

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

АВТОРЕФЕРАТ

диссертации на соискание ученой степени  
кандидата биологических наук

НОВОСИБИРСК – 2018

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук (ИХБФМ СО РАН).

Научный руководитель: **Филипенко Максим Леонидович**, к.б.н.,  
Федеральное государственное бюджетное учреждение науки Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, заведующий лабораторией фармакогеномики

Официальные оппоненты: **Макарова Светлана Ивановна**, д.б.н.,  
Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт молекулярной биологии и биофизики», старший научный сотрудник

**Максимов Даниил Александрович**, к.б.н.,  
Федеральное государственное бюджетное учреждение науки Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, научный сотрудник

Ведущая организация: Федеральное государственное бюджетное научное учреждение «Томский национальный исследовательский медицинский центр Российской академии наук»

Защита состоится «23» марта 2018 г. В 10<sup>00</sup> часов на заседании диссертационного совета Д 003.045.01 при ФГБУН Институте химической биологии и фундаментальной медицины СО РАН по адресу 630090, г. Новосибирск, проспект акад. Лаврентьева, 8.

С диссертацией можно ознакомиться в библиотеке Института химической биологии и фундаментальной медицины СО РАН и на сайте по адресу: <http://www.niboch.nsc.ru/>.

Автореферат разослан «    » \_\_\_\_ 201\_ г.

Ученый секретарь  
диссертационного совета  
кандидат химических наук

Коваль В. В.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы.** 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). Одними из основных факторов, предрасполагающих к развитию данных заболеваний являются мутации в генах *BRCA1* и *BRCA2*, повышающие риск возникновения опухоли до 80 лет до 72% для РМЖ и до 44% для РЯ. 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 встречаемости («hotspot»-мутации) делает эту процедуру слишком дорогой. Использование массового параллельного секвенирования (MPS или NGS) позволяет ускорить, упростить и значительно снизить стоимость исследования данных генов. В связи с этим, разработка эффективных методов исследования генов *BRCA1/2* на основе NGS и их совершенствование являются актуальными задачами. Кроме того, применение NGS позволяет получить более полную и репрезентативную картину распространенности и спектра мутаций *BRCA1/2* у пациенток с РЯ, проживающих в Российской Федерации.

**Целью 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.

**Научная новизна 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.

**Практическая значимость.** 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 таких 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.

**Основные положения диссертации, выносимые на защиту.** Разработанный метод выявления клинически значимых вариантов в генах *BRCA1* и *BRCA2* с использованием NGS позволяет сократить число процедур приготовления библиотеки по сравнению с коммерчески доступным набором реагентов Qiagen в три раза и получить минимально допустимое покрытие всех экзонов генов *BRCA1/2* (более 20 прочтений на позицию при медианном покрытии более 300 прочтений) при исследовании 96 образцов и секвенировании библиотеки на MiSeq Illumina. Использование разработанной программы cutPrimers приводит к увеличению покрытия целевых фрагментов минимум на 8% и до 100% по сравнению с некоторыми программами сравнения. Программа BRCA-analyzer имеет специфичность 100% для клинически значимых вариантов и более высокую чувствительность (100% против 81%) по сравнению с коммерчески доступным веб-сервисом. Программа по выявлению CNV в генах *BRCA1/2* имеет чувствительность 100% и специфичность 82,7%, что позволяет использовать данный метод для тестирования пациентов на CNV. Суммарная частота встречаемости пациентов с клинически значимыми герминальными вариантами в генах *BRCA1* и *BRCA2* среди больных РМЖ, имеющих указания на наследственный характер заболевания, составляет 14,4%, из которых 11,3% – пациенты с клинически значимыми вариантами с повышенной частотой в популяции (5382insC, 4154del4, C61G и др). 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 составляет 30,3%, из которых 15,4% – пациенты с клинически значимыми герминальными вариантами с повышенной частотой в

популяции (5382insC, 4154del4, C61G и другие). Потеря гетерозиготности наблюдается в 70% всех случаев РЯ, среди несущих герминальный клинически значимый вариант – в 83%. Наличие потери гетерозиготности не может являться достаточным критерием для классификации вариантов по клиническому значению. Основным механизмом возникновения потери гетерозиготности, по-видимому, является генная конверсия. Формирование ЛОН через CNV наблюдается только в 14% случаев.

**Личный 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 Тромэншлегер И.Н. Образцы ДНК из крови были выделены Задорожным А.В. Разработка программ, анализ данных NGS и статистический анализ полученных данных сделаны лично автором.

**Апробация работы.** Результаты выполненной работы обсуждались на семинарах Лаборатории фармакогеномики ИХБФМ СО РАН. Отдельные части работы были доложены на международных конференциях: Международный биотехнологический конгресс (Москва, 20–22 февраля 2017 г.), IX Всероссийская научно-практическая конференция с международным участием «Молекулярная диагностика 2017» (Москва, 18–20 апреля 2017 г.), Вторая международная научно-практическая конференция «NGS в медицинской генетике» (Суздаль, 26–28 апреля 2017 г.).

**Публикации.** По материалам диссертации в рецензируемых журналах опубликовано 3 работы, получен 1 патент.

**Объем и структура работы.** Диссертационная работа изложена на 118 страницах и состоит из перечня используемых сокращений, введения, литературного обзора, материалов и методов, результатов и их обсуждений, выводов, списка литературы. Список цитируемой литературы включает 199 наименований. Диссертация содержит 11 таблиц и 22 рисунка. Литературный обзор включает в себя сведения о характеристиках рака молочной железы и рака яичников, строении и функциях белков, кодируемых генами *BRCA1* и *BRCA2*, о методах выявления мутаций

в данных генах и о работах, в которых проводилось выявление мутаций в генах *BRCA1/2*, в том числе на выборках пациентов Российских онкологических центров.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

### Оптимизация методов приготовления библиотеки

При разработке методов приготовления библиотеки нами были опробованы два подхода: с использованием праймеров, используемых при секвенировании генов *BRCA1/2* по Сэнгеру и амплификации каждого целевого фрагмента по-отдельности с последующим пулированием амплификационных смесей; и мультиплексный метод, при котором все ампликоны разделены на четыре мультиплекса и имеют одинаковую длину (150 п.о.). Оценка качества полученных библиотек производилась путем вычисления покрытия после секвенирования на платформе MiSeq для каждого из ампликонов и пациентов с помощью samtools mpileup и собственных Python-скриптов.

Для первого подхода медианное значение покрытия каждого нуклеотида по ампликонам (86 шт.) составило 927,0 прочтений (от 182,0 до 2269,9) (рисунок 1). 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 (ввиду большого размера ампликонов), и трудозатратен, был разработан мультиплексный метод с размером ампликонов не более 150 п.о.

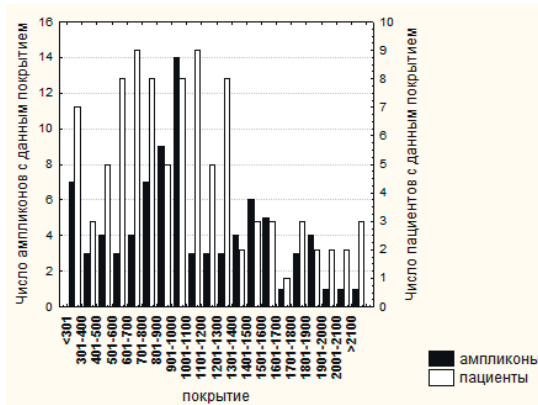


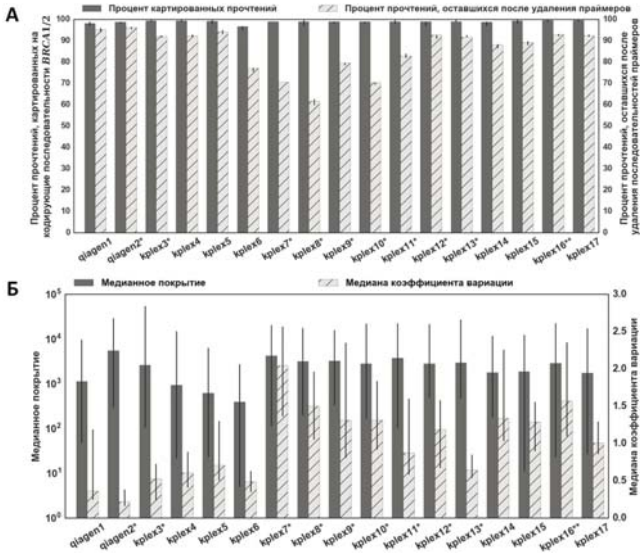
Рисунок 1. Распределение покрытия для ампликонов и пациентов.

Первая версия мультиплексного метода приготовления библиотеки 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 покрытие 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 были покрыты минимум 20 прочтениями, результаты этого запуска не использовались для выявления вариаций.

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 2**). 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 2А**). 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). 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. Таким образом, разработанные методы приготовления библиотек позволяют получить последовательности всех ампликонов с высоким значением покрытия.

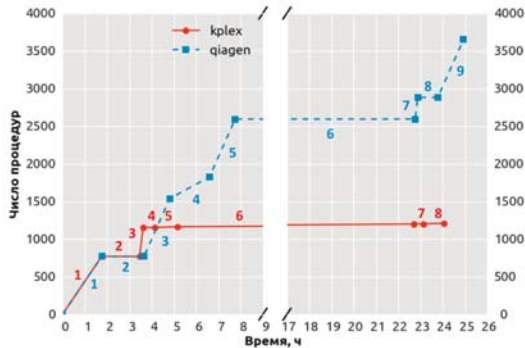




**Рисунок 2.** Сравнение показателей библиотек, приготовленных с помощью коммерческого набора реагентов Qiagen и собственного разработанного метода (обозначены на рисунках как «kplox» с номером запуска). **А.** Сравнение по проценту прочтений, остающихся после удаления последовательностей праймеров и по проценту прочтений, картированных на кодирующие последовательности генов *BRCA1/2* + 100 нуклеотидов рядом **Б.** Сравнение по медиане средних значений покрытия ампликонов и медиане коэффициента вариации покрытия среди пациентов для каждого из ампликонов.

Для разработанного мультиплексного метода приготовления библиотеки и коммерческого набора реагентов были оценены число процедур и время, необходимое для их проведения. 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. 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 3** – это процедура №3). В то же время, для Qiagen индексирующие последовательности вводятся только на этапе №7.



**Рисунок 3.** 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.

### Анализ данных

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). На основе результатов, полученных при анализе данных секвенирования библиотеки на основе праймеров для секвенирования по Сэнгеру, весь алгоритм анализа данных был модифицирован следующим образом:

- Все процедуры обработки были объединены в единую Python-программу (программа BRCA-analyzer <http://github.com/aakechin/brca-analyzer/>), что дало возможность запускать анализ одной командой и в несколько потоков;
- Был введен этап удаления последовательностей праймеров из прочтений с помощью разработанной нами программы cutPrimers, доступной для скачивания с сайта <http://github.com/aakechin/cutprimers/>.

Разработанные нами программы были сравнены с уже существующими. Программа для удаления последовательностей праймеров из прочтений (cutPrimers) была сравнена со следующими аналогичными программами: cutadapt, BBDuk и AlienTrimmer. Сравнение с Trimmomatic – одной из наиболее известных программ для

вырезания технических последовательностей из праймеров – приведено не будет, поскольку с ней не удалось достичь каких-либо приемлемых результатов. 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 FASTQ-файлах одного целого запуска (42 образца, всего 16429892 прочтения). 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 1**.

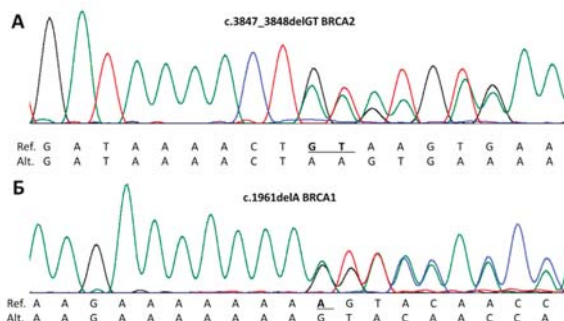
**Таблица 1.** Сравнительный анализ четырех программ для удаления последовательностей из прочтений. Наилучшие варианты по каждому параметру сравнения выделены жирным шрифтом. Якорные символы (^ и \$) позволяют cutadapt искать последовательности праймеров только на самых концах прочтений.  $k=20$  – длина «к-меров» при поиске последовательности праймера составляет 20 букв.  $edist=2$  и  $err=5$  – при поиске последовательности праймера допускается две или пять ошибок (замены, инсерции и/или делеции), соответственно.

Программа →	cutadapt	BBDuk	AlienTrimmer	cutPrimers
Режим →	С якорными символами	8 потоков, $k=20$ , $edist=2$	Парные прочтения	8 потоков, $err=5$
Время обработки, секунды	40823	<b>1109</b>	3254	13739
Число оставшихся прочтений (% от всех)	14345458 (87,3%)	15222691 (92,6%)	2385222 (14,5%)	<b>15231108 (92,7%)</b>
Число покрытых ампликонов (>30)	181,0 (174–181)	183,0 (152–187)	41,5 (7–80)	<b>189,0 (182–189)</b>
Медианное покрытие	1743,0	1387,0	3,0	<b>1885,5</b>

Программы cutadapt и AlienTrimmer оставляли значительно меньше прочтений по сравнению с cutPrimers и BBDuk. 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.

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 TC<sub>3</sub>), два образца (один пациент) с мутацией c.1961delA (последовательность GA<sub>8</sub>) и один образец с мутацией c.3847\_3848delGT (CTGTAAGTGA<sub>7</sub>TAATA<sub>3</sub>). 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 4**). Клинически значимых мутаций, которые были бы выявлены Qiagen и не выявлены BRCA-analyzer, обнаружено не было.



**Рисунок 4.** 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.

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 для 96 образцов больных РМЖ

Для исследования кодирующих последовательностей генов *BRCA1/2* были отобраны 96 образцов 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 клинически значимых вариантов с повышенной частотой в популяции (5382insC, C61G, 4154delA, 185delAG, 2080delA, 3819del5, 3875del4) и *BRCA2* (6174delT, 9318del4, 1528del4). Все найденные SNV, инсерции и делеции были проверены секвенированием по Сэнгеру. Из подтвержденных вариантов были отобраны те, которые является клинически значимыми или вероятно клинически значимыми, а также несинонимичные замены, для которых клиническая значимость неизвестна (отсутствуют в 1000Genomes или частота менее 0,5%, отсутствуют или имеют статус «unknown» в BIC и ClinVar) (таблица 2). Таким образом, из 96 проанализированных образцов ДНК было найдено три клинически значимых варианта (3,1%) и три варианта с неизвестным клиническим значением (3,1%).

**Таблица 2.** SNV, инсерции и делеции, найденные при исследовании кодирующих последовательностей генов *BRCA1/2* у 96 больных РМЖ. BIC/ClinVar – клиническая значимость для мутации, приведенная в BIC или ClinVar. CPC – сдвиг рамки считывания.

Ген	CDS	Тип мутации	Экз.	BIC	ClinVar	1000G	in silico	Число обр
<i>BRCA2</i>	c.223G>C	миссенс	3	unknown	unknown	0,01%	наруш.	1
	c.5659A>G	миссенс	11	-	unknown	-	нейтр.	1
	c.7254_7255delAG	CPC	14	-	-	-	-	1
<i>BRCA1</i>	c.1251T>G	миссенс	10	unknown	unknown	-	наруш.	1
	c.343_344delCC	CPC	6	-	-	-	-	2

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

Для исследования распространенности клинически значимых вариантов в кодирующих последовательностях генов *BRCA1/2* у больных РЯ были отобраны 301

пациент (331 образец, для 30 пациентов ДНК была выделена из крови и из гистологического блока), все из которых имели позитивный ответ на терапию цисплатином. Всего были найдены 53 клинически значимых варианта (40 уникальных, 29 – в гене *BRCA1* и 24 – в *BRCA2*) (таблица 3). Клинически значимые варианты 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 др.). Полученные частоты соответствуют частотам, полученным другими авторами в Германии (23% только герминальных мутаций), Бразилии (17% герминальных среди всех больных) и России (20,9% только «hotspot»-мутаций), но выше, чем в работе китайских коллег (9,1% среди случаев хотя бы с одним фактором риска).

**Таблица 3.** Клинически значимые мутации, найденные среди больных РЯ. 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.

Ген	CDS	Тип мутации	Экз.	ВIC	ClinVar	Число обр/пац *
<b>BRCA2</b>	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	СРС	11	-	Да	1/1
	c.3865_3868delAAAT	СРС	11	-	Да	2/1
	c.4211_4215delCAAAT	СРС	11	-	Да	1/1
	c.5073dupA	СРС	11	-	Да	1/1
	c.5286T>G	нонсенс	11	-	-	2/2
	c.6443_6447delCTATT	СРС	11	-	-	1/1
	c.6468_6469delTC	СРС	11	-	Да	1/1
	c.6475C>T	нонсенс	11	-	Да	1/1
c.6494delT	СРС	11	-	Да	1/1	

	c.6602_6603dupCT	CPC	11	-	-	1/1
	c.6998dupT	CPC	13	-	Да	2/2
	c.8331+1G>A	HCC	18	-	Да	1/1
	c.9253delA	CPC	24	-	-	1/1
<b>BRCA1</b>	c.5444G>A	нонсенс	22	Да	Да	3/2
	c.5347delA	CPC	21	-	-	2/1
	c.5278-1G>C	HCC	19	-	Да	1/1
	c.5266dupC	CPC	19	-	Да	3/3
	c.5152+1G>T	HCC	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	

### 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. Полученные значения клинических показателей метода близки к тем, которые часто выдвигаются к различным тест-системам (порядка 90%), что при дальнейшей доработке метода может привести к его использованию при клиническом тестировании пациентов.

## 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. 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. 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 4). 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*.



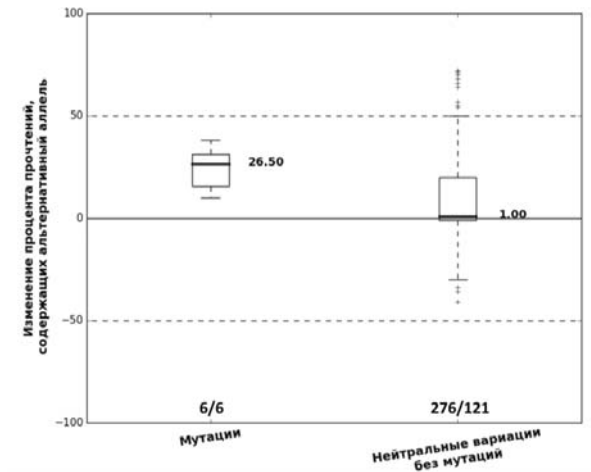
Потеря гетерозиготности наблюдалась у 21 (70%) пациента из 30, из которых пять несли герминальный клинически значимый вариант. У оставшегося одного пациента с клинически значимым вариантом частота альтернативного аллеля для некоторых вариантов также была повышена (максимально – на 19%). Чтобы оценить 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.

**Таблица 4.** 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	ВС (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.

Из группы нейтральных вариантов были исключены те, которые расположены в том же гене (независимо от того, на той же копии хромосомы или другой), что и клинически значимый вариант, у соответствующих пациентов. 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 BIC или обозначены как «unknwn» 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. Сравнение было проведено по абсолютному изменению процента прочтений для всех клинически значимых вариантов и нейтральных вариантов, для которых процент прочтений увеличился (**рисунок 5**), а также по числу вариантов, для которых произошло статистически значимое увеличение процента прочтений.



**Рисунок 5.** Изменение процента прочтений, содержащих аллель клинически значимого варианта (на рисунке обозначены как мутации) или нейтральных вариантов. 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 ( $p$ -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 5**. 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. При сравнении клинически значимых вариантов с нейтральными вариантами статистически значимых отличий найдено не было ( $p$ -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.

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. При их сравнении было обнаружено, что регионы с потерей гетерозиготности перекрываются CNV только в 3 (14%) случаях из 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). 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.

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

**Основные результаты диссертации опубликованы в следующих работах:**

1. 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.
2. 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.
3. 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.
4. 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; приоритет от 11.12.2015 г.
5. Кечин А. А., Боярских У. А., Ермоленко Н. А., Тюляндина А. С., Лазарева Д. Г., Авдальян А. М., Тюляндин С. А., Кушлинский Н. Е., Филипенко М. Л. Потеря гетерозиготности в генах *BRCA1* и *BRCA2* у больных раком яичников и её применимость для клинической классификации вариаций // Бюллетень экспериментальной биологии и медицины. – 2018. – Т. 165. – № 1. – С. 110–117.