каждой позиции); 4) медианное покрытие. Все процедуры проводились на персональном компьютере (Intel Core i4-2700K 3,5 ГГц, 32 Гб RAM).

Сравнение было выполнено на трех парах FASTQ-файлов трех образцов, которые были взяты из трех различных запусков секвенирования генов *BRCA1/2*, а также одного целого запуска (42 образца). Все наборы прочтений содержали следующее число пар прочтений, соответственно: 12757 (1 образец), 24376 (1 образец), 89919 (1 образец) и 16429892 (42 образца). После удаления последовательностей праймеров каждой из программ, проводили картирование оставшихся прочтений на референсную последовательность генома человека (версия сборки hg19), используя BWA, и оценивали полученное покрытие. Результаты сравнения представлены в **таблице 8**.

Таблица 8. Сравнительный анализ четырех программ, используя четыре набора последовательностей из четырех независимых запусков NGS. Cutadapt была использована в двух режимах: с якорными символами (^ и \$), которые заставляют cutadapt искать последовательности праймеров только на самых концах прочтений, и без них. BBDuk также запускалась в двух режимах: для поиска последовательностей праймеров с допуском только замен (hdist=3) и с допуском и замен, и инсерций/делеций (edist=2). Наилучшие варианты по каждому параметру сравнения выделены жирным шрифтом.

Число прочтений (число образцов)	Программа	cutadapt		BBDuk		AlienTrimmer	cutPrimers
	Режим	с якорными символами	без якорных символов	8 потоков, k=20, edist=2	8 потоков, k=20, hdist=3	парные прочтения	8 потоков, eгг=5
12757 (1)	Время обработки, секунды	32	113	9	56	3	7
24376 (1)		59	216	9	59	5	16
89919 (1)		211	824	13	61	18	70
16429892 (42)		40823	143653	1109	3969	3254	13739
12757	Число оставшихся прочтений ≥90 н., % от всех прочтений	8259 (64,7%)	3913 (30,7%)	9781 (76,7%)	9993 (78,3%)	4849 (38,0%)	11182 (87,6%)
24376 (1)		21854 (89,6%)	11897 (48,8%)	22733 (93,2%)	21876 (89,7%)	7829 (32,1%)	22707 (93,2%)
89919 (1)		64283 (71,5%)	31662 (35,2%)	67753 (75,3%)	65638 (73,0%)	27612 (30,7%)	66885 (74,4%)
16429892 (42)		14345458 (87,3%)	7800514 (47,5%)	15222691 (92,6%)	15217255 (92,5%)	2385222 (14,5%)	15231108 (92,7%)
12757 (1)	Число покрытых ампликонов (>30 на позицию)	105	65	111	137	6	146
24376 (1)		174	89	159	181	8	183
89919 (1)		181	116	185	188	21	189
16429892 (42)		181,0 (174-181)	118,5 (82-130)	188,0 (149-189)	183,0 (152-187)	41,5 (7-80)	189,0 (182-189)
12757 (1)	Медианное покрытие	37,0	9,0	35,0	60,0	0,0	68,0
24376 (1)		135,5	19,0	74,5	141,5	0,0	145,5
89919 (1)		431,0	97,0	247,0	456,5	3,0	473,0
16429892 (42)		1743,0	165	974,0	1387,0	3,0	1885,5

Программы cutadapt и AlienTrimmer оставляли значительно меньше прочтений по сравнению с cutPrimers и BBDuk. Основной проблемой AlienTrimmer и cutadapt (режим без якорных символов) было то, что они удаляли все последовательности, которые соответствуют какому-либо из 190 праймеров, которые вводит пользователь, в том числе находящиеся в середине прочтения (**рисунок 15**). Это приводит к потере большого числа прочтений пересекающихся ампликонов и снижению числа правильно обработанных прочтений. В то же время, cutadapt с якорными символами находит последовательности праймеров, удаленных только на несколько нуклеотидов от конца, что может встречаться довольно часто: в случае, если длина ампликона меньше длины прочтения. В этом случае прибор NGS прочитывает последовательность адаптера и его символы рассматриваются программой cutadapt как ошибки (инсерции) при поиске последовательности праймера в прочтении.

Вариант расположения последовательности праймера в прочтении	cutadapt ¹	AlienTrimmer, cutadapt ² , Trimmomatic	cutPrimers, BBDuk

Рисунок 15. Сравнение программ для удаления последовательностей праймеров из прочтений с допущением 5 ошибок (только 3 – для BBDuk и 30% - для cutadapt) и максимальной удаленности праймера от конца прочтения на 10 нуклеотидов (этот параметр доступен только для cutPrimers и BBDuk). Первая колонка содержит различные варианты расположения последовательности праймера на 3'-конце прочтения. Первая строка показывает пример, когда последовательность праймера находится на самом конце прочтения. Вторая строка – когда праймер расположен на самом конце, но содержит некоторое число ошибок (замен или инсерций/делеций). Третья строка – последовательность праймера удалена на три нуклеотида от 3'-конца прочтения. Четвертая – последовательность праймера удалена от 3'-конца на три нуклеотида и содержит три ошибки. Пятая – последовательность праймера удалена от конца прочтения на 6 нуклеотидов. Последняя строка – обрабатываемое прочтение содержит последовательность, гомологичную последовательности соседнего ампликона (малый серый прямоугольник). * - ошибка в последовательности праймера; 1 – режим cutadapt с якорными символами; 2- режим cutadapt без якорных символов).

В то же время, cutPrimers и BBDuk ищут последовательность праймера только в части прочтения и не удаляют последовательности, соответствующие последовательностям праймеров соседних ампликонов. Как следствие, такой подход дает больший выход правильно обрезанных прочтений. Особенно эта разница заметна при обработке прочтений целого запуска по числу ампликонов, покрытых минимум 30 прочтениями: для cutadapt не было ни одного образца, для которого были бы покрыты все 189 ампликонов (один ампликон не был покрыт совсем).

Несмотря на то, что BBDuk показал значительно более высокую скорость обработки, его использование имеет некоторые трудности. Во-первых, BBDuk имеет ограничение по числу допустимых ошибок (только 3 – для замен и только 2 – для замен + инсерций/делеций). Во-вторых, использование параметра «edist» приводит к вырезанию вместе с праймером лишней последовательности от 3'-конца (см. **таблицу 8** и **рисунок 16А**), что изменяет значения покрытия крайней позиции в ампликоне. В-третьих, BBDuk использует для поиска последовательностей праймеров так называемые k-меры, что приводит к вырезанию лишней последовательности из прочтения (**рисунок 16Б** и **16В**).

Таким образом, cutPrimers представляет собой новую программу, позволяющую быстро и эффективно удалять последовательности праймеров из прочтений. Несмотря на то, что программа проигрывает некоторым из аналогичных программ по скорости работы, она превосходит их по точности идентификации и удаления последовательностей праймеров.

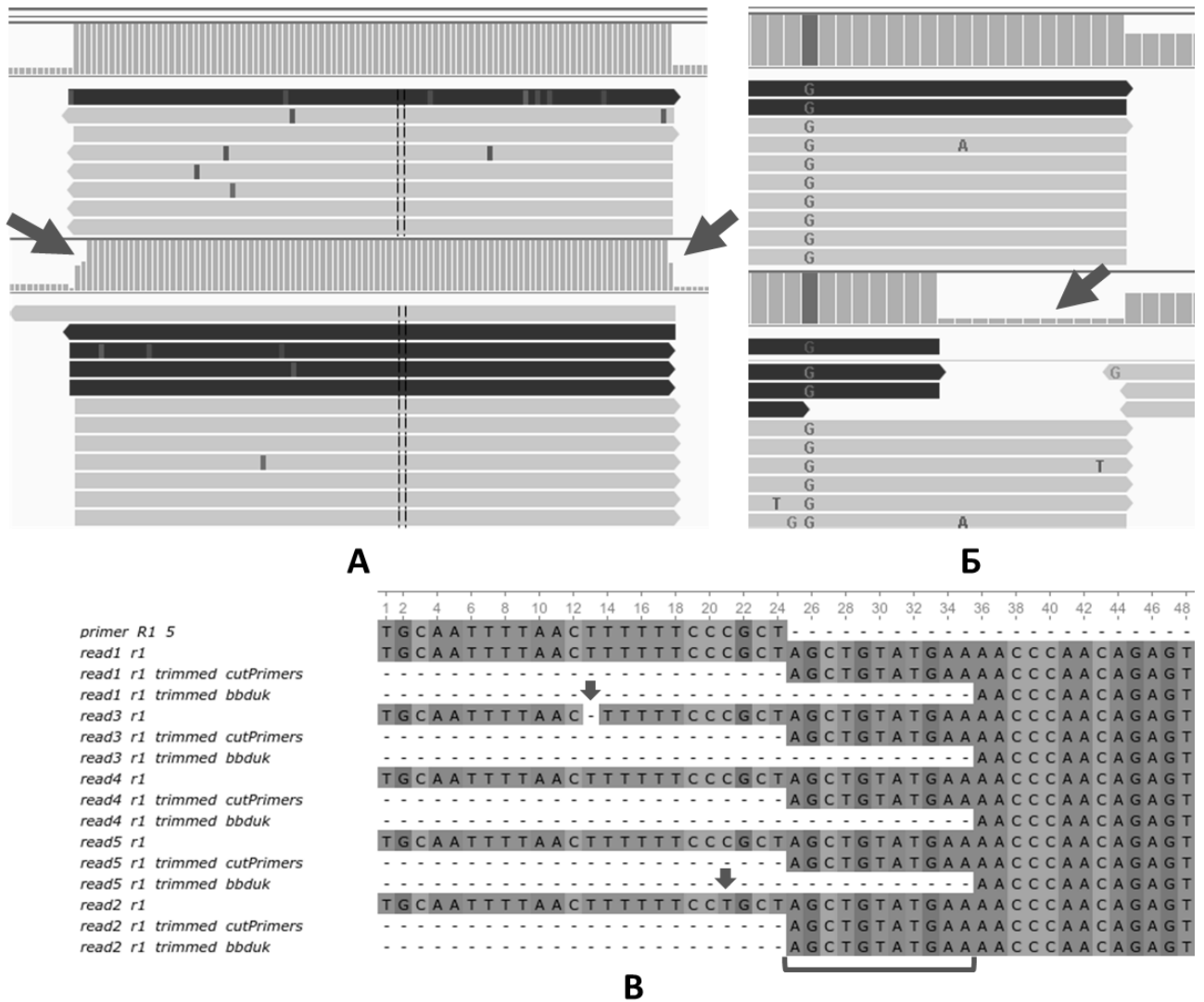


Рисунок 16. А. Сравнение прочтений, картированных на *BRCA1* и обрезанных с помощью *cutPrimers* (верхнее выравнивание) и *BBDuk* (режим «edist», нижнее выравнивание). Стрелки указывают на изменение покрытия крайних позиций в прочтениях. Данное снижение покрытия вызвано вырезанием *BBDuk* вместе с последовательностью праймера рядом стоящих символов. Данная проблема характерна и для других ампликонов. **Б.** Сравнение прочтений, картированных на *BRCA1* и обрезанных с помощью *cutPrimers* (верхнее выравнивание) и *BBDuk* (режим «hdist», нижнее выравнивание). Стрелка указывает на измененное покрытие последних 11 нуклеотидов в прочтении. Данное уменьшение характерно только для этого ампликона. **В.** Сравнение прочтений, обрезанных с помощью *cutPrimers* и *BBDuk* для ампликона из рисунка Б. Стрелки указывают на нуклеотиды в прочтении, которые отличаются от нуклеотидов в праймере. Скобка указывает на часть последовательности прочтения, которая неверно вырезана программой *BBDuk*.

Сравнение BRCA-analyzer с аналогичными пакетами

Сравнение BRCA-analyzer с коммерческим веб-сервисом было проведено на двух запусках NGS 105 образцов ДНК от 85 пациентов. Всего было выявлено 27 клинически значимых мутаций, все из которых были найдены с помощью BRCA-analyzer (чувствительность 100%). Веб-сервис Qiagen выявил только 22 из них (чувствительность 81%). Все пропущенные варианты располагались в гомополимерных районах: два образца с мутацией c.5266dupC (гомополимерная последовательность TC₃), два образца (один пациент) с мутацией c.1961delA (последовательность GA₈) и один образец с мутацией c.3847_3848delGT (CTGTAAGTGA₇TAATA₃). Веб-сервис Qiagen отфильтровал их именно по причине нахождения в гомополимере. В то же время, все 27 мутаций, выявленные BRCA-analyzer, были подтверждены секвенированием по Сэнгеру (**рисунок 17**). Клинически значимых мутаций, которые были бы выявлены Qiagen и не выявлены BRCA-analyzer, обнаружено не было.

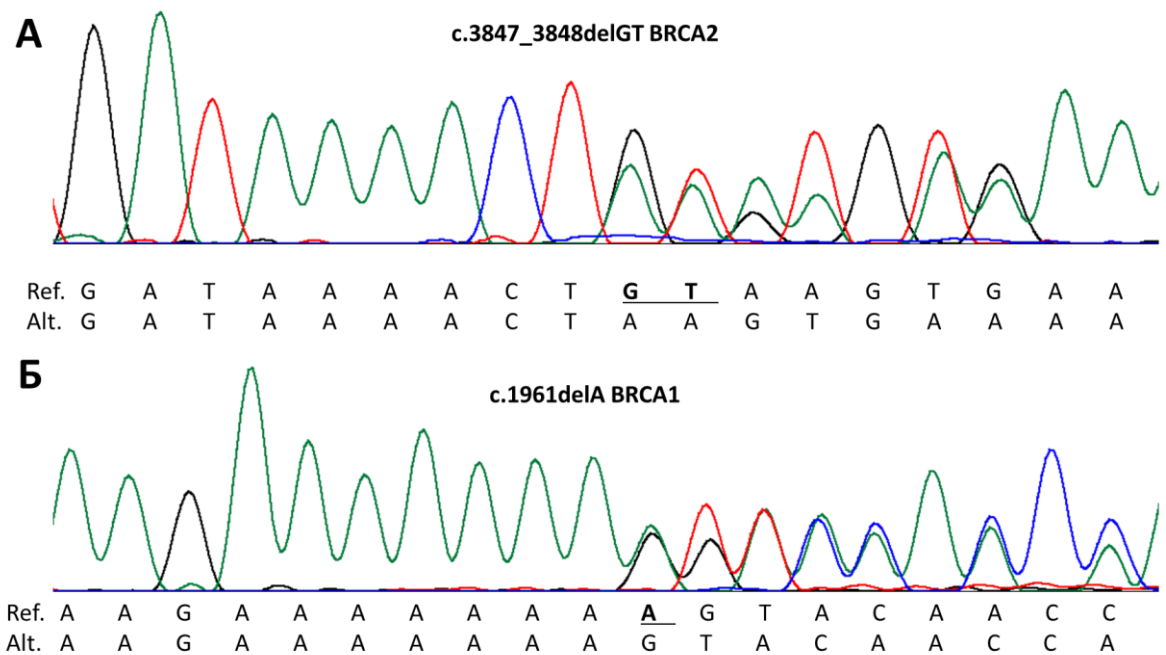


Рисунок 17. Хроматограммы некоторых из подтвержденных клинически значимых мутаций, выявленных программой BRCA-analyzer и пропущенных веб-сервисом Qiagen. Удаленные нуклеотиды выделены жирным шрифтом и подчеркнуты. **А.** c.3847_3848delGT была найдена в образце ДНК, выделенном из крови. **Б.** c.1961delA была выявлена в образцах ДНК, выделенных из крови и из ПГБ.

Помимо чувствительности, было проведено сравнение производительности программ. Для запуска BRCA-analyzer использовались одновременно 5–8 серверов (в зависимости от загруженности) на Суперкомпьютерном центре Новосибирского государственного университета (<http://nusc.nsu.ru>). Каждый сервер включал два

шестиядерных процессора Intel Xeon X5670 с тактовой частотой 2932 МГц и 24 Гб RAM. Время обработки, используя BRCA-analyzer, составило 55 минут. В то же время, протоколом Qiagen, не учитывая времени ожидания в очереди, составило 155 минут. Поскольку недоступной остается информация о производительности сервера Qiagen, нельзя наверняка сделать выводы о превосходстве BRCA-analyzer над веб-сервисом Qiagen. Однако факт того, что при данном сравнении результат оказался не хуже, говорит о высокой эффективности работы программы. Другой параметр, по которому возможно сравнение программ – удобство выдаваемых результатов: веб-сервис Qiagen выдает простой список неаннотированных вариаций, по которым нельзя даже выбрать мутации, имеющие клиническое значение. В свою очередь, BRCA-analyzer не только выдает аннотацию вариаций необходимую для их клинической интерпретации, но и разделяет их по клиническому значению. Кроме того, BRCA-analyzer – первый свободно доступный автоматический протокол для обработки данных NGS генов *BRCA1/2*.

В работе мы показали, что использование обычного неспециализированного протокола обработки данных NGS, как это сделано в сервисе Qiagen, может приводить к значительному снижению клинических показателей (до 81%). Для генов *BRCA1/2* хорошо известно, что большое число клинически значимых мутаций находятся именно в гомополимерных районах, и данный факт учтен в программе BRCA-analyzer. Подобная проблема показана и для гена *CHEK2*, для которого в геноме есть несколько псевдогенов с высокой гомологией [190].

Всего программой BRCA-analyzer было выявлено 16074 вариации для 1016 образцов 781 пациентов больных РЯ и 7 РМЖ, из которых 190 были классифицированы как клинически значимые для 146 пациентов. Все выявленные клинически значимые мутации были подтверждены секвенированием по Сэнгеру (специфичность метода 100%).

Таким образом, нами была разработана первая свободно доступная программа по обработке и анализу данных NGS генов *BRCA1* и *BRCA2* – BRCA-analyzer – и показана ее более эффективная работа по сравнению с существующим коммерческим сервисом.

Выявление коротких SNV, инсерций и делеций в 331 образце ДНК больных раком яичников

Ниже будут приведены мутации, выявленные для пациентов из следующих выборок (все пациенты имели позитивный ответ на терапию цисплатином):

- В (за исключением 41 образца для 41 пациента, образцы ДНК которых попали в первую версию приготовления библиотеки). Всего вариации были проанализированы для 90 образцов 66 пациентов: 40 образцов – только из

крови, 2 – только ПГБ, 24 – и из крови, и из ПГБ;

- Г (за исключением 13 пациентов, образцы ДНК которых попали в первую версию приготовления библиотеки). Всего вариации были проанализированы для 94 образцов 90 пациентов: 67 – только из крови; 19 – только из ПГБ; 4 – и из крови, и из ПГБ;
- Е (ни один из пациентов не попал в первую версию). Всего вариации были проанализированы для 147 образцов 145 пациентов: 111 – только из крови, 32 – только из ПГБ; 2 – и из крови, и из ПГБ.

Всего были найдены 53 клинически значимых мутации (40 уникальных, 29 – в гене *BRCA1* и 24 – в *BRCA2*) (таблица 9). Клинически значимые мутации были найдены для 53 пациентов из проанализированных 301 (17,6%). Для 34 (10,3%) пациентов найденные мутации были герминальными, для 4 (1,3%) пациентов – соматические, для оставшихся 15 это невозможно было точно определить по причине отсутствия материала нормальной ткани. Таким образом, суммарная частота встречаемости пациентов, несущих клинически значимые варианты составляет 30,3% (для 55 образцов, не включенных в анализ с помощью NGS, количественным ПЦР были выявлены «hotspot»-мутации), из которых 15,4% – герминальные клинически значимые варианты с повышенной частотой в популяции (5382insC, 4154del4, C61G и другие). Полученные частоты соответствуют частотам, полученным другими авторами в Германии (23% только герминальных мутаций) [113], Бразилии (17% герминальных среди всех больных) [117] и России (20,9% только «hotspot»-мутаций) [5], но выше, чем в работе китайских коллег (9,1% среди случаев хотя бы с одним фактором риска) [115].

Таблица 9. Клинически значимые мутации, найденные среди больных РЯ. **CDS** – обозначение мутации в кодирующей последовательности гена. **ВIC/ClinVar** – клиническая значимость для мутации, приведенная в ВIC или ClinVar. **Число обр/пац.** – число образцов и пациентов. **НСС** – нарушение сайта сплайсинга. **СРС** – сдвиг рамки считывания. Здесь обозначение экзонов приведено с учетом только кодирующих экзонов транскрипта NM_007294.3.

Ген	CDS	Тип мутации	Экз.	ВIC	ClinVar	Число обр/пац
<i>BRCA2</i>	c.51_52delAC	СРС	2	-	Да	1/1
	c.262_263delCT	СРС	3	-	Да	1/1
	c.658_659delGT	СРС	8	-	Да	1/1
	c.693delC	СРС	9	-	-	1/1
	c.1759dupA	СРС	10	-	-	2/2
	c.2808_2811delACAA	СРС	11	-	Да	2/2

	c.3847_3848delGT	CPC	11	-	Да	1/1
	c.3865_3868delAAAT	CPC	11	-	Да	2/1
	c.4211_4215delCAAAT	CPC	11	-	Да	1/1
	c.5073dupA	CPC	11	-	Да	1/1
	c.5286T>G	нонсенс	11	-	-	2/2
	c.6443_6447delCTATT	CPC	11	-	-	1/1
	c.6468_6469delTC	CPC	11	-	Да	1/1
	c.6475C>T	нонсенс	11	-	Да	1/1
	c.6494delT	CPC	11	-	Да	1/1
	c.6602_6603dupCT	CPC	11	-	-	1/1
	c.6998dupT	CPC	13	-	Да	2/2
	c.8331+1G>A	НСС	18	-	Да	1/1
	c.9253delA	CPC	24	-	-	1/1
BRCA1	c.5444G>A	нонсенс	22	Да	Да	3/2
	c.5347delA	CPC	21	-	-	2/1
	c.5278-1G>C	НСС	19	-	Да	1/1
	c.5266dupC	CPC	19	-	Да	3/3
	c.5152+1G>T	НСС	17	Да	Да	1/1
	c.5136G>A	нонсенс	17	Да	Да	1/1
	c.4689C>G	нонсенс	15	Да	Да	2/2
	c.4675G>A	миссенс	14	Да	Да	2/1
	c.4426A>T	нонсенс	13	-	-	1/1
	c.4161_4162delTC	CPC	11	-	Да	1/1
	c.3637delG	CPC	10	-	-	1/1
	c.3627dupA	CPC	10	-	Да	2/1
	c.3442delG	CPC	10	-	Да	1/1
	c.3247_3251delATGCT	CPC	10	-	-	1/1
	c.3143delG	CPC	10	-	-	1/1
	c.3083delG	CPC	10	-	-	1/1
	c.2285_2286delGA	CPC	10	-	-	1/1
	c.1961delA	CPC	10	-	Да	1/1
	c.1961dupA	CPC	10	-	Да	2/2
	c.181T>G	миссенс	4	Да	Да	1/1
c.126_127delAT	CPC	3	-	-	1/1	

Также у шести пациентов нами была найдена мутация c.10094_10095insGAATTATATC, расположенная в 27 (последнем) экзоне гена *BRCA2*. Данная мутация располагается в кодирующей последовательности за другой известной мутацией c.9976A>T (p.Lys3326*), которая на данный момент считается скорее нейтральной, чем клинически значимой (базы данных BIC, ClinVar, [191, 192]). В связи с этим, мы классифицировали мутацию c.10094_10095insGAATTATATC как вариацию с неизвестным клиническим значением.

Программа для выявления CNV в генах BRCA1/2

Для выявления CNV нами была разработана собственная программа, алгоритм которой на текущий момент не опубликован. Для определения чувствительности и

специфичности метод был проверен на 197 образцах ДНК, выделенных из крови (из разных выборок, большая часть – Ж), для которых CNV также были выявлены с помощью MLPA. Из 197 пациентов CNV были обнаружены в 12 образцах. Используемый нами алгоритм выявил все 12 CNV, однако, вместе с тем, им были ошибочно определены еще 32 CNV. Таким образом, чувствительность метода составила 100,0%, специфичность – 82,7%. Специфичность метода может быть повышена при самостоятельном просмотре формируемых программой изображений с нормализованными значениями покрытия (рисунок 18). Полученные значения клинических показателей метода близки к требованиям, которые часто выдвигаются к различным тест-системам (порядка 90%), что при дальнейшей доработке метода может привести к возможности его использования при клиническом тестировании пациентов.

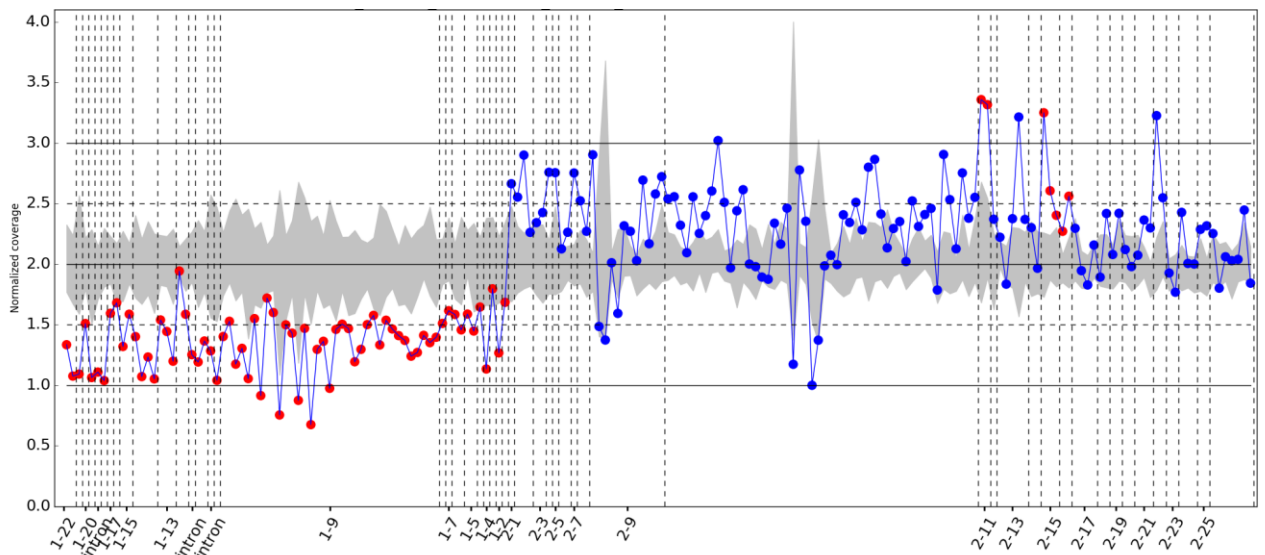


Рисунок 18. Пример изображения, выдаваемого программой по выявлению CNV. Каждая точка – нормализованное значение покрытия ампликона. Красные точки – ампликоны, составляющие экзон, которые по предсказанию программы делетированы или дуплицированы. Синие – ампликоны с нормальным покрытием. По горизонтальной оси – номера ампликонов и номера экзонов: сначала гена *BRCA1* (с 22 по 1 в обратном порядке), а затем гена *BRCA2* (с 1 по 26). Вертикальные пунктирные прямые обозначают границы экзонов. Серая зона вокруг прямой $y=2,0$ обозначает разброс значений покрытия ампликонов от 20 до 80% всех пациентов. У пациента наблюдается полная потеря одной из копий гена *BRCA1*.

Исследование CNV и потери гетерозиготности в генах *BRCA1/2*

Относительно феномена «потеря гетерозиготности» до сих пор остаются

нерешенными два вопроса. Во-первых, что является механизмом возникновения потери гетерозиготности: геновая конверсия, приводящая к замене одной из родительских копий другой, что ведет к тому, что у пациента становится две одинаковых копии; или CNV, приводящие к делеции одной из копий, в результате чего соотношение числа родительских копий изменяется до 1:0 [6]. Во-вторых, может ли потеря гетерозиготности быть показателем клинического значения вариации. В некоторых работах авторы утверждают, что клиническое значение исследуемой вариации может быть доказано, показав только, что для данного локуса наблюдается потеря гетерозиготности [123, 193]. В то же время, ряд других исследователей считают, что такой критерий не может быть использован [194]. Кроме того, такого критерия нет в рекомендациях по классификации вариаций в пять категорий: клинически значимые, вероятно клинически значимые, вариации с неопределенным клиническим значением, вероятно нейтральные и нейтральные [183]. С развитием технологии NGS и методов таргетного секвенирования генов *BRCA1/2*, появилась возможность исследования многих видов изменений в геноме одновременно. Целью данной части работы стали оценка применимости наличия потери гетерозиготности для классификации вариантов по их клиническому значению, а также исследование механизма формирования потери гетерозиготности в опухоли больных раком яичников.

Для этого нами были выбраны 30 больных РЯ, для которых с помощью NGS уже были проанализированы образцы ДНК из крови и из ПГБ. Всего для 30 пациентов было найдено 518 герминальных и соматических точечных и коротких вариантов, 61 из которых были исключены из дальнейшего анализа, так как были представлены менее, чем 50 прочтениями. Девять вариантов были классифицированы как клинически значимые, ассоциированные с развитием РМЖ и РЯ (таблица 10). Также было найдено восемь соматических CNV: пять – дупликаций (клиническое значение неизвестно, поскольку остается неясным, куда была дублирована последовательность; три из них – у пациентов с мутациями) и три – делеции, затрагивающие большую часть гена (вероятно имеют клиническое значение). Для остальных 16 пациентов не было выявлено ни одного варианта, который мог быть классифицирован как клинически значимый или возможно клинически значимый (для всех частота в 1000Genomes была выше 0,1% либо они встречались у всех пациентов и находились в интроне, либо в базе данных VIC были записаны как нейтральные). Таким образом, клинически значимые варианты были найдены у 40% образцов (у 20% – герминальные и у 20% – соматические); 75% найденных клинически значимых вариантов – в гене *BRCA1* (25% – CNV), 25% – *BRCA2*.

Таблица 10. Клинически значимые варианты и CNV, выявленные в генах *BRCA1/2* у 30 больных раком яичников. CPC – сдвиг рамки считывания. Все приведенные варианты отсутствовали в базе данных 1000Genomes. Герминальные (герм.) были найдены в образцах ДНК из крови и гистологического блока; соматические (сомат.) – только из гистологического блока.

ИД пац.	Ген	Тип варианта	CDS	ВІС (клин. знач.)	№ экзона или интрона	Сомат или герм.
2	<i>BRCA1</i>	нонсенс	c.5444G>A	Да	21	герм.
5	<i>BRCA1</i>	CPC	c.126_127delAT	Н/Д	2	сомат.
6	<i>BRCA1</i>	CPC	c.3083delG	Н/Д	9	сомат.
7	<i>BRCA1</i>	миссенс	c.4675G>A	Да	13	герм.
8	<i>BRCA2</i>	CPC	c.3865_3868delAAAT	Н/Д	10	герм.
11	<i>BRCA1</i>	CNV	дупликация экзонов 19–22	Н/Д	19–22	сомат.
12	<i>BRCA1</i>	CNV	дупликация экзонов 15–21	Н/Д	15–21	сомат.
14	<i>BRCA1</i>	CNV	делеция всего гена <i>BRCA1</i>	Н/Д	1–22	сомат.
15	<i>BRCA1</i>	CNV	делеция экзонов 7–14	Н/Д	7–14	сомат.
16	<i>BRCA1</i>	CPC	c.968delG	Н/Д	9	герм.
19	<i>BRCA1</i>	CNV	делеция экзонов 9–12	Н/Д	9–12	сомат.
22	<i>BRCA2</i>	нонсенс	c.6475C>T	Н/Д	10	сомат.
	<i>BRCA1</i>	CNV	дупликация гена <i>BRCA1</i>	Н/Д	1–22	сомат.
27	<i>BRCA1</i>	CPC	c.3627dupA	Н/Д	9	герм.
	<i>BRCA1</i>	CNV	дупликация экзонов 1–3	Н/Д	1–3	сомат.
29	<i>BRCA2</i>	CPC	c.6405_6408delCTTA	Н/Д	10	герм.
	<i>BRCA1</i>	CNV	дупликация экзонов 17–22	Н/Д	17–22	сомат.

Потеря гетерозиготности и мутации, ассоциированные с развитием РМЖ/РЯ

Потеря гетерозиготности наблюдалась у 21 (70%) пациента из 30, из которых пять несли герминальный клинически значимый вариант. У оставшегося одного пациента с клинически значимым вариантом частота альтернативного аллеля для некоторых вариантов также была повышена (максимально – на 19%). Чтобы оценить применимость потери гетерозиготности в качестве критерия классификации вариантов по их клиническому значению, мы сравнили изменение процента прочтений, содержащих альтернативный аллель (т.е. аллель, отличный от референсной последовательности генома человека) для клинически значимых и нейтральных вариантов. Из группы нейтральных вариантов были исключены те, которые расположены в том же гене, что и клинически значимый вариант, у соответствующих пациентов (**рисунок 19**). Это было сделано по той причине, что такие нейтральные варианты часто изменяют свою частоту вместе с клинически значимым вариантом. Для того, чтобы исключить попадание потенциально клинически значимых вариантов в группы нейтральных, из нейтральных вариантов были

исключены варианты с неизвестным клиническим значением (VUS, несинонимичные замены, которые отсутствуют в базе данных VIC или обозначены как «unknown» и имеют частоту в 1000Genomes менее 0,1%), а также те варианты, которые находятся в то же гене, что и VUS у несущих их пациентов.

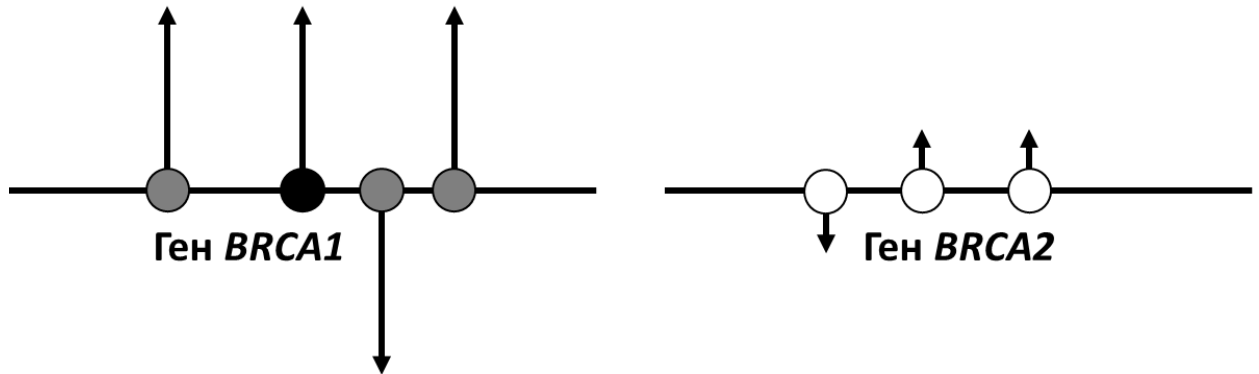


Рисунок 19. Три категории вариантов. Черным кругом обозначен клинически значимый вариант (категория 1), серыми кругами – нейтральные варианты, которые расположены в том же гене, что и клинически значимый вариант (категория 2 – были исключены из анализа). Белыми кругами обозначены варианты, которые расположены в другом гене и, следовательно, не могут быть сцеплены с клинически значимым вариантом (категория 3). Часто наблюдается одновременное изменение процента прочтений, несущих альтернативный аллель, для клинически значимого варианта и вариантов, расположенных в том же гене. Изменение процента прочтений для разных вариантов может быть разнонаправленным, в зависимости от того, какого аллеля (референсного или альтернативного) стало больше.

Сравнение было проведено по абсолютному изменению процента прочтений для всех клинически значимых вариантов и нейтральных вариантов, для которых процент прочтений увеличился (**рисунок 20**), а также по числу вариантов, для которых произошло статистически значимое увеличение процента прочтений.

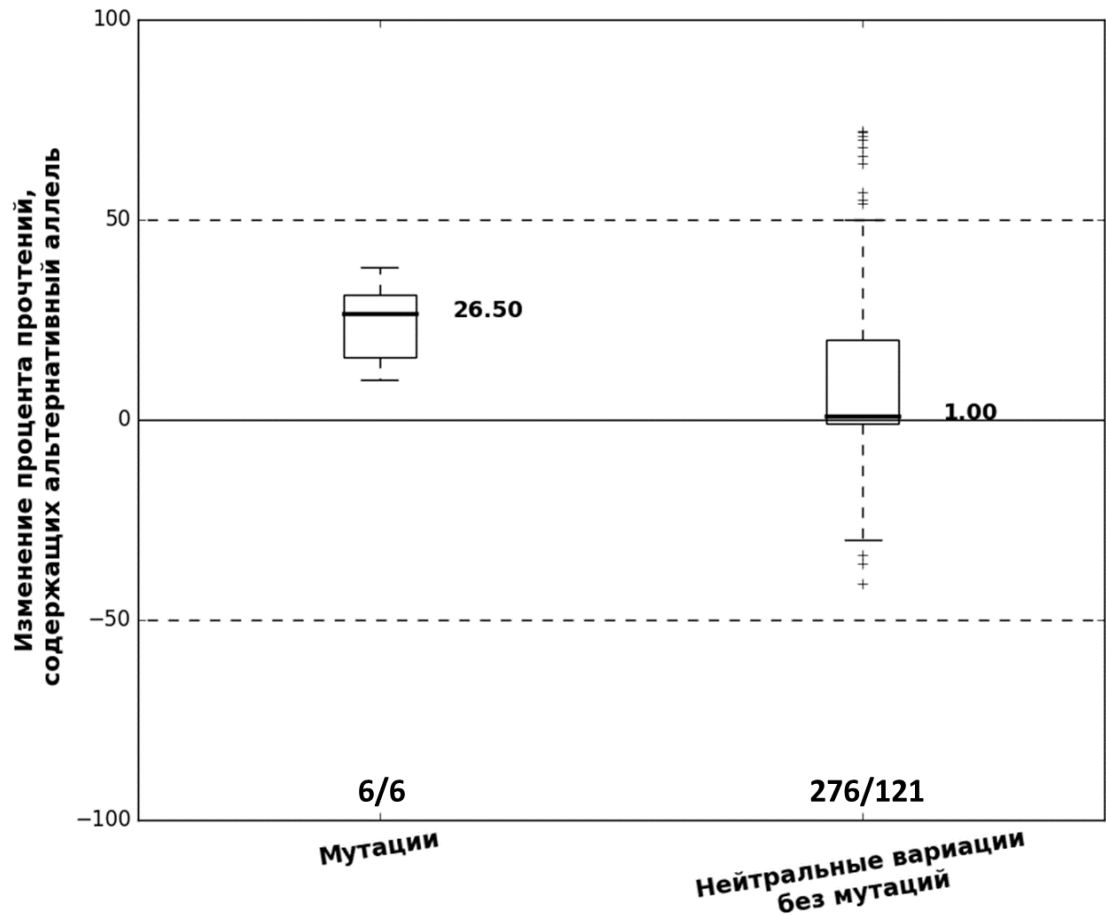


Рисунок 20. Изменение процента прочтений, содержащих аллель клинически значимого варианта (на рисунке обозначены как мутации) или нейтральных вариантов. «Нейтральные вариации без мутаций» – варианты, которые в базе данных VIC обозначены как нейтральные либо имеют частоту в 1000Genomes более 0,1% и которые находятся либо в другом гене относительно герминального клинически значимого варианта или варианта с неизвестным клиническим значением, либо у этого пациента отсутствуют герминальные клинически значимые варианты и варианты с неизвестным клиническим значением. Над горизонтальной осью подписано число таких вариантов и число вариантов, для которых процент прочтений увеличился. Жирная полоса в прямоугольнике – медиана, справа подписано соответствующее значение; прямоугольник отражает межквартильный диапазон (между первым и третьим квартилями, т.е. ~25% и 75%); усы отражают ~99,3% всех значений; знаки плюса отражают оставшиеся ~0,7% значений.

Для герминальных клинически значимых вариантов процент прочтений, несущих альтернативный аллель, всегда увеличивался (медианное изменение составило 26,5%). Для оставшихся 276 нейтральных вариантов увеличение процента прочтений наблюдалось для 142 вариантов (51,4%) с медианным изменением 20,0%. Статистически значимое увеличение процента прочтений (p-value при одностороннем точном тесте Фишера менее

0,05 и изменение процента прочтений не менее 20%) наблюдалось для 121 варианта (43,8%). Разница в увеличении процента прочтений, несущих альтернативный аллель (отличный от референсного генома человека), для клинически значимых вариантов и нейтральных вариантов оказалась статистически незначимой ($p\text{-value}=0,156$ в одностороннем тесте Манна-Уитни). Таким образом, для клинически значимых вариантов процент прочтений, несущих мутантный аллель, увеличивается статистически незначимо больше (медианы 26,5% против 20,0%), чем для нейтральных вариантов, что подтверждает диаграмма размаха, приведенная на **рисунке 20**. Для многих нейтральных вариантов наблюдается большее увеличение процента прочтений, чем для клинически значимых вариантов.

Далее мы сравнили, насколько чаще для клинически значимых вариантов увеличивается процент прочтений, несущих альтернативный аллель, по сравнению с нейтральными вариантами (**таблица 11**). При сравнении клинически значимых вариантов с нейтральными вариантами статистически значимых отличий найдено не было ($p\text{-value}=0,242$ при использовании одностороннего точного теста Фишера). Следовательно, при использовании потери гетерозиготности в качестве критерия для классификации выявленных вариантов по клиническому значению, исследуемый вариант может почти равновероятно оказаться как клинически значимым, так и нейтральным.

Таблица 11. Сравнение доли вариантов, для которых произошло статистически значимое увеличение процента прочтений ($p\text{-value}$ при одностороннем точном тесте Фишера менее 0,05 и изменение процента прочтений не менее 20%). Здесь мутация – это клинически значимый вариант.

	Клинически значимые варианты (мутации)	Нейтральные варианты без мутаций и VUS
Статистически значимое увеличение процента прочтений	4	121
Процент прочтений не изменился или уменьшился	2	155

Механизм возникновения регионов с LOH: CNV или генная конверсия?

Для того, чтобы понять, какой механизм превалирует при формировании на хромосомах районов с потерей гетерозиготности, необходимо было совместить карты наблюдаемых крупных перестроек и районов с потерей гетерозиготности (**рисунки 21–22**). При их сравнении было обнаружено, что регионы с потерей гетерозиготности перекрываются CNV только в 3 (14%) случаях из 21 (пример на **рисунке 21**). В

большинстве случаев (5 из 8) выявленные соматические CNV либо не перекрывали район потери гетерозиготности (3 пациента), либо в данном районе не были выявлены никакие вариации (2 пациента) (пример на **рисунке 22**). Также часто наблюдались варианты, когда в образце была найдена потеря гетерозиготности, но не было выявлено CNV (всего 13 пациентов – 62%). Для 9 пациентов потеря гетерозиготности не наблюдалась.

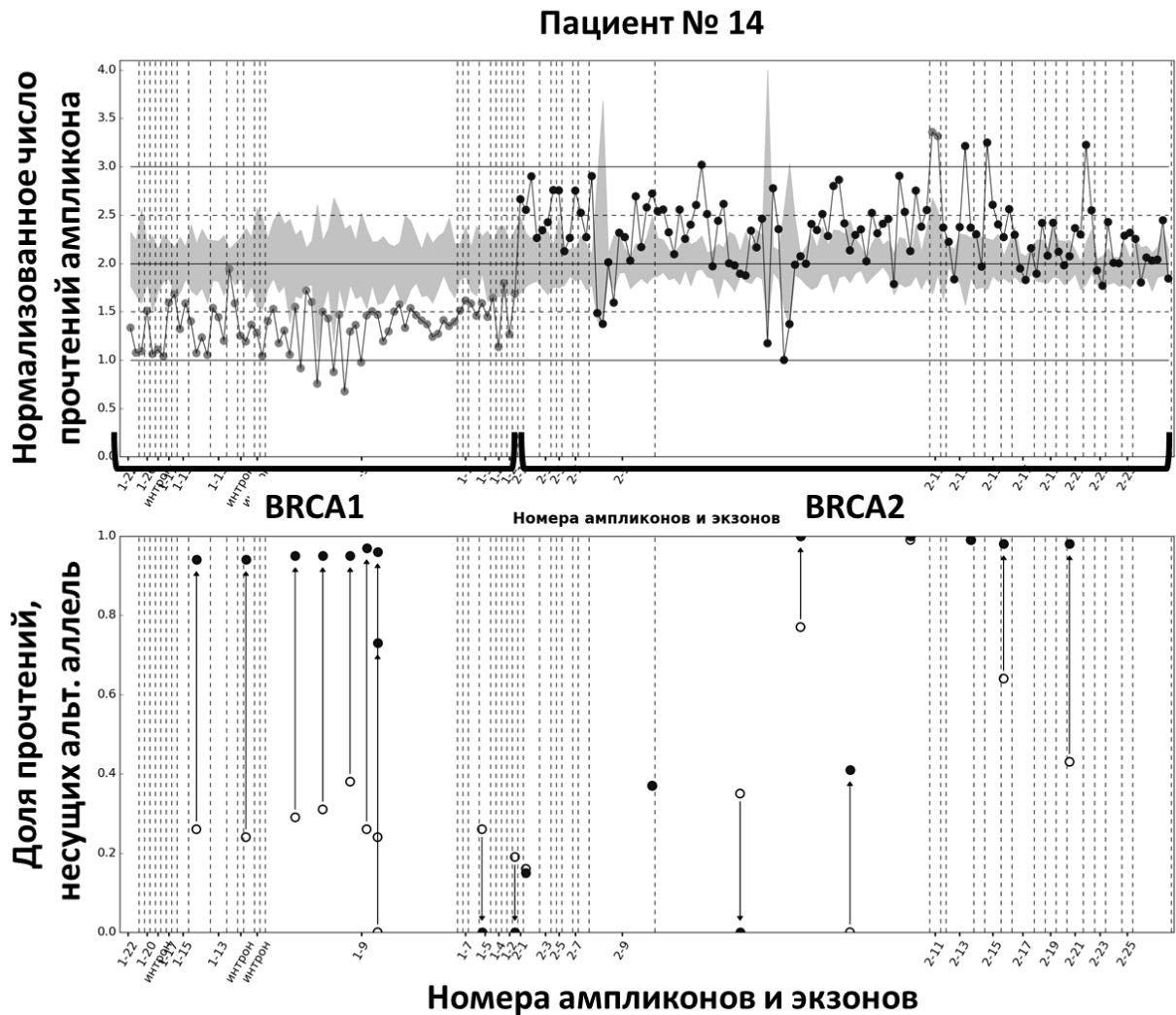


Рисунок 21. Объединенная карта ампликонов и экзонов генов *BRCA1* и *BRCA2* для пациента №14, на которую нанесены нормализованное число прочтений ампликонов (сверху) и изменения процента прочтений, содержащих альтернативный аллель (снизу). Сверху каждая точка отражает нормализованное число прочтений. Серым выделены точки, которые программа определила, как потенциальная делеция или дупликация. Снизу не закрашенные точки обозначают процент прочтений для варианта в образце ДНК из крови, закрашенные – в гистологическом блоке. Для данного пациента видна делеция всего гена *BRCA1* (левая часть карты) и соответствующие ей районы потери гетерозиготности. В гене *BRCA2* (правая часть карты) также наблюдаются ЛОН.

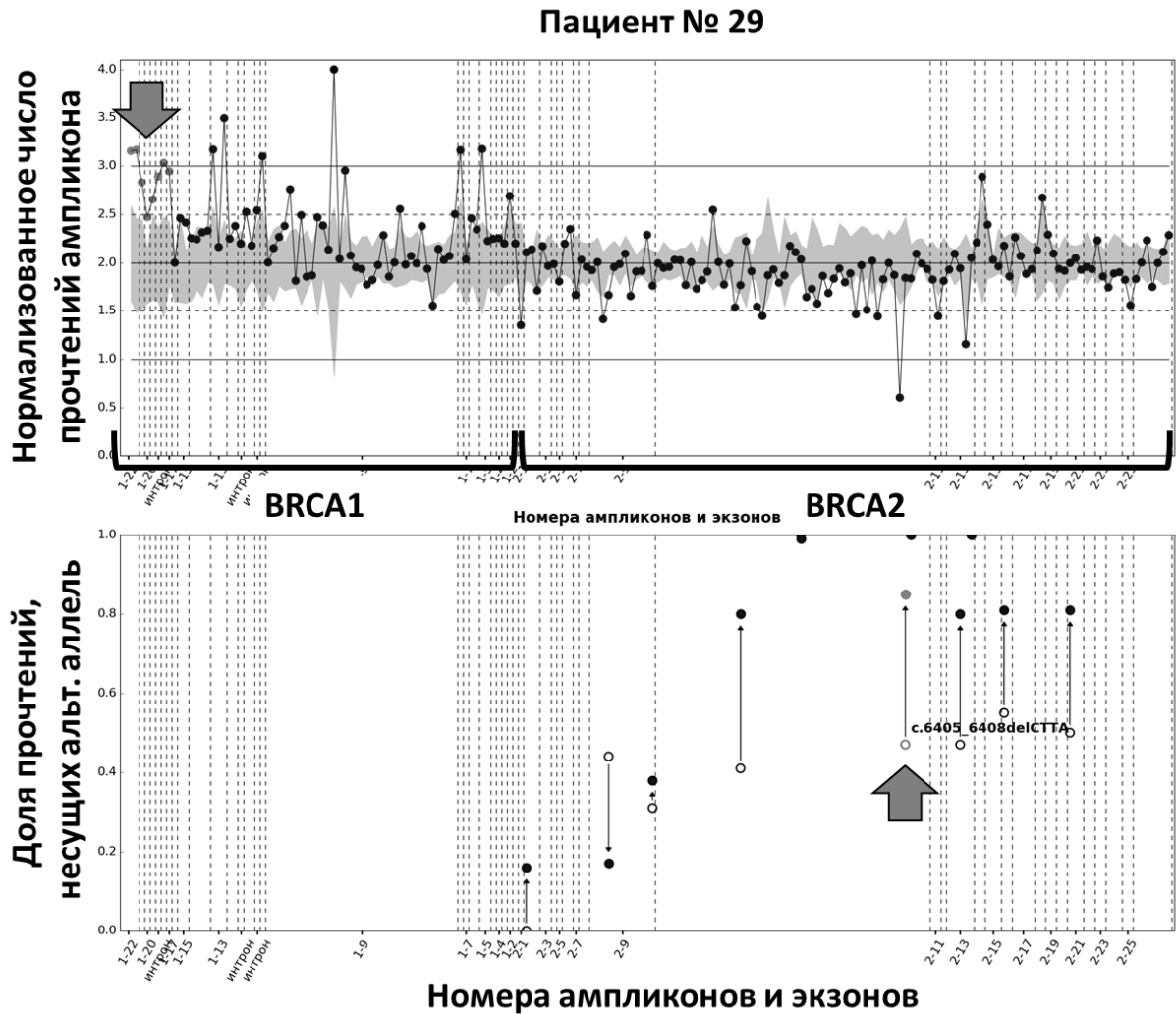


Рисунок 22. То же, что и на рисунке 21, но для пациента №29. Показана дупликация с интрона 19 по экзон 22 (отмечена стрелкой сверху) в гене BRCA1, однако эта CNV не коррелирует с потерей гетерозиготности. Кроме того, обнаружен клинически значимый герминальный вариант c.6405_6408delCTTA (отмечен стрелкой снизу) – обозначен точкой серого цвета и подписан сбоку. Вместе с клинически значимым вариантом наблюдается потеря гетерозиготности практически во всем гене.

В литературе описано множество случаев потери гетерозиготности в опухолевых клетках, часто достигающие до 100% всех образцов, несущих клинически значимый вариант [146, 195]. При этом данный феномен связывают с инактивацией второй копии гена, несущей аллель дикого типа и часто выявляют с помощью количественной ПЦР на клинически значимый вариант [123, 193]. С другой стороны, потеря гетерозиготности имеет широкую представленность на уровне всего генома, приводя к появлению районов со сниженным количеством гетерозиготных локусов размером до нескольких миллионов пар нуклеотидов [196]. В данной работе мы провели исследование феномена потери гетерозиготности во всей кодирующей последовательности и прилегающих участках

интронов генов *BRCA1/2* на вариациях, выявленных с помощью таргетного NGS. Как и описано в литературе, нами было показано наличие ЛОН у большинства пациентов, несущих герминальный клинически значимый вариант в каком-либо из двух генов (83%), причем у оставшегося одного пациента процент прочтений с альтернативным аллелем некоторых вариантов также был увеличен. Как и ожидалось, вместе с клинически значимым вариантом гетерозиготность теряли и локусы, находящиеся рядом. Однако неожиданным оказался результат, что потеря гетерозиготности наблюдалась и в большинстве образцов без клинически значимого герминального варианта (67%).

Одной из гипотез, которую мы проверяли, было предположение о том, что потеря гетерозиготности может свидетельствовать о клинической значимости выявленного варианта. Мы показали, что для клинически значимых вариантов увеличение процента прочтений, несущих альтернативный аллель, статистически незначимо выше, чем для нейтральных вариантов ($p\text{-value}=0,156$). Кроме того, при сравнении частоты случаев, когда процент прочтений с альтернативным аллелем увеличивается, не было обнаружено статистически значимого различия между вариантами, ассоциированными с РЯ, и нейтральными вариантами. Более того, для многих нейтральных вариантов увеличение процента прочтений оказалось больше, чем для клинически значимых вариантов (см. **рисунок 20**). Все это говорит о том, что потеря гетерозиготности вряд ли может использоваться в качестве критерия для определения клинического значения VUS.

Важным для понимания механизмов развития опухоли является определение механизма появления районов с потерей гетерозиготности. Данный вопрос до сих пор оставался малоизученным из-за отсутствия методов, которые бы позволяли его адекватно решать. На сегодняшний день технология NGS дает возможность определять множество видов событий в геноме человека за один запуск: от SNV и коротких инсерций/делаций до крупных перестроек и потери гетерозиготности [197]. В нашей работе мы показали, что соматические CNV выявляются в 27% (8) случаев, и только 38% (3) из них объясняют наблюдаемую потерю гетерозиготности. Для остальных 18 случаев ЛОН мы предположили механизм их появления через генную конверсию, что совпадает с результатами, полученными ранее для нескольких других видов рака другими исследователями [198]. Неясной остается природа вариабельности значений нормализованного покрытия в некоторых образцах (см. **рисунки 21 и 22**): является ли это следствием изменения числа копий в геноме или только неравномерности амплификации. Необходимо дальнейшее исследование феномена «потеря гетерозиготности», так как это может иметь значение и для прогноза эффективности лечения пациентов препаратами платины [199].

Таким образом, из результатов данной части работы можно предположить, что 1) потеря гетерозиготности вряд ли может быть использована в качестве критерия для классификации вариантов по клиническому значению; 2) геновая конверсия является основным механизмом наблюдаемой потери гетерозиготности.

Выводы

1. Разработан метод выявления клинически значимых вариантов в генах *BRCA1* и *BRCA2* с использованием NGS, позволяющий сократить число процедур приготовления библиотеки по сравнению с коммерчески доступным набором реагентов Qiagen в три раза. Разработанный метод приготовления библиотеки позволяет получить минимально допустимое покрытие всех экзонов генов *BRCA1/2* (более 20 прочтений на позицию при медианном более 300) при исследовании 96 пациентов и секвенировании библиотеки на MiSeq Illumina.
2. Для разработанной программы cutPrimers показаны более точная идентификация и эффективное удаление последовательностей праймеров из прочтений по сравнению с существующими программами, что приводит к увеличению числа прочтений целевых фрагментов минимум на 8% и до 100% по сравнению с некоторыми программами.
3. Для разработанной программы BRCA-analyzer на 1016 образцах ДНК, выделенных из крови и из парафинированных гистологических блоков, показана специфичность 100% для клинически значимых вариантов. Показана более высокая чувствительность (100% против 81%) по сравнению с коммерчески доступным веб-сервисом.
4. Для разработанной программы по выявлению CNV в генах *BRCA1/2* на 197 образцах ДНК, выделенных из крови, показана чувствительность 100% и специфичность 82,7%, что позволяет использовать данный метод для тестирования пациентов на CNV.
5. Суммарная частота встречаемости пациентов с клинически значимыми герминальными вариантами в генах *BRCA1* и *BRCA2* среди больных РМЖ, имеющих указания на наследственный характер заболевания, составила 14,4%, из которых 11,3% – пациенты с клинически значимыми вариантами с повышенной частотой в популяции (5382insC, 4154del4, C61G и другие).
6. Суммарная частота встречаемости пациентов с клинически значимыми герминальными и соматическими вариантами в генах *BRCA1* и *BRCA2* среди больных РЯ с позитивным ответом на лечение цисплатином составила 30,3%, из которых 15,4% – пациенты с клинически значимыми герминальными вариантами с повышенной частотой в популяции (5382insC, 4154del4, C61G и другие).
7. Показано, что потеря гетерозиготности наблюдается в 70% всех случаев рака яичников, среди несущих герминальный клинически значимый вариант – в 83%

случаев. Наличие потери гетерозиготности не может являться достаточным критерием для классификации вариантов по клиническому значению.

8. Из 21 случая LOH, только для трех (14%) в качестве механизма формирования можно предположить CNV. Для остальных случаев, основным механизмом возникновения LOH, по-видимому, является генная конверсия, что согласуется с литературными данными.

Список литературы

1. Каприн А.Д., Старинский В.В., Петрова Г.В. Злокачественные новообразования в России в 2015 году (заболеваемость и смертность). // МНИОИ им. П.А. Герцена - филиал ФГБУ «НМИРЦ» Минздрава России. – 2017.
2. Kuchenbaecker K.B., Hopper J.L., Barnes D.R., et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for *BRCA1* and *BRCA2* Mutation Carriers // *JAMA*. – 2017. – V. 317. – No. 23. – P. 2402.
3. Fachal L., Blanco A., Santamariña M., Carracedo A., Vega A. Large Genomic Rearrangements of *BRCA1* and *BRCA2* among Patients Referred for Genetic Analysis in Galicia (NW Spain): Delimitation and Mechanism of Three Novel *BRCA1* Rearrangements // Toland AE, editor. *PLoS One*. – 2014. – V. 9. – No. 3. – P. 93306.
4. Preobrazhenskaya E. V., Bizin I. V., Kuligina E.S., et al. Detection of *BRCA1* gross rearrangements by droplet digital PCR // *Breast Cancer Res. Treat.* – 2017.
5. Батенева Е.И., Филиппова М.Г., Тюляндина А.С., и др. Результаты генетического скрининга герминальных мутаций в генах *BRCA1* и *BRCA2* у больных раком молочной железы и больных раком яичников в российской популяции // *Онкогинекология*. – 2015. – №3. – С. 34–39.
6. Ryland G.L., Doyle M.A., Goode D., et al. Loss of heterozygosity: what is it good for? // *BMC Med Genomics*. – 2015. – V. 8. – No. 1. – P. 45.
7. Лазарев А.Ф., Задонцева Н.С., Гофман А.А. Наследственный рак молочной железы // *Российский онкологический журнал*. – 2014. – Т. 19 – №2. – P. 40–46.
8. Стенина М.Б., Жукова Л.Г., Королева И.А., и др. Практические рекомендации по лекарственному лечению инвазивного рака молочной железы // *Злокачественные опухоли*. – 2016. – Т. 4. – №2. – С. 97–122.
9. Robert M., Frenel J.-S., Gourmelon C., Patsouris A., Augereau P., Campone M. Olaparib for the treatment of breast cancer // *Expert Opin. Investig. Drugs*. – 2017. – V. 26. – No. 6. – P. 751–759.
10. Солопова А.Г., Бицадзе В.О., Солопова А.Е., Макацария А.Д., Розанов И.А. Рак яичника: современные подходы к классификации, диагностике, стадированию и дифференцированной тактике ведения больных // *Журнал акушерства и женских болезней*. – 2017. – Т. 66. – №2. – С. 55–66.
11. Тюляндин С.А., Деньгина Н.В., Коломиец Л.А., и др. Практические рекомендации по лекарственному лечению рака яичников / первичного рака брюшины / рака маточных труб // *Злокачественные опухоли*. – 2016. – Т. 4. – №2. – С. 123–134.

12. Eng K.H., Hanlon B.M., Bradley W.H., Szender J.B. Prognostic factors modifying the treatment-free interval in recurrent ovarian cancer. // *Gynecol. Oncol.* – 2015. – V. 139. – No. 2. – P. 228–235.
13. McLachlan J., George A., Banerjee S. The current status of PARP inhibitors in ovarian cancer // *Tumori J.* – 2016. – V. 102. – No. 5. – P. 433–440.
14. Anon. Rucaparib Approved for Ovarian Cancer // *Cancer Discov.* – 2017. – V. 7. – No. 2. – P. 120-121.
15. Smithers D.W. Family histories of 459 patients with cancer of the breast. // *Br. J. Cancer.* – 1948. – V. 2. – No. 2. – P. 163–167.
16. Gusella J.F., Wexler N.S., Conneally P.M., et al. A polymorphic DNA marker genetically linked to Huntington's disease // *Nature.* – 1983. – V. 306. – No. 5940. – P. 234–238.
17. Hall J.M., Lee M.K., Newman B., et al. Linkage of early-onset familial breast cancer to chromosome 17q21. // *Science* – 1990. – V. 250. – No. 4988. – P. 1684–1689.
18. Schildkraut J.M., Risch N., Thompson W.D. Evaluating genetic association among ovarian, breast, and endometrial cancer: evidence for a breast/ovarian cancer relationship. // *Am. J. Hum. Genet.* – 1989. – V. 45. – No. 4. – P. 521–529.
19. Easton D.F., Bishop D.T., Ford D., Crockford G.P. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. // *Am. J. Hum. Genet.* – 1993. – V. 52. – No. 4. – P. 678–701.
20. Smith S.A., DiCioccio R.A., Struewing J.P., et al. Localisation of the breast-ovarian cancer susceptibility gene (BRCA1) on 17q12-21 to an interval of ≤ 1 cM. // *Genes. Chromosomes Cancer.* – 1994. – V. 10. – No. 1. – P. 71–76.
21. Albertsen H.M., Smith S.A., Mazoyer S., et al. A physical map and candidate genes in the BRCA1 region on chromosome 17q12-21. // *Nat. Genet.* – 1994. – V. 7. – No. 4. – P. 472–479.
22. Miki Y., Swensen J., Shattuck-Eidens D., et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. // *Science.* – 1994. – V. 266. – No. 5182. – P. 66–71.
23. Wooster R., Neuhausen S.L., Mangion J., et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. // *Science.* – 1994. – V. 265. – No. 5181. – P. 2088–2090.
24. Wooster R., Bignell G., Lancaster J., et al. Identification of the breast cancer susceptibility gene BRCA2. // *Nature.* – V. 378. – No. 6559. – P. 789–792.
25. Nagy R., Sweet K., Eng C. Highly penetrant hereditary cancer syndromes // *Oncogene.* – 2004. – V. 23. – No. 38. – P. 6445–6470.
26. Соколенко А.П., Иевлева А.Г., Митюшкин Н.В., и др. Синдром наследственного рака молочной железы и яичников в Российской Федерации // *Acta Naturae.* – 2010. – Т. 2 – №4

(7). – C. 35-39.

27. Welch P.L., King M.C. BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. // *Hum. Mol. Genet.* – 2001. – V. 10. – No. 7. – P. 705–713.

28. Aziz S., Kuperstein G., Rosen B., et al. A Genetic Epidemiological Study of Carcinoma of the Fallopian Tube // *Gynecol. Oncol.* – 2001. – V. 80. – No. 3. – P. 341–345.

29. Zweemer R.P., van Diest P.J., Verheijen R.H.M., et al. Molecular Evidence Linking Primary Cancer of the Fallopian Tube to BRCA1 Germline Mutations // *Gynecol. Oncol.* – 2000. – V. 76. – No. 1. – P. 45–50.

30. Thompson D., Easton D.F., Breast Cancer Linkage Consortium. Cancer Incidence in BRCA1 mutation carriers. // *J. Natl. Cancer Inst.* – 2002. – V. 94. – No. 18. – P. 1358–1365.

31. Vaughn J.P., Davis P.L., Jarboe M.D., et al. BRCA1 expression is induced before DNA synthesis in both normal and tumor-derived breast cells. // *Cell Growth Differ.* – 1996. – V. 7. – No. 6. – P. 711–715.

32. Chen Y., Farmer A.A., Chen C.F., Jones D.C., Chen P.L., Lee W.H. BRCA1 is a 220-kDa nuclear phosphoprotein that is expressed and phosphorylated in a cell cycle-dependent manner. // *Cancer Res.* – 1996. – V. 56. – No. 14. – P. 3168–3172.

33. Gudas J.M., Li T., Nguyen H., Jensen D., Rauscher F.J., Cowan K.H. Cell cycle regulation of BRCA1 messenger RNA in human breast epithelial cells. // *Cell Growth Differ.* – 1996. – V. 7. – No. 6. – P. 717–723.

34. Orban T.I., Olah E. Expression Profiles of BRCA1 Splice Variants in Asynchronous and in G1/S Synchronized Tumor Cell Lines // *Biochem. Biophys. Res. Commun.* – 2001. – V. 280. – No. 1. – P. 32–38.

35. Larson J.S., Tonkinson J.L., Lai M.T. A BRCA1 mutant alters G2-M cell cycle control in human mammary epithelial cells. // *Cancer Res.* – 1997. – V. 57. – No. 16. – P. 3351–3355.

36. Yarden R.I., Pardo-Reoyo S., Sgagias M., Cowan K.H., Brody L.C. BRCA1 regulates the G2/M checkpoint by activating Chk1 kinase upon DNA damage. // *Nat. Genet.* – 2002. – V. 30. – No. 3. – P. 285–289.

37. Okada S., Ouchi T. Cell cycle differences in DNA damage-induced BRCA1 phosphorylation affect its subcellular localization. // *J. Biol. Chem.* – 2003. – V. 278. – No. 3. – P. 2015–2020.

38. Dasika G.K., Lin S.C., Zhao S., Sung P., Tomkinson A., Lee E.Y. DNA damage-induced cell cycle checkpoints and DNA strand break repair in development and tumorigenesis. // *Oncogene.* – 1999. – V. 18. – No. 55. – P. 7883–7899.

39. Xu X., Qiao W., Linke S.P., et al. Genetic interactions between tumor suppressors Brcal and p53 in apoptosis, cell cycle and tumorigenesis. // *Nat. Genet.* – 2001. – V. 28. – No. 3. – P. 266–271.

40. Xu B., Kim St., Kastan M.B. Involvement of Brca1 in S-phase and G(2)-phase checkpoints after ionizing irradiation. // *Mol. Cell. Biol.* – 2001. – V. 21. – No. 10. – P. 3445–3450.
41. Xu B., O'Donnell A.H., Kim S.-T., Kastan M.B. Phosphorylation of serine 1387 in Brca1 is specifically required for the Atm-mediated S-phase checkpoint after ionizing irradiation. // *Cancer Res.* – 2002. – V. 62. – No. 16. – P. 4588–4591.
42. Kerr P., Ashworth A. New complexities for BRCA1 and BRCA2 // *Curr. Biol.* – 2001. – V. 11. – No. 16. – P. 668–676.
43. Venkitaraman A.R., Anderson S.F., Schlegel B.P., et al. Cancer Susceptibility and the Functions of BRCA1 and BRCA2 // *Cell.* – 2002. – V. 108. – No. 2. – P. 171–182.
44. Saha J., Davis A.J. Unsolved mystery: the role of BRCA1 in DNA end-joining // *J. Radiat. Res.* – 2016. – V. 57. – No. S1. – P. 18–24.
45. Monteiro A.N., August A., Hanafusa H. Evidence for a transcriptional activation function of BRCA1 C-terminal region. // *Proc. Natl. Acad. Sci. U. S. A.* – 1996. – V. 93. – No. 24. – P. 13595–13599.
46. Zheng L., Pan H., Li S., et al. Sequence-Specific Transcriptional Corepressor Function for BRCA1 through a Novel Zinc Finger Protein, ZBRK1 // *Mol. Cell.* – 2000. – V. 6. – No. 4. – P. 757–768.
47. Bochar D.A., Wang L., Beniya H., et al. BRCA1 Is Associated with a Human SWI/SNF-Related Complex: Linking Chromatin Remodeling to Breast Cancer // *Cell.* – 2000. – V. 102. – No. 2. – P. 257–265.
48. Monteiro A.N.. BRCA1: exploring the links to transcription // *Trends Biochem. Sci.* – 2000. – V. 25. – No. 10. – P. 469–474.
49. Krum S.A., Miranda G.A., Lin C., Lane T.F. BRCA1 associates with processive RNA polymerase II. // *J. Biol. Chem.* – 2003. – V. 278. – No. 52. – P. 52012–52020.
50. Scully R., Anderson S.F., Chao D.M., et al. BRCA1 is a component of the RNA polymerase II holoenzyme. // *Proc. Natl. Acad. Sci. U. S. A.* – 1997. – V. 94. – No. 11. – P. 5605–5610.
51. Clark S.L., Rodriguez A.M., Snyder R.R., Hankins G.D. V., Boehning D. Structure-Function Of The Tumor Suppressor BRCA1. // *Comput. Struct. Biotechnol. J.* – 2012. – V. 1. – No. 1. – P. e201204005.
52. Lorick K.L., Jensen J.P., Fang S., Ong A.M., Hatakeyama S., Weissman A.M. RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. // *Proc. Natl. Acad. Sci. U. S. A.* – 1999. – V. 96. – No. 20. – P. 11364–11369.
53. Hashizume R., Fukuda M., Maeda I., et al. The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. // *J. Biol. Chem.* – 2001. – V. 276. – No. 18. – P. 14537–14540.

54. Chen C.-F., Li S., Chen Y., Chen P.-L., Sharp Z.D., Lee W.-H. The Nuclear Localization Sequences of the BRCA1 Protein Interact with the Importin- Subunit of the Nuclear Transport Signal Receptor // *J. Biol. Chem.* – 1996. – V. 271. – No. 51. – P. 32863–32868.
55. Brzovic P.S., Rajagopal P., Hoyt D.W., King M.C., Klevit R.E. Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. // *Nat. Struct. Biol.* – 2001. – V. 8. – No. 10. – P. 833–837.
56. Rodríguez J.A., Henderson B.R. Identification of a functional nuclear export sequence in BRCA1. // *J. Biol. Chem.* – 2000. – V. 275. – No. 49. – P. 38589–38596.
57. Henderson B.R. Regulation of BRCA1, BRCA2 and BARD1 intracellular trafficking. // *Bioessays.* – 2005. – V. 27. – No. 9. – P. 884–893.
58. Zhang F., Ma J., Wu J., et al. PALB2 links BRCA1 and BRCA2 in the DNA-damage response. // *Curr. Biol.* – 2009. – V. 19. – No. 6. – P. 524–529.
59. Zhang F., Fan Q., Ren K., Andreassen P.R. PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2. // *Mol. Cancer Res.* – 2009. – V. 7. – No. 7. – P. 1110–1118.
60. Sy S.M.H., Huen M.S.Y., Chen J. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. // *Proc. Natl. Acad. Sci. U. S. A.* – 2009. – V. 106. – No. 17. – P. 7155–7160.
61. Zhang J., Willers H., Feng Z., et al. Chk2 phosphorylation of BRCA1 regulates DNA double-strand break repair. // *Mol. Cell. Biol.* – 2004. – V. 24. – No. 2. – P. 708–718.
62. Cantor S.B., Bell D.W., Ganesan S., et al. BACH1, a Novel Helicase-like Protein, Interacts Directly with BRCA1 and Contributes to Its DNA Repair Function // *Cell.* – 2001. – V. 105. – No. 1. – P. 149–160.
63. Greenberg R.A., Sobhian B., Pathania S., Cantor S.B., Nakatani Y., Livingston D.M. Multifactorial contributions to an acute DNA damage response by BRCA1/BARD1-containing complexes. // *Genes Dev.* – 2006. – V. 20. – No. 1. – P. 34–46.
64. Kim H., Chen J., Yu X. Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. // *Science.* – 2007. – V. 316. – No. 5828. – P. 1202–1205.
65. Kim H., Huang J., Chen J. CCDC98 is a BRCA1-BRCT domain-binding protein involved in the DNA damage response. // *Nat. Struct. Mol. Biol.* – 2007. – V. 14. – No. 8. – P. 710–715.
66. Yu X., Fu S., Lai M., Baer R., Chen J. BRCA1 ubiquitinates its phosphorylation-dependent binding partner CtIP. // *Genes Dev.* – 2006. – V. 20. – No. 13. – P. 1721–1726.
67. Cortez D., Wang Y., Qin J., Elledge S.J. Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. // *Science.* – 1999. – V. 286. – No. 5442. – P. 1162–1166.

68. Yun M.H., Hiom K. CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. // *Nature*. – 2009. – V. 459. – No. 7245. – P. 460–463.
69. Ford D., Easton D.F., Stratton M., et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. // *Am. J. Hum. Genet.* – 1998. – V. 62. – No. 3. – P. 676–689.
70. Breast Cancer Linkage Consortium. Cancer risks in BRCA2 mutation carriers. // *J. Natl. Cancer Inst.* – 1999. – V. 91. – No. 15. – P. 1310–1316.
71. Roy R., Chun J., Powell S.N. BRCA1 and BRCA2: different roles in a common pathway of genome protection. // *Nat. Rev. Cancer*. – 2012. – V. 12. – No. 1. – P. 68–78.
72. Choi E., Park P.G.P.-G., Lee H.-O.H.C.H.O.H., et al. BRCA2 fine-tunes the spindle assembly checkpoint through reinforcement of BubR1 acetylation. // *Dev. Cell*. – 2012. – V. 22. – No. 2. – P. 295–308.
73. Daniels M.J., Wang Y., Lee M., Venkitaraman A.R. Abnormal cytokinesis in cells deficient in the breast cancer susceptibility protein BRCA2. // *Science*. – 2004. – V. 306. – No. 5697. – P. 876–879.
74. Shamoo Y. Structural insights into BRCA2 function // *Curr. Opin. Struct. Biol.* – 2003. – V. 13. – No. 2. – P. 206–211.
75. Laboratories M.G. BRACAnalysis® Technical Specifications // 2012. – P. 1–2.
76. Baldwin A.L., Cook-Deegan R. Constructing narratives of heroism and villainy: case study of Myriad's BRACAnalysis® compared to Genentech's Herceptin®. // *Genome Med.* – 2013. – V. 5. – No. 1. – P. 8.
77. Chan P.C., Wong B.Y., Ozcelik H., Cole D.E. Simple and rapid detection of BRCA1 and BRCA2 mutations by multiplex mutagenically separated PCR. // *Clin. Chem.* – 1999. – V. 45. – No. 8 Pt 1. – P. 1285–1287.
78. Hauss O., Müller O. The Protein Truncation Test in Mutation Detection and Molecular Diagnosis // In: *In Vitro Transcription and Translation Protocols*. – V. 375. – Totowa, NJ. – Humana Press. – 2007. – P. 151–164.
79. Walsh T., Lee M.K., Casadei S., et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. // *Proc. Natl. Acad. Sci. U. S. A.* – 2010. – V. 107. – No. 28. – P. 12629–12633.
80. De Leeneer K., Helleman J., De Schrijver J., et al. Massive parallel amplicon sequencing of the breast cancer genes BRCA1 and BRCA2: opportunities, challenges, and limitations // *Hum. Mutat.* – 2011. – V. 32. – No. 3. – P. 335–344.
81. Ozcelik H., Shi X., Chang M.C., et al. Long-Range PCR and Next-Generation Sequencing of BRCA1 and BRCA2 in Breast Cancer // *J. Mol. Diagnostics*. – 2012. – V. 14. – No. 5. – P. 467–

475.

82. Dacheva D., Dodova R., Popov I., et al. Validation of an NGS Approach for Diagnostic BRCA1/BRCA2 Mutation Testing // *Mol. Diagn. Ther.* – 2015. – V. 19. – No. 2. – P. 119–130.

83. Vaca-Paniagua F., Alvarez-Gomez R.M., Fragoso-Ontiveros V., et al. Full-exon pyrosequencing screening of BRCA germline mutations in Mexican women with inherited breast and ovarian cancer. // *PLoS One.* – 2012. – V. 7. – No. 5. – P. e37432.

84. Michils G., Hollants S., Dehaspe L., et al. Molecular Analysis of the Breast Cancer Genes BRCA1 and BRCA2 Using Amplicon-Based Massive Parallel Pyrosequencing // *J. Mol. Diagn.* – 2012. – V. 14. – No. 6. – P. 623–630.

85. Hernan I., Borràs E., de Sousa Dias M., et al. Detection of genomic variations in BRCA1 and BRCA2 genes by long-range PCR and next-generation sequencing. // *J. Mol. Diagn.* – 2012. – V. 14. – No. 3. – P. 286–293.

86. Feliubadaló L., Lopez-Doriga A., Castellsagué E., et al. Next-generation sequencing meets genetic diagnostics: development of a comprehensive workflow for the analysis of BRCA1 and BRCA2 genes. // *Eur. J. Hum. Genet.* – 2013. – V. 21. – No. 8. – P. 864–870.

87. Tarabeux J., Zeitouni B., Moncoutier V., et al. Streamlined ion torrent PGM-based diagnostics: BRCA1 and BRCA2 genes as a model. // *Eur. J. Hum. Genet.* – 2014. – V. 22. – No. 4. – P. 535–541.

88. D'Argenio V., Esposito M.V., Telese A., et al. The molecular analysis of BRCA1 and BRCA2: Next-generation sequencing supersedes conventional approaches // *Clin. Chim. Acta.* – 2015. – V. 446. – P. 221–225.

89. Lai S., Brookes C., Prosser D.O., Lan C.-C., Doherty E., Love D.R. Diagnostic Screening Workflow for Mutations in the BRCA1 and BRCA2 Genes. // *Sultan Qaboos Univ. Med. J.* – 2015. – V. 15. – No. 1. – P. 58-70.

90. Ellison G., Huang S., Carr H., et al. A reliable method for the detection of BRCA1 and BRCA2 mutations in fixed tumour tissue utilising multiplex PCR-based targeted next generation sequencing. // *BMC Clin. Pathol.* – 2015. – V. 15. – P. 5.

91. Koczkowska M., Zuk M., Gorczynski A., et al. Detection of somatic BRCA1/2 mutations in ovarian cancer - next-generation sequencing analysis of 100 cases. // *Cancer Med.* – 2016. – V. 5. – No. 7. – P. 1640–1646.

92. Schenk D., Song G., Ke Y., Wang Z. Amplification of overlapping DNA amplicons in a single-tube multiplex PCR for targeted next-generation sequencing of BRCA1 and BRCA2 // Palsson A, editor. *PLoS One.* – 2017. – V. 12. – No. 7. – P. e0181062.

93. Neveling K., Mensenkamp A.R., Derks R., et al. BRCA Testing by Single-Molecule Molecular Inversion Probes // *Clin. Chem.* – 2017. – V. 63. – No. 2. – P. 503–512.

94. Huijismans C.J.J., Damen J., van der Linden J.C., et al. Comparative analysis of four methods to extract DNA from paraffin-embedded tissues: effect on downstream molecular applications // *BMC Res. Notes.* – 2010. – V. 3. – No. 1. – P. 239.
95. Li G., Guo X., Tang L., et al. Analysis of BRCA1/2 mutation spectrum and prevalence in unselected Chinese breast cancer patients by next-generation sequencing // *J. Cancer Res. Clin. Oncol.* – 2017.
96. Buzolin A.L., Moreira C.M., Sacramento P.R., et al. Development and validation of a variant detection workflow for BRCA1 and BRCA2 genes and its clinical application based on the Ion Torrent technology. // *Hum. Genomics.* – 2017. – V. 11. – No. 1. – P. 14.
97. Goodwin S., McPherson J.D., McCombie W.R. Coming of age: ten years of next-generation sequencing technologies. // *Nat. Rev. Genet.* – 2016. – V. 17. – No. 6. – P. 333–351.
98. Kang H.P., Maguire J.R., Chu C.S., et al. Design and validation of a next generation sequencing assay for hereditary *BRCA1* and *BRCA2* mutation testing // *PeerJ.* – 2016. – V. 4. – P. e2162.
99. Shin S., Hwang I.S., Lee S.-T., Choi J.R. Evaluation of an amplicon-based next-generation sequencing panel for detection of BRCA1 and BRCA2 genetic variants // *Breast Cancer Res. Treat.* – 2016. – V. 158. – No. 3. – P. 433–440.
100. DePristo M.A., Banks E., Poplin R., et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. // *Nat. Genet.* – 2011. – V. 43. – No. 5. – P. 491–498.
101. Li H., Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform // *Bioinformatics.* – 2009. – V. 25. – No. 14. – P. 1754–1760.
102. Li H., Handsaker B., Wysoker A., et al. The Sequence Alignment/Map format and SAMtools. // *Bioinformatics.* – 2009. – V. 25. – No. 16. – P. 2078–2079.
103. Ebbert M.T.W., Wadsworth M.E., Staley L.A., et al. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. // *BMC Bioinformatics.* – 2016. – V. 17 Suppl 7. – No. Suppl 7. – P. 239.
104. McKenna A., Hanna M., Banks E., et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. // *Genome Res.* – 2010. – V. 20. – No. 9. – P. 1297–1303.
105. Cingolani P., Platts A., Wang L.L., et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. // *Fly (Austin).* – V. 6. – No. 2. – P. 80–92.
106. Wang K., Li M., Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. // *Nucleic Acids Res.* – 2010. – V. 38. – No. 16. – P. e164.

107. Abecasis G.R., Auton A., Brooks L.D., et al. An integrated map of genetic variation from 1,092 human genomes. // *Nature*. – 2012. – V. 491. – No. 7422. – P. 56–65.
108. Sherry S.T., Ward M.H., Kholodov M., et al. dbSNP: the NCBI database of genetic variation. // *Nucleic Acids Res.* – 2001. – V. 29. – No. 1. – P. 308–311.
109. Landrum M.J., Lee J.M., Benson M., et al. ClinVar: public archive of interpretations of clinically relevant variants. // *Nucleic Acids Res.* – 2016. – V. 44. – No. D1. – P. 862-868.
110. Ng P.C., Henikoff S. SIFT: predicting amino acid changes that affect protein function // *Nucleic Acids Res.* – 2003. – V. 31. – No. 13. – P. 3812–3814.
111. Adzhubei I., Jordan D.M., Sunyaev S.R. Predicting functional effect of human missense mutations using PolyPhen-2. // *Curr. Protoc. Hum. Genet.* – 2013. – V. Chapter 7. – P. Unit7.20.
112. Rebbeck T.R., Mitra N., Wan F., et al. Association of type and location of BRCA1 and BRCA2 mutations with risk of breast and ovarian cancer. // *JAMA*. – 2015. – V. 313. – No. 13. – P. 1347–1361.
113. Meisel C., Sadowski C.E., Kohlstedt D., et al. Spectrum of genetic variants of BRCA1 and BRCA2 in a German single center study // *Arch. Gynecol. Obstet.* – 2017. – V. 295. – No. 5. – P. 1227–1238.
114. Wu X., Wu L., Kong B., et al. The First Nationwide Multicenter Prevalence Study of Germline BRCA1 and BRCA2 Mutations in Chinese Ovarian Cancer Patients // *Int. J. Gynecol. Cancer*. – 2017. – P. 1.
115. Lang G.-T., Shi J.-X., Hu X., et al. The spectrum of BRCA mutations and characteristics of BRCA-associated breast cancers in China: Screening of 2,991 patients and 1,043 controls by next-generation sequencing. // *Int. J. Cancer*. – 2017. – V. 141. – No. 1. – P. 129–142.
116. Zhang J., Sun J., Chen J., et al. Comprehensive analysis of BRCA1 and BRCA2 germline mutations in a large cohort of 5931 Chinese women with breast cancer // *Breast Cancer Res. Treat.* – 2016. – V. 158. – No. 3. – P. 455–462.
117. Maistro S., Teixeira N., Encinas G., et al. Germline mutations in BRCA1 and BRCA2 in epithelial ovarian cancer patients in Brazil // *BMC Cancer*. – 2016. – V. 16. – No. 1. – P. 934.
118. Alvarez C., Tapia T., Perez-Moreno E., et al. BRCA1 and BRCA2 founder mutations account for 78% of germline carriers among hereditary breast cancer families in Chile // *Oncotarget*. – 2017.
119. Torres D., Bermejo J.L., Rashid M.U., et al. Prevalence and Penetrance of BRCA1 and BRCA2 Germline Mutations in Colombian Breast Cancer Patients // *Sci. Rep.* – 2017. – V. 7. – No. 1. – P. 4713.
120. Jalkh N., Chouery E., Haidar Z., et al. Next-generation sequencing in familial breast cancer patients from Lebanon // *BMC Med. Genomics*. – 2017. – V. 10. – No. 1. – P. 8.

121. Наседкина Т.В., Громыко О.Е., Емельянова М.А., и др. Определение герминальных мутаций в генах BRCA1, BRCA2 и CHEK2 с использованием биочипов у больных раком молочной железы // Молекулярная биология. – Т. 48. – №2. – С. 243–250.
122. Suspitsin E.N., Sherina N.Y., Ponomariova D.N., и др. High frequency of BRCA1, but not CHEK2 or NBS1 (NBN), founder mutations in Russian ovarian cancer patients // Hered. Cancer Clin. Pract. – 2009. – V. 7. – No. 1. – P. 5.
123. Iyevleva A.G., Suspitsin E.N., Kroeze K., et al. Non-founder BRCA1 mutations in Russian breast cancer patients // Cancer Lett. – 2010. – V. 298. – No. 2. – P. 258–263.
124. 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. – V. 10. – No. 5. – P. 467–473.
125. Tulay P., Doshi A., Serhal P., SenGupta S.B. Differential expression of parental alleles of BRCA1 in human preimplantation embryos. // Eur. J. Hum. Genet. – 2016. – V. 25. – No. 1. – P. 37–42.
126. Vos J.R., Oosterwijk J.C., Aalfs C.M., et al. Bias Explains Most of the Parent-of-Origin Effect on Breast Cancer Risk in BRCA1/2 Mutation Carriers // Cancer Epidemiol. Prev. Biomarkers. – 2016. – V. 25. – No. 8. – P. 1251–1258.
127. Peterlongo P., Chang-Claude J., Moysich K.B., et al. Candidate Genetic Modifiers for Breast and Ovarian Cancer Risk in BRCA1 and BRCA2 Mutation Carriers // Cancer Epidemiol. Biomarkers Prev. – 2015. – V. 24. – No. 1. – P. 308–316.
128. Couch F.J., Wang X., McGuffog L., et al. Genome-Wide Association Study in BRCA1 Mutation Carriers Identifies Novel Loci Associated with Breast and Ovarian Cancer Risk // Hunter KW, editor. PLoS Genet. – 2013. – V. 9. – No. 3. – P. e1003212.
129. Kuchenbaecker K.B., McGuffog L., Barrowdale D., et al. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers // JNCI J. Natl. Cancer Inst. – 2017. – V. 109. – No. 7.
130. Walker L.C., Marquart L., Pearson J.F., et al. Evaluation of copy-number variants as modifiers of breast and ovarian cancer risk for BRCA1 pathogenic variant carriers. // Eur. J. Hum. Genet. – 2017. – V. 25. – No. 4. – P. 432–438.
131. Hamdi Y., Soucy P., Adoue V., et al. Association of breast cancer risk with genetic variants showing differential allelic expression: Identification of a novel breast cancer susceptibility locus at 4q21 // Oncotarget. – 2016. – V. 7. – No. 49. – P. 80140–80163.
132. Hamdi Y., Soucy P., Kuchenbaecker K.B., et al. Association of breast cancer risk in BRCA1 and BRCA2 mutation carriers with genetic variants showing differential allelic expression: identification of a modifier of breast cancer risk at locus 11q22.3 // Breast Cancer Res. Treat. –

2017. – V. 161. – No. 1. – P. 117–134.

133. Gowen L.C., Johnson B.L., Latour A.M., Sulik K.K., Koller B.H. Brca1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities. // *Nat. Genet.* – 1996. – V. 12. – No. 2. – P. 191–4.

134. Jeong J.-H., Jo A., Park P., Lee H., Lee H.-O. Brca2 deficiency leads to T cell loss and immune dysfunction. // *Mol. Cells.* – 2015. – V. 38. – No. 3. – P. 251–8.

135. Yan D.-H., Wen Y., Su L.-K., et al. A delayed chemically induced tumorigenesis in Brca2 mutant mice // *Oncogene.* – 2004. – V. 23. – No. 10. – P. 1896–1901.

136. Evers B., Jonkers J. Mouse models of BRCA1 and BRCA2 deficiency: past lessons, current understanding and future prospects // *Oncogene.* – 2006. – V. 25. – No. 43. – P. 5885–5897.

137. Wong-Brown M., McPhillips M., Gleeson M., et al. When is a mutation not a mutation: the case of the c.594-2A>C splice variant in a woman harbouring another BRCA1 mutation in trans // *Hered. Cancer Clin. Pract.* – 2016. – V. 14. – No. 1. – P. 6.

138. Kuschel B., Gayther S.A., Easton D.F., Ponder B.A., Pharoah P.D. Apparent human BRCA1 knockout caused by mispriming during polymerase chain reaction: implications for genetic testing. // *Genes. Chromosomes Cancer.* – 2001. – V. 31. – No. 1. – P. 96–98.

139. Domchek S.M., Tang J., Stopfer J., et al. Biallelic deleterious BRCA1 mutations in a woman with early-onset ovarian cancer. // *Cancer Discov.* – 2013. – V. 3. – No. 4. – P. 399–405.

140. Orban T.I., Olah E. Emerging roles of BRCA1 alternative splicing. // *Mol. Pathol.* – 2003. – V. 56. – No. 4. – P. 191–197.

141. Sawyer S.L., Tian L., Kähkönen M., et al. Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype. // *Cancer Discov.* – 2015. – V. 5. – No. 2. – P. 135–142.

142. Mathew C.G. Fanconi anaemia genes and susceptibility to cancer // *Oncogene.* – 2006. – V. 25. – No. 43. – P. 5875–5884.

143. Rebbeck T.R., Friebel T.M., Mitra N., et al. Inheritance of deleterious mutations at both BRCA1 and BRCA2 in an international sample of 32,295 women // *Breast Cancer Res.* – 2016. – V. 18. – No. 1. – P. 112.

144. Sokolenko A.P., Voskresenskiy D.A., Iyevleva A.G., et al. Large family with both parents affected by distinct BRCA1 mutations: implications for genetic testing. // *Hered. Cancer Clin. Pract.* – 2009. – V. 7. – No. 1. – P. 2.

145. Moschetta M., George A., Kaye S.B., Banerjee S. BRCA somatic mutations and epigenetic BRCA modifications in serous ovarian cancer // *Ann. Oncol.* – 2016. – V. 27. – No. 8. – P. 1449–1455.

146. Chao A., Chang T.-C., Lapke N., et al. Prevalence and clinical significance of BRCA1/2 germline and somatic mutations in Taiwanese patients with ovarian cancer // *Oncotarget.* – 2016.

– V. 7. – No. 51. – P. 85529–85541.

147. Winter C., Nilsson M.P., Olsson E., et al. Targeted sequencing of BRCA1 and BRCA2 across a large unselected breast cancer cohort suggests that one-third of mutations are somatic. // *Ann. Oncol.* – 2016. – V. 27. – No. 8. – P. 1532–1538.

148. Nik-Zainal S., Davies H., Staaf J., et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. // *Nature.* – 2016. – V. 534. – No. 7605. – P. 47–54.

149. Hino O., Kobayashi T. Mourning Dr. Alfred G. Knudson: the two-hit hypothesis, tumor suppressor genes, and the tuberous sclerosis complex. // *Cancer Sci.* – 2017. – V. 108. – No. 1. – P. 5–11.

150. John E.M., Miron A., Gong G., et al. Prevalence of pathogenic BRCA1 mutation carriers in 5 US racial/ethnic groups. // *JAMA.* – 2007. – V. 298. – No. 24. – P. 2869–2876.

151. Kang E., Seong M.-W., Park S.K., et al. The prevalence and spectrum of BRCA1 and BRCA2 mutations in Korean population: recent update of the Korean Hereditary Breast Cancer (KOHBRA) study. // *Breast Cancer Res. Treat.* – 2015. – V. 151. – No. 1. – P. 157–168.

152. Stadler Z.K., Salo-Mullen E., Patil S.M., et al. Prevalence of BRCA1 and BRCA2 mutations in Ashkenazi Jewish families with breast and pancreatic cancer. // *Cancer.* – 2012. – V. 118. – No. 2. – P. 493–499.

153. Karami F., Mehdipour P. A comprehensive focus on global spectrum of BRCA1 and BRCA2 mutations in breast cancer. // *Biomed Res. Int.* – 2013. – V. 2013. – P. 928562.

154. Gayther S.A., Harrington P., Russell P., Kharkevich G., Garkavtseva R.F., Ponder B.A. Frequently occurring germ-line mutations of the BRCA1 gene in ovarian cancer families from Russia. // *Am. J. Hum. Genet.* – 1997. – V. 60. – No. 5. – P. 1239–1242.

155. Tereshchenko I. V., Basham V.M., Ponder B.A.J., Pharoah P.D.P. BRCA1 and BRCA2 mutations in Russian familial breast cancer // *Hum. Mutat.* – 2002. – V. 19. – No. 2. – P. 184–184.

156. Tereshchenko I. V., Béshém V.M., Slonimskaja E.M., et al. [Investigation of mutations of BRCA1 and BRCA2 genes in 52 breast cancer patients]. // *Vopr. Onkol.* – 2002. – V. 48. – No. 1. – P. 24–28.

157. Loginova A.N., Pospekhova N.I., Lyubchenko L.N., et al. Spectrum of mutations in BRCA1 gene in hereditary forms of breast and ovarian cancer in Russian families. // *Bull. Exp. Biol. Med.* – 2003. – V. 136. – No. 3. – P. 276–278.

158. Грудинина Н.А., Голубков В.И., Тихомирова О.С., и др. Преобладание широко распространенных мутаций в гене BRCA1 у больных семейными формами рака молочной железы Санкт-Петербурга // *Генетика.* – 2005. – Т. 41. – №3. – С. 405–410.

159. Krylova N., Lobeiko O.S., Sokolenko A.P., et al. BRCA1 4153delA founder mutation in

- Russian ovarian cancer patients // *Hered. Cancer Clin. Pract.* – 2006. – V. 4. – No. 4. – P. 193–196.
160. Sokolenko A.P., Rozanov M.E., Mitiushkina N. V., et al. Founder mutations in early-onset, familial and bilateral breast cancer patients from Russia // *Fam. Cancer.* – 2007. – V. 6. – No. 3. – P. 281–286.
161. Митрофанов Д.В., Часовникова О.В., Коваленко С.П., Ляхович В.В. Метод выявления мутации 5382insc в гене BRCA1 человека с помощью флуоресцентномеченых олигонуклеотидов // *Молекулярная биология.* – 2009. – Т. 43. – №6. – С. 999–1005.
162. Часовникова О.Б., Митрофанов Д.В., Анисименко М.С., Воевода М.И., Коваленко С.П., Ляхович В.В. Анализ распространенности мутаций BRCA1 5382insC и CHEK2 1100delC у жителей Сибирского региона // *Генетика.* – 2012. – Т. 48. – №6. – С. 768–772.
163. Наседкина Т.В., Громыко О.Е., Емельянова М.А., и др. Определение герминальных мутаций в генах BRCA1, BRCA2 и CHEK2 с использованием биочипов у больных раком молочной железы // *Молекулярная биология.* – 2014. – Т. 48. – №2. – С. 243.
164. Cherdyntseva N., Gervas P., Voropaeva E., et al. New variants in the BRCA1 gene in Buryat Mongol breast cancer patients: Report from two families // *Cancer Biomarkers.* – 2017. – V. 18. – No. 3. – P. 291–296.
165. Liu X., Harada S. DNA isolation from mammalian samples. // *Curr. Protoc. Mol. Biol.* – 2013. – V. Chapter 2. – P. Unit2.14.
166. Diego S. Entire document and oligonucleotide sequences © 2007-2013 Illumina, Inc. All rights reserved. – 2014.
167. Nguyen-Dumont T., Pope B.J., Hammet F., Southey M.C., Park D.J. A high-plex PCR approach for massively parallel sequencing // *Biotechniques.* – 2013. – V. 55. – No. 2. – P. 69–74.
168. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data.
169. Garrison E., Marth G. Haplotype-based variant detection from short-read sequencing // 2012. – P. 1–9.
170. Cock P.J.A., Antao T., Chang J.T., et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. // *Bioinformatics.* – 2009. – V. 25. – No. 11. – P. 1422–1423.
171. Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. // *Bioinformatics.* – 2014. – V. 30. – No. 15. – P. 2114–2120.
172. Szabo C., Masiello A., Ryan J.F., Brody L.C. The Breast Cancer Information Core: Database design, structure, and scope // *Hum. Mutat.* – 2000. – V. 16. – No. 2. – P. 123–131.
173. Lek M., Karczewski K.J., Minikel E. V., et al. Analysis of protein-coding genetic variation

in 60,706 humans // *Nature*. – 2016. – V. 536. – No. 7616. – P. 285–291.

174. 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. – V. 337. – No. 6090. – P. 64–69.

175. Glusman G., Caballero J., Mauldin D.E., Hood L., Roach J.C. Kaviar: an accessible system for testing SNV novelty. // *Bioinformatics*. – 2011. – V. 27. – No. 22. – P. 3216–3217.

176. Landrum M.J., Lee J.M., Riley G.R., et al. ClinVar: public archive of relationships among sequence variation and human phenotype. // *Nucleic Acids Res.* – 2014. – V. 42. – No. Database issue. – P. 980–985.

177. Forbes S.A., Bhamra G., Bamford S., et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). // *Curr. Protoc. Hum. Genet.* – 2008. – V. Chapter 10. – P. Unit 10.11.

178. Chun S., Fay J.C. Identification of deleterious mutations within three human genomes // *Genome Res.* – 2009. – V. 19. – No. 9. – P. 1553–1561.

179. Schwarz J.M., Rödelsperger C., Schuelke M., Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. // *Nat. Methods*. – 2010. – V. 7. – No. 8. – P. 575–576.

180. Reva B., Antipin Y., Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. // *Nucleic Acids Res.* – 2011. – V. 39. – No. 17. – P. e118.

181. Shihab H.A., Gough J., Cooper D.N., et al. Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models // *Hum. Mutat.* – 2013. – V. 34. – No. 1. – P. 57–65.

182. Dong C., Wei P., Jian X., et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies // *Hum. Mol. Genet.* – 2015. – V. 24. – No. 8. – P. 2125–2137.

183. 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. – V. 17. – No. 5. – P. 405–424.

184. Guimera R.V. bcbio-nextgen: Automated, distributed next-gen sequencing pipeline // *EMBnet.journal*. – 2012. – V. 17. – No. B. – P. 30.

185. Koboldt D.C., Zhang Q., Larson D.E., et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. // *Genome Res.* – 2012. – V. 22. – No. 3. – P. 568–576.

186. Maxwell K.N., Wubbenhorst B., Wenz B.M., et al. BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. // *Nat. Commun.* – 2017. – V. 8. – No. 1. – P. 319.

187. Kanchi K.L., Johnson K.J., Lu C., et al. Integrated analysis of germline and somatic variants in ovarian cancer. // *Nat. Commun.* – 2014. – V. 5. – P. 3156.
188. Samorodnitsky E., Datta J., Jewell B.M., et al. Comparison of Custom Capture for Targeted Next-Generation DNA Sequencing // *J. Mol. Diagnostics.* – 2015. – V. 17. – No. 1. – P. 64–75.
189. Rehm H.L., Bale S.J., Bayrak-Toydemir P., et al. ACMG clinical laboratory standards for next-generation sequencing // *Genet. Med.* – 2013. – V. 15. – No. 9. – P. 733–747.
190. Sodha N., Williams R., Mangion J., Bullock S.L., Yuille M.R., Eeles R.A. Screening hCHK2 for mutations. // *Science.* – 2000. – V. 289. – No. 5478. – P. 359.
191. Arbustini E., Sgarella A., Ferrari A., et al. RE: BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers // *J. Natl. Cancer Inst.* – 2016. – V. 108. – No. 12. – P. 172.
192. Meeks H.D., Song H., Michailidou K., et al. BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers. // *J. Natl. Cancer Inst.* – 2016. – V. 108. – No. 2.
193. Osorio A., Milne R.L., Honrado E., et al. Classification of missense variants of unknown significance in *BRCA1* based on clinical and tumor information // *Hum. Mutat.* – 2007. – V. 28. – No. 5. – P. 477–485.
194. Beristain E., Guerra I., Vidaurrazaga N., Burgos-Bretones J., Tejada M.I. LOH analysis should not be used as a tool to assess whether UVs of BRCA1/2 are pathogenic or not // *Fam. Cancer* – 2010. – V. 9. – No. 3. – P. 289–290.
195. Rzepecka I.K., Szafron L., Stys A., et al. High frequency of allelic loss at the BRCA1 locus in ovarian cancers: clinicopathologic and molecular associations // *Cancer Genet.* – 2012. – V. 205. – No. 3. – P. 94–100.
196. Žilina O., Koltšina M., Raid R., Kurg A., Tõnisson N., Salumets A. Somatic mosaicism for copy-neutral loss of heterozygosity and DNA copy number variations in the human genome // *BMC Genomics.* – 2015. – V. 16. – No. 1. – P. 703.
197. Margraf R.L., VanSant-Webb C., Sant D., et al. Utilization of Whole-Exome Next-Generation Sequencing Variant Read Frequency for Detection of Lesion-Specific, Somatic Loss of Heterozygosity in a Neurofibromatosis Type 1 Cohort with Tibial Pseudarthrosis // *J. Mol. Diagnostics.* – 2017. – V. 19. – No. 3. – P. 468–474.
198. Zhang J., Lindroos A., Ollila S., et al. Gene Conversion Is a Frequent Mechanism of Inactivation of the Wild-Type Allele in Cancers from MLH1/MSH2 Deletion Carriers // *Cancer Res.* – 2006. – V. 66. – No. 2. – P. 659–664.
199. Maxwell K.N., Wubbenhorst B., Wenz B.M., et al. BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. // *Nat. Commun.* – 2017. – V. 8. – No.

1. – P. 319.

Приложение 1

Структуры праймеров для приготовления библиотеки

Таблица 1. Структуры ампликон-специфичной части праймеров для амплификации экзонов гена *BRCA1*.

BRCA1	Прямой праймер	Обратный праймер
1ex	5'-AGGCACTTTATGGCAAACCTCA-3'	5'-CATCTGTCAGCTTCGGAAATC-3'
2ex	5'-GGACGTTGTCATTAGTTCTTTGG-3'	5'-TGCATAGGAGATAATCATAGGAATC-3'
3ex	5'-TCAGTTCCTGACACAGCAGAC-3'	5'-TCCTGGGTTATGAAGGACAAA-3'
4ex	5'-ATGGCTCTTAAGGGCAGTTGT-3'	5'-TCCTACTGTGGTTGCTTCCAA-3'
6ex	5'-TCACACGGTTTATACAGATGTCA-3'	5'-GTTTTCATGGACAGCACTTGA-3'
7ex	5'-GAGCATACATAGGGTTTCTCTTGG-3'	5'-GAAGAAGAAGAAGAAAACAAATGG-3'
8ex	5'-CAGGAGGAAAAGCACAGAAC-3'	5'-CACTTCCCAAAGCTGCCTAC-3'
9ex	5'-CTGCCACAGTAGATGCTCAGTA-3'	5'-GGAAAATACCAGCTTCATAGACA-3'
10ex	5'-TGGTCAGCTTTCTGTAATCGAA-3'	5'-CTCTTTTCAGTGCCTGTTAAGTTG-3'
11a	5'-TAGCCAGTTGGTTGATTTCCA-3'	5'-TTCCTTACTTCCAGCCCATCT-3'
11b	5'-CAAACAGCCTGGCTTAGCA-3'	5'-CTCTCTACTGATTTGGAGTGAATC-3'
11c	AGTAGCTGATGTATTGGACGTTCT-3'	5'-CTTTTGTTTTATTCTCATGACCACTA-3'
11d	5'-GGAGCAGAATGGTCAAGTGAT-3'	5'-GCAGGTTCTTTACCTTCCATGA-3'
11e	5'-TGATAGTTGTTCTAGCAGTGAAGAG-3'	5'-CTGAGTGCCATAATCAGTACCA-3'
11f	5'-AAGACCCCAAAGATCTCATGT-3'	5'-ATGTTGCACATTCCTCTTCTG-3'
11g	5'-TTTAAGTATCCATTGGGACATGA-3'	5'-GCCTCTGAACTGAGATGATAGAC-3'
11h	5'-GTCAGAAAGATAAGCCAGTTGA-3'	5'-TGCTTGAATGTTTTTCATCACTG-3'
11i	5'-TTAAAGAAGCCAGCTCAAGCA-3'	5'-ACAGCAGAACTTTCCTTAATGTCA-3'
11k	5'-TGCATCTCAGGTTTGTCTGA-3'	5'-TGCTTTGCCAATATTACCTG-3'
11m	5'-CATTGAAGAATAGCTTAAATGACTG-3'	5'-AATAGACTGGGGCAAACACA-3'
12ex	5'-CAGCAAGTTGCAGCGTTTATAG-3'	5'-GAATGCAAAGGACACCACA-3'
13ex	5'-GTAATTCATTTTCTTGGTGCCAT-3'	5'-TGTGCTGAGCAAGGATCATAA-3'
14ex	5'-TCTGTCTGTTGCATTGCTTGT-3'	5'-AGATGTCAGATACCACAGCATC-3'
15ex	5'-CAGTCATTTCTGATCTCTCTGACA-3'	5'-TCAAAGTGTTTGTCCAATACAG-3'
16ex	5'-GTAATTCACATTCATCGTTGTG-3'	5'-TGTTAAGTCTTAGTCATTAGGGAGA-3'
17ex	5'-GCTGTGTGCTAGAGGTAACCTCA-3'	5'-ATGTGGTTTTTATGCAGCAGAT-3'
18ex	5'-GGCTCTTTAGCTTCTTAGGACA-3'	5'-CAATTCTGAGGTGTTAAAGGGA-3'
19ex	5'-AAGGACCTCTCCTCTGTCATTC-3'	5'-AAAGTGGTGCATTGATGGAAG-3'
20ex	5'-GTGTCTGCTCCACTTCCATTG-3'	5'-CCTGTGTGAAAGTATCTAGCACTG-3'
21ex	5'-CTGGACATTGGACTGCTTGT-3'	5'-AGGCTGGTGTGGAACCTCT-3'
22ex	5'-CCCATTGAGAGGTCTTGCTA-3'	5'-GGTGCCAGTCTTGCTCACA-3'
23ex	5'-GAAGTGACAGTTCCAGTAGTCCT-3'	5'-ATGAGTGATAAACCAAACCCAT-3'
24a	5'-GTCCAGGAGAATGAATTGACAC-3'	5'-TTGTGCTCATGGCAGATTTCT-3'
24b	5'-CCCTTAAAGATTTTCTGCTTGAAG-3'	5'-CATGCCCAGGTTTCAAGTTTC-3'
24c	5'-AACATTGGGGAGGAAATTCTG-3'	5'-CGGTTCTTGAAAATCTTCTGCT-3'
24d	5'-TGCTAGATTTCTAAAGAATGTGTTTC-3'	5'-TGAAGCTGTATGGTTTCAGCA-3'
24e	5'-GCTGTTGCTTTCTTTGAGGTG-3'	5'-CACAGTGAAAAGGCTCTGAGA-3'

Таблица 2. Структуры ампликон-специфичной части праймеров для амплификации экзонов гена *BRCA2*.

BRCA2	Прямой праймер	Обратный праймер
1ex	5'-CCTTTCGCCACACTGAGA-3'	5'-GCAGAGACAAAAGGGCAAGA-3'
2ex	5'-GTTCCAGGAGATGGGACTGA-3'	5'-ТСТААГСААСАСТГТГАСГТАСТГ-3'
3ex	5'-GATCTTTAACTGTTCTGGGTCACA-3'	5'-AGATTTTAAACACAGGTTTGCCTA-3'
4ex	5'-TCCCTATACATTCTCATTCCCAGTA-3'	5'-CTTCTACCAGGCTCTTAGCCA-3'
5-6	5'-CACTTGATGATTATTTAATGCTTCA-3'	5'-GGCAAAGGTATAACGCTATTGTC-3'
7ex	5'-CCTTAATGATCAGGGCATTTC-3'	5'-TGACACCACTGGACTACCACT-3'
8ex	5'-GCCATATCTTACCACCTTGTGA-3'	5'-CACATATAGGACCAGGTTTAGAGAC-3'
9ex	5'-GGGACTACTACTATATGTGCATTGAGA-3'	5'-AGGAGCAATCCTTCAATGGT-3'
10a	5'-CTGTTTCTATGAGAAAGGTTGTGA-3'	5'-GCTACATTTGAATСТААТGGATCA-3'
10b	5'-CTCATTGTATCTGAAGTGGAAC-3'	5'-GAATGCAGTCTGTATGAGATTCA-3'
10c	5'-GGTAAATAAGAGAGATGAAGAGCA-3'	5'-CCTGCATTCTTCAAAGCTACAGA-3'
10d	5'-GGAGGACTCCTTATGTCCAA-3'	5'-ACACAGAAGGAATCGTCATCTAT-3'
11a	5'-TTAGTGAATGTGATTGATGGTACTT-3'	5'-GAAAGTCAGTATCACTGTATTCCACT-3'
11b	5'-GCAGCATGTCACCCAGTACA-3'	5'-CTCTTAGATTTGTGTTTGGTTGA-3'
11c	5'-CATGAGAGTAGCATCACCTTCA-3'	5'-CCGATTTTAAATCTTGACCTAGAGT-3'
11d	5'-GGTTTATGTTCTTGCAGAGGAGA-3'	5'-CACTACTCTGAAATGTGCAGATAC-3'
11e	5'-GAAACTGAGCAAGCCTCAGTCA-3'	5'-GCTGCTGTCTACCTGACCAATC-3'
11f	5'-AGAGATGCTGATCTTCATGTCA-3'	5'-ACAAAAGTGCCAGTAGTCATTTTC-3'
11g	5'-CTTCAAGTAAATGTCATGATTCTGT-3'	5'-GCTTCTTGAGCTTTCGCAAC-3'
11h	5'-CACTCAGATTAAAGAAGATTGTGCA-3'	5'-CCCTGGAAGGTCACTAGTTGA-3'
11i	5'-AAGAAAGTGTCACAGTTGGTACT-3'	5'-CATAAATTATCACTTAAAGAGCTTAGGT-3'
11k	5'-CCTTGTTTCTATTGAGACTGTGGT-3'	5'-ATAGCTGTTAGACATGCTACTGTTAC-3'
11m	5'-ATTGATGGTCAACCAGAAAGA-3'	5'-TGGACAATTTAATGGCTGCAT-3'
11o	5'-TGCGTTGAGGAACCTTGTGACT-3'	5'-GAAACTTTCTCCAATCCAGACAT-3'
11p	5'-TTGCTGACATTCAGAGTGAAGA-3'	5'-GAGAAAGCAGATGAATTTACCACA-3'-3'
11s	5'-ТАСТGCTATACGТАСТCCAGAACА-3'	5'-GTGACACTTTGGTTCCTAATACC-3'
11t	5'-CTCTCTCAATTTCAACAAGACAAAC-3'	5'-AAAATAGTGATTGGCAACACGA-3'
12	5'-CCTGTTTAGACCCGTGTTAAATAGTG-3'	5'-CCATACCTATAGAGGGAGAACAGA-3'
13	5'-TTGAGCATCTGTTACATTCACTGA-3'	5'-CGGGAAGTGTTAACTTCTTAACG-3'
14a	5'-CCATGTAGCAAATGAGGGTCT-3'	5'-CTAACACACTGTTCAACTCTGTGA-3'
14b	5'-ACAGGCAGACCAACCAAGT-3'	5'-CGGAAATATCTAACTGAAAGGCA-3'
15	5'-GGGGTTGTGCTTTTTAAATTTTC-3'	5'-CCATTCCTGCACTAATGTGTTTC-3'
16	5'-TGTTTTTATTGTGTGATACATGTTTAC-3'	5'-AGAAAGAGGGATGAGGGGAATAC-3'
17	5'-AATTCAGTATCATCCTATGTGGTT-3'	5'-CATGGAAGTCACAGACTACACAGA-3'
18a	5'-CTAGAGTCACTTCCTAAAATATGC-3'	5'-ACCAACTGTCAGTCTGCCAT-3'
18b	5'-ACTTACAGATGGGTGGTATGCT-3'	5'-CTAGAATTTAACTGAATCAATGACTGA-3'
19	5'-GGCAGTCCTAGAAGAATGAAAATC-3'	5'-TGTATCAAAGAAAGAAATATATGGTAAAG-3'
20	5'-GAATGTTATATATGTGACTTTTTGGT-3'	5'-GTCTCTAAGACTTTGTTCTCATATTAG
21	5'-AATCTCCCTTCTTTGGGTGT-3'	5'-CCAGAGAGTCTAAAACAGCTTCTCA-3'
22u	5'-TGTTCTGATTGCTTTTTATCCAA-3'	5'-GGCATTAGTAGTGGATTTTGCT-3'
23	5'-GATAATCACTTCTTCCATTGCATC-3'	5'-GGAGATTCCATAAACTAACAAGCA
24	5'-GTGCTTGTAGTTTATGGAATCTC-3'	5'-TGCCAACCTGGTAGCTCCAACCT-3'
25	5'-GCATCTTAAAATTCATCTAACACATCT-3'	5'-CCTTGATACTGGACTGTCAAATAG-3'
26	5'-GGTCCAAACTTTTCAATTTCTGCT-3'	5'-GGAGCCACATAACAACCACA-3'
27a	5'-GACTGTGTGTAATATTTGCGTGCT-3'	5'-CTTTATGGGTGTTTCGTATTTGGT-3'
27b	5'-CATTTCAGCCACCAAGGAGT-3'	5'-CAGAGATGTAGTACAACGTCGTTTC-3'
27c	5'-GCTCCCACCAGTTCAGAAGA-3'	5'-GAAGCAAAAAGTATACCAATACGGA-3'
27d	5'-CACACATTAGTACTTATGTTGCACA-3'	5'-AGAAAATGGTATGTTTATTTCACTTG-3'
27e	5'-TCTTTGGATTTGATCACTACAAGT-3'	5'-GGAATTAGAGTTCACTGAGGGTTCT-3'
27f	5'-TCAGGATGAATATGAAGAGTGGT-3'	5'-AGGCTCTGGGTCCCTGTGATAG-3'

Таблица 3. Нуклеотидные последовательности праймеров-адаптеров (общий вид)

<i>Праймер</i>	<i>Общий вид последовательности</i>
D5	5'-aatgatacggcgaccaccgagatctacac/i5/acactctttccctacacgacgctcttccgatct-3'
D7	5'-caagcagaagacggcatacagat/i7/gtactggagttcagacgtgtgctcttccgatc-3'

Таблица 4. Нуклеотидные последовательности баркодов

Индекс баркода	Нуклеотидная последовательность баркода
D501	TATAGCCT
D502	ATAGAGGC
D503	CCTATCCT
D504	GGCTCTGA
D505	AGGCGAAG
D506	TAATCTTA
D507	CAGGACGT
D508	GTA CTGAC
D701	ATTACTCG
D702	TCCGGAGA
D703	CGCTCATT
D704	GAGATTCC
D705	ATTCAGAA
D706	GAATTCGT
D707	CTGAAGCT
D708	TAATGCGC
D709	CGGCTATG
D710	TCCGCGAA
D711	TCTCGCGC
D712	AGCGATAG

Приложение 2

Последовательности праймеров для мультиплексного метода приготовления библиотеки

Таблица 1. Последовательности праймеров для мультиплексного метода приготовления библиотеки.

Номер	Идентификатор	Последовательность праймера
1	BRCA1_1_F1	ctctctatgggcagtcggtgattACTTTGTAAGCTCATTCTTGGG
2	BRCA1_1_R1	ctgcgtgtctccgactcagAGAGTGGGTGTTGGACAGTGT
3	BRCA1_1_F2	ctctctatgggcagtcggtgattTCCTGGCACTGGTAGAGTGCT
4	BRCA1_1_R2	ctgcgtgtctccgactcagGATTAGAGCCTAGTCCAGGAG
5	BRCA1_2_F1	ctctctatgggcagtcggtgattCAAACCCATGCAAAAGGACCC
6	BRCA1_2_R1	ctgcgtgtctccgactcagTTGAATGCTCTTTCCTTCCTGG
7	BRCA1_3_F1	ctctctatgggcagtcggtgattATTGTGTCTCCCTCTCTGAC
8	BRCA1_3_R1	ctgcgtgtctccgactcagGCAGTGATTTTACATCTAAATGTC
9	BRCA1_4_F1	ctctctatgggcagtcggtgattTACTCCACTATGTAAGACAAAGG
10	BRCA1_4_R1	ctgcgtgtctccgactcagTGTCCCTCTCTCTCCTCTCT
11	BRCA1_5_F1	ctctctatgggcagtcggtgattTGTGAAAGTATCTAGCACTGTGT
12	BRCA1_5_R1	ctgcgtgtctccgactcagCCCAGGACAGAAAGGTAAAGC
13	BRCA1_5_F2	ctctctatgggcagtcggtgattTTTTGTCAACTTGAGGGAGGGA
14	BRCA1_5_R2	ctgcgtgtctccgactcagATCCTGATGGGTTGTGTTTGGT
15	BRCA1_6_F1	ctctctatgggcagtcggtgattGAAGGAAGCAAATACATTTTAACT
16	BRCA1_6_R1	ctgcgtgtctccgactcagTGTATGTAACCTGTCTTTTCTATG
17	BRCA1_7_F1	ctctctatgggcagtcggtgattTAAAGGGAGGAGGGGAGAAATA
18	BRCA1_7_R1	ctgcgtgtctccgactcagCTTCTAATCCTTTGAGTGTTTTTC
19	BRCA1_8_F1	ctctctatgggcagtcggtgattCTCCCAAAGTGCTGCGATTAC
20	BRCA1_8_R1	ctgcgtgtctccgactcagAАСТААТСТААТТАСТГААГАСТ
21	BRCA1_8_F2	ctctctatgggcagtcggtgattACCTGTTTTTCATAACAACATGAGT
22	BRCA1_8_R2	ctgcgtgtctccgactcagCTGAGCTGTGTGCTAGAGGTA
23	BRCA1_9_F1	ctctctatgggcagtcggtgattTTCCAGAATGTTGTТААГТСТТАГ
24	BRCA1_9_R1	ctgcgtgtctccgactcagAGCAGGGAGAAGCCAGAATTG
25	BRCA1_9_F2	ctctctatgggcagtcggtgattTGACCCTTCTGTTGAAGCTGT
26	BRCA1_9_R2	ctgcgtgtctccgactcagGCATTGAAAGTTCCCAATTGAA
27	BRCA1_9_F3	ctctctatgggcagtcggtgattCTCTGGGCAGATTCTGCAACT
28	BRCA1_9_R3	ctgcgtgtctccgactcagCTGGAATCAGCCTCTTCTCTG
29	BRCA1_9_F4	ctctctatgggcagtcggtgattGAAGGATCAGATTСАГГГТСАТ
30	BRCA1_9_R4	ctgcgtgtctccgactcagTCTTTTTAATTCTTAAACAGAGACCA
31	BRCA1_10_F1	ctctctatgggcagtcggtgattATAACAAAAGTGTCCATGATAGAC
32	BRCA1_10_R1	ctgcgtgtctccgactcagACATCTTACTTGCCAAGGCAAG
33	BRCA1_10_F2	ctctctatgggcagtcggtgattCAGCAGATGAAATATTACCTAGAT
34	BRCA1_10_R2	ctgcgtgtctccgactcagGTCTTCAGAATAGAACTACCCA
35	BRCA1_10_F3	ctctctatgggcagtcggtgattCCTTAAATGAGCTCCTCTTGAGA
36	BRCA1_10_R3	ctgcgtgtctccgactcagGTGGCGATGGTTTTCTCCTTC
37	BRCA1_11_F1	ctctctatgggcagtcggtgattTGGAGAAAGTATGGTGAAAAAATT

38	BRCA1_11_R1	ctgcgtgtctccgactcagTTGAGGTGTCTGCAGATAGTTC
39	BRCA1_11_F2	ctctctatgggcagtcggtgattCCTGGTTCTTTATTTTACTGGTA
40	BRCA1_11_R2	ctgcgtgtctccgactcagAAGTAGATTTGTTTTTCTCATTCCAT
41	BRCA1_12_F1	ctctctatgggcagtcggtgattTTCAGTAATTTGCCAAAATGACGA
42	BRCA1_12_R1	ctgcgtgtctccgactcagCCTTGATGTCTACAATTTACCT
43	BRCA1_13_F1	ctctctatgggcagtcggtgattTTTGCTTAAGATATCAGTGTGG
44	BRCA1_13_R1	ctgcgtgtctccgactcagGTTAGAACAGCATGGGAGCCA
45	BRCA1_13_F2	ctctctatgggcagtcggtgattGAAGGGTAGCTGTTAGAAGGC
46	BRCA1_13_R2	ctgcgtgtctccgactcagGTATTTCAATTTCTTGGTGCCATT
47	BRCA1_14_F1	ctctctatgggcagtcggtgattAACCACAAACAATTGTGCCATTAA
48	BRCA1_14_R1	ctgcgtgtctccgactcagAGTGACATTTTAACCACTCAGGT
49	BRCA1_14_F2	ctctctatgggcagtcggtgattTGCACACACACACGCTTTTT
50	BRCA1_14_R2	ctgcgtgtctccgactcagTTTACATCTGAACCTCTGTTTTTG
51	BRCA1_15_F1	ctctctatgggcagtcggtgattAAACATTTAGCTCACTTCTATAAATA
52	BRCA1_15_R1	ctgcgtgtctccgactcagTGACAAGGAATTGGTTTCAGATG
53	BRCA1_15_F2	ctctctatgggcagtcggtgattAAGCCCGTTCCTCTTCTTCAT
54	BRCA1_15_R2	ctgcgtgtctccgactcagTGAAGACTTGACTGCAAATACA
55	BRCA1_15_F3	ctctctatgggcagtcggtgattTCAAGAAAGGATCCTGGGTGTT
56	BRCA1_15_R3	ctgcgtgtctccgactcagGCAGTAACCAGGTAATATTGGC
57	BRCA1_15_F4	ctctctatgggcagtcggtgattGTGATGTTCTGAGATGCCTTT
58	BRCA1_15_R4	ctgcgtgtctccgactcagATACCTTCTCAGTCTACTAGGC
59	BRCA1_15_F5	ctctctatgggcagtcggtgattACTCGGTAGCAACGGTGCTAT
60	BRCA1_15_R5	ctgcgtgtctccgactcagAAGAGGGGCCAAGAAATTAGAG
61	BRCA1_15_F6	ctctctatgggcagtcggtgattCTAGATAAGTTCTCTTCTGAGGA
62	BRCA1_15_R6	ctgcgtgtctccgactcagCATTAAAGGAAAGTTCTGCTGTTTT
63	BRCA1_15_F7	ctctctatgggcagtcggtgattTCCTTTCTGGACGCTTTTGCTA
64	BRCA1_15_R7	ctgcgtgtctccgactcagATGGGAAGTAGTCATGCATCTC
65	BRCA1_15_F8	ctctctatgggcagtcggtgattTCAGGTGTCTCAGAACAAACCT
66	BRCA1_15_R8	ctgcgtgtctccgactcagTTGTAAGCATCCTGAAATAAAAAAG
67	BRCA1_15_F9	ctctctatgggcagtcggtgattCTGAACTACTTCTTCATATTCTTG
68	BRCA1_15_R9	ctgcgtgtctccgactcagCAGAACTAGGTAGAAACAGAGG
69	BRCA1_15_F10	ctctctatgggcagtcggtgattTAAGCATAGCATTCAATTTTGGC
70	BRCA1_15_R10	ctgcgtgtctccgactcagTAAAGAAGCCAGCTCAAGCAATA
71	BRCA1_15_F11	ctctctatgggcagtcggtgattTAGTACTGGAACCTACTTCATTAA
72	BRCA1_15_R11	ctgcgtgtctccgactcagTTGAGGAACATTCAATGTCACCT
73	BRCA1_15_F12	ctctctatgggcagtcggtgattTTCTCATTTCCTATTTCTCTTTC
74	BRCA1_15_R12	ctgcgtgtctccgactcagTAAACATGGACTTTTACAAAACCC
75	BRCA1_15_F13	ctctctatgggcagtcggtgattGAAAAAGTGGTGGTATACGATAT
76	BRCA1_15_R13	ctgcgtgtctccgactcagATAAGCCAGTTGATAATGCCAAAT
77	BRCA1_15_F14	ctctctatgggcagtcggtgattCTAGAGCCTCCTTTGATACTAC
78	BRCA1_15_R14	ctgcgtgtctccgactcagATGTGAACAAAAGGAAGAAAATCAA
79	BRCA1_15_F15	ctctctatgggcagtcggtgattCTTGATATTAGACTCATTCTTCC
80	BRCA1_15_R15	ctgcgtgtctccgactcagCTCCGTTTTCAAATCCAGGAAA
81	BRCA1_15_F16	ctctctatgggcagtcggtgattTGTTGCACATTCTCTTCTGCA
82	BRCA1_15_R16	ctgcgtgtctccgactcagGTTAACCACAGTCGGGAAACAA
83	BRCA1_15_F17	ctctctatgggcagtcggtgattTTCATTTCTTCCATTTCTATGC

84	BRCA1_15_R17	ctgcgtgtctccgactcagTCAGTGTGCAGCATTTGAAAAC
85	BRCA1_15_F18	ctctctatgggcagtcggtgattAACCATGAATTAGTCCCTTGGG
86	BRCA1_15_R18	ctgcgtgtctccgactcagTCATTGGTACCTGGTACTGATTA
87	BRCA1_15_F19	ctctctatgggcagtcggtgattGAGATACTTTCCTGAGTGCCA
88	BRCA1_15_R19	ctgcgtgtctccgactcagGAAACAGTTAAAGTGTCTAATAATG
89	BRCA1_15_F20	ctctctatgggcagtcggtgattATGAGATCTTTGGGGTCTTCAG
90	BRCA1_15_R20	ctgcgtgtctccgactcagCAAATGCACCTGGTCTTTTACT
91	BRCA1_15_F21	ctctctatgggcagtcggtgattAAGTTCACTGGTATTTGAACACTT
92	BRCA1_15_R21	ctgcgtgtctccgactcagAAGGTAAGAACCTGCAACTGG
93	BRCA1_15_F22	ctctctatgggcagtcggtgattTGGCTTGTTACTCTTCTGGCT
94	BRCA1_15_R22	ctgcgtgtctccgactcagTGTAATGCAAAATTGATAGTT
95	BRCA1_15_F23	ctctctatgggcagtcggtgattCTTTATCTTCTACTGCTAGAAC
96	BRCA1_15_R23	ctgcgtgtctccgactcagTCAAAGCACCTAAAAAGAATAGG
97	BRCA1_15_F24	ctctctatgggcagtcggtgattGGTAGAAGACTTCCTCCTCAG
98	BRCA1_15_R24	ctgcgtgtctccgactcagCCAATAGAATCACTCGAAAAAGA
99	BRCA1_15_F25	ctctctatgggcagtcggtgattCAGCTTTCGTTTTGAAAGCAGAT
100	BRCA1_15_R25	ctgcgtgtctccgactcagACCAAACGGAGCAGAATGGTC
101	BRCA1_15_F26	ctctctatgggcagtcggtgattCCACTATTAGTAATATTCATCACTT
102	BRCA1_15_R26	ctgcgtgtctccgactcagGCGTAAAAGGAGACCTACATCA
103	BRCA1_15_F27	ctctctatgggcagtcggtgattTAAAATCCTCAGGATGAAGGCC
104	BRCA1_15_R27	ctgcgtgtctccgactcagCCTCCCCAACTTAAGCCATGT
105	BRCA1_15_F28	ctctctatgggcagtcggtgattTGCTCCTATAATTAGATTTTCAGTT
106	BRCA1_15_R28	ctgcgtgtctccgactcagCTCATGAGGCTTTAATATGTAAAA
107	BRCA1_15_F29	ctctctatgggcagtcggtgattTGATTTGGAGTGAACCTTTTCAC
108	BRCA1_15_R29	ctgcgtgtctccgactcagCTGAATCAAATGCCAAAGTAGCT
109	BRCA1_15_F30	ctctctatgggcagtcggtgattTCATTTAGAACGTCCAATACATC
110	BRCA1_15_R30	ctgcgtgtctccgactcagTGTTCTTGGATAAACTAAATAG
111	BRCA1_15_F31	ctctctatgggcagtcggtgattACTCATTAACTTTCTGAATGCTG
112	BRCA1_15_R31	ctgcgtgtctccgactcagCAGAAAAAAAGGTAGATCTGAATG
113	BRCA1_15_F32	ctctctatgggcagtcggtgattTTCTCTCACACAGGGGATCAG
114	BRCA1_15_R32	ctgcgtgtctccgactcagTGAATTCTGTAATAAAAGCAAACAG
115	BRCA1_15_F33	ctctctatgggcagtcggtgattTTGGCTCCTTGCTAAGCCAGG
116	BRCA1_15_R33	ctgcgtgtctccgactcagTGTGGAGCCATGTGGCACAAA
117	BRCA1_15_F34	ctctctatgggcagtcggtgattCTGTAATGAGCTGGCATGAGTA
118	BRCA1_15_R34	ctgcgtgtctccgactcagTGAACATCATCAACCCAGTAATA
119	BRCA1_15_F35	ctctctatgggcagtcggtgattCTTCTCAGTGGTGTTCAAATCAT
120	BRCA1_15_R35	ctgcgtgtctccgactcagATGAAGTATGTATTTTTTTAATGACAA
121	BRCA1_16_F1	ctctctatgggcagtcggtgattTTCAGTGCCTGTAAAGTTGGCA
122	BRCA1_16_R1	ctgcgtgtctccgactcagGTTTCTTATTAGGACTCTGTCTTT
123	BRCA1_17_F1	ctctctatgggcagtcggtgattGAAACATCAATCCTTAATATTAACTA
124	BRCA1_17_R1	ctgcgtgtctccgactcagTATTGATTTATTTTTTGGGGGGAAA
125	BRCA1_18_F1	ctctctatgggcagtcggtgattAAGCTGCCTACCACAAATACAAA
126	BRCA1_18_R1	ctgcgtgtctccgactcagGGATACAACCTCAAAGACGTC
127	BRCA1_18_F2	ctctctatgggcagtcggtgattCCTTACCCAATTCAATGTAGACA
128	BRCA1_18_R2	ctgcgtgtctccgactcagCAATTGCTTGACTGTTCTTTACC
129	BRCA1_19_F1	ctctctatgggcagtcggtgattACTCCATCTCAAAAAAAAAAAAAAGAAA

130	BRCA1_19_R1	ctgcgtgtctccgactcagTCTATCATCCAAAGTATGGGCTA
131	BRCA1_19_F2	ctctctatgggcagtcggtgattAAGTCTTTTGGCACGGTTTCTG
132	BRCA1_19_R2	ctgcgtgtctccgactcagAACTGCAAACATAATGTTTTCCCT
133	BRCA1_20_F1	ctctctatgggcagtcggtgattGATACAGAACTAAAATTAACCTAGA
134	BRCA1_20_R1	ctgcgtgtctccgactcagGCTTGACACAGGTTTGGAGTG
135	BRCA1_20_F2	ctctctatgggcagtcggtgattCATTCTTGGGATATTCAACACTTA
136	BRCA1_20_R2	ctgcgtgtctccgactcagGCTGAGTGTGTTTCTCAAACAAT
137	BRCA1_21_F1	ctctctatgggcagtcggtgattTTGCTTCCAACCTAGCATCATTA
138	BRCA1_21_R1	ctgcgtgtctccgactcagTAAATTAATTTGTTCTTTCTTTCTTTAT
139	BRCA1_22_F1	ctctctatgggcagtcggtgattTGGGTTATGAAGGACAAAAACAAA
140	BRCA1_22_R1	ctgcgtgtctccgactcagTATTTTCTTTTTCTCCCCCCTA
141	BRCA1_23_F1	ctctctatgggcagtcggtgattGAATCCCAAATTAATACACTCTTG
142	BRCA1_23_R1	ctgcgtgtctccgactcagTCTAATGTGTTAAAGTTCATTGGAA
143	BRCA2_1_F1	ctctctatgggcagtcggtgattTTTGCAGACTTATTTACCAAGCAT
144	BRCA2_1_R1	ctgcgtgtctccgactcagCGGTGTAATTTATAAAGTTATATAAAA
145	BRCA2_2_F1	ctctctatgggcagtcggtgattTGTTCTGGGTCACAAATTTGTCT
146	BRCA2_2_R1	ctgcgtgtctccgactcagGATTCTTCTGCAGGTTCAGAATT
147	BRCA2_2_F2	ctctctatgggcagtcggtgattTTCTTCAGAAGCTCCACCCTAT
148	BRCA2_2_R2	ctgcgtgtctccgactcagGCTCTTTGAATATTATTGGAGTTG
149	BRCA2_2_F3	ctctctatgggcagtcggtgattACCATCTTATAATCAGCTGGCTT
150	BRCA2_2_R3	ctgcgtgtctccgactcagTTCTCCCAGTCTACCATATTG
151	BRCA2_3_F1	ctctctatgggcagtcggtgattTCATTCCCAGTATAGAGGAGAC
152	BRCA2_3_R1	ctgcgtgtctccgactcagGTGCGAAGACTTTTATGTCTACT
153	BRCA2_3_F2	ctctctatgggcagtcggtgattTTTCAGGAAGGAATGTTCCCAAT
154	BRCA2_3_R2	ctgcgtgtctccgactcagTTCATTTAAATATTTTAATATAATAGCTT
155	BRCA2_4_F1	ctctctatgggcagtcggtgattAATAAGATAAACTAGTTTTTGCCAG
156	BRCA2_4_R1	ctgcgtgtctccgactcagATTCTAAGAATAAAAAGCATTGTTTTT
157	BRCA2_5_F1	ctctctatgggcagtcggtgattAATGCTTTTTATTCTTAGAATACTAG
158	BRCA2_5_R1	ctgcgtgtctccgactcagAAGTCATAAAAATAAACCAGGTAGA
159	BRCA2_6_F1	ctctctatgggcagtcggtgattAGTAAACGTTAAGTGAAATAAAGAG
160	BRCA2_6_R1	ctgcgtgtctccgactcagTCCACCTCAGCTCCTAGACTT
161	BRCA2_6_F2	ctctctatgggcagtcggtgattTCAGACACCAAACATATTTCTGA
162	BRCA2_6_R2	ctgcgtgtctccgactcagTCTGCTCTTTCTTGTAATACACA
163	BRCA2_7_F1	ctctctatgggcagtcggtgattTCTGACAAAAAATAAGTTTTTGCATT
164	BRCA2_7_R1	ctgcgtgtctccgactcagTTTCAACAGTCTAATCAATGTCATA
165	BRCA2_8_F1	ctctctatgggcagtcggtgattGATAACTGAAATCACCAAAGTGA
166	BRCA2_8_R1	ctgcgtgtctccgactcagTTCTTCAGACTTTCATCATGATTG
167	BRCA2_8_F2	ctctctatgggcagtcggtgattGCAGAATGTGAAAAGCTATTTTTT
168	BRCA2_8_R2	ctgcgtgtctccgactcagAAACCTGTAGTTCAACTAAACAGA
169	BRCA2_9_F1	ctctctatgggcagtcggtgattTGAGAAAGGTTGTGAGAATAATATA
170	BRCA2_9_R1	ctgcgtgtctccgactcagTGGCATTGACTTTCCAATGTGG
171	BRCA2_9_F2	ctctctatgggcagtcggtgattCATTTAAAGTAAATAGCTGCAAAGA
172	BRCA2_9_R2	ctgcgtgtctccgactcagACTTTTTGTAGATTTTTTGTCTACA
173	BRCA2_9_F3	ctctctatgggcagtcggtgattGATAGTTTTTCATTATGTTTTTCTAAA
174	BRCA2_9_R3	ctgcgtgtctccgactcagAGATACAAATGAGTATTTTTCTTTCA
175	BRCA2_9_F4	ctctctatgggcagtcggtgattGAATGTGAAAAATCTAAAAACCAAG

176	BRCA2_9_R4	ctgcgtgtctccgactcagGTACAACCTTCCTTGGAGATTTTG
177	BRCA2_9_F5	ctctctatgggcagtcggtgattGCCCTTTGAGAGTGGAAAGTGA
178	BRCA2_9_R5	ctgcgtgtctccgactcagTATTTTGGTCACATGAAGAAATATG
179	BRCA2_9_F6	ctctctatgggcagtcggtgattAGATGGAGAAAATACCCCTATTG
180	BRCA2_9_R6	ctgcgtgtctccgactcagTGATTTTGGTAGGCTAGAAATAC
181	BRCA2_9_F7	ctctctatgggcagtcggtgattTACTTCAGAGAATTCTTTGCCAC
182	BRCA2_9_R7	ctgcgtgtctccgactcagATATTGCCTGCTTTACTGCAAGA
183	BRCA2_9_F8	ctctctatgggcagtcggtgattCTTGAATCTCATACTAGACTGCAT
184	BRCA2_9_R8	ctgcgtgtctccgactcagACCTGAAAAACTTGCATTGAAAGT
185	BRCA2_9_F9	ctctctatgggcagtcggtgattAGAATAAGAGAATCACCTAAAGAG
186	BRCA2_9_R9	ctgcgtgtctccgactcagGACATAAGGAGTCCTCCTTCT
187	BRCA2_9_F10	ctctctatgggcagtcggtgattGGAAATACATACTGTTTGCTCAC
188	BRCA2_9_R10	ctgcgtgtctccgactcagACTTATTTGTTTTCTTTTTCAAAGTG
189	BRCA2_9_F11	ctctctatgggcagtcggtgattTTTGAAGAATGCAGGTTTAATATC
190	BRCA2_9_R11	ctgcgtgtctccgactcagCTTCAAACCTGGGCTGAACAGTT
191	BRCA2_9_F12	ctctctatgggcagtcggtgattGAAAGACCAAAAATCAGAACTAATT
192	BRCA2_9_R12	ctgcgtgtctccgactcagGGAATCGTCATCTATAAAACTATAT
193	BRCA2_10_F1	ctctctatgggcagtcggtgattTGAATGTGATTGATGGTACTTTAA
194	BRCA2_10_R1	ctgcgtgtctccgactcagCTCAGAATTGTCCCAAAGAGC
195	BRCA2_10_F2	ctctctatgggcagtcggtgattAAGAACCAACTTTGTCCTTAACTA
196	BRCA2_10_R2	ctgcgtgtctccgactcagTAATAAATAACTGTAGTTTTTCCTTA
197	BRCA2_10_F3	ctctctatgggcagtcggtgattCTTGATTATAAAGAAGCAAAATGTAA
198	BRCA2_10_R3	ctgcgtgtctccgactcagGCCAAGACCTCTTCTTTTATATC
199	BRCA2_10_F4	ctctctatgggcagtcggtgattTGATCCAAAAAGCAAAAAGTTTCA
200	BRCA2_10_R4	ctgcgtgtctccgactcagAGTGCTGGCATTTCATGATCAT
201	BRCA2_10_F5	ctctctatgggcagtcggtgattTTTCAATCCAGAAAAGTCTTTTAT
202	BRCA2_10_R5	ctgcgtgtctccgactcagGTTACCTTTGAGCTTGTCTGAC
203	BRCA2_10_F6	ctctctatgggcagtcggtgattTAGAGGCAAAGAATCATACAAAAT
204	BRCA2_10_R6	ctgcgtgtctccgactcagGGCAACAGCTCAACGTTTTTATA
205	BRCA2_10_F7	ctctctatgggcagtcggtgattAAGATGTATGTGCTTTAAATGAAAAT
206	BRCA2_10_R7	ctgcgtgtctccgactcagTGAAGTAGTTTCTTCTTGATTTTTTT
207	BRCA2_10_F8	ctctctatgggcagtcggtgattAAAACACAAATCTAAGAGTAATCC
208	BRCA2_10_R8	ctgcgtgtctccgactcagGCAAGATTATTCCTTTCATTAGCT
209	BRCA2_10_F9	ctctctatgggcagtcggtgattATGAGAATAATTTTGTCTTCCAAGT
210	BRCA2_10_R9	ctgcgtgtctccgactcagTCACCTGTGTCTCCATATAAAAC
211	BRCA2_10_F10	ctctctatgggcagtcggtgattCCATTTTCAAGAACTCTACCATG
212	BRCA2_10_R10	ctgcgtgtctccgactcagGACCTAGAGTCATTTTTATATGCT
213	BRCA2_10_F11	ctctctatgggcagtcggtgattAGGAGAACAATAATAGTGTAAGC
214	BRCA2_10_R11	ctgcgtgtctccgactcagTTGAAATTGGACCTAAGAGTCTT
215	BRCA2_10_F12	ctctctatgggcagtcggtgattTAATGATTACATGAACAAATGGGC
216	BRCA2_10_R12	ctgcgtgtctccgactcagTCTTCAATATCTTTGAAGAACATTTT
217	BRCA2_10_F13	ctctctatgggcagtcggtgattTCTGAACATAACATTAAGAAGAGC
218	BRCA2_10_R13	ctgcgtgtctccgactcagATACAGTATTAATTGACTGAGGCT
219	BRCA2_10_F14	ctctctatgggcagtcggtgattTTAGATAATCAAAGAACTGAGCA
220	BRCA2_10_R14	ctgcgtgtctccgactcagATTATGGTTTGAATTAATAATCCTGC
221	BRCA2_10_F15	ctctctatgggcagtcggtgattACCCCTCAGATGTTATTTTCCAA

222	BRCA2_10_R15	ctgcgtgtctccgactcagTAGCTTGGTTTTCTAAACTGAGT
223	BRCA2_10_F16	ctctctatgggcagtcggtgattGAATCAGGAAGTCAGTTTGAATTT
224	BRCA2_10_R16	ctgcgtgtctccgactcagTCATTATGACATGAAGATCAGCAT
225	BRCA2_10_F17	ctctctatgggcagtcggtgattCCACTTCTGAGGAATGCAGAG
226	BRCA2_10_R17	ctgcgtgtctccgactcagTTGTTACAGTCATTTTTCAACAGG
227	BRCA2_10_F18	ctctctatgggcagtcggtgattGAAATTAACGGAAGTTTGCTGG
228	BRCA2_10_R18	ctgcgtgtctccgactcagTGCAGAGCTTCAGTAGAAACATT
229	BRCA2_10_F19	ctctctatgggcagtcggtgattTTCTGCTCATGGCACAAAACCTG
230	BRCA2_10_R19	ctgcgtgtctccgactcagCAGAATCATGACATTTACTTGAAG
231	BRCA2_10_F20	ctctctatgggcagtcggtgattCAGAGGTACATCCAATAAGTTTAT
232	BRCA2_10_R20	ctgcgtgtctccgactcagCAGTAGTCATTTCAATATTATTTTGT
233	BRCA2_10_F21	ctctctatgggcagtcggtgattAAAAAATAATAAATGCCAACTGATATT
234	BRCA2_10_R21	ctgcgtgtctccgactcagAATTCTAAGTTATGAGAATTTCTACT
235	BRCA2_10_F22	ctctctatgggcagtcggtgattAAGATAACAAATATACTGCTGCC
236	BRCA2_10_R22	ctgcgtgtctccgactcagGATAATTTAAGACATATGTTGTGCT
237	BRCA2_10_F23	ctctctatgggcagtcggtgattAACGGACTTGCTATTTACTGATC
238	BRCA2_10_R23	ctgcgtgtctccgactcagACCATGACATGCTTCTTGAGCT
239	BRCA2_10_F24	ctctctatgggcagtcggtgattATTTAACTTTTTTGGGAAGTTGCGAA
240	BRCA2_10_R24	ctgcgtgtctccgactcagTTTTTCCCCTTGCACTGCTGAAA
241	BRCA2_10_F25	ctctctatgggcagtcggtgattAGATTTTGAGACTTCTGATACATTT
242	BRCA2_10_R25	ctgcgtgtctccgactcagTCAGAATGTAATTCAGAATTTAAGG
243	BRCA2_10_F26	ctctctatgggcagtcggtgattAAACCAGAAGAATTGCATAACTTTT
244	BRCA2_10_R26	ctgcgtgtctccgactcagATTTCCAGTACCAACTGGGACA
245	BRCA2_10_F27	ctctctatgggcagtcggtgattTTAAACACAAAATACTGAAAGAAAG
246	BRCA2_10_R27	ctgcgtgtctccgactcagTGCAATTTTAACTTTTTTCCCGCT
247	BRCA2_10_F28	ctctctatgggcagtcggtgattCTCTATTGGGTTTTTCATACAGCT
248	BRCA2_10_R28	ctgcgtgtctccgactcagTTTAGGGTCTTTGCCCATTGAT
249	BRCA2_10_F29	ctctctatgggcagtcggtgattTAGTGAAATCACCAGTTTTAGCC
250	BRCA2_10_R29	ctgcgtgtctccgactcagTTTATCATTATTGAGAGAATTCTGC
251	BRCA2_10_F30	ctctctatgggcagtcggtgattCTGCCCAAAGTGTAAGAAAT
252	BRCA2_10_R30	ctgcgtgtctccgactcagTTTCAAAAAGATACTTTTTTGATGTTTT
253	BRCA2_10_F31	ctctctatgggcagtcggtgattTTTATGTAGACAACTGAAAATCTC
254	BRCA2_10_R31	ctgcgtgtctccgactcagAGGCTGAATTTTCAATGACTGAAT
255	BRCA2_10_F32	ctctctatgggcagtcggtgattTTGTTACACAAATCAGTCCCCTT
256	BRCA2_10_R32	ctgcgtgtctccgactcagGTTGACCATCAAATATTCCTTCT
257	BRCA2_10_F33	ctctctatgggcagtcggtgattACTTGAAGCAAAAAAATGGCTTAG
258	BRCA2_10_R33	ctgcgtgtctccgactcagTTTTTCGGAGAGATGATTTTTGTC
259	BRCA2_10_F34	ctctctatgggcagtcggtgattTTCAAACAGTACTATAGCTGAAAAT
260	BRCA2_10_R34	ctgcgtgtctccgactcagATCAAGTTTATTTTTTGAGAGATATC
261	BRCA2_10_F35	ctctctatgggcagtcggtgattCTGATGAGGTATATAATGATTCAG
262	BRCA2_10_R35	ctgcgtgtctccgactcagGTGGGTATGCATTTGCATCTTTT
263	BRCA2_10_F36	ctctctatgggcagtcggtgattGTTTTTCCAAAGTAATATCCAATGT
264	BRCA2_10_R36	ctgcgtgtctccgactcagATTACTATTAGATATGGACAATTT
265	BRCA2_10_F37	ctctctatgggcagtcggtgattTGCAAAAATAAAAATGCAGCCATT
266	BRCA2_10_R37	ctgcgtgtctccgactcagTGTCTGTAAATATGCTTTTCACTT
267	BRCA2_10_F38	ctctctatgggcagtcggtgattTTGTGTTTACATGAAACAATAAAA

268	BRCA2_10_R38	ctgcgtgtctccgactcagTCTGAATCATCCAATGCCTCG
269	BRCA2_10_F39	ctctctatgggcagtcggtgattAACGAAAATTATGGCAGGTTGTTA
270	BRCA2_10_R39	ctgcgtgtctccgactcagATATTTTGGTTATGTTGTAAAATTTC
271	BRCA2_10_F40	ctctctatgggcagtcggtgattTTTTTGCTGACATTCAGAGTGAA
272	BRCA2_10_R40	ctgcgtgtctccgactcagTGACTTATGAAGCTTCCCTATAC
273	BRCA2_10_F41	ctctctatgggcagtcggtgattGAAACTTCAGATATATGTAAATGTA
274	BRCA2_10_R41	ctgcgtgtctccgactcagCAGAAAACACTTGTCTTGCCTTT
275	BRCA2_10_F42	ctctctatgggcagtcggtgattCAGGTATCAGATGCTTCATTACA
276	BRCA2_10_R42	ctgcgtgtctccgactcagAGTACGTATAGCAGTATTTTCTTC
277	BRCA2_10_F43	ctctctatgggcagtcggtgattACATTCAGACCAGCTCACAAGA
278	BRCA2_10_R43	ctgcgtgtctccgactcagAATGGAACTTGCTTTCCACTTG
279	BRCA2_10_F44	ctctctatgggcagtcggtgattTGCTTTCTCTGGATTTAGTACAG
280	BRCA2_10_R44	ctgcgtgtctccgactcagCATTTTGTCTAGACGTAGGTGAA
281	BRCA2_10_F45	ctctctatgggcagtcggtgattAGAACTGAGCATAGTCTTCACTA
282	BRCA2_10_R45	ctgcgtgtctccgactcagTTTAAGTTATTTGATAATTTAAATTCTTT
283	BRCA2_10_F46	ctctctatgggcagtcggtgattCAGAAATGGAAAAACCTGCAGT
284	BRCA2_10_R46	ctgcgtgtctccgactcagTTCCTAATAACCAACTGTTGTTTGT
285	BRCA2_10_F47	ctctctatgggcagtcggtgattCATATCTCTCTCAATTTCAACAAG
286	BRCA2_10_R47	ctgcgtgtctccgactcagATCAGAAAAAGTTTCAGTTTTACCA
287	BRCA2_10_F48	ctctctatgggcagtcggtgattCACCTAAAAACGTAAAAATGGAAAT
288	BRCA2_10_R48	ctgcgtgtctccgactcagCATAAAAGCTTTAGCAATTTCTAC
289	BRCA2_10_F49	ctctctatgggcagtcggtgattGAAAACACTTTGAAACAGAAGCA
290	BRCA2_10_R49	ctgcgtgtctccgactcagAATTCTTGAATTTGACAAAACCATTT
291	BRCA2_10_F50	ctctctatgggcagtcggtgattTTTTACATGTCCCGAAAATGAGG
292	BRCA2_10_R50	ctgcgtgtctccgactcagAAGTCTACTGAATAAACACTTTAAAA
293	BRCA2_11_F1	ctctctatgggcagtcggtgattAAGTATTTTTGTTTAAACATTTAAAGAG
294	BRCA2_11_R1	ctgcgtgtctccgactcagTAATAAGTTTCTTTTGGATTGAGGGT
295	BRCA2_11_F2	ctctctatgggcagtcggtgattATATATGAAATATTTCTTTTTAGGAGA
296	BRCA2_11_R2	ctgcgtgtctccgactcagGGAGAACAGATATAAATAAAAAGCT
297	BRCA2_12_F1	ctctctatgggcagtcggtgattATTCTCTTAGATTTTAACTAATATGTA
298	BRCA2_12_R1	ctgcgtgtctccgactcagTGTATTAGAATTTAGAAAAATTTAAACA
299	BRCA2_13_F1	ctctctatgggcagtcggtgattGGGTCTGCAACAAAGGCATATT
300	BRCA2_13_R1	ctgcgtgtctccgactcagACAGAAATTCTTGACCAGGTGC
301	BRCA2_13_F2	ctctctatgggcagtcggtgattGAGATACAGAATCCAAATTTTACC
302	BRCA2_13_R2	ctgcgtgtctccgactcagTTTTCAATTTCTTGTAGCAGAACTT
303	BRCA2_13_F3	ctctctatgggcagtcggtgattAGTTTCAGGACATCCATTTTATC
304	BRCA2_13_R3	ctgcgtgtctccgactcagCCTAACACACTGTTCAACTCTG
305	BRCA2_13_F4	ctctctatgggcagtcggtgattCCTTTTAAAACATAATCACATTTTCA
306	BRCA2_13_R4	ctgcgtgtctccgactcagGTTAAACTGATGAATCTCATTGTC
307	BRCA2_13_F5	ctctctatgggcagtcggtgattTCTGATGATAGTAAAAATAAGATTAAT
308	BRCA2_13_R5	ctgcgtgtctccgactcagGAAAGGCAAAAATTCATCACACAA
309	BRCA2_14_F1	ctctctatgggcagtcggtgattTTCAATTTTATTTTTGCTAAGTATTTAT
310	BRCA2_14_R1	ctgcgtgtctccgactcagGGATGTTTTTGGCAAGATACAGAC
311	BRCA2_14_F2	ctctctatgggcagtcggtgattGCGTCTTTCCACAGCCAGGCA
312	BRCA2_14_R2	ctgcgtgtctccgactcagAAAAGCCATCAGTATTGTAGACAA
313	BRCA2_15_F1	ctctctatgggcagtcggtgattGAAGATTCTAGTAGTTAATGAAAATT

314	BRCA2_15_R1	ctgcgtgtctccgactcagTTAGAAACGCCATACGTATACAG
315	BRCA2_15_F2	ctctctatgggcagtcggtgattTTTTGTGTGTGTTTATTTTGTGTAG
316	BRCA2_15_R2	ctgcgtgtctccgactcagCTGTATTCCTTTTCCAGTCCATA
317	BRCA2_15_F3	ctctctatgggcagtcggtgattGAAGATTATTTTGGTAAGGAAAGTT
318	BRCA2_15_R3	ctgcgtgtctccgactcagACATAAAAGTTAACACACAATCTTTT
319	BRCA2_16_F1	ctctctatgggcagtcggtgattCATCCTATGTGGTTTTTATGATAAT
320	BRCA2_16_R1	ctgcgtgtctccgactcagATTCCATAGCTGCCAGTTTCCA
321	BRCA2_16_F2	ctctctatgggcagtcggtgattTTTATAATCACTATAGATGGATCATA
322	BRCA2_16_R2	ctgcgtgtctccgactcagTATATGATTACGTAATGTAATGCTTT
323	BRCA2_17_F1	ctctctatgggcagtcggtgattCAGTGGAATTCTAGAGTCACAC
324	BRCA2_17_R1	ctgcgtgtctccgactcagAGAGAACAAGTGTTTTTGCAGCT
325	BRCA2_17_F2	ctctctatgggcagtcggtgattAAGATAATGGAAAGGGATGACAC
326	BRCA2_17_R2	ctgcgtgtctccgactcagTGTAAGTTCAATAATGGCCACTTT
327	BRCA2_17_F3	ctctctatgggcagtcggtgattAACTAGTAGTGCAGATACCCAA
328	BRCA2_17_R3	ctgcgtgtctccgactcagTCCATGAAGAATAATCTTCTGAC
329	BRCA2_17_F4	ctctctatgggcagtcggtgattAAGAATGGCAGACTGACAGTTG
330	BRCA2_17_R4	ctgcgtgtctccgactcagGACTGATTTTTACCAAGAGTGCA
331	BRCA2_18_F1	ctctctatgggcagtcggtgattATCTGTAATAGAATTGAATACATATTT
332	BRCA2_18_R1	ctgcgtgtctccgactcagTGATAAGGGCAGAGGAAAAGGT
333	BRCA2_18_F2	ctctctatgggcagtcggtgattTTGGATTCTTTCCTGACCCTAG
334	BRCA2_18_R2	ctgcgtgtctccgactcagAGAAATATATGGTAAGTTTCAAGAAT
335	BRCA2_19_F1	ctctctatgggcagtcggtgattTGGCCTGATACAATTAAGTTGAAT
336	BRCA2_19_R1	ctgcgtgtctccgactcagCCACATATTTTGCTGCTTCTTT
337	BRCA2_19_F2	ctctctatgggcagtcggtgattTTTCGCAATGAAAGAGAGGAAGA
338	BRCA2_19_R2	ctgcgtgtctccgactcagTTAGAAATAACAATGTGTACCATATA
339	BRCA2_20_F1	ctctctatgggcagtcggtgattTTCTTTGGGTGTTTTATGCTTGG
340	BRCA2_20_R1	ctgcgtgtctccgactcagCACGAACTTGCTGTCTTGTAG
341	BRCA2_20_F2	ctctctatgggcagtcggtgattAACCATATTTACCATCACGTGCA
342	BRCA2_20_R2	ctgcgtgtctccgactcagCATCAAGCCTCATTATATGTCTT
343	BRCA2_21_F1	ctctctatgggcagtcggtgattTAAGATGAGCTCTAATTTTGTGTGA
344	BRCA2_21_R1	ctgcgtgtctccgactcagATTCAAGGCTCTTAACTGCTCTT
345	BRCA2_21_F2	ctctctatgggcagtcggtgattTGGTCACAGGGTTATTTTCAAGTG
346	BRCA2_21_R2	ctgcgtgtctccgactcagCTTGATAAACCTTGTTCTTTTGT
347	BRCA2_21_F3	ctctctatgggcagtcggtgattGAAGGCCATGGAATCTGCTGA
348	BRCA2_21_R3	ctgcgtgtctccgactcagGTTAATAAACTGATAAAAACAAAGC
349	BRCA2_22_F1	ctctctatgggcagtcggtgattATTTAAATGATAATCACTTCTTCCAT
350	BRCA2_22_R1	ctgcgtgtctccgactcagTTAGATTTTGAAGTTGCAAGATGAT
351	BRCA2_22_F2	ctctctatgggcagtcggtgattCAGAAGGAAAGAGATACAGAATTT
352	BRCA2_22_R2	ctgcgtgtctccgactcagTCAAACTGAAAAATTACAATGAAAG
353	BRCA2_23_F1	ctctctatgggcagtcggtgattTGATAAGTGCTTGTTAGTTTATGG
354	BRCA2_23_R1	ctgcgtgtctccgactcagATGGCTGAAAGTCTGGATCTAAA
355	BRCA2_23_F2	ctctctatgggcagtcggtgattAGCCCCTTCACTTCAGCAAATT
356	BRCA2_23_R2	ctgcgtgtctccgactcagCAATGTTTTTTAAAAGAATCAATAAAAA
357	BRCA2_24_F1	ctctctatgggcagtcggtgattCATCTTAAAATTCATCTAACACATC
358	BRCA2_24_R1	ctgcgtgtctccgactcagCTTAATAATGTCCTCATTAAGGTC
359	BRCA2_24_F2	ctctctatgggcagtcggtgattTTTACTGGCAATAAAGTTTTGGATA

360	BRCA2_24_R2	ctgcgtgtctccgactcagTAGCAGAAAACACAGAAAAATCTC
361	BRCA2_24_F3	ctctctatgggcagtcggtgattAGGCCTTCTTACTTTATTTGCTG
362	BRCA2_24_R3	ctgcgtgtctccgactcagAGAAAAATACCAAAATGTGTGGTG
363	BRCA2_25_F1	ctctctatgggcagtcggtgattAAACATAAATATGTGGGTTTGCAAT
364	BRCA2_25_R1	ctgcgtgtctccgactcagTCTTTAGTTGGGGTGGACCAC
365	BRCA2_25_F2	ctctctatgggcagtcggtgattATACTGCATGCAAATGATCCCAA
366	BRCA2_25_R2	ctgcgtgtctccgactcagCTCCATATATACTTCTTATAATATTC
367	BRCA2_26_F1	ctctctatgggcagtcggtgattGTGTGTAATATTTGCGTGTCTAAA
368	BRCA2_26_R1	ctgcgtgtctccgactcagATAAAGGACTTTGATAATATATCTCA
369	BRCA2_26_F2	ctctctatgggcagtcggtgattTTATCAGATGTCTTCTCCTAATTG
370	BRCA2_26_R2	ctgcgtgtctccgactcagTTTTGGTCATCAATCTCTTTCTC
371	BRCA2_26_F3	ctctctatgggcagtcggtgattTGACTTCAAAGTCTTGTAAGGG
372	BRCA2_26_R3	ctgcgtgtctccgactcagTCTGTGCAGCCGGAGAAACAA
373	BRCA2_26_F4	ctctctatgggcagtcggtgattACCTGTTAGTCCCATTTGTACAT
374	BRCA2_26_R4	ctgcgtgtctccgactcagTTTCATTGAATTTTTTAAATGGAGTC
375	BRCA2_26_F5	ctctctatgggcagtcggtgattAAAAGAACTGAATTCTCCTCAGAT
376	BRCA2_26_R5	ctgcgtgtctccgactcagATATAAATTGTTTTTCTCCTGTTGA
377	BRCA2_26_F6	ctctctatgggcagtcggtgattTACCCAAGCTCTTTTGTCTGGT
378	BRCA2_26_R6	ctgcgtgtctccgactcagCTCTCCTGTTCTTTGATCAGAG
379	BRCA2_26_F7	ctctctatgggcagtcggtgattACTGAAACGACGTTGTACTACAT
380	BRCA2_26_R7	ctgcgtgtctccgactcagCGTCAATAATTTATTGTCGCCTTT