

ՀԱՅԱՍՏԱՆԻ ՀԱՆՐԱՊԵՏՈՒԹՅԱՆ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԶԳԱՅԻՆ ԱԿԱԴԵՄԻԱ  
ՄՈԼԵԿՈՒԼԱՅԻՆ ԿԵՆՍԱԲԱՆՈՒԹՅԱՆ ԻՆՍՏԻՏՈՒՏ

ՅՈՒԴԻՆԱ ԱՆԱՍՏԱՍԻԱ ՍԵՐԳԵՅԻ

ՈՒՌՈՒՑՔՆԵՐԻ ԿԼԻՆԻԿՈՐԵՆ ՆՇԱՆԱԿԱԼԻ ՕՆԿՈԳԵՆ ՄՈԼԵԿՈՒԼԱՅԻՆ  
ԽԱՆԳԱՐՈՒՄՆԵՐԻ ՀԱՅՏՆԱՐԵՐՈՒՄ ԳԵՆՈՄԱՅԻՆ ԵՎ ՏՐԱՆՍԿՐԻՊՏՈՄԱՅԻՆ  
ՏՎՅԱԼՆԵՐԻ ՀԻՄԱՆ ՎՐԱ

Գ.00.03. – «Մոլեկուլային և բջջային կենսաբանություն» մասնագիտությամբ  
կենսաբանական գիտությունների թեկնածուի գիտական աստիճանի հայցման  
ատենախոսության

ՍԵՂՄԱԳԻՐ

ԵՐԵՎԱՆ – 2026

---

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК РЕСПУБЛИКИ АРМЕНИЯ  
ИНСТИТУТ МОЛЕКУЛЯРНОЙ БИОЛОГИИ

ЮДИНА АНАСТАСИЯ СЕРГЕЕВНА

ИДЕНТИФИКАЦИЯ КЛИНИЧЕСКИ ЗНАЧИМЫХ ОНКОГЕННЫХ МОЛЕКУЛЯРНЫХ  
НАРУШЕНИЙ ОПУХОЛЕЙ НА ОСНОВЕ ГЕНОМНЫХ И ТРАНСКРИПТОМНЫХ  
ДАНЫХ

АВТОРЕФЕРАТ

диссертации на соискание ученой степени  
кандидата биологических наук по специальности  
03.00.03 – «Молекулярная и клеточная биология»

ЕРЕВАН – 2026

Ատենախոսության թեման հաստատվել է ՀՀ ԳԱԱ մոլեկուլային կենսաբանության  
ինստիտուտի գիտական խորհրդում:

Գիտական ղեկավար՝ Կ.գ.թ. Բագան Ալեքսանդր Վլադիսլավի

Պաշտոնական ընդդիմախոսներ՝ տեխ.գ.դ. Բորիսով Նիկոլայ Միխայիլի  
Կ.գ.թ. Բաբայան Նելլի Սամվելի

Առաջատար կազմակերպություն՝ Երևանի պետական համալսարան

Ատենախոսության պաշտպանությունը տեղի կունենա 2026 թ. մայիսի 21-ին, ժամը  
14:00-ին, ՀՀ ԳԱԱ մոլեկուլային կենսաբանության ինստիտուտում, 042  
մասնագիտական խորհրդի նիստում (ՀՀ 0014, ք. Երևան, Հասրաթյան 7):

Ատենախոսությանը կարելի է ծանոթանալ ՀՀ ԳԱԱ մոլեկուլային կենսաբանության  
ինստիտուտի գրադարանում և <https://imb.am/> կայքում:

Ատենախոսության սեղմագիրն առաքվել է 2026 թ. ապրիլի 20-ին:

042 մասնագիտական խորհրդի գիտական քարտուղար,  
կենս. գիտ. թեկնածու



Ռ.Վ. Զախարյան

---

Тема диссертации утверждена на заседании ученого совета Института молекулярной  
биологии НАН РА.

Научный руководитель: к.б.н. Багаев Александр Владиславович

Официальные оппоненты: д.тех.н. Борисов Николай Михайлович  
к.б.н. Бабаян Нелли Самвеловна

Ведущая организация: Ереванский государственный университет

Защита диссертации состоится 21 мая 2026 г. в 14:00, на заседании специализированного  
совета 042, в Институте молекулярной биологии НАН РА (РА, 0014, г. Ереван, ул. Асратяна  
7).

С диссертацией можно ознакомиться в библиотеке Института молекулярной биологии НАН  
РА и на сайте <https://imb.am/>.

Автореферат диссертации разослан 20 апреля 2026 г.

Ученый секретарь специализированного совета 042  
кандидат биол. наук.



Захарян Р.В.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Постановка задачи.** Онкологические заболевания остаются одной из ведущих причин заболеваемости и смертности в мире, что определяет высокую значимость разработки точных методов молекулярной диагностики и персонализированной терапии [Siegel *et al.*, 2025]. Комплексное геномное профилирование опухолей стало ключевым инструментом современной онкологии для поиска биомаркеров [ASCO, 2019].

Стандартом остается анализ ДНК из тканевых биопсий, обеспечивающий стабильность материала и равномерность покрытия регионов интереса, однако такой подход не отражает функциональную реализацию выявленных мутаций, что может быть дополнено данными транскриптомного секвенирования [Hussen *et al.*, 2022]. Дополнительные сложности анализа связаны с наличием в геноме гомологичных последовательностей и последовательностей низкой сложности [Mandelker *et al.*, 2016], скорректировать находки из которых также можно за счет объединения данных РНК и ДНК.

В последние годы активно развивается направление анализа внеклеточной опухолевой ДНК из плазмы крови. Этот малоинвазивный метод позволяет проводить мониторинг ремиссии и рецидива заболевания, а также использоваться при невозможности получения тканевого материала. Также использование данных внеклеточной опухолевой ДНК может повысить чувствительность поиска значимых для терапий биомаркеров по сравнению с анализом только материала первичного очага опухоли, т. к. в кровь выходят фрагменты ДНК метастатических сайтов [Brannon *et al.*, 2021]. Вместе с тем низкая концентрация входной ДНК (5-10 нг/мл плазмы) и малая доля опухолевой фракции (1-10%) требуют увеличения глубины покрытия и применения алгоритмов подавления артефактов, возникающих при сверхглубоком секвенировании [Maruzani *et al.*, 2024].

Интеграция данных ДНК- и РНК-секвенирования рассматривается как перспективный путь повышения точности диагностики. Добавление транскриптомных данных позволяет подтвердить экспрессию мутантного аллеля, уточнить функциональную значимость варианта, выявить нарушения сплайсинга и повысить чувствительность детекции мутаций с низкой аллельной частотой [Wilkerson *et al.*, 2014]. Кроме того, анализ внеклеточной ДНК может выявлять клинически значимые альтерации, пропущенные при исследовании тканевых образцов вследствие опухолевой гетерогенности [Park *et al.*, 2021].

Несмотря на развитие технологий, в настоящее время отсутствует единый стандарт аналитической валидации интегрированных тестов на основе полноэкзомного и транскриптомного секвенирования как для тканевых образцов, так и для внеклеточных нуклеиновых кислот. Существующие рекомендации преимущественно ориентированы на таргетные панели [Frampton *et al.*, 2013], а руководства для полноэкзомных тестов остаются ограниченными по методологии ортогональной проверки и оценке специфичности [Shah *et al.*, 2024].

Таким образом, разработка и аналитическая валидация подходов, объединяющих данные полноэкзомного и транскриптомного секвенирования с анализом внеклеточных нуклеиновых кислот, является актуальной и недостаточно разработанной задачей. Реализация таких методов направлена на повышение чувствительности молекулярной диагностики, корректную интерпретацию соматических альтераций и обеспечение назначения эффективной таргетной терапии.

### Цель и задачи работы

Цель диссертационной работы: выявить клинически значимые онкогенные мутации для широкого спектра онкологических заболеваний на основе интегрированного анализа данных секвенирования ДНК и РНК из тканевых и жидкостных биопсий с целью повышения информативности молекулярного профилирования для выбора таргетных терапий.

Для достижения этой цели были поставлены следующие задачи:

1. Разработать и валидировать подходы к интегрированному анализу данных секвенирования ДНК и РНК опухолевых тканей, фиксированных в парафине (FFPE), с учетом молекулярных особенностей деградации нуклеиновых кислот и фильтрацией артефактов фиксации с использованием методов машинного обучения.
2. Разработать и валидировать подходы по выявлению и анализу распределения соматических мутаций с низкой аллельной частотой (до 1%) по данным секвенирования циркулирующей опухолевой ДНК плазмы крови.
3. Исследовать взаимосвязь функциональных групп генов и типов точечных мутаций (GOF, LOF, мутации горячих точек, изменения по всему экзому) в данных секвенирования ДНК и РНК у когорты пациентов ( $n > 3000$ ) с онкологическими заболеваниями различной нозологии для выявления категорий генов, в которых интеграция транскриптомных и геномных данных повышает биологическую и клиническую значимость обнаруженных альтераций.
4. Оценить предиктивный потенциал интегрированного анализа транскриптомных и геномных данных для когорты образцов FFPE и жидкостных биопсий ( $n = 100$ ) с целью повышения точности выявления клинически значимых соматических мутаций.
5. Расширить функциональную аннотацию VUS-вариантов посредством анализа данных РНК-секвенирования, включая оценку влияния мутаций на сплайсинг, с целью предсказания потенциальной устойчивости опухолей к терапии.

**Научная новизна.** В настоящей работе предложен комплексный подход к исследованию молекулярного профиля опухолевой гетерогенности, основанный на интеграции данных секвенирования нуклеиновых кислот из деградированных клинических материалов (FFPE и плазма крови). Подход обеспечивает объединённое геномно-транскриптомное профилирование опухолей и учитывает ограничения, связанные с качеством исходного материала и низкой долей опухолевой ДНК. В рамках методологической части создан референсный набор соматических мутаций модельных клеточных линий с учётом особенностей клеточного состава, предназначенный для молекулярной калибровки и аналитической валидации тестов, приближенных к реальным клиническим условиям.

Разработан и экспериментально валидирован высокочувствительный подход к выявлению соматических мутаций из образцов FFPE тканей и из циркулирующей опухолевой ДНК. Показано, что объединение данных тканевых и жидкостных биопсий повышает полноту молекулярного профилирования, расширяя спектр клинически значимых мутаций и улучшая выявление потенциальных драйверных событий в реальной когорте пациентов.

Получены новые биологические результаты, подтверждающие функциональную значимость интеграции транскриптомных данных. Показано, что объединённый анализ ДНК и РНК позволяет подтвердить значимые онкогенные события, а также выявить дополнительные ранее пропущенные мутации. Установлено, что GOF-мутации могут отсутствовать в данных РНК, в том числе подобные находки выявлены в гене *PIK3CA*, критичном для выбора таргетной терапии рака молочной железы. Кроме того, продемонстрировано, что ассоциация мутаций с нарушениями сплайсинга позволяет функционально классифицировать варианты неопределённой значимости (например, *RBI* c.1389+2\_1389+3insG) и объяснять механизмы резистентности к терапии в клинической практике.

**Практическая ценность работы** заключается в разработке и внедрении валидированного интегративного подхода к клиническому геномному профилированию опухолей на основе анализа ДНК и РНК из FFPE-образцов и плазмы крови, что расширяет

возможности молекулярной диагностики при ограниченном или деградированном материале. Сформированная референсная база соматических мутаций модельных клеточных линий и предложенные алгоритмы биоинформатической обработки обеспечивают стандартизацию, калибровку и воспроизводимость NGS-тестов, включая выявление низкочастотных вариантов. Показана клиническая значимость интеграции РНК- и ДНК-секвенирования ткани и жидкостной биопсии для повышения полноты мутационного профиля, выявления дополнительных терапевтически значимых событий, что способствует более обоснованному выбору таргетной терапии и персонализации лечения пациентов.

**Степень достоверности и апробация результатов.** Достоверность полученных результатов обеспечивается применением разработанных подходов к реальным клиническим образцам тканевых в рамках клинического испытания BIGR, BEGIN (NCT04991922, NCT06272864,  $n > 3800$ ) и жидкостных биопсий ( $n = 238$ ). Результаты подтверждены многочисленными экспериментами на модельных клеточных линиях и референсных генетических стандартах.

Основные результаты диссертации апробированы в международном научном сообществе: опубликовано 2 статьи в рецензируемых журналах – Communications Medicine, Human Genetics и 1 статья в сборнике Докладов НАН РА. Результаты докладывались на 4 международных конференциях, включая съезды ASH (2023), AACR (2023), AMP (2023, 2025) (США), ESMO (2023) (Испания), а также на заседаниях Ученого совета Института молекулярной биологии НАН РА.

**Объем и структура работы.** Диссертация изложена на 158 страницах машинописного текста и содержит следующие разделы: введение, обзор литературы, материалы и методы, результаты и обсуждение, заключение, выводы, список литературы, а также приложение (с. 159-174). Работа содержит 59 оригинальных рисунков и 20 таблиц. Список литературы включает 215 библиографических ссылок.

**Публикации.** Результаты работы отражены в 3-х научных публикациях и в 4-х тезисах представленных на международных конференциях.

## **МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ**

**Клеточные линии FFPE.** Для оценки точности выявления соматических SNV и коротких инделов использованы опухолевые клеточные линии (HCC1143, HCC1937, COLO829, HCC1395, NCI-H1770) и соответствующие парные нормальные линии (ATCC). Были сформированы FFPE-блоки опухолевых и нормальных клеток с последующим выделением ДНК и РНК. Экстракты смешивали в различных пропорциях (0-100%) для моделирования разной опухолевой клеточности. Для линии COLO829 дополнительно проанализированы 70 технических реплик для оценки воспроизводимости.

**Референсные материалы для жидкостной биопсии.** Для валидации анализа внеклеточной ДНК использованы стандарты SeraSeq® и Twist PanCancer с заданными аллельными частотами (0.1-2.5% VAF). Эксперименты проводили в технических репликах для определения лимита детекции, чувствительности и специфичности теста.

**Лабораторные и публичные клинические данные.** Проанализированы данные TCGA (LUAD и BRCA) с парными экзомным и транскриптомным секвенированием, а также несколько клинических когорт с опухолевыми диагнозами разных нозологий, собранными и просеквенированными в CLIA/CAP сертифицированной лаборатории в Уолтеме, Массачусетс: 3803 пациента для интегративного анализа ДНК/РНК, 181 образец без нормы для разработки алгоритмов фильтрации, и 100 пациентов с парными образцами ткани и плазмы для сопоставления мутационных профилей. Дополнительно использована когорта из 238 пациентов с жидкостной биопсией.

## **Аналитические и биоинформатические методы**

**Выравнивание и контроль качества.** Аннотация регионов интереса выполнялась на основе GENCODE и MANE с учётом сложности последовательностей и качества картирования. Контроль качества FASTQ-файлов проводился с использованием FastQC; дубликаты маркировались Picard. WES и таргетные ДНК-данные выравнивались на hg38 с помощью BWA; для внеклеточной ДНК применялась UMI-дедупликация с формированием консенсусных прочтений. RNA-seq данные выравнивались STAR, количественная оценка экспрессии выполнялась Kallisto.

**Поиск и аннотация мутаций.** Соматические точечные мутации и короткие делеции/вставки в данных экзомного секвенирования (опухоль/норма) определялись Strelka2; в образцах без нормы и в РНК-секвенировании использовались Pindel и Mutect2 с пониженными порогами детекции. Для жидкостной биопсии применялись LoFreq и Mutect2 с дополнительной фильтрацией артефактов. Аннотация вариантов проводилась с использованием VEP с учётом dbSNP, gnomAD и COSMIC; клиническая интерпретация выполнялась по уровням доказательности (LoE 1-4) согласно рекомендациям AMP/ASCO/CAP и NCCN/FDA.

**Фильтрация шума.** Для фильтрации данных полноэкзомного секвенирования исключались гены с высокой частотой полиморфизмов (HLA, VDJ). Основные пороги: прочтения, содержащие мутацию ( $t\_alt\_count \geq 7$ ,  $VAF \geq 0.01$ ,  $VAF \times t\_alt\_count \geq 0.4$ ). Баланс чувствительности и точности (F1) оптимизировался с помощью Grid Search и градиентного бустинга XGBoost на клеточных линиях HCC1143 и COLO829 с различными разведениями. Данные делились на обучающую и тестовую выборки 80/20. Фильтрация мутаций в РНК включала три этапа: базовый —  $VAF \geq 0.2$ ,  $\geq 3$  альтернативных прочтений и покрытие  $>5\times$  с исключением врожденных вариантов и позиций DARNED; кодирующие регионы — экзонные миссенс- и нонсенс-мутации, короткие делеции/вставки; «виртуальная норма» — удаление вариантов с частотой  $>0.5\%$  по gnomAD и подбор порога патогенности MutAnt (0.1). Контроль производился на независимых выборках для предотвращения переобучения. Совместная фильтрация экзомных и РНК мутаций включала классификацию вариантов на врожденные, соматические и артефакты. Обучающая выборка — 151 образец, тестовая — 30; балансировка классов выполнялась SMOTE. Использовался XGBClassifier с признаками качества выравнивания, покрытия, поддержки аллелей и частот gnomAD; вклад признаков оценивался с помощью SHAP. Для таргетного секвенирования внеклеточной ДНК плазмы применялись пороги фильтрации: две реплики —  $\geq 6$  прочтений,  $VAF \geq 0.001$ ,  $VAF \times depth \geq 0.03$ , качество  $\geq 70$ ; одна реплика —  $\geq 6$  прочтений,  $VAF \geq 0.001$ ,  $VAF \times depth \geq 0.06$ , качество  $\geq 80$ . При объединении Mutect2 и LoFreq параметры использовались как независимые признаки, дополнялись фильтрами FilterMutectCalls и частотами gnomAD. Финальная классификация вариантов с низкой аллельной частотой выполнялась XGBoost с порогом вероятности 0.3 для оптимального баланса чувствительности и точности.

**Валидация** алгоритмов поиска точечных мутаций и инделов выполнялась на размеченных наборах данных клеточных линий с использованием реплик, разведений и эталонных контролей, предоставленных поставщиками. Разметка вариантов проводилась по геномной координате и аминокислотной замене, что позволяло учитывать различия аннотации транскриптов. Основными метриками качества были точность ( $TP/(TP+FP)$ ) и чувствительность ( $TP/(TP+FN)$ ), рассчитываемые для разных уровней представленности мутаций и глубины покрытия. Для оценки согласованности находок РНК и ДНК использовался индекс Жаккарда, отражающий долю пересечения вариантов между двумя источниками. Для каждого разведения рассчитывались доверительные интервалы метрик с учётом объёма выборки, что позволяло объективно сравнивать производительность алгоритмов поиска и фильтрации мутаций, а также эффективность фильтров артефактов.

**Анализ событий сплайсинга.** Альтернативные события сплайсинга (AS) анализировались с помощью Spladder v3.0.2 на наборе из 151 гена с использованием эталонных транскриптов MANE Select. Каждый тип AS (пропуск экзона, удержание интрона, альтернативные 3'/5'-сайты) сохранялся отдельно, после чего проводилось пересечение с точечными мутациями для отбора вариантов, влияющих на сплайсинг. Рассматривались только мутации, пересекающие экзоны, вовлечённые в AS-событие ( $\pm 3$  п.н.), с исключением остальных экзонов того же гена. Для оценки влияния мутаций на AS вводилась метрика Splice Efficiency (SE), отражающая интенсивность проявления события. SE рассчитывалась по типу события: для альтернативных сайтов — отношение поддерживающих прочтений к каноническим, для пропуска экзонов — отношение числа прочтений экзона к фланкирующим стыкам, для удержания интрона — отношение покрытия интрона к соседним экзонам. Статистическая значимость проверялась с помощью Z-оценок с поправкой на множественное тестирование (FDR).

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

**Сборка референсов соматических мутаций.** С целью валидации поиска соматических мутаций был собран репрезентативный набор вариантов на основе данных клеточных линий COLO829, HCC1143, HCC1937, HCC1395 и NCI-H1770 с сериями разведений от 10% до 100% опухолевых клеток [Zook *et al.*, 2014, 2019, Wagner *et al.*, 2022]. Истинные мутации определялись по повторяемости в сериях разведений и коэффициентам корреляции VAF с долей опухолевых клеток, что позволило классифицировать варианты на истинные и артефакты для оценки точности алгоритмов детекции и фильтрации (Рисунок 1). Для оценки чувствительности использовались 4-6 технических реплик NGS при 100% опухолевых клетках, с контролем парных нормальных образцов, исключением вариантов с низким VAF (<5%) и минимальным покрытием <80 прочтений.

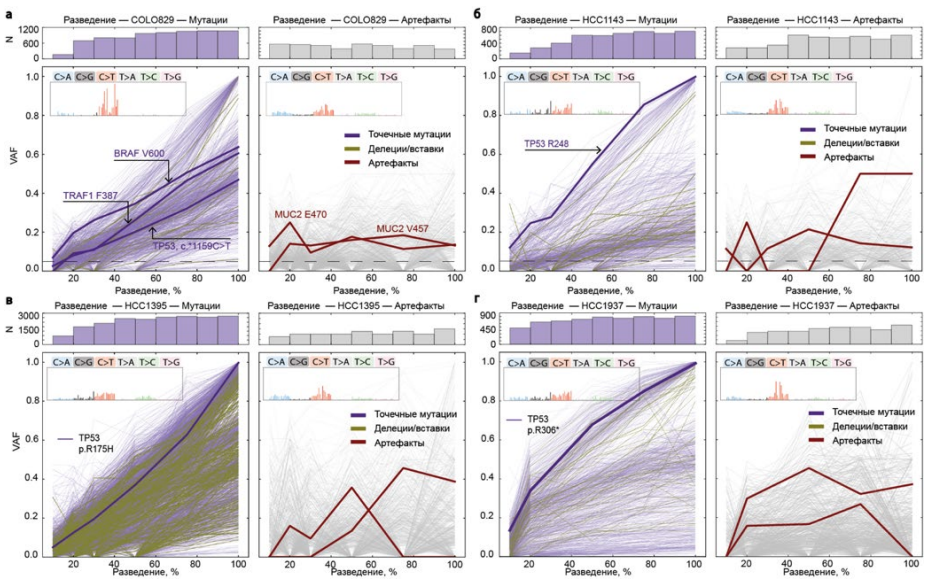


Рисунок 1. Зависимость разведения частоты варианта от разведения опухолевой клеточной линии и частоты геномных контекстов замен с разделением на истинные варианты и предсказанные артефакты. а, б, в, г – клеточные линии COLO829, HCC1143, HCC1935, HCC1937.

В референсный набор вошли 3042 точечных варианта: 1128 миссенс-мутаций, 746 синонимичных замен, 745 интронных вариантов, 140 мутаций сайтов сплайсинга, 58 нонсенс-мутаций, 183 варианта 5' и 3' UTR, 22 делеции со сдвигом рамки считывания, 13 делеций и 7 вставок с сохранением или нарушением рамки, а также 1 мутация со сломанным стоп-кодоном (Рисунок 2). Для каждой линии было определено до 18 354 генов дикого типа, что расширяет ранее доступные наборы данных для валидации платформ масштаба полного экзома. Мутации распределены по экзому со средней плотностью 43 мутации/Мб, что обеспечивает разнообразный эталон соматического экзома. Аллельные частоты 558 точечных мутаций сравнивались с опубликованными данными [Olsson *et al.*, 2015] и проверялись ортогонально по секвенированию методом Сенгера; чувствительность поиска составила 98%, корреляция VAF  $R = 0.87$ ,  $p = 4.9e-177$ .

Мутационная нагрузка на хромосому

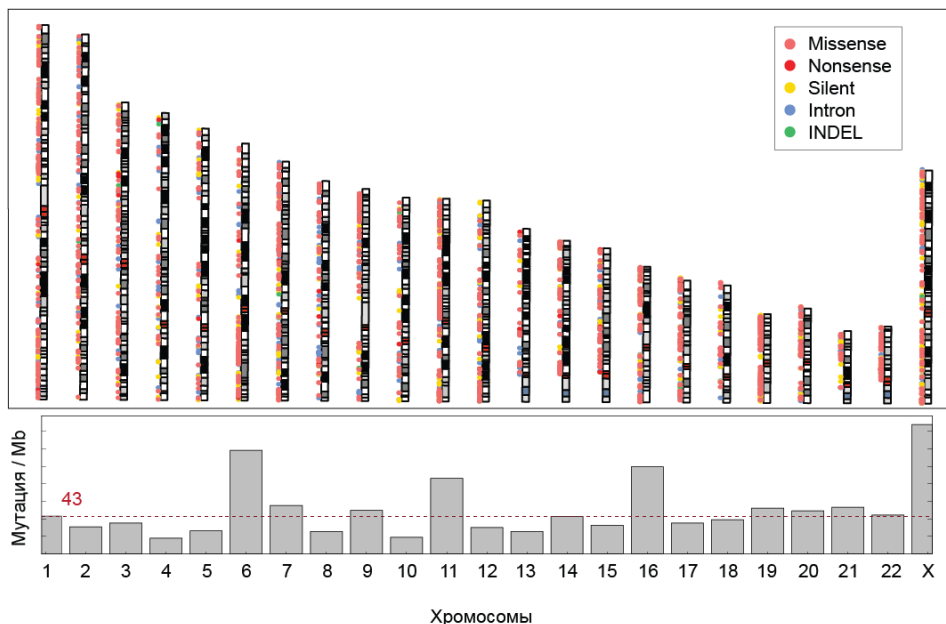


Рисунок 2. Распределение референсных мутаций по геному человека.

Верификация набора с использованием данных РНК показала, что 35% референсных вариантов обнаруживались в транскриптом, при этом пересечение с потенциальными артефактами составило <1% (Рисунок 3). Аллельные частоты экспрессированных вариантов в РНК и ДНК демонстрировали слабую, но статистически значимую корреляцию ( $r = 0.3-0.6$ ,  $p < 0.05$ ), что отражает аллель-специфическую экспрессию и дифференциальную опухолевую экспрессию генов [Методы]. VAF в РНК для генов с высокой опухолевой экспрессией может быть высоким даже при низкой доле опухолевых клеток (<30%) – Рисунок 4. Геномный контекст истинных и ложноположительных вариантов показал биологически значимое распределение для истинных мутаций, в то время как артефакты имели случайный контекст.

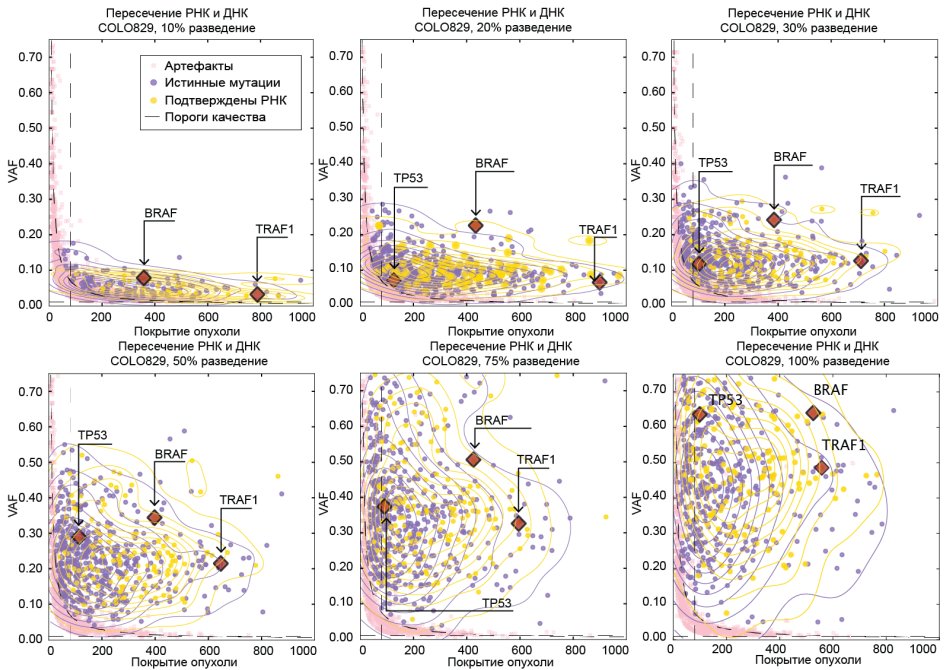


Рисунок 3. Пересечение мутаций в ДНК и РНК с визуализацией зависимости VAF мутации в ДНК от покрытия. Красными маркерами выделены мутации из базы данных Cancerhotspots.

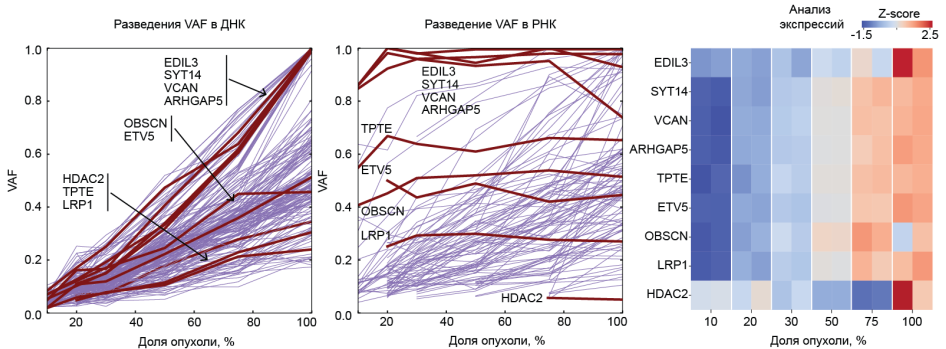


Рисунок 4. Исследование особенности зависимости аллельной частоты и доли опухоли в РНК и ДНК.

Для линии COLO829 стабильность детектирования проверялась на 70 контрольных образцах за 4 месяца; VAF мутаций оставался стабильным с отклонениями  $\leq 2.43\%$  для ключевых мутаций (*BRAF* V600E, *ATR* L2064, *HP1BP3* I188), при этом различались два клональных подтипа, что подтверждает воспроизводимость и надежность набора референсных мутаций.

**Валидация алгоритмов поиска соматических мутаций из FFPE.** Strelka2 показал оптимальный баланс точности и ресурсов, с чувствительностью  $>80\%$  и точностью  $>90\%$  для

точечных мутаций на WES (80×) [Kim et al., 2017]. Первичная валидация на клеточных линиях выявила F1 = 73.7% для точечных мутаций и 76.4% для делеций/вставок из-за артефактов, при этом чувствительность оставалась >90% при низкой доле опухоли (20-30%).

Для снижения артефактов оптимизировали параметры Somatic EVS и QSS с логистической функцией. Оптимальные пороги: VAF  $\geq 0.01$ ,  $\geq 7$  поддерживающих прочтений, коэффициент зависимости  $\geq 0.4$ ; более строгие пороги увеличивали точность. При доле опухоли 20% чувствительность составила  $85.0 \pm 2.5\%$ , точность —  $94.1 \pm 1.9\%$ .

На 3042 референсных вариантах из пяти клеточных линий при  $\geq 20\%$  доли опухоли чувствительность достигла  $95.3 \pm 0.9\%$  для точечных мутаций и  $85.2 \pm 3.0\%$  для делеций/вставок; при  $\geq 30\%$  —  $97.1 \pm 0.8\%$  и  $89.9 \pm 1.8\%$ . Точность по всем разведениям >95%, специфичность  $99.8 \pm 0.04\%$  – Таблица 1.

Сравнение с Mutect2 и Dragen показало преимущество оптимизированного Strelka2: F1 точечных мутаций 95.3%, F1 делеций/вставок 91.9% против 68-92% у других методов [Kim et al., 2017].

Таблица 1. Метрики качества поиска мутаций (точечные мутации и короткие делеции/вставки) в результате раундов оптимизации алгоритма Strelka2 на данных клеточных линий.

--	--	Разведение, %					
		10	20	30	50	75	100
Фильтр макс. Точность	Чувст-сть	37.23%	68.82%	73.58%	88.78%	93.96%	94.41%
	Точность	91.92%	97.70%	98.41%	98.12%	97.54%	96.81%
	F1	50.48%	79.46%	82.74%	93.05%	95.68%	95.55%
Фильтр макс. F1	Чувст-сть	74.01%	88.23%	90.09%	95.83%	97.24%	97.79%
	Точность	80.43%	89.56%	92.90%	94.51%	94.13%	93.77%
	F1	76.35%	88.55%	91.32%	95.13%	95.62%	95.72%
Фильтр EVS, QSS	Чувст-сть	71.85%	85.02%	88.74%	94.23%	94.90%	96.24%
	Точность	88.59%	94.09%	95.90%	96.42%	96.27%	95.41%
	F1	78.65%	89.10%	92.07%	95.27%	95.51%	95.81%

**Поиск и фильтрация мутаций из РНК.** Для выявления соматических мутаций в РНК-секвенировании использовались алгоритмы Pisce и Mutect2 с оптимизированными порогами для подавления FFPE-шума [Quinn et al., 2013, Piskol et al., 2014]. Фильтрация включала оценку патогенности вариантов с помощью MutAnt, SIFT4G, BayesDel и ClinPred, что позволило повысить пересечение мутаций РНК и ДНК, особенно в клинически значимых генах.

На клеточных линиях COLO829 и HCC1143 пороги MutAnt\_AF  $\geq 0.3$  и MutAnt\_noAF  $\geq 0.1$  значительно увеличили точность поиска соматических мутаций (до  $96.4 \pm 2.1\%$ ) при стабильной чувствительности  $57 \pm 8.5\%$ . Фильтрация по популяционной частоте мало влияла на точность, указывая на доминирующий вклад технических артефактов РНК.

Для клинических данных (TCGA-BRCA n = 24, TCGA-LUAD n = 55 и BostonGene когорты) применение порогов MutAnt увеличило индекс Жаккарда пересечения РНК и ДНК с 0.17-0.29 до 0.53-0.55 (p < 0.0001, FDR), подтверждая эффективность фильтрации. Сравнение инструментов MutAnt, SIFT4G, BayesDel и ClinPred показало сопоставимую эффективность в отборе патогенных мутаций (Рисунок 5).

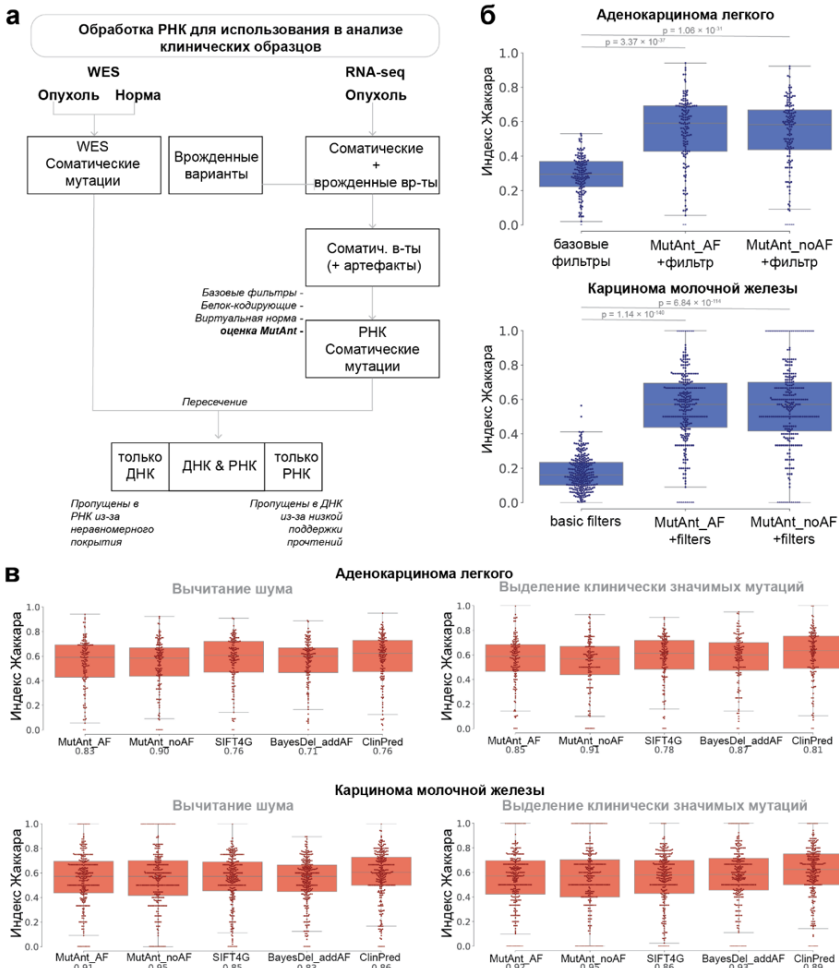


Рисунок 5. Оценка параметров качества поиска и фильтрации мутаций из РНК для клинических образцов пациентов с диагнозами карцинома молочной железы и аденокарцинома легкого. а. Процедура сравнения, б. Влияние фильтрации по оценке патогенности в сравнении с базовыми подходами, в. Сравнение оценки патогенности разных алгоритмов.

Дополнительно, для интеграции данных РНК и ДНК опухоли был оценён VAF и поддержка альтернативного аллеля (AAC) в РНК. Порог VAF = 1.9% и AAC = 7 процентов позволили существенно повысить чувствительность поиска при низкой доле опухолевых клеток: для 10% разведения с  $19.2 \pm 4.9\%$  до  $42.0 \pm 6.1\%$ , для 20% с  $57.0 \pm 3.8\%$  до  $69.8 \pm 3.5\%$ , для 30% с  $65.2 \pm 1.8\%$  до  $76.9 \pm 3.0\%$ , при сохранении высокой точности фильтрации (Рисунок 6).

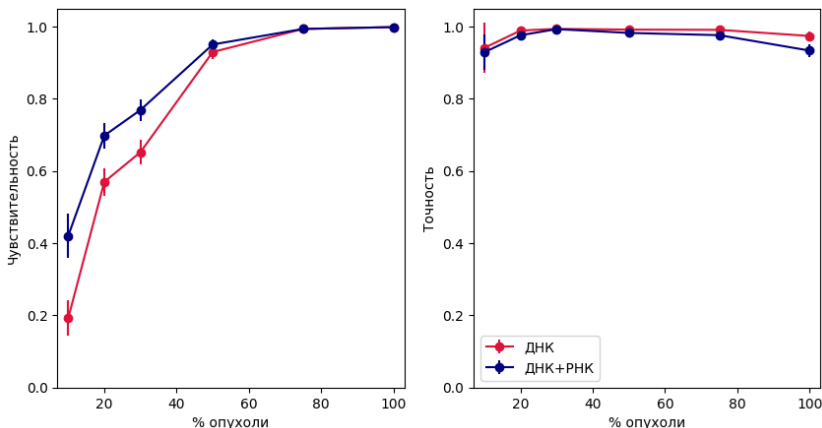


Рисунок 6. Зависимость чувствительности и точности поиска от доли опухоли в образце в режимах поиска мутаций по ДНК опухоли и нормы и по ДНК и РНК опухоли и ДНК нормы.

**Валидация поиска соматических мутаций с низкой аллельной частотой.** Для анализа низкочастотных соматических мутаций во внеклеточной ДНК плазмы использовалась дедупликация по уникальным молекулярным идентификаторам в сочетании с алгоритмом LoFreq и стандартными порогами по аллельной частоте и покрытию. Валидация проводилась на референсных смесях Twist и SeraSeq с двумя техническими репликами для оценки лимита детекции [Методы] – Рисунок 7.

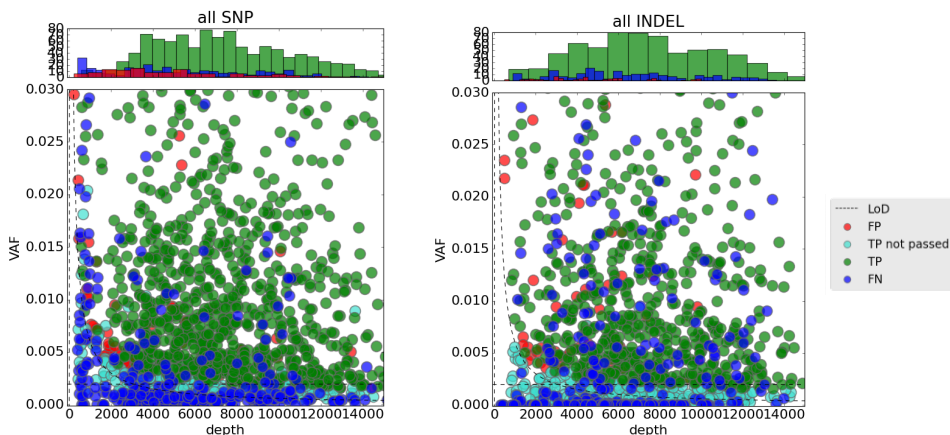


Рисунок 7. Зависимость аллельной частоты от покрытия в референсных смесях Twist и SeraSeq.

При анализе двух реплик лимит детекции точечных мутаций составил 0.2% при покрытии 3500 (Рисунок 8), с чувствительностью 89.8% и точностью 93.5%; для одной реплики лимит увеличивался до 0.35%, чувствительность 89.0%, точность 96.6% (Рисунок 9).

Для коротких делеций и вставок аналогичные значения составили 0.2% (две реплики) и 0.35% (одна реплика), с чувствительностью 81.9% и 79.9% соответственно. Использование двух реплик снижает лимит детекции, но требует удвоенного объёма секвенирования.

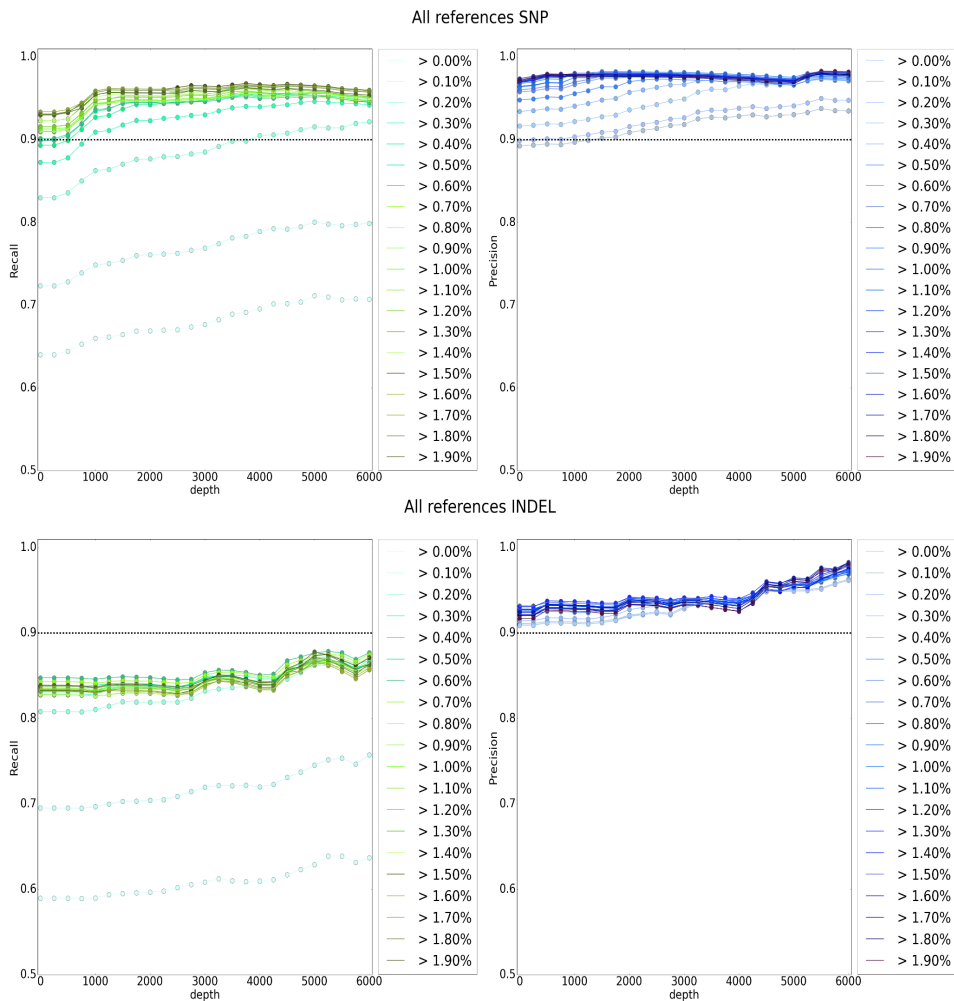
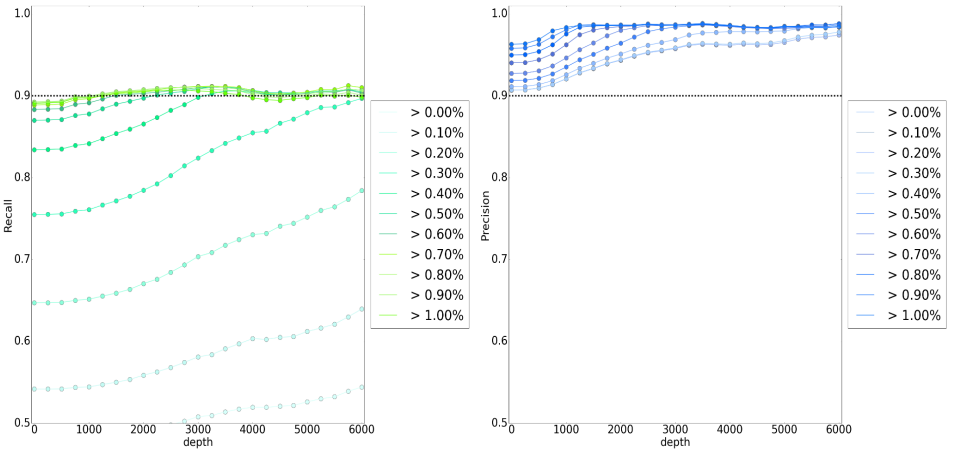


Рисунок 8. Зависимость чувствительности (слева) и точности (справа) поиска мутаций от покрытия для вариантов с разным уровнем аллельной частоты. Алгоритм реализован для совместного поиска мутаций по двум репликам.

### SNP, 1 replicate mode



### INDEL, 1 replicate mode

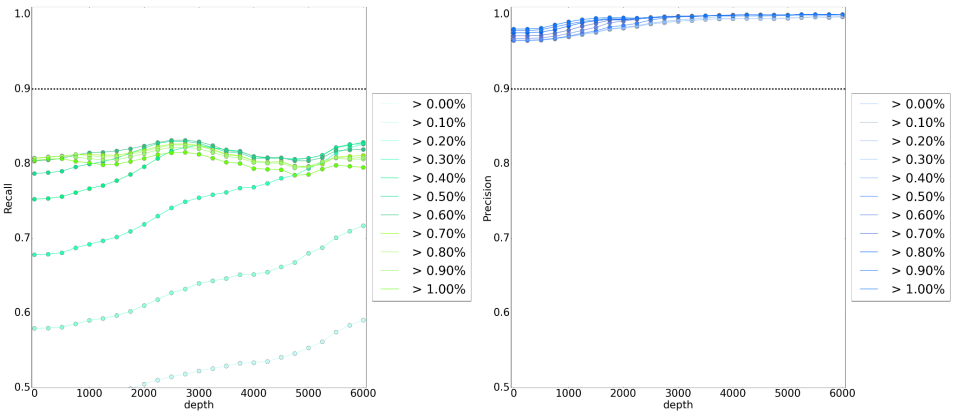


Рисунок 9. Зависимость чувствительности (слева) и точности (справа) поиска мутаций от покрытия для вариантов с разным уровнем аллельной частоты. Алгоритм реализован для анализа одной повторности секвенирования.

Для снижения алгоритм-специфичных ограничений LoFreq был применён ансамбль алгоритмов LoFreq и Mutect2 с фильтрацией артефактов по параметрам покрытия, аллельной частоты и функциональной аннотации с использованием градиентного спуска. На тестовых смесях SeraSeq и Twist лимит детекции достиг 0.25% при покрытии 3500, с чувствительностью 95.6% и точностью 90.5%, превосходя обработку одной реплики LoFreq и близко соответствуя результатам двух реплик.

Применение ансамбля алгоритмов к плазменным образцам пациентов позволило обнаружить все 7/7 мутаций, выявленных в ткани, и 9/10 вариантов по сравнению со стандартной обработкой. Таким образом, интеграция нескольких алгоритмов с оптимизацией порогов позволяет снизить требования к секвенированию без потери качества детекции низкочастотных соматических мутаций.

**Особенности распределения соматических мутаций внеклеточной опухолевой ДНК.** Оптимизированный алгоритм поиска точечных мутаций был применён к ДНК плазмы на таргетной панели из 216 генов для 70 пациентов с парным секвенированием опухолевой ткани. В плазме выявлено 217 соматических замен и коротких делеций/вставок у 44 пациентов, в ткани — 290 вариантов у 57 пациентов. Прямое пересечение результатов составило 23-29%, для горячих точек — до 35%, что указывает на ограниченную сходимость платформ и необходимость независимой валидации [Методы].

Для оценки границ теста и лимита детекции были проанализированы 238 клинических образцов (631 плазменный, 318 нормальных), включая глубину покрытия, число событий и VAF. Всего обнаружено 3936 точечных мутаций, 196 делеций и 35 вставок при средней глубине >5000×. Обнаруженные варианты охватывали широкий диапазон VAF. Частоты мутаций соответствовали ожидаемым по типу опухоли (*TP53*, *KRAS*, *APC*), что подтверждает точность и специфичность теста – Таблица 2.

Таблица 2. Наиболее часто мутированные гены в когорте 238 пациентов с разбиением по диагнозам.

Диагноз	Частоты наиболее часто мутированных генов	Кол-во образцов	% позитивных образцов
Рак молочной железы	<i>TP53</i> :25.83%, <i>NF1</i> :6.67%, <i>POLE</i> :5.83%, <i>PIK3CA</i> :5.83%, <i>PPP2R2A</i> :5.83%	190	63.16
Аденокарцинома поджелудочной железы	<i>TP53</i> :40.0%, <i>KRAS</i> :40.0%, <i>ARID1A</i> :20.0%, <i>CDKN2A</i> :20.0%, <i>ARID1B</i> :20.0%	25	60
Немелкоклеточный рак легкого	<i>TP53</i> :35.0%, <i>NOTCH1</i> :20.0%, <i>KEAP1</i> :20.0%, <i>NF1</i> :15.0%, <i>PGR</i> :15.0%	22	90.91
Опухоль желчного протока	<i>TP53</i> :44.44%, <i>CASR</i> :11.11%, <i>RSPO2</i> :11.11%, <i>GNAS</i> :11.11%, <i>NOTCH1</i> :11.11%	11	81.82
Колоректальный рак	<i>APC</i> :44.44%, <i>TP53</i> :33.33%, <i>TSC2</i> :22.22%, <i>PGR</i> :22.22%, <i>NOTCH3</i> :22.22%	11	81.82
Рак простаты	<i>SPOP</i> :25.0%, <i>FANCD2</i> :25.0%, <i>PPP2RIA</i> :25.0%, <i>SF3B1</i> :25.0%, <i>AR</i> :25.0%	10	80
Опухоль мягких тканей	<i>POLE</i> :40.0%, <i>TP53</i> :40.0%, <i>KDR</i> :40.0%, <i>NRG1</i> :40.0%, <i>PIK3CA</i> :20.0%	8	62.5
Рак почки	<i>TMPRSS2</i> :25.0%, <i>HDAC1</i> :25.0%, <i>RICTOR</i> :25.0%, <i>BAP1</i> :25.0%, <i>MAX</i> :25.0%	5	80
CUP	<i>NRAS</i> :50.0%, <i>BAP1</i> :50.0%, <i>PTCH1</i> :50.0%, <i>KRAS</i> :50.0%, <i>GNAS</i> :50.0%	4	50
Рак мочевыводящих протоков	<i>TP53</i> :100.0%, <i>PIK3CA</i> :100.0%, <i>RIT1</i> :50.0%, <i>PIK3R2</i> :50.0%, <i>SMARCA4</i> :50.0%	2	100

Продолжение Таблицы 2			
Рак пищевода	<i>NOTCH2:50.0%, POLE:50.0%, CDH1:50.0%, NOTCH1:50.0%, PIK3R2:50.0%</i>	3	66.67
Меланома	<i>KDR:100.0%, FLT3:100.0%, ATRX:100.0%, ARID1B:100.0%, POLD1:50.0%</i>	2	100
Рак яичников	<i>TP53:50.0%, PBRM1:50.0%, CTNNA1:50.0%, ESR1:50.0%, NBN:50.0%</i>	2	100
Рак желудка	<i>KDM6A:50.0%, GNAQ:50.0%</i>	2	100
Мелкоклеточный рак легкого	<i>DDR2:50.0%, MAP3K1:50.0%, ERCC6:50.0%, CDH1:50.0%, ATRX:50.0%</i>	3	66.67

Ортогональная проверка с использованием коротких фрагментов (<160 п.о.) показала высокую корреляцию аллельной частоты с полными выравниваниями: R = 0.98 для плазмы и R = 0.91 для ткани, с достоверным повышением VAF на коротких фрагментах (угол наклона 1.16,  $p < 1e-308$ ), подтверждающее опухолевое происхождение детектируемых мутаций.

**Анализ профиля мутаций когорты 3802 пациентов.** На рисунке 10 представлено распределение повторяющихся онкогенных изменений на тепловой карте и показателей TMB и по типам опухолей. Категории обнаруженных изменений, рассмотренные индивидуально в каждой модальности данных, включали: соматические точечные варианты (69.4%), повышенную экспрессию клинически значимых генов (65.4%), амплификации (16.7%), делеции (11.8%), герминальные варианты (5.1%), фьюжены (1.8%), высокую мутационную нагрузку (6.2%) и высокий уровень микросателлитной нестабильности (0.7%).

Комплексный анализ ДНК и РНК секвенирования показал, что как минимум одно клинически значимое драйверное событие выявлено у 97.9% пациентов, включая точечные мутации (69.4%), амплификации (16.7%), делеции (11.8%), фьюжены (1.8%), герминальные варианты (5.1%), повышенную экспрессию генов (65.4%), высокую мутационную нагрузку (6.2%) и микросателлитную нестабильность (0.7%). Наиболее часто встречались *TP53* (>60% опухолей), *KRAS*, в частности известная замена в кодоне G12 (особенно при аденокарциноме поджелудочной железы и колоректальном раке), *PIK3CA* (рак молочной железы и уротелиальная карцинома), *BRAF* p.V600E (меланома, эндокринные опухоли, холангиокарцинома) и *ERBB2* (HER2+ рак молочной железы, рак желудка и пищевода).

Специфические изменения включали слияние *DNAJB1::PRKACA* при фиброламмеллярной карциноме (37.5% слияний в гепатобилиарных опухолях), амплификацию *MDM2* при саркомах (8.6%) и *IDH1* при холангиокарциноме и глиомах. Мутационная нагрузка была наибольшей в меланоме и немелкоклеточном раке легкого, низкой — в саркомах и тимоммах.

Частоты точечных замен, наблюдаемые в исследовании, хорошо согласуются с данными внешних эталонных когорт.

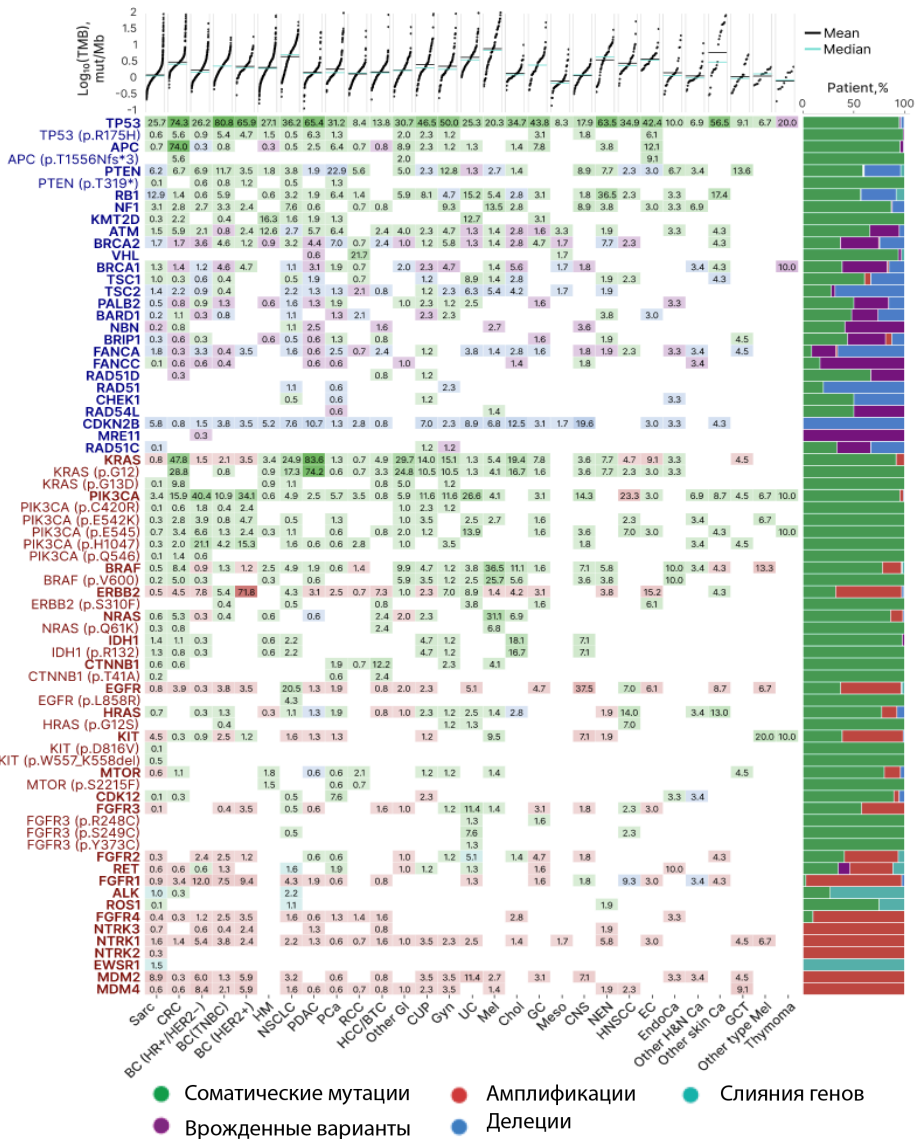


Рисунок 10. Онкоплот повторяющихся генетических изменений в различных типах опухолей, сгруппированных по функции генов и типу варианта. Названия онкосупрессоров выделены синим, а онкогенов – красным цветом. Цветовая шкала отражает преобладающий тип события для каждого гена, тогда как подробное распределение долей различных типов событий представлено на сопутствующей гистограмме справа. В верхней части показаны отсортированные значения  $\log_{10}(\text{TMB})$  для каждого диагноза, а также средние и медианные значения. Красный – амплификации, синий – делеции, зеленый – точечные соматические мутации, фиолетовый – врожденные варианты, цвет морской волны – фьюжены.

**Интеграция данных РНК секвенирования для коррекции частот мутирования генов.** Интеграция данных РНК и ДНК секвенирования в когорте 3802 пациентов повысила чувствительность выявления соматических мутаций, в частности в генах Cancerhotspot, где подтверждение достигало 64.94%, а восстановление мутаций добавляло 7-8% событий (Рисунок 11).

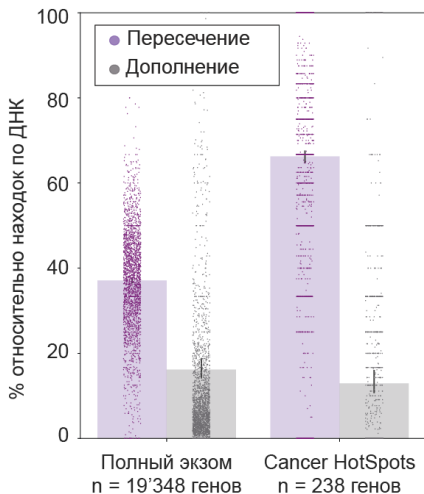


Рисунок 11. Пересечение мутаций в РНК и ДНК, точками показаны индивидуальные образцы пациентов.

Вклад РНК был особенно значим для вариантов с низкой аллельной частотой (<20%, Рисунок 12), типичных для субклональных мутаций или образцов с низким содержанием опухолевой ДНК. За счет дополнительного подтверждения находок в РНК (Рисунок 13) возможно восстановить такие клинически значимые находки с LoE1-4 как *NF1* p.D1644N у пациента с меланомой (LoE 4), для которой при обнаружении рекомендован траметиниб; *KRAS* p.A146T (LoE 1, рекомендован цетуксимаб для колоректального рака); *CHEK2* p.R137\* (LoE 4, рекомендованы PARP ингибиторы).

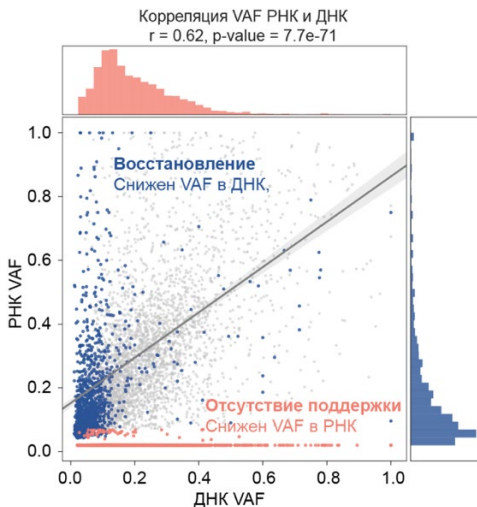


Рисунок 12. Корреляция VAF в данных РНК и ДНК с обозначением групп восстановленных и пропущенных по данным РНК мутаций.

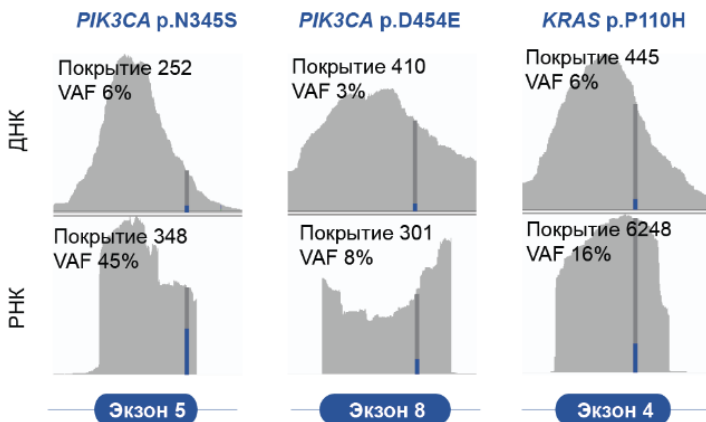


Рисунок 13. Примеры точечных замен с большей поддержкой в РНК, позволяющей обнаружить мутацию.

Также подход позволял корректировать ложноположительные находки в гомологичных регионах, например в *PIK3CA* (Рисунок 14).

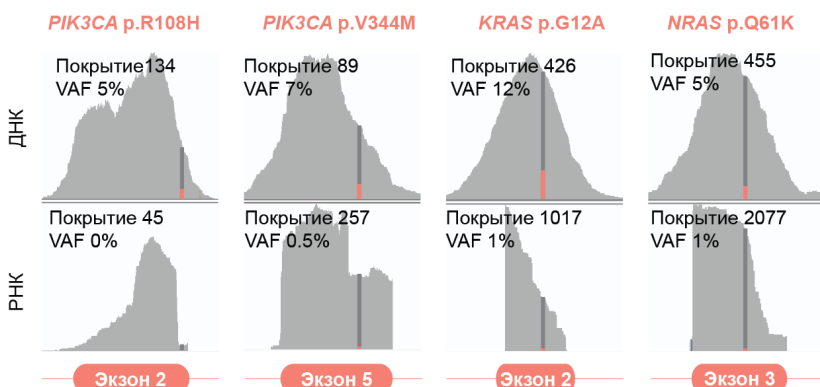


Рисунок 14. Примеры точечных замен с большей поддержкой в ДНК и минимальной экспрессией.

В проспективном анализе 211 пациентов клинически значимые биомаркеры LOE 1-4 выявлялись у 77.7%, из них 38 пациентов (18%) получили рекомендации терапии, а 30 (14.2%) — таргетное лечение на основании геномного профиля, включая 10 случаев, где ключевую роль сыграли данные РНК. Применение интегрированного подхода повышало частоту клинического ответа (65.2% против 30.6%,  $p = 0.0037$ ).

**Применение тестирования ДНК и РНК в клинических испытаниях.** В проспективном анализе 211 пациентов (исследования BIGR и BEGIN) клинически значимые биомаркеры (LoE 1-4) были выявлены у 77.7% пациентов, при этом у 38 пациентов (18.0%) они были признаны клинически релевантными, что обеспечило дополнительные рекомендации по терапии. В итоге 30 пациентов (14.2%) получили таргетную терапию, из которых у 10 пациентов (33%) ключевую роль сыграли находки, выявленные по данным РНК-секвенирования; среди пациентов, уже получавших лечение ( $\geq 2$  линий терапии), РНК

позволила назначить терапию в 10.4% случаев, недоступных при анализе только ДНК. Частота клинического ответа была значительно выше у пациентов, получавших терапию, обоснованную геномными находками (65.2% против 30.6%,  $p = 0.0037$ ), что подтверждает клиническую эффективность использования геномного профилирования при выборе терапии.

**Объединение данных тканевых и жидкостных биопсий.** Для 100 пациентов, имеющих данные ДНК ткани, РНК ткани и внеклеточной ДНК плазмы, прямое пересечение клинически значимых мутаций составило 27.1% между ДНК ткани и плазмы и 82.2% между РНК и ДНК ткани.

Интеграция данных трёх модальностей позволила увеличить число находок в генах CancerHotspot с 107 до 133 событий (+24.3%) и расширила выявление клинически значимых мутаций – дополнительные находки получили еще 7 пациентов (общее число увеличилось с 51 до 58 пациентов). Корреляция аллельных частот была выше между РНК и ДНК ткани ( $R = 0.59$ ) по сравнению с тканевой и жидкостной биопсиями ( $R = 0.41$ ), что отражает более согласованное выявление вариантов внутри ткани и основное расширение профиля мутаций за счет дополнительных сайтов, не доступных при биопсии ткани.

Интеграция данных выявляет гетерогенность опухолевого состава и позволяет восстанавливать мутации, недоступные при анализе одной платформы, включая варианты *KRAS* и *CDKN2A*, а также расширяет профиль мутаций за счёт циркулирующей ДНК плазмы, что важно для оценки метастатических очагов.

**Уточнение функциональной значимости соматических мутаций за счет объединения с данными сплайсинга.** На основе анализа сплайсированных прочтений были выявлены 8 ранее неаннотированных вариантов: *RBI* с.1389+2\_1389+3insG (HR+ HER2– рак молочной железы, пропуск экзона 14, потеря функции белка) – Рисунок 15, *MSH2* p.K122N с.366G>T (мелкоклеточный рак лёгкого, пропуск экзона 2, NMD), *TP53* с.97-27\_97-6del (анapластическая крупноклеточная лимфома, пропуск экзона 4 и активация скрытого сайта, LOF), *SMARCA4* p.K74= с.222G>A (аденокарцинома лёгкого, скрытый акцептор, NMD), *FANCA* p.L244F с.732G>C (нейроэндокринная карцинома, удаление 25 нуклеотидов, NMD), *CDHI* p.G571R с.1711G>C (HR+ HER2– рак молочной железы, пропуск экзона 11, NMD), *NFI* p.V1093A с.3278T>C (миксофибросаркома, удаление 40 нуклеотидов, NMD) и *NFI* p.T954= с.2862T>A (недифференцированная плеоморфная саркома, пропуск экзона 22, NMD). Эти находки демонстрируют способность интеграции РНК и ДНК уточнять функциональное влияние вариантов и выявлять клинически значимые мутации, недоступные при анализе только ДНК.

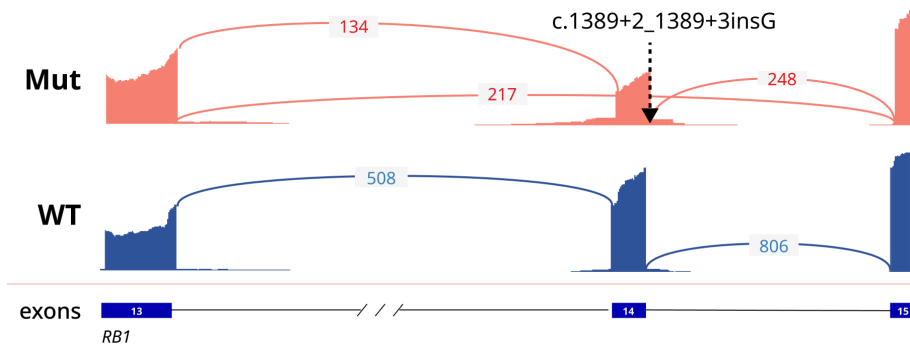


Рисунок 15. Пример альтернативного сплайсинга транскрипта *RBI* гена в результате активации мутации в сайте сплайсинга - *RBI* с.1389+2\_1389+3insG.

## ЗАКЛЮЧЕНИЕ

В работе представлен интегрированный анализ ДНК и РНК секвенирования из FFPE тканей и плазмы крови для профилирования пациентов с онкологическими заболеваниями. Основное внимание уделялось выявлению соматических мутаций и событий альтернативного сплайсинга для оценки прогностических и терапевтических биомаркеров. Анализ проведён на 3802 образцах тканевых биопсий и 238 образцах плазмы крови.

Для верификации методов поиска мутаций из FFPE тканей были собраны референсные стандарты точечных мутаций и коротких инделов на основе пяти модельных клеточных линий с учётом копийности и клеточного состава. Общий набор включал 3042 варианта по клинически значимым генам, что позволяло оценивать чувствительность алгоритмов и калибровать показатели мутационной нагрузки.

Применение методов фильтрации и интеграции данных позволило выявить высокочастотные мутации в ключевых генах: *TP53* (до 40% в различных когортах), *KRAS*, *CDKN2A*, *APC*, *FGFR1-4* и *NOTCH1-4*. В плазме крови обнаружено дополнительно 66% точечных мутаций и 31% коротких делеций/вставок по сравнению с тканевой биопсией, включая низкочастотные варианты с аллельной частотой <1%. Интеграция данных ДНК и РНК позволила выявить до 8% дополнительных клинически значимых мутаций, повысив информативность анализа даже при низком покрытии секвенирования регионов интереса ДНК.

Анализ событий альтернативного сплайсинга показал значимые эффекты мутаций в генах *RBI*, *MSH2*, *TP53* и *SMARCA4*, включая как качественные изменения (появление/утрата сплайсинговых событий), так и количественные изменения экспрессии экзонов. Для пациентов с неполными находками ДНК-секвенирования использование РНК позволило восстановить важные мутации, включая мутации с эффектом GOF в *PIK3CA*, ранее не выявленные на уровне ДНК.

Совместный анализ тканевых и жидкостных биопсий уточнил мутационный профиль у пациентов: число пациентов с выявленными клинически значимыми событиями увеличилось с 51 до 58 из 100 пациентов, в том числе за счёт мутаций в генах *KRAS*, *CDKN2A*.

Таким образом, интеграция ДНК и РНК секвенирования обеспечивает выявление дополнительных терапевтически значимых мутаций, уточняет частоту мутаций в ключевых генах и способствует более точному выбору геномно-обоснованной терапии.

## ВЫВОДЫ

1. Разработаны и валидированы подходы к интегрированному анализу данных экзомного и транскриптомного секвенирования FFPE-образцов, обеспечивающие подавление артефактов фиксации и получение достоверных молекулярных профилей опухолей; показана применимость референсных наборов соматических мутаций, отражающих варибельность копийности и клеточного состава, для стандартизации и валидации NGS-тестов.
2. Разработаны и валидированы подходы к детекции соматических мутаций с низкой аллельной частотой ( $\leq 1\%$ ) в циркулирующей опухолевой ДНК плазмы, обеспечивающие повышение чувствительности и специфичности за счёт оптимизации алгоритмов анализа и подавления технических артефактов.
3. Интеграция геномных и транскриптомных данных повышает точность оценки частот мутаций и выявление клинически значимых альтераций, включая мутации в генах-мишенях таргетной терапии, за счёт учета экспрессии и функциональной релевантности вариантов.
4. Объединенный анализ данных тканевых и жидкостных биопсий расширяет спектр выявляемых соматических мутаций и повышает информативность молекулярного

профилирования опухолей, отражая их гетерогенность и улучшая выявление клинически значимых событий.

5. В клиническом испытании показано, что интеграция геномных и транскриптомных данных и персонализированный выбор терапии увеличивает эффективность лечения. Дополнительно, интеграция данных ДНК- и РНК-секвенирования позволяет уточнять функциональную значимость вариантов, включая VUS, за счёт анализа транскриптомных эффектов, в том числе нарушений сплайсинга.

### СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Yudina A, Tazearslan C, Baisangurov A, Nuzhdina E, Lauziere K, Segodin V, Podsvirova S, Starikov S, Chasse M, Shaposhnikov K, Kaneunyenye L, Klimchuk O, Kuzkina N, English N, Khagai G, Sookiasian D, Shafranskaya D, Fernandez D, Lozinsky Y, Sobolev A, Abdou M, Turova P, Chernyshov K, Efremov A, Andrewes S, Feinberg A, McKenna B, Brown JH, Love A, Curran J, Lennerz J, Bagaev A. Clinical and analytical validation of a combined RNA and DNA exome assay across a large tumor cohort. *Commun Med (Lond)*. 2025 Jun 16;5(1):236. <https://doi.org/10.1038/s43856-025-00934-3>.
2. Sarachakov A, Yudina A, Svekolkin V, Parfenenkova A, Spektor M, Oshchepkova P, Pak M, Antysheva Z, Balabanian L, Brown JH, Feinberg M, Fowler N, Bagaev A. MutAnt: mutation annotation tool predicts deleteriousness of missense mutations and improves mutation calling from transcriptomics. *Hum Genet*. 2025 Dec;144(11-12):1245-1268. <https://doi.org/10.1007/s00439-025-02775-5>.
3. Юдина А.С. Интеграция NGS данных комплексного геномного профилирования солидной и жидкостной биопсии для улучшения молекулярной диагностики онкологических заболеваний. Доклады НАН РА. 2024;124(3-4):42-50. <https://doi.org/10.54503/0321-1339-2024.124.3-4-42>.
4. Yudina A, Tyurin K, Chernyavskij D, Oshchepkova P, Danchurova A, Nemchaninova T, Kosmin A, Fedorova A, Moshareva M, Sagaradze G, Bagaev A. Integrating RNA and DNA Analysis to Enhance Detection, Annotation, and Interpretation of Somatic Variants and Active Driver Mutations in Tumors. Association for Molecular Pathology 2025 Annual Meeting Abstracts. *J Mol Diagn* 2025, 27 Suppl: S1. Abstract I020
5. Yudina A, Starikov S, Efremov A, Sookiasian D, Nuzhdina E, Podsvirova S, Chasse M, Conroy T, English N, Love A, Tabakov D, Baisangurov A, Tazearslan C, Fowler N, Bagaev A. Comparative analysis of cfDNA liquid biopsy and tumorbased next-generation sequencing (NGS) approaches. ESMO Congress 2023, *Annals of Oncology* 2023, v. 34, Issue S2, Abstract 1215P. <https://doi.org/10.1016/j.annonc.2023.09.2304>
6. Yudina A, Bagaev A, Baisangurov A, Love A, Brown J, Khagai G, Kuzkina N, Segodin V, Starikov S, Shafranskaya D, Nuzhdina E, Podsvirova S, Sobolev A, Efremov A, Tazearslan C, Chasse M, Fernandez D, Curran J, Morgan K, Kaneunyenye L, English N, Villarreal A, McKenna B, Fowler N. Developing Bespoke Genomic Reference Standards for Validating Somatic Mutation Detection in Clinical Whole-Exome Sequencing Assays. Association for Molecular Pathology 2023 Annual Meeting Abstracts. *J Mol Diagn* 2023, 25: S1 Abstract I004
7. Yudina A, Efremov A, Sookiasian D, Nuzhdina E, Podsvirova S, Chasse M, Conroy T, English N, Starikov S, Klimchuk O, Tabakov D, Love A, Suryamohan K, Baisangurov A, Tazearslan C, Fowler N, Bagaev A. Analytical validation of a liquid biopsy test using cell-free circulating tumor DNA for mutational profiling [abstract]. In: Proceedings of the American Association for Cancer Research Annual Meeting 2023; Part 1 (Regular and Invited Abstracts); 2023 Apr 14-19; Orlando, FL. Philadelphia (PA): AACR; *Cancer Res* 2023;83(7\_Suppl):Abstract nr 1042.

## **Յուդինա Անաստասիա Սերգեյի**

### **Ուռուցքների կլինիկորեն նշանակալի օնկոգեն մոլեկուլային խանգարումների հայտնաբերում գենոմային և տրանսկրիպտոմային տվյալների հիման վրա**

#### **Ամփոփագիր**

**Բանալի բառեր՝** ճշգրիտ ուռուցքաբանություն, համապարփակ գենոմային պրոֆիլավորում, տրանսկրիպտոմիկա, սոմատիկ մուտացիաներ, ազատ շրջանառվող ԴՆԹ, հեղուկ բիոպսիա

Քաղցկեղը շարունակում է մնալ աշխարհում հիվանդացության և մահացության հիմնական պատճառներից մեկը, ինչը պայմանավորում է ճշգրիտ մոլեկուլային ախտորոշման և անհատականացված թերապիաների մշակման կարևորությունը: Ուռուցքների համապարփակ գենոմային պրոֆիլավորումը դարձել է ժամանակակից ուռուցքաբանության առանցքային գործիքներից մեկը՝ կլինիկորեն նշանակալի կենսամարկերների հայտնաբերման համար:

Հյուսվածքային բիոպսիաներից ստացված ԴՆԹ-ի վերլուծությունը շարունակում է դիտարկվել որպես ստանդարտ մոտեցում, քանի որ այն ապահովում է հետազոտվող նյութի կայունություն և խնդրո առարկա գենոմային տեղամասերի համաչափ ծածկույթ: Սակայն այս մոտեցումը լիարժեքորեն չի արտացոլում հայտնաբերված մուտացիաների ֆունկցիոնալ հետևանքները, որոնց գնահատումը կարող է լրացվել տրանսկրիպտոմային սեքվենավորման տվյալներով: Հավելյալ բարդություններ են առաջանում նաև գենոմում առկա հոմոլոգ և ցածր բարդության հաջորդականությունների մեկնաբանման ընթացքում, ինչը ևս կարող է հստակեցվել ԴՆԹ և ՌՆԹ տվյալների ինտեգրման միջոցով:

Վերջին տարիներին արյան պլազմայից շրջանառվող ուռուցքային ԴՆԹ-ի (շուԴՆԹ) վերլուծությունը դարձել է ակտիվ զարգացող հետազոտական ուղղություն: Այս մոտեցումը հնարավորություն է տալիս վերահսկել ռեմիսիայի ընթացքը և հիվանդության կրկնությունը, ինչպես նաև կիրառելի է այն դեպքերում, երբ հյուսվածքային նյութը հասանելի չէ: Բացի այդ, շուԴՆԹ-ի կիրառումը կարող է բարձրացնել թերապևտիկ նշանակություն ունեցող կենսամարկերների հայտնաբերման զգայունությունը՝ համեմատած միայն առաջնային ուռուցքային հյուսվածքի հետազոտության հետ, քանի որ մետաստատիկ օջախներից ԴՆԹ-ի բեկորները ներթափանցում են արյան շրջանառություն: Միևնույն ժամանակ, մուտքային ԴՆԹ-ի ցածր կոնցենտրացիան (5-10 նգ/մլ պլազմայում) և ուռուցքային բաղադրիչի փոքր մասնաբաժինը (1-10%) պահանջում են սեքվենավորման բարձր խորություն և գերխորը (ultra-deep) սեքվենավորման ժամանակ առաջացող արտեֆակտների ճնշմանն ուղղված ալգորիթմների կիրառում:

ԴՆԹ և ՌՆԹ սեքվենավորման տվյալների ինտեգրումը ներկայումս դիտարկվում է որպես ախտորոշման ճշգրտության բարձրացման հեռանկարային մոտեցում: Տրանսկրիպտոմային տվյալների ներառումը հնարավորություն է տալիս հաստատել մուտանտ ալելների էքսպրեսիան, հստակեցնել տարբերակների ֆունկցիոնալ նշանակությունը, հայտնաբերել սփլայսինգի փոփոխությունները և բարելավել ցածր հաճախականությամբ ալելային մուտացիաների նկատմամբ զգայունությունը: Բացի

այդ, շուրջ վերլուծությունը կարող է բացահայտել կլինիկորեն նշանակալի փոփոխություններ, որոնք ուռուցքի հետերոգենության պատճառով կարող են չարձանագրվել հյուսվածքային նմուշներում:

Չնայած տեխնոլոգիական առաջընթացին, ներկայումս գոյություն չունի ամբողջական էքզոմի և տրանսկրիպտոմի սեքվենավորման վրա հիմնված ինտեգրված թեստերի անալիտիկ վալիդացման միասնական ստանդարտ, որը կիրառելի կլինի ինչպես հյուսվածքային նմուշների, այնպես էլ շրջանառվող նուկլեինաթթուների համար: Գործող ուղեցույցները հիմնականում միտված են թիրախային պանելների կիրառմանը, մինչդեռ ամբողջական էքզոմի վրա հիմնված հետազոտությունների վերաբերյալ հանձնարարականները դեռևս սահմանափակ են՝ հատկապես օրթոգոնալ վալիդացման մեթոդաբանության և սպեցիֆիկության գնահատման առումով:

Այս աշխատանքը ցույց է տալիս, որ ԴՆԹ և ՌՆԹ սեքվենավորման տվյալների ինտեգրումն ընդլայնում է կլինիկորեն կիրառելի մոլեկուլային փոփոխությունների հայտնաբերման հնարավորությունները: Բուժառուների մեծ խմբի ( $n = 3803$ ) հետազոտության ընթացքում ՌՆԹ տվյալները թույլ են տվել հաստատել սոմատիկ մուտացիաների զգալի մասը, ինչպես նաև հայտնաբերել հավելյալ տարբերակներ, որոնք ԴՆԹ-ի մակարդակում չէին արձանագրվել, հատկապես կլինիկորեն նշանակալի գեներում: Կարևոր է նշել, որ տրանսկրիպտոմային վերլուծությունը բարելավել է տարբերակների մեկնաբանումը՝ հաստատելով դրանց էքսպրեսիան, բացահայտելով չարտահայտված գենոմային փոփոխությունները և ի հայտ բերելով սփլայսինգի հետ կապված ֆունկցիոնալ ազդեցությունները՝ դրանով իսկ հստակեցնելով հայտնաբերված մուտացիաների կենսաբանական և կլինիկական նշանակությունը:

Բացի այդ, հյուսվածքային և հեղուկ բիոպսիաների համակցված վերլուծությունը տրամադրել է փոխլրացնող տեղեկություններ ուռուցքի հետերոգենության վերաբերյալ: Այն դեպքում, երբ հյուսվածքից և պլազմայից ստացված փոփոխությունների ուղիղ համընկնումը պահպանվել է ցածր մակարդակի վրա, շուրջ վերլուծությունը հնարավորություն է տվել հայտնաբերել կլինիկորեն նշանակալի հավելյալ տարբերակներ մետաստատիկ օջախներից: Այս տվյալների ինտեգրումը մեծացրել է կլինիկորեն կիրառելի արդյունքների ընդհանուր քանակը և բարելավել բուժառուների ընտրությունը թիրախային թերապիաների համար: Կլինիկական խմբերում այս մոտեցումը հանգեցրել է լրացուցիչ թերապևտիկ առաջարկությունների ձևակերպմանը և նպաստել գենոմային տվյալների հիման վրա բուժում ստացող անձանց շրջանում արձագանքի ավելի բարձր ցուցանիշների արձանագրմանը:

Այսպիսով, ամբողջական էքզոմի և տրանսկրիպտոմի սեքվենավորման տվյալները շրջանառվող նուկլեինաթթուների վերլուծության հետ համակցող մոտեցումների մշակումը և անալիտիկ վալիդացումը արդիական և դեռևս թերի ուսումնասիրված խնդիր է: Նման մեթոդների ներդրումը նպատակ ունի բարձրացնել մոլեկուլային ախտորոշման զգայունությունը, ապահովել սոմատիկ փոփոխությունների առավել ճշգրիտ մեկնաբանումը և նպաստել արդյունավետ թիրախային թերապիաների նշանակմանը:

## Identification of Clinically Relevant Oncogenic Molecular Alterations in Tumors Based on Genomic and Transcriptomic Data

### Summary

**Keywords:** Precision Oncology; Comprehensive Genomic Profiling; Transcriptomics; Somatic mutations; cfDNA; Liquid Biopsy

Cancer remains one of the leading causes of morbidity and mortality worldwide, highlighting the critical importance of developing accurate molecular diagnostics and personalized therapies [Siegel et al., 2025]. Comprehensive genomic profiling of tumors has become a key tool in modern oncology for identifying biomarkers [ASCO, 2019].

DNA analysis from tissue biopsies remains the standard, providing material stability and uniform coverage of regions of interest; however, this approach does not reflect the functional consequences of detected mutations, which can be complemented by transcriptome sequencing data [Hussen et al., 2022]. Additional challenges arise due to homologous and low-complexity sequences in the genome [Mandelker et al., 2016], the interpretation of which can also be refined by integrating DNA and RNA data.

In recent years, the analysis of circulating tumor DNA (ctDNA) from blood plasma has emerged as an active research direction. This minimally invasive approach enables monitoring of remission and disease recurrence and is useful when tissue material is unavailable. Moreover, using ctDNA can improve the sensitivity of detecting therapeutically relevant biomarkers compared to analyzing only primary tumor tissue, as fragments of DNA from metastatic sites enter the bloodstream [Brannon et al., 2021]. At the same time, the low concentration of input DNA (5-10 ng/mL plasma) and the small tumor fraction (1-10%) require increased sequencing depth and the application of algorithms to suppress artifacts that arise during ultra-deep sequencing [Maruzani et al., 2024].

Integration of DNA and RNA sequencing data is considered a promising approach to increase diagnostic accuracy. Incorporating transcriptome data allows confirmation of mutant allele expression, clarification of variant functional significance, detection of splicing alterations, and improved sensitivity for low allele frequency mutations [Wilkerson et al., 2014]. Additionally, ctDNA analysis can reveal clinically significant alterations that may be missed in tissue samples due to tumor heterogeneity [Park et al., 2021].

Despite technological advances, there is currently no unified standard for analytical validation of integrated tests based on whole-exome and transcriptome sequencing, applicable to both tissue samples and circulating nucleic acids. Existing guidelines are mostly oriented toward targeted panels [Frampton et al., 2013], while recommendations for whole-exome tests remain limited regarding orthogonal validation methodology and specificity assessment [Shah et al., 2024].

This work demonstrates that integration of DNA and RNA sequencing data expands the detection of clinically actionable alterations. Across large patient cohort (n = 3803), RNA data

enabled confirmation of a substantial fraction of somatic mutations and recovery of additional variants not detected at the DNA level, particularly in clinically relevant genes. Importantly, transcriptome analysis improved interpretation of variants by confirming their expression, identifying non-expressed genomic alterations, and revealing splicing-associated functional effects, thereby refining the biological and clinical relevance of detected mutations.

Furthermore, the combined analysis of tissue and liquid biopsies provided complementary information on tumor heterogeneity. While direct overlap between tissue-derived and plasma-derived mutations remained limited, ctDNA analysis identified additional clinically relevant variants, alterations originating from metastatic sites. Integration of these data increased the overall number of actionable findings and improved patient stratification for targeted therapies. In clinical cohorts, this approach led to additional therapy recommendations and contributed to higher response rates in patients receiving genomically matched treatments compared to standard therapy.

Thus, the development and analytical validation of approaches that combine whole-exome and transcriptome sequencing data with circulating nucleic acid analysis is a relevant and underexplored task. Implementation of such methods aims to increase the sensitivity of molecular diagnostics, enable accurate interpretation of somatic alterations, and support the administration of effective targeted therapies.

A handwritten signature in blue ink, appearing to be 'N. A. H.', located in the lower right quadrant of the page.