

ОПЫТ СОЗДАНИЯ ПУБЛИЧНОЙ БАЗЫ ДАННЫХ МУТАЦИЙ ONCOBRCA: БИОИНФОРМАЦИОННЫЕ ПРОБЛЕМЫ И РЕШЕНИЯ

А.Г. Никитин¹, О.И. Бровкина¹, Д.С. Ходырев¹, О.А. Гусев^{2,3}, М.Г. Гордиев⁴

¹ Федеральный научно-клинический центр специализированных видов медицинской помощи и медицинских технологий Федерального медико-биологического агентства России, Москва, Российская Федерация

² Казанский (Приволжский) федеральный университет, Казань, Республика Татарстан, Российская Федерация

³ Институт физико-химических исследований RIKEN, Префектура Канагавы, Йокогама, Япония

⁴ Республиканский клинический онкологический диспансер Министерства здравоохранения Республики Татарстан, Казань, Республика Татарстан, Российская Федерация

Обоснование. Развитие наследственных онкологических синдромов обусловлено генетическими нарушениями в системе репарации ДНК, состоящей более чем из 100 генов. Однако в настоящее время в большинстве медицинских центров России диагностика наследственных заболеваний раком яичника и раком молочной железы сводится к определению наиболее частых мутаций (8 точек) в генах BRCA1 и BRCA2 с помощью методов полимеразной цепной реакции. При этом данные мутации являются частыми для славянской популяции, в то время как в остальных популяциях России они встречаются реже или не встречаются вообще. **Цель исследования** — получить представление о ландшафте наследственных патогенных вариантов в генах системы репарации; разработать методы быстрого анализа данных полногеномного секвенирования. **Методы.** Методом секвенирования нового поколения (Next Generation Sequencing, NGS) была проанализирована панель из 34 генов системы репарации в 1644 образцах биоматериала пациентов с наследственными онкологическими синдромами. **Результаты.** Выявлены 119 патогенных мутаций генов BRCA1/BRCA2 у 374 пациентов, при этом обнаружена выраженная разница в частоте встречаемости мутаций между различными этносами — славянами, татарами, башкирами и чувашами. Созданы публичная база данных частоты мутаций и полиморфизмов генов системы репарации ДНК в различных этносах (<https://oncobrca.ru>), а также методы для автоматической параллельной обработки любого количества образцов от сырых данных до готового отчета. **Заключение.** С помощью метода NGS появилась возможность выявления редких мутаций, характерных для различных этносов, что дает возможность назначать оптимальную химиотерапию и разрабатывать диагностические методы популяционного скрининга носительства мутаций генов системы репарации ДНК. Современные подходы к анализу данных полногеномного секвенирования позволили сократить время получения результата до нескольких часов после проведения анализа.

Ключевые слова: онкогенетика, наследственные синдромы, секвенирование, рак.

(Для цитирования: Никитин А.Г., Бровкина О.И., Ходырев Д.С., Гусев О.А., Гордиев М.Г. Опыт создания публичной базы данных мутаций oncoBRCA: биоинформационные проблемы и решения. Клиническая практика. 2020;11(1):In Press. doi: [10.17816/clinpract25860](https://doi.org/10.17816/clinpract25860))

CREATING A PUBLIC MUTATION DATABASE ONCOBRCA: BIOINFORMATIC PROBLEMS AND SOLUTIONS

A.G. Nikitin¹, O.I. Brovkina¹, D.S. Khodyrev¹, O.A. Gusev^{2,3}, M.G. Gordiev⁴

¹ Federal Scientific and Clinical Center of Specialized Types of Medical Care and Medical Technologies of the Federal Medical and Biological Agency of Russia, Moscow, Russian Federation

² Kazan' (Volga region) Federal University, Kazan', Republic of Tatarstan, Russian Federation

³ Institute of Physical and Chemical Research RIKEN, Yokohama, Kanagawa Prefecture, Japan

⁴ Republican Clinical Oncology Center, Kazan', Republic of Tatarstan, Russian Federation

Background. The development of hereditary cancer syndromes is caused by genetic disorders in the DNA repair system, which consists of more than 100 genes. However, at present, in the majority of medical centers in Russia, diagnosis of hereditary OC and BC is to determine the most frequent mutations (8 points) in BRCA1 and BRCA2 genes using PCR methods. Thus the given mutations are frequent for

Slavic population while in other populations of Russia they occur less often or do not occur at all. **Aims.** To reveal the landscape of hereditary pathogenic variants in the genes of reparation system and develop methods of quick analysis of the NGS data. **Methods.** By next generation sequencing (NGS) a panel of 34 genes of the reparation system was analyzed in 1644 samples of patients with inherited cancer syndromes. **Results.** The NGS analysis revealed 119 pathogenic mutations of BRCA1/BRCA2 genes in 374 patients, with a marked difference in mutation frequencies between different ethnic groups — Slavs, Tatars, Bashkirs and Chuvash. A public database of frequencies of mutations and polymorphisms of genes of the DNA reparation system in different ethnoses was created at <https://oncobrca.ru>. Methods were created for automatic parallel processing of any number of samples from raw data to the final report. **Conclusions.** With the NGS method it became possible to identify rare mutations characteristic of different ethnic groups, which makes it possible to prescribe optimal chemotherapy and develop diagnostic methods of population screening of gene mutation carrier of DNA repair system. Modern approaches to the analysis of genome sequencing data allowed reducing the time of obtaining the result to several hours after the analysis.

Keywords: cancer, genetics, heredity, mutation.

(For citation: Nikitin AG, Brovkina OI, Khodyrev DS, Gusev OA, Gordiev MG. Creating a Public Mutation Database oncoBRCA: Bioinformatic Problems and Solutions. *Journal of Clinical Practice*. 2020;11(1):In Press. doi: [10.17816/clinpract25860](https://doi.org/10.17816/clinpract25860))

ОБОСНОВАНИЕ

Рак яичников и рак молочной железы являются основными причинами онкологической смертности среди женщин во всем мире [1]. Оба вида рака высокогетерогенные с сильным наследственным компонентом: примерно 10–15% случаев рака яичников и 5–7% случаев рака молочной железы являются наследственными [2]. Ранняя диагностика снижает смертность, связанную с этими заболеваниями, поэтому генетическое тестирование на предрасположенность к наследственным онкосиндромам будет полезным дополнением к рутинной клинической практике.

Развитие наследственных рака молочной железы и рака яичников обусловлено генетическими нарушениями в системе репарации ДНК [3]. В настоящее время в большинстве медицинских центров Российской Федерации диагностика наследственного рака молочной железы представляет собой определение наиболее частых мутаций (8 точек) в генах *BRCA1*, *BRCA2* с помощью методов полимеразной цепной реакции (ПЦР), хотя в мировой литературе описано более 1000 мутаций гена *BRCA1*, и во многих популяциях присутствует свой собственный набор частых мутаций. Поэтому существующий на данный момент отечественный подход к генетической диагностике рака яичников и рака молочной железы выявляет только мутации, характерные для славянских популяций, населяющих территорию России [4, 5], и не учитывает остальные мутации в генах *BRCA1/BRCA2*, а также

других генах системы репарации, обнаружение которых позволило бы выбрать оптимальную тактику лечения [6]. Население Российской Федерации состоит более чем из ста этнических групп, что дает уникальную возможность изучения наследственных патогенных мутаций.

После внедрения в клиническую практику ингибиторов поли(АДФ-рибоза)-полимеразы (poly ADP ribose polymerase, PARP) открылись новые возможности в лекарственной терапии ряда онкологических заболеваний, обусловленных нарушениями в генах системы репарации двухцепочечных разрывов ДНК, ряд препаратов получил одобрение Управления по санитарному надзору за качеством пищевых продуктов и медикаментов США (Food and Drug Administration, FDA) — нирапариб (Niraparib, Tesaro), рупапариб (Rucaparib, Clovis Oncology) и олапариб (Olaparib, AstraZeneca), но возможность назначения таргетной терапии по результатам генетического анализа предъявляет особые требования к срокам проведения такого исследования: как правило, результат должен быть получен не позднее 14–20 дней с момента назначения, что требует ускорения и автоматизации как лабораторной (подготовка библиотек), так и биоинформационной части (обработка сырых данных, аннотация и интерпретация), так как стандартные сроки выполнения аналогичных анализов на данный момент составляют более 30–40 дней, причем значительное количество времени тратится именно на биоинформационную обработку.

Цель исследования — анализ мутаций в генах системы репарации у пациентов с наследственными онкологическими синдромами; создание публичной базы данных с информацией о встречаемости этих мутаций в различных этнических выборках, что позволит расширить знания о генетических основах рака молочной железы, яичников и других с помощью объединения информации о вариантах генов и соответствующих клинических данных со всей территории России, а также поможет в интерпретации мутаций, встречающихся при анализе российских популяций.

МЕТОДЫ

Дизайн исследования

Пациенты включались в исследование при наличии возможности собрать образцы крови для проведения поиска герминальных мутаций, наличии добровольного информированного согласия об участии в исследовании.

Пациенты исключались из исследования при наличии предшествующей или настоящей системной противоопухолевой или лучевой терапии по поводу выявленного РМЖ, любом доказательстве наличия тяжелой или неконтролируемой системной патологии, активной инфекции, активного геморрагического диатеза или почечного трансплантата, в том числе пациенты с известным гепатитом В, гепатитом С или вирусом иммунодефицита человека (HIV).

В исследуемую группу включались больные РМЖ или РЯ с наличием хотя бы одного из клинических критериев (отягощенный семейный анамнез по злокачественным новообразованиям (случаи РМЖ, РЯ, РПЖ, рака поджелудочной железы у кровных родственников), отягощенный собственный анамнез (наличие в прошлом или настоящем РМЖ контрлатеральной молочной железы, РЯ, рака поджелудочной железы), возраст младше 50 лет, ТНРМЖ).

У пациентов при первичном визите был собран семейный и собственный анамнез, на основании которого пациент включался в данное исследование. Для молекулярно-генетического тестирования производился забор венозной крови в количестве 4–6 мл.

Критерии соответствия

Работа выполнена в соответствии с требованиями международного стандарта этических норм и качества научных исследований GCP (Good

Clinical Practice — *надлежащая клиническая практика*) и Хельсинкской декларации по защите прав человека.

Условия проведения

Исследование включало 1644 образца биоматериала от пациентов с наследственным раком, проходивших обследование и лечение в Республиканском клиническом онкологическом диспансере Министерства здравоохранения Республики Татарстан (Казань) в 2014–2016 гг.

Методы регистрации исходов

Гены, включенные в панель: *ATM, BARD1, BRCA1, BRCA2, CDH1, CDK4, CDK12, CDKN2A, CFTR, CHEK1, CHEK2, CTNNA1, EPCAM, FANCI, FANCL, FANCL, MLH1, MSH2, MSH6, MUTYH, PALB2, PARP1, PDGFRA, PMS2, PPP2R2A, PRSS1, RAD51B, RAD51C, RAD51D, RAD54L, SPINK1, STK11, TP53, XRCC3*.

ДНК из цельной периферической крови выделялась с помощью набора QIAamp DNA Blood Mini Kit (Qiagen, Германия) на автоматической станции QIAcube (Qiagen). Концентрация ДНК измерялась на спектрофотометре NanoVue Plus (GE Healthcare, США) и составляла 30–50 нг/мкл. Подготовка библиотек для секвенирования осуществлялась с помощью NimbleGen SepCapEZ Choice (Roche, Швейцария) по протоколу, рекомендованному производителем. Секвенирование проводилось на приборе MiSeq (Illumina, США). Картирование прочтений на референсную последовательность генома человека (hg19) проводилось при помощи алгоритма BWA-MEM, качество исходных данных, выравнивания, обогащения и покрытия целевых регионов проверялось с помощью FastQC, BAMQC и NGSrich. Среднее покрытие составило 274x, доля корректно картированных прочтений — 99,1%, доля целевых регионов с покрытием выше 100x — 92,8%.

Дедупликация, рекалибровка и поиск нуклеотидных вариаций выполнялись с помощью GATK4+Strelka2, полученный VCF-файл обрабатывался с помощью программы SnpSift (глубина прочтения более 10) и аннотировался с помощью SnpEff (анализ всех транскриптов), ANNOVAR (анализ частот аллелей в ExAC, 1000G и ESP6500, алгоритмы проверки функциональной значимости SIFT, PolyPhen2, MutationTaster, FATMM, CADD, DANN, M-CAP, REVEL), баз данных dbSNP, ClinVar, HGMD Professional 2019.4, BRCA Exchange (блок-схема представлена на [рис. 1](#)).

Рис. 1. Схема универсального пайплайна для версий генома hg19/hg38

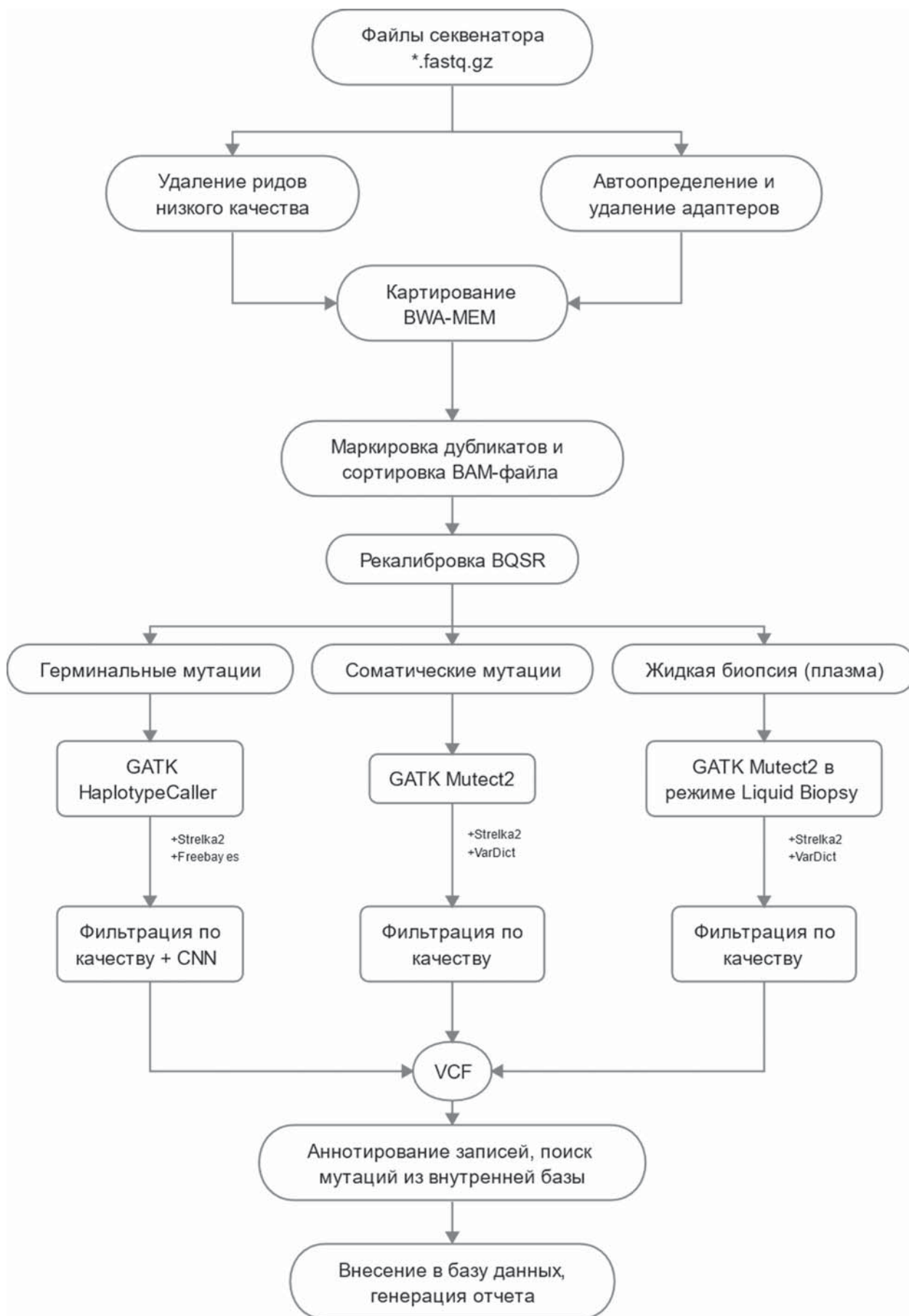


Таблица 1

Представленность в изучаемой выборке 8 самых частых мутаций генов *BRCA1/2*, используемых в ПЦР-диагностике в РФ

Наименование мутации	Число мутаций в выборке
BRCA1: 5382insC	87
BRCA1: 300T>G	9
BRCA1: 4153delA	9
BRCA1: 2080delA	6
BRCA1: 185delAG	4
BRCA1: 3819delGTAAA	4
BRCA1: 3875delGTCT	1
BRCA2: 6174delT	0

Пайплайн¹ построен по модульному принципу с использованием общих вычислительных блоков для облегчения внесения изменений в алгоритмы и контроля версий. Вычислительный конвейер автоматически подстраивается под размер входных файлов и распределяет нагрузку по необходимому числу контейнеров, что дает возможность быстро и эффективно обрабатывать данные любого размера — от таргетных панелей до экзоменов и геномов. Каждый модуль выполняется в изолированном окружении с динамическим выделением

¹ Пайплайн (от англ. pipeline — водопровод) — процесс разработки (подготовки, производства), программный конвейер.

ресурсов, требуемых для обработки загруженного набора файлов, а облачная среда снимает ограничения на количество одновременно анализируемых образцов. Время получения готовых отчетов для 80% пациентов из запуска MiSeq (96 образцов) не более 3 ч, для всех 96 — не более 24 ч.

РЕЗУЛЬТАТЫ

Основные результаты исследования

В результате секвенирования для генов *BRCA1/BRCA2* были обнаружены 29 912 нуклеотидных вариантов (однонуклеотидные замены, короткие вставки/делеции размером до 50 пар нуклеотидов), из них 119 мутаций у 374 пациентов являлись патогенными.

Широко распространенная ПЦР-панель из 8 частых мутаций (табл. 1) не отражает истинного распределения патогенных мутаций *BRCA1/BRCA2* в российских популяциях (например, мутация *BRCA2: 6174delT* из этой панели в нашей выборке не встретилась ни разу), что подтверждается результатами исследования OVATAR компании «АстраЗенека», и необходима разработка новых диагностических тест-систем для скрининга носительства частых мутаций в генах *BRCA1/BRCA2*.

Предлагаемая нами панель из 8 мутаций с учетом частоты их встречаемости в различных этносах приведена в табл. 2.

С учетом полученных результатов мы считаем, что в Российской Федерации необходимо проведение исследования по изучению распространен-

Таблица 2

Распределение частот 15 самых частых мутаций по данным настоящего исследования в различных этносах

Мутация	Славяне (%)	Татары (%)	Башкиры (%)	Чуваши (%)	Все (%)
c.5266dupC BRCA1:NM_007294.3 p.Q1756fs	8,5	6,6	10	3,2	6,7
c.181T>G BRCA1:NM_007294.3 p.C61G	0,6	1,4	0	0	0,7
c.4035delA BRCA1:NM_007294.3 p.E1345fs	0,8	0,4	0	0,3	0,7
c.5251C>T BRCA1:NM_007294.3 p.R1751X	1	0	0	0	0,5
c.5161C>T BRCA1:NM_007294.3 p.Q1721X	0	0,8	6,9	0	0,5

Таблица 2. Окончание

c.1961delA BRCA1:NM_007294.3 p.K654fs	0,8	0	0	0	0,46
c.3749dupA BRCA2:NM_000059.3 p.E1250fs	0	1	0	0	0,38
c.4675G>A BRCA1:NM_007294.3 p.E1559K	0,6	0,2	0	0	0,38
c.8754+1G>A BRCA2:NM_000059.3	0,3	0	0	3,2	0,3
c.961_962insAA BRCA2:NM_000059.3 p.Q321fs	0,1	0,8	3,4	0	0,3
c.3700_3704del BRCA1:NM_007294.3 p.V1234fs	0,1	0,6	0	0	0,3
c.4689C>G BRCA1:NM_007294.3 p.Y1563X	0,4	0	0	0	0,2
c.2897_2898del BRCA2:NM_000059.3 p.T966fs	0	0,4	0	0	0,2
c.3629_3630del BRCA1:NM_007294.3 p.E1210fs	0	0	0	0	0,2
c.4327C>T BRCA1:NM_007294.3 p.R1443X	0,3	0	0	0	0,2
c.1301_1304del BRCA2:NM_000059.3 p.K434fs	0,3	0	0	0	0,2
c.3143delG BRCA1:NM_007294.3 p.G1048fs	0,2	0,2	0	0	0,2
c.5075-1G>A BRCA1:NM_007294.3	0,3	0	0	0	0,2
c.7007+1G>A BRCA2:NM_000059.3	0	0,4	0	0	0,2
c.9117G>A BRCA2:NM_000059.3 p.P3039P	0,2	0,2	0	0	0,2
c.2806_2809del BRCA2:NM_000059.3 p.K936fs	0,3	0	0	3,2	0,2
c.3756_3759del BRCA1:NM_007294.3 p.L1252fs	0,2	0	0	0	0,1

ности герминальных мутаций при наследственных раках в панели генов (в том числе системы репарации), а также для идентификации патогенных вариантов, приводящих к развитию онкозаболеваний, причем это исследование должно включать несколько тысяч пациентов для увеличения статистической мощности и крупную выборку популяционного контроля для оценки популяционных частот.

Проведенный анализ результатов секвенирования нового поколения (Next Generation Sequencing, NGS) 1644 образцов наследственной онкопатологии выявил 119 патогенных мутаций генов *BRCA1/BRCA2* у 374 пациентов, при этом была обнаружена выраженная разница в частотах встречаемости мутаций между различными этносами — славянами, татарами, башкирами и чувашами.

Частота мутаций в генах *BRCA1, BRCA2* существенно отличается между пациентками с раком яичников и раком молочной железы славянского и татарского происхождения, что подтверждает необходимость NGS-анализа в случае отсутствия положительных результатов анализа ПЦР в реальном времени.

ОБСУЖДЕНИЕ

При анализе данных полногеномого секвенирования в клинической практике возникает две основные проблемы — скорость обработки/получения итогового результата и показатели чувствительности/специфичности проводимого исследования. Вариантов проведения анализа у лаборатории существует несколько:

- встроенное ПО прибора (например, MiSeq Reporter);
- облачное ПО производителя прибора (например, Illumina BaseSpace);
- готовое платное решение (например, SOPHiA GENETICS);
- использование собственного пайплайна на локальном сервере;
- использование собственного пайплайна в облаке.

Все варианты, кроме последнего, не соответствовали заданным параметрам:

- 1) время получения результата для 50% образцов в запуске не более 3 ч;
- 2) точность детекции мутаций не менее 99,9%.

Внедрение в практику экзомного и геномного секвенирования остро поставило вопрос аналитических характеристик используемых методов. Для правильной интерпретации результатов теста

важно иметь информацию о его точности, чувствительности и специфичности. К сожалению, из-за сложности NGS-секвенирования этот вопрос далек от окончательного решения, но научное сообщество предприняло ряд усилий для характеристики методик. В рамках проекта «Геном в бутылке» (Genome in a Bottle) J. Zook и соавт. [7] на нескольких платформах были отсекувенированы эталонные клеточные линии от добровольцев, данные секвенирования — объединены для получения кросс-валидированных высокодостоверных наборов вариантов, с которыми можно сравнивать результаты работы пайплайнов. Стандартом становятся показатели Recall и Precision более 99%. ПО MiSeq Reporter, как и приложение BWA Enrichment App в облаке Illumina BaseSpace, используют слишком старые версии программ: BWA 0.7.7-0.7.9a (опубликованы в 2014 г.), Picard 1.79 (октябрь 2012 г.), GATK 1.6 (2012 г.). По данным N. Miller и соавт. [8], чувствительность метода при использовании GATK версии 1.6 составила всего 27,3% по сравнению с 97,3% у версии 3.2. Показатель чувствительности приложения Isaac Enrichment App в облаке Illumina BaseSpace составляет 94–95%, что также не соответствует современным требованиям к анализу данных NGS, так как чувствительность новых версий (например, GATK4) более 98–99%. Данных цифр можно добиться при использовании собственных пайплайнов, запускаемых на локальном сервере или в облачной среде, но главной проблемой анализа больших файлов является отсутствие линейного прироста производительности при масштабировании биоинформационных алгоритмов на многопроцессорных системах. Так, BWA-MEM эффективно распараллеливается не более чем на 12–16 ядер CPU, инструмент HaplotypeCaller из набора GATK4 всегда работает в однопоточном режиме (несколько потоков использует только алгоритм PairHMM), шаги дедупликации, сортировки и рекалибровки BQSR также имеют существенные ограничения в максимальном использовании доступных вычислительных мощностей — все это приводит к тому, что не удается значительно снизить время обработки образца даже при наличии мощного локального сервера без отказа от части вычислений или замены алгоритмов на более быстрые, но менее точные. Например, по данным J. Chen и соавт. [9], оптимизация пайплайна для 24 CPU обеспечивает для генома общее время выполнения двух шагов GATK4-BQSR + GATK4-HaplotypeCaller в диапазоне 35–40 ч и оценку полного времени выполнения

всего пайплайна в диапазоне 70–90 ч с учетом сбора метрик качества, причем дальнейшее увеличение количества CPU не приводит к пропорциональному уменьшению времени обработки. Для решения этой проблемы команда GATK (столкнувшись с необходимостью анализа десятков и сотен геномов в день) разработала ряд инструментов и подходов для распределенных высокопроизводительных вычислений, реализованных в последней версии The Genome Analysis Toolkit [10]. Все шаги, допускающие независимую обработку определенного набора прочтений (BWA-MEM, BQSR, HaplotypeCaller и т.д.), автоматически распределяются по необходимому количеству виртуальных машин с оптимальными для каждой подзадачи ресурсами и настройками, а шаги, требующие анализа полного набора данных (дедупликация, сортировка), выполняются в кластерном окружении с подстройкой размера кластера под объем данных. Дальнейшие способы ускорения геномных вычислений связаны с использованием аппаратного ускорения (графические ускорители для ПО Parabricks и FPGA для ПО DRAGEN), но достигаемые скорости обработки в тысячи экзомов/геномов в сутки пока мало актуальны для Российской Федерации.

Главным преимуществом облачного подхода к биоинформационному анализу является независимость обработки образцов друг от друга, что позволяет одновременно анализировать десятки и сотни образцов, получая на примере панели генов системы репарации общее время биоинформационного анализа всего запуска прибора в пределах 3 ч с обеспечением максимальной точности результата. Используемый пайплайн обладает следующими характеристиками при анализе референсного образца GiaB/NIST NA12878/HG001 на экзомной панели Illumina Nextera Expanded Exome с учетом высокодостоверных регионов:

- Precision 0,99004;
- Recall 0,99993;
- F-measure 0,99496.

Автоматизация поиска ранее известных патогенных мутаций дает возможность формировать готовый отчет об исследовании для 70–80% образцов в запуске сразу по завершению обработки, при этом в нашем проекте после роста размера базы данных свыше 1000 пациентов около 50–60% новых образцов имели ранее обнаруженные и интерпретированные патогенные мутации, 10–15% не имели мутаций вообще и лишь 20% требовали

дальнейшей интерпретации возможно патогенных вариантов. Мы ожидаем, что при росте количества образцов в базе свыше 10 000 более 90% образцов в запуске не потребует затрат времени на интерпретацию.

ЗАКЛЮЧЕНИЕ

С помощью метода NGS появилась возможность в короткие сроки выявлять редкие мутации, характерные для различных этносов, что дает возможность оптимизировать диагностическую и лечебную тактику пациентов из данной популяции.

В результате проведенной работы была создана публичная база данных частоты мутаций и полиморфизмов генов системы репарации в различных этносах (<https://oncobrca.ru>).

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Работа поддержана компанией «АстраЗенека» (грант ESR-17-12934).

КОНФЛИКТ ИНТЕРЕСОВ

Авторы данной статьи подтвердили отсутствие конфликта интересов, о котором необходимо сообщить.

УЧАСТИЕ АВТОРОВ

А.Г. Никитин — написание статьи, биоинформационный анализ данных; О.И. Бровкина, Д.С. Ходырев — биоинформационный анализ данных; О.А. Гусев — лабораторная часть исследования; М.Г. Гордиев — клинический дизайн исследования, сбор биоматериала. Все авторы внесли существенный вклад в проведение поисково-аналитической работы и подготовку статьи, прочли и одобрили финальную версию до публикации.

СПИСОК ЛИТЕРАТУРЫ

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65(2):87–108. doi: 10.3322/caac.21262.
2. Sokolenko AP, Iyevleva AG, Mitushkina NV, et al. Hereditary Breast-Ovarian Cancer Syndrome in Russia. *Acta Naturae*. 2010;2(4):31–35.
3. Foretova L, Machackova E, Navratilova M, et al. BRCA1 and BRCA2 mutations in women with familial or early-onset breast/ovarian cancer in the Czech Republic. *Hum Mutat*. 2004;23(4):397–398. doi: 10.1002/humu.9226.
4. Хасанова А.И., Гордиев М.Г., Ратнер Е.Ю., и др. BRCA-ассоциированный рак молочной железы у представительниц татарской национальности на примере клинического случая // *Приволжский онкологический вестник*. — 2016. — №2. — С. 104–108. [Khasanova AI, Gordiev MG, Ratner EYu, et al. Clinical report of BRCA-associated breast cancer among representative of the tatar nationality group. *Privolzhskiy onkologicheskiy vestnik*. 2016;(2):104–108. (In Russ).]
5. Fackenthal JD, Olopade OI. Breast cancer risk associated

with BRCA1 and BRCA2 in diverse populations. *Nat Rev Cancer*. 2007;7(12):937–948. doi: 10.1038/nrc2054.

6. Matsuda S. Defective DNA repair systems and the development of breast and prostate cancer (Review). *Int J Oncol*. 2013;42(1):29–34. doi: 10.3892/ijo.2012.1696.

7. Zook JM, McDaniel J, Olson ND, et al. An open resource for accurately benchmarking small variant and reference calls. *Nat Biotechnol*. 2019;37(5):561–566. doi: 10.1038/s41587-019-0074-6

8. Miller NA, Farrow EG, Gibson M, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency manage-

ment of genetic diseases. *Genome Med*. 2015;7:100. doi: 10.1186/s13073-015-0221-8.

9. Chen J, Li X, Zhong H, et al. Systematic comparison of germline variant calling pipelines across multiple next-generation sequencers. *Sci Rep*. 2019;9(1):1–13. doi: 10.1038/s41598-019-45835-3.

10. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43.

КОНТАКТНАЯ ИНФОРМАЦИЯ

Никитин Алексей Георгиевич

к. биол. н., заведующий лабораторией генетики ФНКЦ специализированных видов медицинской помощи и медицинских технологий ФМБА России;

адрес: 115682, Москва, Ореховый бульвар, д. 28, **e-mail:** avialn@gmail.com, **SPIN-код:** 3367-0680,

ORCID: <https://orcid.org/0000-0001-9762-3383>

Бровкина Ольга Игоревна

к. биол. н., старший научный сотрудник ФНКЦ специализированных видов медицинской помощи и медицинских технологий ФМБА России;

e-mail: brov.olia@gmail.com, **SPIN-код:** 3631-1397, **ORCID:** <https://orcid.org/0000-0002-0946-7331>

Ходырев Дмитрий Сергеевич

к.биол.н., старший научный сотрудник ФНКЦ специализированных видов медицинской помощи и медицинских технологий ФМБА России;

e-mail: dmkh008@gmail.com, **SPIN-код:** 4292-7072

Гусев Олег Александрович

к.биол.н., заведующий лабораторией экстремальной биологии Казанского (Приволжского) федерального университета; научный сотрудник Института физико-химических исследований (RIKEN)

e-mail: gaijin.ru@gmail.com, **SPIN-код:** 5666-3711, **ORCID:** <https://orcid.org/0000-0002-6203-9758>

Гордиев Марат Гордиевич

заведующий генетической лаборатории Республиканского клинического онкологического диспансера Министерства здравоохранения Республики Татарстан

e-mail: marat7925@gmail.com, **SPIN-код:** 8388-3566