



УТВЕРЖДАЮ
Руководитель отдела сертификации
ООО «ТестГен»
Л.М. Халилова
(по доверенности №95701 от 30 декабря 2022)
«14» июля 2023 г.



МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ по применению набора реагентов «Quasar-BRCA1/2»

Оглавление

1. Проведение анализа.....	3
2. Выбор реагентов для секвенатора.....	5
3. Облачная платформа XplainBio	5
4. Технические артефакты.	6
5. Методы измерения концентрации ДНК.	6
6. Разведение ДНК.....	7
7. Измерение концентрации библиотек и пулирование.....	8
8. Проведение капиллярного электрофореза.	8
9. Повторное секвенирование неудачных образцов.....	9
10. Биоинформационная обработка результатов с помощью платформы XplainBio.	9
11. Автоматическая интерпретация образцов с помощью платформы XplainBio.....	10

1. Проведение анализа

Исследование статуса мутаций генов BRCA1 и BRCA2, ассоциированных с риском развития рака молочной железы и раком яичников, набором реагентов «Quasar-BRCA1/2» состоит из следующих этапов:

Схема 1. Этапы работы.

1 Этап: целевая мультиплексная ПЦР-амплификация ДНК



2 Этап: очистка целевых ампликонов



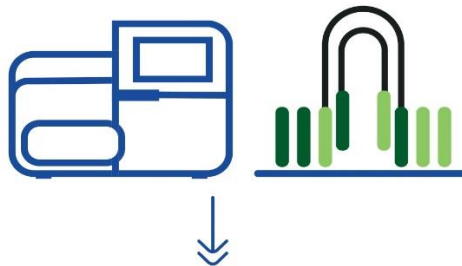
3 Этап: индексная ПЦР с получением ДНК-библиотек



4 Этап: объединение и очистка ДНК-библиотек, подготовка рабочей концентрации пула



5 Этап: секвенирование с применением Illumina MiSeq/NextSeq



6 Этап: интерпретация результатов и выдача заключений

Прочтения

```

ATGGCATTGCAATTTGACAT
TGGCATTGCAATTTG
AGATGGTATTG
ATGGCATTGCAA
ATGGCATTGCAATTTGAC
ATGGCATTGCAATT
AGATGGCATTGCAATTTG
    
```

Референтный геном AGATGGTATTGCAATTTGACAT

Герминальные мутации, Ампликоны	Всего: 25	🔒 🛡️ ✓	08.10.22 19:16	2725x 85% 0%
Герминальные мутации, Ампликоны	Всего: 23	🔒 🛡️ ✓	08.10.22 19:16	3116x 87% 0%
Герминальные мутации, Ампликоны	Всего: 23	🔒 🛡️ ✓	08.10.22 19:16	4083x 84% 0%
Герминальные мутации, Ампликоны	Всего: 22	🔒 🛡️ ✓	08.10.22 19:16	2590x 80% 0%
Герминальные мутации, Ампликоны	Всего: 20	🔒 🛡️ ✓	08.10.22 19:16	4971x 85% 0%

Схема 2. Безопасные точки остановки при работе с набором «Quasar-BRCA1/2»



2. Выбор реагентов для секвенатора.

В зависимости от количества секвенируемых образцов и типа биоматериала, возможно использование различных версий картриджей. Допустимая степень мультиплексирования зависит от требуемого покрытия и общего числа прочтений в запуске. Рекомендуемое число прочтений на образец - 200 тыс. для ДНК из крови и 400 тыс. для ДНК из гистологических блоков.

Секвенирование необходимо проводить в режиме парноконцевых прочтений (PE), минимально допустимая длина прочтений 2x150, для картриджей на 500 и 600 циклов рекомендуется использовать длину прочтений 2x250 или 2x300.

Таблица 1. Расчетные характеристики для ДНК, выделенной из крови.

Набор	Режим	Число прочтений	Прочтений на образец	Образцов на запуск
v2 Nano	2x250 2x150	2 000 000	≈200 000	12
v2 Micro	2x150	8 000 000	≈200 000	48
v2	2x250 2x150	25 000 000	≈200 000	96
v3	2x300 2x150	50 000 000	≈200 000	96-192

Таблица 2. Расчетные характеристики для ДНК, выделенной из FFPE-блоков.

Набор	Режим	Число прочтений	Прочтений на образец	Образцов на запуск
v2 Nano	2x250 2x150	2 000 000	≈400 000	4
v2 Micro	2x150	8 000 000	≈400 000	16
v2	2x250 2x150	25 000 000	≈400 000	48
v3	2x300 2x150	50 000 000	≈400 000	до 64

3. Облачная платформа XplainBio

XplainBio - <https://x-bio.ru/> - безопасная облачная платформа для интерпретации результатов NGS-анализа с автоматизацией всех этапов исследования. Платформа автоматизирует анализ данных NGS-секвенирования с помощью современных биоинформационных алгоритмов по поиску патогенных вариантов генов.

XplainBio превращает сырые данные секвенирования в понятный отчет и не требует знаний биоинформатики. Платформа автоматически обрабатывает геномы, экзомы и таргетные панели.

Использование валидированных международным научным сообществом подходов обеспечивает высокие показатели чувствительности и специфичности проводимого анализа, минимизируя вероятность ошибок. Постоянно обновляемые международные базы данных предоставляют самые актуальные сведения о генетических вариантах. XplainBio использует более 80 различных источников информации для максимально полного описания генов и мутаций.

Платформа XplainBio спроектирована для ускорения алгоритмов обработки геномных данных. Современные решения из мира Big Data позволяют получить результат для десятков образцов менее чем за 30 минут. Облачные вычислительные ресурсы позволяют использовать самые ресурсоемкие и точные алгоритмы анализа, сохраняя высокий уровень производительности - более 3000 образцов в сутки.

Хранение и обработка загруженных файлов производится в российских дата-центрах, соответствующих требованиям о хранении персональных данных 152-ФЗ.

4. Технические артефакты.

Методы секвенирования, включая NGS, имеют ряд технологических недостатков, приводящих к ошибочному определению нуклеотидных последовательностей.

Одной из наиболее частых ошибок является неправильное определение длины гомополимерных фрагментов, что выражается в ложноположительных вставках/делециях одного или нескольких нуклеотидов.

Как правило, аллельный баланс этих ошибок соответствует профилю ошибок ПЦР-амплификации и не превышает 10-15%, что делает их относительно безопасными для герминальных мутаций (кроме случаев мозаицизма), но для соматических вариантов данный тип ошибок представляет проблему, особенно в генах BRCA1/2, где значительная часть истинных мутаций обнаруживается именно в гомополимерных областях.

Технические артефакты характеризуются тем, что крайне редкая по популяционным базам мутация, которая должна встретиться в запуске не более одного раза, присутствует более чем в 10-20% образцов.

Каждая NGS-тест-система имеет свой уникальный набор технических артефактов, который определяется производителем с помощью статистических методов на выборках из нескольких сотен образцов. Список артефактов для набора «Quasar-BRCA1/2» приведен в таблице 3.

Таблица 3. Известные технические артефакты набора реагентов «Quasar-BRCA1/2»

hg38	Ген	кДНК	Белок	Аллельный баланс
chr13:32379885CA>C	BRCA2	NM_000059.3:c.9097delA	p.Thr3033fs	10-15%
chr17:43093569CT>C	BRCA1	NM_007294.4:c.1961delA	p.Lys654fs	5-10%

Интерпретировать данные мутации следует с осторожностью, с учетом соотношения прочтений для каждого из аллелей, а также встречаемости варианта в запуске в целом.

Для облегчения идентификации в интерфейсе платформы XplainBio при анализе образца варианты помечаются соответствующими флагами – HomoP (область гомополимера) и OR/Freq (частота варианта сильно отличается от популяционной).

5. Методы измерения концентрации ДНК.

Для ДНК, выделенной из крови, рекомендуется использовать спектрофлуориметрический метод измерения концентрации с использованием флуориметров Qubit (Thermo Fisher Scientific), или аналогичных.

Для ДНК, выделенной из FFPE-блоков, допустимо использовать спектрофлуориметрический метод измерения концентрации (Qubit, или аналоги). Однако эффективная концентрация ДНК, то есть концентрация фрагментов ДНК, пригодных для амплификации, может значительно отличаться от измеренных этим методом значений. Наиболее частые причины различий – присутствие ингибиторов ПЦР, сильная фрагментация, химическая деградация или модификация ДНК. Если ДНК образца фрагментирована настолько, что средний размер фрагментов меньше, чем длина ампликона ПЦР, эффективность амплификации резко падает. Эффективная концентрация ДНК, выделенной из крови, принимается равной измеренной на Qubit.

Для количественного определения эффективной концентрации ДНК рекомендуется использовать методы измерения, основанные на ПЦР. Например, набор реагентов «Quasar Red» (производство ООО «ТестГен»), который позволяет рассчитать индекс деградации и оценить степень фрагментации ДНК для аналитов двух длин 101 и 224 п.о. Данные длины соответствуют средним длинам ампликонов, используемым для NGS-секвенирования с прочтениями 2x75 и 2x150 соответственно.

6. Разведение ДНК.

Точное измерение концентрации и разведение образца важны для эффективности проведения исследования. Рекомендуется проводить анализ образцов ДНК, выделенных одним методом.

В большинстве случаев концентрация (эффективная) входящей ДНК 1 нг/мкл является достаточной для получения оптимального результата.

Для фрагментированной ДНК, выделенной из парафиновых блоков (FFPE-блоки), рекомендуется использовать измерение эффективной концентрации методом ПЦР (см. раздел 5). Только этот метод может адекватно измерить эффективную концентрацию ДНК!

Тем не менее для таких образцов допустимо оценивать концентрацию спектрофлуориметрическим методом (Qubit). Так как этот метод измерения в данном случае дает лишь оценочный результат, целесообразно увеличить разведение входящей ДНК до 3 нг/мкл для компенсации деградации ДНК. В ряде случаев требуется увеличение входящей концентрации до 5 нг/мкл и более, что может быть установлено экспериментально.

В рамках одного раунда пробоподготовки и запуска секвенатора возможно смешивание образцов разного типа.

Использование ДНК с эффективной концентрацией менее 1 нг/мкл может привести к превалированию неспецифических продуктов мультиплексной амплификации, что негативно скажется на результатах исследования. В этом случае рекомендуется повторное выделение ДНК из биологического материала. В процессе выделения возможно уменьшение объема элюции чистой ДНК для увеличения концентрации в соответствии с рекомендациями производителей наборов для выделения.

При разведении ДНК до необходимой концентрации рекомендуется не использовать аликвоты ДНК менее 1 мкл, чтобы уменьшить ошибки пипетирования.

Для расчета необходимого количества ДНК и деионизованной воды mQ для получения раствора с необходимой входящей концентрацией используйте следующие правила:

ДНК из крови, измеренная на Qubit, или из блоков, измеренная методом ПЦР:

Концентрация менее 1 нг/мкл – требуется повторное выделение.

Концентрация от 1 до 12 нг/мкл:

$$V_{DNA} (\text{объем ДНК}) = \frac{12 \text{ нг}}{\text{концентрация ДНК}^{\text{нг/мкл}}}$$

$$V_{mQ} (\text{объем } mQ) = 12 \text{ мкл} - V_{DNA} (\text{объем ДНК})$$

Концентрация более 12 нг/мкл:

$$V_{DNA} (\text{объем ДНК}) = 1 \text{ мкл}$$

$$V_{mQ} (\text{объем } mQ) = \text{концентрация ДНК}^{\text{нг/мкл}} - 1$$

ДНК из блоков, измеренная на Qubit:

Концентрация менее 1 нг/мкл – требуется повторное выделение.

Концентрация от 1 до 3 нг/мкл – желательно повторное выделение, результат может быть сомнительным. Для расчета воспользуйтесь формулами для ДНК из крови (выше)

Концентрация от 3 до 36 нг/мкл:

$$V_{DNA} (\text{объем ДНК}) = \frac{36 \text{ нг}}{\text{концентрация ДНК } \text{нг/мкл}}$$

$$V_{mQ} (\text{объем } mQ) = 12 \text{ мкл} - V_{DNA} (\text{объем ДНК})$$

Концентрация более 36 нг/мкл:

$$V_{DNA} (\text{объем ДНК}) = 1 \text{ мкл}$$

$$V_{mQ} (\text{объем } mQ) = \frac{\text{концентрация ДНК } \text{нг/мкл}}{3 \text{ нг/мкл}} - 1$$

7. Измерение концентрации библиотек и пулирование.

При работе с ДНК, выделенной из одного типа образцов, равномерность покрытия прочтений NGS-секвенирования между образцами обеспечивается нормализацией концентрации входящей ДНК и протоколом исследования.

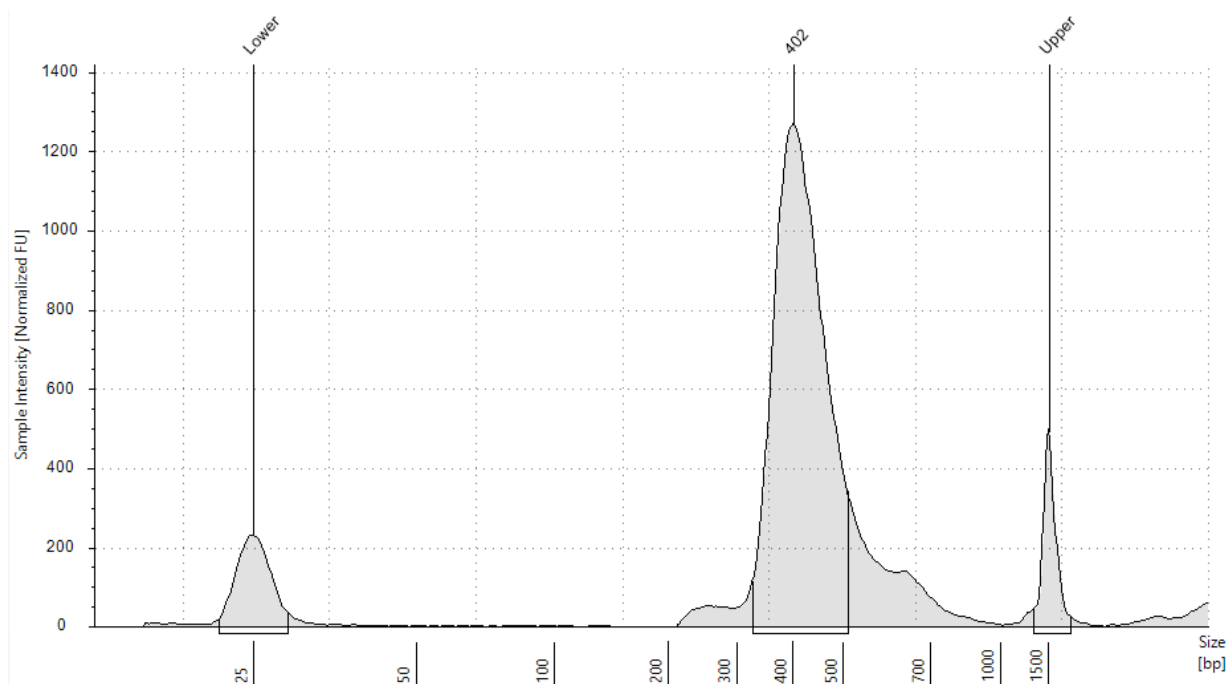
Стандартный протокол исследования оптимизирован для поточного анализа: все образцы после индексного ПЦР смешиваются и очищаются в виде единой смеси. Различия в финальном покрытии между образцами до 2-2.5 раз является нормальными и компенсируются избыточностью необходимого количества данных на образец.

В случае необходимости возможно провести очистку каждой индексированной библиотеки и измерение концентрации индивидуально для последующего смешивания в эквимольных количествах для секвенирования. Такой подход может увеличить равномерность покрытия прочтений между образцами, но является значительно более трудоемким и требующим дополнительных затрат реактивов. Для получения рекомендаций по индивидуальной очистке индексированных библиотек свяжитесь со службой технической поддержки «ТестГен».

8. Проведение капиллярного электрофореза.

Перед секвенированием готовой ДНК-библиотеки рекомендуется всегда проводить оценку ее качества с использованием капиллярного гель-электрофореза. Например, на Agilent TapeStation или BioAnalyzer. При использовании данного оборудования рекомендуется использовать наборы высокой чувствительности (High Sensitivity, HS), анализирующие ДНК в размере длин 35-1000 п.о.

Типичный вид кривой представлен на рисунке ниже. Фракция длиной 300-700 п.о. должна составлять не менее 90% площади кривой (Integrated Area), а фракция длины 100-200 п.о. - не более 5-7%. В случае большей доли фракции длины 100-200 п.о. секвенирование библиотеки не рекомендуется, так как это приведет к превалированию несмысловой ДНК в данных сиквенса (димеров). Для получения рекомендаций по работе с такими библиотеками свяжитесь со службой технической поддержки «ТестГен».



9. Повторное секвенирование неудачных образцов.

При получении низкого покрытия образца, в случае уверенности в качестве входящей ДНК, рекомендуется его повторное секвенирование. В этом случае нет необходимости проведения повторной ПЦР. Секвенирование допустимо производить из готовой библиотеки после индексирования (при наличии) с контролем уникальности используемых индексов, либо проводить только вторую индексную ПЦР очищенного ранее продукта целевой амплификации (см. схему 2).

10. Биоинформационная обработка результатов с помощью платформы XplainBio.

Из-за особенностей метода подготовки библиотек набором «Quasar-BRCA1/2» (праймеры остаются в составе прочтений и должны быть удалены на этапе биоинформационной обработки) рекомендуется использовать веб-портал XplainBio <https://x-bio.ru/>.

В случае необходимости использования собственных подходов следует учитывать следующее:

- Адаптеры должны удаляться с помощью программного обеспечения прибора (MiSeq или NextSeq).
- Необходимо фильтровать прочтения по минимальной длине со значением 100 п.н. – например, в случае использования утилиты Trimmomatic:

```
ILLUMINACLIP:TruSeq3-PE-2.fa:4:30:5:2:keepBothReads LEADING:3
TRAILING:3 MINLEN:100
```

- Необходимо выравнивать прочтения без предварительной отрезки праймеров и выполнять тримминг в BAM-файле:

```
fgbio TrimPrimers \  
-i входной BAM-файл \  
-o выходной BAM-файл \  
-p BED-файл \  
--first-of-pair false \  
--hard-clip true \  
-r референсный геном (fasta)
```

BED-файл с координатами праймеров для работы утилиты fgbio предоставляется по запросу в службу технической поддержки компании.

- Версия генома, используемая в fgbio – hg38/GRCh38.
- После биоинформационной обработки необходимо провести валидацию результатов с помощью сравнения с порталом XplainBio, обращая особое внимание на совпадения границ прочтений после тримминга и параметров QC (особенно равномерность покрытия и количество нуклеотидов, покрытых менее 4x).

- Равномерность вычисляется по формуле $Pct > 0.2 * \text{среднее}$ (доля нуклеотидов с покрытием не менее 20% от среднего).

- Количество нуклеотидов с покрытием менее 4x при правильных параметрах обработки должно совпадать с результатами XplainBio с точностью до одного нуклеотида.

- Рекомендуется использовать несколько алгоритмов поиска вариантов, особенно для соматических мутаций – например, GATK4 HaplotypeCaller/Mutect2, Strelka2 и DRAGEN. В случае применения Mutect2 использование альтернативного алгоритма является обязательным, так как данная утилита имеет тенденцию к пропуску вариантов, присутствующих в популяционных базах данных.

11. Автоматическая интерпретация образцов с помощью платформы XplainBio.

Для таргетных панелей, особенно с малым количеством генов (например, BRCA1/2), при анализе герминальных мутаций часто после фильтрации либо не остается вариантов вообще, либо 2-3 варианта уже классифицированы в локальной базе данных – это позволяет автоматически интерпретировать образцы при выполнении нескольких условий:

- Среднее покрытие не менее граничного значения в настройках панели.
- Равномерность покрытия не менее граничного значения в настройках панели.
- Количество непокрытых нуклеотидов равно 0.

В этом случае образец не нуждается в ручном просмотре и может быть автоматически интерпретирован в зависимости от результатов фильтрации:

- Как отрицательный, если все оставшиеся варианты имеют категорию Benign или Likely Benign (класс патогенности 1 или 2).

- Как положительный, если в списке вариантов есть патогенный вариант из локальной базы данных, обнаруженный всеми тремя инструментами поиска – GATK4, Strelka2 и DRAGEN, не являющийся при этом техническим артефактом (класс патогенности 4 или 5).

При большом количестве классифицированных вариантов в локальной базе мутаций доля автоматически закрытых образцов может достигать 80% от запуска, что значительно снижает время до выдачи результата и ускоряет интерпретацию более сложных оставшихся образцов.