

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ
«НАЦИОНАЛЬНЫЙ МЕДИЦИНСКИЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
РАДИОЛОГИИ» МИНИСТЕРСТВА ЗДРАВООХРАНЕНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

На правах рукописи

Шегай Петр Викторович

**СОЗДАНИЕ ФАРМАКОГЕНОМНОЙ ПЛАТФОРМЫ ДЛЯ ПОДДЕРЖКИ
ПРИНЯТИЯ ПЕРСОНАЛИЗИРОВАННЫХ ВРАЧЕБНЫХ РЕШЕНИЙ В
ОНКОЛОГИИ**

3.1.6. Онкология, лучевая терапия

1.5.8. Математическая биология, биоинформатика

**ДИССЕРТАЦИЯ
на соискание ученой степени
доктора медицинских наук**

**Научный консультант:
академик РАН, академик
РАО, д.м.н., профессор
Каприн Андрей Дмитриевич**

Москва – 2025 г.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	6
ГЛАВА 1 - Обзор литературы.....	19
1.1. Применение эволюционного подхода в онкологии	19
1.2. Понятие персонализированной и прецизионной медицины в мировой клинической практике.....	22
1.3. Анализ существующих методов прецизионной медицины и индивидуализированного подбора терапевтических препаратов для онкологических больных	24
1.4. Трижды негативный рак молочной железы как модель для испытания эффективности платформ прецизионной медицины.....	31
1.4.1. Молекулярно-генетическая классификация трижды негативного рака молочной железы на основании анализа экспрессии генов	32
1.4.2. Генетические особенности трижды негативного рака молочной железы на основании данных секвенирования нового поколения	35
1.4.3. Особенности диагностики трижды негативного рака молочной железы.....	37
1.4.4. Подходы к терапии трижды негативного рака молочной железы	39
1.5. Рак поджелудочной железы как модель для испытания эффективности платформ прецизионной медицины	45
1.5.1. Генетические особенности рака поджелудочной железы.....	47
1.5.2. Особенности диагностики рака поджелудочной железы.....	54
1.5.3. Подходы к терапии рака поджелудочной железы	58
ГЛАВА 2 - Материалы и методы исследования	61
2.1. Мета-анализ клинических исследований персонализированного лечения онкологических заболеваний в сравнении с неперсонализированным лечением	61
2.1.1. Стратегия поиска и отбора исследований	61
2.1.2. Извлечение и категоризация данных	62
2.1.3. Синтез данных и статистическая обработка	63
2.2. Экзомное секвенирование	63
2.2.1. Разработка адаптированного протокола пробоподготовки и секвенирования	63
2.2.2. Методология сравнения адаптированного протокола пробоподготовки и секвенирования с референсным протоколом	64
2.3. Разработка фармакогеномной платформы и системы поддержки принятия врачебных решений для подбора персонализированного лекарственного лечения онкологических заболеваний	66
2.3.1. Архитектура фармакогеномной платформы и системы поддержки принятия врачебных решений	66
2.3.2. Методология создания озера данных.....	68

2.3.2.1. Выбор решений для управления озером данных	68
2.3.2.2. Методология наполнения озера данных	69
2.3.2.2.1 Регистрационные и клинико-фенотипические данные больных.....	70
2.3.2.2.2. Лабораторные метрики пробоподготовки и секвенирования биологического материала больных.....	70
2.3.2.2.3. Нуклеотидная последовательность ДНК больных (необработанные данные секвенирования в формате FASTQ и их метрики качества)	72
2.3.2.2.4. Выявленные геномные варианты, показатели микросателлитной нестабильности, опухолевой мутационной нагрузки, вариации числа копий и данные о рекомендациях персонализированного лекарственного лечения	76
2.3.2.2.5. Метаданные, метрики качества работы герминального и соматического биоинформатических конвейеров, общие сводные таблицы.....	79
2.3.2.2.6. Информация из открытых мировых баз данных.....	80
2.3.3. Разработка биоинформатических конвейеров для обработки данных полноэкзомного секвенирования	90
2.3.3.1. Определение требований к биоинформатическим конвейерам	90
2.3.3.2. Разработка биоинформатических конвейеров для обработки данных полноэкзомного секвенирования.....	92
2.3.4. Оценка показателей эффективности выявления геномных вариантов созданной фармакогеномной платформой.....	101
2.3.5. Создание системы поддержки принятия врачебных решений для подбора персонализированного лекарственного лечения онкологических заболеваний на базе фармакогеномной платформы	102
2.4. Общая характеристика анализируемого биологического материала и когорт больных раком поджелудочной железы и трижды негативным раком молочной железы, биологический материал которых подвергался анализу	106
2.4.1. Критерии включения и исключения биологического материала больных раком поджелудочной железы и трижды негативным раком молочной железы	106
2.4.2. Характеристика когорты больных раком поджелудочной железы, биологический материал которых подвергался анализу	107
2.4.3. Характеристика когорты больных трижды негативным раком молочной железы, биологический материал которых подвергался анализу	109
2.5. Морфологическое и иммуногистохимическое исследование образцов опухолевой ткани	111
ГЛАВА 3 - Результаты исследования	113
3.1. Мета-анализ клинических исследований персонализированного лечения онкологических заболеваний в сравнении с неперсонализированным лечением	113
3.1.1. Исследования, включенные в мета-анализ.....	113
3.1.2. Сравнение исходов лечения онкологических заболеваний	114
3.2. Адаптированный протокол пробоподготовки и секвенирования.....	118
3.2.1 Описание адаптированного протокола пробоподготовки и секвенирования.....	118

3.2.2 Сравнение адаптированного протокола пробоподготовки и секвенирования с референсным протоколом	121
3.3. Фармакогеномная платформа и система поддержки принятия врачебных решений для подбора персонализированного лекарственного лечения онкологических заболеваний	123
3.3.1. Структура разработанной фармакогеномной платформы и её взаимодействие с системой поддержки принятия врачебных решений	123
3.3.2. Выбранный инструмент управления озером данных	126
3.3.3. Наполнение озера данных	131
3.3.3.1. Структура созданного озера данных	131
3.3.3.2 Наполнение собственными данными	133
3.3.3.3. Наполнение опубликованными данными из открытых мировых источников....	141
3.3.4. Анализ существующих биоинформатических конвейеров	147
3.3.5. Разработанные биоинформатические конвейеры для определения герминального и соматического геномного профиля	149
3.3.6. Интерфейс рабочей зоны управления биоинформатическими расчетами.....	161
3.3.7. Система поддержки принятия врачебных решений для подбора персонализированного лекарственного лечения онкологических заболеваний.....	165
3.4. Морфологические и иммуногистохимические характеристики образцов опухолевой ткани	172
3.4.1. Морфологические и иммуногистохимические характеристики образцов опухолевой ткани больных раком поджелудочной железы	172
3.4.2. Морфологические и иммуногистохимические характеристики образцов опухолевой ткани больных трижды негативным раком молочной железы.....	177
3.5. Эффективность выявления геномных вариантов созданной фармакогеномной платформой	182
3.6. Геномный профиль при раке поджелудочной железы и трижды негативном раке молочной железы.....	188
3.6.1. Геномный профиль при раке поджелудочной железы.....	188
3.6.1.1. Герминальный геномный профиль.....	188
3.6.1.2. Соматический геномный профиль.....	193
3.6.2. Геномный профиль при трижды негативном раке молочной железы.....	197
3.6.2.1. Герминальный геномный профиль.....	197
3.6.2.2. Соматический геномный профиль.....	202
3.7. Рекомендации персонализированного лекарственного лечения по итогам анализа биологического материала больных раком поджелудочной железы и трижды негативным раком молочной железы.....	208
ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ	246
ЗАКЛЮЧЕНИЕ	262
ВЫВОДЫ.....	271
ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ.....	273

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ	274
СПИСОК ЛИТЕРАТУРЫ.....	277
ПРИЛОЖЕНИЯ.....	304
Приложение 1. Клинико-фенотипические сведения о больных трижды-негативным раком молочной железы, содержащиеся в своде.....	304
Приложение 2. Метрики качества FASTQ файлов.....	306
Приложение 3. Собираемые метрики качества работы биоинформатического конвейера	308
Приложение 4. Набор параметров, собирающих информацию о патогенных и вероятно-патогенных вариантах, выявленных в результате работы герминального биоинформатического конвейера	311
Приложение 5. Колонки в аннотации герминального VCF файла	315
Приложение 6. Колонки в аннотации парного VCF файла	320
Приложение 7. Учебно-методическое пособие “Молекулярно-генетические исследования в онкологии: учебно-методическое пособие” (г. Обнинск, 2023 г.).....	322
Приложение 8. Патенты и свидетельства РИД.....	323

ВВЕДЕНИЕ

Актуальность темы исследования

На сегодняшний день показатели выживаемости при целом ряде онкологических заболеваний значительно выше, чем 20 лет назад, однако количество выявленных случаев злокачественных новообразований (ЗНО) продолжает неуклонно расти. По данным Всемирной организации здравоохранения онкологические заболевания остаются одной из ведущих причин смерти в мире. Так, только в 2022 г. ЗНО унесли жизни 9,7 млн человек, из них почти 250 тыс. - в России [3, 122].

Такие онкологические заболевания, как рак поджелудочной железы (РПЖ) и трижды негативный рак молочной железы (ТНRMЖ), характеризуются плохим прогнозом ввиду агрессивного характера опухоли и ограниченного спектра действенных лечебных стратегий. Вместе с тем, по состоянию на 2022 год в Российской Федерации на учете находилось более 20 тыс. больных РПЖ, а также более 750 тыс. больных раком молочной железы (РМЖ) [3], при этом на долю ТНRMЖ приходится порядка 10-20% от всех случаев РМЖ [9].

Начиная с середины прошлого века лекарственное лечение ЗНО было основано преимущественно на применении цитостатических/цитотоксических агентов (химиотерапии) с общим подходом к лечению каждого заболевания. Действие большинства этих химиопрепаратов направлено на активно делящиеся клетки организма, что является причиной их побочных эффектов, например, нейтропении и анемии [262]. В то же время, стало очевидным, что одни и те же препараты, применяемые в равных дозах, демонстрируют различную эффективность даже в отношении опухолей, не различающихся по гистогенезу, локализации и степени дифференцировки. Это явилось основанием для разработки новой стратегии подбора терапии, основанной, прежде всего, на молекулярных характеристиках опухоли. Данному подходу способствовали результаты исследований молекулярной генетики злокачественных опухолей, а также механизмов первичной и приобретенной химиорезистентности. Американский транснациональный институт исследования данных в области здравоохранения (IQVIA Institute for Human Data Science) обозначил поиск новых терапевтических комбинаций для преодоления лекарственной резистентности в числе новых глобальных трендов исследований и разработок в онкологии.

Концепция персонализированного подбора лекарственной терапии, основанной на генетических особенностях пациента, появилась в конце прошлого столетия и всего за одно десятилетие эволюционировала в новое интенсивно растущее направление - фармакогеномику [76, 167]. Решающую роль в прорыве, достигнутом в фармакогеномике за последние десятилетия, сыграло развитие методов высокопроизводительного секвенирования. Масштабные

международные проекты, такие как Международный консорциум генома рака (International Cancer Genome Consortium) и Атлас генома рака (The Cancer Genome Atlas), позволили сформировать подробные каталоги т.н. «соматических драйверных мутаций», ассоциированных с инвазивностью и метастазированием опухолей [16, 133, 368]. Благодаря этому стало возможным развитие трех направлений онкогеномики: диагностического - выявление ЗНО и, прежде всего, выявление минимальной остаточной болезни; предиктивного - предсказание клинического исхода; и персонализированного терапевтического – наиболее высокотехнологичного направления, использующего анализ генетических нарушений для разработки и создания новых препаратов направленного действия. В результате изучения генетических особенностей опухолевых клеток в 90-х годах XX века были созданы первые таргетные препараты, способные “превратить смертельный рак в хроническое контролируемое заболевание”. Эффективное применение таргетных и иммунотерапевтических препаратов непосредственно связано с выявлением герминальных и соматических молекулярно-генетических характеристик больных ЗНО. Тем не менее, несмотря на значительные достижения в изучении молекулярной биологии опухолей, в настоящее время в клинической практике терапевтические решения в значительной степени основаны на патоморфологических (анатомическое расположение, стадия, гистология), а не на молекулярных характеристиках (мутационный профиль, соматические вариации генома, экспрессия генов и т.д.) опухолевых клеток.

Ряд клинических центров России располагает инфраструктурой и достаточными ресурсами для внедрения технологии высокопроизводительного секвенирования в рамках государственных программ прецизионной медицины и высокотехнологичной помощи. В то же время существует ряд трудностей, без преодоления которых невозможно осуществить внедрение геномных технологий в рутинную клиническую практику. К таким трудностям можно отнести: (I) - недостаток подготовленного персонала, опыт и квалификация которого должны охватывать несколько дисциплин: патоморфологию, клиническую биоинформатику, медицинскую генетику; (II) - отсутствие единой доступной и понятной для практикующего онколога централизованной базы данных мутаций, выявляемых в геноме опухоли, и клинических особенностей течения заболевания (в настоящее время данная информация представлена множеством отдельных источников в разнородном формате); (III) - отсутствие специальной IT-инфраструктуры для хранения и обработки значительного объема накопленных к настоящему времени молекулярно-генетических данных.

В эпоху геномной медицины для подбора эффективной персонализированной лечебной стратегии необходима работа с большими и постоянно пополняемыми массивами информации [105, 214]. В связи с этим приобретает актуальность использование фармакогеномных платформ,

способных к непрерывному обновлению и интеграции как структурированных, так и неструктурированных данных и обеспечивающих эффективную работу с большими объемами геномной, клинико-фенотипической и справочной информации. Фармакогеномная платформа может быть охарактеризована как комплекс инструментов для интеграции, хранения и эффективной работы с большим массивом различных данных, включая результаты генетических тестирований, данные необходимые для аннотирования геномных вариантов и для обеспечения функций по подбору эффективного персонализированного лечения [277]. Фармакогеномная платформа способна обеспечить автоматизацию учёта геномных особенностей каждого пациента, что способствует выбору наиболее эффективного, а в некоторых случаях и безопасного лечения. Это снижает риск побочных эффектов и увеличивает вероятность успешного исхода терапии. Интеграция индивидуальных геномных характеристик больных и справочных данных об ассоциациях молекулярно-генетических биомаркеров (далее - биомаркеры) с чувствительностью к лекарственным препаратам в фармакогеномные платформы способствует более широкому внедрению персонализированного подхода к лечению ЗНО. Использование фармакогеномных данных также может помочь в идентификации подгрупп пациентов, которые наиболее вероятно ответят на новое лечение, тем самым ускоряя процесс клинических испытаний и сокращая их стоимость [99].

В мировой практике описаны примеры создания фармакогеномных платформ, аккумулирующих массивы генетической информации для проведения последующего анализа и подбора персонализированной терапии. В 2023 году коллективом авторов из Национального онкологического центра Республики Корея предложена фармакогеномная платформа для выявления основных механизмов резистентности у пациентов с распространенным или рефрактерным раком легкого и подбора персонализированного лечения. Скрининг чувствительности к лекарственным средствам проводился на основе краткосрочных культур опухолевых клеток, выделенных из новообразований пациента. На основе исследования генетических вариантов и экспрессии генов определялись восприимчивые к конкретному лекарственному средству кандидаты. Важной особенностью разработки стала возможность учитывать не только молекулярно-генетические характеристики, но и микроокружение опухоли [271].

Несмотря на наличие зарубежных решений, актуально создание отечественной фармакогеномной платформы, интегрирующей данные о больных ЗНО из российской популяции. Применение таких платформ способствуют более эффективному внедрению персонализированной геномной медицины в клиническую практику, при этом существует высокий потенциал их применения и для фундаментальных исследований. Современная онкологическая наука уделяет большое внимание изучению молекулярно-генетических

особенностей больных из определенных популяций. Было показано, что среди больных аденокарциномой поджелудочной железы мутации в гене *TP53* более распространены в марокканской популяции по сравнению с египетской [292]. Агрегация геномных данных больных ЗНО из различных популяций способствует выявлению особенностей развития онкологических заболеваний, а также характеризует частоту встречаемости тех или иных геномных вариантов. Аккумуляция в рамках фармакогеномной платформы информации о распространенности молекулярно-генетических биомаркеров при различных типах ЗНО у пациентов из российской популяции обладает значительной научной и прикладной значимостью. Этот процесс способствует созданию обширной базы данных, которая может использоваться для разработки персонализированных подходов к диагностике и лечению, повышению эффективности противоопухолевой терапии и оптимизации стратегии ведения пациентов.

Наполнение фармакогеномных платформ может не ограничиваться молекулярно-генетическими данными. Добавление клинично-фенотипической информации также вносит весомый вклад в понимание особенностей ЗНО в определенных популяциях. Оказание медицинской помощи большому количеству больных в условиях ФГБУ «НМИЦ Радиологии» Минздрава России предоставляет уникальные возможности для аккумуляции существенных объемов клинично-фенотипических данных и проведения ретроспективных исследований.

Объем функций, реализуемых фармакогеномными платформами, может различаться в зависимости от их структуры. Одним из подходов является интеграция в платформу биоинформатических конвейеров для выявления молекулярно-генетических характеристик больных ЗНО. Вместе с тем, биоинформатическая обработка данных секвенирования является комплексной задачей, эффективность реализации которой может быть повышена с помощью передовых технологий искусственного интеллекта, в частности - машинного обучения. Искусственный интеллект — это более широкая концепция, которая охватывает создание систем, способных выполнять задачи, требующие интеллекта, таких как распознавание речи, принятие решений, обработка изображений, решение проблем и многое другое. Искусственный интеллект включает в себя любые методы и технологии, которые позволяют компьютеру имитировать человеческий интеллект. Это могут быть как системы на основе простых правил, так и сложные нейронные сети. Машинное обучение — это подмножество искусственного интеллекта, которое фокусируется на разработке алгоритмов, способных обучаться на данных и улучшать свои результаты без необходимости явно программировать их для решения конкретной задачи. В отличие от традиционного программирования, где разработчики пишут код для выполнения каждой задачи, в машинном обучении системы обучаются выявлять закономерности в данных и делать предсказания на основе них. Машинное обучение использует статистические методы для

построения моделей, которые могут адаптироваться и улучшаться со временем. В машинном обучении ключевым аспектом является обучение на данных и создание прогнозных моделей.

Выявление геномных вариантов является одним из этапов биоинформатического анализа сведений, при этом он может быть ассоциирован с такими сложностями как корректное различение найденных вариантов и артефактов секвенирования, а также некорректных выравниваний прочтений [79]. Начало применения машинного обучения, а именно сверточной нейронной сети, для выявления геномных вариантов способствовало дальнейшему совершенствованию их анализа [282]. На сегодняшний день опубликован массив работ, демонстрирующих эффективность применения программы по выявлению герминальных геномных вариантов DeepVariant, использующей технологии машинного обучения [17, 340, 395]. Включение программ на основе машинного обучения в биоинформатические конвейеры, являющиеся частью фармакогеномной платформы, представляет перспективный подход к обеспечению высокого уровня анализа данных секвенирования.

Фармакогеномные платформы могут аккумулировать большие массивы различных данных, способствующих изучению ЗНО и подбору персонализированного лекарственного лечения. Вместе с тем, для обеспечения более эффективного применения указанных данных в реальной клинической практике необходимы специализированные инструменты, обеспечивающие взаимодействие лечащего врача с фармакогеномной платформой. Кроме того, необходимы инструменты, позволяющие генерировать отчеты о подборе персонализированного лечения ЗНО, отвечающие требованиям понятности и удобства для практикующего клинициста. Названные функции на настоящий момент способны реализовывать системы поддержки принятия врачебных решений (СППВР).

СППВР предназначены для улучшения качества диагностики и повышения эффективности терапии за счет цифровой трансформации нозологических знаний, клинико-диагностической и другой медицинской информации о пациентах. Традиционная СППВР состоит из программного обеспечения, предназначенного для прямой помощи в принятии клинических решений, при этом характеристики отдельного пациента сопоставляются с компьютеризированной клинической базой данных, а оценки или рекомендации для конкретного пациента в итоге предоставляются врачу для принятия решения.

Актуальность применения СППВР, учитывающих геномные особенности онкологических больных, подтверждается их широким применением в мировой и отечественной практике. Так, например, система FoundationOne[®]CDx (Foundation Medicine Inc, США) одобрена Управлением по санитарному надзору за качеством пищевых продуктов и медикаментов США (англ. U.S. Food and Drug Administration, FDA) для подбора персонализированной терапии солидных опухолей на основе выявленного индивидуального геномного профиля опухоли больного. Вместе с тем

нужно отметить высокую стоимость использования данной системы, а также ее ограниченную доступность в Российской Федерации.

На настоящий момент в Российской Федерации имеется доступ к более чем 20 различным СППВР, при этом немногие из них предусматривают возможность подбора персонализированного лечения онкологических заболеваний. Среди отечественных СППВР целесообразно отметить платформу Oncobox, функционирующую на основе алгоритма количественной оценки степени активации каждого исследуемого молекулярного пути с целью определения эффективности различных противоопухолевых препаратов [330]. Однако платформа Oncobox предусматривает подбор таргетной терапии из ограниченного перечня, состоящего из 160 препаратов, не охватывая, таким образом, все современные возможности. Функции по подбору персонализированного лечения онкологических заболеваний также могут быть реализованы с использованием генетического теста Atlas Solo Комплекс Плюс [403]. В указанном тесте производится секвенирование панели из 160 генов, иммуногистохимические (ИГХ) исследования и флуоресцентная гибридизация *in situ* (англ. fluorescent *in situ* hybridization, FISH) для анализа эффективности различных таргетных противоопухолевых препаратов, однако ограниченный список исследуемых генов сужает диапазон подбора лекарственных препаратов.

Один из наиболее авторитетных исследователей СППВР проф. Дж. Ошерофф (Jerome A. Osheroff) еще более 10 лет назад так сформулировал пять ключевых правил работы СППВР (“CDS Five Rights”) для успешного решения клинической задачи: 1 - правильная информация (подбор источников и обновление баз данных, основываясь на методах доказательной медицины); 2 - правильный пользователь (ориентация СППВР на всю цепочку участников лечебного процесса: лечащий врач - консультирующие специалисты - пациент); 3 - правильный формат вмешательства СППВР (предупреждение угроз, информационная поддержка, рекомендации лечащему врачу); 4 - правильный канал взаимодействия (электронная информационная медицинская система, электронная медицинская карта и т.д.); 5 - правильный момент для интеграции в лечебный процесс (принятие решения о назначении терапии, момент постановки диагноза и др.) [263]. Поскольку доступ к актуальной доказательной информации является одним из основных условий успешного функционирования СППВР, обеспечение работы системы на базе фармакогеномной платформы обладает преимуществами, улучшающими качество и эффективность медицинской помощи.

Ввиду изложенного выше, для полномасштабного внедрения результатов геномных исследований в медицинскую онкологическую практику приобретает актуальность разработка отечественной фармакогеномной платформы, в рамках которой будут объединены сразу несколько важнейших компонентов:

- биоинформатические конвейеры для определения герминального и соматического геномного профиля, включающие инструменты на основе технологий машинного обучения;
- система интеграции, централизованного хранения и структурирования значительных объемов собственных геномных, клинико-фенотипических и иных данных;
- пополняемое хранилище справочных данных, позволяющих устанавливать функциональную значимость выявляемых геномных вариантов, а также их влияние на чувствительность к широкому спектру противоопухолевых лекарственных препаратов, основанное на интеграции сведений из открытых мировых баз данных.

Вместе с тем, необходимо обеспечить работу СППВР для подбора эффективного персонализированного лекарственного лечения ЗНО в форме рекомендаций, получающей необходимые данные из фармакогеномной платформы, и включающей автоматизированную систему генерации отчетов. Интеграция больших массивов данных в отечественную фармакогеномную платформу позволит биоинформатическим конвейерам осуществлять аннотирование геномных вариантов, а СППВР - формировать рекомендации персонализированного лекарственного лечения в автономном режиме, при отсутствии подключения к сети Интернет.

Всё вышесказанное и предопределило выполнение данной работы.

Цель исследования

Разработать фармакогеномную платформу и систему поддержки принятия врачебных решений для подбора персонализированного лекарственного лечения онкологических заболеваний на основе анализа данных высокопроизводительного полноэкзомного секвенирования биологического материала больных с применением методов машинного обучения.

Задачи исследования

1. Провести мета-анализ клинических исследований медикаментозного лечения различных злокачественных новообразований, реализованных с применением персонализации на основе индивидуальных геномных характеристик больных либо без персонализации, для определения клинической эффективности указанных подходов;

2. Разработать адаптированный протокол пробоподготовки и секвенирования экзомных ДНК-библиотек и получить сырые данные высокопроизводительного секвенирования из парных образцов крови и опухолевого материала больных раком поджелудочной железы и трижды негативным раком молочной железы из российской популяции;

3. Разработать биоинформатические конвейеры обработки данных высокопроизводительного секвенирования, включающие инструменты на основе технологий машинного обучения, для выявления геномных вариантов, показателей опухолевой мутационной нагрузки и микросателлитной нестабильности;

4. Обработать данные полноэкзомного секвенирования с применением разработанных биоинформатических конвейеров и описать ландшафт геномных вариантов больных раком поджелудочной железы и трижды негативным раком молочной железы для расширения карты знаний по российской популяции;

5. Создать единое локальное пополняемое “озеро данных” (data lake) с включением результатов выявления геномных вариантов, клинико-фенотипической информации, информации из открытых мировых баз данных, необходимой для аннотирования геномных вариантов и формирования рекомендаций персонализированного лечения, и иных сведений;

6. Спроектировать и разработать фармакогеномную платформу на основе озера данных и биоинформатических конвейеров и провести оценку её эффективности в сравнении с эталонными решениями на примере трижды негативного рака молочной железы путем сопоставительного анализа результатов выявления геномных вариантов;

7. Разработать систему поддержки принятия врачебных решений, обеспечивающую взаимодействие врача с фармакогеномной платформой и формирующую рекомендации персонализированного лекарственного лечения онкологических заболеваний, и оценить её пригодность на примере данных полноэкзомного секвенирования образцов рака поджелудочной железы и трижды негативного рака молочной железы.

Научная новизна

Создана отечественная фармакогеномная платформа, позволяющая интегрировать большие объемы геномных и клинико-фенотипических данных больных ЗНО из российской популяции, а также мировые данные об ассоциации биомаркеров с чувствительностью к противоопухолевой терапии.

Разработаны оригинальные биоинформатические конвейеры для определения геномного профиля больных ЗНО, включающие инструменты на основе технологий машинного обучения.

Получены и обработаны биоинформатическими конвейерами, включающими инструменты на основе технологий машинного обучения, данные полноэкзомного секвенирования парных образцов опухоли и крови для российских когорт больных РПЖ (n=40) и ТНРМЖ (n=100).

В фармакогеномную платформу добавлен массив геномных, клинико-фенотипических и иных данных, позволивший расширить каталог распространенности геномных вариантов и иных показателей при РПЖ и ТНРМЖ в российской популяции в научно-исследовательских целях.

Создана новая СППВР для подбора персонализированного лекарственного лечения онкологических заболеваний на основе биомаркеров больных, получающая информацию об ассоциациях биомаркеров с чувствительностью к противоопухолевым лекарственным препаратам из озера данных, находящегося в составе фармакогеномной платформы.

Таким образом, в результате настоящей работы реализована цифровая трансформация персонализированной медицины по направлению фармакогеномики в онкологии.

Практическая значимость исследования

Разработанная отечественная фармакогеномная платформа позволяет:

- с высокой точностью анализировать данные высокопроизводительного секвенирования и определять геномные профили больных при помощи биоинформатических конвейеров, включающих инструменты на основе технологий машинного обучения.

- аккумулировать большие объемы геномных, клинико-фенотипических и иных данных, которые в дальнейшем могут быть использованы для анализа распространенности геномных вариантов у российской популяции больных различными ЗНО, поиска предиктивных и прогностических биомаркеров, а также стратификации больных.

Разработанная СППВР позволяет:

- анализировать интегрированные в фармакогеномную платформу мировые данные об ассоциации выявленных биомаркеров с чувствительностью к различным таргетным, иммунотерапевтическим и химиотерапевтическим препаратам и формировать рекомендации персонализированного лечения ЗНО.

По итогам проведенной работы получены патент на изобретение, 2 свидетельства о регистрации программы для ЭВМ и 10 свидетельств о регистрации баз данных.

На основании полученных результатов разработано учебно-методическое пособие «Молекулярно-генетические исследования в онкологии: учебно-методическое пособие» (г. Обнинск, ФГБУ «НМИЦ радиологии» Минздрава России, 2023. 82 с. Серия «Библиотека врача-онколога»). В данном пособии приведена информация о молекулярно-генетических исследованиях, входящих в рекомендации Минздрава России, а также о дополнительных исследованиях, которые могут способствовать назначению персонализированного лечения ЗНО.

Методология и методы исследования

Настоящая работа представляет собой биоинформатическое и клиническое ретроспективное исследование. В рамках диссертационной работы был проанализирован биологический материал (образцы опухолевой ткани и периферической крови) и собраны клинико-фенотипические данные больных РПЖ и ТНРМЖ, проходивших лечение в ФГБУ «НМИЦ Радиологии» Минздрава России в период с 2016 по 2023 гг.

При выполнении диссертационной работы использованы следующие методы исследования:

1. Высокопроизводительное полноэкзомное секвенирование парных образцов крови и опухолевой ткани больных РПЖ и ТНРМЖ, выполненное на высокопроизводительном секвенаторе MGISEQ-G400 (MGI, Китай);

2. Биоинформатическая обработка данных высокопроизводительного секвенирования, в том числе с применением инструментов на основе технологий машинного обучения, для выявления геномных вариантов, показателей опухолевой мутационной нагрузки и микросателлитной нестабильности у больных РПЖ и ТНРМЖ. Обработка данных велась на языках Bash, Python, R, многие плагины сделаны на языках Perl и C++.

3. Разработка оригинальной СППВР для подбора персонализированного лекарственного лечения онкологических заболеваний с использованием языков программирования JavaScript, NodeJS, R, Python, Bash, системы управления пакетами conda, nrm, Linux (Debian 10) в качестве основной операционной системы для разработки и выполнения программ, а также открытых программных пакетов NodeJS, Playwright, TAPES и Chromium.

4. Гистологическое и ИГХ исследования. ИГХ окрашивание опухолевого материала больных РПЖ с использованием антител против CK7, ЕМА, СА125, СЕА осуществляли для верификации и определения гистологического типа опухоли, против Her Par-1 - для дифференциальной диагностики при исследовании метастатических узлов РПЖ в печени, против HER2/neu - для определения чувствительности опухоли к таргетной терапии анти-HER2/neu. ИГХ окрашивание опухолевого материала больных ТНРМЖ проводили с применением антител против рецепторов эстрогенов, рецепторов прогестерона и HER2/neu для подтверждения трижды негативного подтипа РМЖ и с антителами против Ki-67 - для оценки митотической активности.

5. Статистическая обработка данных в формате *.csv с использованием облачного сервиса Google Collab и Visual Studio Code с визуализацией результатов исследования в виде графиков типа “Хит-мэп” (англ. heat map). Для осуществления вычислительных операций при выполнении статистического анализа использовали язык программирования Python-4.0 на базе Google Compute Engine с объемом оперативной памяти 34 Гб, библиотеки Pandas, Numpy, Matplotlib,

Seaborn. Для пропорционального мета-анализа использовали программы RStudio, применяя язык программирования R и пакеты «Meta», «Metafor» и «Tidyverse».

6. Общенаучный метод анализа, сравнения и систематизации информации.

Основные положения, выносимые на защиту

1. Применение персонализации на основе геномных характеристик больных при подборе лекарственного лечения онкологических заболеваний ассоциировано с улучшением клинических исходов за счет своевременной рекомендации наиболее эффективных лекарственных препаратов.

2. Разработанный адаптированный протокол пробоподготовки и секвенирования, а также биоинформатические конвейеры обеспечивают получение корректных данных высокопроизводительного секвенирования, не уступая эталонным референсным решениям.

3. Разработанная фармакогеномная платформа предоставляет возможность формировать ландшафты геномных вариантов, отражающие герминальный и соматический геномный профиль при онкологических заболеваниях у больных из российской популяции.

4. Созданная фармакогеномная платформа содержит информацию из различных мировых баз данных, что позволяет проводить аннотирование выявленных геномных вариантов и предоставлять СППВР информацию об ассоциациях биомаркеров с чувствительностью к противоопухолевым лекарственным препаратам.

5. Интеграция больших объемов геномных, клинико-фенотипических и иных данных о больных из российской популяции в разработанную фармакогеномную платформу способствует формированию комплексной базы знаний об онкологических заболеваниях, которая может быть использована для дальнейших исследований;

6. Созданная СППВР обеспечивает персонализированный подход к лечению онкологических заболеваний и позволяет формировать рекомендации по подбору таргетного, иммунотерапевтического и химиотерапевтического лечения с учетом индивидуальных геномных характеристик больного.

Внедрение результатов работы в практику

Созданная фармакогеномная платформа и функционирующая на её основе СППВР внедрены в диагностический и лечебный процесс отдела трансляционной онкологии МНИОИ им. П.А. Герцена - филиала ФГБУ «НМИЦ Радиологии» Минздрава России и отдела молекулярной генетики МРНЦ им. А.Ф. Цыба - филиала ФГБУ «НМИЦ Радиологии» Минздрава России (акт внедрения от 02.04.2024 № 01-12-958).

Результаты диссертационной работы внедрены в процесс обучения студентов 1-2 курсов на кафедре биологии им. академика В.Н. Ярыгина Института биологии и патологии человека ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России (справка о внедрении от 06.12.2024).

Разработанное и опубликованное учебно-методическое пособие “Молекулярно-генетические исследования в онкологии: учебно-методическое пособие”. Обнинск: ФГБУ «НМИЦ радиологии» Минздрава России, 2023. 82 с. (Серия «Библиотека врача-онколога») используется при повышении квалификации и первичной подготовке врачей-онкологов.

Степень достоверности и апробация результатов

Высокая степень достоверности результатов, полученных в настоящей диссертационной работе, обусловлена большим объемом экспериментального материала, полученного с использованием геномных и постгеномных методов исследований. Результаты работы подтверждены статистически и согласуются с результатами проведенного автором мета-анализа на более чем 7000 больных. С учетом мировой практики, размер выборки соответствует задачам описания молекулярно-генетического ландшафта при данных ЗНО и апробации разработанной СППВР [156, 190, 349]. Достоверность результатов биоинформатического анализа, полученных с применением разработанных биоинформатических конвейеров, подтверждается сравнением с популярными мировыми аналогами.

Основные результаты настоящего диссертационного исследования были доложены и обсуждены на следующих конференциях и конгрессах: Научная конференция NGS в медицинской генетике (Суздаль, 2021), Научная конференция III Научно-практическая онлайн конференция “Парадигмы лекарственной терапии у онкологических больных” (Москва, 2021), OpenBio-2021 (Москва, 2021), III Научно-практическая онлайн конференция “Парадигмы лекарственной терапии у онкологических больных” (Москва, 2021), 10-я Всероссийская научно-практическая конференция по геномному секвенированию и редактированию (Москва, 2022), Первая ежегодная конференция московского общества медицинских генетиков с международным участием (Москва, 2022), Конгресс с международным участием “Молекулярная диагностика и биобезопасность - 2022” (Москва, 2022), V юбилейный международный форум онкологии и радиотерапии (Москва, 2022), “Геномное секвенирование и редактирование. XI Всероссийская научная конференция” (Москва, 2023), V Юбилейная научно-практическая конференция “Парадигмы лекарственной терапии у онкологических больных” (Москва, 2023), Конгресс с международным участием “Молекулярная диагностика и биобезопасность — 2024” (Москва, 2024), Научно-практическая конференция “Экспериментальная и клиническая онкология” (Москва, 2024).

Апробация диссертационной работы состоялась на заседании межотделенческой конференции отделения трансляционной онкологии, отделения персонифицированных биотехнологических методов лечения, отделения прогноза эффективности консервативного лечения, отделения экспериментальной фармакологии и токсикологии, отдела онкоморфологии, отделения онкологии и реконструктивно-пластической хирургии молочной железы и кожи МНИОИ им. П.А.Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России, центра инновационных радиологических и регенеративных технологий ФГБУ «НМИЦ радиологии» Минздрава России, лаборатории анализа постгеномных данных ФГБНУ «Научно-исследовательский институт биомедицинской химии» им. В.Н. Ореховича РАН, кафедры онкологии и рентгенодиагностики им. В.П. Харченко Медицинского института ФГАОУ ВО «РУДН им. Патриса Лумумбы» Министерства науки и высшего образования Российской Федерации, протокол № 10 от 23.04.2025 г.

Публикация результатов исследования

По теме настоящей диссертационной работы опубликованы 34 работы, в том числе 10 статей, опубликованных в зарубежных рецензируемых журналах первого и второго квартиля (Q1/Q2), и 1 статья в зарубежном рецензируемом журнале третьего квартиля (Q3), индексируемых в базе Scopus, и 10 статей в российских рецензируемых журналах, рекомендованных ВАК РФ, получены 1 патент на изобретение, 2 свидетельства о регистрации программы для ЭВМ и 10 свидетельств о регистрации баз данных.

Структура и объем диссертации

Настоящая диссертация состоит из введения, главы “обзор литературы”, главы “материалы и методы исследования”, главы “собственные результаты”, обсуждения результатов исследования, заключения, выводов, практических рекомендаций, а также 8 приложений. Диссертация изложена на 335 страницах машинописного текста, содержит 34 таблицы, 67 рисунков. Список цитируемой литературы содержит 406 источников.

ГЛАВА 1 - Обзор литературы

1.1. Применение эволюционного подхода в онкологии

За последние 20 лет подход к лечению онкологических заболеваний перешел от выбора универсальной терапии для определенного типа ЗНО на подбор терапии для определенного пациента с учетом его индивидуальных особенностей (персонализированная и прецизионная медицина). Проведение такого лечения онкологических заболеваний требует понимания как механизмов развития ЗНО, так и механизмов противоопухолевой защиты. Применение на практике данных подходов может использовать методы сравнительной геномики и эволюционной онкологии [134]. Эволюционная онкология — это подраздел эволюционной медицины, который использует методы и принципы эволюционной биологии для изучения, предотвращения и лечения онкологических заболеваний человека [6, 147].

Несмотря на то, что по данным Национального института рака США за последние 40 лет было зарегистрировано общее снижение показателей смертности онкологических больных, наблюдается непрерывный и устойчивый рост заболеваемости всеми типами рака. Снижение показателей смертности связано с постоянным прогрессом в области медицины и фармакологии, а также улучшением методов диагностики, способствующих более раннему выявлению злокачественных новообразований, и внедрением в терапию более эффективных лекарств и терапевтических подходов. В частности, за последние 40 лет произошло снижение показателей смертности от рака предстательной железы, РМЖ, толстой кишки, мочевого пузыря, рака шейки матки и желудка. Почти неизменные показатели смертности наблюдаются при меланоме, РПЖ и пищевода, для которых еще нет действительно эффективных фармакологических методов лечения. Увеличение смертности наблюдается от карциномы печени, в то время как рак легких и неходжкинская лимфома демонстрируют переменную тенденцию на протяжении многих лет с небольшим увеличением показателей смертности с 1975 по 1995 гг. и общим снижением показателей смертности, зарегистрированных за последние 20 лет [118].

Рост новых случаев онкологических заболеваний связан не только с улучшением методов выявления ЗНО, но также и увеличением продолжительности жизни. Кроме того, несмотря на развитие современных методов лечения, диагностики и профилактики онкологических заболеваний, показатели выживаемости по многим типам рака в России и в мире остаются невысокими [3]. Все это говорит о необходимости более полного изучения процессов, которые лежат в основе развития ЗНО.

В некропсийном материале крупных млекопитающих наблюдается более низкий уровень ЗНО в сравнении с человеческой популяцией. Это может быть обусловлено более эффективными

механизмами противоопухолевой защиты, поскольку повреждение или недостаточная активность именно этих механизмов приводит к возникновению опухолей [29]. Можно выделить две основные группы генов, обеспечивающих противоопухолевую защиту: гены-регуляторы клеточного цикла и гены репарации ДНК. Гены-регуляторы клеточного цикла представляют собой более консервативные последовательности, их повреждение приводит к значительному повышению риска развития ЗНО за счет нарушения контроля клеточного цикла (контрольная точка G1/S) и неограниченного деления клеток [30, 128]. Гены репарации ДНК более разнообразны и узкоспециализированы, обеспечивают репарацию ДНК на различных уровнях в течение всего жизненного цикла клетки [63, 175].

У разных отрядов млекопитающих обнаружены разнообразные молекулярные механизмы противоопухолевой защиты, что является проявлением параллельной эволюции и может быть связано с адаптацией к разнообразным условиям среды. Для хоботных характерно увеличение числа копий антионкогенов, у голубого землекопа противоопухолевая защита обеспечивается балансом про- и анти-апоптогенов, в то время как у рукокрылых подобный механизм связан с комбинацией увеличения копий гена *p53* и увеличением синтеза новых форм мРНК, которые обеспечивают элиминацию мутантных мРНК. Более эффективное восстановление ДНК и более высокая чувствительность к повреждению ДНК могут быть ключом к низкой частоте опухолей у крупных и долгоживущих млекопитающих [350]. Кроме того, у некоторых животных не наблюдается увеличения риска развития ЗНО при увеличении их размера или росте продолжительности жизни, что может быть связано с низкой вероятностью появления соматических мутаций [69]. Например, такие животные как слоны, крупные морские млекопитающие и представители рукокрылых с высокой продолжительностью жизни имеют среднюю частоту развития ЗНО, сопоставимую с другими позвоночными [58]. Это может быть связано и с повышением эффективности механизмов противоопухолевой защиты [28, 138].

Сорок лет назад Питер Ноуэлл впервые официально описал рак как эволюционный процесс, связанный с клональной эволюцией популяций опухолевых клеток [257]. С тех пор эта гипотеза получила подтверждение при открытии внутриопухолевой субклональной гетерогенности и клональной селекции при разных типах рака [33, 62, 135]. Признание фундаментальной эволюционной природы рака и представление об опухолях как динамично развивающихся и адаптирующихся структурах требует переоценки возможностей и ограничений, которые это дает прецизионной медицине рака.

В настоящее время рак рассматривается как динамический эволюционный процесс, определяемый сложными взаимодействиями между клональными вариантами, которые включают как конкуренцию, так и сотрудничество [59, 145, 386]. Раковые клоны образуют гетерогенные популяции, которые включают несколько субпопуляций, постоянно

развивающихся в конкуренции [145, 386]. Клональная гетерогенность играет ключевую роль в прогрессировании опухоли [234], поэтому имеет важное значение для диагностики и терапии, поскольку редкие лекарственно-устойчивые варианты могут стать доминирующими и привести к рецидиву у пациента [110, 206]. Например, в случае предракового пищевода Барретта клональное разнообразие является важным предиктором прогрессирования опухоли до аденокарциномы пищевода [228]. В случае же РМЖ, рака яичников, желудка и немелкоклеточного рака легкого (далее - НМРЛ) существует парадоксальная связь между хромосомной нестабильностью и прогнозом прогрессирования рака. Наихудшие результаты наблюдаются при опухолях с промежуточным значением хромосомной нестабильности, тогда как при экстремальных показателях результат улучшается [55, 296]. В недавнем исследовании 12-ти различных типов рака риск наступления летального исхода увеличивался, когда более двух клонов сосуществовали в одном образце опухоли, и снижался при сосуществовании более четырех клонов, демонстрируя, что уровень гетерогенности в опухолях может существенно влиять на результат лечения [34]. Понимание механизмов, лежащих в основе клональной гетерогенности, необходимо для разработки новых терапевтических стратегий, направленных на подавление наиболее агрессивных клонов и предотвращение рецидивов.

Дополнительными гетерогенными факторами резистентности, помимо устойчивости опухоли к лекарственным препаратам, возможно являются увеличение количества регуляторных Т-клеток и миелоидных супрессорных клеток в микросреде опухоли, аномальная васкуляризация и образование опухоль-ассоциированных фибробластов [52, 196, 311, 313]. Иммунотерапия, включая вакцины или ингибиторы контрольных точек, менее эффективна при опухолях большого размера, или при метастатическом заболевании, поскольку иммуносупрессивные сигналы, экспрессируемые клетками в микросреде опухоли, не позволяют улучшить показатели долгосрочной выживаемости пациентов, особенно если иммунотерапия используется в качестве монотерапии [64]. Таргетные методы лечения, такие как ингибиторы BRAF для меланомы, оказываются неэффективными при вторичных мутациях BRAF V600, активации N-RAS и путей выживания через рецептор тирозинкиназы [254, 370]. Gopal с соавторами показали, что динамика распространения опухоли влияет на успех лечения анти-BRAF/MEK, а полученная в результате клональная эволюционная динамика может помочь в разработке более эффективных терапевтических режимов [143]. В связи с этим можно сделать предварительный вывод, что комбинированные подходы, включающие одновременно таргетную терапию и иммунотерапию, могут предложить более перспективные результаты за счет преодоления различных механизмов резистентности и улучшения общей выживаемости пациентов.

Эволюция рака концептуально аналогична эволюции микроорганизмов [332], и находится в зависимости от динамического взаимодействия трех основных процессов: (1) генерация

наследственных вариаций, (2) влияние случайных рождений и смертей на судьбу новых генотипов, называемое генетическим дрейфом, и (3) дарвиновский отбор, который изменяет частоту генотипов в популяции в зависимости от их преимущества для относительной приспособленности. Приобретение наследственных изменений и генетический дрейф являются случайными процессами, в то время как дарвиновский отбор является детерминированным по своей природе. Использование особенностей и закономерностей эволюционных процессов может помочь в предсказании развития и клинического течения, а также ответа на лечение различных типов рака.

1.2. Понятие персонализированной и прецизионной медицины в мировой клинической практике

Прецизионная медицина в отношении онкологии определяется как «использование терапевтических средств, которые потенциально могут принести пользу подгруппе пациентов, у которых рак проявляет определенные молекулярные или клеточные особенности (чаще всего геномные изменения и изменения в паттернах экспрессии генов или белков) [385]. Для некоторых опухолей молекулярное профилирование может дать клинически значимую диагностическую и прогностическую информацию. Ввиду геномной сложности рака прецизионная медицина стала возможной благодаря растущему объему знаний, определяющих ключевые движущие силы онкогенеза в сочетании с достижениями в анализе опухолей с помощью секвенирования нового поколения и других технологий профилирования, а также благодаря доступности новых терапевтических средств.

Персонализированная и прецизионная медицина особенно важны в онкологии по нескольким причинам. Во-первых, в онкологии уделяется повышенное внимание профилактике: существуют скрининговые программы для выявления генетической предрасположенности к формированию ЗНО, что позволяет предлагать пациентам индивидуальные рекомендации для модификации риска. Во-вторых, при лечении онкологических пациентов присутствует значительная краткосрочная токсичность, а долгосрочные функциональные последствия связаны со стратегиями хирургического вмешательства и химиолучевой терапии. Надлежащий отбор пациентов для лечения с целью максимизации эффективности и минимизации токсичности долгое время был фундаментальной частью рутинной клинической практики, но до недавнего времени у клиницистов было ограниченное количество инструментов, с помощью которых можно было бы осуществить этот отбор. Исследования и разработки в области персонализированной медицины рака, включая использование ключевых геномных и омиксных

данных пациента (транскриптомика, метаболомика, протеомика и т.д.), улучшают показатели выживаемости и быстро становятся важной частью повседневной клинической практики [191, 243, 388].

Для персонализированного подхода к лечению онкологических заболеваний ключевую роль играет определение молекулярно-генетического профиля опухоли для назначения индивидуального лечения. Молекулярное профилирование опухолей заключается в исследовании нуклеиновых кислот (ДНК и РНК) и/или белков в образце опухоли отдельного пациента с использованием полимеразной цепной реакции (ПЦР), FISH, секвенирования по Сэнгеру, секвенирования нового поколения (англ. next generation sequencing, NGS), ИГХ и других методов [114]. Каждый из этих методов имеет свои преимущества и недостатки, связанные, прежде всего, со стоимостью и чувствительностью исследования. В настоящее время прогноз эффективности традиционных противоопухолевых агентов основывается, прежде всего, на измерении экспрессии белка с помощью ИГХ. Назначение же недавно разработанных и одобренных препаратов для таргетной терапии происходит после обнаружения подходящих для лечения генетических изменений с помощью NGS. Хотя ИГХ и FISH являются основными инструментами прецизионной медицины в повседневной клинической практике, подходы NGS все чаще заменяют эти традиционные методы.

Современный подход прецизионной медицины к диагностике и лечению злокачественных новообразований предполагает комплексное использование геномного, протеомного и метаболомного анализа. Геномный анализ при этом используется как основной, а протеомное и метаболомное исследования несут уточняющую и дополняющую функции. Использование протеомики может помочь тогда, когда было обнаружено несколько генетических изменений и трудно определить наиболее релевантный драйвер. Более того, генетические мутации не всегда приводят к предсказуемому изменению соответствующего белка, и существует множество других факторов, которые определяют характеристики опухоли, например, модификации белков [235]. Метаболомный анализ помогает определить фармакологический фенотип пациента и понять механизмы вариабельности ответа на медикаментозную терапию [37].

За последние несколько лет значительно увеличилось количество опухолевых маркеров, имеющих значение для разработки лекарств. Благодаря использованию биомаркеров в лечении многих типов рака значительно улучшились показатели выживаемости пациентов. FDA поддерживает подход прецизионной медицины и выдает разрешения на применение таргетных препаратов с 1998 г., когда трастузумаб был одобрен для лечения HER2-положительного РМЖ. В 2020 г. FDA одобрило 39 новых лекарств или их комбинаций для лечения различных видов рака: медуллярного рака щитовидной железы, нейробластомы, метастатического ТНРМЖ, множественной миеломы и других. Большинство из этих препаратов направлено на конкретный

биомаркер или конкретное генетическое изменение в опухоли. При этом многие новые лекарства находятся на стадии разработки или клинических испытаний.

Терапия, учитывающая данные молекулярного и генетического профилирования опухоли, может иметь большое значение для лечения больных ЗНО [116, 275, 335]. Ценность прецизионной медицины для повседневной клинической практики в значительной степени зависит от наличия эффективных инструментов для преобразования данных пациента в диагностические критерии и таргетное лечение. Выбор препарата должен основываться на совокупности медицинской информации о пациенте, включая данные геномики и/или протеомики, в сочетании с характеристиками, описывающими свойства необходимого препарата. Персонализированные подходы к диагностике клинически значимых изменений в опухоли могут обеспечить быстрое назначение наиболее адекватных методов лечения, адаптированных к индивидуальным особенностям конкретных опухолей.

1.3. Анализ существующих методов прецизионной медицины и индивидуализированного подбора терапевтических препаратов для онкологических больных

Персонализированный подход к диагностике и лечению онкологических заболеваний широко применяется во многих странах, например, в США, Японии, Китае, странах Евросоюза. Этот подход включает, помимо оптимизации тактики ведения пациентов и процесса принятия клинических решений, применение NGS для улучшения диагностики и назначения эффективной терапии.

NGS позволяет одновременно идентифицировать наличие геномных аномалий, таких как однонуклеотидные варианты (англ. single nucleotide variant, SNV), короткие вставки/делеции, вариации числа копий и слияние нескольких генов [20, 217]. Выявление конкретных молекулярных характеристик (например, микросателлитной нестабильности (англ. microsatellite instability, MSI) в опухоли) [209], соматических и герминальных мутаций (например, в генах *BRCA*, *BRAF* или *ERBB2*) [31, 142], или слияния генов (например, *FGFR* и *NTRK*) [103, 111], может определить выбор лечения и улучшить исход для пациентов независимо от расположения опухоли. Чтобы правильно применять молекулярные методы лечения, необходимо идентифицировать группы пациентов, для которых польза от данной терапии наиболее вероятно, превысит риски серьезных нежелательных реакций и побочных эффектов. Такая оценка может быть осуществлена с помощью определения конкретных биомаркеров, позволяющих

прогнозировать ответ на терапию или устойчивость к ней [194]. Например, для РМЖ выделяются пять основных подтипов (Рисунок 1).



Рисунок 1. Пять основных молекулярных подтипов рака молочной железы. Создано с помощью BioRender.com

Обнаружение геномных изменений потенциально позволяет выявлять опухоли с наилучшими шансами на реакцию на традиционные схемы цитотоксической химиотерапии, а также на более новые методы лечения моноклональными антителами, низкомолекулярными ингибиторами тирозинкиназы и иммунотерапию. Кроме того, идентификация геномных аномалий может привести к разработке лекарств против новых белков (неоантигенных противоопухолевых вакцин), участвующих в альтернативных сигнальных путях.

Наиболее показательные результаты были достигнуты при применении персонализированного подхода к лечению РМЖ и легких, а также меланомы [35, 289, 327]. При РМЖ, помимо определения рецепторов гормонов в качестве терапевтических агентов, нацеливание на другие биомаркеры, такие как мутации в генах *PIK3CA* (альфа-изоформа каталитической субъединицы фосфатидилинозитол-4,5-бисфосфат-3-киназы) или *ERBB2* (рецептор 2 эпидермального фактора роста человека), полностью изменило терапевтический подход [35, 124, 178, 324]. Более того, обнаружение амплификации гена *HER2* также внесло огромный вклад в выявление другой важной подгруппы пациентов, которых можно успешно лечить путем ингибирования HER2 [109, 199, 358]. Фундаментальный сдвиг также наблюдается

в лечении пациентов с диагнозом НМРЛ. Идентификация мутаций в генах *EGFR* (рецептор эпидермального фактора роста) [267], и *EML4-ALK* [325], повлияло на результаты лечения пациентов с поздней стадией заболевания. Выявление мутации *BRAF-V600E* и последующее лечение ингибиторами BRAF и MEK значительно улучшило клинические исходы у пациентов с меланомой [124]. Это лишь несколько примеров применения прецизионной медицины в повседневной клинической практике.

Применение новых молекулярно-направленных агентов может быть ограничено первичной или вторичной резистентностью к терапии. Молекулярные дефекты в опухолевых клетках в дополнение к дефектам, на которые нацелены лекарства, могут привести либо к неспособности реагировать на лекарство с самого начала, либо к раннему прогрессированию заболевания после первоначального ответа. Вероятно, это связано как с исходными, так и с приобретенными механизмами устойчивости. Выявление приобретенной резистентности, которая развивается после первоначального применения кризотиниба у пациентов с *EML4-ALK* НМРЛ, способствовало разработке ингибиторов *EML4-ALK* второго поколения. Использование ингибиторов *EGFR* при НМРЛ в конечном итоге приводит к устойчивости, в некоторых случаях за счет вторичных мутаций в гене *EGFR* или амплификации онкогена *MET*. Использование ингибиторов *MET* в сочетании с ингибиторами *EGFR* может обеспечить продление эффективности терапии [269]. Аналогичным образом, у пациентов с меланомой *BRAF V600E*, получавших вемурафениб, резистентность и рецидив опухоли обычно развиваются через 5-7 месяцев. Механизмы приобретенной резистентности включают появление дополнительных мутаций в гене *BRAF*, а также в последующих элементах сигнального пути, таких как *NRAS*, *MEK* и *ERK*, в субпопуляциях опухолевых клеток, что приводит к молекулярной эволюции опухоли, результатом которой её гетерогенность и, как следствие, снижение эффективности терапии в этих субпопуляциях [339]. Особо можно отметить возможность диагностирования генетических маркеров ТНРМЖ (Рисунок 2). Таким образом, разработки в области рационального терапевтического дизайна должны быть направлены на выявление, понимание и нацеливание этих дополнительных молекулярных мутаций, модифицирующих болезнь, для разработки эффективных методов второй линии или комбинированной терапии.

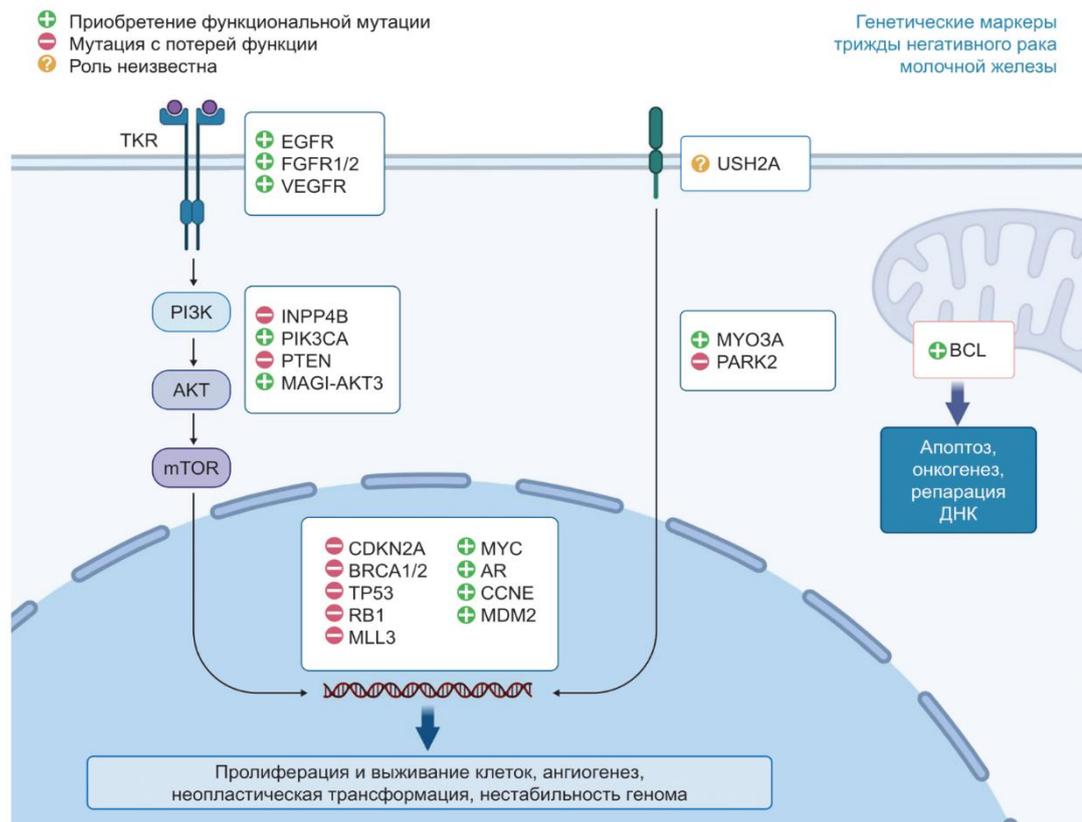


Рисунок 2. Генетические маркеры трижды негативного рака молочной железы. Создано с помощью BioRender.

Существует несколько панельных генетических тестов, которые используются для назначения наиболее подходящей терапии и различаются по количеству исследуемых генов [61, 360]. Большинство таких тестов включает несколько наиболее изученных генов, которые являются мишенью для эффективной терапии. Панель, включающая большое количество генов, не обязательно представляет дополнительную ценность для пациента, поскольку не все варианты имеют одинаковую клиническую значимость. Некоторые варианты могут быть чрезвычайно редкими за пределами определенных популяций, но могут быть распространены в пределах определенной группы. Например, вариант HLA-B*15:02, связанный с повышенным риском развития тяжелых кожных побочных реакций у пациентов, получавших карбамазепин, имеет частоту аллелей 0,04 % у пациентов европейского происхождения и 6,88 % у пациентов восточноазиатского происхождения [280]. Панельный тест, таким образом, может быть более полезным для пациента, если он более точно анализирует варианты, соответствующие его этническому происхождению. В связи с этим разработка фармакогеномной платформы, основанной на анализе частоты встречаемости различных генетических вариантов в российской популяции и ее генетических ландшафтах, является актуальной задачей.

Достижения последних лет в области молекулярной биологии и онкогеномики привели к выявлению сложных и уникальных биологических особенностей, связанных с канцерогенезом. Для выявления этих характеристик и оптимизации на этом основании противоопухолевой терапии у отдельных пациентов в настоящее время используются профилирование опухолевой и внеклеточной (циркулирующей) ДНК, иммунные маркеры, протеомный анализ и анализ экспрессии генов. Таким образом, клиническая диагностика эволюционировала, перейдя от ориентированной на тип опухоли к генно-ориентированным, гистологически агностическим, адаптированным к профилированию биомаркеров исследованиям с целью улучшения результатов лечения. Было проведено множество испытаний прецизионных диагностических инструментов. Большинство этих испытаний продемонстрировали, что индивидуально подобранная терапия обеспечивает лучшие результаты по сравнению со стандартным лечением при разных типах опухолей.

Первым клиническим исследованием в области прецизионной медицины солидных опухолей с использованием транскриптомики было исследование «WINTHER» [295], в котором сравнивалась экспрессия генов в опухолях с экспрессией в прилегающих нормальных тканях, и было показано, что такой анализ позволяет увеличить число пациентов, которым может быть назначена терапия, и увеличить продолжительность жизни без прогрессирования заболевания для более чем 20% пациентов. В других исследованиях транскриптомный анализ использовался для выбора таргетной терапии пациентов с солидными опухолями на поздних стадиях [367]. Так, в исследовании двадцати пациентов с рефрактерными саркомами были идентифицированы гиперэкспрессированные гены и aberrантно функционирующие сигнальные пути, которые были предложены в качестве мишеней для подбора индивидуальной таргетной терапии (ТТ). ТТ проводилась 9 пациентам; не получали ТТ 11 пациентов. Общая выживаемость (ОВ) и выживаемость без прогрессирования (ВБП) были значительно выше в группе пациентов, получавших ТТ: медиана ОВ составила 8,83 против 4,93 месяца, а медиана ВБП составила 6,17 против 1,6 месяца в группе ТТ и без ТТ, соответственно (ОВ: $P = 0,0014$, ВБП: $P = 0,0011$) [367].

Однако основой большинства модулей интерпретации СППВР, разработанных к настоящему времени в мире, можно считать межинституциональное проспективное исследование «I-PREDICT», в котором использовалось секвенирование опухолевой ДНК и формировались рекомендации по индивидуальной комбинированной терапии.

Исследование «I-PREDICT» проводилось в двух центрах (San Diego Moores Cancer Center и Avera Cancer Institute), в исследовании было использовано тканевое геномное профилирование с помощью методов NGS (Foundation Medicine; 236–405 генов) и, в некоторых случаях, ИГХ анализ на наличие маркера PD-L1, опухолевой мутационной нагрузки (англ. tumor mutational burden, TMB), состояния MSI и NGS циркулирующей опухолевой ДНК из крови. Основываясь

на полученной информации, специальный совет, состоящий из онкологов, фармакологов, генетиков, хирургов, рентгенологов, патологов и экспертов по биоинформатике, подбирает индивидуальную комбинацию нескольких препаратов для воздействия на большинство геномных изменений в опухоли каждого пациента [321]. Для 83 пациентов было подобрано лечение и проведена оценка его эффективности. Было продемонстрировано, что таргетное воздействие на большую часть выявленных молекулярных изменений, коррелировало со значительно лучшими показателями контроля заболевания, а также с более длительными показателями безрецидивной выживаемости и ОВ по сравнению с нацеливанием на меньшее количество соматических изменений.

Разработчики I-PREDICT предлагают оптимизировать текущую парадигму прецизионной онкологии, которая предполагает поиск соответствия одной драйверной мутации одному препарату, и использование комбинаций индивидуально подобранных терапевтических агентов. Однако можно отметить сложность применяемой I-PREDICT методологии, а именно необходимость сочетать одновременно множество диагностических инструментов, что значительно увеличивает стоимость теста и усложняет интерпретацию результатов, необходимость в итоговом консультировании со множеством специалистов для выбора терапевтической стратегии и возможное увеличение количества побочных реакций от неканонических сочетаний нескольких препаратов.

В мире существует много примеров как разработки, так и успешного внедрения в клиническую практику СППВР. Но лишь небольшая их часть позволяет осуществлять их “online” совершенствование за счет дополнения генетических данных пациентов. Следует отметить существующую систему FoundationOne®CDx, которая стала первой одобренной FDA диагностической программой широкого спектра, клинически и аналитически подтвержденной для всех солидных опухолей. Она предназначена для предоставления врачам клинически значимой информации — как для выбора подходящей терапии для пациентов, так и для понимания результатов с доказательствами релевантности — на основе индивидуального геномного профиля опухоли каждого пациента. Среди недостатков данной платформы следует отметить высокую стоимость поддержания ввиду ее устаревших технических решений, что лимитирует ее широкое внедрение в систему онкологической помощи. Кроме того, доступность этой системы в России ограничена.

Несмотря на наличие в России более 20 различных СППВР, только несколько из них фактически предусматривает возможность подбора персонализированной терапии в онкологии - Платформа OncoBox и Oncofinder, использующая алгоритм, который количественно анализирует степень активации каждого исследуемого молекулярного пути путем расчета значения силы его активации, а далее - индивидуального рейтинга эффективности 160 противоопухолевых

препаратов, и генетический тест “Atlas Solo Комплекс Плюс”, в рамках которого анализируется в том числе панель из 160 генов для определения эффективности таргетных противоопухолевых препаратов [403]. Основная задача СППВР в системе Galenos, созданной ИТ-компанией “ТехЛАБ”, состоит в контроле выполнения медицинских стандартов. Другая СППВР Lexema-Medicine была разработана для оценки эффективности лечения отдельного пациента и формирования оптимальных рекомендаций по назначению терапии, а также оценки долгосрочных рисков для здоровья пациентов.

Среди отечественных СППВР наибольший интерес представляет вышеупомянутая биоинформатическая платформа Oncobox [330]. В исследовании, проведенном на клеточных культурах рака яичников человека SKOV-3 и нейробластомы NGP-127, были определены профили экспрессии генов чувствительных и резистентных клеток с помощью микрочипов и проанализированы изменения в 378 сигнальных путях, ассоциированных с канцерогенезом, с использованием биоинформатической платформы Oncobox. В соответствии с принципом работы системы в качестве мишеней для воздействия таргетных препаратов были выбраны белки, участвующие в как можно большем числе сигнальных путей, активированных в резистентных клетках. Было протестировано 13 комбинаций лекарств и/или селективных ингибиторов, предсказанных платформой Oncobox, и 10 случайных комбинаций. Показатели синергизма для прогнозов Oncobox были значительно выше, чем для случайно выбранных комбинаций препаратов [330]. Однако следует отметить, что результаты работы данной системы не были подтверждены в клинических испытаниях. Немаловажно и то, что Oncobox исследует только опухолевый материал без контрольного исследования здоровой ткани (крови), что не позволяет наложить прочитанный геном опухоли на геном здоровой ткани, и, соответственно, идентифицировать герминальные мутации и дифференцировать их от соматических. Основным же ограничивающим фактором СППВР Oncobox остается предусмотренный подбор таргетной терапии всего из 160 препаратов, заранее включенных в систему. Крайне ограниченный перечень препаратов в ряде случаев не охватывает все терапевтические возможности, доступные пациенту сегодня, и становится сдерживающим фактором, что не отвечает задачам трансляции новых достижений персонифицированной терапии в клиническую практику.

Еще одним отечественным аналогом является система “Atlas Solo Комплекс Плюс”. В основе данного теста лежит проведение комплексного исследования, включающего в себя комбинацию из NGS исследования панели из 160 генов опухолевого материала, FISH исследований, а также ИГХ исследований [44, 403]. Главным недостатком данной системы является ограниченная зона поиска, что существенно ограничивает спектр подбора таргетной терапии.

В 2019 г. в России был запущен Федеральный проект «Борьба с онкологическими заболеваниями», целью которого является рост продолжительности жизни за счет улучшения диагностики и повышения эффективности лечения ЗНО. Предполагается достигнуть поставленных целей с помощью, в том числе, применения персонализированного подхода к ведению онкологических больных. Этот подход сочетает в себе диагностику на основе анализа различных данных пациента, использование прецизионной терапии (таргетной и иммунотерапии) и применение СППВР. Это важно для российской системы здравоохранения, поскольку проблемы профилактики и лечения рака сформулированы как национальный приоритет, и первые результаты реализации федерального проекта позволяют сделать определенные выводы. В рамках проекта государственное финансирование лечения злокачественных новообразований было увеличено более чем вдвое по сравнению с 2018 г. Это позволило применять современные таргетные и иммунологические препараты, включенные в международные и национальные клинические руководства. Новые противоопухолевые препараты назначают не только в республиканских, но и в региональных онкологических центрах. Федеральный проект «Борьба с онкологическими заболеваниями» в 2019-2023 гг. показал хорошие результаты, но мы предполагаем большой потенциал СППВР, которая обеспечит рациональное использование финансовых средств и увеличит эффективность применения противоопухолевых лекарственных препаратов.

1.4. Трижды негативный рак молочной железы как модель для испытания эффективности платформ прецизионной медицины

РМЖ включает несколько молекулярных подтипов, имеющих разную чувствительность к лекарственным воздействиям и разный прогноз. ТНРМЖ характеризуется отсутствием экспрессии рецепторов эстрогенов и прогестерона (РЭ/РП) и амплификации онкогена *HER2/neu* [73, 256, 345]. Радиологически ТНРМЖ часто идентифицируется как гиперплотное образование без связанных кальцификаций [107]. ТНРМЖ имеет клинические особенности, которые включают высокую инвазивность и высокий метастатический потенциал, ранние рецидивы и высокую смертность [176]. Поскольку клетки ТНРМЖ не обладают экспрессией РЭ, РП и HER2, они нечувствительны к эндокринной терапии, таргетной терапии, направленной на HER2, а стандартизированные схемы лечения ТНРМЖ относительно малоэффективны. Неoadъювантная химиотерапия при этом является предпочтительным подходом к лечению [66].

ТНРМЖ чаще всего встречается у молодых женщин в пременопаузе в возрасте до 40 лет; он составляет примерно 15–20 % всех случаев РМЖ [246]. В одном исследовании было показано, что риск ТНРМЖ у женщин младше 40 лет вдвое выше, чем у женщин старше 40 лет: отношение

шансов в возрасте 20-39 лет составляет 2,13, в возрасте 40-54 лет – 1,00-1,09 [352]. По сравнению с другими подтипами РМЖ, время выживания пациентов с ТНРМЖ меньше, а уровень смертности составляет 40 % в течение первых 5 лет после постановки диагноза [97]. Примерно у 46 % пациентов с ТНРМЖ обнаруживаются метастазы. У женщин с ТНРМЖ в 4 раза чаще формировались висцеральные метастазы в течение 5 лет после постановки диагноза, чем у женщин с другими типами рака [98]. Частота рецидивов после операции достигает 25 %, а среднее время выживания при метастазировании составляет всего 13,3 месяца. Метастазы часто поражают головной мозг и внутренние органы и в основном на 3-м году после постановки диагноза [218]. Среднее время до рецидива у пациентов с неТНРМЖ составляет 35–67 месяцев, а у пациентов с ТНРМЖ – только 19–40 месяцев [394].

Большинство образцов ТНРМЖ гистологически классифицируются как инвазивные протоковые карциномы высокой степени злокачественности без специфичности с чертами базальноклеточного рака. Центральный некроз, заметный лимфоцитарный инфильтрат и фиброз являются общими гистологическими признаками ТНРМЖ [221]. Следует отметить, что не все ТНРМЖ являются базальными.

1.4.1. Молекулярно-генетическая классификация трижды негативного рака молочной железы на основании анализа экспрессии генов

Анализ профилей экспрессии генов часто классифицирует ТНРМЖ как подтип базально-подобного РМЖ (далее - БПРМЖ). Примерно 56% профилей экспрессии генов ТНРМЖ и БПРМЖ перекрываются. Коэффициент перекрытия может достигать 60–90 % между ТНРМЖ и БПРМЖ, по сравнению с 11,5 % между неТНРМЖ и БПРМЖ [278, 283].

В ТНРМЖ с базальными признаками не экспрессируются РЭ, РП и рецептор Her2/neu, но обычно экспрессируются высокомолекулярные базальные цитокератины (цитокератин 5/6, цитокератин 14 и цитокератин 17), виментин, p-кадгерин, αВ-кристаллин, фасцин и кавеолы 1 и 2 [287]. В 2011 г. Lehmann и соавторы выполнили профилирование экспрессии генов в образцах опухолей 587 пациентов с ТНРМЖ и разделили ТНРМЖ на 6 подтипов: базальноподобный 1 (БП1), базальноподобный 2 (БП2), иммуномодулирующий (ИМ), мезенхимальный (М), мезенхимальный стволоподобный (МСП) и люминальный с рецептором андрогенов (ЛАР) [211].

Подтип БП1 характеризуется высокой частотой гетерозиготных или гомозиготных делеций генов, связанных с репарацией ДНК, таких как *BRCA2*, *PTEN*, *MDM2*, *RB1* и *TP53*, а также значительной амплификацией генов *MYC*, *PIK3CA*, *CDK6*, *AKT2*, *KRAS*, *FGFR1*, *IGF1R*, *CCNE1* и *CDKN2A/B*, которые регулируют клеточный цикл и участвуют в репарации ДНК. В отличие от него, подтип БП2 характеризуется выраженной активацией сигнальных путей *EGFR*,

MET, *NGF*, *Wnt/β-катенин* и *IGF-1R* [211]. Признаком ИМ подтипа является активация сигнальных путей, связанных с функционированием иммунных клеток, включая сигнальные пути дендритных клеток, рецептора Т-клеток, IL-12 и IL-7, а также сигнальные пути Th1/Th2, сигнальные пути НК-клеток и рецептора В-клеток [387].

Выраженная активация сигнальных путей, ассоциированных с миграцией клеток и дифференцировкой (сигнальные пути киназы анапластической лимфомы, *Wnt*, трансформирующего фактора роста (*TGF-β*)) является признаком подтипа М [212]. Слабая экспрессия генов, связанных с пролиферацией клеток, отмечается в подтипе МСП, в то время как экспрессия *HOX* генов (*HOXA5*, *HOXA10*, *MEIS1*, *MEIS2*, *MEOX1*, *MEOX2* и *MSX1*) и генов, ассоциированных со стволовостью (*ABCA8*, *PROCR*, *ENG*, *ALDH1*, *PER1*, *ABCBI*, *TERT2IP*, *BCL2*, *BMP2* и *THY*), выражена в данном подтипе ТНРМЖ. Кроме того, выражены маркеры мезенхимальных стволовых клеток (*BMP2*, *ENG*, *ITGAV*, *KDR*, *NGFR*, *NT5E*, *PDGFR*, *THY1* и *VCAM1*). Активация гормональных сигнальных путей (включая синтез стероидов, метаболизм порфиринов и метаболизм андрогенов/эстрогенов) характерная для подтипа ЛАР. Стоит отметить, что вместе с отсутствием экспрессии РЭ в подтипе ЛАР наблюдается выраженная экспрессия рецептора андрогенов (РА) [387]. Пациенты с подтипом ЛАР демонстрировали более высокие показатели ОВ и выживаемости без метастазов по сравнению с пациентами, имеющими подтипы М и БП2. Кроме того, частота рецидивов у пациентов с подтипом ЛАР ниже по сравнению с подтипами М и БП2 в течение 3-х лет после постановки диагноза [231, 387]. Необходимо также отметить, что недавние исследования показывают, что подтип ЛАР может быть более чувствителен к терапиям, нацеленным на РА, что открывает новые перспективы для персонализированного лечения пациентов с этим подтипом ТНРМЖ [241, 348]. В частности, ингибиторы РА, такие как энзалутамид, демонстрируют обнадеживающие результаты в клинических испытаниях [38, 351], что может существенно улучшить прогноз и качество жизни пациентов с подтипом ЛАР.

По итогам последующих молекулярно-генетических исследований, проведенных Burstein и соавторами, была предложена следующая классификация ТНРМЖ: подтип ЛАР, экспрессирующий РА и муцин *MUC1*; М, экспрессирующий рецептор фактора роста тромбоцитов α-типа и рецептор *c-Kit*; БПИС (базальноподобный иммуносупрессивный), экспрессирующий иммуносупрессивную молекулу *VTCN1*; и БПИА (базальноподобный иммуноактивный), экспрессирующий молекулы передачи сигнала *STAT* и высвобождающий цитокины. Было продемонстрировано, что показатели выживаемости без признаков заболевания для названных подтипов уменьшались в следующем порядке: БПИА>М>ЛАР>БПИС [67, 387]. Кроме того, недавние исследования показали, что каждый из этих подтипов может обладать уникальной чувствительностью к различным терапевтическим подходам, что открывает новые

возможности для разработки персонализированных стратегий лечения. Например, подтип ЛАР может быть более восприимчив к гормональной терапии, в то время как БПИС и БПИА могут требовать иммунотерапевтических подходов [242].

В 2016 году, после категоризации 165 образцов ТНРМЖ в сочетании с анализом экспрессии мРНК и коэкспрессии для выявления взаимодействий между мРНК и длинными некодирующими РНК, была предложена новая классификация подтипов ТНРМЖ [220]. Подтип ИМ участвует в регуляции иммунных клеток, включая сигнальные пути цитокинов, передачу сигналов иммунных клеток (рецепторы Т-клеток и В-клеток), процессинг и презентацию антигенов, сигнальные пути хемокинов и путь передачи иммунных сигналов (сигнальный путь NF-κB). Все гены, высоко экспрессируемые в подтипе ИМ, связаны с иммунными функциями. Некоторые гены, участвующие в иммунном ответе (*CCR2*, *CXCL13*, *CXCL11*, *CD1C*, *CXCL10* и *CCL5*), также значительно экспрессируются, что указывает на тесную связь подтипа ИМ с иммунной регуляцией. Подтип ЛАР характеризуется уникальными активными путями гормональной регуляции, включая метаболизм андрогенов и эстрогенов, биосинтез стероидных гормонов и метаболизм порфирина и хлорофилла. Несмотря на то, что ИГХ исследование показало, что подтип ЛАР является РЭ-отрицательным, профиль экспрессии демонстрирует активацию сигнального пути эстрогена, что предполагает возможность ответа подтипа ЛАР на антиандрогеновую и традиционную антиэстрогенную терапию [67]. Мезенхимоподобный подтип (МП) характеризуется множеством уникальных сигнальных путей, таких как взаимодействие внеклеточного матрикса и рецепторов, щелевые контакты, сигнальные пути TGF-β и пути, связанные с факторами роста (переносчики ABC и адипокиновые сигнальные пути). Подтип БПИС характеризуется активацией путей клеточного деления, клеточного цикла, репликации ДНК и регуляции репарации ДНК. В подтипе БПИС значительно усиливается экспрессия генов, связанных с пролиферацией, включая *CENPF*, *BUB1* и *PRC1*. Таким образом, подтип БПИС проявляет высокие пролиферативные свойства [67]. Гены, участвующие в иммунном ответе, передаче сигналов иммунных клеток (костимуляция Т-клеток, сигнальный путь рецепторов Т-клеток, активация В-клеток и хемотаксис дендритных клеток), и процесс активации комплемента значительно подавляются в подтипе БПИС. Прогностический анализ показал, что пациенты с подтипом БПИС имели худшие показатели выживаемости без рецидива и более высокий риск рецидива по сравнению с другими подтипами [387].

В настоящее время большинство исследований молекулярных подтипов ТНРМЖ основывается на оценке мРНК различных генов. Однако уровень экспрессии мРНК не всегда коррелирует с уровнем экспрессии белка, поскольку процесс трансляции белка включает множество этапов модификации и регуляции, которые могут существенно влиять на терапевтический эффект и прогноз заболевания у некоторых пациентов. В то же время, точное

определение молекулярного подтипа ТНРМЖ на основе результатов иммуногистохимического окрашивания в рутинной клинической практике при большом количестве клинических образцов остается неопределенным, и существующие данные часто противоречивы [387]. В связи с этим, в последнее время все большее внимание уделяется молекулярно-генетической классификации ТНРМЖ с использованием методов NGS.

Это позволяет не только точно классифицировать клинический образец, но и определить подходящие для лечения определенного пациента таргетные препараты.

1.4.2. Генетические особенности трижды негативного рака молочной железы на основании данных секвенирования нового поколения

Примерно от 11 до 20 % пациентов с ТНРМЖ независимо от семейного анамнеза имеют герминальную мутацию в генах *BRCA1* или *BRCA2* (*BRCA1/2*) [232]. Ген *BRCA1*, определяющий предрасположенность к РМЖ и раку яичников, был идентифицирован в 1994 г. [237], а вскоре после этого был обнаружен ген *BRCA2*, также вовлеченный в развитие опухолей данных локализаций [378].

При РМЖ и раке яичников, возникающих в условиях герминальной мутации генов *BRCA1* или *BRCA2*, нарушается путь, который функционирует при репарации двухцепочечных разрывов ДНК и обеспечивает поддержку репликации ДНК после восстановления дефектных вилок репликации [216]. Первые исследования риска развития злокачественной опухоли, связанной с мутациями в этих генах, были ограничены случаями с большим размером опухоли и молодым возрастом пациента на момент постановки диагноза. Эти работы дали очень высокие оценки значимости мутаций *BRCA1/2*, приближаясь к 90%-ному риску развития РМЖ [126, 195]. После этого появились более репрезентативные популяционные исследования, которые включали пациентов с диагнозом РМЖ и рак яичников без известного отягощенного семейного анамнеза или ранней манифестации заболевания. Анализ данных подгрупп дал более низкие оценки значимости исследуемых мутаций в развитии злокачественных опухолей. По результатам метаанализа было определено, что среди носителей мутаций в генах *BRCA1* и *BRCA2* риск развития РМЖ составляет 57 % и 49 %, а рака яичников – 40 % и 18 % соответственно [39, 81, 204]. В более позднем проспективном когортном исследовании кумулятивный риск РМЖ к 80 годам составил 72 % для носителей мутации в гене *BRCA1* и 69 % для носителей мутации в гене *BRCA2*. Это исследование также продемонстрировало, что риск РМЖ для носителей мутации варьирует в зависимости от семейного анамнеза РМЖ у родственников первой и второй степени родства, что свидетельствует о существовании других генетических факторов, которые модифицируют риски РМЖ [202]. Современные исследования позволили уточнить влияние

различных факторов на риск развития ЗНО у носителей мутаций в генах *BRCA1/2*. Например, было выявлено, что такие факторы, как образ жизни, диета, использование гормональной терапии, могут существенно влиять на риск развития рака, связанный с мутациями в этих генах [96, 136]. Это подчеркивает необходимость индивидуализированного подхода к оценке риска и профилактике рака у носителей мутаций *BRCA1/2*.

РМЖ, ассоциированный с герминальной мутацией в гене *BRCA1*, в 50-88 % случаев имеет трижды негативный подтип (Musolino et al., 2007; Atchley et al., 2008; Byrski et al., 2008) [42, 68, 250]. Мутации в гене *BRCA1* обнаруживаются у 23% пациентов с ТНРМЖ, диагностированных в молодом возрасте и/или с семейным анамнезом РМЖ [291], и у 8-20 % пациентов с ТНРМЖ, которые не были отобраны по раннему возрасту или семейному анамнезу при постановке диагноза [36, 95, 131, 142, 197, 291, 312]. Хотя герминальные мутации в гене *BRCA2* реже ассоциированы с ТНРМЖ, чем мутации в гене *BRCA1*, они обнаруживаются у 4-9 % больных ТНРМЖ [91, 142, 154].

Мутации в генах *BRCA1* и *BRCA2* вызывают дефицит гомологичной рекомбинации (ГР), что приводит к повышенной нестабильности генома [366]. Опухоли без мутаций в генах *BRCA* также могут демонстрировать геномную нестабильность высокой степени. Признаки дефицита гомологичной рекомбинации и повышенной нестабильности генома в совокупности называются *BRCAness* [222]. Опухоли, которые имеют общие молекулярные характеристики с мутантными *BRCA* опухолями, то есть опухоли с «*BRCAness*», также могут реагировать на аналогичные терапевтические подходы. Раковые клетки с мутациями *BRCA1/2* проявляют чувствительность к химиотерапии на основе платины, и к ингибиторам поли(АДФ-рибоза)полимеразы (англ. poly ADP ribose polymerase, PARP). В 2020 г. Lin и соавторы выявили, что образцы ТНРМЖ с мутациями в генах ГР имеют сходное большое количество и большую длину крупномасштабных структурных aberrаций и аналогичные измененные хромосомные области, как и образцы с мутациями в генах *BRCA*. Это указывает на чувствительность таких ТНРМЖ к платине и ингибиторам PARP. Опухоли с мутацией генов ответа на повреждение ДНК, отличных от ГР (в генах *PTEN* и *MSH6*), также содержали значительные структурные aberrации (*BRCAness*), однако у них были участки геномных изменений, отличные от таковых в опухолях с мутациями генов *BRCA* и ГР, что может объяснить предыдущие данные о том, что раковые клетки с мутациями генов *PTEN* и *MSH6* нечувствительны к ингибиторам PARP [219].

В настоящий момент для клинической оценки риска развития РМЖ возможно генетическое тестирование от 6 до более чем 100 генов [318]. Изначально при генетическом тестировании помимо *BRCA1/2* исследовались еще 4 гена, ассоциированных с РМЖ – *CDH1*, *PTEN*, *STK11* и *TP53*. Впоследствии генетические панели расширились, чтобы охватить дополнительные 15-20 генов аналогичных по функциям *BRCA1/2* (например, *ATM*, *BARD1*,

CHEK2, *PALB2* и другие), и дополнительные 25-40 генов, охватывающих риск развития рака других локализаций (например, *CDKN2A*, *MEN1*, *MLH1*, *MSH2*, *MSH6* и *MUTYH*). В последнее время панели расширились до более чем 100 генов, которые потенциально ассоциированы с развитием злокачественных опухолей (например, *AXIN2*, *CYLD*, *SLX4*) [22].

В 2015 г. Консорциум по ТНРМЖ оценил частоту различных патогенных мутаций у более чем 1800 пациентов с ТНРМЖ. Были собраны образцы ДНК данной группы пациентов, независимо от отягощенности их семейного анамнеза по РМЖ или раку яичников, и секвенированы 122 гена репарации ДНК, включая 17 генов предрасположенности к РМЖ (*ATM*, *BRCA1*, *BRCA2*, *BARD1*, *BRIP1*, *CHEK2*, *CDH1*, *MRE11A*, *NBN*, *PALB2*, *PTEN*, *RAD50*, *RAD51C*, *RAD51D*, *STK11*, *TP53* и *XRCC2*). У 267 пациентов (14,6%) была выявлена 271 патогенная мутация, в то время как мутации в генах *BRCA1* и *BRCA2* составляли большую их часть (155 мутаций или 57 % и 49 мутаций или 18 %, соответственно); 67 мутаций (25 %) были обнаружены в 12 из 15 других генов предрасположенности, включая *PALB2* (7,7 %), *BARD1* (3,3 %), *RAD51D* (2,5 %), *RAD50* (2,2 %) и *RAD51C* (2,2 %) [94]. Эти вероятно патогенные мутации в генах, не относящихся к *BRCA1/2*, наблюдались у 3,7 % пациентов с ТНРМЖ, а гены *BRCA1/2* – 11,2 %. Примечательно, что не было обнаружено мутаций в гене *CHEK2*, что согласуется с его ассоциацией с гормонально-рецепторным РМЖ [94, 347, 369]. Распространенность мутаций варьировала в зависимости от возраста на момент постановки диагноза (5 % - у пациентов с диагностированным заболеванием в возрасте 60 лет или старше без отягощенного семейного анамнеза, 18 % - в младших возрастных группах независимо от семейного анамнеза) [94]. Эти результаты свидетельствуют о том, что для пациентов с ТНРМЖ тестирование более широкой группы генов предрасположенности к развитию злокачественных опухолей может быть более эффективным.

В ходе полногеномного поиска ассоциаций (англ. genome-wide association study, GWAS) в общей популяции было идентифицировано 179 однонуклеотидных полиморфизмов (англ. single nucleotide polymorphism, SNP) или небольших вставок или делеций, связанных с РМЖ. Хотя аллели риска в отдельных SNP связаны с умеренным увеличением риска РМЖ, было показано, что они мультипликативно сочетаются, что приводит к значительному его повышению [88].

1.4.3. Особенности диагностики трижды негативного рака молочной железы

Определение индивидуального риска развития рака может быть сложной задачей, и для оценки риска развития РМЖ возможно использование математических инструментов. Существуют модели BRCAPRO и Тайрера-Кузика для определения риска развития РМЖ. Модель BRCAPRO часто предсказывает средний или слегка повышенный риск развития РМЖ

по сравнению с общей популяцией для тех, у кого неинформативен тест *BRCA1/2*; при этом модель Тайрера-Кузика показывает прогнозируемые риски часто намного выше, чем у других количественных моделей [266].

Женщины, которые считаются подверженными высокому риску развития РМЖ в течение жизни (определяемому как риск не менее 20 %), должны проходить ежегодную скрининговую маммографию и МРТ молочной железы, а также клиническое обследование молочной железы каждые 6-12 месяцев, начиная с возраста на 10 лет младше, чем у самого младшего заболевшего члена семьи, но не ранее 30 лет для маммографии и не раньше 25 лет для МРТ (National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology). Например, если у женщины, соответствующей порогу риска, есть сестра, у которой РМЖ был выявлен в возрасте 30 лет, может быть рекомендовано проведение МРТ в возрасте 25 лет. Для женщин с риском от 15 до 20 % также возможно назначение МРТ молочных желез [226, 301].

Пациенты, имеющие в анамнезе РМЖ, имеют риск развития контралатерального РМЖ от 4 до 7 % в течение первых 5-10 лет после установления диагноза [255]. Те, у кого есть семейный анамнез по РМЖ, имеют более высокий риск даже с отрицательным результатом теста на мутации в генах *BRCA1/2* [286]. Эти риски еще более значительны для молодых женщин.

Для некоторых пациентов, имеющих отягощенный семейный анамнез по РМЖ, желательно усиленное наблюдение с помощью МРТ молочных желез в дополнение к маммографии, хотя по данным Американского общества клинической онкологии (англ. American Society of Clinical Oncology, ASCO) на данный момент доказательств для таких рекомендаций недостаточно [301]. Необходимы исследования, чтобы определить, будет ли скрининг МРТ у пациентов с наличием в анамнезе РМЖ обеспечивать снижение смертности.

Тем не менее, в международных практических рекомендациях по скринингу РМЖ указано, что пациентам с повышенным риском развития РМЖ в качестве дополнения к скрининговой маммографии нужно предлагать ежегодную МРТ молочной железы [226, 301]. МРТ-скрининг позволяет выявлять значительно большую долю злокачественных опухолей на стадиях 0-I, чем использование только маммографии, особенно среди молодых женщин высокого риска. Важно отметить, что в настоящий момент нет результатов ни одного рандомизированного исследования МРТ-скрининга у носителей мутации *BRCA1/2* или других женщин высокого риска, и выводы об улучшении выживаемости МРТ-скрининга особенно у пациентов с повышенным риском развития ТНРМЖ основаны на клинических наблюдениях и имитационном моделировании [131, 158].

Остаются нерешенными вопросы о величине кумулятивного пожизненного риска, связанного с другими генами предрасположенности к РМЖ. Предполагаемый кумулятивный пожизненный риск развития РМЖ с наследственными мутациями в генах *ATM*, *CHEK2*, *NBN* и

PALB2 приближается или превышает 30 %. Соответственно, женщины-носители мутаций в этих генах имеют право на проведение МРТ молочной железы в соответствии с современными принципами [226]. Недостаточно данных, чтобы рекомендовать МРТ молочной железы для скрининга носителей нескольких мутаций в других генах, которые могут увеличить риск развития ТНРМЖ, таких как *BARD1*, *RAD51C* и *RAD51D* [112, 203]. Неопределенность также остается относительно соответствующего возраста начала скрининга среди носителей мутаций умеренной пенетрантности, поскольку возрастная заболеваемость РМЖ с мутациями в генах, таких как ген *PALB2*, недостаточно изучена [203]. Таким образом, отягощенный семейный анамнез по РМЖ носителя мутации является решающим фактором в персонализации плана скрининга пациента, как с точки зрения используемых методов, так и с точки зрения возраста манифестации заболевания. Для женщин с семейным анамнезом ТНРМЖ, но без обнаруживаемой мутации при генетическом тестировании, семейный анамнез является ключевым фактором при разработке соответствующего режима скрининга РМЖ [354].

1.4.4. Подходы к терапии трижды негативного рака молочной железы

ТНРМЖ имеет тенденцию к более агрессивному течению, чем другие типы РМЖ. Также в отличие от других подтипов (гормонально-положительных, *HER2*-положительных) утвержденных целевых методов лечения ТНРМЖ не существует, хотя иммунотерапия (в сочетании с химиотерапией) доступна для пациентов с прогрессирующим ТНРМЖ, который экспрессирует лиганд запрограммированной гибели клеток 1 (PD-L1).

У пациентов с распространенным спорадическим метастатическим ТНРМЖ при отсутствии экспрессии PD-L1 лечение заключается в использовании одного агента при химиотерапии. Стратегии комбинированной химиотерапии могут быть подходящими для некоторых пациентов с быстро прогрессирующим метастатическим заболеванием. Возможно применение схем лечения на основе платины или без нее: выбор зависит от профиля токсичности. В метаанализе 10 рандомизированных исследований, сравнивающих платиносодержащую химиотерапию со схемами, не содержащими платину, у 958 женщин с метастатическим ТНРМЖ, уровень смертности за один год в группе пациентов, принимающих препараты платины, составил 46 % против 51 % в группе со схемами, не содержащими платину. Однако токсичность 3 и 4 степени была выше среди пациентов в группе терапии с платиной, включая тошноту/рвоту и анемию [113].

При наличии экспрессии PD-L1 в опухоли возможно добавление к химиотерапии ингибиторов иммунных контрольных точек. Атезолизумаб одобрен FDA для использования в комбинации с набпаклитакселом для пациентов с нерезектабельным местнораспространенным

или метастатическим ТНРМЖ с PD-L1-положительными инфильтрирующими опухоль иммунными клетками, покрывающими по ИГХ ≥ 1 % площади опухоли. Кроме того, пембролизумаб одобрен в сочетании с химиотерапией для пациентов с метастатическим ТНРМЖ, опухоли которых экспрессируют PD-L1 с комбинированным положительным баллом (англ. combined positive score, CPS) ≥ 10 (процент от общего количества клеток с PD-L1). Это разумная альтернатива атезолизумабу и набпаклитакселу, особенно для тех, у кого таксаны не могут быть предпочтительными, например, у пациентов с плохой переносимостью предыдущей терапии таксанами или с коротким интервалом прогрессирования на данной терапии (менее 12 месяцев). Однако данные ОВ для этого подхода еще не представлены.

Для женщин, у которых ожидаемая продолжительность жизни составляет не менее 10 лет, также могут быть предложены варианты снижения риска развития РМЖ. Например, для женщин с 5-летним риском развития РМЖ не менее 1,7 % или 20 % пожизненного риска возможно проведение химиопрофилактики с помощью таких агентов, как тамоксифен, ралоксифен или ингибиторы ароматазы, которые снижают риск РМЖ примерно на 50 % (NCCN Clinical Practice Guidelines in Oncology). Женщинам с очень высоким риском развития РМЖ возможно проведение мастэктомии.

Хотя такие методы лечения были одобрены, подтверждающие данные основаны на опыте пациентов, получавших лечение первой линии по поводу метастатического заболевания. Преимущества более позднего лечения неизвестны. Учитывая данное ограничение, пациенты, прошедшие предшествующее лечение таксанами (либо в неоадьюванте, либо в условиях метастазирования), по-прежнему являются кандидатами на комбинацию атезолизумаб/набпаклитаксел.

В рандомизированном исследовании (IMpassion 130) 902 больных метастатическим ТНРМЖ, были рандомизированы в группу набпаклитаксела либо атезолизумаба, либо плацебо [306]. Примерно половина больных получала таксаны в ранней стадии заболевания. Для включения в исследование должно было пройти не менее 12 месяцев после неоадьювантной химиотерапии. Статус генов *BRCA* не входил в критерии отбора. При среднем сроке наблюдения 13 месяцев было выявлено лишь небольшое, но статистически значимое различие в ВВП в пользу включения атезолизумаба. ВВП для тех, кто получал атезолизумаб, по сравнению с теми, кто не получал атезолизумаб, составлял 7,2 против 5,5 месяцев (отношение рисков (ОР) для прогрессирования или смерти 0,80), с незначительной тенденцией к улучшению ОВ (21,3 против 17,6 месяцев; ОР для смерти 0,84). Однако в проспективно запланированном подмножественном анализе результатов в соответствии с PD-L1-экспрессирующими иммунными эффекторными клетками в опухолях атезолизумаб улучшил как ВВП (7,5 против 5 месяцев; ОР 0,62), так и ОВ (25 месяцев против 15,5 месяцев; ОР 0,62). Нежелательные явления степени ≥ 3 имели место у 49

% пациентов, получавших атезолизумаб, и у 42 % пациентов, получавших плацебо, при этом нейропатия степени 3 или 4 чаще встречалась у больных, получавших атезолизумаб (5,5 % против 2,8 %). Среди 451 больного, получавшего атезолизумаб, было 3 смерти, связанных с лечением (0,7 %), что согласуется с другими исследованиями ингибиторов контрольных точек. Побочные эффекты привели к прекращению лечения в 16 % случаев в группе атезолизумаба по сравнению с 8 % в контрольной группе [115].

В другом исследовании, IMpassion 131, изучалось применение атезолизумаба в сочетании с паклитакселом при метастатическом ТНРМЖ первой линии с акцентом на PD-L1-положительные опухоли, определенные аналогично IMpassion 130. В отличие от IMpassion 130, значительного улучшения ВБП в PD-L1 не наблюдалось (5,7 против 6 месяцев) [240]. При этом поскольку основное различие между IMpassion 130 и 131 заключалось в химиотерапии, предпочтительной комбинацией с атезолизумабом остается набпаклитаксел.

В исследовании KEYNOTE-355 847 пациентов с местно рецидивирующим неоперабельным или метастатическим ТНРМЖ, для которых время без признаков заболевания составляло ≥ 6 месяцев, были случайным образом распределены для получения химиотерапии (набпаклитаксел, паклитаксел или гемцитабин/карбоплатин) с пембролизумабом или без него [93]. Рандомизация была стратифицирована по типу химиотерапии в исследовании (таксан или гемцитабин-карбоплатин), экспрессии PD-L1 на исходном уровне ($CPS \geq 1$ или < 1) и предыдущему лечению тем же классом химиотерапии в неоадьюванте или адьюванте. Среди больных с $CPS \geq 10$ или более медиана ВБП составила 9,7 месяцев в группе химиотерапии в сочетании с пембролизумабом и 5,6 месяцев в группе химиотерапии в сочетании с плацебо (ОР для прогрессирования или смерти 0,65). Медиана ВБП составила 7,6 и 5,6 месяцев (ОР 0,44) среди пациентов с $CPS < 10$ или более. Эффект лечения пембролизумабом увеличивался, если опухоль была обогащена PD-L1⁺ - клетками. Результаты предполагают, что польза ограничена теми, у кого $CPS \geq 10$. Нежелательные явления 3-5 степени были сопоставимы между двумя группами (примерно 70 %), хотя 2 больных в группе пембролизумаба умерли от токсичности, связанной с лечением [93].

В дополнение к испытаниям комбинированной химиотерапии для ТНРМЖ есть исследования применения иммунотерапии, например, пембролизумаба - моноклонального антитела, селективно блокирующего взаимодействие между белком программируемой клеточной гибели 1 (PD-1) и его лигандами PD-L1 и PD-L2, а также анти-PD-L1, - авелумаба и атезолизумаба [21], с частотой ответа обычно < 20 % даже в опухолях, отобранных по PD-L1. Дополнительные стратегии, включая комбинированную иммунотерапию с другой системной терапией или лучевой терапией, а также иные подходы, находятся в стадии разработки. Кроме

того, активно исследуется оптимизация биомаркеров, позволяющих прогнозировать ответ на иммунотерапию.

PARP ингибиторы могут быть использованы при РМЖ, ассоциированном с мутацией в генах *BRCA*, преимущественно ТНРМЖ. Для большинства пациентов с ТНРМЖ с мутациями в генах *BRCA*, которые ранее получали химиотерапию в условиях неoadъювантного, адъювантного или метастатического заболевания, возможно применение перорального ингибитора PARP, а не химиотерапии, поскольку имеющиеся данные указывают на его более высокую эффективность и меньшее количество побочных эффектов. Кроме того, комбинация иммунотерапии и химиотерапии является приемлемой альтернативой ингибитору PARP для пациентов с PD-L1-положительным заболеванием.

В исследовании OlympiAD в подгруппе из 121 носителя мутации в генах *BRCA* с метастатическим ТНРМЖ, где все пациенты получали антрациклины и таксаны либо в рамках адъювантного лечения, либо при метастатическом процессе, в группе пациентов, получавших олапариб, наблюдалось улучшение ВВП по сравнению с теми, кто получал химиотерапию (ОР прогрессирования или смерти 0,43) [293]. Общее исследование, в которое вошли пациенты с гормон-положительным и HER2-отрицательным заболеванием, показало положительный результат, но улучшения, отмеченные при применении олапариба, были значительно выше в популяции ТНРМЖ. Аналогичным образом в подгруппе ТНРМЖ в исследовании EMBRACA, в которое были включены пациенты с распространенным РМЖ и герминальной мутацией в генах *BRCA*, талазопариб улучшил ВВП по сравнению с химиотерапией (ОР 0,60). Следует отметить, что варианты химиотерапии с одним агентом сравнения не включали ни таксаны, ни препараты платины в этих исследованиях: сравнивали только PARP ингибиторы с терапией второй линии. Неизвестно, какой эффект PARP ингибиторы покажут в сравнении с препаратами первой линии.

В стадии клинической разработки находится несколько других PARP ингибиторов [356]. Например, велипариб (ABT-888) был протестирован в сочетании с темозоломидом, алкилирующим агентом, среди 41 женщины с распространенным ТНРМЖ (из которых 8 имели герминальную мутацию гена *BRCA*) в исследовании фазы II [180]. Хотя общий ответ и показатели клинической пользы составляли 7 % и 17 % от всей исследуемой популяции, наиболее значимые показатели были зарегистрированы среди пациентов с мутациями *BRCA*, у которых показатели общего ответа и клинической пользы составили 37,5% и 62,5% соответственно.

Применение ингибирования PARP в качестве противоопухолевой терапии обосновано тем, что PARP участвует в молекулярных процессах, ведущих к восстановлению клеток после повреждения ДНК. Когда PARP1, самый распространенный член семейства PARP, ингибируется, двухцепочечные разрывы ДНК накапливаются и в нормальных условиях репарируются посредством механизма гомологичной рекомбинации, зависящего от пути *BRCA* [346].

Предполагалось, что ингибирование PARP в сочетании с химиотерапевтическими препаратами, повреждающими ДНК, сделает опухоли, лишенные функции BRCA, чрезвычайно чувствительными, и эта гипотеза подтвердилась как в доклинических, так и в клинических исследованиях [120]. Учитывая общие клинико-патологические характеристики BRCA-мутированных РМЖ и ТНРМЖ, эффективность и безопасность ингибирования PARP проверяется в обоих условиях. Кроме того, PARP ингибиторы продолжают исследоваться на спорадическом ТНРМЖ.

Для пациентов, ранее не получавших химиотерапию, или с прогрессированием на ингибиторах PARP, химиотерапия является предпочтительным вариантом. Для подгруппы пациентов с BRCA-ассоциированным РМЖ, которые также экспрессируют PD-L1, рекомендовано применение схемы набпаклитаксел/атезолизумаб в качестве начальной схемы химиотерапии по сравнению с другими вариантами химиотерапии. Однако для пациентов с BRCA-ассоциированными и PD-L1-отрицательными опухолями как платина, так и таксаны считаются подходящими вариантами химиотерапии, причем выбор зависит от графика и соображений токсичности. В рекомендациях ASCO для носителей мутаций в генах BRCA1/2 с прогрессирующим РМЖ рекомендуется использовать препараты платины вместо таксанов [355], основываясь на рандомизированном исследовании карбоплатина по сравнению с доцетакселом в терапии первой линии ТНРМЖ.

В рандомизированном исследовании TNT напрямую сравнивали карбоплатин и доцетаксел в условиях терапии первой линии для женщин с метастатическим ТНРМЖ. Общие показатели ответа были аналогичными в группе в целом, но среди 43 женщин с известной мутацией в генах BRCA1/2 карбоплатин привел к более высокому уровню ответа (68 % против 33%) и ВБП (6,8 против 4,4 месяца) [357]. Тем не менее, исследование имело перекрестный дизайн, и не было обнаружено статистически значимой разницы в ОБ (12,8 месяцев и 12 месяцев) для тех, кто получал карбоплатин или доцетаксел, соответственно, что позволяет предположить, что любой агент может быть введен первым. Токсичность ≥ 3 степени среди получавших карбоплатин по сравнению с доцетакселом включала фебрильную нейтропению у 2 % и 20 %, диарею у 3 % и 7 %, тромбоцитопению у 7 % и 0 % соответственно. Токсичность любой степени карбоплатина по сравнению с доцетакселом включала алопецию у 35 % и 89 %, артралгии у 4 % и 21 %, диарею у 34 % и 64 %, периферическую нейропатию у 33 % и 71 % соответственно.

При более поздней линии терапии возможно применение сацитумаба говитекана. Белок Троп-2 экспрессируется в большинстве ТНРМЖ. Сацитумаб говитекан представляет собой конъюгат антитело-лекарственное средство, нацеленный на Троп-2 для селективной доставки SN-38, активного метаболита иринотекана. Он одобрен FDA для лечения взрослых пациентов с метастатическим ТНРМЖ, которые ранее получали как минимум два курса лечения

метастатического заболевания. При приеме этого агента может развиваться тяжелая нейтропения и диарея, включая случаи нейтропенического колита. В исследовании 108 пациентов с метастатическим ТНРМЖ, получавших ранее лечение, частота объективного ответа на сацитузумаб говитекан составила 34 % при медиане ВБП 5,5 месяцев, продолжительности ответа 9,1 месяца и ОВ 13 месяцев. Нежелательные явления степени ≥ 3 включали нейтропению (42 %), лейкопению (11 %), анемию (11 %) и диарею (8 %) [48].

Ингибитор иммунных контрольных точек пембролизумаб одобрен FDA для лечения нерезектабельных или метастатических опухолей с высокой MSI (англ. *microsatellite instability - high, MSI-H*) или дефицитом репарации ошибочно спаренных нуклеотидов (англ. *deficient mismatch repair, dMMR*), а также опухолей с высокой ТМВ, которые прогрессировали после предшествующего лечения и не имеют удовлетворительных альтернативных вариантов лечения. Терапия пембролизумабом возможна для пациентов с этими молекулярными маркерами, ранее не получавшим иммунотерапию, при условии, когда химиотерапия (и ингибиторы *PARP* для носителей мутаций в генах *BRCA*) больше не эффективна или при развитии неприемлемой токсичности. Исследования, подтверждающие одобрение пембролизумаба по приведенным показаниям, не включали пациентов с РМЖ, но эффективность была продемонстрирована при других типах рака, включая рак шейки матки, эндометрия и яичников.

Рецептор эпидермального фактора роста (англ. *epidermal growth factor receptor, EGFR/HER1*) является наиболее известным белком, сверхэкспрессируемым в ТНРМЖ, для которого существует несколько моноклональных антител и низкомолекулярных ингибиторов. На сегодняшний день 3 клинических исследования фазы II оценили эффективность цетуксимаба, моноклонального антитела против EGFR, в комбинации с химиотерапией при поздних стадиях ТНРМЖ и показали, что он имеет лишь умеренную активность [74]. Точно так же исследования ингибиторов EGFR не показали клинического воздействия на ТНРМЖ и, соответственно, не рекомендуются для лечения ТНРМЖ.

РА экспрессируются как в нормальной, так и в злокачественной ткани молочной железы [372]. В то время как РМЖ, положительный по гормональным рецепторам, чаще экспрессирует РА (91 %), при ТНРМЖ РА встречаются примерно в 30 % случаев [89]. Прогноз для пациентов с РА-положительным ТНРМЖ оказался более благоприятным, чем для пациентов с РА-отрицательным [211]. В нескольких исследованиях была предпринята попытка определить противоопухолевую активность ингибирования РА при распространенном ТНРМЖ. В одном исследовании сообщалось о 6-месячном клиническом эффекте 19 % для антагониста РА бикалутамида среди 50 пациентов (12 % из 424 прошедших скрининг) с метастатическим РА-положительным ТНРМЖ [148]. Эффективный ингибитор РА энзалутамид был исследован среди 118 пациентов с РА-положительным метастатическим ТНРМЖ (55 % из 404 обследованных)

[351]. Наблюдались 2 полных ответа и 5 частичных ответов. В настоящий момент данная лечебная стратегия не до конца изучена, однако предпринимаются попытки для выявления надежного биомаркера для группы пациентов, которые будут реагировать на ингибирование РА.

Считается, что ангиогенез является важной мишенью для лечения злокачественных опухолей. Однако на сегодняшний день проспективные исследования не показали, что включение ингибиторов ангиогенеза влияет на ОВ пациентов с ТНРМЖ. Поэтому использование ингибиторов ангиогенеза в адъювантном режиме или при метастатическом заболевании для данных пациентов не показано. Из препаратов этого класса бевацизумаб изучен наиболее широко, при этом данные исследований показывают, что хотя включение бевацизумаба может улучшить ВВП, оно практически не влияет на ОВ [71]. Это, по-видимому, применимо и для пациентов с ТНРМЖ, когда бевацизумаб использовался в адъюванте и при прогрессировании в качестве первой линии [239].

Доказательства того, что комбинация иммунотерапии и химиотерапии на ранней стадии заболевания может быть полезна, получены из небольшого исследования неoadъювантной терапии, в ходе которого был обнаружен улучшенный патологический полный ответ с применением пембролизумаба с добавлением химиотерапии на основе антрациклина/таксана [253]. Было отмечено появление аутоиммунных осложнений, что показывает необходимость получения окончательных результатов исследования для возможного включения ингибиторов иммунных контрольных точек в стандарты лечения неметастатического РМЖ.

1.5. Рак поджелудочной железы как модель для испытания эффективности платформ прецизионной медицины

Рак поджелудочной железы (РПЖ) — злокачественное новообразование, которое развивается из эпителиальной ткани различных отделов поджелудочной железы. Из локализаций преобладает головка (63,8 %), реже рак формируется в теле (23,1 %) или хвосте (7,1 %) железы [4, 5]. Примерно 95 % случаев РПЖ представляют собой опухоли экзокринных клеток, чаще всего это протоковая аденокарцинома поджелудочной железы. Эндокринный РПЖ, как правило, представляет собой более медленно развивающуюся опухоль с более благоприятным прогнозом [86]. Средний возраст при постановке диагноза РПЖ составляет 70 лет.

Эпидемиологические данные указывают на то, что РПЖ занимает 12-е место по распространенности у мужчин, 11-е место по распространенности у женщин и 7-е место среди причин смертей, связанных с раком [193]. В России частота РПЖ достигает 8,8 случая на 100 тыс. населения [13]. Предраковые изменения поджелудочной железы получили название панкреатической интраэпителиальной неоплазии (англ. pancreatic intraepithelial neoplasia, PanIN).

Их подразделяют на 3 группы: PanIN-1 — интраэпителиальная протоковая гиперплазия; PanIN-2 — дисплазия низкой степени и PanIN-3 — выраженная дисплазия протокового эпителия и/или аденокарцинома [13].

РПЖ имеет наихудший прогноз из всех распространенных солидных ЗНО: 5-летняя ОВ составляет примерно 10%. Уровень заболеваемости и смертности для РПЖ увеличивался в среднем на 0,3% в год в течение последнего десятилетия [168]. Хотя наблюдается умеренное улучшение выживаемости пациентов с РПЖ, абсолютное число людей, умирающих от этого заболевания, продолжает расти. В основе таких тенденций лежит сочетание старения населения, увеличения ожидаемой продолжительности жизни и распространения пандемий ожирения и диабета среди населения, а также особенности онкогенеза РПЖ.

Существует несколько фундаментальных проблем, которые лежат в основе высокой смертности от РПЖ. Во-первых, поджелудочная железа расположена глубоко в верхней части живота, за желудком и между аортой и ее главными верхними ветвями. Это не только усложняет выявление новообразований, но и, поскольку распространение опухоли захватывает и сосуды, только в 15-20% случаев возможна хирургическая резекция, которая является основным способом лечения РПЖ. На ранних стадиях РПЖ протекает бессимптомно. По данным анализа канцер-регистра России в 2021 г., РПЖ выявлен на IV стадии у 59,5% пациентов [3]. Кроме того, РПЖ демонстрирует агрессивное течение, характеризующееся ранним метастазированием. Более 50% пациентов с РПЖ имеют отдаленные метастазы при поступлении, и у большинства пациентов, подвергшихся резекции, метастазы развиваются в течение последующих 4 лет после операции, что свидетельствует о фактическом наличии микрометастазов у пациентов с локализованными опухолями. Именно поэтому показатели выживаемости драматичны: смертность в течение первого года у пациентов с метастатическим РПЖ составляет 67,3%, 5-летняя выживаемость - 3% [3]. И, наконец, патофизиологические процессы при РПЖ могут резко ослаблять пациентов, ограничивая их способность противостоять агрессивному лечению. Так, синдром кахексии присутствует почти у 80% пациентов с РПЖ на момент постановки диагноза, и это может дополнительно осложняться экзокринной и эндокринной дисфункцией поджелудочной железы. Пациенты с кахексией плохо переносят лечение, о чем свидетельствует снижение выживаемости после панкреатэктомии или химиотерапии [146].

Полногеномное и экзомное секвенирование, проведенные Международным консорциумом по геному рака (англ. International Cancer Genome Consortium, ICGC) и Атласом генома рака (англ. The Cancer Genome Atlas, TCGA), способствовали изучению генетических особенностей РПЖ [54, 361]. В настоящее время результаты анализа генов, участвующих в молекулярном патогенезе РПЖ, позволили по-новому классифицировать новообразования поджелудочной железы. В новой классификации учитываются как морфологические, так и

генетические особенности [50, 54, 186, 361]. Однако выявление огромного количества генетических изменений, в том числе соматических мутаций, выходит за рамки рутинных клинико-генетических лабораторных исследований, что тормозит развитие персонализированного лечения этой категории больных.

1.5.1. Генетические особенности рака поджелудочной железы

В аденокарциноме поджелудочной железы обычно обнаруживаются множественные комбинации генетических мутаций [54, 186, 361]. Генетические изменения, характеризующие РПЖ можно подразделить на три основные категории: мутационная активация онкогенов, таких как *KRAS*; инактивация генов-супрессоров опухолей, таких как *TP53*, *p16*/ингибитор циклин-зависимой киназы 2A (*CDKN2A*) и *SMAD4*; инактивация генов поддержания целостности генома, таких как гомолог 1 *mutL* (*MLH1*), гомолог 2 *mutS* (*MSH2*) и гомолог 6 *mutS* (*MSH6*), которые контролируют восстановление повреждений ДНК. Хотя большинство этих генетических aberrаций представляют собой соматические мутации, другие присутствуют в зародышевой линии родственников, несущих семейную предрасположенность к раку поджелудочной железы [290, 397].

Патогенные варианты герминальных мутаций в генах предрасположенности к развитию злокачественной опухоли выявляются примерно у 4-20 % пациентов с РПЖ, включая пациентов с явно спорадическими опухолями [139, 174]. Однако примерно половина пациентов с РПЖ, у которых обнаружен патогенный генетический вариант, не имеют семейного анамнеза РПЖ и/или не соответствуют клиническим критериям наследственного синдрома. В целом, при развитии протоковой аденокарциномы поджелудочной железы могут быть задействованы 4 основных гена-драйвера: 1 онкоген *KRAS* и 3 гена-супрессора опухоли *CDKN2A*, *TP53* и *SMAD4*.

Ген *KRAS*, расположенный на хромосоме 12p, является одним из наиболее часто мутирующих генов при РПЖ. Этот ген является человеческим гомологом трансформирующего гена, выделенного из вируса саркомы крыс Кирстен [174]. Большинство мутаций в этом гене происходят в кодоне 12, что приводит к аномальной активации белкового продукта гена [169]. Более 90 % случаев РПЖ содержат мутацию гена *KRAS* [169, 186, 365]. Кроме того, эти мутации, по-видимому, возникают на очень ранних стадиях канцерогенеза, на что указывает их присутствие в неинвазивных - предопухолевых образованиях. Так, мутации гена *KRAS* были выявлены в неинвазивных внутрипротоковых папиллярных муцинозных новообразованиях (ВПМН), PanIN и в неинвазивных муцинозно-кистозных новообразованиях, причем распространенность этих мутаций повышается с увеличением степени этих неинвазивных поражений [32, 166, 183, 247]. Исследования на мышах предоставляют убедительные

доказательства того, что онкогенные мутации в гене *KRAS* необходимы для образования PanIN, наиболее частого предшественника РПЖ, а также для инициации и развития инвазивного РПЖ [24, 90, 161]. Поскольку мутации гена *KRAS* являются частыми и ранними событиями при неоплазии поджелудочной железы, ген *KRAS* является мишенью для разработки диагностического теста для раннего выявления РПЖ.

При карциноме поджелудочной железы была выявлена потеря функционирования нескольких генов-супрессоров опухоли. Чтобы нарушить функцию гена, необходимо инактивировать обе копии (как материнские, так и отцовские аллели) гена. Это может происходить по одному из трех механизмов: инактивация одной копии гена внутригенной мутацией в сочетании с потерей второго аллеля (так называемая потеря гетерозиготности); гомозиготная делеция (делеция обеих копий гена) или гиперметилирование промотора гена. Более чем в 50 % всех случаев РПЖ инактивируются следующие гены-супрессоры опухолей: *p16/CDKN2A*, *TP53* и *SMAD4* (ранее известный как *DPC4*) [46, 186].

Ген *p16*/ингибитор циклин-зависимой киназы 2А (*CDKN2A*) на хромосоме 9р соматически инактивирован приблизительно в 95 % случаев РПЖ [46, 70, 186, 307]. Большинство этих инактивирующих мутаций приводят к потере функции *p16*, белкового продукта гена *CDKN2A*. В 40 % случаев рака ген инактивирован гомозиготной делецией; в 40 % это внутригенная мутация, связанная с делецией второго аллеля; и в 15 % случаев инактивация гена происходит за счет гиперметилирования промотора гена *p16/CDKN2A*. Инактивация гена *p16/CDKN2A* при РПЖ важна по нескольким причинам. Во-первых, потеря функции этого гена отменяет контроль клеточного цикла в этих опухолях. Во-вторых, унаследованные мутации в гене *p16/CDKN2A* являются одной из причин синдрома семейной атипичной множественной меланомы [397]. Пациенты с данным синдромом имеют повышенный риск развития меланомы и повышенный в 20–34 раза риск развития РПЖ. И, в-третьих, при гомозиготных делециях, которые инактивируют *p16/CDKN2A*, часто также инактивируется соседний ген *MTAP* [233, 338]. Данные, полученные при использовании клеточных линий, предполагают, что инактивация гена *MTAP* при РПЖ может быть использована для разработки таргетных препаратов [200, 233].

Ген *TP53* на хромосоме 17р является геном, мутации в котором часто встречаются при ЗНО, включая РПЖ. Ген *TP53* инактивируется в 75–85 % случаев РПЖ, почти всегда в результате внутригенной мутации (которая чаще всего приобретает соматически) в сочетании с потерей второго аллеля [46, 54, 186, 361]. Инактивация гена *TP53* нарушает две важные функции клетки: регуляцию клеточной пролиферации и гибель клеток (апоптоз) в ответ на повреждение ДНК.

Ген *SMAD4*, расположенный на хромосоме 18q, инактивирован примерно в 50 % случаев РПЖ [46, 54, 186, 361]. В 30 % опухолей ген инактивирован гомозиготной делецией, а еще в 20 % случаев наблюдается внутригенная мутация, связанная с потерей второго аллеля. Белковый

продукт гена *SMAD4* участвует в передаче внутриклеточных сигналов от рецепторов трансформирующего фактора роста бета (TGF β) внутриклеточной мембраны к ядру [101]. Мутации в генах, кодирующих другие компоненты сигнального пути TGF β , такие как *SMAD3*, *TGF β R1* и *TGF β R2*, также были обнаружены при РПЖ, что подчеркивает важность этого сигнального пути для развития РПЖ [46, 54, 186, 361]. Инактивация гена *SMAD4* при РПЖ важна по нескольким причинам. Во-первых, для белка *SMAD4* было разработан метод ИГХ мечения срезов тканей, результаты которого сильно коррелируют со статусом гена *SMAD4* [373]. Это означает, что ИГХ окрашивание на наличие или отсутствие белка *SMAD4* может использоваться с диагностической целью в качестве дополнительного метода при интерпретации спорных результатов биопсий тканей. Также это позволяет предполагать РПЖ как возможный первичный диагноз у пациентов с метастатической аденокарциномой без первично выявленного очага [171, 373]. Например, если у пациента был РПЖ, при котором ген *SMAD4* генетически инактивирован, а позже в легких развилась аденокарцинома, которая имела интактную экспрессию *SMAD4*, можно сделать вывод, что опухоль в легком не являлась метастазом РПЖ. РПЖ с утратой экспрессии *SMAD4* имеет более высокую частоту отдаленных метастазов и худший прогноз [56, 179]. Хотя необходимы дальнейшие исследования, эти данные предполагают, что статус гена *SMAD4* может стать полезным для прогностической стратификации и принятия терапевтических решений.

Гены системы репарации неспаренных оснований ДНК мутируют приблизительно в 2-3% случаев РПЖ [46, 54, 177, 361]. Некоторые из таких видов рака имеют характерный «медуллярный» гистологический вид, поэтому РПЖ с дефектами в генах репарации несоответствия ДНК иногда может быть распознан гистологически с помощью диагностического микроскопа [374].

До 20 % пациентов с диагнозом РПЖ имеют герминальные мутации в генах предрасположенности к развитию злокачественных опухолей. Так, например, мутации в генах *BRCA1* и *BRCA2* (одна из причин синдрома наследственного РМЖ и яичников) на хромосоме 13q связаны с повышенным риском РПЖ [163, 397]. Унаследованная герминальная мутация затем сочетается с соматической потерей второго аллеля, полностью инактивируя функцию гена в опухоли [172]. Герминальные мутации в гене *BRCA2* увеличивают риск РПЖ в 3,5–10 раз [157]. Это составляет одну из наиболее важных причин наследственного РПЖ. Мутации в гене *BRCA2* обнаруживаются у 17 % пациентов, у которых несколько членов семьи имели РПЖ. Белковый продукт гена *BRCA2* функционирует при анемии Фанкони и играет важную роль в восстановлении повреждений, вызванных сшивающими агентами ДНК [157]. Было высказано предположение, что эту функцию *BRCA2* можно использовать для лечения. Исследования *in vitro* показали, что РПЖ с генетически инактивированным геном *BRCA2* значительно более

восприимчив к агентам «сшивания» ДНК, чем РПЖ с генетически интактным геном *BRCA2*. Действительно, в нескольких отчетах документально подтверждены значительный терапевтический ответ у пациентов с РПЖ, содержащим инактивированный ген *BRCA2*, на агенты сшивания ДНК (такие как митомицин), на производные платины (такие как цисплатин) или на PARP ингибиторы [141, 189, 223]. Препарат на основе платины оксалиплатин входит в состав химиотерапевтической комбинации FOLFRINOX (фторурацил, лейковорин, иринотекан и оксалиплатин) для лечения распространенного РПЖ. Герминальные мутации гена *BRCA1* также могут предрасполагать к РПЖ, но пенетрантность меньше, чем при герминальных мутациях гена *BRCA2* [172].

Ген *PALB2* расположен на хромосоме 16p и кодирует белок, связывающий белок *BRCA2* [382]. Герминальные мутации в гене *PALB2* увеличивают риск развития РМЖ, и были выявлены примерно у 3 % пациентов с наследственной формой РПЖ [323, 384, 398]. Герминальные мутации в гене *PALB2* увеличивают риск РПЖ в 15 раз [173]. Поскольку ген *PALB2*, подобно генам *BRCA1* и *BRCA2*, является членом пути анемии Фанкони, РПЖ, при котором *PALB2* был генетически инактивирован, должен иметь такую же специфическую терапевтическую чувствительность, которая наблюдалась при РПЖ с мутациями в гене *BRCA2* [359].

Ген серин/треонинкиназы 11 (*STK11*) на хромосоме 19p функционирует как ген-супрессор опухоли и кодирует этот фермент, который регулирует полярность клеток (Jenne et al., 1998). Герминальные мутации в гене *STK11* связаны с синдромом Пейтца-Егерса (СПЕ), аутосомно-доминантным заболеванием, при котором развиваются гамартоматозные полипы желудочно-кишечного тракта, пигментные пятна на губах и слизистой оболочке щек, а также различные ЗНО желудочно-кишечного тракта [137, 302]. Пациенты с СПЕ имеют существенно повышенный риск развития РПЖ [137]. Кроме того, соматические мутации в гене *STK11* наблюдались примерно в 4 % РПЖ, особенно в возникающих из ВПМН [32, 302]. Утрата гетерозиготности по *STK11* сообщается в 25 % ВПМН у пациентов с отсутствием признаков СПЕ [302]. Таким образом, инактивация гена *STK11* играет роль как в наследственном, так и в спорадическом РПЖ. Скрининг лиц с СПЕ и другими наследственными заболеваниями высокого риска с использованием методов визуализации на основе эндоУЗИ или МРТ может выявить бессимптомные ранние новообразования поджелудочной железы, включая ВПМН, на этапе, когда они потенциально излечимы [140, 207].

Ген, мутировавший при атаксии-телеангиэктазии (*ATM*), на хромосоме 11q кодирует белок семейства PI3/PI4-киназ. Продукт гена *ATM* играет важную роль в ответе клетки на повреждение ДНК [337]. Герминальные мутации в гене *ATM* вызывают атаксию-телеангиэктазию, нейродегенеративное заболевание, характеризующееся плохой координацией (атаксией) и расширенными сосудами (телеангиэктазиями). У людей с атаксией-

телеангиэктазией риск развития рака в течение жизни составляет 25 %. Герминальные мутации в гене *ATM* также были зарегистрированы у 3% семей с семейным РПЖ, а соматические (приобретенные) мутации в гене *ATM* - в протоковых аденокарцинах [54, 144, 290]. Герминальные мутации в гене *ATM* увеличивают риск РПЖ в 8-9 раз [172]. Эти мутации представляют особый интерес, так как РПЖ, при котором ген *ATM* двуваллельно инактивирован, может быть более чувствительным к определенным методам лечения [83, 375].

Гены репарации несоответствия ДНК (mismatch repair, MMR), такие как гомолог 1 *mutL* (*MLH1*), гомолог 2 *mutS* (*MSH2*) и гомолог 6 *mutS* (*MSH6*), хорошо известны своей важной ролью в патогенезе колоректального рака, особенно при синдроме Линча (наследственный неполипозный колоректальный рак). Пациенты с синдромом Линча имеют унаследованные (герминальные) мутации в одном из нескольких генов восстановления несоответствия ДНК и повышенный риск развития нескольких видов рака желудочно-кишечного тракта, включая РПЖ [173, 188]. Карциномы поджелудочной железы с инактивированным геном репарации несоответствия ДНК важно распознавать по нескольким причинам. Во-первых, как и в случае с толстой кишкой, соматическая инактивация гена репарации ДНК у пациента с диагностированным РПЖ предполагает возможность герминальной мутации и синдрома Линча [188, 373, 376]. Все пациенты с диагностированным РПЖ с MSI-H или dMMR должны пройти тестирование на наличие герминальных мутаций на предмет синдрома Линча, независимо от возраста или семейного анамнеза [208]. Эти пациенты (и члены их семей) подвержены риску различных новообразований желудочно-кишечного тракта. Во-вторых, РПЖ с MSI-H, отличительной чертой которой является наличие инактивированного гена репарации ошибочного спаривания ДНК, по-видимому, имеет несколько лучший прогноз, чем стандартные протоковые аденокарциномы [251]. В-третьих, РПЖ с MSI-H также может быть менее чувствительным к химиотерапевтическим агентам, таким как фторурацил, как и рак толстой кишки, по крайней мере, в условиях адьюванта [288]. Вместо этого некоторые виды рака с MSI-H чувствительны к иммунной терапии, такой как ингибиторы иммунных контрольных точек. Объективные, в некоторых случаях полные и стойкие ответы на ингибиторы иммунных контрольных точек, такие как пембролизумаб, были зарегистрированы у различных пациентов со статусом MSI-H [209].

Мутации и в ряде других генов были идентифицированы при РПЖ [186]. Эти гены, как правило, менее важны, чем *KRAS*, *p16/CDKN2A*, *TP53* и *SMAD4*, но также вносят вклад в патогенез РПЖ. Это такие гены как *AKT2*, *AIB1*, *ALK4*, *ARID1A*, *AXIN1*, *BRAF*, *CDH1*, *CDK4*, *CTNNB1*, *EP300*, *EPHA3*, *ERBB2*, *FBXW7*, *GATA6*, *MAP2K4*, *MKK4*, *MLL3*, *MYB*, *MYC*, *Notch1*, *PBRM1*, *PIK3CA*, *Robo2*, *SMARCA2*, *SMARCA4*, *TGFbR2* и некоторые другие [54, 186, 361]. Кроме того, инактивирующие мутации в гене *RNF43*, компоненте убиквитин-зависимых путей, и *GNAS*,

который кодирует альфа-субъединицу белка, связывающего гуанин-нуклеотид, также были идентифицированы в ВПМН поджелудочной железы, а также в инвазивных типах рака, возникающих из ВПМН [32, 380, 381]. Сходным образом были идентифицированы специфические транслокации с участием генов *PRKACA* и *PRKACB*, которые вызывают внутрипротоковые онкоцитарные папиллярные новообразования и связанные с ними инвазивные карциномы [322]. Несмотря на большое количество генов, продукты многих из них включены в общие сигнальные пути, и сложность спектра мутаций при РПЖ можно упростить, если рассматривать клеточные пути, на которые нацелены эти генетические изменения. Например, в ходе секвенирования 24 образцов РПЖ было обнаружено, что гены, регулирующие апоптоз, фазовый переход G1/S во время деления, сигнальные пути hedgehog, KRAS, TGF β и Wnt/Notch, содержали мутации во всех случаях [186].

В дополнении к мутационной инактивации, экспрессия ряда генов при РПЖ может быть подавлена aberrантным метилированием промотора, т.е. эпигенетическим способом прекращения работы генов [261, 303, 304]. К генам, работа которых при РПЖ может быть подавлена таким образом, относятся *UCHL1*, *NPTX2*, *SARP2*, *CLDN5*, *REPRIMO (RPRM)*, *LHX1*, *Wnt7a*, *FOXE1*, *TJP2*, *CDH3*, *ST14* и *p16/CDKN2A* [304]. Напротив, некоторые из генов, которые сверхэкспрессируются при РПЖ, гипометилированы [304]. Сейчас исследуют различия в паттернах метилирования между новообразованиями поджелудочной железы и нормальными клетками для разработки тестов ранней диагностики и новых подходов к терапии [164, 261, 274, 303, 304]. Необходимы дальнейшие исследования, прежде чем любой из этих подходов можно будет использовать в повседневной клинической практике.

Мутации чаще всего происходят в ядерной ДНК, но в митохондриальной ДНК также обнаруживаются соматические мутации при РПЖ [185, 227]. Однако остается неясным, играет ли какая-либо из этих мутаций роль в онкогенезе. В то же время, митохондриальные мутации могут быть использованы для разработки тестов ранней диагностики РПЖ. Митохондриальные мутации являются особенно привлекательной мишенью для раннего выявления РПЖ, поскольку клетки содержат тысячи митохондрий, и поэтому клональную мутацию в митохондриях легче обнаружить, чем клональную мутацию в ядерной ДНК [185].

МикроРНК (англ. microRNA, miRNA) – короткие, некодирующие сегменты РНК, которые регулируют экспрессию других генов. Ряд miRNAs по-разному экспрессируется при РПЖ и предшествующих поражениях, которые приводят к РПЖ [57, 152, 297, 342]. Например, MiR-34a часто подвергается делеции при РПЖ, при этом эта miRNA, по-видимому, играет роль в экспрессии генов, связанных с TP53 [78].

Генетические исследования небольших внутрипротоковых поражений в поджелудочной железе, помогли установить, что PanIN может быть предшественником инвазивной протоковой

аденокарциномы [166, 170, 249]. Полное экзомное секвенирование PanIN и других протоковых аденокарцином показало значительную долю общих соматических мутаций в большинстве случаев, подтверждая предположение о том, что PanIN создает риск протоковых аденокарцином [249]. PanIN несут в себе многие из тех же мутаций, что и при инвазивных протоковых аденокарциномах, и увеличение количества мутаций ассоциировано с увеличением степени дисплазии в PanIN. Мутации в гене *KRAS*, по-видимому, возникают первыми, при самом низком уровне PanIN, называемом PanIN-1. Дальше появляются мутации в гене *p16/CDKN2A* (в поражениях PanIN-2), а инактивация генов *SMAD4* и *TP53* происходит только при высшей степени PanIN - поражениях PanIN-3 или, возможно, даже позже, при инвазивных карциномах [166, 170].

Разработка модели онкогенеза РПЖ имеет важное значение для определения стратегий профилактики, раннего обнаружения и лечения этого типа ЗНО. Например, очень ранние молекулярные события могут быть подходящими мишенями для химиопрофилактики, тогда как более поздние события могут быть полезны для стратегий раннего выявления неоплазии поджелудочной железы. Кроме того, понимание онкогенеза РПЖ имеет решающее значение для разработки генно-инженерных моделей РПЖ на животных. Осознание того, что мутации в гене *KRAS* являются одним из самых ранних событий в онкогенезе поджелудочной железы, побудило исследователей ввести мутантный *KRAS* в поджелудочные железы мышей. В результате у этих мышей развились опухоли поджелудочной железы, которые имитируют аналогичное заболевание у человека [161, 162].

Генетические изменения, определяемые в новообразованиях поджелудочной железы, начинают исследоваться в сопоставлении с морфологией опухоли и прогнозом больного, а также новой «молекулярной классификацией» неоплазий поджелудочной железы [319]. Так, почти все солидные псевдопапиллярные новообразования поджелудочной железы содержат мутации гена бета-катенина (*CTNNB1*), и эти мутации могут объяснять недостаточные контакты клеток между собой, что характерно для неопластических клеток при данном типе опухолей [19, 344, 381]. Серозные цистаденомы поджелудочной железы почти всегда являются доброкачественными и медленно растущими новообразованиями; для них характерны мутации в гене фон Хиппеля-Линдау (*VHL*). В спорадических условиях они приобретаются, тогда как у пациентов с синдромом Хиппеля-Линдау они передаются по наследству. ВПМН часто содержат мутации в генах *GNAS*, *RNF43*, *KRAS*, *TP53* и *SMAD4*, вместе с тем, для муцинозных кистозных новообразований характерен сходный мутационный профиль, за исключением мутации в гене *GNAS* [381]. Транслокации с участием либо гена *PRKACA*, либо гена *PRKACB* в свою очередь свойственны внутрипротоковым папиллярным новообразованиям поджелудочной железы [322]. Результаты проведенных исследований свидетельствуют о том, что между основными

кистозными новообразованиями поджелудочной железы наблюдаются отличия в выявляемых мутациях [381]. Ожидается, что молекулярное профилирование жидкости таких кист, полученное во время эндоскопии, поможет в клинической классификации типа исследуемого новообразования [331, 389].

Подавляющее большинство недифференцированных карцином поджелудочной железы демонстрируют потерю экспрессии гена кадгерина *CDH1*, утрата которого может коррелировать с инвазивным фенотипом и плохим прогнозом [377]. Также описана недифференцированная карцинома поджелудочной железы с «рабдоидными» признаками с потерей SMARCB1 (INI1) [23].

Большинство панкреатобластом, которые возникают в детстве, утратили материнскую копию хромосомы 11p [18]. Такая генетическая характеристика объединяет данное заболевание с другими опухолями детского возраста, такими как нефробластома (опухоль Вильмса) и гепатобластома, которые также обычно утрачивают материнскую копию хромосомы 11p.

Направление дифференцировки новообразований поджелудочной железы с выраженными остеокластоподобными гигантскими клетками остается неясным. Однако теперь очевидно, что недифференцированные плеоморфные клетки, которые находятся между остеокластоподобными гигантскими клетками, несут генетические изменения, которые идентичны тем, которые обнаруживаются в соседних эпителиальных предшественниках [225, 371]. Эти наблюдения помогли установить, что такие новообразования на самом деле являются карциномами, и поэтому название этих опухолей было изменено на недифференцированные карциномы с остеокластоподобными гигантскими клетками.

Несколько более поздних исследований использовали экспрессию генов для более подробной классификации РПЖ [46, 244]. Было предложено разделить РПЖ на 4 подтипа (плоскоклеточный, иммуногенный, панкреатический предшественник, aberrantly дифференцированный эндокринно-экзокринный (Aberrantly Differentiated Endocrine Exocrine, ADEX)), или 2 подтипа (базальный, классический). Некоторые из них, например, плоскоклеточный подтип, соответствуют хорошо установленным гистологическим подтипам (аденосквамозная карцинома), в то время как иммуногенный подтип и подтип ADEX могут быть артефактами в первоначальных исследованиях [155].

1.5.2. Особенности диагностики рака поджелудочной железы

Всем пациентам с диагнозом РПЖ целесообразно пройти обследование для предварительного определения риска семейной предрасположенности к данному заболеванию [336]. Эта оценка включает подробный анамнез и определение наследственной отягощенности

(у родственников первой и второй степени родства), гистологический подтип опухоли, возраст на момент постановки диагноза.

Оптимальная стратегия скрининга пациентов с риском наследственного РПЖ неясна. Рекомендации по скринингу РПЖ были опубликованы несколькими группами исследователей [41, 140]. В соответствии с рекомендациями Американской гастроэнтерологической ассоциации, Международного Консорциума по скринингу рака поджелудочной железы (*англ.* Cancer of the Pancreas Screening Study - CAPS), Американского колледжа гастроэнтерологии, общий популяционный скрининг пациентов со средним риском (примерно 1 из 60 в США) РПЖ не рекомендуется [41, 265].

Кандидатами на скрининг РПЖ являются следующие группы больных:

- 1) пациенты с СПЕ, независимо от семейного анамнеза;
- 2) пациенты с мутациями в гене *CDKN2A* (p16), независимо от семейного анамнеза;
- 3) пациенты с мутацией в гене *BRCА 2* и с хотя бы 1 заболевшим родственником первой степени родства или хотя бы с 2 больными родственниками любой линии родства;
- 4) пациенты с мутацией в генах *BRCА1*, *PALB2*, *ATM* (кодирует мутантный при атаксии-телеангиэктазии белок) и генах, связанных с синдромом Линча при условии наличия 1 или нескольких родственников с диагнозом РПЖ;
- 5) пациенты с наследственным панкреатитом и мутацией в гене *PRSS1*.

Также кандидатами на скрининг РПЖ независимо от мутационного статуса являются больные, которые имеют:

- 1) по крайней мере, 2 родственников с диагнозом РПЖ по одной стороне родословной, из которых хотя бы 1 является родственником первой линии;
- 2) по крайней мере, 2 родственников с диагнозом РПЖ, которые являются родственниками первой линии друг для друга, из которых хотя бы 1 является родственником первой линии для рассматриваемого пациента.

В настоящий момент нет единого мнения относительно наилучшего метода визуализации для скрининга лиц с высоким риском РПЖ, и имеется недостаточно проспективных данных, сравнивающих результаты различных стратегий скрининга. Однако эндоУЗИ и/или визуализация на основе МРТ, по-видимому, дает самый высокий результат при отсутствии лучевой нагрузки [41]. Доказательств, подтверждающих использование биомаркеров для скрининга РПЖ, недостаточно.

В нескольких исследованиях оценена эффективность визуализации в обнаружении раннего РПЖ и преинвазивных поражений поджелудочной железы. При этом сравнение диагностической эффективности и точности скрининговых тестов было ограничено вариабельностью исследуемых популяций, типом методов скрининга и протоколами

визуализации. Результаты показали наличие ограничений у УЗИ и МРТ/МРХПГ (МР-холангиопанкреатография); в то же время они обнаруживают больше поражений по сравнению с КТ (компьютерная томография). Так, в сравнительном исследовании эндоУЗИ, МРТ/МРХПГ и КТ на 225 бессимптомных пациентах с высоким риском РПЖ (исследование CAPS3) у 216 с помощью эндоУЗИ и МРТ было выявлено значительно больше (в основном, небольших кистозных) поражений поджелудочной железы по сравнению с КТ (43, 33 и 11%, соответственно) [72]. При кистозных поражениях МРХПГ обеспечивает лучшую визуализацию связи кисты с основным протоком поджелудочной железы.

В другом проспективном исследовании анализировали результаты обследования 139 человек из группы высокого риска, которые прошли эндоУЗИ и МРТ в первоначальном скрининге РПЖ [153]. Клинически значимыми считались любые солидные образования, кисты размером ≥ 10 мм, внутрипротоковые папиллярные муцинозные новообразования главного протока. Всего с помощью эндоУЗИ и МРТ было обнаружено 11 клинически значимых поражений у 9 лиц из группы высокого риска (6 %); 6 поражений из 11 были обнаружены обоими способами; 2 дополнительных поражения были очевидны только на эндоУЗИ (единственные 2 солидных поражения, которые были обнаружены при первоначальном скрининге), и 3 дополнительных поражения (все три кисты ≥ 10 мм) были обнаружены только с помощью МРТ.

Несколько маркеров могут быть также использованы с целью улучшения ранней диагностики новообразований поджелудочной железы, не определяемых с помощью методов визуализации, например, для определения высокого риска развития РПЖ [213]. Наиболее используемый опухолевый маркер сыворотки для РПЖ – антиген СА 19-9. Имеются ограниченные данные об эффективности использования СА 19-9 для диагностики у пациентов с высоким риском РПЖ. В одном исследовании скрининга лиц, у которых есть хотя бы 1 родственник с диагнозом РПЖ, сывороточный СА 19-9 был повышен у 27 человек из 546 (4,9 %). Неопластические или ЗНО были обнаружены при последующем эндоУЗИ только у 5 пациентов (0,9 %), и только у 1 была аденокарцинома поджелудочной железы (0,2%) [402]. Несмотря на ограниченные данные о клинической пользе СА 19-9 у лиц с высоким риском, текущие международные рекомендации указывают на возможность использования данного анализа при наличии подозрений на новообразование поджелудочной железы [140].

Аберрантная экспрессия miRNA является потенциальными диагностическим маркером для нескольких солидных опухолей, включая РПЖ [184]. Исследование множественных микроРНК, выполняемое в сочетании с СА 19-9, может повысить точность диагностики РПЖ. Однако необходимы дополнительные исследования в когортах высокого риска.

Исследования, изучающие комбинации опухолеспецифических циркулирующих белков (включая СА 19-9) и метаболитов, а также циркулирующих белков и мутаций в свободной ДНК

в крови, предполагают раннее выявление потенциально операбельного РПЖ, но необходимы дополнительные исследования в когортах высокого риска, особенно у пациентов с панкреатитом. Регулярный мониторинг уровня глюкозы в крови (уровень глюкозы в крови натощак и/или гемоглобин А1С) может рассматриваться как начальный анализ и при постоянном наблюдении, поскольку отклонения от нормы при тестировании требуют дальнейшей оценки [140]. Секрет поджелудочной железы, жидкость из кисты, полученная с помощью тонкоигольной биопсии под контролем эндоУЗИ, могут быть проанализированы на молекулярные маркеры. NGS может быть полезно для обнаружения низкочастотных мутаций в секрете поджелудочной железы и жидкости кисты поджелудочной железы [32]. Несмотря на то, что данных недостаточно, генетические изменения (аберрантное метилирование, мутантный онкоген *Kras*, инактивированные гены-супрессоры опухолей *SMAD4* и *TP53*), по-видимому, аналогичны тем, которые обнаруживаются при спорадическом РПЖ [187]. Возможными маркерами являются мутации в генах *GNAS*, *KRAS* и *TP53* [380]. В проспективном исследовании родственников с наследственным РПЖ и контрольной группы мутантная ДНК *TP53* присутствовала в панкреатическом секрете у 29 из 43 пациентов с протоковой аденокарциномой поджелудочной железы и у 4 из 8 пациентов с интраэпителиальной неоплазией высокой степени [187].

У подавляющего большинства пациентов РПЖ диагностируется на поздних стадиях при наличии отдаленных метастазов. Даже если диагноз поставлен в то время, когда он потенциально резектабелен, прогноз инвазивного РПЖ неудовлетворительный. Учитывая эти факты, необходим точный и чувствительный тест для диагностики потенциально излечимой ранней неинвазивной неоплазии поджелудочной железы. Некоторые методы позволяют обнаружить редкие мутантные гены, даже если они смешаны с тысячами нормальных копий гена. Приобретенные генетические мутации, связанные с определенными видами рака, могут стать новыми мишенями для разработки чувствительных скрининговых тестов на данное заболевание. Генетические изменения являются привлекательными мишенями для таких тестов раннего выявления, потому что многие из этих изменений специфичны для злокачественных опухолей. Например, мутации гена *KRAS* могут быть обнаружены в крови больных РПЖ [279]. Еще большую чувствительность можно получить, если эти тесты на циркулирующую опухолевую ДНК объединить с тестами на белки и раковые антигены, экспрессируемые опухолевыми клетками [87]. Тем не менее, прежде чем какой-либо молекулярный диагностический тест можно будет использовать в повседневной клинической практике, требуется проведение дальнейших исследований.

1.5.3. Подходы к терапии рака поджелудочной железы

Несмотря на то, что химиотерапия является основной стратегией лечения метастатического РПЖ, все применяемые режимы не смогли существенно улучшить выживаемость пациентов [127, 399]. Недостаточный прогресс в разработке новых эффективных препаратов, вероятно, связан с непониманием геномных и молекулярных механизмов химиорезистентности и прогрессии опухолей [205]. Несмотря на то, что таргетная терапия РПЖ находится на начальном этапе развития, все больше данных свидетельствует о том, что индивидуализированные методы лечения, основанные на конкретных генетических изменениях РПЖ у отдельного пациента, в ближайшем будущем войдут в повседневную клиническую практику. Один из подходов к персонализации лечения РПЖ заключается в тестировании на наличие патогенных герминальных вариантов в генах наследственного РПЖ, например, *BRCA2*, *BRCA1* и *PALB2* и на показатели MSI.

РПЖ, который возникает при наличии наследственной мутации в генах *BRCA1*, *BRCA2*, *PALB2*, а также опухолевой (соматической) мутации в этих генах, связан с дефектным ответом на повреждение ДНК, что увеличивает чувствительность к повреждающим ДНК агентам, таким как производные платины, и к препаратам, нацеленным на сигнальные пути ответа на повреждение ДНК, включая ингибиторы PARP [272]. Большинство пациентов, у которых присутствовали герминальные или соматические мутации в генах *BRCA* или *PALB2*, или другие изменения в генах, связанных с репарацией дефицита гомологичной рекомбинации (например, в генах *ATM*, *BAP1*, *BARD1*, *BLM*, *BRIP1*, *CHEK2*, *FAM175A*, *FANCA*, *FANCC*, *NBN*, *RAD50*, *RAD51*, *RAD51C* и *RTEL1*) могут получать платиносодержащий режим для терапии первой линии (например, гемцитабин плюс цисплатин, лейковорин плюс оксалиплатин и краткосрочная инфузия фторурацила (FOLFOX) или FOLFIRINOX). Для пациентов с герминальными мутациями в генах *BRCA* или *PALB2* без прогрессирования заболевания после не менее 16 недель химиотерапии на основе платины может быть начата поддерживающая терапия PARP ингибитором олапарибом. Во время прогрессирования заболевания варианты включают повторное лечение исходной схемой на основе платины (если пациент получал PARP ингибитор) или введение схемы, содержащей платину, но отличной от той, которая была назначена в первой линии (например, FOLFOX для пациентов, которые ранее получали гемцитабин плюс цисплатин, или наоборот). Для пациентов, которые не получали начальную химиотерапию на основе платины, рекомендовано лечение схемой химиотерапии на основе платины для терапии второй линии (например, гемцитабин плюс цисплатин, FOLFOX, FOLFIRINOX или модифицированный FOLFIRINOX).

Имеющиеся данные по эффективности применения PARP ингибиторов в настоящий момент ограничены [25, 60, 362]. В предварительном отчете клинического исследования TAPUR (Targeted Agent and Profiling Utilization Registry, таргетные агенты и реестр профилирования) из 30 больных с распространенным РПЖ, инактивирующими герминальными или соматическими мутациями в генах *BRCA1/2* и отсутствием стандартных вариантов лечения (53 % получали 3 или более предшествующих линий терапии), получавших монотерапию олапарибом (300 или 400 мг в день), был выявлен объективный ответ у 1 больного и еще у 7 больных - стабилизация процесса (уровень контроля заболевания 31 %, уровень объективного ответа 4 %). Медианная ВБП составила 8,1 недели, годовая ОБ - 47 % [26]. Однако лечение PARP ингибитором у этой категории больных пока не одобрено FDA и еще не считается стандартным подходом после неудачной исходной цитотоксической химиотерапии. В настоящее время проводится несколько исследований, посвященных изучению этой стратегии.

В текущих исследованиях “Молекулярный анализ для выбора терапии” Национального института рака (англ. Molecular Analysis for Therapy Choice - MATCH) и “Таргетные препараты и реестр применения профилирования” ASCO (англ. Targeted Agent and Profiling Utilization Registry - TAPUR) осуществляется высокопроизводительное секвенирование с использованием панели генов для идентификации генетических вариантов в опухолях больных рефрактерным раком. Результаты секвенирования помогают подобрать лечебные стратегии с учетом выявленных молекулярных характеристик, находящиеся на этапе клинических исследований, либо одобренные для лечения других типов ЗНО (офф-лейбл). Два таких генетических панельных теста: MSK-IMPACT (Комплексное профилирование мутаций эффективных мишеней рака Мемориального центра им. Слоуна-Кеттеринга) и FoundationOne CDx (F1CDx) одобрены FDA в США.

В условиях распространенного РПЖ герминальное или соматическое тестирование может выявить потенциально значимые для подбора лечения геномные варианты, для которых доступно клиническое исследование эффективности таргетной терапии, одобренной для другого типа рака. Подсчитано, что примерно в 25 % случаев РПЖ наблюдаются потенциально значимые для подбора лечения молекулярные изменения, и внедрение прецизионной медицины с использованием молекулярно-направленных методов лечения, нацеленные на онкогенные факторы, а также ответ на повреждение ДНК и пути репарации, может оказать существенное влияние на выживаемость пациентов с прогрессирующим РПЖ [45, 281].

Иммунотерапевтический подход к лечению злокачественных опухолей основан на предположении, что иммунная система играет ключевую роль в надзоре и ликвидации злокачественной опухоли, и что опухоли развиваются путем ускользания от иммунной системы. Некоторые виды рака с dMMR и MSI-H особенно чувствительны к иммунной терапии. Опухоли,

имеющие высокий уровень ТМВ, считаются более иммуногенными и чувствительными к иммунотерапии ингибитором иммунных контрольных точек. Объективные, в некоторых случаях полные и стойкие ответы на ингибиторы иммунных контрольных точек, нацеленные на рецептор PD-1, например, пембролизумаб, были выявлены у различных пациентов со статусом опухоли MSI-H, включая РПЖ [209, 210, 229]. Например, в клиническом исследовании KEYNOTE-158 фазы II, в которое вошли 22 пациента с диагностированным РПЖ, наблюдалось 4 объективных ответа (18 %), один из которых был полным; средняя продолжительность ответа составила 13,4 месяца [229].

В США пембролизумаб в настоящее время одобрен для лечения различных распространенных солидных опухолей, включая РПЖ с MSI-H или dMMR, которые прогрессировали после предшествующего лечения, и для которых нет удовлетворительных альтернативных вариантов лечения. Обновленные рекомендации ASCO на 2020 год рекомендуют использовать пембролизумаб в качестве терапии второй линии для пациентов с опухолями MSI-H/dMMR [87]. Однако лишь у ≤ 2 % метастатических форм РПЖ наблюдается dMMR [210, 236]. У других пациентов с распространенным РПЖ преимущества иммунотерапии не установлены, и такое лечение остается исследовательским [258]. Важным моментом является то, что MSI-H или dMMR могут указывать на наличие синдрома Линча, наследственного состояния, которое предрасполагает к нескольким видам рака, включая РПЖ. Всем пациентам с РПЖ и MSI-H/dMMR целесообразно пройти герминальное генетическое тестирование на предмет синдрома Линча, независимо от возраста или семейного анамнеза [208].

В то время как подавляющее большинство опухолей имеют высокую ТМВ, не все из них характеризуются dMMR. Появляется все больше доказательств преимущества ингибиторов иммунных контрольных точек в этой популяции, и пембролизумаб теперь также одобрен FDA для лечения взрослых и детей с неоперабельными или метастатическими солидными опухолями с высокой ТМВ (≥ 10 мут/Мб). Но очень немногие виды РПЖ характеризуются высокой степенью репарации неспаренных оснований (*англ.* proficient mismatch repair - pMMR)/ТМВ, и не установлено, реагируют ли они на иммунотерапию ингибиторами иммунных контрольных точек [264]. Ингибиторы тропомиозин-рецепторной киназы (ТРК) ларотректиниб и энтректиниб одобрены для ТРК-позитивных злокачественных опухолей, независимо от гистологического подтипа или локализации. Хотя такие опухоли редко обнаруживаются среди протоковых аденокарцином поджелудочной железы (в 0,34 % и 0,56 % случаев в двух отдельных сериях [260, 328], в таких случаях сообщается об эффекте ингибирования ТРК [106, 258]. Обновленные рекомендации ASCO на 2020 г. предусматривают использование ларотректиниба или энтректиниба в качестве терапии второй линии для пациентов с ТРК-позитивными опухолями [87].

ГЛАВА 2 - Материалы и методы исследования

2.1. Мета-анализ клинических исследований персонализированного лечения онкологических заболеваний в сравнении с неперсонализированным лечением

2.1.1. Стратегия поиска и отбора исследований

Мета-анализ и систематический обзор, включая поиск данных, отбор исследований, извлечение данных и категоризацию с дальнейшим статистическим анализом, были выполнены в соответствии с рекомендациями PRISMA (англ. Preferred reporting items for systematic reviews and meta-analyses, Предпочтительные элементы отчетности для систематических обзоров и мета-анализов) [268, 299]. Протокол исследования был зарегистрирован в Международном проспективном регистре систематических обзоров PROSPERO (идентификатор записи CRD42024504021).

Поиск публикаций производился в базе данных PubMed (<https://PubMed.ncbi.nlm.nih.gov/>) в период с 20.11.2023 по 20.12.2023. Поиск основывался на терминах словаря MeSH, названии статьи и типе публикации. Был сформирован следующий поисковый запрос: (("Clinical Trials, Phase II as Topic"[Mesh] AND cancer*[tiab])) NOT (review*[pt] OR Meta-analysis[pt] OR Systematic review[pt]), даты публикации - с 2010 по 2021 год. В качестве дополнительных фильтров были выбраны исследования людей.

Критериями включения были: 1. Клиническое исследование 2-й фазы; 2. Применение химиотерапевтических, иммунотерапевтических и таргетных препаратов для лечения онкологических заболеваний; 3. Возраст пациентов старше 18 и меньше 85 лет; 4. Наличие полного текста опубликованной статьи на английском языке. Критериями исключения были: 1. Проведение паллиативной терапии, лечение онкологических заболеваний у детей, гормональной терапии, вакцинотерапии; 2. Недостаточность статистических данных, необходимых для анализа и оценки качества исследования, в том числе - отсутствие оптимальных конечных точек, таких как частота объективного ответа (ЧОО), ВВП или ОВ; 3. Возраст до 18 или более 85 лет; 4. Лечение сопутствующих заболеваний у онкологических больных; 5. Наличие персонализированного и неперсонализированного лечения в одной группе исследования; 6. Отсутствие какого-либо специфического лекарственного лечения онкологического заболевания; 7. Отсутствие подробной информации о параметрах включения пациентов; 8. Публикации следующих типов: обзор литературы, клинический случай, письмо в редакцию, оригинальная статья, не относящаяся ко 2-й фазе исследования. На первом этапе отбора для идентификации

публикаций, не соответствующих критериям включения, оценивали название и реферат работ, выявленных с применением поискового запроса. На втором этапе отбора оценивали полные тексты работ. В ходе отбора отдельно оценивали каждую группу клинических исследований.

Поиск публикаций производили в соответствии с рекомендациями для проведения систематических обзоров и мета-анализов [65, 299, 333]. Отбор осуществляли два исследователя независимо друг от друга с применением критериев включения и исключения. Затем третий исследователь проверял результаты отбора. До завершения отбора исследователи не знали о решениях друг друга. Все разногласия здесь и далее разрешались в аргументированной дискуссии при модерации автора настоящей диссертационной работы.

2.1.2. Извлечение и категоризация данных

На настоящем этапе извлекали и фиксировали такие данные клинических исследований как количество пациентов, медиана возраста, ранее проведенное лечение, дизайн исследования (рандомизация, одноцентровое/многоцентровое), тип терапии (химиотерапия/иммунотерапия/таргетная терапия), импакт-фактор журнала и показатели конечных точек такие как ЧОО, ОВ и ВВП (медианы выживаемости, доли выживших больных за время наблюдения при недостижении медианных значений и время наблюдения). Результаты фиксировали в таблице Microsoft Excel. Три исследователя извлекали данные, а четвертый исследователь проверял полученные результаты.

Персонализацию оценивали на основе данных об изучении геномных характеристик больных и дизайне исследования. Исследования, в которые больные были включены с учетом определенных геномных характеристик, мы считали персонализированными, иные исследования считали неперсонализированными.

Недостающие данные в опубликованных работах запрашивали у авторов. Если недостающие данные не были получены, включали только имеющиеся данные. При отсутствии данных, важных для оценки качества исследования, исследование исключалось.

После независимой оценки отбирали исследования, в которых применялся персонализированный подход к лечению онкологических заболеваний, и исследования с неперсонализированным подходом к лечению для группы сравнения. Если в клиническом исследовании присутствовало как персонализированное лечение, так и неперсонализированное лечение, то в последующем мета-анализе указанные подходы оценивались в разных группах [299, 333].

2.1.3. Синтез данных и статистическая обработка

Был осуществлён синтез всех параметров исследований, включая такие, как количество пациентов, предшествующее лечение, дизайн исследования, тип лечения (химиотерапия/иммунотерапия/таргетная терапия), импакт-фактор журнала, ЧОО, а также 1-летняя и медианная ОВ и ВБП. Однолетнюю выживаемость рассчитывали на основе медианы выживаемости и времени наблюдения или процента выживших пациентов (при недостижении медианы выживаемости) за время наблюдения и времени наблюдения.

Для осуществления пропорционального мета-анализа использовали программы RStudio, применяя язык программирования R и программные пакеты «meta», «metafor» и «tidyverse» [49]. Относительные риски рассчитывали, используя алгоритм Тер-Симонян-Лэрд [100]. Гетерогенность оценивали на основе τ^2 -статистики [100, 396], и I^2 статистики Хиггинса-Томпсона [160, 309]. Доверительные интервалы τ^2 рассчитывали методом Джексона [181]. Эффекты сравнили с помощью точного критерия Фишера и использовали t-критерий или критерий Манна-Уитни в зависимости от типа распределения, $p < 0,05$. Нормальность распределения проверяли с помощью критерия Шапиро-Уилка, $\alpha = 0,05$. Проводили оценку гетерогенности; было принято решение об использовании модели случайных эффектов [151]. Использовали мета-регрессию для оценки влияния применения персонализированного подхода и других параметров при $p < 0,05$, 95% доверительного интервала, значения p-value были оценены с помощью метода Кнаппа-Хартунга [47, 159]. Каждая подгруппа включала не менее 20 исследований.

2.2. Экзомное секвенирование

2.2.1. Разработка адаптированного протокола пробоподготовки и секвенирования

Разработку адаптированного протокола пробоподготовки и секвенирования осуществляли для обеспечения высокого качества пробоподготовки, а также адаптации реагентов, предназначенных для работы на платформе Illumina, к работе на высокопроизводительном NGS секвенаторе MGISEQ-G400 (MGI, Китай). Адаптированный протокол предназначался для дальнейшего проведения молекулярно-генетического анализа образцов биологического материала больных РПЖ и ТНРМЖ.

Разработка адаптированного протокола пробоподготовки и секвенирования включала 2 блока задач: 1 - выбор наиболее оптимальных коммерчески доступных наборов для выделения ДНК из периферической крови и опухолевой ткани, заключенной в парафин; 2 - модификация

существующего протокола обогащения экзомных библиотек с использованием реактивов Roche (США) и Agilent (США) для его адаптации к секвенатору MGISEQ-G400 (MGI, Китай).

Было произведено сравнение различных коммерчески доступных наборов для выделения ДНК с целью выбора набора, в наибольшей степени отвечающего целям настоящей работы. При выделении ДНК из периферической крови были опробованы такие наборы как QIAamp DNA Blood Mini Kit (QIAGEN, США), MagPure Blood DNA kit (Magen Biotech, Китай) и “Проба-ГС-генетика” (“ДНК-технология”, Россия).

При выделении ДНК из опухолевой ткани исследовали такие наборы как QIAamp DNA FFPE Tissue kit (QIAGEN, США), AVENIO Tumor DNA Isolation Kit (Roche, США) и Cobas DNA Sample Preparation Kit (Roche, США). Анализировали такие параметры, как количество выделенной ДНК из заданного объёма, а также её чистота (определялась на спектрофотометре Nanodrop 2000 путем определения показателей на длинах волн 260/280 и 260/230). Количество выделенной ДНК определяли путём измерения концентрации в 1 мкл и умножения на объём элюции; концентрацию выделенной ДНК измеряли с использованием флуориметра Qubit 2.0 (Life Technologies, США). Приготовление библиотек из выделенной ДНК осуществляли по стандартному протоколу производителя без модификаций с использованием MGIEasy Universal DNA Library Prep Set (MGI Tech, Китай).

Для реализации этапа обогащения экзомных библиотек производили модифицирование протокола обогащения зондами Roche (США) или зондами Agilent (США), предназначенными для работы на платформе Illumina, с целью их адаптации к секвенированию на платформе MGI (Китай). Было произведено модифицирование (полная замена) блокирующих олигонуклеотидов для обеспечения работы зондов Roche и Agilent на библиотеках MGI. В рамках модифицирования протокола обогащения библиотек были разработаны и синтезированы олигонуклеотиды Block 1 и Block 2 с LNA-модификацией (добавлением замкнутых нуклеотидов).

Циркуляризацию осуществляли по стандартному протоколу производителя без модификаций с использованием MGIEasy Circularization Kit (MGI, Китай). Секвенирование осуществляли на секвенаторе MGISEQ-G400 (MGI, Китай) по стандартному протоколу производителя без модификаций.

2.2.2. Методология сравнения адаптированного протокола пробоподготовки и секвенирования с референсным протоколом

Для оценки разработанного в рамках настоящего исследования протокола пробоподготовки и секвенирования осуществляли его сравнение с широко применяемой линией наборов AVENIO (Roche, США) [392].

Для проведения сравнительного анализа образцы биологического материала 30 больных ТНРМЖ исследовали с применением разработанных протоколов пробоподготовки и секвенирования, далее эти же 30 образцов ТНРМЖ исследовали с применением референсных протоколов AVENIO (Roche, США).

Пробоподготовку и секвенирование с использованием разработанного протокола осуществляли как указано в разделе 3.2.1 “Описание адаптированного протокола пробоподготовки и секвенирования”. Пробоподготовку и секвенирование с применением наборов AVENIO (Roche, США) проводили в соответствии с инструкцией производителя [392]. Выделение ДНК из опухолевой ткани, залитой парафином, осуществляли с использованием наборов AVENIO Tumor Isolation and QC Kit (Roche, США). Для контроля качества выделенной ДНК использовали ПЦР, концентрацию выделенной ДНК измеряли на флуориметре Qubit 2.0 (Life Technologies, США).

Библиотеки из выделенной ДНК получали с использованием набора AVENIO Tumor Library Prep Kit (Roche, США). Баркодирование библиотек осуществляли при помощи ПЦР с использованием наборов AVENIO Tumor Sample PrimersKit (Roche, США). Для обогащения библиотек использовали наборы AVENIO Tumor Enrichment Kit (Roche, США) с панелью AVENIO Tumor Tissue Expanded Panel Kit (Roche, США), в состав которой входят 77 онкоассоциированных генов. Концентрацию ДНК в изготовленных библиотеках измеряли с использованием флуориметра Qubit 2.0 (Life Technologies, США), контроль качества осуществляли с применением биоанализатора Agilent Bioanalyzer 2100 (Agilent Technologies, США). Секвенирование проводили с использованием платформы NextSeq500 (Illumina, США).

Выявление отличий между анализируемыми протоколами проводили на различных этапах по следующим качественным и количественным характеристикам: 1. Способ экстракции ДНК, выход ДНК; 2. Длительность лигирования; 3. Способ обогащения библиотек; 4. Количество исследуемых генов; 5. Особенности секвенирования. Сравнение количественных характеристик между разработанным протоколом пробоподготовки и секвенирования и референсным протоколом AVENIO (Roche, США) осуществляли как описано далее:

1. Выход ДНК определяли путём измерения концентрации в 1 мкл и умножения на объём элюции. Концентрацию выделенной ДНК измеряли с использованием флуориметра Qubit 2.0 (Life Technologies, США);
2. Длительность лигирования сравнивали, исходя из инструкций производителя к тестируемым наборам;
3. Количество исследуемых генов определялось выбранной панелью для таргетного обогащения.

2.3. Разработка фармакогеномной платформы и системы поддержки принятия врачебных решений для подбора персонализированного лекарственного лечения онкологических заболеваний

2.3.1. Архитектура фармакогеномной платформы и системы поддержки принятия врачебных решений

Для осуществления функции подбора эффективного персонализированного лекарственного лечения онкологических заболеваний СППВР должна функционировать на основе фармакогеномной платформы, способной поддерживать постоянный доступ к необходимым данным, логическую интеграцию данных и надлежащее курирование данных, а также располагать новейшими инструментами обработки и структурирования данных. При этом, архитектура фармакогеномной платформы должна обеспечивать решение следующих основных задач:

- обеспечивать возможность интеграции и хранения больших объемов геномных, справочных, клинических и иных релевантных данных;
- предоставлять СППВР возможность анализировать накопленные данные;
- осуществлять актуализацию накапливаемой информации ввиду высокой скорости появления новых молекулярно-генетических данных в области онкологии;
- использовать алгоритмы машинного обучения для анализа данных;
- поддерживать пользовательские модули, которые медицинский персонал сможет использовать в своих ежедневных рабочих процессах для принятия более эффективных решений.

Опираясь на данную концепцию, в фармакогеномную платформу, разрабатываемую в настоящем исследовании, были включены следующие структурные элементы: озеро данных, биоинформатические конвейеры для определения герминального и соматического геномного профиля (далее - герминальный биоинформатический конвейер, соматический биоинформатический конвейер), инструменты для организации вычислений и программный интерфейс приложения (англ. application programming interface, API) доступа к данным [123]. Структура фармакогеномной платформы и взаимосвязь с СППВР отражены на рисунке 3.



Рисунок 3. Структура комплекса “фармакогеномная платформа - система поддержки принятия врачебных решений”, разрабатываемого в настоящем исследовании. API (англ. application programming interface) - программный интерфейс приложения, MSI (англ. microsatellite instability) - показатели микросателлитной нестабильности, TMB (англ. tumor mutational burden) - опухолевая мутационная нагрузка

В рамках фармакогеномной платформы разрабатывались следующие элементы:

1. Герминальный и соматический биоинформатические конвейеры для обработки данных высокопроизводительного секвенирования и выявления геномных вариантов, показателей вариации числа копий (англ. copy number variation, CNV), TMB, MSI, функционирующие с применением инструментов на основе технологий машинного обучения;
2. Озеро данных, способное аккумулировать большие объемы различных геномных, клинико-фенотипических, справочных и иных сведений;
3. Инструменты для организации вычислений и API. В качестве инструментов для организации вычислений использовали программы Apache Airflow 2.9.1, Kubernetes 1.28 и docker-контейнер. Apache Airflow 2.9.1 применяли для организации очередей задач в биоинформатических конвейерах и распределения нагрузки. Kubernetes 1.28 использовали для организации управления контейнеризованными рабочими нагрузками. Под API подразумевали стандартизованный интерфейс для взаимодействия с внешними источниками данных (базы данных, инструменты и прочее).

Согласно разработанной архитектуре, СППВР является отдельным структурным элементом, не входящим в состав фармакогеномной платформы. Вместе с тем, СППВР опирается на данные, аккумулированные в фармакогеномной платформе, для подбора

персонализированного лекарственного лечения ЗНО с учётом геномных характеристик больного. Структурно СППВР состоит из модуля подбора персонализированного лекарственного лечения на основе геномных вариантов и ТМВ, модуля определения мутационного профиля опухоли и модуля визуализации, обеспечивающего формирование отчетов о подборе персонализированного лекарственного лечения онкологических заболеваний.

2.3.2. Методология создания озера данных

2.3.2.1. Выбор решений для управления озером данных

Для создания фармакогеномной платформы было необходимо произвести интеграцию большого объема различных структурированных и неструктурированных данных, включая клинико-фенотипическую информацию о больных, необработанные данные секвенирования, выявленные геномные характеристики, информацию из мировых открытых баз данных и т.д. Ввиду наличия большого объема информации, было принято решение о её размещении в озере данных.

Для выбора инструмента управления озером данных применялся метод двухэтапного анализа различных инструментов. Были проанализированы следующие инструменты управления озером данных: Dataiku, Dremio, Knime, Delta Lake, а также стек технологий (Apache NIFI + Apache Kafka + Apache Airflow).

На первом этапе анализа производили тестовое использование указанных инструментов управления озером данных для выявления их возможностей, оценки пользовательского интерфейса и гибкости платформы. Также по итогам первого этапа анализа инструментов были определены требования, которым должен удовлетворять инструмент управления озером данных, на базе которого оно будет создаваться.

На втором этапе проводили анализ соответствия инструментов управления озером данных Dataiku, Dremio, Knime, Delta Lake, а также стека технологий (Apache NIFI + Apache Kafka + Apache Airflow) требованиям, определенным в ходе первого этапа анализа. Исходя из задачи интеграции различных данных, были определены следующие требования, которым должен удовлетворять инструмент управления озером данных, который будет использоваться в настоящем исследовании:

1. Возможность организовать систему на кластере машин;
2. Возможность динамического добавления новых источников данных;
3. Наличие открытого исходного кода;
4. Поддержка процессов извлечения, трансформации и загрузки данных;

5. Наличие интуитивно понятного, удобного графического интерфейса;
6. Интерфейс с функцией перетаскивания элементов (drag-and-drop);
7. Совместимость с реляционными базами данных;
8. Совместимость с NoSQL базами данных;
9. Наличие встроенной системы управления процессами (оркестратор);
10. Простота реализации;
11. Поддержка различных языков программирования;
12. Интеграция с LDAP (англ. Lightweight Directory Access Protocol, облегчённый протокол доступа к каталогам).

2.3.2.2. Методология наполнения озера данных

Озеро данных является одним из структурных элементов фармакогеномной платформы, необходимым для аккумуляции и хранения больших объемов разнородных (структурированных и неструктурированных) данных. Для построения озера данных в настоящей работе использовали следующий стек технологий: хранилище неструктурированных данных под архитектурой Amazon S3, PostgreSQL для хранения структурированных данных, платформа Dataiku для проведения ETL процессов (англ. extract, transform, load - процессы извлечения, преобразования и загрузки), трансформации данных и управления перемещением данных.

В создаваемое озеро данных загружали следующие типы собственных данных:

1. Регистрационные и клинико-фенотипические данные больных;
2. Лабораторные метрики пробоподготовки и секвенирования биологического материала больных;
3. Нуклеотидная последовательность ДНК больных (необработанные данные секвенирования в формате FASTQ и их метрики качества);
4. Герминальные и соматические геномные варианты, показатели TMB, MSI и CNV, выявленные по итогам работы биоинформатических конвейеров;
5. Метаданные, метрики качества работы герминального и соматического биоинформатических конвейеров;
6. Результаты подбора персонализированного лекарственного лечения онкологических заболеваний, сформированные СППВР.

Также в озеро данных интегрировали опубликованную информацию из открытых мировых баз данных, необходимую для аннотирования геномных вариантов и аккумуляции сведений об ассоциациях биомаркеров с чувствительностью к противоопухолевым лекарственным препаратам. Аккумуляция в озеро данных сведений об ассоциациях биомаркеров с чувствительностью к противоопухолевым препаратам в дальнейшем использовались СППВР для подбора персонализированного лекарственного лечения ЗНО.

2.3.2.2.1 Регистрационные и клинико-фенотипические данные больных

На первом этапе в озеро данных добавляли регистрационную и клинико-фенотипическую информацию о больных с целью формирования свода о характеристиках выявленных злокачественных новообразований, применяемых методах лечения, возникающих осложнениях и исходах лечения. К регистрационной информации относили возраст, пол, идентификационный номер (ID) больного, номер истории болезни и т.д. Фиксировали также клинико-фенотипическую информацию, как общую для больных с различными типами ЗНО (данные о стадии заболевания согласно TNM-классификации, локализации опухоли, данные гистологического и ИГХ исследований, сведения о проводимом лечении и т.д.), так и специфичную для конкретных типов ЗНО. Так, о больных ТНРМЖ помимо общей клинико-фенотипической информации собирали сведения о менструальном статусе, эндопротезировании, экспандере, симметризирующей операции, профилактической мастэктомии и т.д. Перечень клинико-фенотипических сведений о больных ТНРМЖ приведен в Приложении 1.

Регистрационные и клинико-фенотипические данные больных попадали в озеро при путем синхронизации с медицинской информационной системой ФГБУ «НМИЦ радиологии» Минздрава России Асклепиус, осуществляемой по протоколу REST API.

2.3.2.2.2. Лабораторные метрики пробоподготовки и секвенирования биологического материала больных

В озеро данных также загружали лабораторные метрики пробоподготовки и секвенирования биологического материала больных (Таблица 1).

Таблица 1. Лабораторные метрики пробоподготовки и секвенирования, загружаемые в озеро данных

№	Параметр	№	Параметр
1	№ Образца	27	Адаптер
2	Филиал, направляющий материал	28	Количество циклов ПЦР
3	Дата поступления	29	Концентрация библиотеки, нг/мкл
4	Тип материала	30	Длина фрагментов в библиотеке
5	Чей пациент?	31	Объем образца, взятый в пул, мкл
6	Номер истории болезни	32	Номер пула в запуске
7	Patient ID	33	Длина фрагментов в пуле
8	Доля опухоли в образце	34	Концентрация обогащенного пула, нг/мкл
9	Статус	35	Объем пула, взятый в циркуляризацию, мкл
10	Пол	36	Объем LOW-TE буфера, для добавления циркуляризации, мкл
11	Комментарий	37	Кит при обогащении
12	Диагноз	38	Концентрация пула после циркуляризации, нг/мкл
13	Консультация	39	Концентрация пула, после циркуляризации фемтомоль/мкл
14	Проект	40	Объем пула взятый в DNB, мкл
15	Направляющий врач	41	Объем LOW-TE буфера, для добавления при подготовке DNB, мкл
16	Исследование	42	Концентрация DNB, нг/мкл
17	ID Образца	43	Номер запуска секвенатора MGI
18	Нуклеиновая кислота	44	Исполнитель анализа
19	Что делается технически	45	Исполнитель биоинформатической обработки
20	Чем выделена НК	46	Крайний срок получения результата
21	Концентрация НК, нг/мкл,	47	Дата ответа
22	Взято в пробу НК мкл, макс=45 мкл	48	Время выполнение анализа (факт)
23	AE Буфер в пробу	49	Результат
24	Тип фрагментации	50	Комментарий
25	Тип очистки после фрагментации	51	Номер блока
26	Набор для приготовления библиотеки		

Аббревиатуры: НК - нуклеиновая кислота, ПЦР - полимеразная цепная реакция, DNB (англ. DNA nanoball) - ДНК наносферы

2.3.2.2.3. Нуклеотидная последовательность ДНК больных (необработанные данные секвенирования в формате FASTQ и их метрики качества)

В озеро данных также добавляли информацию о нуклеотидной последовательности ДНК больных, представляющую собой необработанные данные высокопроизводительного секвенирования в формате FASTQ, при этом загрузку осуществляли в хранилище Amazon S3. Наполнение хранилища осуществляли при помощи специально созданных скриптов для автоматической загрузки и проверки данных из различных источников. Скрипты для автоматической загрузки и проверки данных (сырых “ридов” - прочтений) создавали с использованием инструментов разработки Python, AWS CLI.

Данные попадали в хранилище с удаленных FTP серверов (англ. file transfer protocol, протокол передачи файла) или же напрямую с секвенатора MGISEQ-G400 (MGI, Китай) после автоматической проверки на наличие справочной информации о том, откуда, каким путем, с использованием какого оборудования были получены данные. Проверку осуществляли по следующим параметрам: наличие всех необходимых данных, совпадение размера данных с реальным. После проведения секвенирования на секвенаторе в папке OutputFq создавали папку, содержащую сырые данные сиквенса в формате FASTQ. Название папки при этом соответствовало номеру проточной ячейки, на которой производили запуск секвенатора (Рисунок 4А).

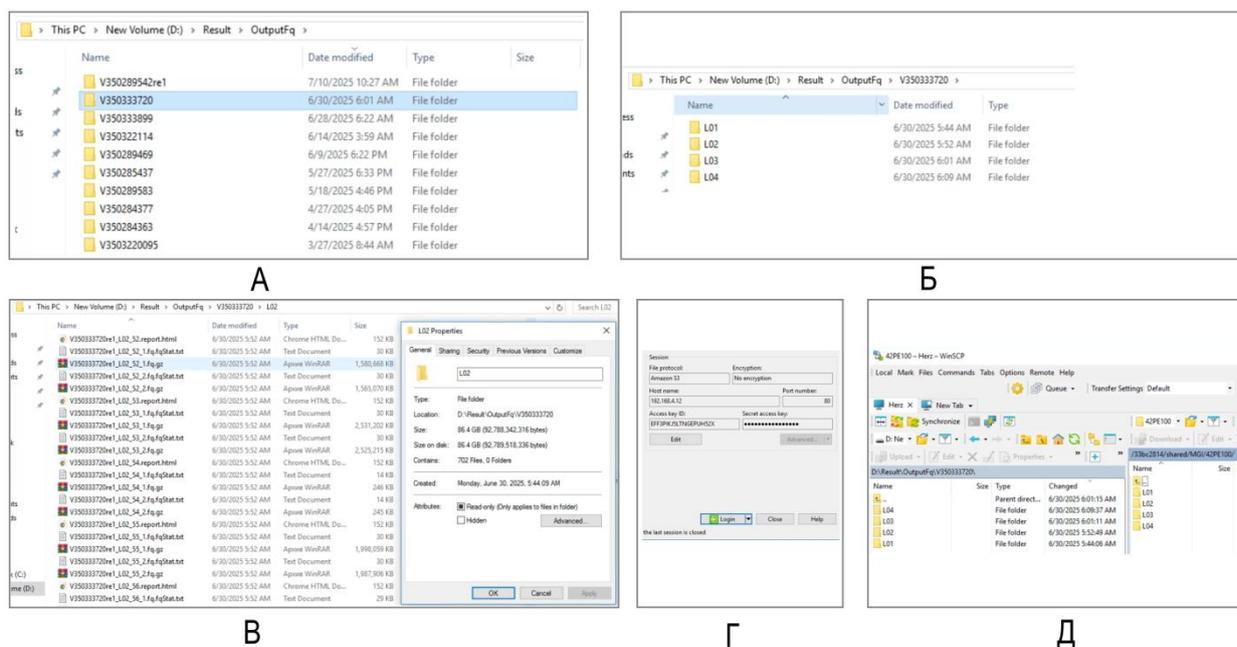


Рисунок 4. А) Процесс присвоения имен в папке секвенатора; Б) Структура папки, содержащей информацию об успешно проведенном запуске секвенатора на проточной ячейке №V350333720, завершившемся 30.06.2025 года; В) Структура папки, содержащей информацию об итоге секвенирования по дорожке номер L02 в запуске секвенатора на проточной ячейке

№V350333720, завершившимся 30.06.2025 г. На рисунке представлены наборы сырых данных секвенирования образцов (формата fq.gz), и сервисная информация о качестве секвенирования. Данные содержат информацию о номере проточной ячейки, номере дорожки, номере адаптера, соответствующего анализируемому образцу, а так же префикса _1 и _2, где _1 - прямые прочтения, а _2 - обратные; Г) Внешний вид окна приложения WinSCP с настройками для прямого подключения к серверу для передачи данных; Д) Внешний вид окна приложения WinSCP в режиме переноса данных с секвенатора на удаленный сервер: в левой части окна открыт запуск секвенатора на проточной ячейке №V350333720, завершившийся 30.06.2025 г., в правой части - он же, но с номером 42PE100 на удаленном сервере после переноса данных

На рисунке 4А представлена папка OutputFq, расположенная на диске D в операционной системе Microsoft Windows 10, установленной на секвенаторе MGI-G400. Папка содержит набор папок, каждая из которых соответствует успешно проведенному запуску секвенатора. Выделен запуск секвенатора, успешно завершившийся 30.06.2025 года, произведенный на проточной ячейке № V350333720. Внутри папки с запуском содержатся 4 папки по каждой дорожке с проточной ячейки, на которой производился запуск секвенатора (рисунок 4Б). На данном рисунке представлена структура распределения данных секвенирования после успешного завершения запуска секвенатора MGI-G400. Проточная ячейка данного секвенатора состоит из 4 независимых дорожек, на каждой из которых происходит реакция в процессе работы. Каждая папка на иллюстрации содержит данные, полученные в рамках секвенирования с дорожек №№1, 2, 3 и 4.

Внутри папки с данными с каждой дорожки секвенатора располагались исходные данные секвенирования образцов, пулированных для данной дорожки, а также технические файлы с информацией о качестве сиквенса. Данные имели формат V350333720re1_L02_52_1.fq.gz, где сначала указывался номер проточной ячейки, после - номер дорожки, далее - номер адаптера, после чего - прямое или обратное прочтение (_1 и _2 соответственно) (Рисунок 4В).

Для начала обработки исходные данные выгружали в хранилище Amazon S3 по протоколу S3 connect при помощи приложения WinSCP с настройками, приведенными на Рисунке 4Г. Далее данные загружали в логическую единицу хранения (бакет) на сервере в папку Runs, где осуществлялось хранение сырых необработанных запусков NGS секвенатора для их последующей обработки (Рисунок 4Д).

Далее производили разделение необработанного запуска секвенатора на отдельные образцы, которые анализировались в конкретном запуске секвенатора (парсинг запуска). Для этого использовали загруженный в озеро данных свод лабораторных метрик пробоподготовки и секвенирования биологического материала больных, содержащий, в том числе, идентификационную информацию больных.

Процесс парсинга происходил с помощью запуска кода в экосистеме Dataiku, который производит сопоставление всего содержимого папок L1, L2, L3, L4, по номерам адаптера в

названии соответствующих файлов, с информацией в сводке. Интерфейс запуска данного процесса представлен на Рисунке 5.

Возможно несколько различных режимов парсинга данных из хранилища.

- 1) Парсинг отдельного образца из всех запусков, в которых он анализировался, и погружение итога в общую папку на образец. Для этого необходимо установить наименование образцов через запятую, а в графе “run” установить значение *:

```
{
  "run": "*",
  "sample": "9754EFCR"
}
```

run - номер запуска

sample - id образца из регистрационной таблицы.

- 2) Парсинг отдельного запуска без указания конкретных образцов, которые в нем содержались. Для этого необходимо установить номер запуска, а в графе “sample” установить значение *:

```
{
  "run": "42PE100",
  "sample": "*"
}
```

- 3) Парсинг конкретных образцов, анализирующихся в конкретном запуске секвенатора. Для этого необходимо указать как конкретный “run”, так и конкретный sample id:

```
{
  "run": "42PE100",
  "sample": "5560LXMY,7458VJNX"
}
```

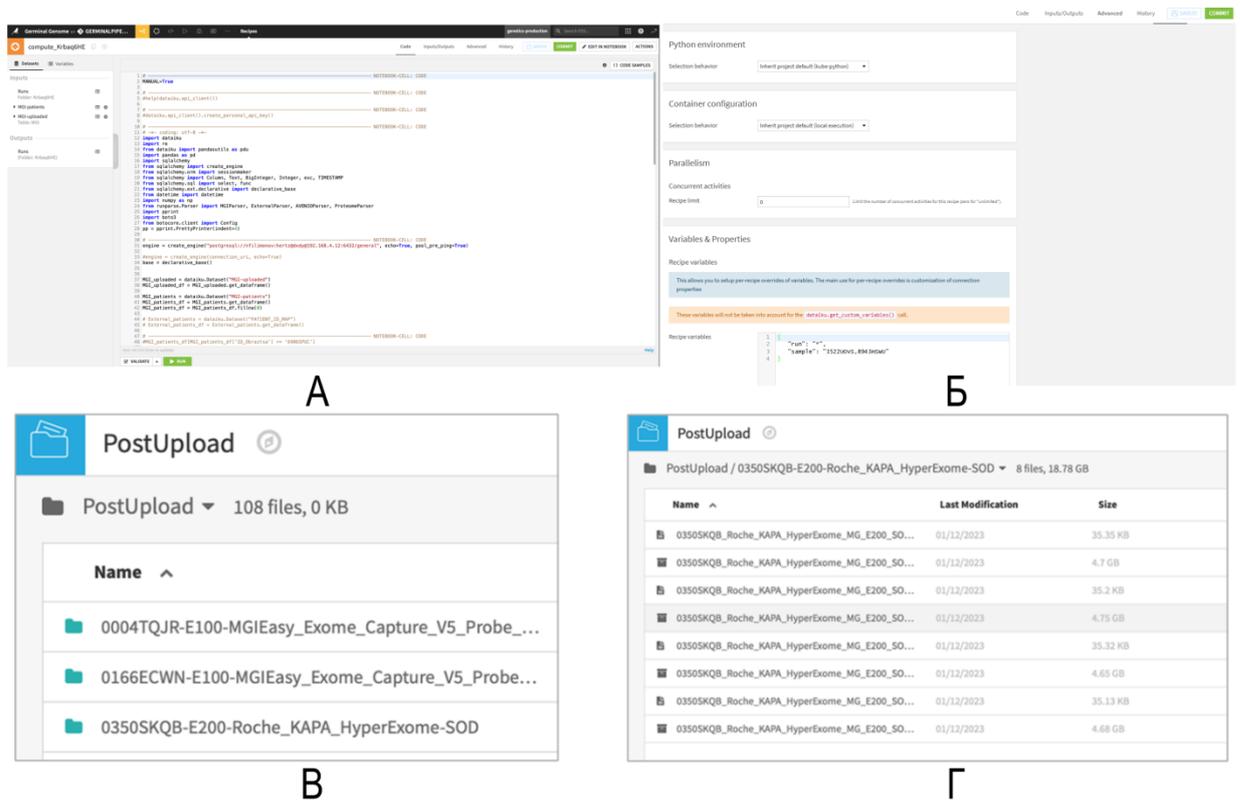


Рисунок 5. Интерфейс платформы Dataiku: А) Внешний вид интерфейса платформы Dataiku, содержащий код для запуска процесса парсинга; Б) Окно с гибкими настройками параметров запуска парсинга; В) Вид репозитория PostUpload; Г) Вид папки, содержащей образец для обсчета

Запуск процесса осуществляли на платформе Dataiku при нажатии кнопки Run (Рисунок 5А), детали настройки задавали при нажатии на меню Advanced, когда можно задать параметры “run” и “sample” (Рисунок 5Б). После завершения парсинга запуска, образцы группировали по папкам таким образом, что в репозитории PostUpload создавалась отдельная папка на каждый образец (Рисунок 5В), который в дальнейшем передавался в обсчет (Рисунок 5Г).

При данной операции происходило переименование исходных данных секвенирования больного из формата V350333720re10212087_L021_52122_12.fq.gz, в котором его предоставляет секвенатор, в формат 0350SKQB_Roche_KAPA_HyperExome_MG_E200_SOD_a77_L01_31PE100_1.fq.gz. Далее, необработанные “риды” (прочтения) приводились к общему наименованию, с использованием таблицы лабораторных метрик пробоподготовки и секвенирования биологического материала больных им присваивались уникальные идентификаторы, содержащие номер образца и метаданные.

Непосредственно по FASTQ файлам производили сбор метрик качества для отсеивания бракованных образцов (Приложение 2).

2.3.2.2.4. Выявленные геномные варианты, показатели микросателлитной нестабильности, опухолевой мутационной нагрузки, вариации числа копий и данные о рекомендациях персонализированного лекарственного лечения

В озеро данных добавляли информацию о герминальных и соматических геномных вариантах, значениях TMB, MSI, CNV, полученную по итогам обработки данных высокопроизводительного секвенирования разработанными биоинформатическими конвейерами. Биоинформатические конвейеры функционируют за пределами озера данных, однако информация о выявляемых геномных вариантах, значениях TMB, MSI и CNV вносится в озеро данных.

Для обеспечения работы биоинформатических конвейеров осуществляли разработку программного модуля, принимающего на вход соматические и/или герминальные FASTQ файлы для дальнейшего запуска микросервисов с биоинформатическими шагами. Для разработки программного модуля использовали языки программирования Python и Bash, и Docker-контейнеры. Для разработки панели запуска биоинформатических конвейеров внутри инструмента управления озером данных Dataiku использовали языки программирования JavaScript, Python, HTML/CSS и фреймворк React.

На микроуровне механизм выявления герминальных и соматических геномных вариантов в исследуемых образцах был реализован для всех типов биоинформатических конвейеров как показано на рисунке 6.

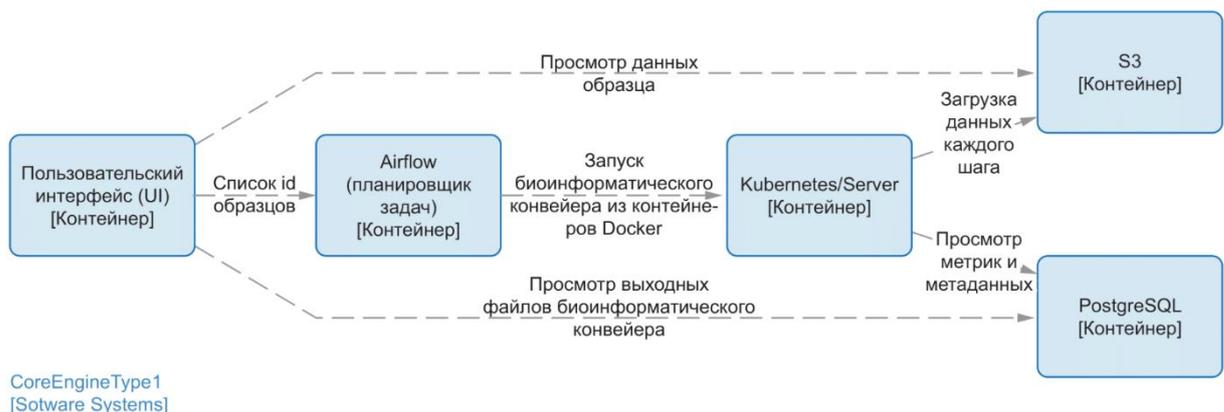


Рисунок 6. Вычислительная архитектура биоинформатических конвейеров. На данном рисунке представлена схема взаимодействия пользовательского интерфейса с данными внутри фармакогеномной платформы. Можно видеть, что через пользовательский интерфейс (UI), команды поступают в Apache AirFlow, который распределяет нагрузку, что приводит к запуску биоинформатического конвейера и последующей обработке данных

На рисунке 6 визуализированы следующие этапы: Этап 1: через панель управления производился выбор образцов, которые необходимо проанализировать. Этап 2: данный запрос направлялся в Apache Airflow, при помощи которого поднимались необходимые docker контейнеры. Apache Airflow - это платформа управления рабочими процессами с открытым исходным кодом для конвейеров данных. Именно там задается последовательность, в которой запускаются докер-контейнеры с биоинформатическими шагами. В настоящей работе Apache Airflow использовали как систему очередей задач для управления биоинформатическими конвейерами (Рисунок 7).

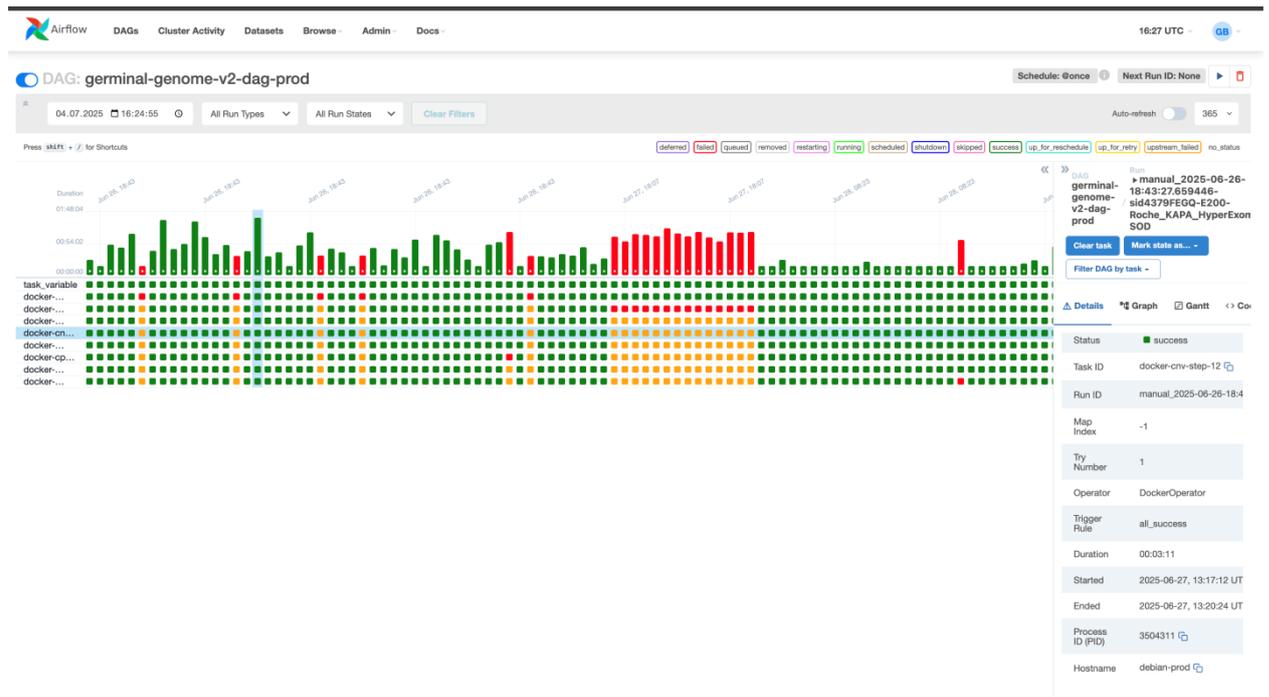


Рисунок 7. Интерфейс Apache Airflow в процессе расчета партии образцов герминальным биоинформатическим конвейером. Отображена очередь образцов, этап обсчета и другая информация.

В процессе обсчета под каждый шаг биоинформатического конвейера создавался свой докер. Для реализации выполнения каждого конкретного шага биоинформатического конвейера собирался докер-контейнер, в котором запускались необходимые команды для осуществления информатических операций с данными. Для этого необходимо запустить команду `docker build -t germinal_pipeline_merge --rm -f ./Dockerfile.step0/ .` Далее (как видно в примерах кода докерфайлов, в каждый образ скачивались необходимые файлы, например BED файлы и другие зависимости). Затем внутри запущенного докер-контейнера, автоматически запускалась необходимая исполняемая команда. Пример докерфайла для создания контейнера для шага QC (англ. quality control, контроль качества) BAM файла приведен на рисунке 8.

```

# syntax=docker/dockerfile:1
FROM python:3.9.17-bullseye
WORKDIR /code
# Install OpenJDK-11
RUN apt-get update && \
    apt-get install -y openjdk-17-jre && \
    apt-get install -y openjdk-17-jdk && \
    apt-get install -y ant && \
    apt-get install -y libpq-dev && \
    apt-get clean;

ENV JAVA_HOME /usr/lib/jvm/java-17-openjdk-amd64/
RUN export JAVA_HOME

RUN wget https://dlcdn.apache.org/spark/spark-3.3.3/spark-3.3.3-bin-hadoop3.tgz
RUN tar xvf spark-3.3.3-bin-hadoop3.tgz && mv spark-3.3.3-bin-hadoop3/ /opt/spark
RUN export SPARK_HOME=/opt/spark && export PATH=$PATH:$SPARK_HOME/bin:$SPARK_HOME/sbin

RUN apt-get -y install locales && sed -i 'en_US.UTF-8/s/^#//g' /etc/locale.gen && locale-gen
# Set the locale
ENV LANG en_US.UTF-8
ENV LANGUAGE en_US:en
ENV LC_ALL en_US.UTF-8

RUN apt install sudo
RUN curl "https://awscli.amazonaws.com/awscli-exe-linux-x86_64.zip" -o "awscliv2.zip" && unzip awscliv2.zip && sudo ./aws/install

COPY .aws /root/.aws
RUN apt-get install -y build-essential
COPY requirements.txt requirements.txt
RUN pip3 install -r requirements.txt
EXPOSE 5000
EXPOSE 6432

RUN mkdir -p /data/
RUN mkdir -p /output/
RUN mkdir -p /BED/
RUN mkdir -p /annotations

RUN aws s3 cp --recursive s3://#####/gatk-4.4.0.0/code/gatk-4.4.0.0 --region ru-central1 --endpoint https://storage.yandexcloud.net
RUN chmod -R 777 /code/gatk-4.4.0.0

ENV GATK /code/gatk-4.4.0.0/gatk
RUN export GATK
COPY ./annotations/KAPA_HyperExome_hg19_primary_targets_short_sort.interval_list
/BED/KAPA_HyperExome_hg19_primary_targets_short_sort.interval_list
COPY ./annotations/MGI_Exome_Capture_V5.hg19_sort.interval_list/BED/MGI_Exome_Capture_V5.hg19_sort.interval_list
COPY ./annotations/SSELXT2_Human_All_Exon_V8_sort.interval_list/BED/SSELXT2_Human_All_Exon_V8_sort.interval_list
COPY ./annotations/SSELXT2_Human_All_Exon_V7_sort.interval_list/BED/SSELXT2_Human_All_Exon_V7_sort.interval_list

RUN export SAMPLEID
RUN export STEP
RUN export MODE

WORKDIR /code
COPY ./somatic-pipe/somatic-bamstat.sh /code
COPY ./somatic-pipe/step4_run.py /code
COPY ./somatic-pipe/Utils.py /code
COPY ./somatic-pipe/ActivityLog.py /code
COPY ./somatic-pipe/Cleaner.py /code
COPY ./somatic-pipe/OutputFiles.py /code
COPY ./somatic-pipe/Parser.py /code
COPY ./somatic-pipe/ParserLog.py /code
COPY ./somatic-pipe/Config.py /code
COPY ./config.ini /code

ENTRYPOINT [ "/bin/bash", "-l", "-c" ]
CMD ["python3 /code/step4_run.py $SAMPLEID $STEP $MODE"]

```

Рисунок 8. Пример докерфайла для создания контейнера для шага QC (англ. quality control, контроль качества) BAM (англ. binary alignment map, карта двоичного выравнивания файла)

На каждом этапе биоинформатического конвейера в Amazon S3 хранилище загружались промежуточные файлы, которые необходимы для воспроизведения результата, для поиска ошибок и аномалий, если таковые обнаружатся. Также это было необходимо для пересчета образца не с нулевого шага при возникновении ошибки.

В озеро данных также вносили информацию о результатах подбора эффективного персонализированного лекарственного лечения ЗНО, находящуюся в отчетах, сформированных СППВР. Кроме того, из сформированных отчетов в озеро данных добавлялась информация об уровне доказательности тех или иных рекомендаций (уровень А - утвержденная рекомендация, уровень В - клиническое доказательство, уровень С - клинический случай, уровень D - доклиническое доказательство, уровень Е - статистический анализ). Также добавлялась информация о регистрации/одобрении рекомендованных препаратов в Российской Федерации, ЕС или США для лечения ЗНО, выявленного у анализируемого больного (он-лейбл), либо для лечения иных ЗНО (офф-лейбл).

2.3.2.2.5. Метаданные, метрики качества работы герминального и соматического биоинформатических конвейеров, общие сводные таблицы

Биоинформатический конвейер состоит из последовательности реализуемых этапов. На каждом шаге биоинформатического конвейера, с использованием языка программирования Python осуществляли сбор метаданных (время начала шага биоинформатического конвейера, время затраченное на шаг, версия биоинформатического конвейера, тип биоинформатического конвейера (соматический/герминальный), статус выполнения, ID образца) и их запись в базу данных PostgreSQL, так как данные являлись структурированными. Неструктурированные метаданные сохраняли в хранилище Amazon S3 в той же папке, где расположены FASTQ файлы.

По FASTQ файлам также получали информацию о том, какой набор реагентов использовался, о типе данных (герминальные, соматические), когда образец был принят в обработку. Все эти данные были зашифрованы в названии образца после того, как он попадал в репозиторий PostUpload. Например, из названия образца 0350SKQB_Roche_KAPA_HyperExome_MG_E200_SOD_a77_L01_31PE100_1.fq.gz можно видеть, что 0350SKQB - это ID образца, Roche_KAPA_HyperExome - набор реагентов для таргетного обогащения, MG - тип секвенатора, E200 - покрытие, SOD - тип данных (в данном случае опухоль), a77 - адаптер, L01 - номер дорожки, 31PE100 - номер запуска, _1 - прямой или обратный “рид” (прочтение).

Осуществляли сбор метрик качества работы биоинформатического конвейера (Приложение 3). Метрики качества считывали с использованием языка программирования Python, например, процент удаленных дубликатов, корректность базовой информации и другие стандарты качества данных, например по BAM или FASTQ файлам, которые также представляли собой структурированный набор данных. Для сбора метрик качества использовали инструмент

PICARD. Результат записывался в PostgreSQL для дальнейшего использования в построении сводной таблицы.

По итогам добавления различных типов информации, указанных в предыдущих пунктах, в рамках озера данных формировали общие сводные таблицы, содержащие весь массив информации, полученной в процессе анализа биологического материала больных. Описанное выше множество данных виртуально комплектовывалось по ID образца с использованием инструментов для проведения ETL процессов, встроенных в инструмент управления озером данных Dataiku, предоставляя возможность для просмотра сводной информации специалистам.

Программирование логики внутри выбранного инструмента управления озером данных, а именно ETL процессы внутри инструмента (процессы слияния таблиц метрик, регистратуры, найденных мутаций) осуществляли с использованием языка программирования Python, а также в no-code режиме с использованием встроенных инструментов, которые предоставляет Dataiku.

2.3.2.2.6. Информация из открытых мировых баз данных

Для осуществления аннотирования выявленных геномных вариантов биоинформатическими конвейерами и дальнейшего подбора эффективного персонализированного лечения СППВР в озеро данных добавляли информацию из различных открытых мировых баз данных, таких как GENCODE, dbNSFP, COSMIC, ClinVar, 1000 Genome project, ExAC, dbSNP, OMIM, gnomAD, Pfam, UniProtKB, DoCM, CancerHotspots, CancerMine, OpenTargetsPlatform, ChEMBL, CIViC, CGI (Таблица 2).

Таблица 2. Открытые мировые базы данных, информацию из которых интегрировали в создаваемое озеро данных для аннотирования геномных вариантов и аккумуляции сведений о влиянии различных биомаркеров на чувствительность к широкому спектру противоопухолевых лекарственных препаратов

№	База данных	Описание	Ссылка
1	GENCODE, Human Release 43	<p>Проект, предоставляющий аннотированные последовательности генов и другие геномные данные для разных видов. Содержит следующие типы информации:</p> <ul style="list-style-type: none"> ● Аннотации генов: информация о структуре генов, таких как их экзоны, интроны, 5'-UTR и 3'-UTR. Информация о продуктах генов, таких как белки и кодирующие РНК. Информация о регуляторных элементах генов, таких как промоторы и энхансеры. ● Данные о транскриптах: информация о различных транскриптах генов, которые могут быть произведены. Информация об экспрессии генов в различных тканях и условиях. ● Данные о белках: информация о белках, кодируемых генами. Информация о функциях белков. ● Данные о вариативности генов: информация о вариантах генов, таких как SNP и вставки/делеции. Информация о влиянии вариантов генов на фенотип. ● Данные о сравнительной геномике: информация о консервативных элементах генов в разных видах. Информация об эволюции генов. 	https://www.encodegenes.org/
2	dbNSFP - Database of Non-synonymous Single Nucleotide Polymorphism Functional Predictions, v5.1.0	<p>База данных, предоставляющая функциональные предсказания и аннотации для всех потенциальных несинонимичных однонуклеотидных вариантов (nsSNVs) в геноме человека. Содержит следующие типы информации:</p> <ul style="list-style-type: none"> ● Информация о вариантах: позиция в геноме. Тип мутации (например, замена С на Т). Эффект мутации на аминокислоту (например, замена аланина на валин). Частота встречаемости мутации в популяциях. ● Функциональные прогнозы: расчет SIFT: предсказывает вероятность того, что мутация повлияет на функцию белка. “Скор” PolyPhen-2: предсказывает вероятность того, что мутация повлияет на структуру белка. “Скор” LRT: предсказывает вероятность того, что мутация повлияет на экспрессию белка. “Скор” MutationTaster: предсказывает вероятность того, что мутация повлияет на стабильность белка. ● Аннотации генов: идентификатор гена. Название гена. Тип гена (например, кодирующий белок, некодирующий РНК). Функции гена. Экспрессия гена. 	https://sites.google.com/site/jpopgen/dbNSFP
3	COSMIC - Catalogue of Somatic Mutations In Cancer, v91	<p>База данных, содержащая информацию о соматических мутациях в раковых генах. Представляет собой ценный ресурс для исследователей, которые изучают причины и лечение рака. Содержит следующие типы информации:</p> <ul style="list-style-type: none"> ● о различных типах мутаций, включая замены нуклеотидов, вставки и делеции. ● о генах, в которых происходят мутации, а также о функциях этих генов. ● о типах рака, в которых встречаются мутации, а также о частоте встречаемости мутаций в этих типах рака. 	https://cancer.sanger.ac.uk/cosmic

№	База данных	Описание	Ссылка
4	ClinVar, v 04_2024	<p>Общедоступная база данных, которая агрегирует и каталогизирует информацию о взаимосвязи между геномными вариантами и их влиянием на здоровье человека. Информация в ClinVar включает:</p> <ul style="list-style-type: none"> ● Варианты генов: SNP, вариации числа копий (CNV) и другие структурные изменения в геноме человека. ● Клинические и фенотипические данные: описания заболеваний, симптомов и других клинически значимых признаков, связанных с вариантами генов. ● Функциональные аннотации: информация о том, как варианты генов могут влиять на функции генов и белков. Данные об интерпретации: Заключение экспертов о клинической значимости вариантов генов. 	https://www.ncbi.nlm.nih.gov/clinvar/
5	1000 Genome project, Phase 3b	<p>Международная исследовательская инициатива, направленная на создание подробного каталога генетических вариаций человека. Целью проекта было секвенировать геномы 1000 человек из разных популяций мира. База данных проекта "1000 Геномов" содержит:</p> <ul style="list-style-type: none"> ● Данные секвенирования: секвенированные геномы 1000 человек, а также геномы дополнительных людей, секвенированные в рамках других проектов. ● Варианты генов: информация о миллионах генетических вариантов, обнаруженных в секвенированных геномах. ● Аннотации генов: информация о генах, включая их функции и расположение в геноме. ● Данные о частоте аллелей: информация о том, насколько распространены различные варианты генов в разных популяциях. ● Функциональные аннотации: информация о том, как варианты генов могут влиять на функции генов и белков. 	https://www.internationalgenome.org/
6	ExAC - Exome Aggregation Consortium, v2.1.1	<p>Представляет собой общедоступную базу данных с информацией о секвенировании экзонов человека. Цель проекта - стать исчерпывающим ресурсом для исследователей, изучающих генетику заболеваний человека. ExAC объединила данные секвенирования экзонов более 60 000 человек из разных популяций. Секвенирование экзонов фокусируется на кодирующих областях генов, где встречается большинство вызывающих заболевания мутаций. Включает следующие данные:</p> <ul style="list-style-type: none"> ● Варианты: генетические изменения, выявленные в экзонах. ● Частота аллелей: как часто каждый вариант встречается в популяции. ● Функциональные аннотации: прогнозируемое влияние вариантов на гены и белки. 	https://exac.broadinstitute.org/
7	dbSNP - Database of Single Nucleotide Polymorphisms, v155	<p>Общедоступная база данных, содержащая информацию о SNP в ДНК человека. SNP - это наиболее распространенный тип генетических вариаций, и они могут играть роль в развитии многих заболеваний. dbSNP содержит следующую информацию:</p> <ul style="list-style-type: none"> ● Местоположение SNP: хромосома и положение в хромосоме, где расположен SNP. ● Изменения нуклеотидов: какой нуклеотид (A, C, T или G) заменен в SNP. ● Частота аллелей: как часто встречаются разные аллели SNP в популяции. ● Функциональные аннотации: как SNP может влиять на функции генов и белков. ● Связь с заболеваниями: связан ли SNP с какими-либо заболеваниями или фенотипами. 	https://www.ncbi.nlm.nih.gov/snp/

№	База данных	Описание	Ссылка
8	OMIM - Online Mendelian Inheritance in Man, 6.40	<p>В переводе с английского языка означает "Онлайн-каталог менделевского наследования у человека". Представляет собой общедоступную базу данных, в которой собрана информация о известных наследственных заболеваниях человека и генах, ответственных за их развитие.</p> <p>Информация в OMIM включает:</p> <ul style="list-style-type: none"> ● Названия и номера OMIM: уникальные идентификаторы для каждого заболевания или гена. ● Синонимы: различные названия заболевания или гена. ● Наследование: тип наследования заболевания (аутосомно-доминантный, аутосомно-рецессивный, сцепленный с X-хромосомой и т. д.). ● Клинические проявления: подробное описание симптомов и признаков заболевания. ● Молекулярную генетику: информация о гене, вызывающем заболевание, включая его хромосомное расположение, структуру и функцию. ● Ссылки на литературу: список научных статей, посвященных данному заболеванию или гену. 	https://www.omim.org
9	gnomAD - Genome Aggregation Database, v3.1.1	<p>Общедоступный ресурс, содержащий информацию о вариативности геномов и экзомов человека. Он был создан международной коалицией исследователей и объединяет данные секвенирования геномов и экзомов более чем 150 000 человек. gnomAD - ценный ресурс для исследователей, изучающих генетику человека, заболевания и эволюцию. База данных gnomAD содержит следующие типы информации:</p> <ul style="list-style-type: none"> ● Частота аллелей: как часто встречаются различные варианты генов в популяции. ● Функциональные аннотации: как варианты генов могут влиять на функции генов и белков. ● Связь с заболеваниями: связан ли вариант гена с каким-либо заболеванием или фенотипом. ● Данные по популяциям: информация о частоте аллелей и функциональных аннотациях в различных популяциях мира. 	https://gnomad.broadinstitute.org/
10	Pfam, v 35.0	<p>База данных белковых семейств и доменов. База данных Pfam содержит следующие типы информации:</p> <ul style="list-style-type: none"> ● Информация о семействах белков: название: уникальное имя, присвоенное каждому семейству. Описание: функциональное описание семейства и его эволюционная история. Члены: список белков, которые были отнесены к данному семейству. Выравнивание: множественное выравнивание последовательностей белков из семейства. Профиль НММ: скрытая марковская модель (НММ), которая представляет собой статистическую модель семейства белков. Структура: информация о трехмерной структуре белков из семейства, если она доступна. ● Аннотации: онтологии: связь семейств белков с онтологиями Gene Ontology (GO) и InterPro. Домены: информация о доменах белков, которые входят в состав членов семейства. Литература: ссылки на научные публикации, в которых описывается семейство белков. ● Инструменты: инструменты для поиска семейств белков по названию, последовательности или аннотации. Инструменты для просмотра информации о семействах белков, включая выравнивания, профили НММ и структуры. 	http://pfam.xfam.org/

№	База данных	Описание	Ссылка
11	UniProtKB, v 2024_04	<p>База данных белковых последовательностей и их функциональной аннотации. База данных содержит следующую информацию:</p> <ul style="list-style-type: none"> ● Идентификаторы: UniProtKB ID: уникальный идентификатор каждого белка. Accession numbers: несколько идентификаторов из других баз данных. ● Названия: рекомендуемое название: краткое и информативное название белка. Альтернативные названия: синонимы и другие названия белка. ● Функции: описание функций белка, включая его каталитическую активность, субстраты и продукты. Биологические процессы, в которых участвует белок. ● Молекулярные характеристики: аминокислотная последовательность. Молекулярная масса. Посттрансляционные модификации. ● Структура: ссылки на базы данных структур белков (например, PDB). Предсказания вторичной структуры. ● Взаимодействия: Ссылки на базы данных белок-белковых взаимодействий (например, IntAct). Взаимодействия с другими биомолекулами (например, ДНК, РНК). ● Субклеточная локализация: где в клетке находится белок. ● Организм: из какого организма получен белок. ● Таксономия: биологическая классификация организма. ● Литература: ссылки на научные публикации, описывающие белок. ● Геномная информация: ссылки на базы данных генов (например, Ensembl). Информация о транскриптах. Варианты генов. ● Болезни: связь белка с генетическими заболеваниями. 	https://www.uniprot.org/
12	DoCM, v2.1	<p>Курируемая база данных, содержащая информацию о мутациях, связанных с раком. База содержит следующую информацию:</p> <ul style="list-style-type: none"> ● Информация о мутациях: позиция в геноме. Тип мутации (например, замена С на Т). Эффект мутации на белок-шаперон (например, замена глицина на аргинин). Частота встречаемости мутации в популяциях ● Функциональные характеристики белков-шаперонов: Тип белка-шаперона (например, HSP70, HSP90). Функции белка-шаперона (например, сворачивание белков, защита белков от агрегации). Взаимодействующие белки ● Клинические сведения: связь мутации с заболеваниями (например, нейродегенеративные заболевания, рак). Механизмы, лежащие в основе патогенности ● Подходы к лечению: экспериментальные данные: Влияние мутации на функцию белка-шаперона. Влияние мутации на агрегацию белков. Влияние мутации на взаимодействие с другими белками ● Структурные данные: 3D-структуры белков-шаперонов. Влияние мутации на структуру белка. Молекулярные механизмы патогенности 	http://docm.info/

№	База данных	Описание	Ссылка
13	CancerHotspots, v2.1	<p>База данных, содержащая информацию о генетических "горячих точках", которые часто мутируют в раковых клетках. База содержит следующую информацию:</p> <ul style="list-style-type: none"> ● Информация о мутациях: позиция в геноме. Тип мутации (например, замена С на Т). Эффект мутации на аминокислоту (например, замена аланина на валин). Частота встречаемости мутации в опухолях. Тип рака, в котором чаще всего встречается мутация ● Аннотации генов: идентификатор гена. Название гена. Тип гена (например, кодирующий белок, некодирующий РНК). Функции гена. Экспрессия гена ● Экспериментальные данные: данные о функциональных исследованиях мутаций. Данные о взаимодействии белков. Данные о сигнальных путях ● Клинические данные: прогностический статус мутаций для онкологических больных. Ответ на терапию. 	http://www.cancerhotspots.org/
14	CancerMine, v 12_2023	<p>База данных, содержащая информацию о генах, связанных с раком, и их функциях. База содержит следующую информацию:</p> <ul style="list-style-type: none"> ● Соотношения ген-рак: определяет, является ли ген драйвером, онкогеном или супрессором опухоли в данном типе рака. Уровень доказательности для каждого соотношения. Источники информации, на которых основано соотношение (например, научные публикации). ● Информация о гене: название гена. Символ гена. Онтология болезней. ID HUGO. ID Entrez Gene ● Информация о раке: тип рака. Онтология болезней. ID DOID. ● Функциональные аннотации: функции гена. Доменные аннотации. Пути GO. Взаимодействия белков 	https://pubmed.ncbi.nlm.nih.gov/31110280/
15	OpenTargets Platform, v GRCh38	<p>Интегративная база данных, которая связывает гены с болезнями для идентификации и валидации мишеней для лекарств. База содержит следующую информацию:</p> <ul style="list-style-type: none"> ● Данные о генах: информация о генах: ID гена, название гена, синонимы, хромосома, положение гена, транскрипты, белковые продукты. Функции генов: Функции генов, основанные на аннотациях GO, KEGG и других базах данных. Регулировка экспрессии: Данные о экспрессии генов из различных источников, таких как RNA-seq и микроматрицы. Взаимодействия генов: Данные о физических и функциональных взаимодействиях генов из различных источников, таких как DIP и BioGRID. Ассоциации с заболеваниями: Данные о генетических ассоциациях с заболеваниями из различных источников, таких как GWAS и eQTL. ● Данные о болезнях: информация о болезнях: Название болезни, синонимы, OMIM ID, описание болезни. Генетические факторы: Гены, связанные с болезнью, основаны на данных GWAS, eQTL и других исследованиях. 	https://platform.opentargets.org/

№	База данных	Описание	Ссылка
16	ChEMBL, v 31	<p>База данных, содержащая информацию о молекулах, связанных с лекарствами, их свойствах, активности и взаимодействиях. База содержит следующую информацию:</p> <ul style="list-style-type: none"> • Химические данные: структурные данные: 2D и 3D структуры молекул. Химические названия и синонимы. SMILES и InChI коды. Физико-химические свойства (растворимость, рКа, логарифмическое распределение октанола/воды). Данные о синтезе: Маршруты синтеза. Реактивы и условия. Данные о спектрах: Спектры ЯМР (1H, 13C, etc.). Масс-спектры. • Геномные данные: информация о генах: Идентификаторы генов. Названия генов. Расположение генов в геноме. Белковые продукты генов. Данные о вариациях: SNPs. CNVs. Вставки и делеции. • Данные об испытаниях на людях: клинические исследования: Фазы исследования. Дизайн исследования. Исходные и конечные точки. Побочные эффекты. Данные о фармакокинетике: Всасывание, распределение, метаболизм и выведение. Период полувыведения. Связывание с белками. Данные о безопасности: Серьезные нежелательные явления. Противопоказания. Взаимодействия с лекарствами. 	https://www.ebi.ac.uk/chembl/
17	CIViC, v1.0.1	<p>База данных, содержащая информацию о клинической интерпретации геномных вариантов. База содержит следующую информацию:</p> <ul style="list-style-type: none"> • Информация о вариантах: позиция в геноме: Координата мутации в геномной последовательности. Тип мутации: Тип мутации (например, замена С на Т, вставка, делеция). Эффект мутации: Влияние мутации на аминокислоту (например, замена аланина на валин). Частота встречаемости мутации: Частота мутации в популяциях. Наследственность: Является ли мутация унаследованной или приобретенной. • Клинические доказательства: тип рака: Тип рака, для которого мутация изучена. Вид доказательств: Тип доказательств, поддерживающих клиническую значимость мутации (например, функциональные исследования, клинические исследования, данные о пациентах). Уровень доказательств: Уровень доказательств, подтверждающих клиническую значимость мутации (например, патогенный, вероятно патогенный, неопределенный). Литература: Ссылки на научные публикации, в которых представлены доказательства. • Аннотации генов: идентификатор гена: Уникальный идентификатор гена. Название гена: Название гена. Функция гена: Известные функции гена. Сигнальные пути: Сигнальные пути, в которых участвует ген. Взаимодействия белков: Белки, с которыми взаимодействует продукт гена. • Информация о фармакогеномике: лекарства: Лекарства, на которые может влиять мутация. Механизм действия: Как мутация влияет на действие лекарства. Клинические рекомендации: Рекомендации по лечению пациентов с этой мутацией. 	https://civicdb.org/

№	База данных	Описание	Ссылка
18	CGI - Cancer genome interpreter, v 12_2023	<p>База данных, содержащая информацию о связи опухолевых геномных вариантов и потенциальных терапевтических мишеней. База содержит следующую информацию:</p> <ul style="list-style-type: none"> ● Информация о мутациях: тип мутации (например, замена нуклеотида, вставка, делеция). Позиция мутации в геноме. Эффект мутации на белок (например, миссенс-мутация, нонсенс-мутация). Частота встречаемости мутации в различных типах рака. ● Функциональная аннотация генов: функции гена. Экспрессия гена. Взаимодействия генов. Сигнальные пути. ● Клиническая информация: тип рака. Стадия рака. Прогноз выживания. Ответ на лечение. ● Данные о лекарствах: механизмы действия лекарств. Побочные эффекты лекарств. Резистентность к лекарствам. ● Исследования: научные публикации, связанные с мутациями и генами, связанными с раком. Клинические испытания, изучающие новые методы лечения рака. 	https://www.cancergenomeinterpreter.org/home

Для интеграции информации из открытых мировых баз данных реализовывали 3 этапа: извлечение, трансформацию и сохранение данных.

Этап 1. Извлечение данных:

- Определение источников. Производили идентификацию всех источников данных, которые необходимо извлечь, в том числе внутренние базы данных, файлы CSV, веб-API и т.д. Производили документирование каждого источника данных, включая его тип, формат, частоту обновления и иные важные сведения.

- Выбор метода извлечения. Был выбран метод потокового извлечения, при котором данные извлекались из источника в режиме реального времени по мере их поступления.

- Реализация извлечения. Использовали инструменты ETL, которые автоматизируют процесс извлечения данных. В настоящей работе использовали Dataiku. ETL-процессы были реализованы с помощью сценариев на языках программирования, таких как Python или Java. API использовали для извлечения данных из веб-приложений и других источников. Пример использования функции request в Python: `x=requests.get('http://url_source');`

- Обеспечение точности и полноты. Осуществляли проверку качества данных, в рамках которой проверяли качество извлеченных данных с использованием языка программирования Python и библиотеки Pandas на наличие ошибок, отсутствующих значений и дубликатов. Данные загружались в библиотеку Pandas. Далее над каждым столбцом выполняли следующие операции: удаление пустых строк с использованием функции `dropna` в Pandas, удаление дубликатов, корректировка данных. За основу выбирали 4 источника, данным которым мы доверяем. Если в 5 источнике по каким-либо позициям совпадало 90%, то записывали данные, приведенные в 4 источниках, которым мы доверяем. Например, если ген *ABCBI* в новом источнике был обозначен как *ABCBI*, то мы расценивали это как ошибку и исправляли на корректное обозначение. Пропуски заполняли медианным значением. Обработку ошибок, возникающих при извлечении данных, осуществляли как было описано выше. Производили аудит данных: отслеживали и документировали процесс извлечения данных для обеспечения его прозрачности.

Этап 2. Трансформация данных:

- Очистка данных. Осуществляли удаление всех неверных, неполных или дублирующихся данных. Производили стандартизацию форматов, в рамках которой данные были преобразованы в единый формат JSON, и обработку отсутствующих значений.

- Преобразование формата. Осуществляли преобразование данных из одного типа в другой, например, из строкового в числовой и агрегирование данных.

- Фильтрация данных. Применяли фильтры к данным, чтобы выбрать только те, которые соответствуют определенным критериям.

- Обеспечение согласованности. Осуществляли проверку целостности данных путем проверки на противоречия и несоответствия. Производили стандартизацию справочных данных путем разработки справочников и осуществляли документирование преобразований.

Этап 3. Сохранение данных:

- Выбор целевого хранилища. Были разработаны требования к хранилищу данных, такие как объем данных, производительность, масштабируемость и безопасность. Производили сравнение производительности различных систем управления базами данных (СУБД), по итогам которого была выбрана реляционная СУБД PostgreSQL (Таблица 3). Осуществляли подготовку целевого хранилища: создавали необходимые таблицы, индексы и другие структуры в целевом хранилище данных.

Таблица 3. Сравнение производительности различных систем управления базами данных

СУБД	Максимальный размер таблицы	Операция	Примерное количество операций в секунду
MySQL	8 ТБ (до 32 ТБ с использованием файловой системы XFS)	INSERT	1000-5000
		SELECT (простой)	5000-10000
		SELECT (сложный)	100-1000
		UPDATE	500-2000
		DELETE	500-2000
		JOIN	100-500
PostgreSQL	32 ТБ	INSERT	2000-10000
		SELECT (простой)	10000-20000
		SELECT (сложный)	500-2000
		UPDATE	1000-5000
		DELETE	1000-5000
		JOIN	500-1000
MSSQL	2 ТБ (до 16 ТБ с использованием файловой системы NTFS)	INSERT	500-2000
		SELECT (простой)	1000-5000
		SELECT (сложный)	100-500
		UPDATE	200-1000
		DELETE	200-1000
		JOIN	100-500

Аббревиатуры: СУБД - система управления базами данных.

С периодичностью 4 раза в год или при получении данных о добавлении информации в источник осуществляли актуализацию информации, посредством загрузки новых сведений. Для осуществления актуализации реализовывали такие же этапы, как и при процессе интеграции информации, при этом использовали следующие инструменты:

- Для извлечения данных: язык Python,
- Для трансформации: язык Python, SQL

- Для сохранения данных в хранилище: язык Python, SQL.

2.3.3. Разработка биоинформатических конвейеров для обработки данных полноэкзомного секвенирования

2.3.3.1. Определение требований к биоинформатическим конвейерам

Биоинформатический конвейер представляет собой совокупность аналитических и вычислительных инструментов, предназначенных для обработки и анализа биоинформатических данных [104]. На настоящий момент существует широкий ряд доступных биоинформатических конвейеров, обладающих различными особенностями и функциями.

В соответствии со спецификацией, конвейер включает в себя совокупность инструментов и программ, которые автоматически выполняют необходимые задачи по обработке данных, начиная от их первичной очистки и заканчивая сложным анализом и визуализацией результатов. Каждый шаг в конвейере выполняет определенную функцию, например, фильтрацию низкокачественных данных, выравнивание последовательностей по референсному геному, идентификацию геномных вариантов и т.д. [298, 383].

В рамках настоящего исследования была поставлена задача создать собственные оригинальные биоинформатические конвейеры при несоответствии существующих конвейеров выдвигаемым требованиям. Существующие конвейеры Sarek [130], Vcbio-Nextgen [150], SpeedSeq [82], DNAP [75], и DNAscan2 [230] были проанализированы на соответствие сформулированным нами требованиям.

Для анализа биоинформатических конвейеров были нами выдвинуты общие и специальные требования. Требования к биоинформатическим конвейерам для анализа данных полноэкзомного секвенирования, особенно в контексте анализа нормальной и опухолевой ДНК, включают несколько ключевых аспектов. Эти требования касаются корректной обработки данных, их качества, анализа мутаций и интерпретации результатов.

Лучшие практики построения биоинформатических конвейеров GATK, разработанные в Институте Броуда, — это один из наиболее признанных стандартов для анализа данных секвенирования, включая полноэкзомное секвенирование. Лучшие практики GATK включают этапы для нормальных и опухолевых образцов, начиная от предобработки до вызова соматических геномных вариантов [404]. Требования к структуре и функциям биоинформатических конвейеров, которые могли бы быть использованы для целей настоящей работы, были основаны на требованиях, изложенных в лучших практиках GATK. Кроме того,

при выдвижении общих требований принимались во внимание особенности используемой инфраструктуры.

Общие требования являются относительно схожими, однако необходимо учитывать также частные случаи для каждого образца, как в случае поиска герминальных вариантов, так и соматических вариантов в опухолевых образцах. Сформулированные нами общие и специальные требования к биоинформатическим конвейерам приведены в таблице 4.

Таблица 4. Общие и специальные требования, предъявляемые к биоинформатическим конвейерам

Общие требования		
№	Требование	Характеристика
1	Масштабируемость	Биоинформатический конвейер должен быть масштабируемым для обработки больших объемов данных.
2	Гибкость	Возможность настройки параметров и добавления дополнительных шагов анализа в зависимости от особенностей конкретного исследования.
3	Документация	Подробное описание всех шагов биоинформатического конвейера, включая параметры и используемые инструменты.
4	Воспроизводимость	Обеспечение воспроизводимости результатов с возможностью повторного запуска биоинформатического конвейера на тех же данных.
5	Возможность использовать docker	Использование контейнеров docker для "оборачивания" шагов биоинформатического конвейера для его отказоустойчивости.
Специальные требования		
6	Сортировка и контроль качества Fastq	Оценка основных статистик качества (Phred scores) и общей чистоты данных.
7	Тримминг адаптеров секвенирования	Удаление адаптерных последовательностей и низкокачественных участков.
8	Выравнивание на референс	Обеспечение высокого процента выравнивающихся чтений.
9	Метрики выравнивания	Вычисление равномерности и глубины покрытия для обнаружения областей недостаточного покрытия.
10	Маркирование дубликатов	Возможность отмечать оптические ПЦР-дубликаты в выровненном файле, но не удалять их.
11	Удаление дубликатов	Удаление дубликатов для предотвращения искажения результатов анализа.
12	Метрики дубликатов	Оценка количества оптических дубликатов и их распределения в геноме.
13	Конвертация SAM в BAM и фильтрация BAM	Конверсия из формата SAM в формат BAM и фильтрация по параметрам качества
14	Обнаружение герминальных вариантов	Герминальные варианты, которые присутствуют как в опухолевой, так и в нормальной ткани, могут указывать на генетическую предрасположенность, но не являются соматическими геномными вариантами.

Специальные требования		
№	Требование	Характеристика
15	Обнаружение соматических вариантов	Обычно анализ опухоль-норма фокусируется на обнаружении соматических SNV/INDEL, поэтому фильтрация герминальных вариантов является первостепенной задачей также, как и использование для этого панели “нормалей”.
16	Перекрёстная проверка вызова SNV/INDEL	После объединения результатов трех программ для выявления вариантов, произвести поиск только тех вариантов, которые были вызваны всеми тремя.
17	Обнаружение CNV	Поиск вариантов копийности генов.
18	Аннотация вариантов	Дополнительная аннотация вариантов с использованием баз данных о геномных вариантах и их функциональных последствиях.
19	Метрики вызова вариантов	Обязательно наличие метрик: глубина покрытия прочтений (Depth of Coverage, DP), Качество “ридов” (прочтений) (Read Quality, GQ/QUAL), частота аллелей Allele Frequency (AF).
20	Фильтрация вариантов	Установка порогов качества и фильтров для минимизации ложноположительных результатов.
21	Сравнение с внешними базами данных	Для герминальных вариантов особенно важно провести проверку внешних баз данных, таких как ClinVar, HGMD, gnomAD и других, для оценки их клинической значимости и связанных с ними фенотипов.
22	Клиническая значимость	Оценка клинической значимости герминальных вариантов важна для определения их возможного влияния на здоровье пациента или его риск развития заболеваний.

Аббревиатуры: CNV (англ. copy number variation) - вариация числа копий, INDEL (англ. insertions/deletions) - вставки/делеции, SNV (англ. single nucleotide variant) - однонуклеотидный вариант

Помимо общих требований, к биоинформатическим конвейерам применяются требования к метрикам. Требования к метрикам и их значениям могут варьироваться в зависимости от конкретного исследования и целей анализа. Биоинформатические конвейеры для обработки данных полноэкзомного секвенирования должны соответствовать этим требованиям, чтобы обеспечить точность, надежность и воспроизводимость результатов анализа. Однако нужно учитывать также частные случаи для каждого образца, как в случае поиска герминальных вариантов, так и в соматических вариантах опухолевых образцов.

2.3.3.2. Разработка биоинформатических конвейеров для обработки данных полноэкзомного секвенирования

По итогам анализа биоинформатических конвейеров Sarek, Bcbio-Nextgen, SpeedSeq, DNAP и DNAscan2 было выявлено, что ни один из проанализированных конвейеров не соответствовал всем сформулированным общим и специальным требованиям (результаты анализа подробно описаны в разделе 3.3.4. “Анализ существующих биоинформатических конвейеров” настоящей работы).

С учетом несоответствия существующих решений предъявляемым требованиям, осуществлялся подбор программных инструментов для разработки собственных биоинформатических конвейеров, соответствующих выдвинутым нами требованиям. Выбранные программные инструменты представлены в таблице 5.

Таблица 5. Выбранные программные инструменты, применяемые на различных шагах биоинформатических конвейеров

Определенные специальные требования	Выбор программного инструмента, выполняющего предъявляемые требования	Ссылка
Сортировка и контроль качества Fastq	BBmap FastP v0.23.4	https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbmap-guide/ (https://github.com/OpenGene/fastp)
Тримминг адаптеров секвенирования	FastP v0.23.4	(https://github.com/OpenGene/fastp)
Выравнивание на референс	BWA v2.2	https://github.com/bwa-mem2/bwa-mem2
Метрики выравнивания	Picard v3.1.1	(https://github.com/broadinstitute/picard)
Маркирование дубликатов	Streammd v4.3.0	(https://github.com/delocalizer/streammd)
Удаление дубликатов	Streammd v4.3.0	(https://github.com/delocalizer/streammd)
Метрики дубликатов	Streammd v4.3.0	(https://github.com/delocalizer/streammd)
Конвертация SAM в BAM и фильтрация BAM	Sambamba v1.0.1.	(https://lomereiter.github.io/sambamba/)
Герминальные SNV/INDEL	Google DeepVariant v1.6.1	(https://github.com/google/deepvariant)
Соматические SNV/INDEL	GATK Mutect2 v4.1.4.1 Rfcaller v1.2.0 Varnet v1.1.0	(https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2) (https://github.com/xa-lab/RFcaller/) (https://github.com/skandlab/VarNet)
перекрёстная проверка вызова SNV/INDEL	bcftools v1.9	(https://samtools.github.io/bcftools/bcftools.html)
Обнаружение CNV	CNVkit v0.9.11 Sequenza 3.0.0	(https://cnvkit.readthedocs.io/en/stable/) (https://bitbucket.org/sequenzatools/sequenza-utils/src/master/)
Метрики вызова вариантов	bcftools v1.9	(https://samtools.github.io/bcftools/bcftools.html)
Фильтрация вариантов	bcftools v1.9	(https://samtools.github.io/bcftools/bcftools.html)
Аннотация вариантов	Variant Effect Predictor v111.0	(https://www.ensembl.org/info/docs/tools/vep/index.html)
Сравнение с внешними базами данных	Variant Effect Predictor v111.0	(https://www.ensembl.org/info/docs/tools/vep/index.html)
Клиническая значимость, MSI, TMB	PCGR v1.4.1 CPSR v1.0.0	(https://github.com/sigven/pcgr) (https://github.com/sigven/cpsr)

Аббревиатуры: формат BAM (англ. Binary Alignment Map) - карта двоичного выравнивания, CPSR (Cancer predisposition sequencing reporter) - программа оценки предрасположенности к онкологии на основе секвенирования), CNV (англ. copy number variation) - вариация числа копий, GATK (англ. Genome Analysis Toolkit) - Набор инструментов для анализа генома, INDEL - (англ. insertions/deletions) - вставки/делеции, PCGR (англ. Personal Cancer Genome Reporter) - программа Персональный репортёр генома рака, SNV (англ. single nucleotide variant) - однонуклеотидный вариант.

С использованием выбранных программных инструментов осуществляли разработку собственных биоинформатических конвейеров для определения герминального и соматического

геномного профиля, соответствующих также общим требованиям масштабируемости, гибкости, документации, воспроизводимости и возможности использовать docker (<https://hub.docker.com/>).

Разработка биоинформатических конвейеров основывалась на лучших практиках построения биоинформатических конвейеров GATK, разработанных в Инстите Броуда (<https://www.broadinstitute.org>), с соответствующими необходимыми модификациями. Вносимые модификации были связаны с тем, что разрабатываемые биоинформатические конвейеры должны показывать высокий уровень анализа при работе с данными секвенирования, полученными при использовании секвенатора MGISEQ-G400 (MGI, Китай), в то время как лучшие практики построения биоинформатических конвейеров GATK в первую очередь применимы к данным секвенирования, получаемым с использованием оборудования Illumina, США [404].

В соответствии с лучшими практиками построения биоинформатических конвейеров GATK, в состав разрабатываемых биоинформатических конвейеров включали 3 блока:

1. Предобработка. В рамках данного блока осуществлялись процессы фильтрации необработанных FASTQ файлов и их выравнивание на референсный геном для формирования BAM-файла, используемого для дальнейшего выявления геномных вариантов.
2. Выявление геномных вариантов. Этап, в рамках которого использовали различные программные инструменты для выявления различных типов вариантов генома, включая SNV, вставки/делеции, варианты числа копий, и формирования итогового VCF-файла.
3. Постпроцессинг и формирование отчетов. В рамках данного этапа осуществлялась обработка полученного VCF-файла, в том числе фильтрация вариантов, аннотирование и формирование отчетов по итогам анализа.

Обработка данных велась на языках Bash, Python, R, многие плагины сделаны на языках Perl и C++.

Для создания собственных биоинформатических конвейеров были выбраны различные программные инструменты, обеспечивающие соответствие изложенным выше специальным и общим требованиям. Все выбранные для использования программные продукты находятся в открытом доступе.

Для сортировки FASTQ файлов использовали программный пакет BbMap. BbMap - это универсальный набор инструментов для биоинформатического анализа. Переформатирование и сортировка осуществлялись скриптом `reformat.sh`, позволяющим сортировать считываемые данные по названию или по другим критериям. Этот шаг обеспечивал правильную упорядоченность считываемых данных в сопряженном состоянии, что имеет решающее значение

для выравнивания и других последующих процессов, поскольку в нашей лаборатории часто используется “досеквенирование” образцов, после чего происходит их слияние [405].

Предварительную обработку и контроль качества FASTQ файлов осуществляли с использованием FastP v0.23.4 для обрезки адаптеров, фильтрации некачественных считываний для повышения качества прочтений. Альтернативой являлся программный пакет Trimmomatic. Но для обеспечения непрерывности биоинформатического конвейера и скорости обработки информации была выбрана программа Fastp v0.23.4, т.к. полученные по итогам её работы данные можно сразу посылать на конвейер выравнивания.

Выравнивание на референсный геном человека GRCh37 предварительно обработанных считываний как для герминальных, так и для опухолевых образцов реализовывали с использованием пакета BWA-MEM v2.2. Маркирование и удаление ПЦР-дубликатов в разработанных биоинформатических конвейерах осуществляли в программе Streammd v4.3.0. Контроль качества выравнивания выполняли путем конвертации SAM файла в сортированный и профильтрованный BAM-файл (бинарный формат, имеющий меньший размер, и обрабатываемый быстрее из-за наличия индекса .bai) после чего удаляли последовательности, где хотя бы один из парных “ридов” не подвергся выравниванию, не имеющие пары на другой молекулярной цепи, а также с качеством выравнивания меньше 20. Операции, описанные выше, а также сортировку и перекодирование SAM файла в BAM реализовывали в пакете Sambamba v1.0.1. Для контроля качества выравнивания использовали Picard Tools v3.1.1.

Выявление герминальных геномных вариантов осуществляли с применением программы на основе машинного обучения Google DeepVariant v1.6.1. DeepVariant представляет собой предобученную модель для высокоточного вызова вариантов с использованием методов машинного обучения [282]. DeepVariant позволяет выявлять варианты со значительно большей точностью, чем предыдущие классические методы [282]. Полученные в ходе коллинга варианты автоматически помечались программой как PASS и принимались к дальнейшей обработке.

DeepVariant является результатом более чем двухлетних исследований команды Google Brain в сотрудничестве с Verily Life Sciences. DeepVariant преобразует саму методику вызова варианта в задачу классификации изображений, хорошо подходящую для существующих технологий и опыта Google. Работа была начата с эталонных геномов GIAB, для которых существует высококачественная исходная информация (или самое близкое приближение, возможное в настоящее время). Используя множественные копии этих геномов, были созданы десятки миллионов обучающих примеров в виде многоканальных тензоров, кодирующих данные, полученные с секвенатора, а затем обучена модель классификации изображений на основе TensorFlow для идентификации истинной последовательности генома по экспериментальным данным, полученным приборами. Хотя результирующая модель глубокого

обучения DeepVariant не обладала специальными знаниями о геномике, в течение года она получила высшую награду за точность выявления SNP на конкурсе precisionFDA Truth Challenge, превзойдя самые современные методы. С тех пор дополнительно была снижена частота ошибок более чем на 50%. DeepVariant может быть развернута на доступных сегодня GCP в конфигурациях, оптимизированных для недорогих и быстрых изменений с использованием масштабируемых технологий, таких как Pipelines API.

Для выявления соматических геномных вариантов использовали программу Mutect2 v4.1.4.1, далее осуществляли кросс-валидацию полученных результатов с применением программ, основанных на использовании технологии машинного обучения RFcaller v1.2.0 и Varnet v1.1.0.

Mutect2 v4.1.4.1 от GATK применяли для выявления соматических вариантов путем сравнения опухолевых и нормальных образцов, определяли потенциальные соматические мутации малой и средней длины. Mutect2 v4.1.4.1 осуществляет выявление соматических коротких вариантов посредством локальной сборки гаплотипов. Коллинг включает в себя как поиск коротких повторов SNV, так и длинных вставок и делеций. Mutect2 использует байесовскую модель соматического генотипирования, которая отличается от оригинальной версии Mutect [85], и использует механизмы HaplotypeCaller, входящий в пакет GATK. Следует отметить, что Mutect2 версии 4.1.0.0 и далее позволяет проводить совместный анализ нескольких образцов. Для Mutect2 была также добавлена панель нормальных образцов (от англ. *"Panel of Normals"*), включающая герминальные варианты, что предотвращает получение ложноположительных данных о герминальных мутациях.

Помимо классических статистических подходов, в последние несколько лет наблюдается расширение использования машинного обучения для различных целей, включая разработку новых инструментов для выявления вариантов.

Для обеспечения эффективного выявления соматических геномных вариантов на настоящем этапе производилась кросс-валидация с применением программных инструментов RFcaller v1.2.0 и Varnet v1.1.0, использующих технологии машинного обучения.

RFcaller использует характеристики на уровне прочтений вместе со стратегиями машинного обучения для идентификации соматических вариантов (SNVs и вставок/делеций) из парных образцов нормальных и опухолевых клеток. Данный инструмент был обучен на данных полногеномного секвенирования (англ. WGS, whole genome sequencing), а его результаты были сравнены с результатами, полученными в исследовании PCAWG (англ. Pan-Cancer Analysis of Whole Genomes, Пан-онкологический анализ полного генома), и оказались весьма схожими с теми, которые получают при комбинировании нескольких инструментов [102]. Для разработки алгоритмов использовались два разных набора мутаций - обучающий и тестовый. Для их

создания с помощью bcftools были найдены все возможные соматические мутации из четырех образцов мантийноклеточной лимфомы секвенированной WGS, с покрытием 30× (M032 и M439 для обучения; M065 и M431 для тестирования). Для первоначального обучения ранее опубликованные мутации были определены как истинно положительные т.е. контрольные. После девяти раундов обучения алгоритмов и курирования набора мутаций все дискордантные варианты были изучены, что позволило получить надежный набор данных для обучения и тестирования финальной версии алгоритмов.

В качестве второго инструмента для кросс-валидации применяли программу VarNet v1.1.0, которая визуализирует выровненные прочтения опухолей и соответствующих нормальных геномов, включая качество секвенирования оснований, качество их отображения и смещение цепей ДНК [198]. Поскольку контролируемое глубокое обучение требует доступа к большим наборам данных с метками, которые в геномике рака обычно присутствуют в небольшом количестве, VarNet v1.1.0 использует слабо контролируемый подход к обучению, при котором генерируются высокодостоверные псевдометки по 7 типам рака и более чем 300 целым геномам рака. Оценка производительности VarNet на реальных и синтетических эталонных наборах данных опухолей продемонстрировала стабильную производительность, часто превосходящую существующие методы. Программа обучена на основе изображений 4,6 млн высокодостоверных соматических вариантов, аннотированных в 356 целых геномах опухолей. VarNet был протестирован на ряде общедоступных наборов данных, демонстрируя производительность, часто превышающую современные методы. В целом результаты демонстрируют, как масштабируемый подход глубокого обучения может дополнить и потенциально вытеснить разработанные человеком признаки и эвристические фильтры при поиске соматических вариантов [198].

Перекрытую проверку между программами для выявления вариантов в соматическом биоинформатическом конвейере проводили с помощью BCFtools v1.9. - intersec.

Герминальные и соматические CNV имеют большое значение при различных синдромах и онкологических заболеваниях. NGS все чаще используется для получения информации о числе копий на основе вариаций глубины чтения в данных секвенирования. Однако этот подход имеет ограничения в случае целевого повторного секвенирования, которое оставляет пробелы в покрытии между регионами, выбранными для обогащения, и вносит погрешности, связанные с эффективностью захвата мишени и подготовки библиотеки. В разрабатываемом герминальном биоинформатическом конвейере использовали новый метод определения CNV, реализованный в программном пакете CNVkit v0.9.11, который использует как целевые чтения, так и неспецифически захваченные внецелевые чтения для определения числа копий равномерно по всему геному. Такая комбинация позволяет достичь как хорошего разрешения на уровне экзонов

в целевых регионах, так и достаточного разрешения в более крупных интронных и межгенных регионах для выявления изменений числа копий. В частности, успешно было определено число копий с разрешением, эквивалентным 100-килобазному, в масштабах всего генома с помощью панели, охватывающей всего 293 гена [343].

Для определения CNV в рамках соматического парного биоинформатического конвейера был использован программный пакет Sequenza v3.0.0. Он использует данные парного секвенирования ДНК опухоли и нормальной ткани, а также имеет возможность анализировать только опухолевую ткань для оценки клеточности и пloidности опухоли, а также для расчета аллель-специфических профилей числа копий и профилей мутаций. Тестирование пакета Sequenza, было проведено другой научной группой на данных экзомных последовательностей 30 опухолей из The Cancer Genome Atlas. Оценка эффективности этих алгоритмов, показала схожесть результатов обработанными алгоритмом анализа аллель-специфического числа копий в опухолях (англ. allele-specific copy number analysis of tumors, ASCAT). Сравнение между Sequenza/exome и SNP/ASCAT выявило сильную корреляцию в оценках клеточности (r Пирсона = 0,90) и пloidности ($r = 0,42$, или $r = 0,94$ после ручной проверки альтернативных решений). Эта производительность заметно превосходила ранее опубликованные алгоритмы. Кроме того, в искусственно симулированных данных, имитирующих нормальные и опухолевые образцы, Sequenza определяла правильную пloidность в образцах с содержанием опухоли до 30 % [121].

Оценка TMB в рамках соматического биоинформатического конвейера определялась как количество соматических, кодирующих, единичных нуклеотидных замен и вставок/делеций на мегабазу исследуемого генома. Все замены оснований и вставки/делеции в кодирующей области целевых генов, включая синонимичные изменения, подсчитывали перед фильтрацией. Синонимичные мутации учитывали для снижения “шума” в выборке. Некодирующие изменения не учитывали. Изменения, перечисленные как известные соматические изменения в базе данных COSMIC, и изменения в генах-супрессорах опухолей не учитывали, поскольку анализируемые гены ориентированы на гены с функциональными мутациями в опухолевой ткани. Изменения, определённые как герминальные, также не учитывали. Известные герминальные мутации в dbSNP не учитывали. Не учитывали также мутации, встречающиеся в базе данных ExAC в числе 2 и более. Для расчета TMB на мегабазу общее количество подсчитанных мутаций делили на размер кодирующей целевой области. В дальнейшем для проверки значимости различий средних между двумя популяциями использовался непараметрический U-тест Манна-Уитни. Методика используемого расчета подробно описана в масштабном исследовании мутационной нагрузки для разных опухолей Chalmers и соавторов [77]. Данный алгоритм был имплементирован в пакете PCGR, расчёт производился там же.

Также в соматическом биоинформатическом конвейере производили определение MSI. MSI является результатом нарушенной системы мисматч-репарации ДНК и представляет собой клеточный фенотип, который имеет клиническое значение для многих типов рака. Анализ MSI имеет наибольшее значение для колоректального рака, рака желудка, раке эндометрия и рака яичников [92].

Для расчета MSI в соматическом биоинформатическом конвейере использовали опубликованный статистический классификатор MSI, который на основе анализа соматических мутаций может разделять MSH (высокая нестабильность) и MSS (стабильные) опухоли [252]. Классификатор был обучен с использованием 999 экзомных датасетов из TCGA (The Cancer Genome Atlas) с известным статусом MSI.

Статистический анализ и фильтрацию производили путем последующего анализа для каждого варианта по различным метрикам, такими как глубина прочтений ($DP > 10$), общий контроль качества ($QUAL > 30$), качество генотипирования ($GQ > 30$) в программе BCFtools v1.9 для обоих конвейеров.

Аннотирование полученных результатов и их интерпретацию осуществляли с помощью Variant Effect Predictor (VEP) v111.0, программного обеспечения для функциональной аннотации локусов количественных признаков, разработанного Европейским Институтом Биоинформатики Европейской молекулярно-биологической лаборатории. VEP представляет собой набор инструментов для анализа, аннотирования и определения приоритетов геномных вариантов в кодирующих и некодирующих областях. Он предоставляет доступ к обширной коллекции геномных аннотаций, с множеством интерфейсов, удовлетворяющих различным требованиям, и простыми опциями для настройки и детального анализа, имеет открытый исходный код и поддерживает полную воспроизводимость результатов, а его использование позволяет упростить и ускорить интерпретацию вариантов в широком спектре проектов исследований.

Инструменты VEP аннотируют две широкие категории геномных вариантов: (1) варианты последовательностей со специфическими и четко определенными изменениями (включая SNV, вставки, делеции, множественные замены пар оснований, микросателлиты и тандемные повторы); и (2) более крупные структурные варианты (длиной более 50 нуклеотидов), включая те, которые имеют изменения в количестве копий или вставки и делеции ДНК. Для всех входных вариантов VEP даёт подробную аннотацию о воздействии на транскрипты, белки и регуляторные области. Для известных или перекрывающихся вариантов включаются частоты аллелей и информация о заболевании или фенотипе.

В отличие от других инструментов, VEP также может обрабатывать идентификаторы вариантов (например, из dbSNP) и обозначения номенклатуры HGVS (например, HGVS с использованием Ensembl, RefSeq или LRG-транскриптов и белков

'ENST00000615779.4:c.102944T>C'; 'BRCA2:p.Val2466Ala'; 'Q15118:стр.Val42Phe'). Эти идентификаторы обычно используются в публикациях и отчетах. Эта функциональность также может быть использована для “обратного отображения” вариантов из координат кДНК или белка в геном и наоборот.

VEP может использоваться для анализа данных любого вида с собранной последовательностью генома и аннотированным набором генов. Файлы данных, необходимые для аннотирования у 80 видов позвоночных и многих беспозвоночных, распределены по геномам Ensembl и Ensembl Genomes, соответственно. Они регулярно обновляются, обеспечивая возможность проведения анализа с использованием современных биологических баз данных. VEP также поддерживает как последнюю версию GRCh38, так и предыдущие сборки GRCh37 для человека. Важно отметить, что все результаты полностью воспроизводимы с использованием архивных версий Ensembl. Наконец, исследователи могут использовать свои собственные данные расшифровки для анализа, например, у видов, еще не входящих в состав Ensembl, или для новых или частных аннотаций. В пакет скриптов VEP включен скрипт для создания набора аннотаций из пары файлов general feature format (GFF) и FASTA.

Отчеты по итогам работы биоинформатических конвейеров формировались с помощью инструмента Cancer Predisposition Sequencing Reporter v1.0.0 (CPSR v1.0.0) в герминальном биоинформатическом конвейере и с помощью Personal Cancer Genome Reporter (PCGR v1.4.1) в соматическом биоинформатическом конвейере.

CPSR основывается на структуре, разработанной для PCGR. Он принимает на вход файл в формате VCF, проводит анализ и сообщает о рисковых генах, связанных с определенным типом рака или синдромом. Он реализует интерпретацию генетических изменений в соответствии с рекомендациями ACMG (American College of Medical Genetics and Genomics) для ЗНО, чтобы классифицировать варианты в соответствии с патогенностью, и создает структурированный и интерактивный отчет о предрасположенности, в котором выделяются мутации с терапевтическими последствиями.

PCGR - Personal Cancer Genome Reporter - это автономный программный пакет для расшифровки индивидуальных опухолевых геномов, предназначенный для использования в диагностике. PCGR интерпретирует преимущественно соматические SNV/вставки/делеции и aberrации числа копий. Программное обеспечение расширяет базовые аннотации генов и вариантов из Ensembl's Variant Effect Predictor (VEP) с помощью онкологически-релевантных, современных аннотаций, получаемых через vcfanno, и создает интерактивные HTML-отчеты, предназначенные для клинической интерпретации. PCGR может выполнять множество типов анализов, включая:

1. Классификацию соматических вариантов (ACMG/AMP);

2. Картирование терапевтических и прогностических последствий соматических aberrаций ДНК;
3. Оценку TMB;
4. Анализ только опухоли (фильтрация вариантов);
5. Классификацию MSI.

2.3.4. Оценка показателей эффективности выявления геномных вариантов созданной фармакогеномной платформой

Оценку показателей эффективности работы созданной фармакогеномной платформы осуществляли путем сравнительного анализа выявленных значимых при ТНРМЖ патогенных/вероятно патогенных герминальных геномных вариантов и соматических геномных вариантов 1-2 классов патогенности. В отношении образцов периферической крови и опухолевой ткани больных ТНРМЖ ($n=100$) были осуществлены этапы пробоподготовки и секвенирования с использованием разработанного адаптированного протокола (п.3.2.1). Также была произведена биоинформатическая обработка сырых данных секвенирования с использованием разработанных биоинформатических конвейеров, содержащихся в фармакогеномной платформе (п.3.3.5). По результатам проведенного полноэкзомного секвенирования, были отобраны 30 образцов ТНРМЖ для сравнительного анализа: 28 образцов с выявленными геномными вариантами и 2 образца - без геномных вариантов в качестве отрицательного контроля.

Эти же 30 образцов были проанализированы с использованием наборов AVENIO (Roche, США), секвенатора NextSeq500 (Illumina, США) и программы AVENIO Oncology Analysis.

Пробоподготовку и секвенирование с применением наборов AVENIO (Roche, США) осуществляли в соответствии с инструкцией производителя [392]. При выделении ДНК из опухолевой ткани, залитой парафином, использовали наборы AVENIO Tumor Isolation and QC Kit (Roche, США). Контроль качества выделенной ДНК осуществляли при помощи ПЦР, концентрацию выделенной ДНК измеряли с использованием флуориметра Qubit 2.0 (Life technologies, США). Библиотеки из выделенной ДНК получали с использованием набора AVENIO Tumor Library Prep Kit (Roche, США). Баркодирование библиотек осуществляли при помощи ПЦР с наборами AVENIO Tumor Sample Primers Kit (Roche, США). Для обогащения библиотек использовали наборы AVENIO Tumor Enrichment Kit (Roche, США) с панелью AVENIO Tumor Tissue Expanded Panel Kit (Roche, США), в состав которой входят 77 онкоассоциированных генов. Концентрацию ДНК в изготовленных библиотеках измеряли на флуориметре Qubit 2.0 (Life Technologies, США), контроль качества осуществляли на биоанализаторе Agilent Bioanalyzer 2100 (Agilent Technologies, США). Секвенирование

проводили с использованием платформы NextSeq500 (Illumina, США). Биоинформатический анализ производили с применением программы AVENIO Oncology Analysis.

2.3.5. Создание системы поддержки принятия врачебных решений для подбора персонализированного лекарственного лечения онкологических заболеваний на базе фармакогеномной платформы

При разработке СППВР для подбора персонализированного лекарственного лечения онкологических заболеваний использовали языки программирования JavaScript, NodeJS, R, Python, Bash и системы управления пакетами conda, npm. В качестве основной операционной системы для разработки и выполнения программ использовали Linux (Debian 10).

В процессе разработки ПО были использованы следующие открытые программные пакеты, распространяемые по лицензии MIT (Массачусетского технологического института), позволяющие копировать, изменять, а также в любом виде распространять созданный с их помощью программный продукт:

1. NodeJS - <https://nodejs.org/en>
2. Playwright - фреймворк (приложение) для веб-автоматизации - <https://github.com/microsoft/playwright>
3. TAPES (Tool for Assessment and Prioritisation in Exome Studies) - <https://github.com/axavier/tapes>
4. Chromium - <https://www.chromium.org/chromium-projects/>

В структуре СППВР производили разработку трех элементов:

1. Модуля определения мутационного профиля опухоли;
2. Модуля подбора эффективного персонализированного лекарственного лечения на основе геномных вариантов и ТМВ, выявленных у больных ЗНО;
3. Модуля для формирования и визуализации отчета с результатами подбора персонализированного лекарственного лечения, выявленными геномными вариантами, показателями ТМВ, MSI, прогнозом течения заболевания и иной информацией.

В составе СППВР разрабатывали модуль определения мутационного профиля опухоли. Набор соматических мутаций, представленных в опухоли, является отражением мутационных процессов, происходивших в ней в течении ее существования, что, в свою очередь, позволяет получить информацию о путях, которые могли привести к канцерогенезу. Различные экзогенные мутагены, как, например, табачный дым или УФ-излучение, а также эндогенные мутагены, такие как активность ферментативного семейства АРОВЕС или дефекты системы мисматч-репарации

ДНК - приводят к специфичным (в каждом случае) профилям (паттернам) мутаций (то есть к различным паттернам типов замен в специфическом контексте последовательности ДНК). Пакет `MutationalPatterns` (<https://bioconductor.org/packages/release/bioc/html/MutationalPatterns.html>) был использован для оценки относительного вклада известных мутационных профилей в единичный образец опухоли. `MutationalPatterns` обеспечивает реконструкцию мутаций, наблюдаемых в данном образце, используя эталонную коллекцию из 67 мутационных профилей.

В рамках модуля подбора эффективного персонализированного лекарственного лечения ЗНО был создан трехуровневый алгоритм фильтрации. Алгоритм фильтрации был основан на поиске исследуемых генов и геномных вариантов в открытых мировых базах данных, содержащих информацию об ассоциациях биомаркеров с чувствительностью к лекарственным препаратам. С учетом того, что информация из различных открытых мировых баз данных была интегрирована в созданное в рамках настоящего исследования озеро данных, далее мы будем использовать термин “поиск в озере данных”. Обеспечение доступа СППВР к информации, содержащейся в озере данных, осуществлялось через встроенный протокол `Dataiku`, запускаемый кастомными скриптами на языке программирования `Python`.

В рамках алгоритма фильтрации были установлены следующие уровни фильтрации для подбора эффективного персонализированного лекарственного лечения на основе геномных вариантов, выявляемых у больных ЗНО:

1-й уровень фильтрации был реализован как фильтрация по типу гена. Формировали перечень генов, в которых у определенного больного были выявлены геномные варианты. Далее производили поиск наличия определенного гена в созданном озере данных. При отсутствии записей об определенном исследуемом гене в озере данных, дальнейший анализ гена прекращался. При наличии записей об определенном исследуемом гене в озере данных, приступали к следующему уровню фильтрации.

2-й уровень фильтрации был реализован как фильтрация по типу геномного варианта. Производили поиск наличия конкретного типа вариантов в созданном озере данных. К типам вариантов относили вставки, делеции, однонуклеотидные и иные геномные варианты. При отсутствии записей о конкретном типе геномного варианта в озере данных, дальнейший анализ прекращался. При наличии записей о конкретном типе геномного варианта в озере данных, приступали к следующему уровню фильтрации.

3-й уровень фильтрации был реализован как фильтрация по выявленному геномному варианту. Формировали перечень геномных вариантов, выявленных в генах, прошедших предыдущие уровни фильтрации. Далее производили скрининг озера данных на наличие записей о выявленном геномном варианте и нахождение клинических аннотаций к соответствующим вариантам. При наличии записей об определенном геномном варианте в озере данных,

производили экстракцию сведений об ассоциации данного геномного варианта с лекарственным препаратом, чувствительность к которому была продемонстрирована в различных исследованиях.

Схема применяемого алгоритма фильтрации для подбора эффективного персонализированного лекарственного лечения на основе геномных вариантов, выявляемых у больных ЗНО представлена на рисунке 9.

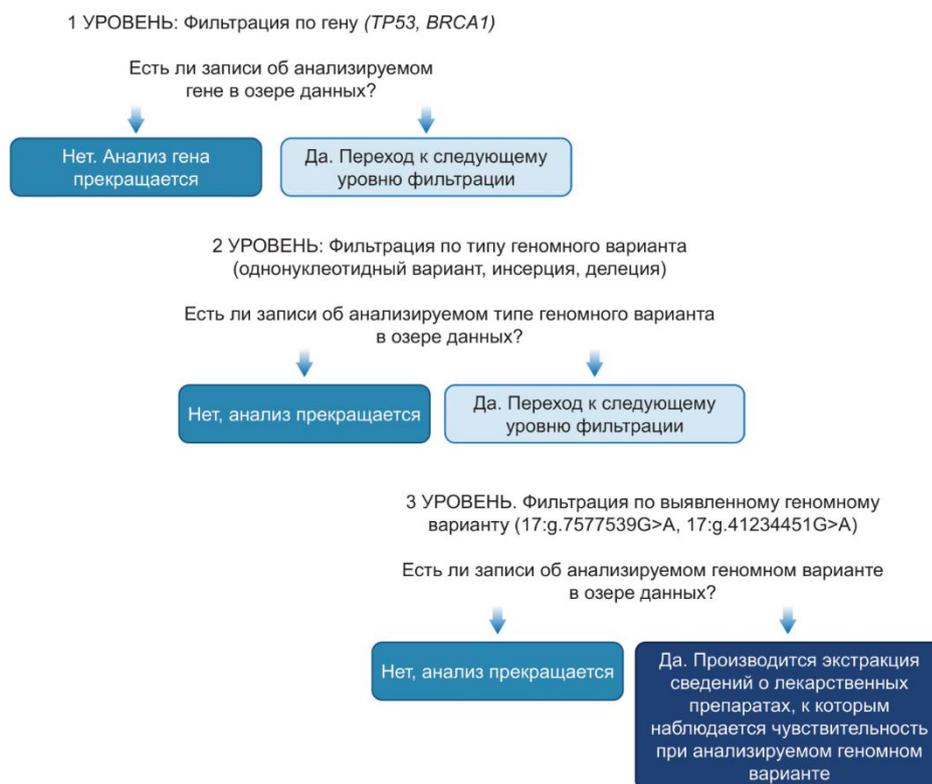


Рисунок 9. Алгоритм фильтрации, применяемый для подбора эффективного персонализированного лекарственного лечения на основе геномных вариантов, выявляемых у больных ЗНО

Для подбора персонализированного лекарственного лечения на основе показателей ТМВ СППВР анализировала информацию в озере данных на предмет наличия сведений о чувствительности к различным лекарственным препаратам при уровне ТМВ равном или превышающем показатель, выявленный в анализируемом опухолевом материале.

В рамках модуля формирования и визуализации отчета проводили сбор данных в единый JSON-файл и осуществляли генерацию динамического отчета. В процессе формирования аннотации и клинической интерпретации представленных геномных вариантов, запускалась программа JSON extractor, которая собирает всю необходимую для генерации отчета информацию в общий текстовый файл в формате JSON. Таким образом данные, получаемые

различными программами с использованием различных баз в разной структуре, приводили к единому (стандартизированному) виду. Этот файл использовали на следующем этапе для генерации динамического отчета в формате PDF для медицинского специалиста.

Для создания персонализированного динамического отчета в формате PDF использовали файл в формате JSON, полученный на предыдущем этапе. На основании информации из данного файла выстраивалось дерево элементов, которые распределялись по страницам. Далее полученное дерево преобразовывалось в одностраничный файл в формате HTML, после чего запускался пакет chromium в headless режиме, что позволяло вычислить размеры всех элементов (блоков) отчета. Исходя из размеров каждого элемента, элементы распределялись по страницам и генерировался новый HTML с зафиксированными позициями элементов, которые затем преобразовывались в отчет формата PDF за счет еще одного запуска пакета chromium. Подобная схема генерации отчета позволяет распределять данные по страницам вне зависимости от количества данных, представленных в отчете.

Верстку макета отчета осуществляли с помощью языка разметки - HTML, а генерацию PDF - с помощью программного пакета Chromium <https://chromium.googlesource.com/chromium/src/>. Динамический программный механизм был разработан с использованием кросс-платформенной среды выполнения JavaScript — Node.js.

В создаваемых отчетах о подборе персонализированного лекарственного лечения также приводили информацию о рейтинге рекомендаций. Для формирования рейтинга принимали в учет такие факторы как уровень доказательности для препарата (уровень А - утвержденная рекомендация, уровень В - клиническое доказательство, уровень С - клинический случай, уровень D - доклиническое доказательство, уровень Е - статистический анализ). Также учитывали данные о наличии регистрации/одобрения препарата в Российской Федерации, ЕС, США для лечения ЗНО, выявленного у больного (“он-лейбл”)/исследования для данного ЗНО, либо для лечения иных ЗНО (“офф-лейбл”)/исследования для других ЗНО. Самый высокий рейтинг присваивали рекомендациям препаратов, обладающим уровнем доказательности А - утвержденная рекомендация, и зарегистрированных/одобренных для лечения ЗНО, выявленного у больного (“он-лейбл”). Наименьший рейтинг присваивали препаратам с уровнями доказательства D - доклиническое доказательство/Е - статистический анализ и изучаемых для ЗНО, отличных от выявленного у больного.

2.4. Общая характеристика анализируемого биологического материала и когорт больных раком поджелудочной железы и трижды негативным раком молочной железы, биологический материал которых подвергался анализу

2.4.1. Критерии включения и исключения биологического материала больных раком поджелудочной железы и трижды негативным раком молочной железы

Проведено ретроспективное исследование биологического материала из двух когорт: больных РПЖ и больных ТНРМЖ.

При обосновании размера необходимых выборок был использован опыт проведения молекулярно-биологических исследований. Распределение параметров больных ТНРМЖ и РПЖ принимали как нормальное. Уровень доверия определен как 99,999% (10^{-5}), что соответствует Z-значению $\sim 4,42$ для нормального распределения. Оценку размера необходимых для проведения исследования выборок пациентов с учетом ранее опубликованных данных по молекулярным характеристикам опухолей определяли по формуле $n=Z^2 \cdot p \cdot (1-p)/E^2$, где n - необходимый размер выборки, Z - Z-значение для выбранного уровня доверия, p - оценка дисперсии, E - допустимая погрешность. По результатам расчетов получены следующие граничные значения для требуемых размеров выборок: $n=93$ (группа больных ТНРМЖ) и $n=39$ (группа больных РПЖ). Таким образом, в исследование должны быть включены размеры когорт пациентов, превышающие данные граничные значения. Для выбранного количества участников исследования ($n=100$ для группы больных ТНРМЖ и $n=40$ для группы больных РПЖ) выполненный расчет дал следующие значения уровней статистической мощности: 0,93 (для группы больных ТНРМЖ) и 0,89 (для группы больных РПЖ), что является приемлемым уровнем статистической мощности для клинических исследований.

В исследовании был проанализирован биологический материал больных, проходивших лечение в ФГБУ «НМИЦ Радиологии» Минздрава России в период с 2016 по 2023 г. Для проведения настоящей работы было получено одобрение Независимого Совета по Этике при МНИОИ им. П.А. Герцена - филиале ФГБУ «НМИЦ Радиологии» Минздрава России (выписка из протокола № 810а от 15.04.2022 г.).

В анализ был включен биологический материал, соответствующий следующим характеристикам: возраст больного старше 18 лет, патоморфологически верифицированный РПЖ либо ТНРМЖ, наличие установленной стадии заболевания по TNM классификации, достаточное количество опухолевого материала для проведения исследования, отсутствие синхронных злокачественных новообразований.

Итогово в исследование был включен биологический материал от 140 больных РПЖ и ТНРМЖ, а также их клинические данные (Рисунок 10).

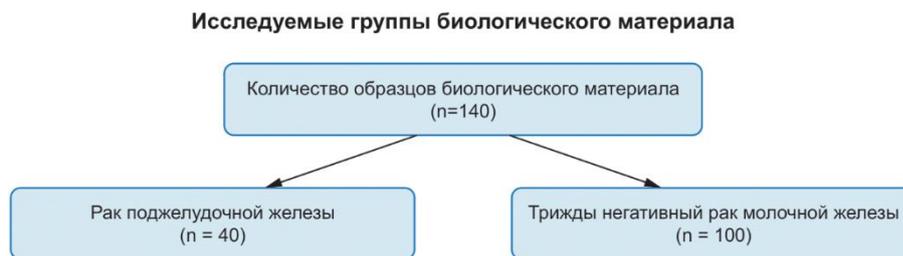


Рисунок 10. Распределение биологического материала, исследуемого в настоящей работе

2.4.2. Характеристика когорты больных раком поджелудочной железы, биологический материал которых подвергался анализу

В исследование были включены парные образцы опухолевой ткани и крови, полученные от 40 больных с патоморфологически верифицированным РПЖ. Диагноз РПЖ устанавливался на основе данных осмотра, анамнеза, результатов лабораторных и инструментальных исследований, включавших общий анализ крови и мочи, биохимический анализ крови, коагулограмму, ЭКГ, КТ органов брюшной полости и малого таза с внутривенным контрастированием, эзофагогастродуоденоскопию, ультразвуковое сканирование вен нижних конечностей, исследование онкомаркера СА 19-9. Для диагностики сопутствующих заболеваний все пациенты были осмотрены терапевтом. По показаниям для уточнения степени распространенности онкологического процесса выполнялась МРТ органов брюшной полости и малого таза с внутривенным контрастированием и ПЭТ/КТ с 18ФДГ. Для постановки диагноза проводилось патоморфологическое исследование. Стадирование осуществлялось согласно классификации TNM (American Joint Committee on Cancer, AJCC - Американская объединенная комиссия по злокачественным новообразованиям, 8-е издание от 2017 года).

Образцы опухолевой ткани были получены в ходе планового хирургического вмешательства и трепанобиопсии под контролем эндо-УЗИ. Перед принятием решения о включении биологического материала в настоящее исследование было осуществлено дополнительное гистологическое исследование для подтверждения наличия опухоли во фрагменте, отобранном для исследования, а также ИГХ исследование для верификации и определения гистологического типа опухоли, дифференциальной диагностики при метастазах РПЖ в печень и определения чувствительности к таргетной терапии анти-HER2/neu.

Образцы биологического материала РПЖ были получены от 26 мужчин и 14 женщин в возрасте от 40 до 79 лет, медиана возраста составила 62,5 года, среднее значение - $61,6 \pm 9,4$. У 19 больных (47,5%) опухоль была обнаружена в головке поджелудочной железы, у 8 больных (20%) - в хвосте, у 5 больных (12,5%) опухоль распространялась на тело и хвост поджелудочной железы, у 8 больных (20%) - только на тело поджелудочной железы.

Заболевание стадировали в соответствии с классификацией TNM. Стадия IB была выявлена у 1 больного (2,5%) T2N0M0; IIA – у 2 (5%) T3N0M0; IIB – у 4 (10%) при этом T2N1M0 - у 1 больного, T3N1M0 - у 3 больных; III – у 8 (20%) при этом T2N2M0 - у 3 больных, T3N2M0 - 1, T4N0M0 - 2, T4N1M0 - 1, T4N2M0 - 1; IV – у 25 больных (62,5%) при этом T2N0M1 - 2, T2N1M1 - 2, T3N0M1 - 5, T3N1M1 - 8, T4N1M1 - 7, T4N2M1 - 1.

Больные получали химиотерапию в различных режимах: 28 больных (70%) получали лечение по схеме FOLFIRINOX, 6 (15%) - Гемцитабин+Абраксан, 3 (7,5%) - Гемцитабин, 2 (5%) - Капецитабин, 1 (2,5%) - по схеме GEMOX. Общие и клинические характеристики больных РПЖ, биологический материал которых исследовался в настоящей работе, представлены в таблице 6.

Таблица 6. Общие и клинические характеристики больных РПЖ, биологический материал которых исследовали в настоящей работе

Характеристика	<i>n</i> = 40 больных
Возраст, годы	
Средний	$61,6 \pm 9,4$
Медиана	62,5
Минимальный	40
Максимальный	79
Пол	
Женщины	14 (35%)
Мужчины	26 (65%)
Локализация опухоли	
Головка	19 (47,5%)
Тело	8 (20%)
Хвост	8 (20%)
Тело+хвост	5 (12,5%)
Стадия	
IB	1 (2,5%)
IIA	2 (5%)
IIB	4 (10%)
III	8 (20%)
IV	25 (62,5%)
Проведенное химиотерапевтическое лечение	
FOLFIRINOX	28 (70%)
Гемцитабин+Абраксан	6 (15%)
Гемцитабин	3 (7,5%)
Капецитабин	2 (5%)
GEMOX	1 (2,5%)

2.4.3. Характеристика когорты больных трижды негативным раком молочной железы, биологический материал которых подвергался анализу

В исследование были включены парные образцы опухолевой ткани и крови, полученные от 100 больных с патоморфологически верифицированным ТНРМЖ. В рамках диагностики РМЖ выполняли осмотр, сбор анамнеза, лабораторные и инструментальные исследования, включавшие общий анализ крови и мочи, биохимический анализ крови, коагулограмму, пальпацию молочных желез и регионарных зон (над-, подключичные, подмышечные лимфатические узлы), ЭКГ, ЭХО-КГ, УЗИ вен нижних конечностей, маммографию, МРТ молочных желез с внутривенным контрастированием (для пациентов младше 40 лет и мультицентричной форме заболевания), УЗИ молочных желез и регионарных зон (на данном этапе выполнялась трепан-биопсия опухоли молочной железы, также при выявлении измененных лимфатических узлов выполнялась тонкоигольная аспирационная биопсия или трепан-биопсия), рентгенографию органов грудной клетки или КТ органов грудной полости (последнее в приоритете у пациентов с трижды-негативным фенотипом, а также при местно-распространенном варианте РМЖ), УЗИ органов брюшной полости, забрюшинного пространства и малого таза или КТ/МРТ органов брюшной полости и малого таза с внутривенным контрастированием, радиоизотопное исследование костей скелета. МРТ головного мозга с в/в контрастированием выполнялось по показаниям (при наличии общемозговой симптоматики, при высокой стадии процесса).

Для постановки диагноза биопсийный материал подвергали гистологическому и ИГХ исследованию. Диагноз ТНРМЖ был выставлен на основе отрицательной экспрессии РЭ, РП и эпидермального фактора роста 2 типа при ИГХ исследовании. Стадирование осуществлялось согласно классификации TNM (AJCC, 8-е издание от 2017 года). Образцы опухолевой ткани, использованные в рамках настоящего исследования, были получены в ходе планового хирургического вмешательства или диагностической core-биопсии. Перед принятием решения о включении биологического материала в настоящее исследование было осуществлено дополнительное гистологическое исследование для подтверждения наличия опухоли во фрагменте, отобранном для исследования, а также ИГХ исследование подтверждения трижды негативного подтипа РМЖ и оценки митотической активности раковых клеток. В когорте ТНРМЖ 100 больных (100%) являлись женщинами в возрасте от 21 до 72 лет, медиана возраста при выявлении заболевания составила 49 лет, среднее значение - 49,9 лет \pm 12,26

У 19 больных ТНРМЖ (19%) были сохранены менструации, 37 больных (37%) были в состоянии менопаузы, 7 больных (7%) - в состоянии медикаментозной менопаузы, у 3 больных менструации отсутствовали (3%), так как была выполнена надвлагалищная ампутация матки; у

34 больных (34%) данные о менструальном статусе отсутствовали. Заболевание стадировали в соответствии с TNM классификацией, при этом распределение больных было следующим: стадия IA у 21 больной (стадия была определена как T1N0M0 - 21), стадия IIA - у 40 больных (T1N1M0 - 8, T2N0M0 - 32), стадия IIB - у 13 больных (T2N1M0 - 10, T3N0M0 - 3), стадия IIIA - у 5 больных (T2N2M0 - 4, T3N1M0 - 1), стадия IIIB - у 7 больных (T4N0M0 - 4, T4N1M0 - 1, T4N2M0 - 2), стадия IIIC - у 6 больных (T2N3M0 - 2, T3N3M0 - 1, T4N3M0 - 3), стадия IV - у 8 больных (T2N0M1 - 1, T2N2M1 - 1, T4N1M1 - 2, T4N2M1 - 1, T4N3M1 - 3).

Больным ТНРМЖ проводилось неoadъювантное/адъювантное химиотерапевтическое лечение по следующим схемам (сгруппированы): 62 больные получали лечение по схеме 4АС+4D/4P/12P, 4ТС, 19 больных по схеме 4АС+4P(с Карбоплатином)/12P(с Карбоплатином), для 19 больных - иное. Лучевая терапия была проведена 51 больной. Общие и клинические характеристики больных ТНРМЖ, биологический материал которых использовали в настоящей работе, представлены в таблице 7.

Таблица 7. Общие и клинические характеристики больных ТНРМЖ, биологический материал которых исследовали в настоящей работе (N=100).

Характеристика	N = 100 больных
Возраст, годы Средний Медиана Минимальный Максимальный	49,9 ± 12,3 49 21 72
Пол Женщины Мужчины	100 (100%) 0
Менструальный статус менструации сохранены менопауза медикаментозная менопауза отсутствуют (надвлагалищная ампутация матки) нет данных	19 (19%) 37 (37%) 7 (7%) 3 (3%) 34 (34%)
Стадия IA IIA IIB IIIA IIIB IIIC IV	21 (21%) 40 (40%) 13 (13%) 5 (5%) 7 (7%) 6 (6%) 8 (8%)
Проведенное химиотерапевтическое лечение 4АС+4D/4P/12P, 4ТС 4АС+4P(с Карбоплатином)/12P(с Карбоплатином) Иное	62 (62%) 19 (19%) 19 (19%)
Лучевая терапия	51 (51%)

2.5. Морфологическое и иммуногистохимическое исследование образцов опухолевой ткани

На парафиновых блоках всех образцов операционного или биопсийного материала РПЖ и ТНРМЖ было проведено гистологическое и ИГХ исследование. Фрагменты исследуемых ЗНО фиксировали в 10% нейтральном растворе формалина с фосфатным буфером (около 72 часов), дегидратировали в батарее спиртов восходящей концентрации в аппарате гистологической проводки тканей фирмы «Pool Scientific Instruments» (Швейцария), заливали в парафин. После этого производили нарезку (полуавтоматический микротом «Galileo», DiaPath, Италия) парафиновых блоков (не менее 5 срезов с каждого блока), срезы толщиной 2 мкм с каждого блока помещали на обычные стекла или на специальные. Срезы окрашивали гематоксилином Майера и эозином или подготавливали для ИГХ исследований на специальных адгезивных полилизинных стеклах «Super Frost Plus» («Mainzel Glaser», Polylysine, Германия). Для окрашивания гематоксилином и эозином в исследовании использовалась автоматическая система гистологической окраски с функцией заключения препарата под покровное стекло Dako Cover Stainer (Agilent, США) согласно инструкции производителя.

ИГХ исследование образцов РПЖ и ТНРМЖ было проведено на срезах с парафиновых блоков в иммуностейнере Ventana BenchMark XT (Ventana, Ventana Medical Systems, США) с применением системы детекции и визуализации ultraView Universal DAB Detection Kit (Ventana Medical Systems, США) и последующим докрасиванием гематоксилином согласно инструкции производителя.

Для ИГХ подтверждения ТНРМЖ использовали первичные кроличьи моноклональные антитела к РЭ, РП, HER2/neu, а также для оценки митотической активности раковых клеток – Ki-67 (Таблица 8).

Таблица 8. Антитела, используемые для иммуногистохимического исследования образцов ТНРМЖ

Антитела	Клон	Тип антитела	Происхождение	Производитель
Anti-Estrogen receptor (ER)	SP1	моноклональное	кролик	Ventana, Roche (США)
Anti-Progesterone Receptor (PR)	1E2	моноклональное	кролик	Ventana, Roche (США)
Anti-Human epidermal growth factor receptor (HER2/neu)	4B5	моноклональное	кролик	Ventana, Roche (США)
Anti-Ki-67	30-9	моноклональное	кролик	Ventana, Roche (США)

При ИГХ исследовании образцов РПЖ для его верификации и определения гистологического типа использовали маркеры принадлежности к опухолям эпителиального происхождения – CK7, ЕМА, СА125, СЕА. При исследовании метастатических узлов РПЖ в печени для дифференциальной диагностики использовали маркеры гепатоцитов Нер Par-1. Исследование экспрессии HER2/neu проводили с целью определения чувствительности опухоли к таргетной терапии анти-HER2/neu (Таблица 9).

Таблица 9. Антитела, используемые для иммуногистохимического исследования образцов РПЖ

Антитела	Класс маркеров	Клон	Тип антитела	Происхождение	Производитель
Anti-Cytokeratin 7 antibody (CK7)	маркеры эпителиальных клеток, элементов цитоскелета	EP16 / EP25	Моноклональный	Кролик	Epitomics Ink (США)
Anti-Epithelial membrane antigen (ЕМА)		E29	Моноклональный	Мышь	Cell Marque (США)
Anti-Carbohydrate Antigen 125 (CA125)	Маркеры эпителиальных/эмбриональных атипичных клеток	X75, X325	Моноклональный	Мышь	Dako (США)
Anti-carcinoembryonic antigen (CEA)		E29	Моноклональный	Мышь	Dako (США)
Anti-Human epidermal growth factor receptor (HER2/neu)	Маркер тирозинкиназных рецепторов	4B5	Моноклональный	Кролик	Roche Diagnostics (Швейцария)
Anti- Hepatocyte paraffin 1 (Нер Par-1)	Маркер гепатоцитов	EP265	Моноклональный	Кролик	Sigma Aldrich (Германия)

Для концентратов антител разведения были подобраны на основе окрашивания контрольных срезов; использовали готовые к применению антитела. Окрашенные срезы были использованы для изготовления препаратов по стандартной методике.

Микроскопическое исследование образцов ТНРМЖ проводили на микроскопе Zeiss Axio Imager.M2 (Carl Zeiss, Германия), оборудованном цветной фотокамерой AxioCam 506.

Микроскопическое исследование готовых препаратов РПЖ проводили на микроскопе Axio lab.a1 (Carl Zeiss, Германия). Для оцифровки изображений использовали сканер Panoramic (3DHISTECH, Венгрия), для анализа изображений – программное обеспечение для ЭВМ Case Viewer (3DHISTECH, Венгрия).

Подсчет количества иммунопозитивных клеток проводили в 10 случайно отобранных полях зрения при увеличении $\times 400$ (в %).

ГЛАВА 3 - Результаты исследования

3.1. Мета-анализ клинических исследований персонализированного лечения онкологических заболеваний в сравнении с неперсонализированным лечением

3.1.1. Исследования, включенные в мета-анализ

По итогам поиска в базе данных PubMed было выявлено 803 публикации. 451 публикация не соответствовала критериям включения на основании оценки названия и аннотации. Далее, в результате полнотекстового анализа работ 296 исследований были полностью исключены, а 7 исследований были исключены частично: 18 неподходящих групп клинических исследований были удалены из окончательного набора данных. Таким образом, мы выбрали 49 исследований, включающих 101 группу, общее количество пациентов составило 7012. Результаты поиска и отбора исследований представлены в диаграмме Prisma flow chart (Рисунок 11).



Рисунок 11. Диаграмма отбора исследований Prisma Flow Chart. Аббревиатуры: ВБП - выживаемость без прогрессирования, ЗНО - злокачественное новообразование, ОБ - общая выживаемость, ЧОО - частота объективного ответа

В 43 группах (42,6%), включающих 2440 пациентов (34,8%), лечение было персонализированным на основе геномного подхода, в 58 группах (57,4%) включавших 4572 пациентов, лечение было неперсонализированным (65,2%). Распределение параметров представлено в таблице 10.

Таблица 10. Характеристики исследований, включенных в мета-анализ

Параметры	Группы	Группа, Abs	Группа, %
Рандомизация		64	63.8
Одноцентровое/многоцентровое		84/17	82.9/17.1
Предшествовавшее лекарственное лечение		45	44.7
Таргетная/иммунотерапия		19	19.1
Химиотерапия		69	68.1
Количество пациентов в группе > 40		75	74.5
5-летний импакт фактор журнала > 10		30	29.8

Тест Фишера. $p < 0,05$

3.1.2. Сравнение исходов лечения онкологических заболеваний

Было проведено сравнение результатов персонализированного лекарственного лечения онкологических заболеваний, основанного на индивидуальных геномных характеристиках пациентов, и неперсонализированного лекарственного лечения. ЧОО и однолетняя ВБП были значительно выше в группах, получавших персонализированное лечение, $p=0,009$ и $p=0,011$ (Таблица 11). Показатели однолетней ОВ значимо не отличались между группами персонализированного и неперсонализированного лечения.

Таблица 11. Анализ результатов лечения / исходов в группах

Параметры		ЧОО, %	1-летняя ВБП, %	1-летняя ОВ, %	ВБП, медиана, мес	ОВ, медиана, мес
Группы и статистика						
Персонализированное лечение	Me	30.1%	41.3%	100.0%	26.5913	9,9
	Q1	17.0 %	30.2%	48.8%	11.925	7,25
	Q3	45.6 %	56.7%	100.0%	41.04	13,6
Неперсонализированное лечение	Me	12.1%	27.9%	69.4%	16.65	6,7
	Q1	9.4%	19.0%	39.3%	10.425	4,8
	Q3	17.3%	38.8%	89.2%	21.15	9,685
Критерий Манна-Уитни, персонализированное лечение/неперсонализированное, p		0,009 *	0.011 *	0.058	0.018*	0.032*
Рандомизированное исследование	Me	43.0%	32.5%	80.4%	19.15	7,8
	Q1	28.9%	27.9%	50.8%	12.5	6,69
	Q3	45.6%	44.4%	100.0%	27.15	10,65

Параметры		ЧОО, %	1-летняя ВБП, %	1-летняя ОВ, %	ВБП, медиана, мес	ОВ, медиана, мес
Группы и статистика						
Нерандомизированное исследование	Me	43.0%	34.2%	72.9%	17.5	8,35
	Q1	28.9%	24.6%	44.3%	10.62	6,15
	Q3	45.6%	49.4%	100.0%	33.54	11,5725
Критерий Манна-Уитни, рандомизированное исследование/нерандомизированное исследование, р		0,746	0.887	0.983	0.817	0.889
Одноцентровое исследование	Me	42.3%	34.2%	80.4%	19.3	8
	Q1	29.7%	27.8%	52.5%	12.8	6,685
	Q3	51.2%	47.5%	100.0%	30.75	11,35
Многоцентровое исследование	Me	34.0%	27.5%	48.1%	11.55	6,9
	Q1	30.0%	22.5%	42.5%	10.2	5,5
	Q3	38.8%	33.8%	97.3%	26.125	9,6575
Критерий Манна-Уитни, одноцентровое исследование/многоцентровое исследование, р		0.279	0.207	0.401	0.330	0,502
Наличие предшествовавшего лекарственного лечения	Me	31.0%	34.2%	62.9%	15.1	8,6
	Q1	26.6%	17.9%	33.6%	8.075	4,55
	Q3	39.4%	41.7%	100.0%	27.625	10,225
Без предшествовавшего лекарственного лечения	Me	44.3%	31.7%	84.1%	19.788	7,65
	Q1	35.5%	27.9%	57.3%	13.925	6,695
	Q3	51.2%	79.2%	100.0%	38.3	18,75
Критерий Манна-Уитни, наличие предшествующего лечения/без предшествующего лечения, р		0,037 *	0.274	0.077	0.360	0.059
Таргетная/Иммунотерапия	Me	39.2%	38.9%	98.5%	25.35	9,335
	Q1	30.0%	29.5%	47.8%	11.46	7,075
	Q3	44.6%	55.9%	100.0%	37.2	13,425
Химиотерапия	Me	38.0%	32.5%	75.6%	18.572	7,95
	Q1	28.9%	22.0%	42.5%	11.085	5,375
	Q3	68.0%	49.4%	100.0%	28.95	11,5725
Критерий Манна-Уитни, таргетная/иммунотерапия в сравнении с химиотерапией, р		0,520	0.532	0.261	0.670	0.301
Число пациентов в группе > 40	Me	37.0%	30.6%	73.7%	17.232	7,4
	Q1	29.8%	24.1%	46.1%	11.07	5,95
	Q3	44.9%	42.0%	100.0%	25.4913	10,05
Число пациентов в группе ≤ 40	Me	43.0%	35.8%	80.8%	19.392	9
	Q1	31.0%	29.4%	54.4%	13.05	7,1
	Q3	50.0%	96.4%	100.0%	38.6	26,75
Критерий Манна-Уитни, число пациентов в группе, р		0,655	0.123	0.420	0.096	0.306
5-летний импакт-фактор журнала > 10	Me	34.0%	34.2%	71.8%	17.232	8,2
	Q1	25.3%	16.9%	41.0%	9.84	4,05
	Q3	46.8%	41.3%	100.0%	25.4826	9,91
5-летний импакт-фактор журнала ≤ 10	Me	43.0%	31.7%	80.4%	19.15	7,7
	Q1	32.0%	27.8%	47.5%	11.55	6,69
	Q3	45.6%	55.0%	100.0%	33.945	12,415
Критерий Манна-Уитни, 5-летний импакт-фактор журнала, р		0,235	0.363	0.321	0.313	0.330

Примечание: * p<0.05

Для оценки варибельности исследований оценивали статистики гетерогенности между исследованиями, которая была довольно высокой для ЧОО, ВБП и ОВ (Таблица 12). При анализе статистик τ^2 и Хиггинса-Томсона были получены следующие результаты: в случае ЧОО $\tau^2 = 0,0283$ [0,0226; 0,0656] и $I^2 = 88,7\%$ [85,6%; 91,1%], в случае ВБП $\tau^2 = 0,0906$ [0,0736; 0,1758] и $I^2 = 96,7\%$ [96,2%; 97,1%], в случае ОВ $\tau^2 = 0,1503$ [0,1006; 0,2195] и $I^2 = 97,8\%$ [97,6%; 98,1%]. Высокие показатели гетерогенности могут быть связаны с высокой дисперсией абсолютных значений параметров, характеризующих результаты лечения, при различных онкологических заболеваниях.

Таблица 12. Гетерогенность между исследованиями, включенными в мета-анализ

Исходы, изучаемые в мета-анализе	ЧОО	ВБП	ОВ
Статистики гетерогенности			
τ^2 -статистика [95% ДИ]	0.0283 [0.0226; 0.0656]	0.0906 [0.0736; 0.1758]	0.1503 [0.1006; 0.2195]
I^2 -статистика [95% ДИ]	88.7% [85.6%; 91.1%]	96.7% [96.2%; 97.1%]	97.8% [97.6%; 98.1%]
z-тест для гетерогенности, значение p	<0.0001	<0.0001	<0.00001

Пропорциональный мета-анализ и мета-регрессия выявили значительную связь между персонализированным лекарственным лечением на основе индивидуальных геномных характеристик больных и ЧОО. Показатели ЧОО были значительно выше в группах персонализированного лечения (рисунок 12А, таблица 11), $p=0,043$. Аналогичные результаты были получены для показателей однолетней ВБП и однолетней ОВ (рисунок 12Б, 12В): $p=0,037$ и $0,022$. Значительно худшие результаты ВБП были обнаружены у пациентов, ранее получавших лечение, например, получающих вторую и более позднюю линию лечения, $p=0,027$, однако эти результаты могут быть связаны с более поздними стадиями заболевания.

Более низкая ВБП вероятно была связана с меньшим количеством пациентов в группе, и, как следствие, с высоким уровнем гетерогенности и различиями в подходах к отбору пациентов в разных исследованиях [309]. Однако подобные результаты не были получены при изучении ОВ и ЧОО.

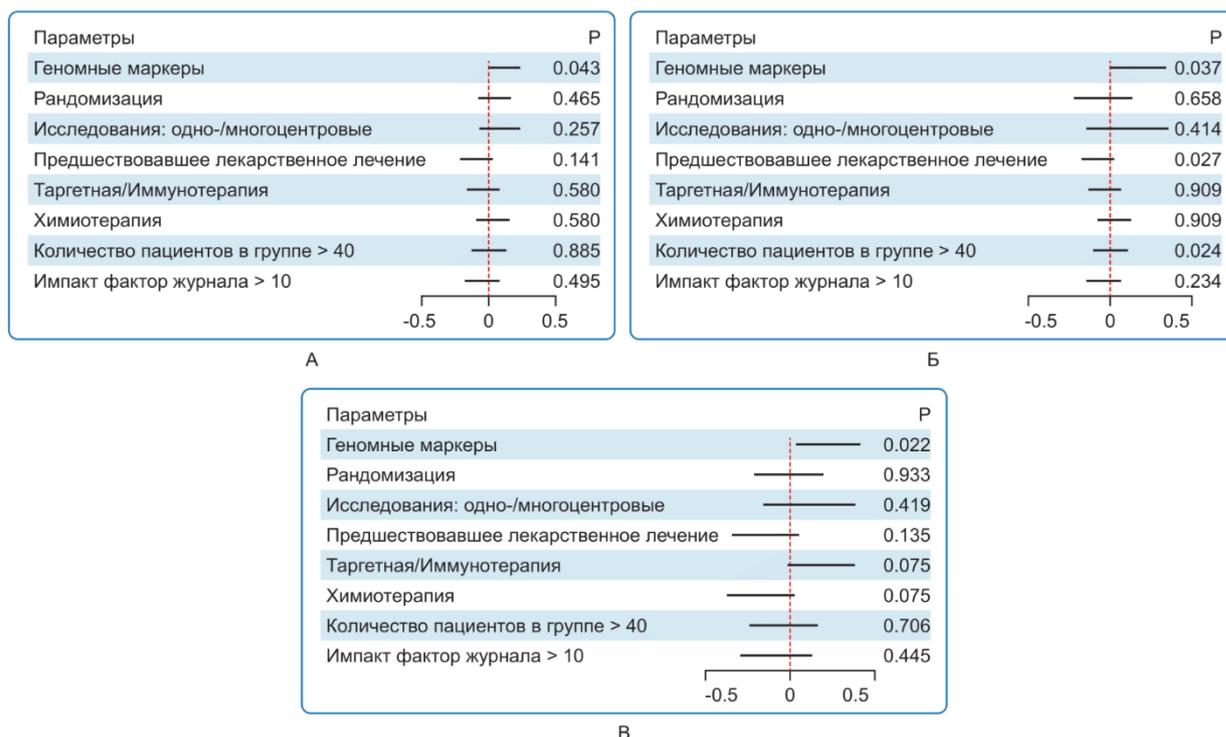


Рисунок 12. Результаты пропорционального мета-анализа и мета-регрессии. Мета-регрессия проводилась между качественными параметрами и размером эффекта исходов. Оценки регрессии и ДИ представлены на лесовидных диаграммах. Значимость коэффициентов регрессии представлена в последних столбцах, $p < 0,05$, z-тест А) ЧОО; Б) 1-летняя ВБП; В) 1-летняя ОВ

Для значимых параметров ВБП была построена модель множественной регрессии. В результате основным прогностическим фактором ВБП стало предшествующее лечение (Таблица 13).

Таблица 13. Множественная мета-регрессия для 1-летней ВБП

Параметры	Статистика	Коэффициент регрессии	р
Предшествовавшее лекарственное лечение		-0.1631	0.048*
Количество пациентов в группе > 40		-0.1567	0.074
Геномные маркеры		0.1235	0.135

Примечание. Были включены только параметры, значимые при одновариантном анализе

Аналогичные результаты были получены при анализе медиан ВБП и ОВ. Мы оценили разницу медиан между группами, получающими персонализированное и неперсонализированное лечение. Медиана ВБП была значительно выше в исследованиях с персонализацией лечения, $p=0,015$, как и медиана ОВ, $p=0,032$. Мета-регрессия является более мощным методом, однако его нельзя использовать для мета-анализа только медиан без дополнительной информации или контрольной группы в исследованиях.

По результатам подгруппового анализа между персонализированными и

неперсонализированными исследованиями не было отличий в изучаемых параметрах (Таблица 14). Таким образом, учёт индивидуальных геномных характеристик больных предложен в качестве основного фактора, улучшающего ЧОО, ОВ и ВБП при медикаментозном лечении онкологических заболеваний по данным исследований II фазы. Дизайн исследования (применение рандомизации и одноцентровой/многоцентровой протокол), предшествующее лечение, количество пациентов в группе и импакт-фактор журнала не влияли на показатели эффективности лечения. Вид лечения - таргетное или иммунотерапевтическое и химиотерапевтическое - также не влияли на отдаленные результаты.

Таблица 14. Сравнение групп персонализированного и неперсонализированного лечения, n=101

Группы	Неперсонализированное лечение. n=58 [%]	Персонализированное лечение n=43 [%]	p-значение*
Параметры			
Рандомизация	47.2	63.8	0.368
Одноцентровое	83.3	82.9	0.521
Многоцентровое	16.7	17.1	
Предшествовавшее лекарственное лечение	38.9	44.7	0.292
Таргетная / иммунотерапия	34.1	37.1	0.768
Химиотерапия	65.9	62.9	
Количество пациентов в группе > 40	50.0	74.5	0.585
5-летний импакт-фактор журнала > 10	25.0	29.8	0.326

Примечание: * - Точный тест Фишера, $p < 0,05$

3.2. Адаптированный протокол пробоподготовки и секвенирования

3.2.1 Описание адаптированного протокола пробоподготовки и секвенирования

В рамках настоящей работы был разработан адаптированный протокол, содержащий последовательность проведения всех этапов пробоподготовки и секвенирования с использованием выбранных в процессе работы коммерчески доступных наборов. Важным результатом в рамках разработанного адаптированного протокола является модификация существующего протокола обогащения экзомных библиотек с применением реактивов Roche (США) и Agilent (США) для его использования при работе на секвенаторе MGISEQ-G400 (MGI, Китай). Разработанный адаптированный протокол был применен для молекулярно-генетического анализа биологического материала больных РПЖ и ТНРМЖ [51, 315].

При разработке адаптированного протокола проводили сравнение различных коммерчески доступных наборов для выделения ДНК из образцов периферической крови и опухолевого материала. По итогам проведенного сравнения были получены сравнительные

данные по выходу ДНК (из расчета на 100 мкл образца для периферической крови, и из расчета на 3 среза толщиной 10 мкм - для образцов опухолевой ткани), а также данные о соотношении поглощения при 260/280 нм и при 260/230 нм для обоих типов биологического материала при применении различных коммерчески доступных наборов для выделения ДНК. Результаты проведенного сравнительного исследования представлены в таблицах 15 и 16.

Таблица 15. Результаты сравнения количества и чистоты ДНК, полученной из образцов цельной венозной крови с применением различных коммерческих наборов (приведены данные по 10 образцам, полученным с применением каждого из наборов)

	QIAamp DNA Blood Mini Kit	MagPure Blood DNA kit	Проба-ГС-генетика
Выход ДНК (из расчета на 100 мкл образца)	2100-3500 нг	700-2000 нг	1500-5000 нг
Соотношение поглощения при 260/280 нм	1,8-2,0	1,8-2,0	1,8-2,0
Соотношение поглощения при 260/230 нм	1,8-2,0	1,6-1,9	1,5-1,9

Таблица 16. Результаты сравнения количества и чистоты ДНК, полученной из образцов опухолевой ткани, заключенной в парафин, с применением различных коммерческих наборов (приведены данные по 10 образцам, полученным с применением каждого из наборов)

	QIAamp DNA FFPE Tissue kit	AVENIO Tumor DNA Isolation Kit	Cobas DNA Sample Preparation Kit
Выход ДНК (из расчета на 3 среза толщиной 10 мкм)	1000-3500 нг	200-800 нг	1000-3000 г
Соотношение поглощения при 260/280 нм	1,8-2,0	1,8-2,0	1,8-2,0
Соотношение поглощения при 260/230 нм	1,8-2,0	1,8-2,0	1,5-1,8

С учетом полученных результатов, как для крови, так и для опухолевой ткани, было принято решение о выборе наборов Qiagen (QIAGEN, США), позволяющих получить из образца максимальное количество ДНК с минимальным количеством примесей, на что указывают высокие значения поглощения при 260/280 нм и 260/230 нм.

Этап обогащения экзомных библиотек был модифицирован с целью адаптации реагентов Roche (США) и Agilent (США), предназначенных для секвенирования на платформе Illumina, к экзому библиотек, приготовленным по протоколу MGI (Китай). Для этого был произведен подбор и синтез новых последовательностей блокирующих олигонуклеотидов Block 1 и Block 2 с LNA-модификацией (добавлением замкнутых нуклеотидов).

Разработанный протокол пробоподготовки и секвенирования подробно представлен далее. Для выделения ДНК из лимфоцитов периферической крови, по результатам проведенного

сравнительного анализа, было решено использовать набор QIAamp DNA Blood Mini Kit (QIAGEN, США) согласно протоколу производителя [364]. Сначала лимфоциты крови обрабатывали лизирующим буфером с добавлением протеиназы К. Затем полученный лизированный образец смешивали со спиртом, наносили на мембранную спин-колонку и центрифугировали. В процессе центрифугирования ДНК связывалась с мембраной колонки. После промывания двумя буферами, очищенную геномную ДНК элюировали со спин-колонки очищенной водой.

Для выделения ДНК из опухолевого материала, залитого в парафин (FFPE блоков), было решено использовать набор QIAamp DNA FFPE Tissue Kit по протоколу производителя [300]. Для выделения ДНК брали 2-3 среза с парафинового блока толщиной 10 мкм. На первом этапе из образцов удаляли парафин с помощью ксилола. Далее образец лизировали в лизирующем буфере, содержащем протеиназу К. После чего полученный раствор инкубировали при 90°C в течение 1 часа для преодоления фиксации нуклеиновых кислот формалином. Затем раствор наносили на мембранную спин-колонку и центрифугировали. В процессе центрифугирования ДНК связывалась с мембраной колонки. После промывания буфером, очищенную геномную ДНК элюировали с колонки очищенной водой. Выбор указанных коммерческих наборов для выделения ДНК из периферической крови и опухолевого материала основан на нашем сравнении нескольких доступных коммерческих наборов для обоих типов образцов.

ДНК была выделена из нескольких срезов FFPE-блоков, содержащих опухоль (процент опухолевых клеток определяли гистологически), а также из образцов периферической крови тех же пациентов для контроля нормальной ДНК. Концентрацию выделенной ДНК измеряли с помощью флуориметра Qubit 2.0 (Life Technologies, США) с использованием набора dsDNA HS Assay Kit (Life Technologies, США) по протоколу производителя. Качество выделенной ДНК оценивали с помощью капиллярного гель-электрофореза на приборе Agilent Bioanalyzer 2100 (Agilent Technologies, США) с использованием набора High Sensitivity DNA (Agilent Technologies, США) по протоколу производителя.

Приготовление библиотек осуществляли из 100-400 нг геномной ДНК, выделенной из лимфоцитов периферической крови и FFPE блоков, с использованием набора MGIEasy Universal DNA Library Prep Set (MGI Tech, Китай). Этап фрагментации ДНК, выделенной из FFPE блоков, выполнялся с помощью протокола [84], основанного на использовании S1 нуклеазы, проводящей разрез второй цепи ДНК по местам локализации ников. Для удаления артефактов фиксации ДНК и более полной фрагментации проводили обработку смесью ферментов USER (New England Biolabs), в состав которой входит урацил-ДНК-гликозилаза и эндонуклеаза VIII, высвобождающие урацил в местах дезаминирования цитозинов с образованием одноцепочечных разрывов цепи ДНК. Измерение концентрации готовых библиотек проводили аналогично

измерению концентрации выделенной ДНК. Длину фрагментов готовых библиотек оценивали с помощью капиллярного гель-электрофореза на приборе Agilent Bioanalyzer 2100 (Agilent Technologies, США) с использованием набора High Sensitivity DNA (Agilent Technologies, США) по протоколу производителя.

Для дальнейших этапов библиотеки объединяли в пулы. Обогащение пулированных образцов проводили по методике с использованием зондов SureSelect Human All Exon V6/V7 (Agilent Technologies, США) и набора для гибридизации SureSelect Target Enrichment System (Agilent Technologies, США) с собственными модификациями, а также зондов Roche Кара HyperExome (Roche, США) и набора для гибридизации KAPA HyperCapture Reagent Kit (Roche, США) с собственными модификациями. Оба типа зондов позволяют исследовать полный экзом человека (более 22 000 генов). Измерение концентраций и длины фрагментов образцов после обогащения оценивали аналогично измерению концентраций и длин фрагментов готовых библиотек. Циркуляризацию пулированных образцов осуществляли с использованием набора MGIEasy Circularization Kit (MGI, Китай) по протоколу производителя. После этапа циркуляризации проводилось измерение концентрации пулированных образцов с помощью флуориметра Qubit 2.0 (Life Technologies, США) с использованием наборов ssDNA Assay Kit (Life Technologies, США). Приготовление наночастиц ДНК (DNB, “DNA nanoballs”) осуществляли с использованием набора DNB Preparation Kit (MGI, Китай) по протоколу производителя. Секвенирование образцов проводили в режимах PE100 и PE150 на платформе MGISEQ-G400 (MGI Tech, Китай) по протоколу производителя со средней величиной покрытия 100x для образцов крови и 200x для опухолевых образцов.

3.2.2 Сравнение адаптированного протокола пробоподготовки и секвенирования с референсным протоколом

В настоящей работе было произведено сравнение разработанного адаптированного протокола пробоподготовки и секвенирования с широко применяемыми коммерческими наборами AVENIO (Roche, США). Было осуществлено сравнение как количественных (выход ДНК, время лигирования, количество исследуемых генов), так и качественных (способ экстракции ДНК, способ обогащения библиотек, особенности секвенирования) параметров. Результаты сравнения различных этапов пробоподготовки и секвенирования представлены в таблице 17.

Таблица 17. Результаты сравнения различных этапов пробоподготовки и секвенирования, реализуемых с использованием референсного протокола AVENIO (Roche, США) и разработанного протокола

№	Параметр	Протокол AVENIO (Roche, США)	Разработанный протокол
1.	Выделение ДНК из опухолевой ткани, заключенной в парафин	AVENIO Tumor Isolation and QC Kit (Roche, США)	QIAamp DNA FFPE Tissue Kit (Qiagen, Германия)
	Результат сравнения	<ul style="list-style-type: none"> • Экстракция ДНК производится на магнитных частицах, • Выход ДНК значительно ниже (200-800 нг из трёх срезов толщиной 10 мкм) 	<ul style="list-style-type: none"> • Экстракция ДНК производится на спин-колонках, • Выход ДНК выше (1000-3500 нг из трёх срезов толщиной 10 мкм)
2.	Приготовление библиотек из выделенной ДНК	AVENIO Tumor Library Prep Kit (Roche, США).	MGIEasy Universal DNA Library Prep Set (MGI, Китай)
	Результат сравнения	Лигирование длится 16-20 часов	Лигирование длится 30 минут
3.	Обогащение библиотек	AVENIO Tumor Enrichment Kit (Roche, США)	KAPA HyperCapture Reagent Kit, Roche, США, и SureSelect Target Enrichment System, Agilent Technologies, США; использовались новые блокирующие олигонуклеотиды, разработанные нами
	Результат сравнения	Обогащение производится индивидуально для каждой библиотеки	Обогащение производится для пула библиотек, что снижает расход реактивов и экономит время работы
4.	Зонды на обогащение	Панель AVENIO Tumor Tissue Expanded Panel Kit (Roche, США)	Kapa HyperExome (Roche, США) и SureSelect Human All Exon V6/V7 (Agilent Technologies, США)
	Результат сравнения	Позволяет исследовать мутации в 77 генах, в том числе шесть транслокаций	Позволяет исследовать на наличие мутаций полный экзом (≈ 22000 генов), но не включает транслокации
5.	Секвенирование	NextSeq500 (Illumina, США)	MGISEQ-G400 (MGI, Китай)
	Результат сравнения	Перед секвенированием не требуется проведения дополнительных подготовительных этапов.	Перед секвенированием необходим дополнительный этап - циркуляризация и приготвление DNB

По сравнению с референсным протоколом, разработанный нами протокол продемонстрировал значительно более высокий выход ДНК (без значимых различий по чистоте и качеству) из одного и того же количества исходного материала. Также в нашем протоколе существенно снизилось время работы и количество затраченных реактивов на этапе обогащения за счёт пулирования библиотек перед этапом обогащения (а не после него, как в референсном протоколе AVENIO (Roche, США)).

3.3. Фармакогеномная платформа и система поддержки принятия врачебных решений для подбора персонализированного лекарственного лечения онкологических заболеваний

3.3.1. Структура разработанной фармакогеномной платформы и её взаимодействие с системой поддержки принятия врачебных решений

В рамках настоящей работы были созданы фармакогеномная платформа и СППВР, позволяющие выполнять подбор персонализированного лекарственного лечения на основе геномных вариантов и показателей ТМВ, выявляемых у больных по итогам проводимого высокопроизводительного секвенирования. СППВР не входит в состав фармакогеномной платформы, однако как фармакогеномная платформа, так и СППВР включают в себя различные элементы, взаимодействующие между собой для реализации поставленных задач.

Фармакогеномная платформа включает следующие элементы:

1. Озеро данных, аккумулирующее большие объемы различных собственных данных, в том числе, регистрационную и клинико-фенотипическую информацию о больных; лабораторные метрики; необработанные данные секвенирования; результаты выявления геномных вариантов, ТМВ, MSI, CNV и иную информацию. Также в озеро интегрирована информация из открытых мировых баз данных, необходимая для аннотирования биоинформатическими конвейерами и подбора эффективного персонализированного лекарственного лечения СППВР (наполнение созданного озера данных подробно описано в разделе 3.3.3. “Наполнение озера данных”);
2. Биоинформатические конвейеры для обработки данных полноэкзомного секвенирования и определения герминального и соматического геномного профиля, включающие инструменты на основе технологий машинного обучения. Структура разработанных биоинформатических конвейеров подробно описана в разделе 3.3.5. настоящей работы “Разработанные биоинформатические конвейеры для определения герминального и соматического геномного профиля”;
3. Комплекс инструментов для организации вычислений и API.

Разработанная СППВР включает следующие элементы:

1. Модуль определения мутационного профиля опухоли;
2. Модуль подбора персонализированного лекарственного лечения на основе геномных вариантов и ТМВ, выявленных у больных ЗНО. Для подбора персонализированного лекарственного лечения созданные алгоритмы анализируют информацию о выявленных

биомаркерах и информацию из различных открытых мировых баз данных, хранящуюся в озере данных фармакогеномной платформы. Алгоритмы подбора персонализированного лекарственного лечения позволяют формировать рекомендации таргетных, иммунотерапевтических и химиотерапевтических препаратов. Также разработанные алгоритмы позволяют формировать прогноз течения заболевания на основе выявленных геномных вариантов;

3. Модуль формирования и визуализации отчета с результатами подбора персонализированного лекарственного лечения, прогнозом течения заболевания, выявленными геномными вариантами, показателями TMB и MSI и иной информацией (более подробно структура и содержание отчетов СППВР описаны в разделе 3.3.7. “Система поддержки принятия врачебных решений для подбора персонализированного лекарственного лечения онкологических заболеваний”).

Структура комплекса “фармакогеномная платформа - СППВР”, а также пути взаимодействия структурных элементов представлены на рисунке 13.

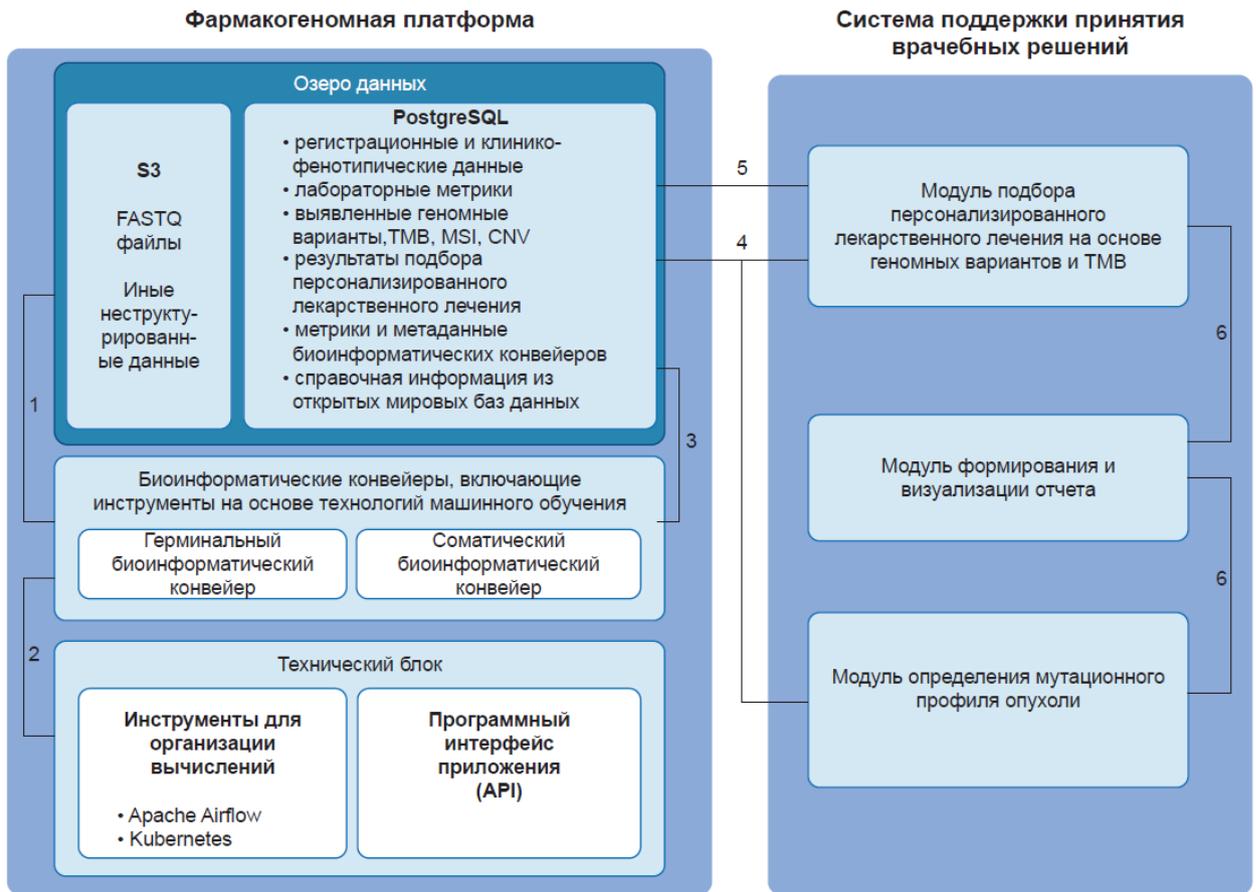


Рисунок 13. Структура комплекса “фармакогеномная платформа - система поддержки принятия врачебных решений” и пути взаимодействия между составными элементами.

Разработанный комплекс “фармакогеномная платформа - СППВР” представляет собой информационную среду, в рамках которой осуществляется взаимодействие между различными элементами. Целесообразно выделить ключевые пути взаимодействия между различными элементами фармакогеномной платформы и СППВР:

- 1) Биоинформатические конвейеры получают исходные данные секвенирования анализируемого образца в формате FASTQ из хранилища S3, находящегося в составе озера данных;
- 2) Технический блок взаимодействует с биоинформатическими конвейерами: Apache Airflow устанавливает очередность задач, Kubernetes управляет контейнеризированными рабочими нагрузками;
- 3) После обработки анализируемого образца биоинформатическими конвейерами, выявленные геномные варианты, показатели TMB, MSI, CNV сохраняются в озеро данных. Также в озере данных сохраняются метрики качества работы биоинформатических конвейеров и метаданные;

Между фармакогеномной платформой и СППВР происходят взаимодействия, описанные далее:

- 4) Модуль подбора персонализированного лекарственного лечения СППВР получает из озера данных фармакогеномной платформы информацию о выявленных геномных вариантах и показателях TMB и MSI. Модуль определения мутационного профиля опухоли СППВР также получает информацию о молекулярно-генетических характеристиках образцов из озера данных фармакогеномной платформы; Для формирования рекомендаций, модуль подбора персонализированного лекарственного лечения СППВР обращается к озеру данных, в котором содержится информация об ассоциациях биомаркеров с чувствительностью к противоопухолевым лекарственным препаратам из открытых мировых баз;
- 5) Результаты подбора персонализированного лекарственного лечения, сформированные СППВР, сохраняются в озере данных фармакогеномной платформы; Между модулями СППВР происходят следующие взаимодействия:
- 6) Инструменты внутри модуля формирования и визуализации отчета объединяют результаты работы модуля подбора персонализированного лекарственного лечения и модуля определения мутационного профиля опухоли в визуальный отчет. В отчет также добавляется информация о всех выявленных геномных вариантах и о показателях TMB и MSI.

Комплекс “фармакогеномная платформа - СППВР” расположен локально на собственных серверах, где происходит обработка. Внешний доступ имеется только у ответственных сотрудников. Модули загрузки, активации и прочее доступны через web - форму.

Таким образом, фармакогеномная платформа и СППВР представляют собой комплексную среду, объединяющую различные элементы, находящиеся во взаимосвязи между собой и реализующие задачи по аккумулярованию больших объемов данных для дальнейшего анализа и подбора эффективного персонализированного лекарственного лечения ЗНО.

3.3.2. Выбранный инструмент управления озером данных

Был проведён анализ инструментов управления озером данных Dataiku, Dremio, Knime, Delta Lake, а также стека технологий (Apache NIFI + Apache Kafka + Apache Airflow) (описание инструментов приведено в таблице 18). Проблематика выбора определялась необходимостью соответствия выдвигаемым требованиям, среди которых возможность организовать систему на кластере машин, возможность динамического добавления новых источников данных, поддержка

процессов извлечения, трансформации и загрузки данных, и другие требования. По результатам двухэтапного анализа было принято решение в пользу инструмента управления озером данных Dataiku на основе его соответствия разработанным требованиям.

Таблица 18. Описание различных инструментов управления озером данных

№	Инструмент управления озером данных	Описание инструмента управления озером данных
1	Dataiku	Платформа для анализа данных и машинного обучения, которая предоставляет возможности для подготовки данных, создания моделей и их развертывания. Dataiku ориентирован на коллаборацию между различными участниками процесса анализа данных и поддерживает интеграцию с различными источниками данных.
2	Dremio	Это платформа для анализа данных, которая позволяет ускорить доступ к данным и упростить их обработку. Dremio предлагает возможности для виртуализации данных и выполнения запросов на больших объемах данных без необходимости их перемещения.
3	Knime	Это инструмент для анализа данных и создания рабочих процессов, который предоставляет визуальный интерфейс для построения моделей и обработки данных. Knime поддерживает интеграцию с различными источниками данных и имеет обширную библиотеку модулей для анализа данных.
4	Delta Lake	Система управления данными, построенная на базе Apache Spark, которая обеспечивает надежность и консистентность данных в озере данных. Delta Lake поддерживает транзакции ACID и позволяет легко управлять версиями данных.
5	Стек технологий (Apache NIFI + Apache Kafka + Apache Airflow)	Apache NIFI: Это инструмент для автоматизации потоков данных, который позволяет собирать, обрабатывать и распределять данные в реальном времени. Apache Kafka: Это платформа для обработки потоков данных, которая обеспечивает высокую производительность и надежность при передаче данных. Apache Airflow: Это инструмент для автоматизации рабочих процессов, который позволяет управлять сложными процессами обработки данных.

Аббревиатуры: Принцип “ACID” - Atomicity (Атомарность), Consistency (Согласованность), Isolation (Изоляция), Durability (Долговечность).

По итогам первого этапа анализа были выявлены ключевые характеристики различных инструментов и стека технологий, имеющие значение для принятия решения о выборе того или иного инструмента.

По результатам анализа платформы Dataiku было установлено, что количество узлов, предоставляемых в бесплатной версии для инициализации источников данных, существенно меньше по сравнению с платной версией. Также был выявлен недостаток в виде ограниченного числа готовых «рецептов» для обработки данных по сравнению с другими системами, что предполагает необходимость реализации их логики преимущественно с использованием языков

программирования Python, R и Shell. На первом этапе анализа также было обнаружено, что Dataiku включает удобный механизм повторного использования «рецептов». Среди положительных характеристик были отмечены интуитивно понятный графический интерфейс и возможность создания веб-приложений для пользовательской работы непосредственно внутри платформы. Важно также отметить, что для управления построенными в потоке конвейерами данных предусмотрена возможность оркестрации запросов.

Dremio представляет собой платформу data-as-a-service с открытым исходным кодом, базирующуюся на технологии Apache Arrow. Эта технология обеспечивает ускорение выполнения запросов к разнообразным источникам данных, включая S3 (Amazon Simple Storage Service), ADLS (Azure Data Lake Storage), HDFS (Hadoop Distributed File System), а также NoSQL и реляционные базы данных. Анализ показал, что Apache Arrow сочетает в себе преимущества столбчатых структур данных и вычислений в оперативной памяти. На основании результатов первого этапа анализа было установлено, что Dremio с использованием Apache Arrow обеспечивает высокую производительность и гибкость при работе со сложными данными и динамическими схемами. Важным аспектом является то, что Apache Arrow позволяет эффективно работать даже с данными, которые не полностью помещаются в оперативную память. Dremio решает задачу быстрой обработки данных, устраняя риски, связанные с избыточным копированием данных. Однако анализ выявил, что Dremio не предоставляет визуальный инструмент для построения конвейеров данных, аналогичный тем, что предлагают платформы Dataiku, KNIME и Nifi. Для подключения новых источников данных в Dremio предусмотрен раздел "Sources".

Основные функции платформы KNIME ориентированы на машинное обучение, интеллектуальный анализ и обработку данных. Было выявлено, что в KNIME доступны дополнительные функции через различные расширения и поддержку от многочисленных сообществ и поставщиков. Передача данных в формате table, который является собственным форматом KNIME, обеспечивает высокую скорость работы с большими таблицами. В KNIME поддерживаются различные типы узлов, включая IO (ввод/вывод данных), Manipulation (преобразование данных, такие как фильтрация строк и столбцов, сортировка, объединение и т.д.), Database (подключение к базам данных, чтение/запись, выполнение запросов), Workflow Control (создание циклов и итерация групп в ходе выполнения workflow), Views (визуализация данных), и scripting (написание скриптов на различных языках программирования). KNIME позволяет объединять узлы в более высокоуровневые компоненты с возможностью создания конфигурационных меню для их настройки, что повышает интерпретируемость программы и облегчает повторное использование частей конвейера данных.

В результате анализа платформы Delta Lake было установлено, что её формат и вычислительные возможности позволяют упрощенно создавать масштабные конвейеры обработки данных. Delta Lake обеспечивает транзакции с соблюдением свойств ACID благодаря масштабируемой обработке метаданных и унификации потоковой и пакетной обработки данных. Журнал транзакций Delta Lake гарантирует согласованную обработку данных, даже при параллельной работе с таблицей других потоковых или пакетных запросов. Однако следует отметить, что Delta Lake не поддерживает транзакции с участием нескольких таблиц и внешние ключи, ограничиваясь транзакциями на уровне отдельной таблицы.

Выявлено, что перед записью данных осуществляется проверка их схемы на предмет совместимости. В случае несовместимости схемы Delta Lake полностью отменяет транзакцию, что позволяет контролировать данные, поступающие в озеро данных, при этом предусмотрена возможность изменения схемы. Отличительной чертой Delta Lake является возможность версионирования данных, что достигается за счёт фиксации всех взаимодействий с таблицей в журнале транзакций. К недостаткам рассматриваемой платформы можно отнести значительные ограничения использования графического интерфейса в рамках бесплатной версии.

В дополнение к использованию готовых инструментов, было принято решение провести анализ возможности реализации системы на основе нескольких низкоуровневых технологий: Apache NIFI для создания конвейеров данных (ETL-процессов), Apache Airflow в качестве оркестратора и Apache Kafka в качестве шины данных. Результаты проведенного анализа показали, что данное решение обеспечивает максимальную гибкость и предоставляет широкие возможности для выбора конкретных подходов. Однако это также приводит к значительному увеличению трудозатрат.

По итогам первого этапа анализа характеристик инструментов для организации озер данных посредством тестового использования были сделаны следующие выводы:

1. Dataiku не имеет ограничений на модель интегрируемых данных и адаптирован для повторного использования конвейеров данных.
2. Dremio более подходит для выполнения запросов и аналитики больших объемов данных внутри озера, но не для загрузки данных из новых источников.
3. KNIME оптимален для задач, использующих методы машинного обучения и интеллектуального анализа данных. Он обеспечивает высокую скорость работы с большими таблицами, поддерживает различные типы узлов, которые можно объединять в верхнеуровневые компоненты с возможностью создания конфигурационных меню для их настройки, что повышает интерпретируемость программы и облегчает повторное использование частей конвейера данных.
4. Delta Lake может быть оптимальным решением для задач, требующих создания больших конвейеров данных. Журнал транзакций Delta Lake гарантирует согласованную

обработку данных, даже при наличии параллельных потоковых запросов. Специальный алгоритм записи данных позволяет контролировать данные, поступающие в озеро, с возможностью их версионирования, так как любое взаимодействие с таблицей фиксируется в журнале транзакций.

5. Использование стека из нескольких технологий обработки данных требует неоправданно высоких трудозатрат.

Итоги первого этапа анализа различных инструментов управления озером данных позволили сформулировать список требований, которым должен удовлетворять инструмент управления озером данных. На втором этапе различные инструменты управления озером данных были проанализированы на соответствие выдвинутым требованиям. В таблице 19 представлены результаты анализа инструментов управления озером данных на соответствие выдвинутым требованиям.

Таблица 19. Соответствие различных инструментов управления озером данных и стека технологий выдвинутым требованиям

	Dremio	Dataiku	KNIME	Delta lake	NIFI + Kafka + Airflow
Возможность организовать систему на кластере машин	Да	Да	Да	Да	Да
Возможность динамического добавления новых источников данных	Нет	Да	Да	Да	Да
Открытый исходный код	Да	Да	Да	Да	Да
Поддержка процессов извлечения, трансформации и загрузки данных (ETL)	Да	Да	Да	Да	Да
Графический интерфейс	Да	Да	Да	Нет	NIFI и Airflow
Интерфейс с функцией перетаскивания элементов (drag-and-drop)	Нет	Да	Нет	Нет	Нет
Совместимость с реляционными базами данных	Да	Да	Да	Да	Да
Совместимость с NoSQL базами данных	Да	Да	Да	Да	Да
Встроенная система управления процессами (оркестратор)	Нет	Да	Да	Нет	Airflow
Простота реализации	Да	Да	Да	Нет	Нет
Поддержка различных языков программирования	SQL	Shell, Python, R, SQL, Spark	Python, Java, SQL	Python, Scala, SQL	практически любые
Интеграция с LDAP	Да	Да	Да	Нет	Нет

Проанализировав характеристики инструментов управления озером данных на соответствие выдвинутым требованиям и учитывая трудозатраты на реализацию, было выявлено, что всем перечисленным критериям отвечает только инструмент Dataiku. Соответственно, было

принято решение использовать Dataiku для реализации системы управления озером биоинформатических данных.

3.3.3. Наполнение озера данных

3.3.3.1. Структура созданного озера данных

В рамках настоящей работы было создано озеро данных под управлением инструмента Dataiku, аккумулирующее собственные данные, а также информацию открытых мировых источников. Оказание медицинской помощи большому количеству больных в ФГБУ «НМИЦ Радиологии» Минздрава России позволило интегрировать в созданное озеро уникальный массив собственных данных о различных характеристиках ЗНО у больных из российской популяции. Были интегрированы геномные и клинико-фенотипические сведения, лабораторные метрики пробоподготовки и секвенирования, различные вспомогательные данные, генерируемые в процессе работы биоинформатических конвейеров (метрики качества, метаданные), информация о подборе персонализированного противоопухолевого лекарственного лечения [14].

Также были аккумулирована опубликованная информация об ассоциациях биомаркеров с чувствительностью к различным противоопухолевым лекарственным препаратам из открытых мировых баз данных. Схема, отображающая элементы наполнения разработанного озера данных, приведена на рисунке 14.

Структура созданного озера данных



Рисунок 14. Структура и элементы наполнения озера данных в рамках фармакогеномной платформы. CNV (англ. copy number variation) - вариация числа копий, MSI (англ. microsatellite instability) - микросателлитная нестабильность, TMB (англ. tumor mutational burden) - опухолевая мутационная нагрузка

Созданное озеро данных входит в состав разработанной фармакогеномной платформы. В её рамках за пределами озера данных функционируют разработанные биоинформатические конвейеры, при этом информация о выявленных геномных вариантах, показателях TMB, MSI и CNV вносится в созданное озеро данных. Разработанная СППВР является отдельным составным элементом и не входит в состав фармакогеномной платформы. Однако результаты подбора персонализированного лекарственного лечения онкологических заболеваний СППВР также аккумулируются в озере данных.

Результаты наполнения созданного озера собственными и опубликованными данными приведены далее.

3.3.3.2 Наполнение собственными данными

По итогам проведённой работы, в созданное озеро данных был интегрирован большой объем собственной информации о больных ЗНО из российской популяции, в том числе, данные секвенирования в формате FASTQ, клинико-фенотипические, геномные сведения. К интегрированным собственным данным также относилась вспомогательная информация, генерируемая в процессе работы биоинформатических конвейеров, сведения о таргетных, иммунотерапевтических и химиотерапевтических препаратах для персонализированного лечения ЗНО, подобранных СППВР.

По состоянию на 18 марта 2024 года, в озеро данных была добавлена регистрационная, клинико-фенотипическая, лабораторная и геномная информация о 3686 пациентах, проходивших лечение и/или обследование в филиалах ФГБУ «НМИЦ радиологии» Минздрава России: 1549 пациентов проходили лечение и/или обследование в Медицинском радиологическом научном центре имени А.Ф. Цыба и 2137 пациентов - в Московском научно-исследовательском онкологическом институте имени П.А. Герцена. У больных, информация о которых была внесена в разработанное озеро данных, наиболее часто встречались следующие заболевания: РМЖ - у 2214 больных, РПЖ - у 331 больного, рак яичников - у 253 больных, рак предстательной железы - у 214 больных, рак легкого - у 103 больных. Информация о количестве пациентов с различными онкологическими и неонкологическими заболеваниями, сведения о которых загружены в созданное озеро данных, представлена в таблице 20.

Таблица 20. Количество пациентов, сведения о которых интегрированы в созданное озеро данных

№	Заболевание	Количество пациентов
1	Рак молочной железы	2214
2	Рак поджелудочной железы	331
3	Рак яичников	253
4	Рак предстательной железы	214
5	Рак легкого	103
6	Колоректальный рак	35
7	Болезнь Моторного Нейрона	31
8	Рак почки	29
9	Рак желудка	28
10	Боковой Амиотрофический Склероз	25
11	Рак тонкого кишечника	22
12	Рак эндометрия	22
13	Глиобластома	20
14	Злокачественные новообразования без первично выявленного очага	16
15	Меланома	12
16	Рак мочевого пузыря	11
17	Сахарный диабет 1-го типа	11

№	Заболевание	Количество пациентов
18	Рассеянный склероз	10
19	Рак печени	8
20	Рак ротоглотки	8
21	Ревматоидный артрит	7
22	Рак забрюшинного пространства	6
23	Рак яичка	6
24	Рак сигмовидной кишки	5
25	Липосаркома	4
26	Мышечная дистрофия Беккера	4
27	Нейроэндокринный рак	4
28	Полинейропатия	4
29	Рак слюнной железы	4
30	Болезнь Паркинсона	3
31	Гастроинтестинальная стромальная опухоль	3
32	Рак щитовидной железы	3
33	Холангиокарцинома	3
34	Аутоиммунный васкулит	2
35	Болезнь Альцгеймера	2
36	Гематологические заболевания	2
37	Дрожательный паралич, системная нейродегенеративная болезнь	2
38	Группа карцином с нейроэндокринной дифференцировкой	2
39	Рак костей	2
40	Синдром Туретта	2
41	Талассемия	2
42	Аутоиммунный атрофический гастрит	1
43	Гидраденокарцинома	1
44	Нейрофиброматоз	1
45	Рак гортани	1
46	Рак носоглотки	1
47	Рак языка	1
48	Саркома Капоши	1
49	Синдром Жильбера	1
50	Синдром Хиппеля-Линдау	1
51	Системная нейродегенеративная болезнь. Последствия травмы мозга	1
52	Хроническая ишемическая болезнь сердца, последствия Инсульта	1
53	Хроническая ишемическая болезнь сердца	1
54	Хронический резистентный к терапии болевой синдром	1
55	Цереброваскулярная болезнь	1
56	Здоров	37
57	Неуточненный диагноз	160
58	Общее количество пациентов	3686

В созданное озеро данных были добавлены результаты NGS биологического материала не только больных онкологическими заболеваниями, но и больных с нейродегенеративными, аутоиммунными заболеваниями и иными неонкологическими заболеваниями. Это позволило распространить область потенциального применения созданной фармакогеномной платформы на более широкий спектр нозологий. Важно отметить, что в созданное озеро данных интегрированы результаты полноэкзомного секвенирования 37 здоровых человек, что также расширяет область

научного применения фармакогеномной платформы и предоставляет возможность для реализации будущих популяционно-генетических исследований.

Из таблицы 20 следует, что в озеро данных были добавлены результаты высокопроизводительного секвенирования крови и/или опухолевого материала 160 больных с неуточненным диагнозом. Методы NGS широко применяются в сложных клинических случаях для постановки диагноза, а также для выявления неустановленного первичного очага онкологического заболевания.

В рамках настоящей работы проводилось полноэкзомное NGS биологического материала когорт больных РПЖ и ТНРМЖ, позволившее выявить герминальные и соматические геномные варианты во всех кодирующих последовательностях. Вместе с тем, в зависимости от клинической ситуации, определенным больным может быть показано проведение молекулярно - генетического исследования с применением генетических панелей, в рамках которых исследуется определенный спектр онкологически значимых генов. В разработанное озеро данных также были добавлены данные о результатах молекулярно - генетического тестирования биологического материала больных, проведенного с использованием панельных тестов. Указанный подход позволяет увеличить количество интегрированной информации о геномных характеристиках пациентов [40]. Необходимо отметить, что в озеро данных, входящее в состав разработанной фармакогеномной платформы, также были добавлены результаты полногеномного секвенирования, секвенирования транскриптома и метагенома. Подходы, применяемые при проведении высокопроизводительного секвенирования, представлены в таблице 21.

Таблица 21. Подходы, применяемые при проведении высокопроизводительного секвенирования

Тип тестирования	Количество пациентов
Полноэкзомное секвенирование с покрытием x 100	2302
Agilent SureSelect v6	6
Agilent SureSelect v7	288
Agilent SureSelect v8	18
MGIEasy Exome Capture V5 Probe set	213
Nanodigmbio NEXome Plus Panel V1.0	242
Roche KAPA HyperExome	1535
Полноэкзомное секвенирование с покрытием x 200	347
Agilent SureSelect v6	8
Agilent SureSelect v7	123
Agilent SureSelect v8	33
Nanodigmbio NEXome Plus Panel V1.0	1
Roche KAPA HyperExome	182

Тип тестирования	Количество пациентов
Панельное тестирование	1013
Nanodigmbio HRR Panel V1.0	532
Nanodigmbio NanOnco Plus Panel V3.0	52
OncoAtlas ABCplus	193
OncoAtlas ABCpro	4
OncoAtlas HRD15	4
Roche AVENIO	228
Полногеномное секвенирование x 0.5	1
Полногеномное секвенирование x 30	1
Полногеномное секвенирование x 4	2
Полногеномное секвенирование x 7	8
Транскриптом	9
Метагеном	3
Общее количество больных	3686

Общий объем внесенных в озеро данных необработанных данных секвенирования в формате FASTQ составляет 26,2 терабайта. После осуществления биоинформатической обработки, в созданное озеро данных была интегрирована информация о 65 438 000 герминальных геномных вариантах и о 42 493 соматических геномных вариантах, выявленных по итогам высокопроизводительного секвенирования биологического материала пациентов ($n=3686$).

Свод регистрационных и клинико-фенотипических данных интегрирован в озеро для обеспечения хранения широкого спектра информации о больных и о характеристиках выявленных ЗНО. Указанный свод содержит как клинико-фенотипические данные, общие для различных заболеваний, так и информацию, специфичную для конкретного типа заболевания. Так, о больных ТНРМЖ в озеро данных добавлена клинико-фенотипическая информация о возрасте выявления опухоли, текущем возрасте больных, гистологическом типе опухоли (инвазивная протоковая/дольковая карцинома, другие типы опухоли), стадии заболевания, показателях KI-67, менструальном статусе, проведенном химиотерапевтическом и лучевом лечении, а также иные сведения (Рисунок 15). Важно отметить, в свод клинико-фенотипических данных больных могут быть интегрированы сведения об ответе больного на лечение, проводимое с учетом рекомендаций СППВР, актуализируемые в озере данных один раз в месяц.

Vozrast	Vozrastnaia_gruppa	Vozrast_vyavleniia_opukhi	Lokalizatsiia_opukhki	Gistologicheski_tip_tumora	Marker_proliferatsii_KI_67	Stadiia	Stadiia_poleinaia_TNM_klass	Stepen_differentsirovki_o	Status_menopauzy	IA_BIS_et_skhema_khimiot	IA_BIS_et_1_NAPRINT_2
ageyr	ageyr	ageyr	okruglang	okruglang	val	stage	stage	stage	stage	ageyr	ageyr
63 40+		56									
63 40+		56	Левая молочная железа	Неинвазивный протоковый рак	100%	IB	T2N0M0	G3	менопауза		8
50 40+		49									
50 40+		49	Правая молочная железа	Неинвазивный протоковый рак	72%	IB	T2N3M0	G3	менопауза		4
72 40+		69									
72 40+		69	Правая молочная железа	Неинвазивный протоковый рак	54%	IIA	T2N2M0	G3	НАМ с преддлечком в 1999 г.		1
53 40+		52									
53 40+		52	Левая молочная железа	Неинвазивный протоковый рак	70%	IA	T2N0M0	G3	менопауза		1
36 1-39		36									
36 1-39		36	Правая молочная железа	Неинвазивный протоковый рак	80%	IA	T2N0M0	G3	нет данных		1
42 40+		40									
42 40+		40	Левая молочная железа	Неинвазивный протоковый рак	90%	IB	T2N1M0	G3	Вторичная анимозия наступила в...		3
36 1-39		33									
36 1-39		33	Левая молочная железа	Неинвазивный протоковый рак	60%	IV	T4N1M1	G3	нет данных		1
49 40+		47									
49 40+		47	правая молочная железа	Неинвазивный протоковый рак	70%	IA	T2N0M0	G2	менопауза		3
47 40+		46									
47 40+		46	Левая молочная железа	Неинвазивная дольчатая карцинома	5%	IA	T1N0M0	G3	нет данных		10
51 40+		49									
51 40+		49	правая молочная железа	Неинвазивный протоковый рак	5%	IA	T2N0M0	G3	нет данных		3
54 40+		52									
54 40+		52	правая молочная железа	Неинвазивный протоковый рак	60%	IA	cT2N0M0	G3	нет данных		1

Рисунок 15. Фрагмент свода регистрационных и клинико-фенотипических данных больных ТНРМЖ, интегрированного в созданное озеро. Агрегированы различные характеристики, в том числе возраст выявления опухоли, текущий возраст больных, менструальный статус, показатели маркера пролиферации KI-67, сведения о проведенном химиотерапевтическом лечении и иные сведения

Лабораторные метрики пробоподготовки и секвенирования биологического материала больных, содержащие подробную информацию о проводимом молекулярно-генетическом тестировании, также интегрированы в разработанное озеро данных в виде свода.

Массив информации, содержащейся в отчетах о выявлении геномных вариантов, значениях TMB, CNV и MSI по результатам работы герминального и соматического биоинформатических конвейеров, также сохранён в озере данных в виде свода. Вес сводов о выявленных геномных вариантах у больных РПЖ и ТНРМЖ составил 369,5 Мб и 1,09 Гб, соответственно. Набор параметров, собирающих информацию о патогенных и вероятно-патогенных вариантах, выявленных в результате работы герминального биоинформатического конвейера, представлен в Приложении 4.

Результаты подбора персонализированного таргетного, иммунотерапевтического и химиотерапевтического лечения, сформированные разработанной СППВР, также вносятся в созданное озеро данных в виде свода. На настоящий момент в озеро данных внесена информация о подборе персонализированного лекарственного лечения онкологических заболеваний по результатам анализа геномных данных 273 больных разработанной СППВР.

Также в созданное озеро данных вносится информация о метаданных, получаемых в процессе работы созданных биоинформатических конвейеров. По мере выполнения этапов, формируются структурированные и неструктурированные данные - выходные файлы (отчеты, аннотации, промежуточные результаты). Структурированные данные сохраняются в PostgreSQL, неструктурированные - в Amazon S3 хранилище в той же папке, где расположены FASTQ файлы.

По окончании каждого шага биоинформатического конвейера, происходит запись времени, затраченного на его выполнение. Это необходимо для выяснения узких мест

биоинформатического конвейера, чтобы понять какой шаг в среднем занимает больше всего времени. Информация о метриках качества также вносится в озеро данных. Внешний вид свода, содержащего собранные метрики качества, представлен на рисунке 16.

id	patient_id	time_stamp	mean_bait_coverage	duplication_rate	fold_enrichment	output	reference_sequence
3344	7245URHI	2023-11-20 22:32:48.887983	99.246371	0.0306125	41.878769	/output/7245URHI_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
2625	7408YPLC	2023-10-14 15:11:23.938808	48.647372	0.0151024	40.363872	/output/7408YPLC_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
2633	3554QXUV	2023-10-15 01:59:55.338736	219.89252	0.0700705	41.437504	/output/3554QXUV_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
2710	1212KLSW	2023-10-27 16:03:17.259671	65.904486	0.0344203	37.214009	/output/1212KLSW_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
2596	1451RPR	2023-10-14 07:39:27.296568	56.653675	0.0181325	37.264077	/output/1451RPR_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
2487	3349NWWG	2023-09-28 06:33:59.944978	57.883942	0.0123655	37.494926	/output/3349NWWG_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
2415	3302MYOH	2023-09-24 23:44:03.885472	47.123937	0.0154069	38.185816	/output/3302MYOH_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
2930	0958HBZC	2023-10-31 00:26:06.973137	113.291175	0.122472	43.751475	/output/0958HBZC_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
2810	9528SCGB	2023-10-28 20:15:46.904704	22.376559	0.00147868	38.19956	/output/9528SCGB_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
2410	2623GHDE	2023-09-24 22:44:58.087962	44.446888	0.014957	37.62211	/output/2623GHDE_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
2455	4242EOJZ	2023-09-27 23:06:30.066857	58.142371	0.00980228	37.314251	/output/4242EOJZ_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
2999	4204QOMC	2023-11-02 01:12:07.158655	63.874769	0.180243	23.241211	/output/4204QOMC_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
3453	4611LAPK	2023-11-24 11:39:14.510529	54.75375	0.0333591	40.670921	/output/4611LAPK_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
1766	5420CKKG	2023-07-30 17:31:26.196434	51.947422	0.0249741	40.289692	/home/huser/data-drive/Science/Core/5420CKKG-...	/home/huser/data-drive/software/index/indexFA_...
2497	4160KBKI	2023-09-28 10:26:20.237364	58.799316	0.0255405	37.791743	/output/4160KBKI_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
2951	1741FSZJ	2023-10-31 16:00:23.066761	108.99138	0.0229203	43.635273	/output/1741FSZJ_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
3060	2512OPLY	2023-11-06 11:48:28.076005	84.916894	0.0156552	42.118356	/output/2512OPLY_HSmetrics.txt	/software/index/indexFA_37_prepA2/GRCh37_ne...
1902	9035LONM	2023-08-07 06:12:50.396796	104.344684	0.0398078	40.056764	/data/9035LONM-E100-Roche_KAPA_HyperExome-...	/root/data-drive/software/index/indexFA_37_prep...

Рисунок 16. Внешний вид свода, собирающего метрики качества, связанные с работой герминального биоинформатического конвейера. Виден результат сводного анализа данных NGS секвенирования 2485 образцов

Сбор метрик качества производится также непосредственно по FASTQ файлам для отсеивания бракованных образцов.

В созданном озере данных формируются общие сводные таблицы, объединяющие весь массив информации (результаты биоинформатического анализа, регистрационные, клинико-фенотипические данные, результаты подбора персонализированного лекарственного лечения, метаданные и тд.), связанной с анализом всей совокупности анализируемых образцов. На рисунке 17 приведен пример построения общей сводной таблицы по всем проанализированным больным в нашем озере данных на платформе Dataiku.

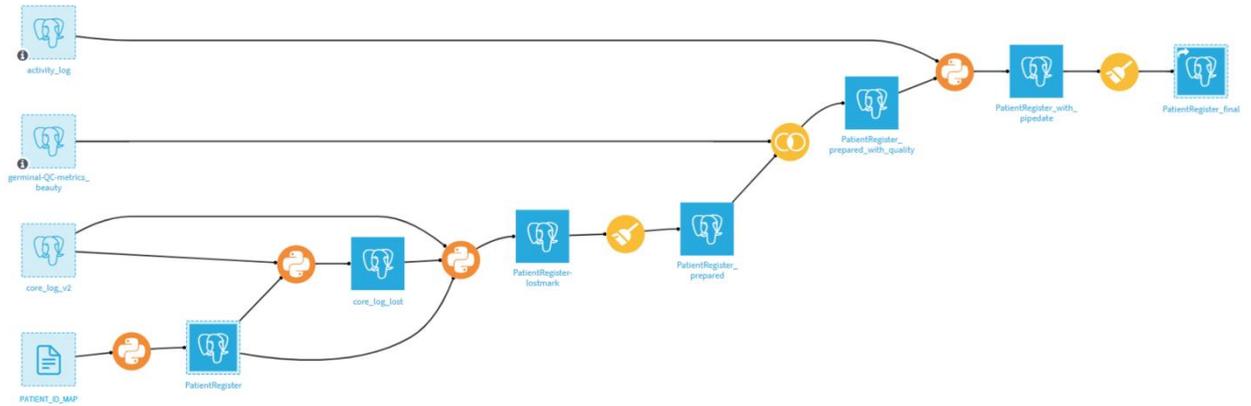


Рисунок 17. Логика формирования общей сводной таблицы по всем найденным геномным вариантам, клинико-фенотипическим и регистрационным данным, результатам подбора персонализированного лекарственно лечения, метрикам качества, метаданным и другим данным

На рисунке 18 представлен внешний вид общей сводной таблицы на платформе Dataiku. В общую сводную таблицу включается большой объем всей собранной информации, отраженной в более чем в 43 столбцах таблицы для герминального и 59 для соматического сводов.

Pol	ID_Obraztsa	Nukleinoivaia_kislota	Что_delaetsia_tekhniches	Chem_vydeleno_NK	Kit_pri_obogashchenii	Nomer_zapuska...	Diagnoz_gruppirovanyi
string	string	string	string	string	string	string	string
Text	Text	Text	Text	Natural lang.	Natural lang.	Text	Text
Ж	2482FLON	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	12PE100	Рак молочной железы
Ж	1209VYVB	Соматическая ДНК	WESx200	Qiagen QIAamp DNA FFPE Tissue Kit	Roche KAPA HyperExome	35PE100	Рак молочной железы
Ж	4356LLCT	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	12PE100	Рак молочной железы
М	8457LDQK	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	38PE100	Рак поджелудочной железы
Ж	9035LONM	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	27PE100	Рак молочной железы
Ж	7637DTZL	Соматическая ДНК	WESx200	Qiagen QIAamp DNA FFPE Tissue Kit	Roche KAPA HyperExome	41PE150	Рак молочной железы
Ж	9585PBXA	Герминальная ДНК	WESx100	Raisiol GM Blood M	Nanodigmbio NEXome Plus Panel v1.0	48PE150	Рак молочной железы
Ж	2899BVSO	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	18PE100	Рак молочной железы
Ж	0625NFED	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Nanodigmbio NEXome Plus Panel v1.0	24PE150	Рак молочной железы
Ж	0408MHSP	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	13PE100	Рак молочной железы
Ж	1789JXRK	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	17PE100	Рак поджелудочной железы
Ж	4540AERX	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Nanodigmbio NEXome Plus Panel v1.0	24PE150	Рак молочной железы
Ж	9697IAQN	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	36PE100	Рак молочной железы
Ж	0530LZAI	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	22PE100	Рак молочной железы
Ж	2569GTGR	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	MGIeasy Exome Capture VS Probe set	1PE100	Рак молочной железы
Ж	9522RRRF	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	22PE100	Рак молочной железы
Ж	0484HBNY	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Nanodigmbio NEXome Plus Panel v1.0	50PE100	Рак молочной железы
Ж	1320FHHP	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	14PE100	Рак молочной железы
М	1427MEED	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Nanodigmbio NEXome Plus Panel v1.0	24PE150	Рак поджелудочной железы
Ж	5676PCAD	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	29PE100	Рак молочной железы
Ж	9248QYHA	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	22PE100	Рак молочной железы
Ж	9755RXZZ	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Nanodigmbio NEXome Plus Panel v1.0	48PE150	Рак молочной железы
Ж	4408KDRG	Соматическая ДНК	WESx200	Qiagen QIAamp DNA FFPE Tissue Kit	Roche KAPA HyperExome	32PE100	Рак молочной железы
Ж	0206RGVU	Герминальная ДНК	WESx100	Qiagen QIAamp DNA Mini Kit	Roche KAPA HyperExome	12PE100	Рак поджелудочной железы

Рисунок 18. Внешний вид части общей сводной таблицы на платформе Dataiku

На рисунке 19 представлена схема структурированных данных из программы Datagrip. Представлен структурированный свод данных на платформе PostgreSQL. Он состоит из разрозненных более маленьких сводов данных, содержащих различную информацию. Как видно из рисунка, связанных между собой по ключам сводов данных достаточно мало, так как слияние происходит виртуально через инструмент Dataiku.

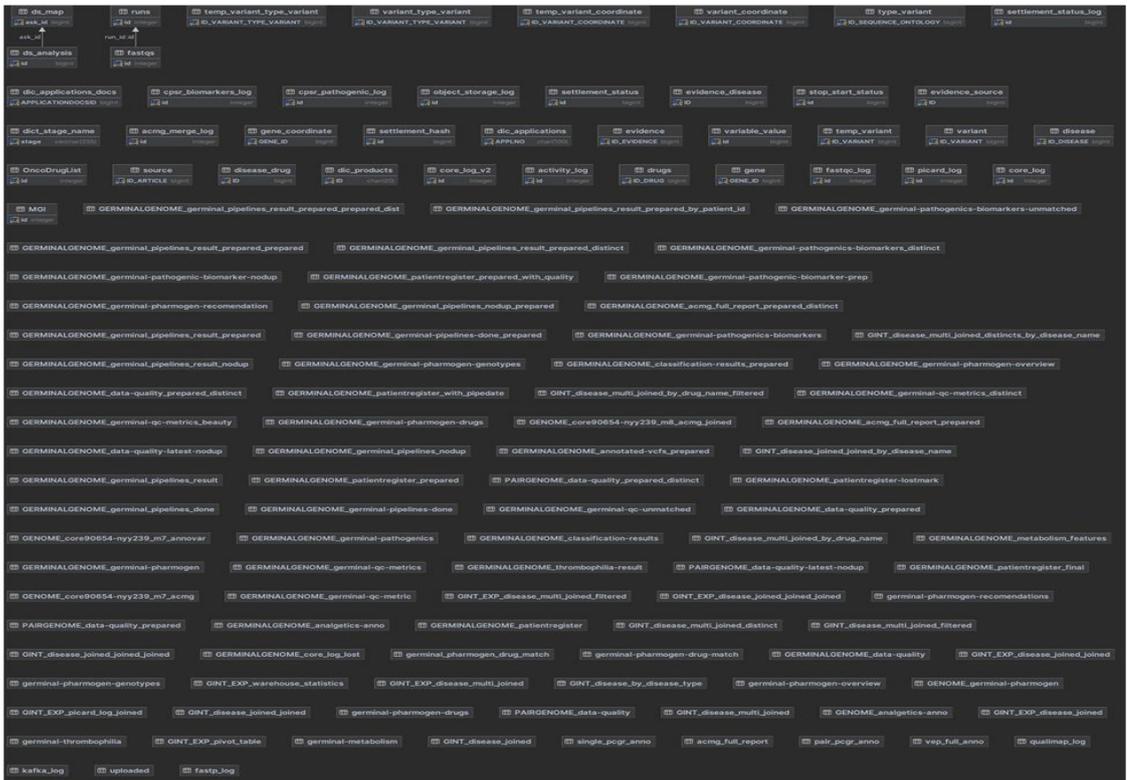


Рисунок 19. PostgreSQL: схема структурированных данных, хранящихся в разработанном озере данных

На рисунке 20 изображена связка бакетов в хранилище Amazon S3 для слабо-структурированных и неструктурированных данных.

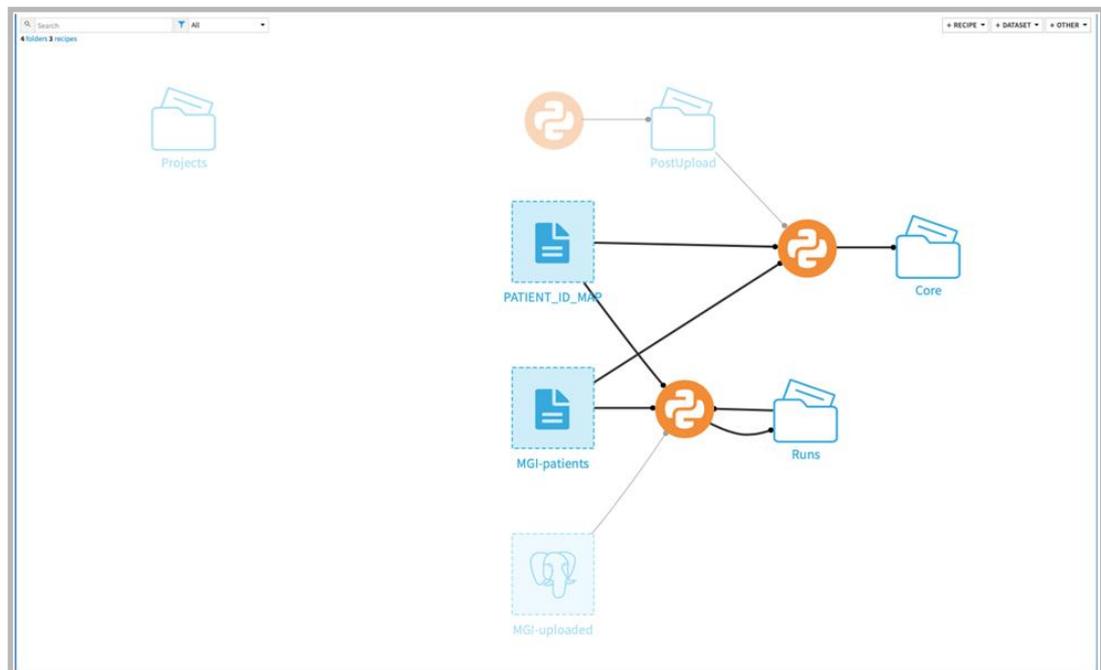


Рисунок 20. Связка бакетов в хранилище Amazon S3 для слабо-структурированных и неструктурированных данных

Показано, что в PostUpload расположены данные образцов в формате FASTQ, которые в данный момент находятся в обработке. Результаты обсчета и промежуточные файлы сохраняются здесь же. “Core” является основным хранилищем. В данном бакете листинг идет по ID пациента, а не образца, так как на одного человека может приходиться несколько различных анализов и, соответственно, образцов. Внутри указанного бакета не производится никаких операций с данными. “Runs” является бакетом для сырых данных. В бакет “Runs” автоматически зачисляются данные с секвенатора и ждут своей очереди, чтобы попасть в PostUpload. В бакете “Projects” расположены артефакты и различные системные файлы. Связки между бакетами разработаны на языке Python, который использует информацию из PostgreSQL и для установления связи между файлом, ID образца и ID пациента и прочими метаданными.

3.3.3.3. Наполнение опубликованными данными из открытых мировых источников

Для обеспечения аннотирования биоинформатическими конвейерами и построения связи “геномный вариант/TMB - лекарственный препарат” СППВР, в созданное в настоящей работе озеро данных была интегрирована информации из открытых мировых баз данных. Суммарно, созданное озеро данных аккумулировало 1132 Гб информации из баз данных GENCODE, dbNSFP, COSMIC, ClinVar, 1000 Genome project, ExAC, dbSNP, OMIM, gnomAD, Pfam, UniProtKB, DoCM, CancerHotspots, CancerMine, OpenTargetsPlatform, ChEMBL, CIViC, CGI (Таблица 22).

Таблица 22. Открытые мировые базы данных, информация из которых была интегрирована в созданное озеро данных

№	База данных	Структура и объем информации, интегрированной в созданное озеро данных
1	GENCODE (Human Release 43)	<p>Общее количество генов: 62,703</p> <p>Белок-кодирующие гены: 19,393</p> <p>Длинные некодирующие РНК (lncRNA): 19,928</p> <p>Малые некодирующие РНК: 7,566</p> <p>Псевдогены: 14,737</p> <p>Общее количество транскриптов: 252,913</p> <p>Белок-кодирующие транскрипты: 89,411</p> <p>Транскрипты lncRNA: 58,023</p> <p>Транскрипты с бессмысленным опосредованным распадом (NMD): 21,354</p> <p>Объем данных: ~10-12 Гб в разархивированном виде</p>
2	dbNSFP (v5.1.0)	<p>84 млн записей несинонимичных SNP (nsSNP)</p> <p>Генов ~20,000</p> <p>Объем данных: ~70 Гб</p>
3	COSMIC (v91)	<p>Мутации: 6 млн соматических мутаций</p> <p>Генов: 700</p> <p>Объем данных: ~50–60 Гб</p>

№	База данных	Структура и объем информации, интегрированной в созданное озеро данных
4	ClinVar (v 04_2024)	Клинически значимых вариантов: ~1,5 млн Генов: ~20,000 Объем данных: ~50–100 ГБ
5	1000 Genome project (Phase 3b)	Количество геномов: 2503 Генетические варианты: 88 миллионов Объем данных: ~150 ГБ
6	ExAC - Exome Aggregation Consortium v2.1.1	Число вариантов: 60 миллионов Индивидов: ~60,000 Объем данных: 200 Гб
7	dbSNP (v155)	Количество записей о SNP: 108 760 547 Объем данных: ~60 ГБ
8	OMIM - Online Mendelian Inheritance in Man, v6.40	Число записей: 6000 Объем данных: 0,1 ГБ
9	gnomAD (v3.1.1)	Общее количество геномов: 150 646 Вариантов: ~240 млн Объем данных: ~50–100 ГБ
10	Pfam(v 35.0)	Белковых семейств: ~19,000 Объем данных: 0,3 ГБ
11	UniProtKB (v 2024_04)	Количество белков: 230 997 167 Объем данных: 136,55 ГБ
12	DoCM (v2.1)	Клинически значимых мутаций: 2503 Объем данных: ~1 ГБ
13	CancerHotspots (v2.1)	Мутаций: 24592 Объем данных: ~0,2 ГБ
14	CancerMine (v 12_2023)	Общее количество мутаций: 15 894 Объем данных: 0,35 ГБ
15	OpenTargetsPlatform (v GRCh38)	Количество записей о заболеваниях: 2,256 Количество записей о мишенях: 1,604 Количество записей о лекарствах: 1,813 Количество записей о биомаркерах: 1,102 Количество записей о генах: 26,523 Количество записей о белках: 22,618 Общий объем данных: ~160 ГБ
16	ChEMBL (v 31)	Общее количество записей: 2 037 548 Количество мишеней: 46 703 Количество биоактивностей: 112 006 587 Объем данных: 22,1 ГБ
17	CIViC (v1.0.1)	Количество геномов: 2672 Объем данных: ~0,16 ГБ
18	CGI (v 12_2023)	Количество геномов: 1869 Генетические варианты: ~60 миллионов Объем данных: ~60 ГБ

Эти наборы данных позволили, в том числе, наполнить собственную базу данных мутаций пациентов различными параметрами, представляющими из себя колонки в аннотации герминального и соматического (парного) VCF файла (Приложение 5, Приложение 6).

Важно отметить, что при достаточном количестве накопленной информации в созданном озере данных в рамках фармакогеномной платформы, её можно начинать использовать для аннотирования и обеспечения подбора персонализированного лечения вместо вышеупомянутых баз данных даже при отсутствии подключения к сети Интернет.

На настоящий момент, загруженное количество информации позволяет производить аннотацию выявляемых геномных вариантов и подбор персонализированного лекарственного лечения ЗНО с использованием только разработанной фармакогеномной платформы без необходимости подключения к внешним базам данных. Механизмы интеграции новой информации из открытых мировых баз данных позволяют осуществлять актуализацию информации, содержащейся в созданном озере данных.

Интеграция информации из названных выше открытых мировых баз данных позволила нам создать собственную базу данных (Рисунок 21).

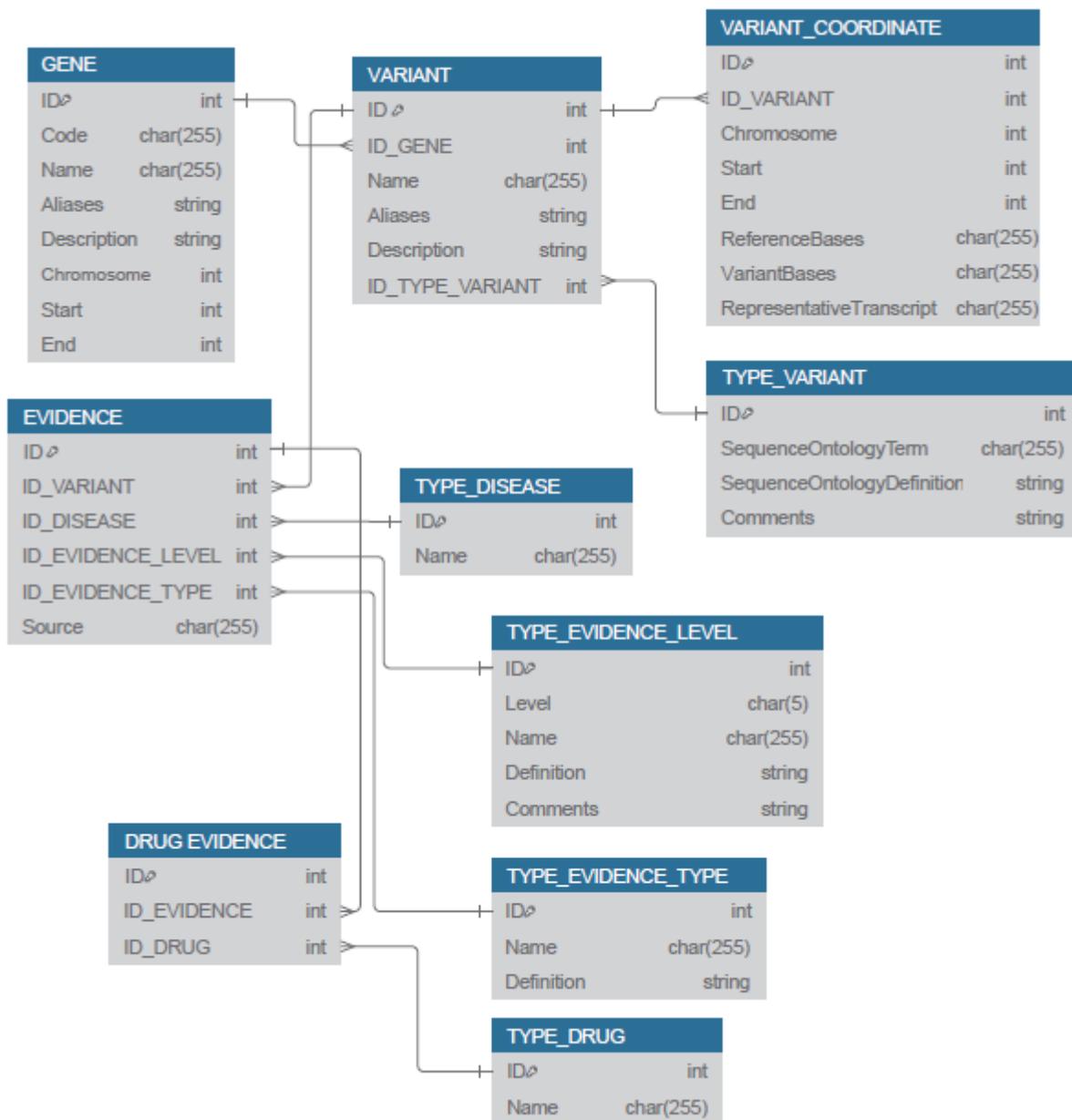


Рисунок 21. Структура созданной базы данных.

Как следует из рисунка 21, в структуре созданной базы данных можно выделить различные таблицы (Таблица 23).

Таблица 23. Таблицы, входящие в созданную базу данных

№	Наименование таблицы	Поля в структуре таблицы
1	GENE (гены)	ID — Идентификатор (ген)ключевое поле), Code - код гена, Name — наименование гена, Aliases — псевдонимы, поскольку у одного гена может существовать существует несколько названий, Description — описание, Chromosome — номер хромосомы, в которой содержится ген, Start — начальная позиция гена в последовательности нуклеотидов, End — конечная позиция гена в последовательности нуклеотидов.
2	VARIANT (мутации)	ID — идентификатор мутации (Ключевое поле), ID_GENE — идентификатор гена (ссылка на таблицу GENE), Name — наименование мутации, Aliases — псевдонимы. У мутации существует несколько названий, Description — описание мутации, ID_TYPE_VARIANT — тип мутации (ссылка на справочник TYPE_VARIANT).
3	VARIANT_COORDINATE (координаты мутации)	ID — идентификатор (ключевое поле), ID_VARIANT - ссылка на мутации (Variant), Chromosome — номер хромосомы, Start — начальная позиция мутации, End — конечная позиция мутации, ReferenceBases — последовательность нуклеотидов до мутации, VariantBases - последовательность нуклеотидов после мутации, RepresentativeTranscript — идентификатор расшифровки гена.
4	Справочник TYPE_VARIANT (типы мутаций)	ID — идентификатор типа (ключевое поле), Sequence Ontology Term - название онтологии, Sequence Ontology Definition — описание онтологии, Comments — дополнительные комментарии.
5	EVIDENCE (диагноз)	ID — идентификатор диагноза (Ключевое поле), ID_VARIANT — идентификатор мутации (ссылка на VARIANT), ID_DISEASE — идентификатор заболевания (ссылка на TYPE_DISEASE), ID_EVIDENCE_LEVEL — ссылка на справочник (TYPE_EVIDENCE_LEVEL), ID_EVIDENCE_TYPE — ссылка на справочник (TYPE_EVIDENCE_TYPE), Source — ссылка на источник (статья, публикации и т.д.).
6	Справочник TYPE_DISEASE (заболевания)	ID - идентификатор болезни (ключевое поле), Name - название.

№	Наименование таблицы	Поля в структуре таблицы
7	Справочник TYPE_EVIDENCE_LEVEL (уровень надежности доказательства)	ID — идентификатор надежности доказательства (ключевое поле), Level — код надежности, Name — наименование, Definition — определение и описание, Comments — дополнительная информация.
8	Справочник TYPE_EVIDENCE_TYPE (типы доказательств)	ID — идентификатор типа (Ключевое поле), Name — наименование, Definition — определение и описание.
9	Сводная таблица DRUG_EVIDENCE (ген)препараты- заклучения)	ID — идентификатор (ключевое поле), ID_EVIDENCE - ссылка на таблицу диагнозов, ID_DRUG — ссылка на таблицу препаратов.
10	TYPE_DRUG (справочник препаратов)	ID — идентификатор (ключевое поле). Name — название препарата.

В рамках созданной базы данных были настроены связи, тем самым обеспечивая возможность отслеживания мутаций, связанных с каждым конкретным геном. Это позволяло эффективно работать с данными о мутациях, анализировать их влияние на гены и проводить дальнейшие исследования генетических вариаций. Связи в структуре базы данных приведены далее:

1. GENE (Гены) ↔ VARIANT (Мутации)

- Тип связи: One-to-Many (один ко многим).
- Связь: Поле GENE.ID (первичный ключ) связано с полем VARIANT.ID_GENE (внешний ключ).

2. VARIANT (Мутации) ↔ VARIANT_COORDINATE (Координаты мутации)

- Тип связи: One-to-One (один к одному).
- Связь: Поле VARIANT.ID (первичный ключ) связано с полем VARIANT_COORDINATE.ID_VARIANT (внешний ключ).

3. VARIANT (Мутации) ↔ TYPE_VARIANT (Типы мутаций)

- Тип связи: Many-to-One (многие к одному).

- Связь: Поле VARIANT.ID_TYPE_VARIANT (внешний ключ) связано с полем TYPE_VARIANT.ID (первичный ключ).
4. EVIDENCE (Диагноз) ↔ VARIANT (Мутации)
- Тип связи: Many-to-One (многие к одному).
 - Связь: Поле EVIDENCE.ID_VARIANT (внешний ключ) связано с полем VARIANT.ID (первичный ключ).
5. EVIDENCE (Диагноз) ↔ TYPE_DISEASE (Заболевания)
- Тип связи: Many-to-One (многие к одному).
 - Связь: Поле EVIDENCE.ID_DISEASE (внешний ключ) связано с полем TYPE_DISEASE.ID (первичный ключ).
6. EVIDENCE (Диагноз) ↔ TYPE_EVIDENCE_LEVEL (Уровень надежности доказательства)
- Тип связи: Many-to-One (многие к одному).
 - Связь: Поле EVIDENCE.ID_EVIDENCE_LEVEL (внешний ключ) связано с полем TYPE_EVIDENCE_LEVEL.ID (первичный ключ).
7. EVIDENCE (Диагноз) ↔ TYPE_EVIDENCE_TYPE (Типы доказательств)
- Тип связи: Many-to-One (многие к одному).
 - Связь: Поле EVIDENCE.ID_EVIDENCE_TYPE (внешний ключ) связано с полем TYPE_EVIDENCE_TYPE.ID (первичный ключ).
8. DRUG_EVIDENCE (Ген)Препараты- заключения ↔ EVIDENCE (Диагноз)
- Тип связи: Many-to-One (многие к одному).
 - Связь: Поле DRUG_EVIDENCE.ID_EVIDENCE (внешний ключ) связано с полем EVIDENCE.ID (первичный ключ).
9. DRUG_EVIDENCE (Ген)Препараты- заключения ↔ TYPE_DRUG (Справочник препаратов)
- Тип связи: Many-to-One (многие к одному).
 - Связь: Поле DRUG_EVIDENCE.ID_DRUG (внешний ключ) связано с полем TYPE_DRUG.ID (первичный ключ).

3.3.4. Анализ существующих биоинформатических конвейеров

Анализ существующих биоинформатических конвейеров Sarek, Bcbio-Nextgen, SpeedSeq, DNAr и DNAscan2 показал, что ни один из них не соответствовал всем требованиям, выдвигаемым нами (аргументация к выдвижению требований к биоинформатическим конвейерам подробно изложена в разделе 2.3.3.1). Соответственно, было принято решение о необходимости разработки собственных биоинформатических конвейеров. Результаты анализа существующих биоинформатических конвейеров представлены в таблице 24.

Таблица 24. Результаты оценки соответствия существующих биоинформатических конвейеров выдвигаемым требованиям

	Выдвигаемые общие требования	Sarek	Bcbio-Nextgen	SpeedSeq	DNAr	DNAscan2
1	Масштабируемость	ДА	ДА	ДА	ДА	ДА
2	Гибкость	НЕТ	НЕТ	НЕТ	НЕТ	НЕТ
3	Документация	ДА	ДА	ДА	ДА	ДА
4	Воспроизводимость	ДА	ДА	ДА	ДА	ДА
5	Возможность использовать docker	ДА	ДА	НЕТ	НЕТ	ДА
	Выдвигаемые специальные требования	Sarek	Bcbio-Nextgen	SpeedSeq	DNAr	DNAscan2
6	Сортировка и контроль качества Fastq	ДА	ДА	ДА	ДА	ДА
7	Тримминг адаптеров секвенирования	ДА	ДА	НЕТ	НЕТ	НЕТ
8	Выравнивание на референс	ДА	ДА	ДА	ДА	ДА
9	Метрики выравнивания	ДА	НЕТ	НЕТ	ДА	НЕТ
10	Маркирование дубликатов	ДА	ДА	ДА	ДА	ДА
11	Удаление дубликатов	НЕТ	НЕТ	НЕТ	НЕТ	НЕТ
12	Метрики дубликатов	ДА	ДА	ДА	ДА	ДА
13	Конвертация SAM в BAM и фильтрация BAM	ДА	ДА	ДА	ДА	ДА
14	Герминальные SNV/INDEL	ДА	НЕТ	НЕТ	НЕТ	ДА
15	Соматические SNV/INDEL	ДА	ДА	ДА	ДА	ДА
16	Перекрёстная проверка вызова SNV/INDEL	НЕТ	НЕТ	НЕТ	НЕТ	НЕТ
17	Обнаружение CNV	ДА	НЕТ	НЕТ	НЕТ	НЕТ
18	Фильтрация вариантов	НЕТ	НЕТ	НЕТ	НЕТ	НЕТ
19	Метрики вызова вариантов	ДА	ДА	ДА	ДА	ДА
20	Аннотация вариантов	ДА	ДА	НЕТ	НЕТ	ДА
21	Сравнение с внешними базами данных	ДА	ДА	НЕТ	НЕТ	ДА
22	Клиническая значимость	ДА	ДА	НЕТ	НЕТ	ДА

Аббревиатуры: CNV (англ. copy number variation) - вариация числа копий, SNV (англ. single nucleotide variant) - однонуклеотидный вариант, INDEL (англ. insertions/deletions) - вставки/делеции

Все вышеперечисленные биоинформатические конвейеры используют на входе файлы пары fastq.gz, а именно форвард и реверс при парном секвенировании. Также, все типы программного обеспечения работают в Unix-образных средах, т.е. Ubuntu, Debian и т.д.

По итогам проведенного анализа на соответствие выдвинутым требованиям были сформулированы выводы касательно различных биоинформатических конвейеров.

Биоинформатический конвейер Sarek соответствует большинству критериев, в том числе важно отметить, что осуществляется его поддержка разработчиками и регулярные обновления. Однако имеются сложности, связанные со вторичным использованием docker и оформлением sample-sheet (таблица значений) для образцов, что делает его непригодным для использования совместно с инструментом управления озером данных Dataiku.

Одним из основных недостатков конвейера Vcbio-Nextgen является то, что в данном конвейере docker не предусмотрен в качестве альтернативы. Кроме того, он является слишком массивным для вторичного запуска через внешний docker-контейнер, что также плохо совместимо с Dataiku. В Vcbio-Nextgen присутствуют фактически все компоненты для работы, однако нет возможности определять герминальные варианты. Поэтому он используется, в основном, на Hadoop-кластерах. В данной работе используются нескольких виртуальных машин, и организация мощностей в виде кластера не применяется.

Некоторые рассматриваемые конвейеры к настоящему времени уже устарели. Так, биоинформатический конвейер SpeedSeq устарел: последние обновления датированы 2015 годом. Данный конвейер фактически не поддерживается, поэтому его использование для реализации настоящего проекта нецелесообразно. Биоинформатический конвейер DNAr также устарел. Кроме того, желательная платформа для его запуска - Centos 7, в то время как в нашем случае используется Debian. Биоинформатический конвейер DNAscan2 также устарел: последние обновления версий сделаны более трёх лет назад.

Также важно отметить, что помимо перечисленного выше, все проанализированные биоинформатические конвейеры могут только маркировать оптические дубликаты, но не обладают способностью удалять их. Более того, в проанализированных конвейерах нет настраиваемых фильтров пост-процессированных вариантов, а также перекрёстной проверки SNV и вставок/делеций между несколькими программами по определению вариантов с целью повышения точности. Кроме того, в некоторых биоинформатических конвейерах отсутствует возможность выявления CNV и даже герминальных SNV.

Ввиду несоответствия существующих биоинформатических конвейеров выдвигаемым нами требованиям, было принято решение о разработке собственных биоинформатических конвейеров для определения герминального и соматического геномного профиля, обеспечивающих высокое качество анализа.

3.3.5. Разработанные биоинформатические конвейеры для определения герминального и соматического геномного профиля

Нами были разработаны два оригинальных биоинформатических конвейера: биоинформатический конвейер для определения герминального геномного профиля (герминальный биоинформатический конвейер) и биоинформатический конвейер для определения соматического геномного профиля в паре опухоль-норма (соматический биоинформатический конвейер). В настоящем исследовании разработанные конвейеры были апробированы на нашей собственной выборке из 40 образцов РПЖ и 100 образцов ТНРМЖ для определения их геномного профиля.

Как герминальный, так и соматический биоинформатический конвейер структурно включают 3 аналитических блока: блок предобработки, блок выявления вариантов и блок постпроцессинга и формирования отчетов. Однако в рамках герминального и соматического биоинформатических конвейеров имеются как аналогичные, так и отличающиеся этапы в рамках названных блоков.

В качестве модели хранилища было решено использовать хранилище Amazon S3. Полученные из секвенатора данные в формате FASTQ загружали в специальный S3 бакет, хранящий образцы для дальнейшей обработки. Герминальный биоинформатический конвейер получает на вход герминальные FASTQ файлы. В соматический биоинформатический конвейер на вход поступают герминальные и соматические FASTQ файлы для одновременного биоинформатического анализа образцов периферической крови и образцов опухолевой ткани и выявления опухоль-специфических соматических геномных вариантов. Все этапы, осуществляемые в рамках блока предобработки, и используемые для предобработки программные инструменты сходны как в герминальном, так и в соматическом биоинформатических конвейерах.

Разработанные биоинформатические конвейеры отличаются на этапе выявления вариантов (генерации VCF файлов) как методикой реализации, так и набором используемых программных инструментов. На этапе генерации отчетов о выявленных вариантах конвейеры также отличаются используемыми программными инструментами. Кроме того, отличия между конвейерами состоят в том, что в рамках соматического биоинформатического конвейера производится выявление показателей TMB и MSI. При выявлении CNV в конвейерах также используются разные инструменты. Структура разработанных биоинформатических конвейеров представлена на рисунке 22. Изображённая схема служит для визуализации архитектуры двух разработанных биоинформатических конвейеров, герминального и соматического. Представленная диаграмма позволяет чётко проследить как этапы, общие для обоих конвейеров,

так и их ключевые различия. Это способствует лучшему пониманию методологических подходов, используемых в разных аналитических сценариях.

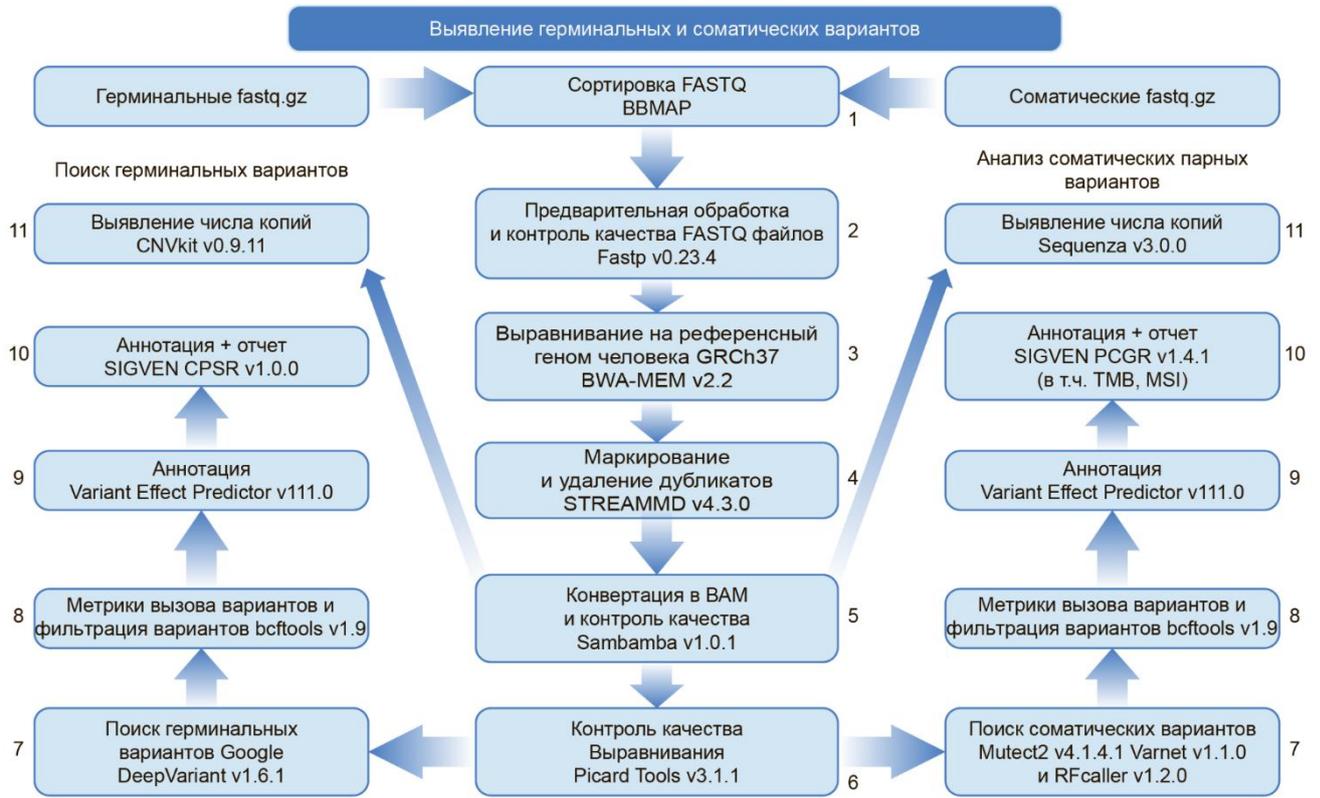


Рисунок 22. Структура разработанных биоинформатических конвейеров для определения герминального и соматического геномного профиля. Для выявления опухоль-специфических соматических геномных вариантов происходит одновременный биоинформатический анализ образцов периферической крови и образцов опухолевой ткани. Блок предобработки (этапы 1-6) является аналогичным для обоих конвейеров. Блок выявления геномных вариантов и CNV, а также блок постпроцессинга и формирования отчетов отличаются между конвейерами.

Функционирование разработанных герминального и соматического парного биоинформатических конвейеров продемонстрировано далее на примере образцов биологического материала больной ТНPMЖ. Биологический материал больной ТНPMЖ был представлен образцами 8928UQFZ – экзом крови (нормальная ткань) и 7270FSIA – экзом опухоли (соматическая ткань). Использовали зонды Roche KAPA HyperExome. До этапа выявления геномных вариантов структура биоинформатических конвейеров (этапы 1-6) совпадает для герминального и соматического парного биоинформатических конвейеров.

Этап 1. Сортировка FASTQ-файлов.

Процесс: Исходные данные в формате FASTQ сортируются для оптимизации последующей обработки. Сортировка упорядочивает чтения для корректного выполнения выравнивания и анализа. Используется в случае необходимости “склейки” FASTQ из разных запусков и

упрощения выравнивания. Инструмент: Используется BБMap Tools для эффективной и точной сортировки файлов FASTQ (Рисунок 23). На рисунке приведены размеры FASTQ-файлов для опухолевого и герминального образцов, что позволяет оценить объём входных данных и требования к вычислительным ресурсам. Демонстрация таких метаданных важна для обоснования выбора инструментов и параметров обработки.

sort_7270FSIA_Roche_KAPA_HyperExome_MG_E200_SOD_a43_L01_44PE100_1.fq.gz	8 812 592
sort_7270FSIA_Roche_KAPA_HyperExome_MG_E200_SOD_a43_L01_44PE100_2.fq.gz	8 768 538
sort_8928UQFZ_Roche_KAPA_HyperExome_MG_E100_GED_a3A_L03_15PE100_1.fq.gz	3 433 077
sort_8928UQFZ_Roche_KAPA_HyperExome_MG_E100_GED_a3A_L03_15PE100_2.fq.gz	3 388 978

Рисунок 23. Пары FASTQ для опухолевого образца 8,8 Гб + 8,7Гб и 3,4Гб + 3,4 Гб у герминальной пары

Этап 2. Предварительная обработка и контроль качества FASTQ файлов.

Процесс: Удаляются адаптеры и основания низкого качества из прочтений. Этот этап снижает смещения при последующем выравнивании и повышает качество выравнивания. Файлы в виде разархивированных FASTQ передаются напрямую в утилиту выравнивания BWA2. Инструмент: Программа Fastp v0.23.4, которая также выполняет фильтрацию чтений по качеству (Рисунок 24).

Before filtering		Before filtering	
total reads:	205.926062 M	total reads:	79.741982 M
total bases:	20.592606 G	total bases:	7.974198 G
Q20 bases:	20.383649 G (98.985281%)	Q20 bases:	7.909301 G (99.186157%)
Q30 bases:	19.838143 G (96.336243%)	Q30 bases:	7.723865 G (96.860710%)
GC content:	47.820843%	GC content:	47.408570%
After filtering		After filtering	
total reads:	205.923442 M	total reads:	79.741982 M
total bases:	20.370403 G	total bases:	7.969410 G
Q20 bases:	20.168570 G (99.009186%)	Q20 bases:	7.904764 G (99.188828%)
Q30 bases:	19.629174 G (96.361243%)	Q30 bases:	7.719477 G (96.863848%)
GC content:	47.754968%	GC content:	47.405352%
Filtering result		Filtering result	
reads passed filters:	205.923442 M (99.998728%)	reads passed filters:	79.741982 M (100.000000%)
reads corrected:	3.157000 M (1.533075%)	reads corrected:	147.976000 K (0.185569%)
bases corrected:	4.426714 M (0.021497%)	bases corrected:	189.309000 K (0.002374%)
reads with low quality:	0 (0.000000%)	reads with low quality:	0 (0.000000%)
reads with too many N:	2.062000 K (0.001001%)	reads with too many N:	0 (0.000000%)
reads too short:	558 (0.000271%)	reads too short:	0 (0.000000%)

А

Б

Рисунок 24: А) Образец опухолевой пары FASTQ, 99% ридов прошли контроль качества; Б) Образец герминальных FASTQ, прошло фильтры 100% ридов.

Данный рисунок иллюстрирует результаты контроля качества ридов после предварительной обработки. Высокий процент прошедших фильтрацию ридов свидетельствует о высоком качестве исходных данных и минимизации ошибок при дальнейшем выравнивании, что критично для точности всего анализа.

Этап 3. Выравнивание на референсный геном человека GRCh37.

Процесс: Обрезанные чтения выравниваются по референсному геному для определения их положения. Результат сохраняется в формате SAM/BAM. Инструмент: BWA-MEM v2.2. обеспечивает высокую точность и эффективность при работе с короткими чтениями. Образец передаётся напрямую по конвейеру в STREAMMD и SAMBAMBA.

Этап 4. Маркирование и удаление дубликатов.

Процесс: Идентифицируются и помечаются дубликаты, возникшие в результате ПЦР. Удаление дубликатов снижает смещения в анализе вариантов. Инструмент: STREAMMD v4.3.0 используется для маркировки и удаления дубликатов из SAM-файла (Рисунок 25). На рисунке отображены результаты удаления дубликатов, включая процент оптических дубликатов в образцах. Разница между опухолевыми и нормальными образцами подчёркивает важность этого этапа для обеспечения надёжности интерпретации вариантов.

<pre>BloomFilter at 6.41% capacity Estimated marginal false positive rate: 4.40e-18 alignments seen: 206692239 alignments marked duplicate: 78536762 templates seen: 102961721 templates unmapped: 25826 templates marked duplicate: 39189295 template duplicate fraction: 0.3807</pre>	<pre>BloomFilter at 3.83% capacity Estimated marginal false positive rate: 2.63e-20 alignments seen: 79766575 alignments marked duplicate: 3590210 templates seen: 39870991 templates unmapped: 10790 templates marked duplicate: 1794787 template duplicate fraction: 0.0450</pre>
---	---

А

Б

Рисунок 25. Результат работы программы STREAMMD (А) Отображено наличие 38% оптических дубликатов (duplicate fraction) в образце опухоли и Б) 4% (duplicate fraction) в норме.

Этап 5. Конвертация в BAM и контроль качества.

Процесс: Конвертация файла из формата SAM в более компактный и удобный формат BAM. Одновременно собираются метрики контроля качества для проверки целостности данных. Данный блок имеет распределённое исполнение внутри конвейера без промежуточных файлов, т.е. обработка идёт в памяти. Инструмент: Sambamba v1.0.1 обеспечивает быструю конвертацию с возможными заданными порогами фильтров, например `sambamba --filter "mapping_quality >=`

20 and not failed_quality_control" и преобразует файлы SAM, находящиеся внутри пайплайна, в заархивированную версию BAM (Рисунок 26).



7270FSIA_dedup.bam	63 610 967
8928UQFZ_dedup.bam	24 340 341

Рисунок 26. Визуализация выровненных и отсортированных BAM-файлов с указанием размера (в данном случае 63,6 и 24,3 Гб)

Рисунок демонстрирует размеры выровненных BAM-файлов после сортировки и конвертации. Эти параметры позволяют судить об объёме данных, покрытии, а также потенциальных различиях в сложности и глубине секвенирования.

Этап 6. Контроль качества выравнивания.

Процесс: Обработка входных BAM-файлов вместе с *.bed файлом зондов. Инструмент: Picard Tools v3.1.1. на основе входного *.bed файла (в данном случае KAPA HyperExome) оценивает до 68 метрик (Рисунок 27).

Sample	7270FSIA	8928UQFZ
MEAN_BAIT_COVERAGE	296.064281	102.869099
FOLD_ENRICHMENT	47.115984	41.79588
MEDIAN_TARGET_COVERAGE	117	89
PCT_TARGET_BASES_30X	0.915029	0.899846

Рисунок 27. В настоящей работе интересующие нас метрики - это среднее покрытие по цели (Mean bait coverage), целевое обогащение зондами (fold enrichment), медианное покрытие цели (Median target coverage) и полнота покрытия 30x (PCT target bases)

На рисунке 27 изображены метрики выравнивания, рассчитанные с помощью Picard Tools. Особый интерес представляют показатели обогащения зондами и полноты покрытия, которые являются индикаторами эффективности подготовки библиотек и точности целевого секвенирования.

После завершения блока предобработки, начиная с этапа 7 (выявление геномных вариантов) биоинформатические конвейеры начинают различаться: в парном соматическом конвейере используются оба BAM-файла, а в герминальном - BAM-файл от образца "нормы".

Этапы 7 - 11, реализуемые в рамках разработанного герминального биоинформатического конвейера, представлены далее.

Этап 7. Поиск герминальных вариантов.

Процесс: На основе выравниваний (BAM-файла) идентифицируются геномные варианты. Программа по выявлению вариантов DeepVariant v1.6.1 использует глубокое обучение для повышения точности анализа. Инструмент: Google DeepVariant v1.6.1 генерирует VCF-файл с обнаруженными вариантами. Пример, обнаружение мутации EHP1 (2:63131731G>A) и BRCA2 (13:32911297TAAAC>T) (Рисунок 28). Приведены примеры мутаций, зафиксированных в VCF-файле, включая потенциально патогенные варианты. Данный рисунок важен для иллюстрации эффективности используемого инструментария по выявлению значимых SNV.

2	62067433	.	T	C	50.4	PASS	.	GT:GQ:DP:AD:VAF:PL	0/1:46:117:57,60:0.512821:50,0,47
2	62081012	.	A	C	71.9	PASS	.	GT:GQ:DP:AD:VAF:PL	1/1:67:93:0,93:1:71,69,0
2	62100380	.	G	T	29.8	PASS	.	GT:GQ:DP:AD:VAF:PL	0/1:27:111:60,51:0.459459:29,0,29
2	62729674	.	T	C	60.4	PASS	.	GT:GQ:DP:AD:VAF:PL	1/1:58:93:0,93:1:60,60,0
2	63131731	.	G	A	34.7	PASS	.	GT:GQ:DP:AD:VAF:PL	0/1:34:82:42,40:0.487805:34,0,42
2	65245365	.	G	A	63.8	PASS	.	GT:GQ:DP:AD:VAF:PL	1/1:60:63:0,63:1:63,62,0
2	65296798	.	G	A	63	PASS	.	GT:GQ:DP:AD:VAF:PL	1/1:60:79:1,78:0.987342:62,62,0
13	32776603	.	T	C	67.2	PASS	.	GT:GQ:DP:AD:VAF:PL	1/1:65:107:0,107:1:67,69,0
13	32785086	.	G	A	56.7	PASS	.	GT:GQ:DP:AD:VAF:PL	1/1:54:112:2,110:0.982143:56,56,0
13	32885654	.	T	C	37.4	PASS	.	GT:GQ:DP:AD:VAF:PL	0/1:34:95:62,33:0.347368:37,0,37
13	32885658	.	G	A	42.6	PASS	.	GT:GQ:DP:AD:VAF:PL	0/1:35:99:64,35:0.353535:42,0,35
13	32911297	.	TAAAC	T	50.4	PASS	.	GT:GQ:DP:AD:VAF:PL	0/1:45:97:49,48:0.494845:50,0,46
13	32912299	.	T	C	70.1	PASS	.	GT:GQ:DP:AD:VAF:PL	1/1:67:98:2,96:0.979592:70,70,0
13	32913055	.	A	G	66.2	PASS	.	GT:GQ:DP:AD:VAF:PL	1/1:64:113:0,113:1:66,67,0

Рисунок 28. Срезы VCF-файлов, показывающих наличие патогенных полиморфизмов

Этап 8. Фильтрация вариантов.

Процесс: Фильтрация по основным параметрам PASS, глубина покрытия DP>10, качество генотипа GQ>30. Инструмент: bcftools v.1.9.

Этап 9. Аннотация Variant Effect Predictor.

Процесс: Варианты аннотируются для предоставления функционального и клинического контекста. Это помогает интерпретировать результаты. Инструмент: Ensembl Variant Effect Predictor v111.0 используется для аннотации герминальных и соматических вариантов (Рисунок 29).

index	Allele	Consequence	IMPACT	SYMBOL	Gene	Feature_type
bigint	string	string	string	string	string	string
Integer	Text	Text	Text	Text	Text	Text
13956	C	missense_variant	MODERATE	BRCA2	ENSG00000139618	Transcript

Рисунок 29. Часть таблицы аннотации VEP (полная версия находится в озере данных)

Показан фрагмент аннотации с использованием Ensembl VEP, который демонстрирует биологическую значимость выявленных вариантов. Такой уровень аннотации необходим для последующей клинической интерпретации.

Этап 10. Аннотация SIGVEN CPSR.

Процесс: Варианты аннотируются для предоставления отчёта. Это помогает представлять результаты в должном виде. Инструмент: SIGVEN CPSR v1.0.0. используется для аннотации герминальных вариантов (Рисунок 30). Пример отчета SIGVEN CPSR, представленный на рисунке, отражает стандартизированный подход к представлению информации о герминальных вариантах, включая потенциальную клиническую значимость.

	Position in the genome	GENE	Genotype	cDNA substitution	A/C replacement	Exon	Transcript	Frequency by gnomAD	Depth of read
+	1 2:63131731G>A	EHBP1	o/1						82
+	2 13:32911297TAAAC>T	BRCA2	o/1	c.2808_2811del	p.A938fs	exon11		7.973e-06	97

Рисунок 30. Часть отчёта SIGVEN CPSR.

Этап 11. Анализ CNV (CNVkit для единичных герминальных образцов)

Используются подобные метрики, \log_2 - значимые изменения копийности, p_ttest - статистическая достоверность, $depth$ покрытие региона, $probes$ - минимальное число маркеров $cn = 0$ или 4 - явные изменения копийности. Инструмент: CNVkit 0.9.11 (Рисунок 31). Демонстрируется выходной файл CNVkit, в котором представлены параметры, используемые для анализа изменений копийности. Такие данные играют важную роль в интерпретации структурных геномных изменений.

chromosome	start	end	gene	log2	cn	depth	p_ttest	probes	weight
1	65564	878438	OR4F5,OR4F3,OR4F2	0.507577	3	104.341	0.0125234	25	19.6637
1	878632	1337981	SAMD11,NOC2L,KLF1	-0.422159	1	78.1544	0.0074083	264	193.499
1	1340999	2100432	MRPL20,ANKRD65,ANKRD66	0.0169139	2	84.4448	0.689109	244	179.15
1	2100932	2235542	PRKCZ,PRKCZ,FAAF1	-1.28687	0	68.9828	0.24413	24	18.4122
1	2235731	6206483	SKI,MORN1,RER1,P	-0.0349446	2	78.5656	0.377659	403	300.177
1	6206724	11015243	CHD5,RPL22,RNF20B	-0.11467	2	83.5333	0.0295377	668	493.154
1	11017079	11087450	C1orf127,TARDBP,PR	-1.11514	0	100.355	0.384441	22	16.1758
1	11087450	12820884	MASP2,SRM,EXOSC1	-0.0604923	2	85.2858	0.0014281	450	332.361
1	12821384	13216968	PRAMEF12,PRAMEF13	0.917322	4	167.442	8.34039e-06	76	60.8131
1	13216968	13387543	PRAMEF18,PRAMEF19	-0.315507	1	148.863	0.895932	18	14.6791
1	13388457	13447549	PRAMEF18,PRAMEF19	0.186551	2	129.198	0.375502	15	12.1505

Рисунок 31. Фрагмент файла, выдаваемого CNVkit

Этапы 7 - 11, реализуемые в рамках разработанного соматического парного биоинформатического конвейера представлены далее.

Этап 7. Поиск соматических вариантов в парных образцах.

Процесс: Анализ выполняется для пар "опухоль-норма", что позволяет точно выявлять соматические варианты за счёт сравнения с нормальным образцом. Инструменты: Mutect2 v4.1.4.1 используется без Base Recalibration, так как она может увеличить уровень шума при анализе пар, а для инструментов машинного обучения Varnet v1.1.0 и RFcaller v1.2.0 она не рекомендуется разработчиками ПО (Рисунки 32 и 33).

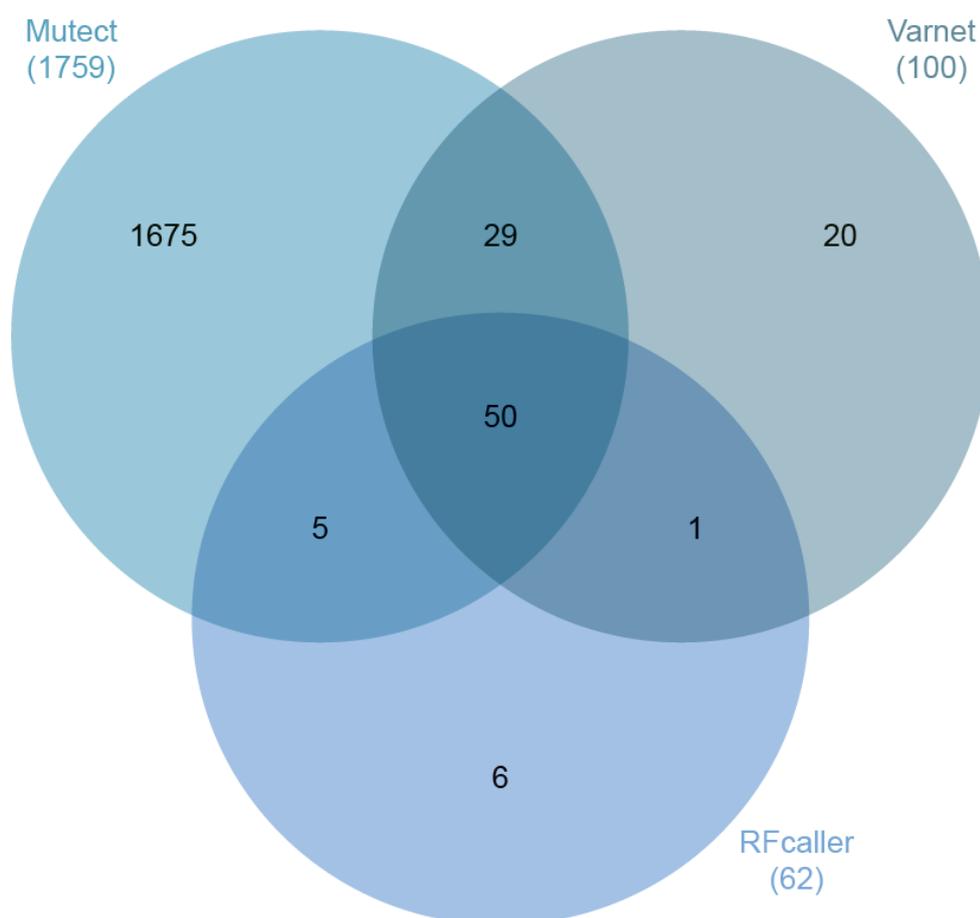


Рисунок 32. Результаты кросс-валидационной проверки между тремя программами по выявлению геномных вариантов, используемыми в настоящем исследовании (Mutect2 v4.1.4.1, RFcaller v1.2.0 и Varnet v1.1.0.) при анализе биологического материала больной ТНРМЖ. Диаграмма Венна показывает пересечения в виде общих SNV между каждой из программ по выявлению геномных вариантов

Диаграмма Венна отображает перекрытия SNV между тремя различными инструментами для выявления вариантов. Она служит доказательством согласованности результатов и повышает доверие к полученным данным.

[Mutect] and [Varnet] and [RFcaller]:
 2237649
 26090442
 110051450
 248756178
 17697399
 160268929
 179826072
 196883945
 210569231
 220172271
 233674441
 32933057
 119156842
 79240119
 104012402
 122836062
 141336569
 147781990
 178039456
 39033979
 75804872
 6456296
 40027637
 82546120
 99051639
 99756078
 128482362
 67894392
 87443910
 95174359
 35308061
 35673964
 135053586
 40265629
 52911444
 57605068
 132404559
 103390972
 75573212
 79750498
 2229834
 83251005
 7577539
 48255878
 6686766
 16548675
 37482162
 56002274
 47850141
 107424174

Рисунок 33. Перечень из 50 SNV, которые являются общими для всех 3 программ по выявлению вариантов. Вариант в гене TP53 (17:7577539) выявлен всеми 3 программами.

На рисунке представлен перечень общих SNV, выявленных всеми тремя инструментами. Такой подход демонстрирует воспроизводимость и валидность полученных результатов.

Этап 8. Фильтрация вариантов.

Процесс: Фильтрация по основным параметрам PASS, глубина покрытия DP>10, качество гено типа GQ>30. Инструмент: bcftools v.1.9

Этап 9. Аннотация Variant Effect Predictor.

Процесс: Варианты аннотируются для предоставления функционального и клинического контекста. Это помогает интерпретировать результаты. Инструмент: Ensembl Variant Effect Predictor v. 111.0 используется для аннотации герминальных и соматических вариантов (Рисунок 34).

SYMBOL	rs	REF	ALT	CLN_SIG	CLINVAR	BIOTYPE_TIER	VARIANT_CLASS
string Text	string Text	string Text	string Text	string Text	string Text	string Text	string Text
TP53	rs121912651			Pathogenic	germline,somatic	TIER 1	SNV

Рисунок 34. Фрагмент аннотации в Ensembl VEP (полная версия находится в озере данных)

Фрагмент аннотации соматических вариантов демонстрирует биологическую и клиническую интерпретацию мутаций, выявленных в опухолевой ДНК. Подобная информация критична для формирования персонализированных рекомендаций.

Этап 10. Аннотация SIGVEN PCGR.

Процесс: Варианты аннотируются с целью интерпретации их воздействия, включая возможные ассоциации с онкологическими заболеваниями. Также происходит определение уровня TMB и микросателлитного статуса. Инструменты: Используется SIGVEN PCGR v1.4.1 (Рисунок 35).

Классы патогенности определены согласно стандартам и руководствам AMP/ASCO/CAP 2017 (Ассоциация молекулярной патологии/Американское общество клинической онкологии, Колледж американских патологов; англ. Association for Molecular Pathology, American Society of Clinical Oncology, College of American Pathologist) [215]. Рекомендации AMP/ASCO/CAP 2017 предлагают систему классификации соматических генетических вариантов в онкологии по четырем уровням («Tier»), основанную на их клинической значимости и надежности доказательств.

Tier I — Strong clinical significance (Высокое клиническое значение)

- Подтверждено одобренными FDA показаниями для этого типа опухоли.
- Поддержано профессиональными рекомендациями (например, NCCN).
- Имеются данные из клинических испытаний с доказанной пользой.

Tier II — Potential clinical significance (Потенциальное клиническое значение)

- Одобрения или данные получены для другого типа опухоли.
- Имеются данные из небольших клинических исследований или ретроспективных анализов.

- Возможно включение в клинические исследования.

Tier III — Unknown clinical significance (Неизвестное клиническое значение)

- Варианты описаны в литературе, но без четкого клинического значения.
- Присутствуют в базах данных, но с ограниченными/противоречивыми данными.

Tier IV — Benign or likely benign (Доброкачественный или вероятно доброкачественный вариант)

- Часто встречаются в популяционных базах данных.
- Нет связи с канцерогенезом по текущим данным.

Таким образом, перечисленные уровни устанавливаются на основе типа доказательств: от клинических руководств и регуляторных одобрений до экспериментальных или наблюдательных исследований.

Варианты с более высоким уровнем Tier (например, Tier I) требуют более весомых доказательств, таких как одобрение FDA или включение в клинические руководства. Варианты с более низким уровнем Tier (например, Tier II или III) могут основываться на ограниченных данных или предварительных исследованиях. Данная система используется в PCGR, позволяя стандартизировать интерпретацию генетических вариантов в онкологии, обеспечивая более точное и последовательное принятие клинических решений.

	SYMBOL	CONSEQUENCE	PROTEIN_CHANGE	VARIANT_CLASS	TIER
+	1 TP53	missense_variant	p.Arg248Trp	SNV	TIER 1
+	2 CDH17	missense_variant	p.Asn438Lys	SNV	TIER 3
+	3 CDH13	missense_variant	p.Val227Gly	SNV	TIER 3
+	4 CA9	missense_variant	p.Pro3Leu	SNV	TIER 4
+	5 C3	missense_variant	p.Thr1213Ala	SNV	TIER 4
+	6 SKI	missense_variant	p.Gly653Asp	SNV	TIER 4
+	7 MAN1C1	missense_variant	p.Ser375Arg	SNV	TIER 4
+	8 AMIGO1	missense_variant	p.Arg29Gly	SNV	TIER 4
+	9 OR2T10	missense_variant	p.Thr298Pro	SNV	TIER 4
+	10 RAD51AP2	missense_variant	p.Asp762Tyr	SNV	TIER 4

Рисунок 35. Отчет PCGR. Отображены выявленные соматические геномные варианты с указанием классов патогенности (Tier). Визуализирован отчет PCGR, включающий классификацию вариантов по уровням клинической значимости (Tier I–IV), что обеспечивает основу для принятия врачебных решений в соответствии с международными стандартами.

При анализе пары “опухоль-норма” соматическим биоинформатическим конвейером у больной ТНРМЖ был обнаружен патогенный соматический вариант в гене *TP53* всеми тремя программами по выявлению геномных вариантов. Данный подход многократно увеличивает точность анализа. Выявленный патогенный вариант в гене *TP53* включается в финальный отчет.

Этап 11. Анализ CNV (Sequenza для парных образцов).

Процесс: Выявление и анализ CNV в одиночных и парных образцах. Этот этап позволяет обнаруживать крупные генетические изменения, такие как делеции или дупликации.

Инструменты: Sequenza v3.0.0 используется для детекции и интерпретации CNV, предоставляя дополняющие друг друга подходы.

Эти этапы обеспечивают комплексный анализ герминальных и соматических вариантов, что необходимо для полноценной интерпретации генетических данных из парных образцов. В данном случае нам представлена таблица с числом копий CNt, которую можно отфильтровать по различным критериям, таким как по глубине покрытия (depth.ratio), BAF (B-allele frequency), копийности (CNt) и стандартному отклонению (Sd.ratio) (Рисунок 36). Приведен фрагмент анализа CNV при помощи Sequenza. Показаны метрики, по которым оцениваются изменения числа копий, включая BAF и depth ratio. Эти параметры важны при комплексной интерпретации профиля опухоли.

chromosome	start.pos	end.pos	Bf	N.BAF	sd.BAF	depth.ratio	N.ratio	sd.ratio	CNt	A	B	LPP
10	42356130	42398211	0.112953688025718	125	0.15310156	8.02546002989725	3363	8.65010371058761	20	18	2	-6.73303219814439
10	42408283	42532576	0.488971825193513	92	0.15669097	8.91672898624917	1988	9.06476513374276	20	10	10	-6.73600079463504
10	38464910	38465037	0.33653436446642	12	0.02557994	4.51082773112402	127	0.690165638668646	18	12	6	-6.74001301591591
19	27736559	28570587	0.0751398830714063	26	0.14116796	4.28577773187858	4269	6.58398656634866	17	16	1	-6.73658639715672
19	27731918	27736551	0.429481075052341	41	0.16000345	5.21972541921253	2437	7.83140681489122	14	8	6	-6.73315906922651
12	37969194	38016436	0.104700634097758	24	0.11256050	1.72337871225111	844	3.25185274994172	12	11	1	-6.73648538976082
11	1092557	1093512	0.369493113626417	14	0.09995782	2.11462481182623	955	0.792037658166766	11	7	4	-6.73390903142867
18	18511077	18520333	0.369694650625927	275	0.13915211	4.12322208109337	4699	3.61082107615324	11	7	4	-6.73313648393562
10	38978243	39128940	0.0310764063827683	9	0.01907514	1.79008433118676	3640	2.70252312549293	10	10	0	-6.74361808466728
12	34403122	34846492	0.22485221340565	13	0.12820191	3.79483766114443	441	1.19409671833033	10	8	2	-6.73990805714355
11	51426088	51426303	0.137759641466212	10	0.08256236	2.20013514010072	215	0.649797699045519	9	8	1	-6.73314927770639
11	51426624	51579367	0.448478781907594	32	0.04722568	1.29550370023198	6984	3.2606161378518	9	5	4	-6.74438729649838

Рисунок 36. Фрагмент файла, выдаваемого Sequenza.

Разработанные биоинформатические конвейеры соответствовали лучшим практикам построения биоинформатических конвейеров GATK Института Броуда. Вместе с тем, разработанные конвейеры были модифицированы для обеспечения высокого качества анализа при использовании секвенатора MGISEQ-G400 (MGI, Китай).

3.3.6. Интерфейс рабочей зоны управления биоинформатическими расчетами

Был разработан интерфейс рабочей зоны управления биоинформатическими расчетами, осуществляемыми герминальным и соматическим биоинформатическими конвейерами. Разработанный интерфейс позволяет специалисту, осуществляющему биоинформатическую обработку данных, получать информацию о статусе обработки, а также управлять параметрами расчета образцов.

На рисунках 37 и 38 изображено главное окно управления расчетами в герминальном и соматическом биоинформатическом конвейерах, соответственно. Показаны элементы управления, которые позволяют осуществлять запуск биоинформатических расчетов, устанавливать параметры расчетов и осуществлять мониторинг статуса процесса в режиме реального времени для каждого образца. Рисунок 37 отображает пользовательский интерфейс системы управления герминальными расчетами. Интерфейс предоставляет специалисту возможность контроля статуса обработки каждого образца, обеспечивая прозрачность биоинформатического процесса.

Name	Type	Step	Last processing date
4112NDOH-E100-Roche_KAPA_HyperExome	GED	DOWNLOAD-MERGE	18.12.2024
5721WBGT-E100-Roche_KAPA_HyperExome	GED	DOWNLOAD-MERGE	18.12.2024
1313RPIG-E100-Roche_KAPA_HyperExome	GED	DOWNLOAD-MERGE	
3725MHHV-E200-Roche_KAPA_HyperExome	SOD	DOWNLOAD-MERGE	
8236RUWT-E200-Roche_KAPA_HyperExome	SOD	DOWNLOAD-MERGE	
6912MKPT-E200-Roche_KAPA_HyperExome	SOD	DOWNLOAD-MERGE	

Stop	Name	Started at	In progress	Current step	Logs	Report
⏸	5721WBGT	2024-12-18 15:28:48	00:00:00			
⏸	4112NDOH	2024-12-18 15:28:48	00:00:00			

Рисунок 37. Окно визуализации процесса расчёта герминальных образцов

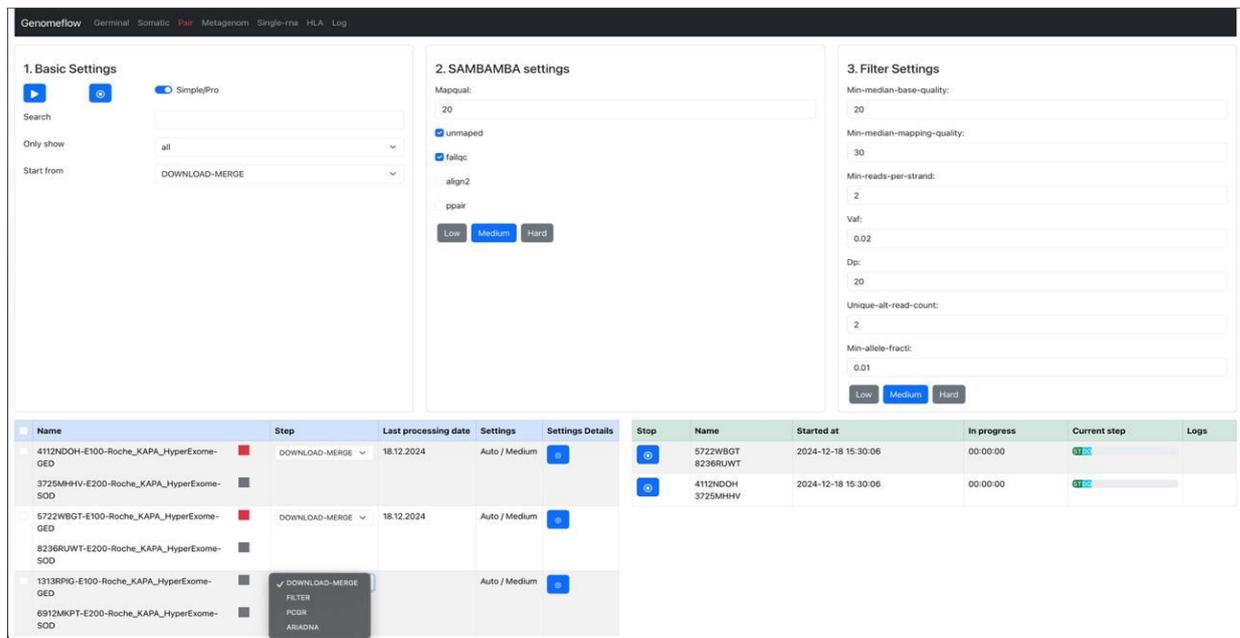


Рисунок 38. Окно визуализации процесса расчёта соматических образцов

На рисунке 38 приведено окно интерфейса для работы с соматическими образцами. Возможность параллельной обработки сразу нескольких пар «опухоль-норма» существенно ускоряет анализ и повышает эффективность использования ресурсов. Для оптимизации времени расчёта его осуществляли для нескольких образцов одновременно. В нижней левой части рисунка 38 можно видеть, что в расчете находились одновременно 3 парных образца. В разделе “PostUpload” указывали директорию исходных данных. Данная директория является корневой для всех расчетов текущего набора данных, алгоритм расчета образцов в качестве входных брал данные из указанной директории, а результаты всех вычислений складывались в данную папку. Для того чтобы запустить обсчёт образцов выбирали тип исследования, исходную директорию набора данных, шаг, с которого система начнет обрабатывать образцы (по умолчанию - с первого шага), а далее выбирали из списка доступных образцов подмножество для расчёта (Рисунок 39А), а затем выбирали параметры запуска и нажимали кнопку запуска расчета (Рисунок 39Б), причем перед запуском возможен выбор различных параметров запуска. Гибкая настройка параметров и выбор подмножества образцов обеспечивает точную адаптацию процесса под конкретные задачи.

<input checked="" type="checkbox"/>	Name	Step	Last processing date	Settings	Settings Details
<input checked="" type="checkbox"/>	4112NDOH-E100-Roche_KAPA_HyperExome-GED 3725MHHV-E200-Roche_KAPA_HyperExome-SOD	DOWNLOAD-MERGE	18.12.2024	Auto / Medium	
<input checked="" type="checkbox"/>	5722WGBT-E100-Roche_KAPA_HyperExome-GED 8236RUWT-E200-Roche_KAPA_HyperExome-SOD	DOWNLOAD-MERGE	18.12.2024	Auto / Medium	
<input checked="" type="checkbox"/>	1313RPIG-E100-Roche_KAPA_HyperExome-GED 6912MKPT-E200-Roche_KAPA_HyperExome-SOD	<input checked="" type="checkbox"/> DOWNLOAD-MERGE <input type="checkbox"/> FILTER <input type="checkbox"/> PCGR <input type="checkbox"/> ARIADNA		Auto / Medium	

А

1. Basic Settings

Search:

Only show:

Start from:

Preset:

Preset flex:

Enddistance: Enddistancemad: Zeroproportion: Minmapqualitydifference:

Minmapquality: Minbasequality: Minaltcount: Maxaltcount:

StrandbiassimpleRepeats: Snvcluster50: Snvcluster100:

Minimumdepth: Maximumdepth: Editdistance: Maxvafnormal:

Minvafumor: Maxsecondumot: Maxbadorient: Foxog:

2. SAMBAMBA settings

Mapqual:

unmaped

failqc

align2

ppair

3. Filter Settings

Min-median-base-quality:

Min-median-mapping-quality:

Min-reads-per-strand:

Vaf:

Dp:

Б

Рисунок 39. А) Интерфейс выбора необходимых образцов для расчета соматических образцов; Б) Внешний вид кнопки запуска расчета соматических образцов (представлена в верхней левой части)

Расчёт одного образца обычно требует большого количества вычислительных и временных ресурсов. Для минимизации вычислительных и временных затрат, при технических сбоях или некорректности исходных данных, расчёт каждого образца разделен на отдельные этапы. При возникновении сбоев пересчитываются только те шаги, на которых произошли сбои. Возможность повторного расчета доступна как для группы выбранных образцов, так и для

каждого образца в отдельности. В системе также существует возможность задавать произвольную начальную позицию расчета для каждой записи из таблицы образцов (Рисунок 39А).

Разработанный интерфейс позволяет отслеживать статус выполнения для каждого образца на каждом этапе биоинформатического анализа. Напротив всех образцов, которые находятся в процессе расчёта, визуализируется этап, на котором находится образец. При возникновении ошибки, соответствующий этап помечался красным цветом, корректно выполненные этапы - зеленым. На рисунке 40 изображён механизм визуального отображения прогресса расчёта и возникающих ошибок; цветовая индикация позволяет оперативно выявлять сбои и повторно запускать только проблемные этапы.

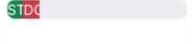
Stop	Name	Started at	In progress	Current step	Logs
	5722WBGT 8236RUWT	2024-12-18 15:30:06	00:00:52		• DOWNLOAD-MERGE
	4112NDOH 3725MHHV	2024-12-18 15:30:06	00:00:52		• DOWNLOAD-MERGE

Рисунок 40. Внешний вид окна, предоставляющего информацию об этапе расчета, на котором находится соматический образец, а также о возникновении ошибок на определенных этапах (отмечены красным цветом)

Для оптимизации времени расчёта “рана” (выполнение определенного скрипта или программы для выравнивания последовательностей ДНК) использовали диспетчер задач Apache Airflow. Хранение и обработку генетических данных осуществляли на серверах с высокими вычислительными характеристиками. Современные сервера позволяют обрабатывать несколько образцов параллельно, что позволяет существенно экономить время при обработке “рана”. Параллельная обработка образцов требует разработки соответствующей параллельной архитектуры со следующими требованиями:

- Оптимальное использование вычислительных ресурсов
- Надежность
- Оптимизация времени выполнения расчётов.

Основная сложность в проектировании многопоточного диспетчера состояла в том, чтобы по вычислительным характеристикам серверного оборудования оптимизировать число образцов, которые обрабатываться параллельно. Работа диспетчера состояла из следующих шагов:

- Анализ вычислительных характеристик серверного оборудования;
- Построение таблицы запусков. Данная таблица описывает какое оптимальное количество образцов возможно запустить на каждом шаге обработки. На данном шаге нужно учитывать две особенности. Указанное количество образцов слишком маленькое. В этом

случае ресурсы сервера используются не в полном объёме и время расчета “рана” может быть в несколько раз больше оптимального. Подразумевается, что некоторые процессы не могут задействовать все ядра в системе или не всю память, тогда, запуская слишком маленькое количество образцов, например 1, на сервере с 100 ядрами, мы всё равно получим производительность равную одному ядру на единицу времени (как, например, Mutect2), при том что шаг выравнивания (BWA2), может задействовать все ядра в системе.

- Запуск образцов в соответствии с состоянием таблицы.

Разработка собственного диспетчера расчетов позволила в среднем сократить время расчета одного “рана” с 7 дней до 3 дней.

3.3.7. Система поддержки принятия врачебных решений для подбора персонализированного лекарственного лечения онкологических заболеваний

Для подбора эффективного персонализированного лекарственного лечения ЗНО на основе геномных вариантов, а также показателей ТМВ, выявляемых у больных, и формирования отчетов для использования лечащим врачом нами была разработана СППВР.

Структурно разработанная СППВР состоит из трех модулей: модуля определения мутационного профиля опухоли, аналитического модуля для подбора эффективного персонализированного лекарственного лечения и формирования прогноза течения заболевания (Рисунок 41), а также модуля формирования и визуализации отчета [2].

```

{
  "id": 18,
  "name": "EID18",
  "description": "Idarubicin increases the overall survival and disease free survival in de novo AML patients with DNMT3A R882 mutation compared to those who do not harbor this mutation.",
  "disease": "DOID:9119",
  "drugs": [
    "Idarubicin"
  ],
  "rating": 4,
  "evidence_level": "B",
  "evidence_type": "Predictive",
  "clinical_significance": "Sensitivity/Response",
  "evidence_direction": "Supports",
  "variant_origin": "Somatic Mutation",
  "status": "accepted",
  "type": "evidence",
  "variant_id": 32,
  "source_id": 53
},

```

Рисунок 41. Пример входных данных СППВР. Связка: “вариант - диагноз - лекарственный препарат”

Рисунок 41 иллюстрирует входные данные СППВР, демонстрируя связку «вариант–диагноз–препарат». Это является центральным элементом в механизме формирования персонализированных рекомендаций.

Разработанная СППВР получает информацию из озера данных, находящегося в составе фармакогеномной платформы, и анализирует следующие типы данных:

- Геномные данные - СППВР проводит анализ вариантов в геноме пациента. Эти данные могут включать в себя SNV, вставки/делеции, структурные варианты и другие изменения в ДНК. СППВР также анализирует показатели ТМВ.
- Клинические данные - СППВР учитывает клинические характеристики больного, в частности, такие как патоморфологический диагноз.
- Данные литературы – СППВР использует информацию из открытых мировых баз данных и научных публикаций, интегрированную в созданное озеро данных в составе фармакогеномной платформы, для установления связи выявленных у больного биомаркеров с различными противоопухолевыми препаратами, в отношении которых наблюдается чувствительность.

После анализа указанных данных разработанная СППВР может определить:

1. Какие биомаркеры, выявленные у больного, могут влиять на ответ на определенные противоопухолевые препараты.

2. Какие противоопухолевые препараты наиболее подходят для конкретного больного, исходя из его геномного профиля, к каким препаратам может наблюдаться резистентность при наличии определенных геномных вариантов.

3. Уровень доказательности рекомендаций того или иного противоопухолевого препарата,

4. Зарегистрирован/одобрен ли определенный противоопухолевый препарат на территории Российской Федерации, ЕС или США.

5. Прогноз течения заболевания на основе выявленных геномных вариантов.

Таким образом, после анализа выявленных геномных вариантов/ТМВ и информации из созданного озера данных, СППВР позволяет формировать персонализированный отчет, состоящий из следующих элементов:

1. На первой странице отчета представлена сводная информация по выполненному исследованию: технические характеристики анализа, персональные данные пациента, кратко представлена общая информация о найденных геномных вариантах, уровне ТМВ, наличии MSI и представлены основные мутационные профили. Эта информация позволяет врачу быстро оценить общую клиническую картину (Рисунок 42).

2. На следующей странице представлено оглавление отчета, позволяющее эффективно ориентироваться в предоставленных результатах и находить необходимые сведения для принятия решений. (Рисунок 43).

3. Далее представлен раздел “Введение” в котором приводится общее описание проводимого исследования.

4. В следующем разделе “Эффективность таргетных и иммунотерапевтических препаратов” представлены клинически релевантные геномные варианты, приведена их классификация. Также представлен список препаратов с указанием уровней доказательности: одобренных для он- и офф-лейбл терапии, а также препаратов, информация по которым была получена в ходе исследований (Рисунок 44). После формирования отчетов осуществляется их модерация для обеспечения включения необходимых рекомендаций. Также приводятся ссылки на исследования соответствующих лекарственных препаратов.

5. В разделе “Прогноз течения заболевания” приводятся данные о прогнозе течения заболевания (благоприятный исход/неблагоприятный исход) при наличии тех или иных геномных вариантов с указанием уровней доказательства, а также ссылками на соответствующие исследования.

6. В разделе “Мутационный профиль и мутационная нагрузка” приводятся основные типы молекулярных профилей, позволяющих определить тип и этиологию опухоли.

7. Раздел “Материалы и методы” содержит описание основных этапов исследования, в том числе выделения ДНК, секвенирования, создания FASTQ файлов и других.

8. Раздел “Список использованных при проведении анализа инструментов и источников” содержит список открытых баз данных и инструментов, которые были использованы для проведения исследования и создания отчета.

9. Раздел “Список выявленных геномных вариантов” содержит классификацию геномных вариантов (Классы I-IV), а также полный список найденных геномных вариантов в геноме опухоли, подвергнутой анализу.

ФИО пациента:
 Диагноз: Рак молочной железы
 Дата отчета: 2.12.2024

Фармакогеномный анализ опухолевого материала

ФИО пациента:	Код образца:	2748TQMX
	Диагноз:	Рак молочной железы
Дата рождения:	Дата отчета:	2.12.2024
	Тип образца:	FFPE

Показатели полученные от лаборатории

Показатель	Значение	Показатель	Значение
Метод исследования	WES	Секвенирующая платформа	MGI-G400
Средняя глубина секвенирования	223X	Доля таргета с покрытием >30X	96%
Размер отсеквенированного региона	35.8Mb	Проанализировано вариантов	6727
Доля опухолевых клеток	0.9	Пloidность опухоли	2

Геномный профиль и выявленные биомаркеры

	TMB (мутационная нагрузка опухоли) Значение TMB составляет 3.1006m/Mb
	MSS (Microsatellite stable) Доля InDels составляет 0.21 (MSS)
	Мутационный профиль не установлен
	Найдено 1 клинически значимых геномных варианта Варианты обнаружены в следующих генах: BRCA2

Рисунок 42. Отчет СППВР о подборе эффективного персонализированного лекарственного лечения онкологических заболеваний. На первой странице представлены технические характеристики анализа, данные о том, что у больной ТНРМЖ показатель TMB составил 3.1006 мут./Мб, а также о том, что был выявлен стабильный микросателлитный статус (MSS) и геномный вариант в гене *BRCA2*

[Пациент: _____ Диагноз: Рак молочной железы _____ Дата отчета: _____]

Оглавление

	Введение	стр.3
	Эффективность таргетных и иммунотерапевтических препаратов	стр.4
	Прогноз течения заболевания	стр.9
	Мутационный профиль и мутационная нагрузка	стр.10
	Материалы и методы	стр.11
	Список использованных при проведении анализа инструментов и источников	стр.14
	Список выявленных геномных вариантов.	стр.16

Предоставленная информация не имеет самостоятельного клинического (медицинского) значения.

Рисунок 43. Отчет СППВР о подборе эффективного персонализированного лекарственного лечения онкологических заболеваний. Оглавление отчета.

Пациент: Диагноз: Рак молочной железы Дата отчета: 2.12.2024



Эффективность таргетных и иммунотерапевтических препаратов

У пациента выявлены следующие герминальные геномные варианты, имеющие терапевтическое значение:

Ген	Вариант	Область	Частота в популяции	Эффект	Глубина прочтений
BRCA2	c.2835del(p.Asp946IlefsTer14)	chr13-32911321-TA-T	0	Патогенный (PM22,PVS11)	

Значимых соматических геномных вариантов не обнаружено

Для данного типа злокачественного образования

Препараты, зарегистрированные в Российской Федерации или одобренные FDA/EMA для лечения при данном типе злокачественного образования («он-лейбл»):

Ген/Вариант	Тип рака	Уровень доказательства	Клиническая значимость	Ссылки на исследование	Рейтинг
Olaparib					
BRCA2 p.Asp946IlefsTer14	Рак молочной железы	A: Утверждено	Чувствительность Ответ	Robson et al., 2017, N Engl J Med	●●●●
Talazoparib					
BRCA2 p.Asp946IlefsTer14	Рак молочной железы	A: Утверждено	Чувствительность/ Ответ	Litton et al., The New England Journal of Medicine, 2018	●●●●

Рисунок 44. Отчет СППВР о подборе эффективного персонализированного лекарственного лечения онкологических заболеваний. Фрагмент раздела “Эффективность таргетных и иммунотерапевтических препаратов”. Можно видеть, что для больной ТНРМЖ с герминальным геномным вариантом в гене *BRCA2* с уровнем доказательности А рекомендованы препараты Olaparib и Talazoparib

На фрагменте отчета приведены геномные варианты и соответствующие препараты, сгруппированные по уровню доказательности. Это позволяет врачу ориентироваться в вариантах терапии, основанных на индивидуальных молекулярных особенностях пациента.

Таким образом, разработанная СППВР позволяет формировать обоснованные рекомендации по подбору лекарственных препаратов для персонализированного лечения больных онкологическими заболеваниями с учетом выявленных геномных характеристик больных.

3.4. Морфологические и иммуногистохимические характеристики образцов опухолевой ткани

3.4.1. Морфологические и иммуногистохимические характеристики образцов опухолевой ткани больных раком поджелудочной железы

Для оценки функционирования разработанных фармакогеномной платформы и СППВР было необходимо осуществить их тестовое применение на образцах биологического материала больных РПЖ. Для подтверждения наличия опухолевого материала во фрагментах, отобранных для исследования (Рисунок 45), а также для верификации и определения гистологического типа опухоли было проведено морфологическое и ИГХ исследование. Результаты проведенных исследований продемонстрировали наличие аденокарциномы в всех образцах биологического материала (n=40) и подтвердили возможность их использования для тестирования фармакогеномной платформы и СППВР.

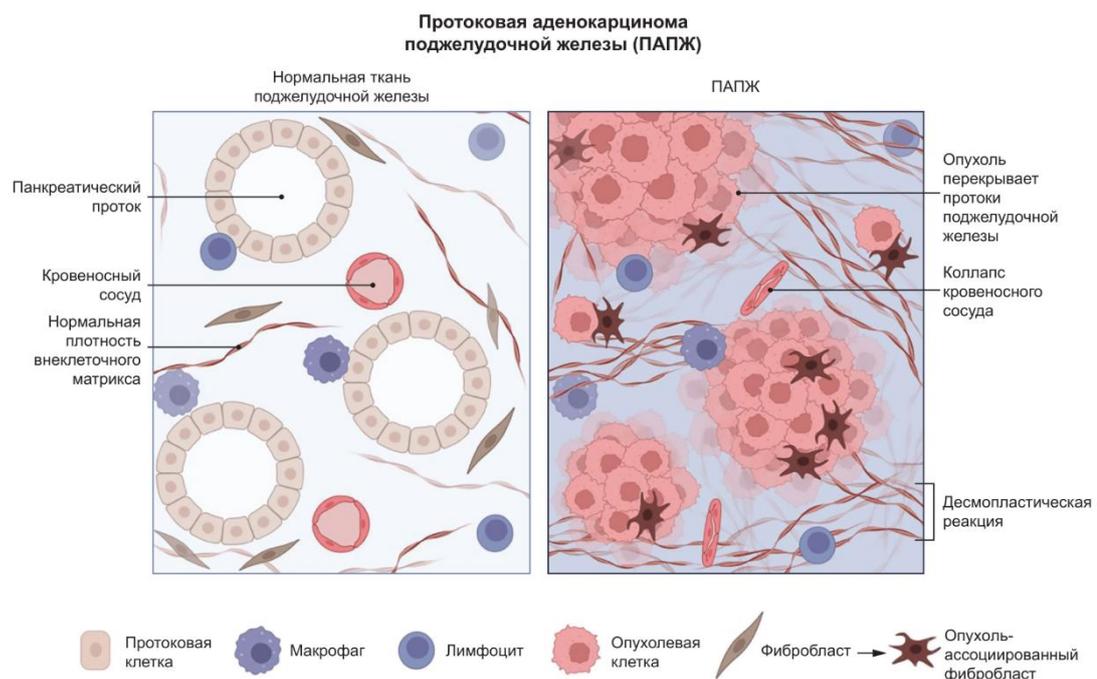


Рисунок 45. Основные морфологические различия нормальной ткани поджелудочной железы и ткани протоковой аденокарциномы поджелудочной железы (ПАПЖ). Создано с помощью Biorender

В когорте больных РПЖ (n = 40), операционный/биопсийный материал которых был использован в настоящем исследовании, первичную опухоль исследовали у 45% больных (n=18).

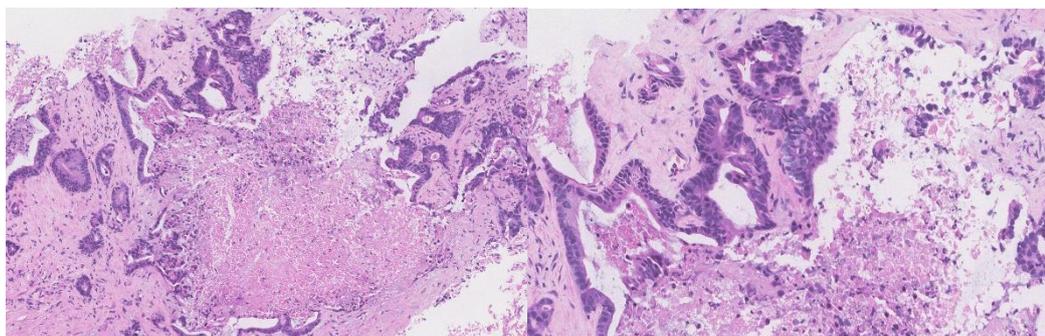
У остальных больных (55%) анализу был подвергнут метастаз РПЖ в печени (n=20; 50%), в лимфатическом узле (n=2; 5%).

По данным морфологического и иммуногистохимического исследования протоковая аденокарцинома диагностирована у 38 больных: у 9 (24%) высокодифференцированная, у 8 (21%) умеренно дифференцированная, еще у 6 (16%) низкодифференцированная; у 2 (5%) недифференцированная опухоль, а у 13 (34%) дифференцировку не удалось определить. У 1 больного установлена карцинома поджелудочной железы смешанного гистологического типа: ее большая часть была представлена низкодифференцированной ацинарной аденокарциномой (около 90%), меньшая часть была представлена умеренно дифференцированной протоковой аденокарциномой. В биоптате у 1 больного обнаружили муцинозную аденокарциному. Морфологические характеристики больных РПЖ представлены в Таблице 25.

Таблица 25. Морфологические характеристики опухолевого материала когорты больных РПЖ

Параметр	N = 40 больных
Гистологический тип	
Протоковая аденокарцинома	38 (95%)
Муцинозная аденокарцинома	1 (2,5%)
Смешанный (ацинарная аденокарцинома и протоковая аденокарцинома)	1 (2,5%)
Степень дифференцировки опухоли	
Не определено	14 (35%)
Недифференцированная	2 (5%)
Низко дифференцированная	6 (15%)
Умеренно дифференцированная	8 (20%)
Высоко дифференцированная	9 (22,5%)
Смешанный тип опухоли: низко + умеренно дифференцированная	1 (2,5%)

При гистологическом исследовании микропрепаратов поджелудочной железы и метастазов РПЖ участки неопластического роста имели вид гнездовидных структур с полигональными клетками с эозинофильной цитоплазмой и плеоморфными гиперхромными ядрами; некоторые кластеры неопластических клеток формировали решётчатую структуру за счёт наличия крупных вакуолей или напоминали перстневидные клетки. Местами высота атипичных glanduloцитов снижена, эпителий – истончён, просветы расширены и содержат муцин (альциан-позитивная реакция). Паратуморальная ткань с умеренной десмопластической реакцией и слабой воспалительной инфильтрацией (Рисунок 46)



А

Б

Рисунок 46. Опухолевая ткань больного Т., 57 лет. Протоковая карцинома поджелудочной железы. Окрашивание гематоксилином и эозином; А – увелич. $\times 200$; Б – увелич. $\times 400$. Плеоморфные атипичные клетки образуют тубулярные структуры. Большинство атипичных клеток полигональной формы, с эозинофильной цитоплазмой и плеоморфными гиперхромными ядрами. Стромальный компонент с умеренной десмопластической реакцией, клеточная воспалительная инфильтрация

При ИГХ исследовании образцов РПЖ для верификации и определения гистологического типа использовали маркеры принадлежности к опухолям эпителиального происхождения – СК7, ЕМА, СА125, СЕА. При исследовании метастатических узлов РПЖ в печени для дифференциальной диагностики использовали маркеры гепатоцитов Нер Par-1. Исследование экспрессии HER2/neu проводили с целью определения чувствительности опухоли к таргетной терапии анти-HER2/neu.

При ИГХ исследовании во всех исследуемых образцах аденокарциномы поджелудочной железы и ее метастазах, выявлены позитивные реакции с антителами к маркерам опухолей эпителиального происхождения СК7, ЕМА, СА125, СЕА (Рисунки 47 - 50), медиана доли окрашенных клеток составила 95%, 70%, 90% и 85% соответственно (Таблица 26). В то же время ИГХ реакции на Her2/neu были негативными, медианы доли окрашенных клеток не превышали 2,5% (Таблица 26, рисунок 51). ИГХ исследование с антителами к Нер-Par1 приведено на рисунке 52.

Таблица 26. Результаты оценки экспрессии иммуногистохимических маркеров в образцах аденокарциномы поджелудочной железы, %

Маркеры	СК7, %	СЕА, %	ЕМА, %	СА125, %	Her2/neu, %
Описательные статистики экспрессии					
Медиана	95	70	90	85	2,5
Квартиль 1	92,5	55	80	77,5	0
Квартиль 3	97,5	80	100	92,5	5

Примечание. Указано количество окрашенных раковых клеток, %

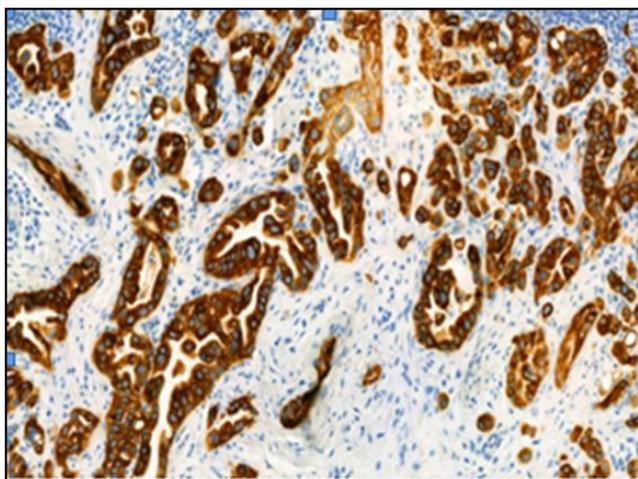


Рисунок 47. Опухолевая ткань больного К., 74 года. Метастаз протокового рака поджелудочной железы в лимфатический узел. Иммуногистохимическая реакция с антителами к СК7. Докрашивание – гематоксилином, увелич. $\times 100$. На фоне иммунонегативных клеток лимфоидной ткани отмечается мембранное окрашивание на СК7 всех атипичных клеток (эпителиальное происхождение опухоли)

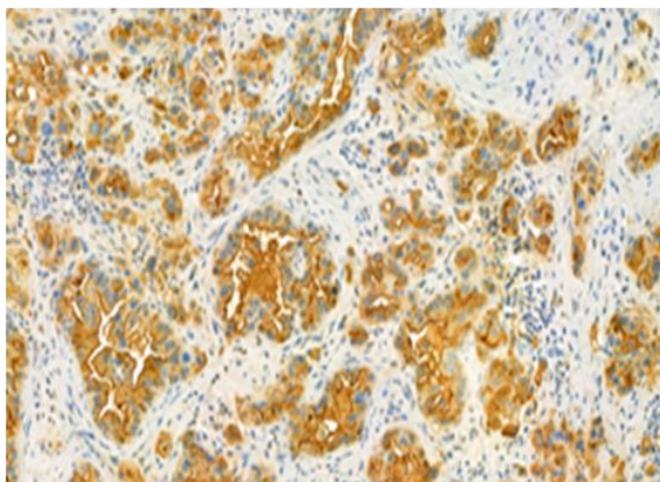


Рисунок 48. Опухолевая ткань больного А., 68 лет. Метастаз протокового рака поджелудочной железы в лимфатический узел. Иммуногистохимическая реакция с антителами к ЕМА. Докрашивание – гематоксилином, увелич. $\times 100$. На фоне иммунонегативных клеток лимфоидной ткани отмечается цитоплазматическое/ мембранное окрашивание на ЕМА всех атипичных клеток (эпителиальное происхождение опухоли)

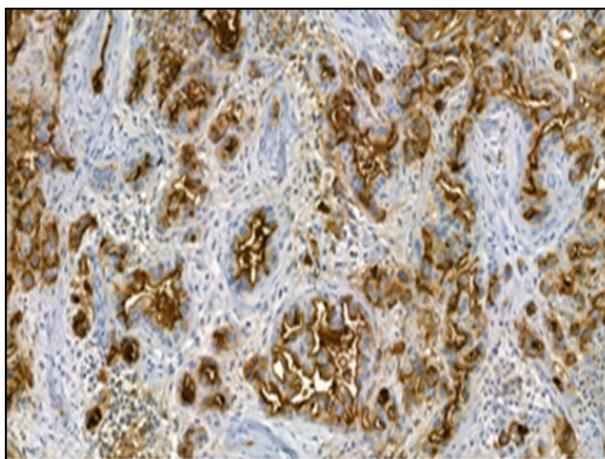


Рисунок 49. Опухолевая ткань больного К., 74 года. Метастаз протокового рака поджелудочной железы в лимфатический узел. Иммуногистохимическая реакция с антителами к СА-125. Докрашивание – гематоксилином, увелич. x100. На фоне иммунонегативных клеток лимфоидной ткани отмечается мембранное окрашивание на СА-125 всех атипичных клеток (эпителиальное происхождение опухоли)

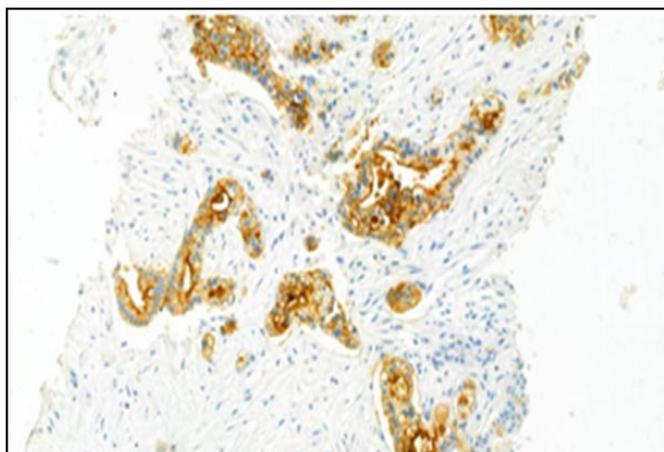


Рисунок 50. Опухолевая ткань больного К., 67 лет. Протоковый рак поджелудочной железы. Иммуногистохимическая реакция с антителами к СЕА. Докрашивание – гематоксилином, увелич. ×100. На фоне иммунонегативных клеток стромы отмечается мембранное окрашивание на СЕА большинства атипичных клеток

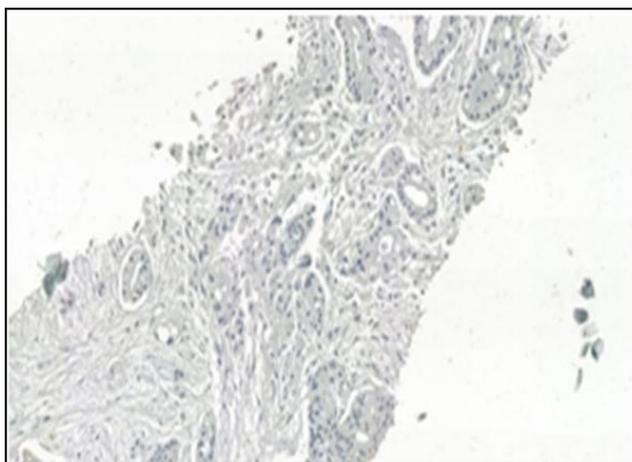


Рисунок 51. Опухолевая ткань больного А., 68 лет. Протоковая карцинома поджелудочной железы. Иммуногистохимическое исследование с антителами к Her2/neu. Докрашивание – гематоксилином, увелич. $\times 100$. Иммуномечение на онкопротеин *cerbB-2* (Her2/neu) не определяется (негативная реакции)

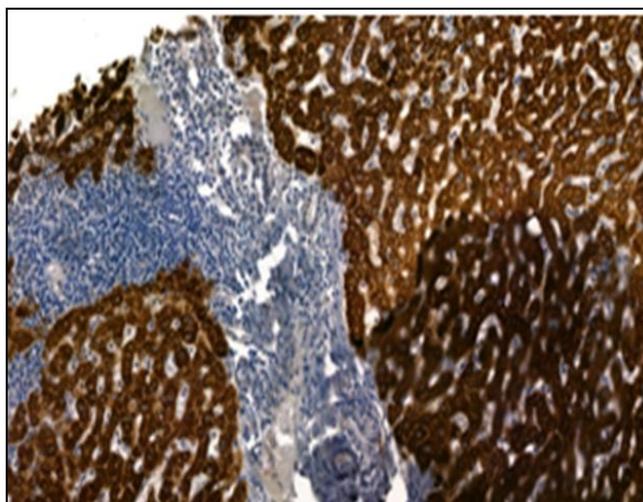


Рисунок 52. Опухолевая ткань больного Т., 63 года. Метастаз протокового рака поджелудочной железы в печень. Иммуногистохимическое исследование с антителами к Her-Par1. Докрашивание – гематоксилином, увелич. $\times 200$. Иммуномечение на маркер гепатоцитов/гепатоцеллюлярную карциному Her Par 1 не определяется (негативная реакции) на фоне выраженного иммуномечения цитоплазмы гепатоцитов (внутренний контроль), окружающих метастаз протокового рака поджелудочной железы

3.4.2. Морфологические и иммуногистохимические характеристики образцов опухолевой ткани больных трижды негативным раком молочной железы

Перед тестированием фармакогеномной платформы и СППВР на образцах биологического материала больных ТНРМЖ, было проведено морфологическое и ИГХ исследование, подтвердившее наличие опухолевого материала во всех фрагментах, отобранных для исследования, а также трижды негативный подтип РМЖ.

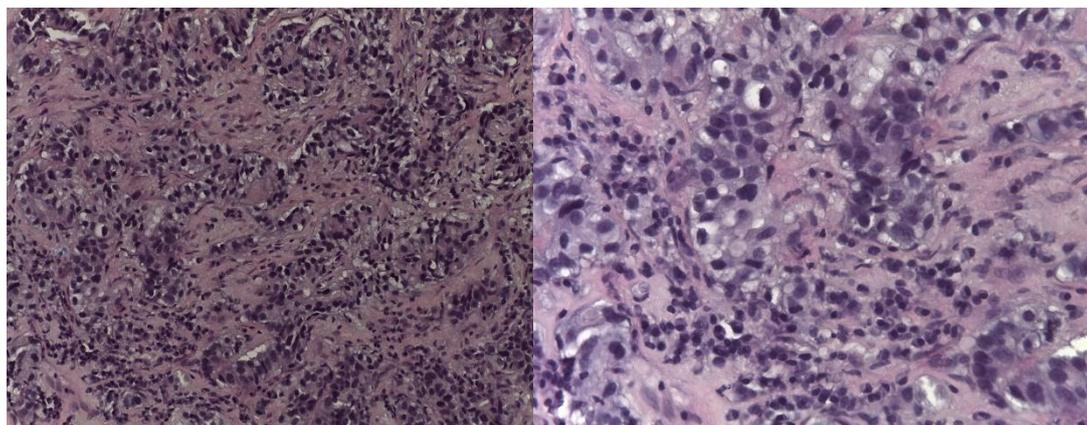
В когорте больных ТНРМЖ ($n=100$) большинство образцов (91%) было представлено инвазивным протоковым раком, 3% – инвазивным дольковым раком, 1% - протоково-дольковым раком, 5% составляли редкие типы ТНРМЖ. Морфологические характеристики опухолевого материала больных ТНРМЖ представлены в таблице 27.

Таблица 27. Морфологические характеристики опухолевого материала больных трижды негативным раком молочной железы

Параметр	N=100 больных
Гистологический тип Инвазивный протоковый рак Инвазивный дольковый рак Протоково-дольковый рак Редкие гистологические типы	91 (91%) 3 (3%) 1 (1%) 5 (5%)
Степень злокачественности G1 G2 G3	1 (1%) 33 (33%) 66 (66%)
KI-67 Среднее значение Медиана значения	63,22% ± 26,46% 70%

Во всех образцах молочной железы наблюдали тканевую и клеточную атипию. Плеоморфные атипичные клетки образовывали солидные и тубулярные структуры. Цитоплазма данных клеток в виде тонкого ободка; их ядра с признаками полиморфизма, расположены эксцентрично, они округлой формы, с «изъеденной» кариолеммой. Стромальный компонент выявлялся десмопластической реакцией различной степени выраженности. В пределах исследуемых микропрепаратов обнаружили множество патологических митозов (Рисунок 53).

Описанная морфологическая картина во всех образцах соответствовала РМЖ инвазивного типа, преимущественно 2 степени или 3 степени злокачественности по Elston и Ellis /G2 или G3 по Nottingham Grading system в модификации Bloom-Richardson.



А

Б

Рисунок 53. Опухолевая ткань больной К., 56 лет. Инвазивный протоковый рак молочной железы, G3. Окрашивание гематоксилином и эозином; А – увелич. $\times 200$; Б – увелич. $\times 400$. Плеоморфные атипичные клетки образуют солидную структуру. Цитоплазма данных клеток в виде тонкого ободка; ядра их с признаками полиморфизма, расположены эксцентрично, они округлой формы, с «изъеденной» кариолеммой. Стромальный компонент с умеренной десмопластической реакцией, клеточная воспалительная инфильтрация

Для верификации иммунофенотипического характера РМЖ было проведено ИГХ исследование с антителами к РЭ, РП, а также рецептору эпидермального фактора роста человека (Her2/neu). Во всех образцах РМЖ (n=100), ИГХ реакции на изучаемые маркеры были негативными (РЭ-, РП-, Her2/neu-), что подтверждает иммунный фенотип ТНРМЖ (Рисунок 54).

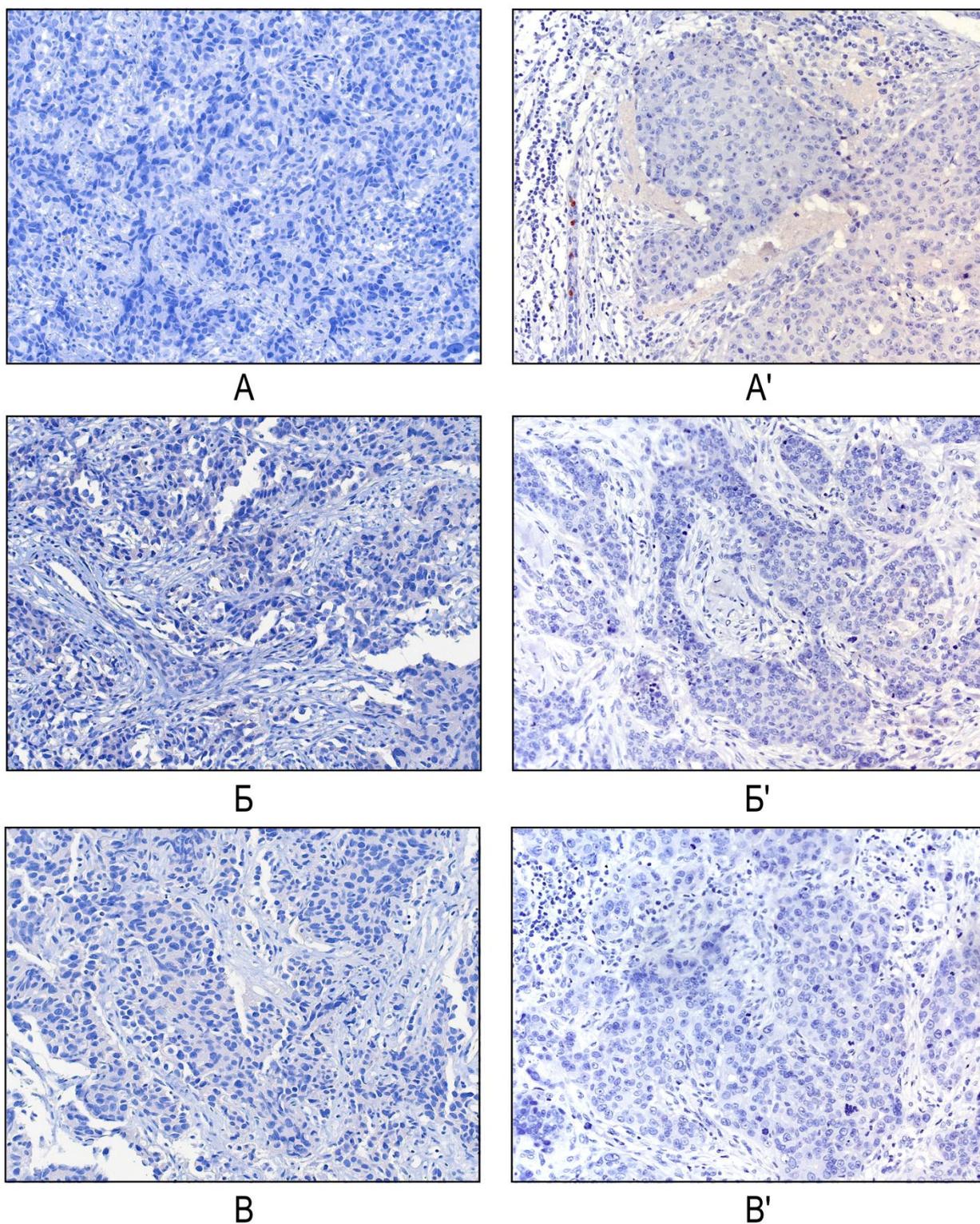


Рисунок 54. Опухолевая ткань, трижды негативный рак молочной железы. Иммуногистохимические реакции с антителами к РЭ (А - увелич. $\times 200$; А' - увелич. $\times 400$), РП (Б - увелич. $\times 200$; Б' - увелич. $\times 400$), Her2/neu (В - увелич. $\times 200$; В' - увелич. $\times 400$). Докрашивание – гематоксилином. При иммуногистохимическом исследовании образцов ТНРМЖ в ядрах атипичных клеток не обнаружено иммуномечения при окраске антителами к РЭ (0%) и РП (0%), а также на онкопротеин *serbB-2* (Her2/neu) (0%)

В то же время во всех образцах ТНРМЖ выявили позитивные ИГХ реакции с антителами к маркеру пролиферации – Ki-67 (рисунок 55, рисунок 56). В соответствии с результатами ИГХ исследования среднее значение уровня экспрессии маркера Ki-67 составило $63,22 \pm 26,46\%$, медиана - 70% (Q1 = 48, Q3 = 80).

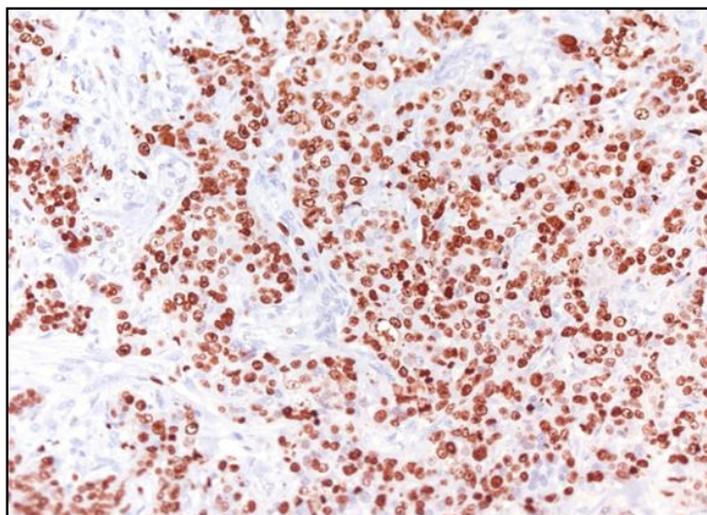


Рисунок 55. Опухолевая ткань больной О., 57 лет. Трижды негативный рак молочной железы. Иммуногистохимическая реакция с антителами к Ki-67. Докрашивание – гематоксилином, увелич. $\times 400$. Окрашивание на Ki-67 отмечается в большинстве ядер атипичных клеток. Средний уровень пролиферативной активности, оцененный по уровню экспрессии Ki-67 по шкале St. Gallen Consensus – 94,8%

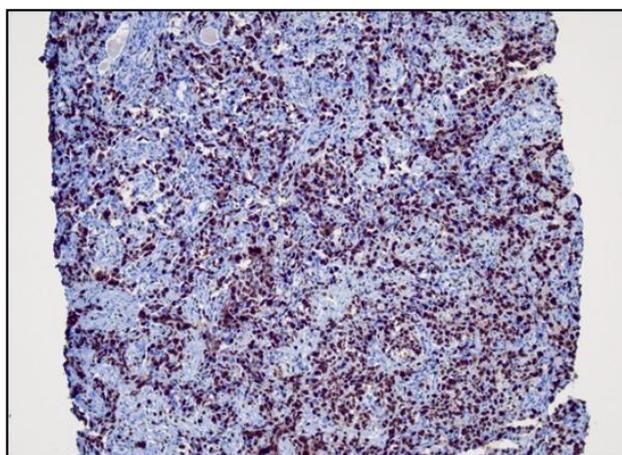


Рисунок 56. Опухолевая ткань больной С., 50 лет. Трижды негативный рак молочной железы. Иммуногистохимическая реакция с антителами к Ki-67. Докрашивание – гематоксилином, увелич. $\times 200$. Окрашивание на Ki-67 отмечается в большинстве ядер атипичных клеток. Средний уровень пролиферативной активности, оцененный по уровню экспрессии Ki-67 по шкале St. Gallen Consensus – 68,7%

3.5. Эффективность выявления геномных вариантов созданной фармакогеномной платформой

Для оценки эффективности созданной фармакогеномной платформы [15], обнаруженные с её помощью геномные варианты при ТНРМЖ были сравнены с геномными вариантами, выявленными с использованием референсных наборов и программ для биоинформатической обработки AVENIO (Roche, США). Для сравнительного анализа из 100 образцов ТНРМЖ, проанализированных фармакогеномной платформой, были отобраны 30 (28 образцов с найденными геномными вариантами и 2 образца, в которых не было выявлено геномных вариантов, в качестве отрицательного контроля). Показатель Q30 (доля таргета с покрытием >30X) для всех образцов в обеих группах составил более 95%.

Все значимые при ТНРМЖ патогенные/вероятно патогенные герминальные геномные варианты и все соматические геномные варианты 1-2 классов патогенности, выявленные с использованием наборов и программ AVENIO (Roche, США), были выявлены и разработанной фармакогеномной платформой (Таблица 28). Кроме того, при применении фармакогеномной платформы было выявлено 15 дополнительных геномных вариантов.

Таблица 28. Результаты выявления патогенных/вероятно патогенных герминальных геномных вариантов, а также соматических геномных вариантов 1-2 классов патогенности созданной фармакогеномной платформой, и наборами и программами AVENIO (Roche, США)

№ Пациента	Тип выявленного геномного варианта	Геномные варианты, выявленные с использованием разработанной фармакогеномной платформы	Геномные варианты, выявленные с использованием наборов прободготовки и программ для биоинформатической обработки AVENIO (Roche, США)
1	Соматический и герминальный	TP53 17:g.7578406C>T p.Arg175His BRCA1 17:g.41209079T>TG p.Gln1777ProfsTer74 rs80357906	TP53 17:g.7578406C>T p.Arg175His BRCA1 17:g.41209079T>TG p.Gln1777ProfsTer74 rs80357906
2	Соматический и герминальный	TP53 17:g.7578534C>G p.Lys132Asn PMS2 7:g.6035165C>T p.Lys301= rs267608153	TP53 17:g.7578534C>G p.Lys132Asn PMS2 7:g.6035165C>T p.Lys301= rs267608153
3	Соматический и герминальный	KMT2C 7:g.151836873GC>G p.Leu4783TrpfsTer27 BRCA2 13:g.32936829A>G p.Arg2659Gly rs80359026	Исследование гена KMT2C не входит в панель AVENIO BRCA2 13:g.32936829A>G p.Arg2659Gly rs80359026

№ Пациента	Тип выявленного геномного варианта	Геномные варианты, выявленные с использованием разработанной фармакогеномной платформы	Геномные варианты, выявленные с использованием наборов пробоподготовки и программ для биоинформатической обработки AVENIO (Roche, США)
4	Соматический и герминальный	<p>KMT2C 7:g.151921114A>T p.Cys1103Ter</p> <p>CDKN2A 9:g.21971053GC>G p.Ala102ArgfsTer44</p> <p>TP53 17:g.7578413C>T p.Val173Met</p> <p>BRCA2 13:g.32915082CTG>C p.Glu2198AsnfsTer4 rs80359605</p>	<p>Исследование гена KMT2C не входит в панель AVENIO</p> <p>Исследование гена CDKN2A не входит в панель AVENIO</p> <p>TP53 17:g.7578413C>T p.Val173Met</p> <p>BRCA2 13:g.32915082CTG>C p.Glu2198AsnfsTer4 rs80359605</p>
5	Соматический и герминальный	<p>PIK3CA 3:g.178951973A>G p.Met1010Val</p> <p>BRCA1 17:g.41244254AAG>A p.Leu1098SerfsTer4 rs80357992</p> <p>FANCI 15:g.89315318C>T p.Arg1285X rs121918164</p>	<p>PIK3CA 3:g.178951973A>G p.Met1010Val</p> <p>BRCA1 17:g.41244254AAG>A p.Leu1098SerfsTer4 rs80357992</p> <p>Исследование гена FANCI не входит в панель AVENIO</p>
6	Соматический и герминальный	<p>TP53 17:g.7577539G>A p.Arg248Trp</p> <p>BRCA1 17:g.41209079T>TG p.Gln1777ProfsTer74 rs80357906</p>	<p>TP53 17:g.7577539G>A p.Arg248Trp</p> <p>BRCA1 17:g.41209079T>TG p.Gln1777ProfsTer74 rs80357906</p>
7	Соматический и герминальный	<p>TP53 17:g.7578406CG>C p.Arg175AlafsTer72</p> <p>BRCA1 17:g.41234451G>A p.Arg1443Ter rs41293455</p>	<p>TP53 17:g.7578406CG>C p.Arg175AlafsTer72</p> <p>BRCA1 17:g.41234451G>A p.Arg1443Ter rs41293455</p>

№ Пациента	Тип выявленного геномного варианта	Геномные варианты, выявленные с использованием разработанной фармакогеномной платформы	Геномные варианты, выявленные с использованием наборов пробоподготовки и программ для биоинформатической обработки AVENIO (Roche, США)
8	Соматический и герминальный	<p>NOTCH1 9:g.139413900C>T p.Trp287Ter</p> <p>TP53 17:g.7578271T>C p.His193Arg</p> <p>STK11 19:g.1207078G>A p.Gly56Arg</p> <p>GNAI1 19:g.3115013G>A p.Arg183His</p> <p>BRCA1 17:g.41243776CCT>C p.Glu1257GlyfsTer9 rs80357579</p>	<p>Исследование гена NOTCH1 не входит в панель AVENIO</p> <p>TP53 17:g.7578271T>C p.His193Arg</p> <p>Исследование гена STK11 не входит в панель AVENIO</p> <p>Исследование гена GNAI1 не входит в панель AVENIO</p> <p>BRCA1 17:g.41243776CCT>C p.Glu1257GlyfsTer9 rs80357579</p>
9	Соматический и герминальный	<p>TP53 17:g.7579389GGGAA>G p.Ser99ArgfsTer23</p> <p>BRCA1 17:g.41209079T>TG p.Gln1777ProfsTer74 rs80357906</p> <p>SLX4 16:g.3658472T>TG p.Gln165ProfsTer121</p>	<p>TP53 17:g.7579389GGGAA>G p.Ser99ArgfsTer23</p> <p>BRCA1 17:g.41209079T>TG p.Gln1777ProfsTer74 rs80357906</p> <p>Исследование гена SLX4 не входит в панель AVENIO</p>
10	Соматический и герминальный	<p>EZH2 7:g.148516705G>A p.Gln328Ter</p> <p>TP53 17:g.7578382G>C p.Ser183Ter</p> <p>BRCA1 17:g.41209079T>TG p.Gln1777ProfsTer74 rs80357906</p>	<p>Исследование гена EZH2 не входит в панель AVENIO</p> <p>TP53 17:g.7578382G>C p.Ser183Ter</p> <p>BRCA1 17:g.41209079T>TG p.Gln1777ProfsTer74 rs80357906</p>
11	Соматический и герминальный	<p>PTEN 10:g.89717609G>C</p> <p>BRCA1 17:g.41209079T>TG p.Gln1777ProfsTer74 rs80357906</p>	<p>PTEN 10:g.89717609G>C</p> <p>BRCA1 17:g.41209079T>TG p.Gln1777ProfsTer74 rs80357906</p>

№ Пациента	Тип выявленного геномного варианта	Геномные варианты, выявленные с использованием разработанной фармакогеномной платформы	Геномные варианты, выявленные с использованием наборов пробоподготовки и программ для биоинформатической обработки AVENIO (Roche, США)
12	Соматический и герминальный	<p>PIK3CA 3:g.178952085A>G p.His1047Arg</p> <p>PTEN 10:g.89685269G>A</p> <p>TP53 17:g.7577536T>C p.Arg249Gly</p> <p>RAD52 12:g.925505C>G p.Gly163Ala</p>	<p>PIK3CA 3:g.178952085A>G p.His1047Arg</p> <p>PTEN 10:g.89685269G>A</p> <p>TP53 17:g.7577536T>C p.Arg249Gly</p> <p>Исследование гена RAD52 не входит в панель AVENIO</p>
13	Соматический	<p>TP53 17:g.7577539G>A p.Arg248Trp</p> <p>TSC1 9:g.135796805G>A p.Arg228Ter</p>	<p>TP53 17:g.7577539G>A p.Arg248Trp</p> <p>Исследование данного региона гена TSC1 не входит в панель AVENIO</p>
14	Соматический	<p>TP53 17:g.7577121G>A p.Arg273Cys</p> <p>NOTCH2 1:g.120458324G>A p.Gln2341Ter</p>	<p>TP53 17:g.7577121G>A p.Arg273Cys</p> <p>Исследование гена NOTCH2 не входит в панель AVENIO</p>
15	Соматический	<p>TP53 17:g.7577121G>A p.Arg273Cys</p>	<p>TP53 17:g.7577121G>A p.Arg273Cys</p>
16	Соматический	<p>TP53 17:g.7577120C>T p.Arg273His</p>	<p>TP53 17:g.7577120C>T p.Arg273His</p>
17	Соматический	<p>TP53 17:g.7577539G>A p.Arg248Trp</p>	<p>TP53 17:g.7577539G>A p.Arg248Trp</p>
18	Соматический	<p>TP53 17:g.7578406C>T p.Arg175His</p>	<p>TP53 17:g.7578406C>T p.Arg175His</p>
19	Соматический	<p>PIK3CA 3:g.178952085A>G p.His1047Arg</p> <p>TP53 17:g.7578406C>T p.Arg175His</p>	<p>PIK3CA 3:g.178952085A>G p.His1047Arg</p> <p>TP53 17:g.7578406C>T p.Arg175His</p>

№ Пациента	Тип выявленного геномного варианта	Геномные варианты, выявленные с использованием разработанной фармакогеномной платформы	Геномные варианты, выявленные с использованием наборов пробоподготовки и программ для биоинформатической обработки AVENIO (Roche, США)
20	Соматический	<p>PIK3CA 3:g.178952085A>G p.His1047Arg</p> <p>PIK3CA 3:g.178936091G>A p.Glu545Lys</p> <p>TP53 17:g.7578212G>A p.Arg213Ter</p>	<p>PIK3CA 3:g.178952085A>G p.His1047Arg</p> <p>PIK3CA 3:g.178936091G>A p.Glu545Lys</p> <p>TP53 17:g.7578212G>A p.Arg213Ter</p>
21	Соматический	<p>ESR1 6:g.152419926A>G p.Asp538Gly</p> <p>PIK3CA 3:g.178936082G>A p.Glu542Lys</p>	<p>ESR1 6:g.152419926A>G p.Asp538Gly</p> <p>PIK3CA 3:g.178936082G>A p.Glu542Lys</p>
22	Соматический	<p>AKT1 14:g.105246551C>T p.Glu17Lys</p>	<p>AKT1 14:g.105246551C>T p.Glu17Lys</p>
23	Соматический	<p>PIK3CA 3:g.178952085A>G p.His1047Arg</p> <p>TP53 17:g.7574013G>GAAGC p.Phe338LeufsTer10</p>	<p>PIK3CA 3:g.178952085A>G p.His1047Arg</p> <p>TP53 17:g.7574013G>GAAGC p.Phe338LeufsTer10</p>
24	Соматический	<p>PIK3CA 3:g.178936094C>A p.Gln546Lys</p>	<p>PIK3CA 3:g.178936094C>A p.Gln546Lys</p>
25	Соматический	<p>PIK3CA 3:g.178936091G>A p.Glu545Lys</p> <p>PIK3CB 3:g.138374245C>A p.Asp1067Tyr</p> <p>TP53 17:g.7579360GAAACC>G p.Gly108ProfsTer39</p>	<p>PIK3CA 3:g.178936091G>A p.Glu545Lys</p> <p>Исследование гена PIK3CB не входит в панель AVENIO</p> <p>TP53 17:g.7579360GAAACC>G p.Gly108ProfsTer39</p>
26	Соматический	<p>PIK3CA 3:g.178936091G>A p.Glu545Lys</p>	<p>PIK3CA 3:g.178936091G>A p.Glu545Lys</p>
27	Соматический	<p>TP53 17:g.7579329T>C p.Lys120Glu</p>	<p>TP53 17:g.7579329T>C p.Lys120Glu</p>

№ Пациента	Тип выявленного геномного варианта	Геномные варианты, выявленные с использованием разработанной фармакогеномной платформы	Геномные варианты, выявленные с использованием наборов пробоподготовки и программ для биоинформатической обработки AVENIO (Roche, США)
28	Соматический	<p>TP53 17:g.7577595C>CG p.Cys229SerfsTer11</p> <p>WT1 11:g.32417910G>A p.Ser381Leu</p> <p>SETBP1 18:g.42532907G>A p.Ser1201Asn</p>	<p>TP53 17:g.7577595C>CG p.Cys229SerfsTer11</p> <p>Исследование гена WT1 не входит в панель AVENIO</p> <p>Исследование гена SETBP1 не входит в панель AVENIO</p>
29	-	Геномные варианты не выявлены	Геномные варианты не выявлены
30	-	Геномные варианты не выявлены	Геномные варианты не выявлены

При использовании созданной фармакогеномной платформы было обнаружено 47 соматических вариантов 1 и 2 классов патогенности в генах *TP53*, *KMT2C*, *CDKN2A*, *PIK3CA*, *NOTCH1*, *STK11*, *GNA11*, *EZH2*, *PTEN*, *TSC1*, *NOTCH2*, *ESR1*, *AKT1*, *PIK3CB*, *WT1*, *SETBP1* и 14 патогенных/вероятно патогенных герминальных вариантов в генах *BRCA1*, *BRCA2*, *PMS2*, *FANCI*, *SLX4* и *RAD52*. У 2 больных (отрицательный контроль) не было выявлено геномных вариантов.

При применении системы AVENIO (Roche, США) было выявлено 35 соматических геномных вариантов 1 и 2 классов патогенности в генах *TP53*, *PIK3CA*, *PTEN*, *ESR1* и *AKT1*, число выявленных патогенных/вероятно патогенных герминальных вариантов в генах *BRCA1*, *BRCA2*, *PMS2* составило 11. У 2 больных, представлявших отрицательный контроль, также не было выявлено геномных вариантов.

Отличия в количестве выявленных геномных вариантов связано с тем, что в панель AVENIO (Roche, США) не входит исследование генов *KMT2C*, *CDKN2A*, *FANCI*, *NOTCH1*, *STK11*, *GNA11*, *SLX4*, *EZH2*, *RAD52*, *NOTCH2*, *PIK3CB*, *WT1*, *SETBP1*, а в гене *TSC1* исследуются не все регионы. При этом в данных генах и были выявлены геномные варианты ($n=15$) с использованием нашей фармакогеномной платформы, в рамках которой осуществлялось полноэкзомное секвенирование.

Применение фармакогеномной платформы показало соответствие референсному протоколу. Кроме того, использование фармакогеномной платформы позволило выявить патогенные и вероятно патогенные герминальные варианты, а также соматические варианты 1 и 2 классов патогенности в большем числе генов, тем самым расширяя возможности для подбора персонализированного лекарственного лечения ЗНО.

3.6. Геномный профиль при раке поджелудочной железы и трижды негативном раке молочной железы

3.6.1. Геномный профиль при раке поджелудочной железы

3.6.1.1. Герминальный геномный профиль

Разработанный адаптированный протокол пробоподготовки и секвенирования, а также биоинформатические конвейеры были применены для определения герминального геномного профиля при исследовании биологического материала больных РПЖ из российской популяции [314]. Исследование герминального геномного профиля РПЖ выявило 16240 генов, содержащих геномные варианты. При этом, из них 8 генов (*ATM*, *BRCA2*, *COG4*, *DNMT1*, *PCNT*, *SLC12A6*, *SYNE2*, *WNK1*) содержали геномные варианты у всех пациентов исследуемой группы. Наибольшее число геномных вариантов (в порядке убывания) наблюдалось в генах: *APC*, *MUC19*, *MUC16*, *BRCA2*, *BRCA1*, *MUC4*, *OBSCN*, *PKD1L2*, *TTN*, *AHNAK2*. С целью обнаружения “выбросов” (экстремально высоких и низких значений) герминального геномного профиля была построена тепловая карта (хитмэп) (Рисунок 57).

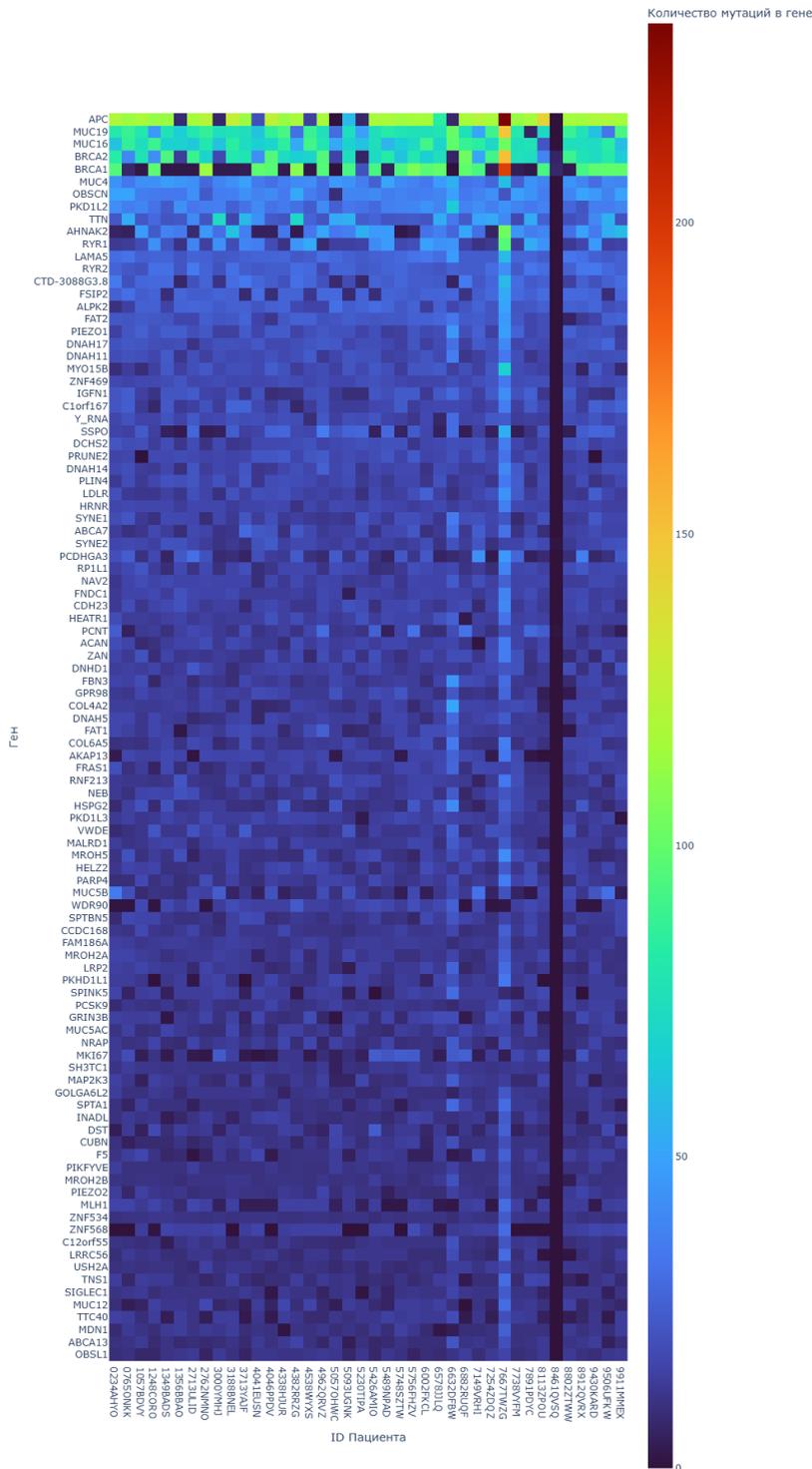


Рисунок 57. Фрагмент тепловой карты типа “Хитмэп” (head) распределения герминальных геномных вариантов для 100 генов (топ 100) с наиболее высоким числом геномных вариантов в исследуемой группе пациентов по данным полноэкзомного NGS в группе больных РПЖ. Гены (по вертикали) отсортированы от большего числа геномных вариантов к меньшему. Набор цветов "торнадо": темные цвета иллюстрируют меньшее число геномных вариантов, светлые и яркие цвета - большее число. Визуализируются экстремально высокие значения: у пациента с ID 7667TWZG отмечается повышенное количество геномных вариантов (преимущественно светлые цвета ячеек), и более 230 геномных вариантов в гене *APC*. У пациента с ID 8461QV9Q напротив отмечается чрезвычайно низкое число геномных вариантов (темный цвет на тепловой карте)

Для лучшей визуализации ландшафта геномных вариантов два пациента с экстремально высоким и низким количеством геномных вариантов (пациент с ID 7667TWZG и пациент с ID 8461QVSQ) были удалены из тепловой карты (хитмэп) (Рисунок 58).

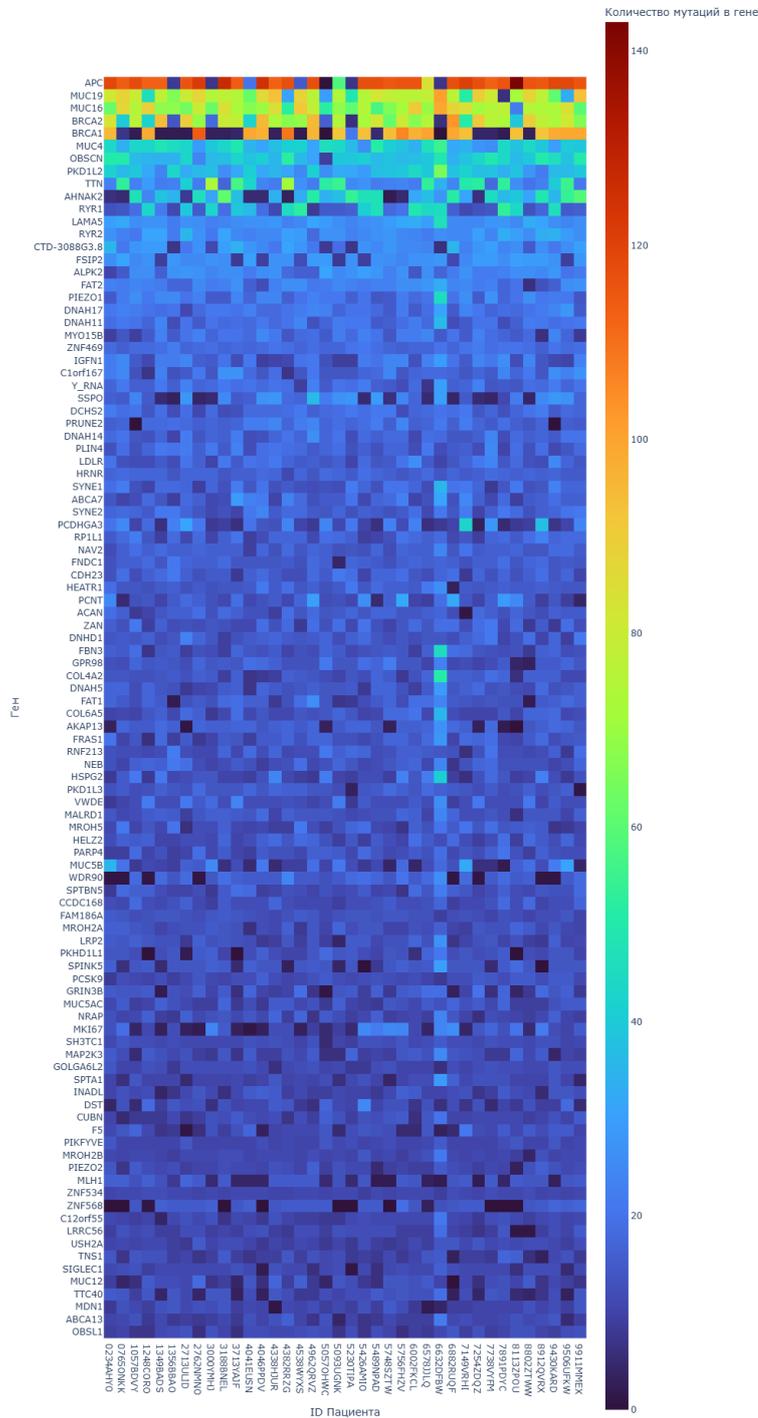


Рисунок 58. Фрагмент ландшафта геномных вариантов - тепловой карты типа “Хитмэп” (head) распределения герминальных геномных вариантов для 100 генов (топ 100) с наиболее высоким числом геномных вариантов в исследуемой группе пациентов по данным полноэкзомного NGS в группе больных РПЖ, после исключения т.н. “выбросов” - двух пациентов с экстремально высоким и низким количеством геномных вариантов. Гены (по вертикали) отсортированы от

большого числа геномных вариантов к меньшему. Набор цветов "торнадо": темные цвета иллюстрируют меньшее число геномных вариантов, светлые и яркие цвета - большее число

Для каждой пары генов была построена корреляционная матрица с применением критерия Пирсона, обнаружив множество значимых положительных корреляций. Так, например, количество геномных вариантов в гене *ZNF534* обнаруживает корреляцию с количеством геномных вариантов в гене *HRNR* с коэффициентом корреляции по Пирсону более 0,92. При этом наличие геномных вариантов в генах *MUC5B*, *WDR90*, *SPINK5* и *MKI67* вообще не обнаруживает положительных корреляций с какими-либо другими генами (Рисунок 59).

SYMBOL	APC	MUC19	MUC16	BRCA2	BRCA1	MUC4	OBSCN	PKD1L2	TTN	AHNAK2	...	C12orf55	LRRCS6	USH2A	TNS1	SIGLEC1	MUC12	TTC40	MDN1	ABCA13	OBSL1
APC	1.000000	0.406557	0.309496	0.806439	0.487023	0.168171	0.251304	0.144690	-0.064192	0.314854	...	0.246966	0.254849	0.496811	0.427760	0.335109	0.284974	0.007090	0.456843	0.198452	0.152874
MUC19	0.406557	1.000000	0.503454	0.495259	0.190839	0.402468	0.247443	0.474525	-0.155940	0.342408	...	0.525137	0.491560	0.629836	0.290047	0.583867	0.412285	0.275159	0.447689	0.409750	0.406194
MUC16	0.309496	0.503454	1.000000	0.406386	0.149118	0.439248	0.350922	0.531325	0.130941	0.357185	...	0.542355	0.611195	0.703428	0.309476	0.592556	0.429803	0.349513	0.308626	0.396747	0.287859
BRCA2	0.806439	0.495259	0.406386	1.000000	0.391412	0.234262	0.175882	0.052146	-0.046741	0.342322	...	0.158721	0.323801	0.476564	0.355861	0.386177	0.315296	-0.112383	0.440038	0.101206	0.259408
BRCA1	0.487023	0.190839	0.149118	0.391412	1.000000	0.003249	0.025644	0.224979	0.039324	0.202874	...	0.261255	0.346981	0.267424	0.307143	0.168370	0.377085	0.088401	0.348861	0.103778	0.044438

Рисунок 59. Часть корреляционной матрицы Пирсона для генов с герминальными геномными вариантами по данным полноэкзомного NGS в группе больных РПЖ.

Для лучшей визуализации положительных корреляций в большом объеме данных была создана графическая иллюстрация в форме тепловой карты или хитмэпа для 100 генов с наибольшим числом геномных вариантов (Рисунок 60).

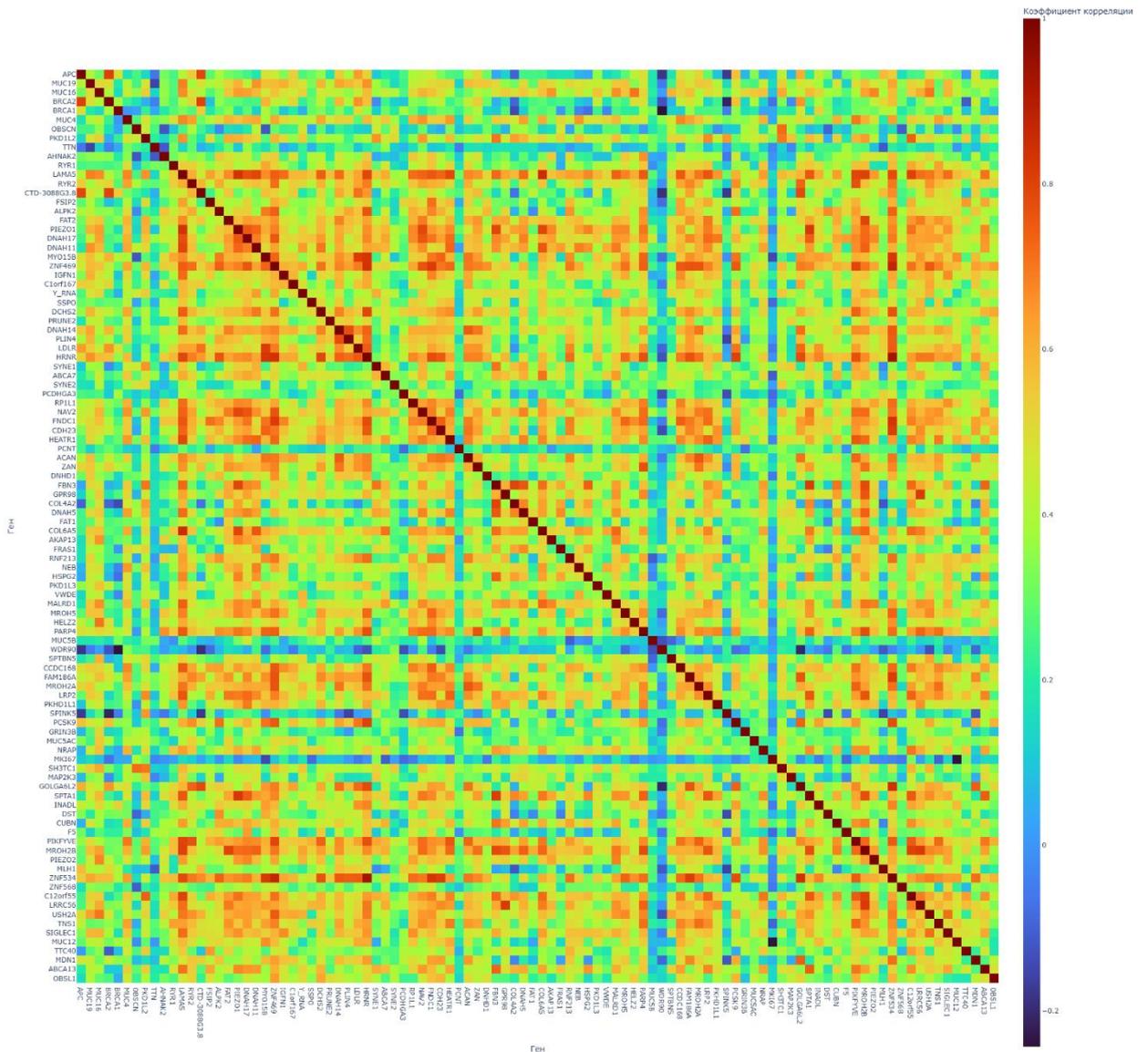


Рисунок 60. Графическая иллюстрация (тепловая карта) корреляционной матрицы Пирсона для 100 генов с наибольшим числом герминальных геномных вариантов по данным полноэкзомного NGS в группе больных РПЖ. Набор цветов "торнадо": темные цвета иллюстрируют отрицательную корреляцию, светлые и яркие цвета - положительные коэффициенты корреляции

При подборе персонализированного лекарственного лечения онкологических заболеваний учитываются патогенные и вероятно патогенные герминальные геномные варианты. Поэтому далее приведено детализированное описание ландшафта патогенных и вероятно патогенных герминальных геномных вариантов, значимых при РПЖ, в когорте 40 больных РПЖ из исследуемой когорты.

Результаты проведенного полноэкзомного NGS показали наличие патогенных и вероятно-патогенных герминальных геномных вариантов у 4 (10%) больных из когорты РПЖ в 3 генах. У 2 больных были выявлены герминальные геномные варианты в гене *ATM*, у 1 больного в гене *ERCC3*, в *FANCM* - также у 1 больного (Таблица 29).

Таблица 29. Патогенные и вероятно патогенные герминальные геномные варианты, выявленные в когорте 40 больных РПЖ

Больные РПЖ с вариантами в определённом гене (n=)	Ген	Геномный вариант	Аминокислотная замена	rs	Патогенность	Количество больных РПЖ с геномным вариантом (n=)
2	<i>ATM</i>	11:g.108175540C>T	p.Gln1879Ter	rs756109905	Патогенный вариант	1
	<i>ATM</i>	11:g.108205832T>C	p.Val2716Ala	rs587782652	Патогенный вариант	1
1	<i>ERCC3</i>	2:g.128050332G>A	p.Arg109Ter	rs34295337	Патогенный вариант	1
1	<i>FANCM</i>	14:g.45636336C>T	p.Arg658Ter	rs368728266	Патогенный вариант	1
Суммарное количество больных РПЖ с патогенными и вероятно патогенными герминальными геномными вариантами.						4 (10%)

3.6.1.2. Соматический геномный профиль

Для определения соматического геномного профиля при РПЖ были применены разработанные адаптированный протокол пробоподготовки и секвенирования и биоинформатический конвейер. Всего выявлено 3223 гена, содержащих геномные варианты во всей когорте из 40 больных.

Для визуализации соматического геномного профиля была построена тепловая карта (хитмэп) (Рисунок 61).

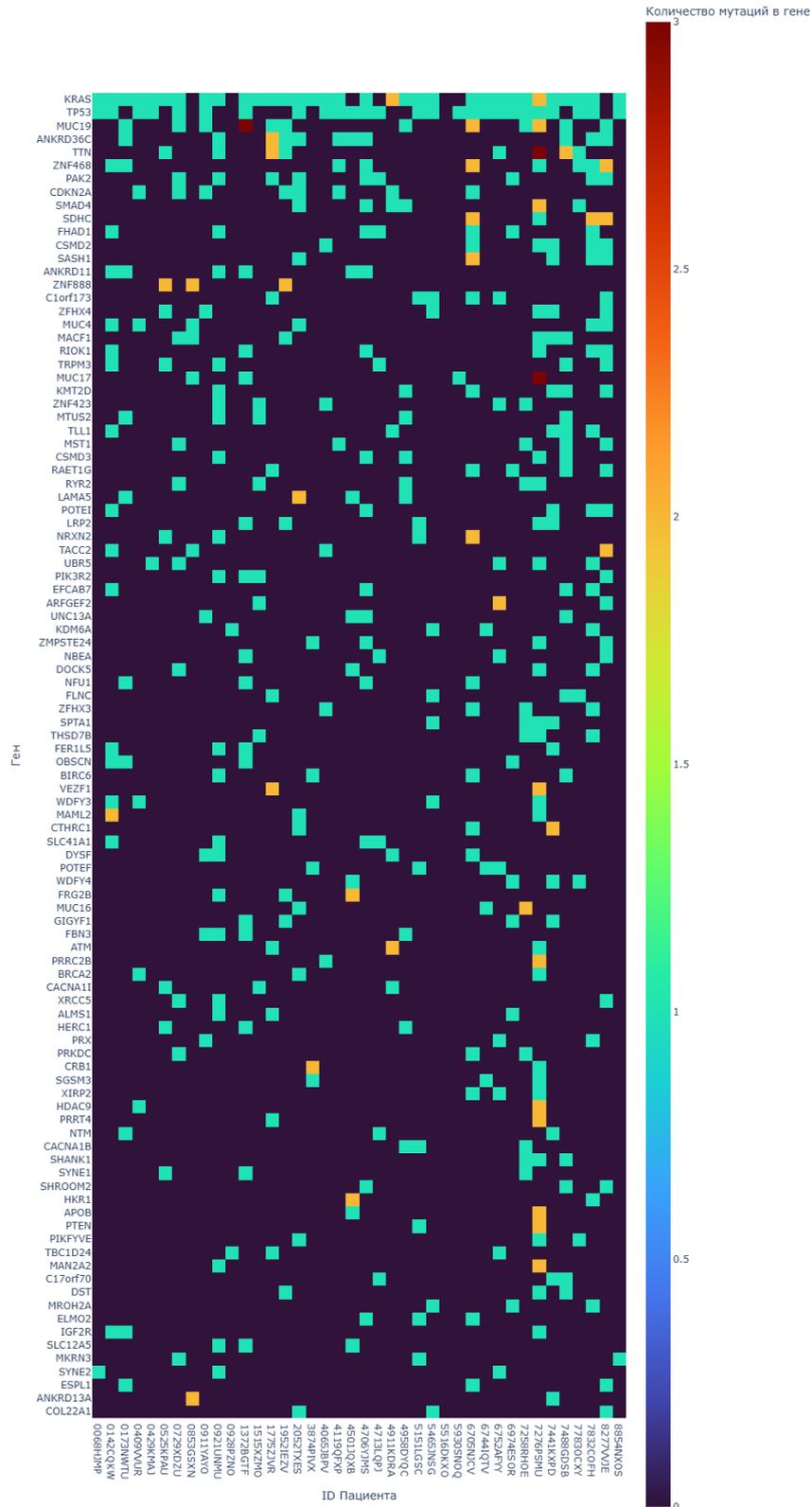


Рисунок 61. Фрагмент тепловой карты (Хитмэп) распределения соматических геномных вариантов по данным полноэкзомного NGS в группе больных РПЖ, показаны 100 генов с наибольшим числом геномных вариантов. Набор цветов "торнадо": темные цвета иллюстрируют меньшее число геномных вариантов, светлые и яркие цвета - большее число; Разброс редких ярко окрашенных точек на графике отражает равномерное распределение единичных геномных вариантов в популяции

Значимыми для подбора персонализированного лекарственного лечения онкологических заболеваний являются соматические геномные варианты 1-го и 2-го классов патогенности. Поэтому далее приведено детализированное описание ландшафта таких соматических геномных вариантов в когорте 40 больных РПЖ (Таблица 30). При этом важно отметить, что у 25 больных наблюдалось наличие одновременно нескольких геномных вариантов (от 2 до 5).

Таблица 30. Соматические геномные варианты 1-го и 2-го класса патогенности, выявленные в когорте 40 больных РПЖ

Больные РПЖ с вариантом в определенном гене (n=)	Ген	Геномный вариант	Аминокислотная замена	Класс патогенности (Tier)	Количество больных РПЖ с геномным вариантом (n=)
33	<i>KRAS</i>	12:g.25398284C>T	p.Gly12Asp	Первый	18
		12:g.25398284C>A	p.Gly12Val	Первый	10
		12:g.25398285C>G	p.Gly12Arg	Первый	2
		12:g.25398281C>T	p.Gly13Asp	Первый	1
		12:g.25398282C>A	p.Gly13Cys	Первый	1
		12:g.25398283A>T	p.Gly12=	Первый	1
		12:g.25398285C>A	p.Gly12Cys	Первый	1
		12:g.25380276T>C	p.Gln61Arg	Второй	1
24	<i>TP53</i>	17:g.7577121G>A	p.Arg273Cys	Второй	2
		17:g.7578212G>A	p.Arg213Ter	Второй	2
		17:g.7574002CG>C	p.Arg342GlufsTer3	Второй	1
		17:g.7577094G>A	p.Arg282Trp	Второй	1
		17:g.7577022G>A	p.Arg306Ter	Второй	1
		17:g.7577106G>A	p.Pro278Ser	Второй	1
		17:g.7577142C>T	p.Gly266Arg	Второй	1
		17:g.7577526AG>A	p.Leu252SerfsTer93	Второй	1
		17:g.7577538C>T	p.Arg248Gln	Второй	1
		17:g.7577547C>A	p.Gly245Val	Второй	1
		17:g.7577566T>TA	p.Asn239Ter	Первый	1
		17:g.7577581A>T	p.Tyr234Asn	Второй	1
		17:g.7577593TAC>T	p.Cys229TyrfsTer10	Второй	1
		17:g.7578190T>C	p.Tyr220Cys	Второй	1
		17:g.7578246CA>C	p.Leu201CysfsTer46	Второй	1
		17:g.7578271T>C	p.His193Arg	Второй	1
		17:g.7578394T>C	p.His179Arg	Второй	1
		17:g.7578406C>T	p.Arg175His	Второй	1
		17:g.7578451ATGGCGC>A	p.Arg158_Met160delinsLeu	Второй	1
		17:g.7578524G>A	p.Gln136Ter	Второй	1
17:g.7578530A>G	p.Phe134Leu	Второй	1		
17:g.7579590A>AAC	p.Ser33ValfsTer12	Второй	1		
3	<i>CDKN2A</i>	9:g.21971024GGCCCCA GGCATCGCGCACGTC CAGCCGC>G	p.Arg103ValfsTer34	Второй	1
		9:g.21971138C>CGGCG	p.Asp74ArgfsTer47	Второй	1
		9:g.21971071AC>A	p.Val96CysfsTer50	Второй	1

Больные РПЖ с вариантом в определённом гене (n=)	Ген	Геномный вариант	Аминокислотная замена	Класс патогенности (Tier)	Количество больных РПЖ с геномным вариантом (n=)
2	<i>ATM</i>	11:g.108121561C>T	p.Arg457Ter	Второй	1
		11:g.108142034A>AT	p.Val994CysfsTer16	Второй	1
		11:g.108150238GAGAT TCTTCC>G	p.Asp1103GlyfsTer3	Второй	1
2	<i>SMAD4</i>	18:g.48593417G>T	p.Glu390Ter	Первый	1
		18:g.48603023C>T	p.Gln442Ter	Первый	1
2	<i>RBI</i>	13:g.48955538C>T	p.Arg552Ter	Второй	1
		13:g.48881511GGGAGA AAGTT>G	p.Trp78CysfsTer30	Второй	1
1	<i>EGFR</i>	7:g.55221822C>T	p.Ala289Val	Второй	1
1	<i>KMT2C</i>	7:g.151859391CTG>C	p.Gln3757GluTer18	Второй	1
1	<i>PDGFRA</i>	4:g.55131161G>A	p.Cys235Tyr	Второй	1
1	<i>SETBP1</i>	18:g.42531425G>A	p.Ser707Asn	Второй	1
1	<i>SF3B1</i>	2:g.198266834T>C	p.Lys700Glu	Второй	1
Суммарное количество больных РПЖ с соматическими геномными вариантами 1-го и 2-го классов патогенности.					35 (87,5%)*

* Количество выявленных геномных вариантов отличается от суммарного количества больных, тк у 25 больных было выявлено одновременно несколько геномных вариантов.

У 35 больных (87,5%) выявлены соматические геномные варианты 1-го и 2-го класса патогенности в 11 генах согласно Руководству по интерпретации клинически значимых соматических мутаций при солидных опухолях, выявленных методом секвенирования следующего поколения (NGS), с целью их клинического использования [1]. У наибольшего количества больных РПЖ (n=33) были выявлены соматические варианты в гене *KRAS*, при этом у 2 больных наблюдалось одновременно 2 различных варианта. У 24 больных были обнаружены соматические варианты в гене *TP53*. Также были выявлены соматические варианты в генах *CDKN2A* у 3 больных, в *ATM* - у 2 больных (при этом у одного больного - одновременно 2 различных варианта), в *SMAD4* - также у 2 больных, в *RBI* - у 2 больных, в *EGFR* - у 1, в *KMT2C* - у 1, в *PDGFRA* - у 1, в *SETBP1* - у 1 и в *SF3B1* - также у 1 больного.

В целом, в проанализированных образцах выявлены различные уровни ТМВ в диапазоне от 0,02 до 14,3 мут/Мб (среднее 2,55±2,32; медиана 2,23). Высокая ТМВ (≥ 10 мут/Мб) выявлена у 1 больного (2,5%) РПЖ. Результаты проведенного NGS показали, что у всех больных РПЖ (n=40) был стабильный микросателлитный статус.

С целью выполнения дальнейшего корректного анализа и для лучшей визуализации ландшафта геномных вариантов пациент с экстремально высоким количеством геномных вариантов в гене *PWRN1* (более 300 вариантов) был удален из тепловой карты (хитмэп) В оставшейся популяции геномные варианты были отсортированы в порядке убывания.

Для визуализации ландшафта герминальных геномных вариантов была построена тепловая карта (хитмэп) (Рисунок 63).

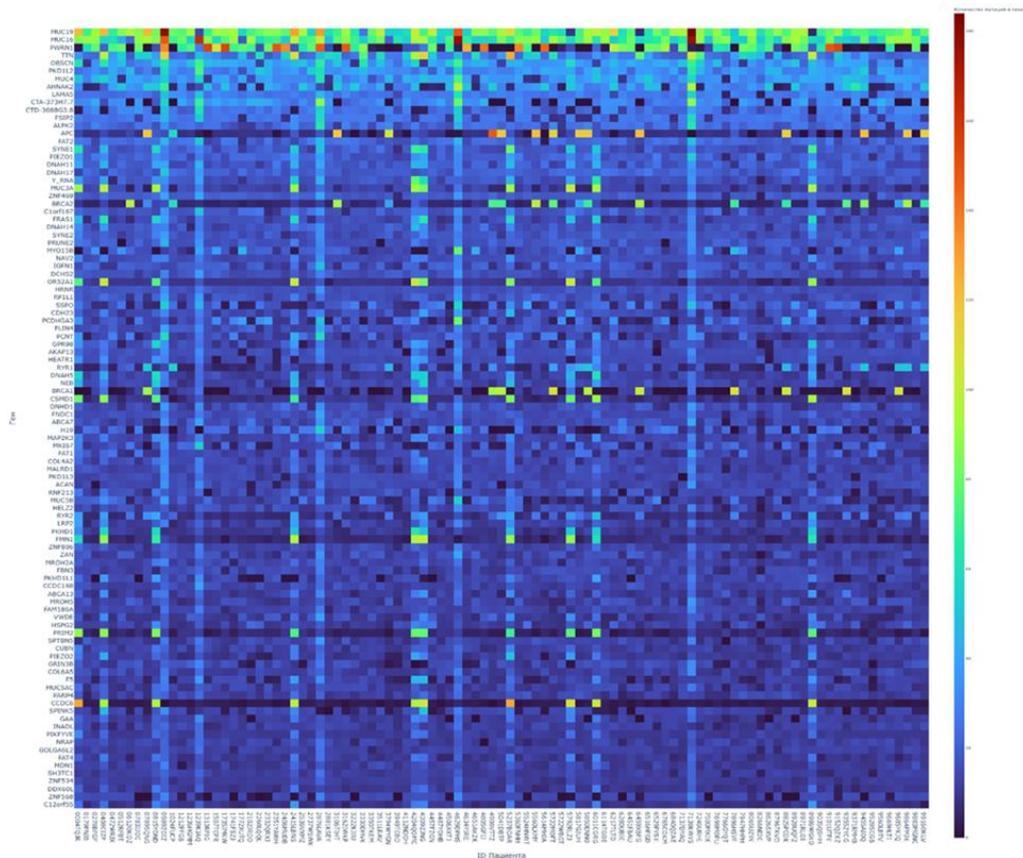


Рисунок 63. Фрагмент ландшафта геномных вариантов - тепловой карты типа “Хитмэп” (heat) распределения герминальных геномных вариантов для 100 генов (топ 100) с наиболее высоким числом геномных вариантов в исследуемой группе пациентов по данным полноэкзомного NGS в группе больных ТНРМЖ, после исключения т.н. “выброса” - пациента с экстремально высоким количеством геномных вариантов в гене *PWRN1*. Гены (по вертикали) отсортированы от большего числа геномных вариантов к меньшему. Набор цветов "торнадо": темные цвета иллюстрируют меньшее число геномных вариантов, светлые и яркие цвета - большее число

Для каждой пары генов была построена корреляционная матрица герминальных геномных вариантов с применением критерия Пирсона, обнаружив множество значимых положительных и отрицательных корреляций. Вид и форма корреляционной матрицы проиллюстрированы на примере фрагмента (Рисунок 64).

SYMBOL	MUC19	MUC16	PWRN1	TTN	OBSCN	PKD1L2	MUC4	AHNAK2	LAMA5	CTA-373H7.7	...	PIKFYVE	NRAP	GOLGA6L2	FAT4	MDN1	SH3TC1	ZNF534	DDX60L	ZNF568	C12orf55
MUC19	1.000000	0.476208	0.203674	0.252443	0.338760	0.176474	0.253802	0.386405	0.442186	0.380667	...	0.288717	0.415398	0.316026	0.273119	0.388767	0.399179	0.501164	0.273442	0.330756	0.180190
MUC16	0.476208	1.000000	0.073464	0.233350	0.399982	0.416885	0.299755	0.367174	0.618874	0.445301	...	0.373086	0.456491	0.560246	0.334006	0.420505	0.574695	0.608879	0.370837	0.164389	0.279374
PWRN1	0.203674	0.073464	1.000000	-0.135913	0.265730	0.108546	0.023316	0.255432	0.164089	0.434498	...	-0.077281	0.131175	0.289430	-0.061867	-0.050296	0.102652	0.092603	-0.077971	-0.056073	-0.120222
TTN	0.252443	0.233350	-0.135913	1.000000	0.263509	0.212165	0.183567	0.051856	0.205422	0.154927	...	0.571624	0.252123	0.230049	0.624324	0.514256	0.433271	0.462267	0.529001	0.220344	0.612206
OBSCN	0.338760	0.399982	0.265730	0.263509	1.000000	0.342171	0.368713	0.435851	0.517393	0.429061	...	0.406595	0.473006	0.539963	0.377854	0.296007	0.522354	0.594280	0.342606	0.172833	0.316047

Рисунок 64. Часть корреляционной матрицы Пирсона для генов с герминальными геномными вариантами по данным полноэкзомного NGS в группе больных ТНRMЖ

Для лучшей визуализации положительных корреляций в большом объеме данных была создана графическая иллюстрация в форме тепловой карты (Хитмэп) (Рисунок 65).

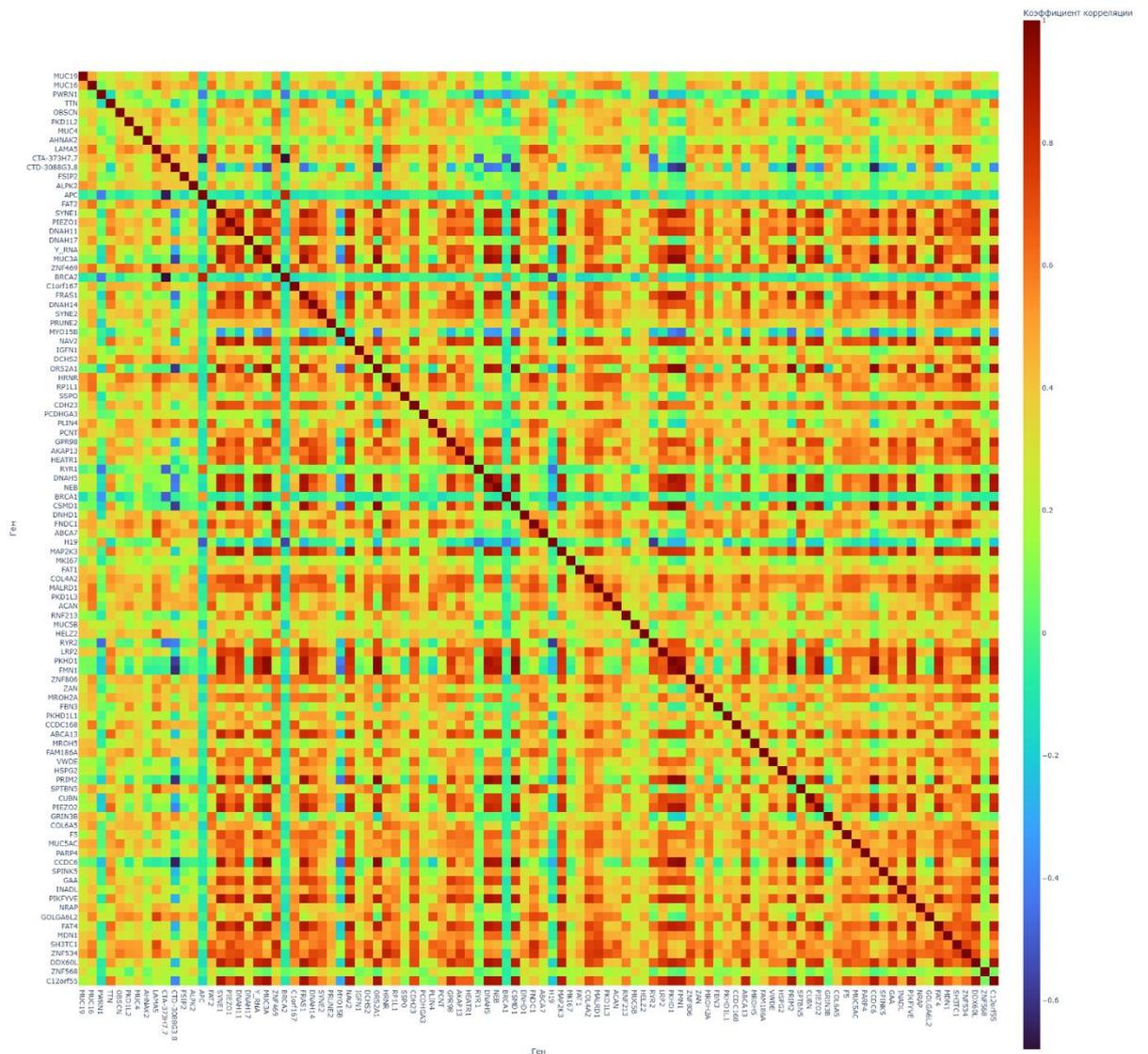


Рисунок 65. Графическая иллюстрация (тепловая карта) корреляционной матрицы Пирсона для 193 генов у 98 больных с герминальными геномными вариантами по данным полноэкзомного NGS в группе больных ТНRMЖ. Темные цвета иллюстрируют отрицательную корреляцию, светлые цвета - положительные коэффициенты корреляции

Значимыми для подбора персонализированного лекарственного лечения ТНРМЖ являются патогенные и вероятно патогенные герминальные геномные варианты в генах, ассоциированных с повышенным риском развития РМЖ и дефицитом репарации путем гомологичной рекомбинации. Далее приведено детализированное описание ландшафта патогенных и вероятно патогенных герминальных геномных вариантов в таких генах в когорте 100 больных ТНРМЖ.

Результаты проведенного полноэкзомного высокопроизводительного секвенирования показали наличие патогенных и вероятно патогенных герминальных геномных вариантов у 28 больных (28%) ТНРМЖ в 10 генах (Таблица 31).

Таблица 31. Патогенные и вероятно патогенные герминальные геномные варианты, выявленные в когорте 100 больных ТНРМЖ разных возрастных подгрупп.

Больные ТНРМЖ с вариантом в гене (n=)	Ген	Геномный вариант	Аминокислотная замена	rs	Патогенность	Больные ТНРМЖ 18-39 лет (n=)	Больные ТНРМЖ ≥40 лет (n=)
18	<i>BRCA1</i>	17:g.41209079T>TG	p.Gln1777ProfsTer74	rs80357906	Патогенный вариант	5	5
		17:g.41234451G>A	p.Arg1443Ter	rs41293455	Патогенный вариант	-	2
		17:g.41197784G>A	p.Arg1788Ter	rs41293465	Патогенный вариант	-	1
		17:g.41215898G>C	p.Ser1668Arg	rs80357094	Патогенный вариант	1	-
		17:g.41243776CCT>C	p.Glu1257GlyfsTer9	rs80357579	Патогенный вариант	-	1
		17:g.41244254AAG>A	p.Leu1098SerfsTer4	rs80357992	Патогенный вариант	-	1
		17:g.41276047C>CT	p.Glu23ArgfsTer18	rs80357783	Патогенный вариант	-	1
		17:g.41276113T>C	p.Met1?	rs80357287	Патогенный вариант	1	-
4	<i>BRCA2</i>	13:g.32911321TA>T	p.Asp946IlefsTer14	rs80359356	Патогенный вариант	1	-
		13:g.32915082CTG>C	p.Glu2198AsnfsTer4	rs80359605	Патогенный вариант	-	1
		13:g.32936829A>G	p.Arg2659Gly	rs80359026	Патогенный вариант	-	1
		13:g.32911297TAAA C>T	p.Ala938ProfsTer21	rs80359351	Патогенный вариант	-	1
1	<i>BARD1</i>	2:g.215595181T>TC AТАCTTTTCTTCC TGTTCА	p.Glu652ValfsTer69	rs587780024	Вероятно патогенный вариант	-	1
1	<i>FANCA</i>	16:g.89858363GCAG A>G	p.Val398AlafsTer16	rs2040090895	Патогенный вариант	1	-
1	<i>FANCI</i>	15:g.89315318C>T	p.Arg1285X	rs121918164	Патогенный вариант	-	1
1	<i>FANCL</i>	2:g.58241312A>G	p.Met1Thr	rs761291501	Патогенный вариант	-	1

Больные ТНРМЖ с вариантами в гене (n=)	Ген	Геномный вариант	Аминокислотная замена	rs	Патогенность	Больные ТНРМЖ 18-39 лет (n=)	Больные ТНРМЖ ≥40 лет (n=)
1	<i>FANCM</i>	14:g.45159263C>T	p.Gln522X	-	Патогенный вариант	-	1
1	<i>PMS2</i>	7:g.6035165C>T	p.Lys301=	rs267608153	Вероятно патогенный вариант	-	1
1	<i>RAD52</i>	12:g.925505C>G	p.Gly163Ala	-	Вероятно патогенный вариант	-	1
1	<i>SLX4</i>	16:g.3658472T>TG	p.Gln165ProfsTer121	-	Патогенный вариант	-	1
Суммарное количество больных ТНРМЖ с патогенными и вероятно патогенными герминальными геномными вариантами. У 2 больных было выявлено одновременно 2 герминальных геномных варианта							28 (28%)

* - Геномные координаты вариантов даны согласно сборке генома hg37.

У наибольшего количества ТНРМЖ (n=18) были выявлены патогенные герминальные геномные варианты в гене *BRCA1*. У 4 больных были выявлены патогенные герминальные варианты в гене *BRCA2*. Также в когорте больных ТНРМЖ были выявлены герминальные геномные варианты в генах *BARD1* - у 1, *FANCA* - у 1, *FANCI* - у 1, *FANCL* - 1, *FANCM* - 1, *PMS2* -1, *RAD52* - 1 и в *SLX4* - также у 1 больного. При этом у 2 больных было выявлено одновременно 2 герминальных геномных варианта (1- *BRCA1* и *FANCI*, 2 - *BRCA1* и *SLX4*).

Далее, мы стратифицировали выявленные патогенные и вероятно патогенные герминальные геномные варианты в зависимости от возраста больных на момент постановки диагноза ТНРМЖ, так как в исследуемой когорте 21 больная была моложе 40 лет, 79 женщин - в возрасте 40 лет и старше на момент постановки диагноза ТНРМЖ.

В подгруппе больных ТНРМЖ моложе 40 лет на момент постановки диагноза у 9 больных (42,85%) из 21 патогенные и вероятно патогенные герминальные геномные варианты были выявлены в 3 генах (*BRCA1*, *BRCA2*, *FANCA*). При этом наиболее часто (у 7 больных) герминальные варианты наблюдались в гене *BRCA1*. В подгруппе больных ТНРМЖ в возрасте более 40 лет на момент постановки диагноза у 19 больных (24,05%) из 79 патогенные и вероятно патогенные герминальные геномные варианты были выявлены в 9 генах (*BRCA1*, *BRCA2*, *BARD1*, *FANCI*, *FANCL*, *FANCM*, *PMS2*, *RAD52*, *SLX4*). Присутствие одновременно 2 герминальных геномных вариантов (1- *BRCA1* и *FANCI*, 2 - *BRCA1* и *SLX4*) было зафиксировано у 2 больных в возрасте более 40 лет. Таким образом, в молодом возрасте (до 40 лет) патогенные и вероятно-патогенные герминальные геномные варианты наблюдались у большей доли больных ТНРМЖ по сравнению с группой больных старше 40 лет.

3.6.2.2. Соматический геномный профиль

Геномный профиль ТНРМЖ, выявленный при помощи разработанных адаптированного протокола пробоподготовки и секвенирования, а также соматического биоинформатического конвейера, оказался неоднородным. Всего выявлено 14328 генов, содержащих геномные варианты в когорте из 100 больных.

Для визуализации соматического геномного профиля была построена тепловая карта (хитмэп) (Рисунок 66).

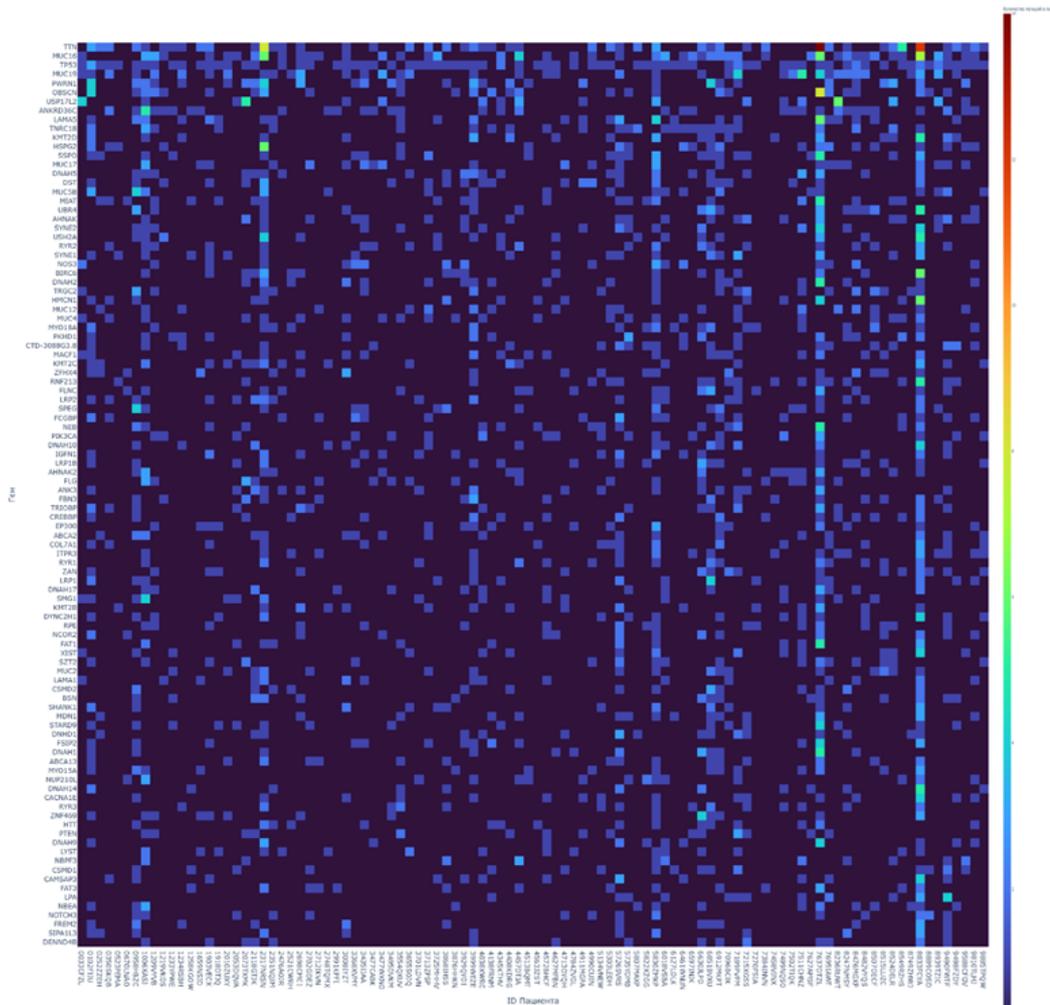


Рисунок 66. Тепловая карта (Хитмэп) распределения соматических геномных вариантов по данным полноэкзомного NGS в группе больных ТНРМЖ, выполнена сортировка генов от большего числа геномных вариантов в группе к меньшему. На рисунке показаны 100 генов с наибольшим числом геномных вариантов. Набор цветов "торнадо": темные цвета иллюстрируют меньшее число геномных вариантов, светлые и яркие цвета - большее число

Значимыми для подбора персонализированного лекарственного лечения ТНРМЖ являются соматические геномные варианты 1-го и 2-го классов патогенности. Далее приведено детализированное описание ландшафта соматических геномных вариантов 1-го и 2-го классов патогенности в когорте больных ТНРМЖ.

Результаты проведенного полноэкзомного высокопроизводительного секвенирования показали наличие 141 соматического геномного варианта 1-го и 2-го классов патогенности в 33 генах у 84 больных (84%) ТНРМЖ. Из них у наибольшего количества больных ТНРМЖ ($n=64$) были выявлены соматические геномные варианты в гене *TP53*, при этом у 3 больных было выявлено одновременно 2 различных варианта в данном гене. У 16 больных выявлены соматические геномные варианты в гене *PIK3CA*, при этом у одной больной было выявлено одновременно 2 варианта). У 9 больных были выявлены геномные варианты в гене *PTEN*, при это у одной больной было выявлено одновременно 2 варианта. В гене *SETBP1* варианты были обнаружены у 5 больных, в гене *RB1* - также у 5 больных.

Также в когорте больных ТНРМЖ были выявлены соматические геномные варианты в генах *KMT2C* - у 3 больных, *ASXL1* - у 1 больной (2 геномных варианта одновременно), *NOTCH1* - у 2, *KRAS* - у 2, *BRCA1* - у 2, *FBXW7* - у 2, *STK11* - у 2, *PTCH1* - у 2, *ESR1* - у 1, *NOTCH2* - у 1, *WT1* - у 1, *AKT1* - у 1, *BAP1* - у 1, *CDKN2A* - у 1, *BIRC3* - у 1, *ERBB2* - у 1, *EZH2* - у 1, *CTNNB1* - у 1, *GNA11* - у 1, *PRDM1* - у 1, *NRAS* - у 1, *PIK3CB* - у 1, *SMARCA4* - у 1, *PTPRB* - у 1, *SF3B1* - у 1, *SMARCB1* - у 1, *TERT* - у 1, *TSC1* - также у 1 больной.

В таблице 32 представлена информация о том, у какого числа больных был выявлен определенный соматический геномный вариант. При этом важно отметить, что у 32 больных наблюдалось наличие одновременно нескольких соматических геномных вариантов в различных генах либо в одном гене (от 2 до 6).

Таблица 32. Соматические геномные варианты 1-го и 2-го класса патогенности, выявленные в когорте 100 больных ТНРМЖ разных возрастных подгрупп

Больные ТНРМЖ с вариантом в гене ($n =$)	Ген	Геномный вариант	Аминокислотная замена	Класс патогенности (Tier)	Больные ТНРМЖ 18-39 лет ($n =$)	Больные ТНРМЖ ≥ 40 лет ($n =$)
64	<i>TP53</i>	17:g.7578406C>T	p.Arg175His	Первый	1	3
		17:g.7577539G>A	p.Arg248Trp	Первый	1	4
		17:g.7577536T>C	p.Arg249Gly	Первый	-	1
		17:g.7577120C>T	p.Arg273His	Первый	-	1
		17:g.7577121G>A	p.Arg273Cys	Первый	1	1
		17:g.7578212G>A	p.Arg213Ter	Второй	-	3
		17:g.7574003G>A	p.Arg342Ter	Второй	2	-
		17:g.7577105G>C	p.Pro278Arg	Второй	-	2
		17:g.7573990T>TCATTCAG	p.Glu346AlafsTer3	Второй	-	1
17:g.7574006ACATCT>A	p.Glu339ValfsTer6	Второй	-	1		

Больные ТНРМЖ с варианто м в гене (n =)	Ген	Геномный вариант	Аминокислотная замена	Класс патогеннос ти (Tier)	Больные ТНРМЖ 18-39 лет (n =)	Больные ТНРМЖ ≥40 лет (n =)
		17:g.7574013G>GAAGC	p.Phe338LeufsTer10	Второй	-	1
		17:g.7577022G>A	p.Arg306Ter	Второй	1	-
		17:g.7577064T>A	p.Lys292Ter	Второй	-	1
		17:g.7577094G>C	p.Arg282Gly	Второй	-	1
		17:g.7577099C>G	p.Arg280Thr	Второй	-	1
		17:g.7577114C>T	p.Cys275Tyr	Второй	-	1
		17:g.7577120C>G	p.Arg273Pro	Второй	-	1
		17:g.7577120CG>C	p.Arg273ValfsTer72	Второй	-	1
		17:g.7577124C>T	p.Val272Met	Второй	-	1
		17:g.7577137C>CCGT	p.Arg267dup	Второй	-	1
		17:g.7577509C>A	p.Glu258Ter	Второй	-	1
		17:g.7577509C>T	p.Glu258Lys	Второй	-	1
		17:g.7577511A>G	p.Leu257Pro	Второй	-	1
		17:g.7577580T>C	p.Tyr234Cys	Второй	1	-
		17:g.7577579G>C	p.Tyr234Ter	Второй	-	1
		17:g.7577579G>T	p.Tyr234Ter	Второй	-	1
		17:g.7577526AG>A	p.Leu252SerfsTer93	Второй	-	1
		17:g.7578191A>G	p.Tyr220His	Второй	1	-
		17:g.7577595C>CG	p.Cys229SerfsTer11	Второй	-	1
		17:g.7578266T>A	p.Ile195Phe	Второй	-	1
		17:g.7577593TAC>T	p.Cys229TyrfsTer10	Второй	1	-
		17:g.7578395G>A	p.His179Tyr	Второй	1	-
		17:g.7578271T>C	p.His193Arg	Второй	-	1
		17:g.7578382G>C	p.Ser183Ter	Второй	-	1
		17:g.7578397TG>T	p.His178ThrfsTer69	Второй	-	1
		17:g.7578403C>A	p.Cys176Phe	Второй	-	1
		17:g.7578406CG>C	p.Arg175AlafsTer72	Второй	-	1
		17:g.7578413C>G	p.Val173Leu	Второй	-	1
		17:g.7578406C>G	p.Arg175Pro	Второй	-	1
		17:g.7578442T>C	p.Tyr163Cys	Второй	1	-
		17:g.7578413C>T	p.Val173Met	Второй	-	1
		17:g.7578441G>GCA	p.Lys164AlafsTer7	Второй	-	1
		17:g.7578431G>A	p.Gln167Ter	Второй	-	1
		17:g.7579329T>C	p.Lys120Glu	Второй	1	-
		17:g.7579366G>T	p.Tyr107Ter	Второй	-	1
		17:g.7578526C>T	p.Cys135Tyr	Второй	-	1
		17:g.7578527A>C	p.Cys135Gly	Второй	-	1
		17:g.7579400GA>G	p.Ser96LeufsTer27	Второй	1	-
		17:g.7579358CG>C	p.Arg110ValfsTer13	Второй	-	1
		17:g.7579360GAAACC>G	p.Gly108ProfsTer39	Второй	-	1
		17:g.7578534C>G	p.Lys132Asn	Второй	-	1
		17:g.7579363ACCGTAGCTGC CCTGGTAGGTTTTCTGGGA AGGGACAGAAGATGACAG GGGCCAGGAGGGGGCTGG TGCAGGGGC>A	p.Ala84PhefsTer40	Второй	1	-
		17:g.7578442T>G	p.Tyr163Ser	Второй	-	1
		17:g.7579389GGGAA>G	p.Ser99ArgfsTer23	Второй	-	1
		17:g.7579358C>G	p.Arg110Pro	Второй	1	-

Больные ТНРМЖ с варианто м в гене (n =)	Ген	Геномный вариант	Аминокислотная замена	Класс патогеннос ти (Tier)	Больные ТНРМЖ 18-39 лет (n =)	Больные ТНРМЖ ≥40 лет (n =)
16	<i>PIK3CA</i>	3:g.178952085A>G	p.His1047Arg	Первый	-	5
		3:g.178936082G>A	p.Glu542Lys	Первый	-	3
		3:g.178936091G>A	p.Glu545Lys	Первый	1	2
		3:g.178936094C>A	p.Gln546Lys	Первый	-	1
		3:g.178951973A>G	p.Met1010Val	Первый	-	1
		3:g.178928071ATGGATTAGA AGATTTGCTGAACCCAT> A	p.Leu452_Gly460del	Первый	-	1
		3:g.178952029A>C	p.Leu1028Phe	Первый	1	-
		3:g.178921553T>A	p.Asn345Lys	Первый	-	1
		3:g.178928220A>G	p.Glu469Gly	Первый	-	1
9	<i>PTEN</i>	10:g.89685269G>A		Первый	-	1
		10:g.89717609G>C		Первый	1	-
		10:g.89720764T>TATAG	p.Glu307AspfsTer6	Первый	-	1
		10:g.89692978CTATGGGGAA G>C	p.Tyr155Ter	Первый	-	1
		10:g.89720693GGACCAGAGG AAACCTCAGAAAAAGTA>G	p.Pro283LysfsTer6	Первый	-	1
		10:g.89717672C>T	p.Arg233Ter	Первый	-	1
		10:g.89690802G>T		Первый	1	-
		10:g.89692793C>T	p.His93Tyr	Второй	-	1
		10:g.89692922T>C	p.Cys136Arg	Второй	-	1
		10:g.89692792C>G	p.Asp92Glu	Второй	-	1
5	<i>SETBP1</i>	18:g.42532907G>A	p.Ser1201Asn	Второй	-	1
		18:g.42530798G>A	p.Arg498Gln	Второй	1	-
		18:g.42531353C>T	p.Ser683Phe	Второй	-	1
		18:g.42531017A>C	p.Tyr571Ser	Второй	-	1
		18:g.42531637C>T	p.Pro778Ser	Второй	-	1
5	<i>RBI</i>	13:g.49027200TTGTCCTCTTA ATCTTCCTCTCCAGAATAA TCACACTGCAGCAGATA>T		Второй	1	-
		13:g.49039399C>G	p.Ser795Ter	Второй	-	1
		13:g.48937095T>C		Второй	-	1
		13:g.48941648C>T	p.Arg320Ter	Второй	-	1
		13:g.48916768G>T	p.Gly100Ter	Второй	-	1
3	<i>KMT2C</i>	7:g.151921114A>T	p.Cys1103Ter	Первый	1	1
		7:g.151836873GC>G	p.Leu4783TrpfsTer2 7	Первый	-	1
1	<i>ASXL1</i>	20:g.31021438CG>C	p.Glu480AsnfsTer22 3	Второй	1	-
		20:g.31021442T>A	p.Phe481Ile	Второй	1	-
2	<i>NOTCH1</i>	9:g.139413900C>T	p.Trp287Ter	Второй	-	1
		9:g.139403482G>T	p.Ser1004Ter	Второй	-	1
1	<i>ESR1</i>	6:g.152419926A>G	p.Asp538Gly	Первый	-	1

Больные ТНРМЖ с варианто м в гене (n =)	Ген	Геномный вариант	Аминокислотная замена	Класс патогеннос ти (Tier)	Больные ТНРМЖ 18-39 лет (n =)	Больные ТНРМЖ ≥40 лет (n =)
1	<i>NOTCH2</i>	1:g.120458324G>A	p.Gln2341Ter	Первый	-	1
2	<i>KRAS</i>	12:g.25398285C>G	p.Gly12Arg	Второй	-	2
2	<i>BRCA1</i>	17:g.41222994AC>A	p.Val1667SerfsTer1 2	Второй	-	1
		17:g.41245278AC>A	p.Val757PhefsTer8	Второй	-	1
2	<i>FBXW7</i>	4:g.153250883G>A	p.Arg393Ter	Второй	-	1
		4:g.153249361T>A	p.Arg473Ter	Второй	-	1
2	<i>STK11</i>	19:g.1207078G>A	p.Gly56Arg	Второй	-	1
		19:g.1207155G>C	p.Lys81Asn	Второй	-	1
2	<i>PTCH1</i>	9:g.98242356AG>A	p.Val322PhefsTer2	Второй	1	-
		9:g.98244230C>T		Второй	-	1
1	<i>WT1</i>	11:g.32417910G>A	p.Ser381Leu	Второй	-	1
1	<i>AKT1</i>	14:g.105246551C>T	p.Glu17Lys	Второй	-	1
1	<i>BAP1</i>	3:g.52443586G>A	p.Gln36Ter	Второй	-	1
1	<i>CDKN2A</i>	9:g.21971053GC>G	p.Ala102ArgfsTer44	Второй	-	1
1	<i>BIRC3</i>	11:g.102206800A>AC	p.Gln477ProfsTer4	Второй	-	1
1	<i>ERBB2</i>	17:g.37880220T>C	p.Leu755Ser	Второй	-	1
1	<i>EZH2</i>	7:g.148516705G>A	p.Gln328Ter	Второй	-	1
1	<i>CTNNB1</i>	3:g.41266104G>A	p.Gly34Glu	Второй	-	1
1	<i>GNAI1</i>	19:g.3115013G>A	p.Arg183His	Второй	-	1
1	<i>PRDM1</i>	6:g.106536207T>A	p.Cys58Ter	Второй	-	1
1	<i>NRAS</i>	1:g.115256530G>T	p.Gln61Lys	Второй	1	-
1	<i>PIK3CB</i>	3:g.138374245C>A	p.Asp1067Tyr	Второй	-	1
1	<i>SMARCA4</i>	19:g.11129654C>A	p.Tyr820Ter	Второй	-	1
1	<i>PTPRB</i>	12:g.70964865AT>A	p.Ile1104PhefsTer4	Второй	-	1
1	<i>SF3B1</i>	2:g.198266834T>C	p.Lys700Glu	Второй	-	1
1	<i>SMARCB1</i>	22:g.24143240C>T	p.Arg158Ter	Второй	-	1

Продолжение Таблицы 32

Больные ТНРМЖ с вариантами в гене (n=)	Ген	Геномный вариант	Аминокислотная замена	Класс патогенности (Tier)	Больные ТНРМЖ 18-39 лет (n=)	Больные ТНРМЖ ≥40 лет (n=)
1	<i>TERT</i>	5:g.1266603AGGTGAGGT>A	p.Pro875HisfsTer16	Второй	1	-
1	<i>TSCI</i>	9:g.135796805G>A	p.Arg228Ter	Второй	-	1
Суммарное количество больных ТНРМЖ с соматическими геномными вариантами 1-го и 2-го классов патогенности						84*

Аббревиатуры: ТНРМЖ - трижды негативный рак молочной железы

* суммарное количество больных ТНРМЖ отличается от суммарного количества выявленных соматических геномных вариантов 1-го и 2-го классов патогенности поскольку у 32 больных было выявлено наличие нескольких соматических вариантов одновременно

Выявленные соматические геномные варианты 1-го и 2-го классов патогенности были стратифицированы в зависимости от возраста больных на момент постановки диагноза ТНРМЖ. Как было указано ранее, 21 больная была моложе 40 лет, 79 - в возрасте 40 лет и старше.

В подгруппе моложе 40 лет соматические геномные варианты 1-го и 2-го классов патогенности были выявлены у 18 больных ТНРМЖ из 21 (85,7%) в 10 генах (*ASXL1*, *KMT2C*, *NRAS*, *PIK3CA*, *PTCH1*, *PTEN*, *RB1*, *SETBP1*, *TERT*, *TP53*).

В подгруппе больных ТНРМЖ в возрасте более 40 лет соматические геномные варианты 1-го и 2-го классов были выявлены в 30 генах (*TP53*, *PIK3CA*, *AKT1*, *VAPB*, *BIRC3*, *BRCA1*, *CDKN2A*, *CTNNA1*, *ERBB2*, *ESR1*, *EZH2*, *FBXW7*, *GNAI1*, *KMT2C*, *KRAS*, *NOTCH1*, *NOTCH2*, *PIK3CB*, *PRDM1*, *PTCH1*, *PTEN*, *PTPRB*, *RB1*, *SETBP1*, *SF3B1*, *SMARCA4*, *SMARCB1*, *STK11*, *TSCI*, *WT1*) у 66 из 79 (83,5%). Таким образом, нами не было обнаружено статистически значимых различий в долях больных с соматическими геномными вариантами 1-го и 2-го классов патогенности в зависимости от возраста ($p = 0,7110$).

Молекулярно-генетическое исследование выявило стабильный микросателлитный статус опухоли у всех больных ТНРМЖ ($n=100$, 100%) В проанализированных образцах опухолевой ткани больных ТНРМЖ выявлены различные уровни ТМВ в диапазоне от 1,3721 мут/Мб до 98,8837 мут/Мб с медианой 5,66 мут/Мб и средним значением $10,02 \pm 14,68$. Высокая ТМВ (≥ 10 мут/Мб) выявлена у 19 больных (19%) ТНРМЖ.

По итогам выполненных молекулярно-генетических исследований было разработано и опубликовано учебно-методическое пособие “Молекулярно-генетические исследования в онкологии: учебно-методическое пособие” (г. Обнинск, ФГБУ «НМИЦ радиологии» Минздрава России, 2023. 82 с. Серия «Библиотека врача-онколога») (Приложение 7).

3.7. Рекомендации персонализированного лекарственного лечения по итогам анализа биологического материала больных раком поджелудочной железы и трижды негативным раком молочной железы

Мы применили разработанную СППВР для анализа данных полноэкзомного секвенирования биологического материала больных РПЖ (n=40) и ТНРМЖ (n=100). Разработанная СППВР использует информацию из озера данных, находящегося в фармакогеномной платформе, для подбора эффективного персонализированного лекарственного лечения, к которому по данным исследований наблюдается чувствительность при наличии выявленных у больного биомаркеров. СППВР также формирует рейтинг рекомендаций для прогнозирования наиболее эффективного терапевтического подхода.

Необходимо отметить, что рекомендации СППВР не являются обязывающими и окончательное решение о назначении того или иного противоопухолевого лекарственного препарата принимается лечащим врачом с учетом различных клинических факторов [400, 401].

Рекомендации для одинаковых вариантов у разных пациентов могут отличаться в связи с параметрами фильтрации СППВР (например, аллельной частотой варианта) и контекстом (например, наличием других вариантов, ассоциированных с резистентностью к препарату). Для вариантов с уровнем доказательности D (экспериментальное доказательство) рекомендации приводятся в случае совпадения типа рака в экспериментальной модели с типом рака, диагностированного у конкретного пациента.

СППВР сформировала рекомендации персонализированного лекарственного лечения для 36 больных РПЖ (90%). При этом у 25 больных было выявлено одновременно несколько биомаркеров (от 2 до 5), с учетом которых были сформированы рекомендации персонализированного лекарственного лечения.

Рекомендации персонализированного лекарственного лечения для больных РПЖ, сформированные разработанной СППВР на основе выявленных биомаркеров, а также данные о регистрации/одобрении рекомендуемых лекарственных препаратов на территории Российской Федерации, ЕС или США представлены в таблице 33.

Необходимо отметить, что препараты, не отмеченные как зарегистрированные для лечения РПЖ (он-лейбл), могут быть зарегистрированы для других типов ЗНО либо находиться на стадии исследований как для данного типа ЗНО, так и для других типов ЗНО. Подробная информация о статусе изучения препаратов приводится в отчетах СППВР.

Таблица 33. Рекомендации персонализированного лекарственного лечения, сформированные СППВР на основе геномных вариантов и показателей ТМВ, выявленных по итогам высокопроизводительного полноэкзомного секвенирования биологического материала больных раком поджелудочной железы

Кол-больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при раке поджелудочной железы*
1	ATM	11:g.108175540C>T (p.Gln1879Ter)	герминальный	Olaparib	D: Доклиническое доказательство	он-лейбл
				Rucaparib	D: Доклиническое доказательство	
				Talazoparib	D: Доклиническое доказательство	
				Niraparib	D: Доклиническое доказательство	
				Препараты платины	D: Доклиническое доказательство	он-лейбл
1		11:g.108205832T>C (p.Val2716Ala)	герминальный	Olaparib	D: Доклиническое доказательство	он-лейбл
				Rucaparib	D: Доклиническое доказательство	
				Talazoparib	D: Доклиническое доказательство	
				Niraparib	D: Доклиническое доказательство	
				Препараты платины	D: Доклиническое доказательство	он-лейбл
1	EGFR	7:g.55221822C>T (p.Ala289Val)	соматический	Afatinib	B: Клиническое доказательство	
				Erlotinib	B: Клиническое доказательство	он-лейбл
				Gefitinib	B: Клиническое доказательство	
				Gefitinib,Erlotinib	B: Клиническое доказательство	
1	KRAS	12:g.25380276T>C (p.Gln61Arg)	соматический	Abemaciclib	B: Клиническое доказательство	
				Atezolizumab,Nivolumab	B: Клиническое доказательство	
				Binimetinib	B: Клиническое доказательство	
				Docetaxel	B: Клиническое доказательство	
				Docetaxel,Selumetinib	B: Клиническое доказательство	

Кол-больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при раке поджелудочной железы*
				Erlotinib, Docetaxel	В: Клиническое доказательство	
				Ингибитор MEK RO4987655	В: Клиническое доказательство	
				Pemetrexed, Trametinib, Docetaxel	В: Клиническое доказательство	
				Salirasib	В: Клиническое доказательство	
				SCH772984, Ингибитор PI3Kbeta AZD8186	D: Доклиническое доказательство	
				Selumetinib, Cetuximab	В: Клиническое доказательство	
				Selumetinib, Erlotinib	В: Клиническое доказательство	
				Sorafenib, Refametinib	В: Клиническое доказательство	
				Temsirolimus, Ridaforolimus	В: Клиническое доказательство	
				Teprotumumab, Erlotinib	В: Клиническое доказательство	
				Trametinib	В: Клиническое доказательство	
1		12:g.25398281C>T (p.Gly13Asp)	соматический	Trametinib	В: Клиническое доказательство	
1		12:g.25398282C>A (p.Gly13Cys)	соматический	Trametinib	В: Клиническое доказательство	
1		12:g.25398283A>T (p.Gly12=)	соматический	Trametinib	В: Клиническое доказательство	
10		12:g.25398284C>A (p.Gly12Val)	соматический	RMC-6236	В: Клиническое доказательство	
				Trametinib	В: Клиническое доказательство	
18		12:g.25398284C>T (p.Gly12Asp)	соматический	RMC-6236	В: Клиническое доказательство	
				RMC-9805	В: Клиническое доказательство	
				Trametinib	В: Клиническое доказательство	
1		12:g.25398285C>A (p.Gly12Cys)	соматический	Adagrasib	A: Утвержденная рекомендация	
				Sotorasib	A: Утвержденная рекомендация	

Кол-больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при раке поджелудочной железы*
				Trametinib	В: Клиническое доказательство	
2		12:g.25398285C>G (p.Gly12Arg)	соматический	RMC-6236	В: Клиническое доказательство	
				Trametinib	В: Клиническое доказательство	
1	PDGFRA	4:g.55131161G>A (p.Cys235Tyr)	соматический	Imatinib	А: Утвержденная рекомендация	
				Regorafenib	А: Утвержденная рекомендация	
				Sunitinib	А: Утвержденная рекомендация	
1	RB1	13:g.48881511GGGA GAAAGTT>G (p.Trp78CysfsTer30)	соматический	Химиотерапия	В: Клиническое доказательство	он-лейбл
1		13:g.48955538C>T (p.Arg552Ter)	соматический	Химиотерапия	В: Клиническое доказательство	он-лейбл
1	TP53	17:g.7574002CG>C (p.Arg342GlufsTer3)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib,Vorinostat	В: Клиническое доказательство	
1		17:g.7577022G>A (p.Arg306Ter)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib,Vorinostat	В: Клиническое доказательство	

Кол-больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при раке поджелудочной железы*
1		17:g.7577094G>A (p.Arg282Trp)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacididine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib,Vorinostat	В: Клиническое доказательство	
1		17:g.7577106G>A (p.Pro278Ser)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacididine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib,Vorinostat	В: Клиническое доказательство	
2		17:g.7577121G>A (p.Arg273Cys)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacididine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib,Vorinostat	В: Клиническое доказательство	
1		17:g.7577142C>T (p.Gly266Arg)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacididine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	

Кол-больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при раке поджелудочной железы*
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577526AG>A (p.Leu252SerfsTer93)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577538C>T (p.Arg248Gln)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577547C>A (p.Gly245Val)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577566T>TA (p.Asn239Ter)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	

Кол-больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при раке поджелудочной железы*
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577581A>T (p.Tyr234Asn)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577593TAC>T (p.Cys229TyrfsTer10)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578190T>C (p.Tyr220Cys)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
2		17:g.7578212G>A (p.Arg213Ter)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл

Кол-больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при раке поджелудочной железы*
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578246CA>C (p.Leu201CysfsTer46)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578271T>C (p.His193Arg)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578394T>C (p.His179Arg)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578406C>T (p.Arg175His)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	

Кол-больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при раке поджелудочной железы*
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacididine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib,Vorinostat	В: Клиническое доказательство	
1		17:g.7578451ATGGC GC>A (p.Arg158_Met160delinsLeu)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacididine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib,Vorinostat	В: Клиническое доказательство	
1		17:g.7578524G>A (p.Gln136Ter)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacididine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib,Vorinostat	В: Клиническое доказательство	
1		17:g.7578530A>G (p.Phe134Leu)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt,Azacididine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	

Кол-больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при раке поджелудочной железы*
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7579590A>AAC (p.Ser33ValfsTer12)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
Кол-во больных	Опухолевая мутационная нагрузка (ТМВ)			Рекомендация персонализированного лекарственного лечения	Степень доказательности	
1	Высокий уровень ТМВ (≥ 10 мутаций/Мб)			Pembrolizumab	А: Утвержденная рекомендация	он-лейбл

Общее количество больных РПЖ, для которых были сформулированы рекомендации персонализированного лекарственного лечения ($n=36$). У 25 больных было выявлено несколько биомаркеров (от 2 до 5), с учетом которых были сформированы рекомендации персонализированного лекарственного лечения.

*В столбце «Статус регистрации препарата в РФ/США/ЕС при раке поджелудочной железы» приведены данные о статусе регистрации препарата для терапии данного ЗНО независимо от наличия или отсутствия описываемого биомаркера.

Наибольшее количество различных рекомендаций персонализированного лечения - 16 различных препаратов/комбинаций препаратов: 1 - Abemaciclib, 2 Atezolizumab, Nivolumab, 3 - Binimetinib, 4 - Docetaxel, 5 - Docetaxel, Selumetinib, 6 - Erlotinib, Docetaxel, 7 - Ингибитор МЕК RO4987655, 8 - Pemetrexed, Trametinib, Docetaxel, 9 - Salirasib, 10 - SCH772984, Ингибитор PI3Kbeta AZD8186, 11 - Selumetinib, Cetuximab, 12 - Selumetinib, Erlotinib, 13 - Sorafenib, Refametinib, 14 - Temsirolimus, Ridaforolimus, 15 - Teprotumumab, Erlotinib, 16 - Trametinib - было сформировано СППВР для больного РПЖ с соматическим геномным вариантом 12:g.25380276T>C (p.Gln61Arg) в гене *KRAS*. Рекомендации для данного геномного варианта имели различные степени доказательности.

СППВР также сформировала рекомендации персонализированного лекарственного лечения для 83 больных ТНРМЖ (83%). При этом у 44 больных было выявлено одновременно несколько биомаркеров (от 2 до 7), с учетом которых были сформированы рекомендации персонализированного лекарственного лечения.

Рекомендации персонализированного лекарственного лечения больным ТНРМЖ, сформированные разработанной СППВР на основе выявленных геномных вариантов и уровня ТМВ, а также данные о степени доказательности и регистрации/одобрении рекомендуемых лекарственных препаратов на территории Российской Федерации, ЕС или США представлены в таблице 34.

Препараты, не отмеченные как зарегистрированные для лечения ТНРМЖ (он-лейбл), могут быть зарегистрированы для других типов ЗНО либо находиться на стадии исследований как для данного типа ЗНО, так и для других типов ЗНО. Подробная информация о статусе изучения препаратов приводится в отчетах СППВР.

Таблица 34. Рекомендации персонализированного лекарственного лечения, сформированные СППВР на основе геномных вариантов и показателей ТМВ, выявленных по итогам высокопроизводительного полноэкзомного секвенирования биологического материала больных ТНРМЖ

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
10	<i>BRCA1</i>	17:g.41209079T>TG (p.Gln1777ProfsTer74)	герминальный	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Niraparib	A: Утвержденная рекомендация	
2		17:g.41234451G>A (p.Arg1443Ter)	герминальный	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Niraparib	A: Утвержденная рекомендация	
1		17:g.41197784G>A (p.Arg1788Ter)	герминальный	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Niraparib	A: Утвержденная рекомендация	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
1		17:g.41215898G>C (p.Ser1668Arg)	герминальный	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Niraparib	A: Утвержденная рекомендация	
1		17:g.41243776CCT>C (p.Glu1257GlyfsTer9)	герминальный	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Niraparib	A: Утвержденная рекомендация	
1		17:g.41244254AAG>A (p.Leu1098SerfsTer4)	герминальный	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Niraparib	A: Утвержденная рекомендация	
1		17:g.41276047C>CT (p.Glu23ArgfsTer18)	герминальный	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Niraparib	A: Утвержденная рекомендация	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
1		17:g.41276113T>C (p.Met1?)	герминальный	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Niraparib	A: Утвержденная рекомендация	
1	<i>BRCA2</i>	13:g.32911321TA>T (p.Asp946IlefsTer14)	герминальный	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Niraparib	A: Утвержденная рекомендация	
1		13:g.32915082CTG>C (p.Glu2198AsnfsTer4)	герминальный	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Niraparib	A: Утвержденная рекомендация	
1		13:g.32936829A>G(p.Arg2659Gly)	герминальный	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Niraparib	A: Утвержденная рекомендация	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
1		13:g.32911297TAAAC>T (p.Ala938ProfsTer21)	герминальный	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Niraparib	A: Утвержденная рекомендация	
1	FANCA	16:g.89858363GCAGA>G (p.Val398AlafsTer16)	герминальный	Talazoparib + Enzalutamide	A: Утвержденная рекомендация	
				Olaparib	D: Доклиническое доказательство	он-лейбл
				Rucaparib	D: Доклиническое доказательство	
				Talazoparib	D: Доклиническое доказательство	он-лейбл
				Niraparib	D: Доклиническое доказательство	
				Препараты платины	D: Доклиническое доказательство	
1	FANCL	2:g.58241312A>G (p.Met1Thr)	герминальный	Olaparib	D: Доклиническое доказательство	он-лейбл
1	BARD1	2:g.215595181T>TCATAC TTTTCTTCCTGTTCA (p.Glu652ValfsTer69)	герминальный	Olaparib	D: Доклиническое доказательство	он-лейбл
				Rucaparib	D: Доклиническое доказательство	
				Talazoparib	D: Доклиническое доказательство	он-лейбл
				Niraparib	D: Доклиническое доказательство	
				Препараты платины	D: Доклиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
1	AKT1	14:g.105246551C>T (p.Glu17Lys)	соматический	Ингибитор Akt MK2206	D: Доклиническое доказательство	
				Capivasertib	B: Клиническое доказательство	
				Tensirolimus	B: Клиническое доказательство	
				Capivasertib + Fulvestrant	A: Утвержденная рекомендация	
1	BAP1	3:g.52443586G>A (p.Gln36Ter)	соматический	Ингибиторы EZH2	D: Доклиническое доказательство	
1	BRCA1	17:g.41222994AC>A (p.Val1667SerfsTer12)	соматический	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Niraparib	A: Утвержденная рекомендация	
1		17:g.41245278AC>A (p.Val757PhefsTer8)	соматический	Olaparib	A: Утвержденная рекомендация	он-лейбл
				Препараты платины	B: Клиническое доказательство	
				Rucaparib	A: Утвержденная рекомендация	
				Talazoparib	A: Утвержденная рекомендация	он-лейбл
				Niraparib	A: Утвержденная рекомендация	
1	CDKN2A	9:g.21971053GC>G (p.Ala102ArgfsTer44)	соматический	Letrozole, Palbociclib	C: Клинический случай	
				Abemaciclib	C: Клинический случай	
				Palbociclib	C: Клинический случай	
				Ribociclib	C: Клинический случай	
1	ERBB2	17:g.37880220T>C (p.Leu755Ser)	соматический	Neratinib	B: Клиническое доказательство	
				Ado-Trastuzumab Emtansine	B: Клиническое доказательство	
				Lapatinib	D: Доклиническое доказательство	
				Margetuximab	D: Доклиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Trastuzumab	D: Доклиническое доказательство	
				Pertuzumab	D: Доклиническое доказательство	
				Tucatinib	D: Доклиническое доказательство	
				Trastuzumab Deruxtecan	A: Утвержденная рекомендация	
1	ESR1	6:g.152419926A>G (p.Asp538Gly)	соматический	Elacestrant	A: Утвержденная рекомендация	
				Letrozole,Palbociclib	B: Клиническое доказательство	
1	EZH2	7:g.148516705G>A (p.Gln328Ter)	соматический	Tazemetostat	A: Утвержденная рекомендация	
1	FBXW7	4:g.153249361T>A (p.Arg473Ter)	соматический	Lunresertib+Camonsertib	B: Клиническое доказательство	
				Ингибитор MTOR	B: Клиническое доказательство	
1		4:g.153250883G>A (p.Arg393Ter)	соматический	Lunresertib+Camonsertib	B: Клиническое доказательство	
				Ингибитор MTOR	B: Клиническое доказательство	
1	GNA11	19:g.3115013G>A (p.Arg183His)	соматический	Cabozantinib	B: Клиническое доказательство	
2	KRAS	12:g.25398285C>G (p.Gly12Arg)	соматический	RMC-6236	B: Клиническое доказательство	
				Trametinib	B: Клиническое доказательство	
1	NOTCH1	9:g.139413900C>T (p.Trp287Ter)	соматический	Ингибиторы гамма-секретазы	D: Доклиническое доказательство	
				Ингибиторы гамма-секретазы (Ro4929097,Pf-03084014,Mk-0752,etc)	B: Клиническое доказательство	
				OMP-52M51	B: Клиническое доказательство	
1	NOTCH2	1:g.120458324G>A (p.Gln2341Ter)	соматический	Mk-0752	B: Клиническое доказательство	
1	NRAS	1:g.115256530G>T (p.Gln61Lys)	соматический	Binimetinib	B: Клиническое доказательство	
				Binimetinib,Ribociclib	B: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Docetaxel, Trametinib	В: Клиническое доказательство	
				Trametinib	В: Клиническое доказательство	
				Cobimetinib	В: Клиническое доказательство	
1	PIK3CA	3:g.178921553T>A (p.Asn345Lys)	соматический	Alpelisib + Fulvestrant	А: Утвержденная рекомендация	
				Capivasertib+Fulvestrant	А: Утвержденная рекомендация	
				Everolimus	В: Клиническое доказательство	
				Tamoxifen	В: Клиническое доказательство	
1		3:g.178928071ATGGATTA GAAGATTTGCTGAACC STAT>A (p.Leu452 Gly460del)	соматический	Alpelisib + Fulvestrant	А: Утвержденная рекомендация	
				Capivasertib+Fulvestrant	А: Утвержденная рекомендация	
				Everolimus	В: Клиническое доказательство	
				Tamoxifen	В: Клиническое доказательство	
3		3:g.178936082G>A (p.Glu542Lys)	соматический	Alpelisib + Fulvestrant	А: Утвержденная рекомендация	
				Capivasertib+Fulvestrant	А: Утвержденная рекомендация	
				Everolimus	В: Клиническое доказательство	
				Tamoxifen	В: Клиническое доказательство	
3		3:g.178936091G>A (p.Glu545Lys)	соматический	Alpelisib + Fulvestrant	А: Утвержденная рекомендация	
				Capivasertib+Fulvestrant	А: Утвержденная рекомендация	
				Everolimus	В: Клиническое доказательство	
				Tamoxifen	В: Клиническое доказательство	
1		3:g.178936094C>A (p.Gln546Lys)	соматический	Alpelisib, Fulvestrant	А: Утвержденная рекомендация	
				Capivasertib+Fulvestrant	А: Утвержденная рекомендация	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Everolimus	В: Клиническое доказательство	
				Tamoxifen	В: Клиническое доказательство	
5		3:g.178952085A>G (p.His1047Arg)	соматический	Alpelisib + Fulvestrant	А: Утвержденная рекомендация	
				Capivasertib+Fulvestrant	А: Утвержденная рекомендация	
				Everolimus	В: Клиническое доказательство	
				Tamoxifen	В: Клиническое доказательство	
1		3:g.178951973A>G (p.Met1010Val)	соматический	Alpelisib + Fulvestrant	В: Клиническое доказательство	
				Capivasertib+Fulvestrant	В: Клиническое доказательство	
				Everolimus	В: Клиническое доказательство	
				Tamoxifen	В: Клиническое доказательство	
1		3:g.178952029A>C(p.Leu1028Phe)	соматический	Alpelisib + Fulvestrant	В: Клиническое доказательство	
				Capivasertib+Fulvestrant	В: Клиническое доказательство	
				Everolimus	В: Клиническое доказательство	
				Tamoxifen	В: Клиническое доказательство	
1		3:g.178928220A>G (p.Glu469Gly)	соматический	Alpelisib + Fulvestrant	В: Клиническое доказательство	
				Capivasertib+Fulvestrant	В: Клиническое доказательство	
				Everolimus	В: Клиническое доказательство	
				Tamoxifen	В: Клиническое доказательство	
1	PIK3CB	3:g.138374245C>A (p.Asp1067Tyr)	соматический	Ингибиторы АКТ	Д: Доклиническое доказательство	
				Ингибиторы MTORC1/2	Д: Доклиническое доказательство	
1	PTCH1	9:g.98244230C>T ()	соматический	Sonidegib	В: Клиническое доказательство	
				Vismodegib	В: Клиническое доказательство	
1		9:g.98242356AG>A (p.Val322PhefsTer2)	соматический	Sonidegib	В: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Vismodegib	В: Клиническое доказательство	
1	PTEN	10:g.89685269G>A ()	соматический	Everolimus	В: Клиническое доказательство	
				Capivasertib+Fulvestrant	А: Утвержденная рекомендация	
				AZD8186	Д: Доклиническое доказательство	
				GSK2636771	Д: Доклиническое доказательство	
1		10:g.89690802G>T ()	соматический	Everolimus	В: Клиническое доказательство	
				Capivasertib+Fulvestrant	А: Утвержденная рекомендация	
				AZD8186	Д: Доклиническое доказательство	
				GSK2636771	Д: Доклиническое доказательство	
1		10:g.89692792C>G (p.Asp92Glu)	соматический	Everolimus	В: Клиническое доказательство	
				Capivasertib+Fulvestrant	А: Утвержденная рекомендация	
				AZD8186	Д: Доклиническое доказательство	
				GSK2636771	Д: Доклиническое доказательство	
1		10:g.89692793C>T (p.His93Tyr)	соматический	Everolimus	В: Клиническое доказательство	
				Capivasertib+Fulvestrant	А: Утвержденная рекомендация	
				AZD8186	Д: Доклиническое доказательство	
				GSK2636771	Д: Доклиническое доказательство	
1		10:g.89692922T>C (p.Cys136Arg)	соматический	Everolimus	В: Клиническое доказательство	
				Capivasertib+Fulvestrant	А: Утвержденная рекомендация	
				AZD8186	Д: Доклиническое доказательство	
				GSK2636771	Д: Доклиническое доказательство	
1		10:g.89692978CTATGGG GAAG>C (p.Tyr155Ter)	соматический	Everolimus	В: Клиническое доказательство	
				Capivasertib+Fulvestrant	А: Утвержденная рекомендация	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				AZD8186	D: Доклиническое доказательство	
				GSK2636771	D: Доклиническое доказательство	
1		10:g.89717609G>C ()	соматический	Everolimus	B: Клиническое доказательство	
				Capivasertib+Fulvestrant	A: Утвержденная рекомендация	
				AZD8186	D: Доклиническое доказательство	
				GSK2636771	D: Доклиническое доказательство	
1		10:g.89717672C>T (p.Arg233Ter)	соматический	Everolimus	B: Клиническое доказательство	
				Capivasertib+Fulvestrant	A: Утвержденная рекомендация	
				AZD8186	D: Доклиническое доказательство	
				GSK2636771	D: Доклиническое доказательство	
1		10:g.89720693GGACCAGAGGAAACCTCAGAAAAAGTA>G (p.Pro283LysfsTer6)	соматический	Everolimus	B: Клиническое доказательство	
				Capivasertib+Fulvestrant	A: Утвержденная рекомендация	
				AZD8186	D: Доклиническое доказательство	
				GSK2636771	D: Доклиническое доказательство	
1		10:g.89720764T>TATAG (p.Glu307AspfsTer6)	соматический	Everolimus	B: Клиническое доказательство	
				Capivasertib+Fulvestrant	A: Утвержденная рекомендация	
				AZD8186	D: Доклиническое доказательство	
				GSK2636771	D: Доклиническое доказательство	
1	RB1	13:g.48916768G>T (p.Gly100Ter)	соматический	Химиотерапия	B: Клиническое доказательство	он-лейбл
				Ингибиторы HDAC	D: Доклиническое доказательство	
				Ингибиторы MDM2/MDMX	D: Доклиническое доказательство	
1		13:g.48941648C>T (p.Arg320Ter)	соматический	Химиотерапия	B: Клиническое доказательство	он-лейбл
				Ингибиторы HDAC	D: Доклиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Ингибиторы MDM2/MDMX	D: Доклиническое доказательство	
1		13:g.49039399C>G (p.Ser795Ter)	соматический	Химиотерапия	B: Клиническое доказательство	он-лейбл
				Ингибиторы HDAC	D: Доклиническое доказательство	
				Ингибиторы MDM2/MDMX	D: Доклиническое доказательство	
1		13:g.48937095T>C	соматический	Ингибиторы HDAC	D: Доклиническое доказательство	
				Ингибиторы MDM2/MDMX	D: Доклиническое доказательство	
1		13:g.49027200TTGTCCTC TTAATCTTCCTCTCCAG AATAATCACACTGCAG CAGATA>T	соматический	Ингибиторы HDAC	D: Доклиническое доказательство	
				Ингибиторы MDM2/MDMX	D: Доклиническое доказательство	
1	SF3B1	2:g.198266834T>C (p.Lys700Glu)	соматический	Spliceostatin A	D: Доклиническое доказательство	
				Emavusertib	D: Доклиническое доказательство	
				Ceralasertib	D: Доклиническое доказательство	
1	SMARCA4	19:g.11129654C>A (p.Tyr820Ter)	соматический	PRT3789	D: Доклиническое доказательство	
1	SMARCB1	22:g.24143240C>T (p.Arg158Ter)	соматический	Tazemetostat	D: Доклиническое доказательство	
1	STK11	19:g.1207078G>A (p.Gly56Arg)	соматический	Вemcentinib+Pembrolizumab	D: Доклиническое доказательство	
1		19:g.1207155G>C (p.Lys81Asn)	соматический	Вemcentinib+Pembrolizumab	D: Доклиническое доказательство	
1	TP53	17:g.7573990T>TCATTCA G (p.Glu346AlafsTer3)	соматический	Адьювантная химиотерапия	B: Клиническое доказательство	он-лейбл
				Alemtuzumab	B: Клиническое доказательство	
				Химиотерапия	B: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	B: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
2		17:g.7574003G>A (p.Arg342Ter)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7574006ACATCT>A (p.Glu339ValfsTer6)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7574013G>GAAGC (p.Phe338LeufsTer10)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577022G>A (p.Arg306Ter)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577064T>A (p.Lys292Ter)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577094G>C (p.Arg282Gly)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
2		17:g.7577105G>C (p.Pro278Arg)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577114C>T (p.Cys275Tyr)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577120C>G (p.Arg273Pro)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577120CG>C (p.Arg273ValfsTer72)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
1		17:g.7577124C>T (p.Val272Met)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577137C>CCGT (p.Arg267dup)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577509C>A (p.Glu258Ter)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577509C>T (p.Glu258Lys)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	B: Клиническое доказательство	
				Pazopanib	B: Клиническое доказательство	
				Pazopanib,Vorinostat	B: Клиническое доказательство	
1		17:g.7577511A>G (p.Leu257Pro)	соматический	Адьювантная химиотерапия	B: Клиническое доказательство	он-лейбл
				Alemtuzumab	B: Клиническое доказательство	
				Химиотерапия	B: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	B: Клиническое доказательство	
				Pazopanib	B: Клиническое доказательство	
				Pazopanib,Vorinostat	B: Клиническое доказательство	
1		17:g.7577526AG>A (p.Leu252SerfsTer93)	соматический	Адьювантная химиотерапия	B: Клиническое доказательство	он-лейбл
				Alemtuzumab	B: Клиническое доказательство	
				Химиотерапия	B: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	B: Клиническое доказательство	
				Pazopanib	B: Клиническое доказательство	
				Pazopanib,Vorinostat	B: Клиническое доказательство	
1		17:g.7577579G>C (p.Tyr234Ter)	соматический	Адьювантная химиотерапия	B: Клиническое доказательство	он-лейбл
				Alemtuzumab	B: Клиническое доказательство	
				Химиотерапия	B: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	B: Клиническое доказательство	
				Pazopanib	B: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577579G>T (p.Tyr234Ter)	соматический	Adjuvant Chemotherapy	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577580T>C (p.Tyr234Cys)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577593TAC>T (p.Cys229TyrfsTer10)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7577595C>CG (p.Cys229SerfsTer11)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578191A>G (p.Tyr220His)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
3		17:g.7578212G>A (p.Arg213Ter)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578266T>A (p.Ple195Phe)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578271T>C (p.His193Arg)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578382G>C (p.Ser183Ter)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578395G>A (p.His179Tyr)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
1		17:g.7578397TG>T (p.His178ThrfsTer69)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578403C>A (p.Cys176Phe)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578406C>G (p.Arg175Pro)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578406CG>C (p.Arg175AlafsTer72)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578413C>G (p.Val173Leu)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578413C>T (p.Val173Met)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578431G>A (p.Gln167Ter)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578441G>GCA (p.Lys164AlafsTer7)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	Д: Доклиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578442T>C (p.Tyr163Cys)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	Д: Доклиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578442T>G (p.Tyr163Ser)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	Д: Доклиническое доказательство	он-лейбл
				Eprenetapopt, Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
1		17:g.7578526C>T (p.Cys135Tyr)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578527A>C (p.Cys135Gly)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7578534C>G (p.Lys132Asn)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprentapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib, Vorinostat	В: Клиническое доказательство	
1		17:g.7579329T>C (p.Lys120Glu)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	B: Клиническое доказательство	
				Pazopanib	B: Клиническое доказательство	
				Pazopanib, Vorinostat	B: Клиническое доказательство	
1		17:g.7579358C>G (p.Arg110Pro)	соматический	Адьювантная химиотерапия	B: Клиническое доказательство	он-лейбл
				Alemtuzumab	B: Клиническое доказательство	
				Химиотерапия	B: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	B: Клиническое доказательство	
				Pazopanib	B: Клиническое доказательство	
				Pazopanib, Vorinostat	B: Клиническое доказательство	
1		17:g.7579358CG>C (p.Arg110ValfsTer13)	соматический	Адьювантная химиотерапия	B: Клиническое доказательство	он-лейбл
				Alemtuzumab	B: Клиническое доказательство	
				Химиотерапия	B: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	B: Клиническое доказательство	
				Pazopanib	B: Клиническое доказательство	
				Pazopanib, Vorinostat	B: Клиническое доказательство	
1		17:g.7579360GAAACC>G (p.Gly108ProfsTer39)	соматический	Адьювантная химиотерапия	B: Клиническое доказательство	он-лейбл
				Alemtuzumab	B: Клиническое доказательство	
				Химиотерапия	B: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	B: Клиническое доказательство	
				Pazopanib	B: Клиническое доказательство	
				Pazopanib, Vorinostat	B: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
1		17:g.7579363ACCGTAGC TGCCCTGGTAGGTTTTC TGGGAAGGGACAGAAG ATGACAGGGGCCAGGA GGGGGCTGGTGCAGGG GC>A (p.Ala84PhefsTer40)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib,Vorinostat	В: Клиническое доказательство	
1		17:g.7579366G>T (p.Tyr107Ter)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib,Vorinostat	В: Клиническое доказательство	
1		17:g.7579389GGGAA>G (p.Ser99ArgfsTer23)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib,Vorinostat	В: Клиническое доказательство	

Кол-во больных	Ген	Выявленный геномный вариант и аминокислотная замена	Тип варианта	Рекомендация персонализированного лекарственного лечения	Степень доказательности	Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы*
1		17:g.7579400GA>G (p.Ser96LeufsTer27)	соматический	Адьювантная химиотерапия	В: Клиническое доказательство	он-лейбл
				Alemtuzumab	В: Клиническое доказательство	
				Химиотерапия	В: Клиническое доказательство	он-лейбл
				Doxorubicin	D: Доклиническое доказательство	он-лейбл
				Eprenetapopt,Azacitidine	В: Клиническое доказательство	
				Pazopanib	В: Клиническое доказательство	
				Pazopanib,Vorinostat	В: Клиническое доказательство	
1	TSC1	9:g.135796805G>A (p.Arg228Ter)	соматический	Everolimus	В: Клиническое доказательство	
Кол-во больных	Опухолевая мутационная нагрузка (ТМВ)			Рекомендация персонализированного лекарственного лечения	Степень доказательности	
19	Высокий уровень ТМВ (≥ 10 мутаций/Мб)			Pembrolizumab	A: Утвержденная рекомендация	он-лейбл

Общее количество больных ТНРМЖ, для которых были сформированы рекомендации персонализированного лекарственного лечения - 83 (83%). У 44 больных ТНРМЖ было выявлено несколько биомаркеров (от 2 до 7), значимых для подбора рекомендаций персонализированного лечения.

*В столбце «Статус регистрации препарата в РФ/США/ЕС при трижды негативном раке молочной железы» приведены данные о статусе регистрации препарата для терапии данного ЗНО независимо от наличия или отсутствия описываемого биомаркера.

Наибольшее количество рекомендаций персонализированного лечения - 8 различных препаратов: 1 - Neratinib, 2 - Ado-Trastuzumab Emtansine, 3 - Lapatinib, 4 - Margetuximab, 5 - Trastuzumab, 6 - Pertuzumab, 7 - Tucatinib, 8 - Trastuzumab Deruxtecan было сформировано СППВР для больной ТНРМЖ с соматическим геномным вариантом 17:g.37880220T>C (p.Leu755Ser) в гене *ERBB2*. При этом сформированные рекомендации имели разные степени доказательности.

Среди лекарственных препаратов, рекомендуемых СППВР, присутствуют как зарегистрированные/одобренные для применения, так и исследуемые препараты. Информация о препаратах, прошедших доклинические и фазы I/II клинических исследований, важна для формирования полного представления о диапазоне перспективных средств.

Созданная в настоящем исследовании СППВР была опробована на образцах РПЖ и ТНРМЖ. Вместе с тем, данная СППВР может применяться для подбора персонализированного лекарственного лечения и при других типах ЗНО, поскольку алгоритмы в её составе выявляют препараты, к которым наблюдается чувствительность, на основе молекулярно-генетических биомаркеров вне зависимости от типа ЗНО [317].

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

Масштабные международные проекты по исследованию молекулярных особенностей опухолей, а также снижение стоимости высокопроизводительного секвенирования способствуют лучшему пониманию механизмов развития онкологических заболеваний, а также делают возможным повысить эффективность лечения рака. Однако сложность интерпретации геномных данных создает препятствия для более активного внедрения высокопроизводительных технологий в клиническую практику. Одной из основных задач в решении этой проблемы является оценка клинической значимости найденных геномных вариантов и их влияния на выбор предпочтительной терапии для конкретного пациента.

Одним из основополагающих шагов по этому пути, реализованных в настоящей работе, стало проведение мета-анализа, в рамках которого были исследованы клинические исходы лечения ЗНО у больных, проходивших генетический анализ для подбора персонализированного лечения, либо не проходивших анализ и получавших неперсонализированное лечение. Результаты были выведены с помощью пропорционального мета-анализа и мета-регрессии, и подтверждены непараметрической статистикой. Показатели ЧОО, медианные и однолетние ВБП и ОВ были значительно выше в исследованиях, где больные были отобраны на основе генетического тестирования. Результаты проведенного мета-анализа свидетельствуют о том, что назначение лечения с учетом генетических особенностей онкологических больных способно улучшить результаты любого вида лекарственного лечения: цитотоксической химиотерапии, таргетной и иммунотерапии. При этом дизайн исследования, импакт-фактор журнала и количество больных в группе не повлияли на исходы лечения. Полученные результаты могут указывать на однородность исследований, но, более вероятно, можно предположить высокую степень связи между персонализацией лечения на основе геномики и лучшими исходами.

Медианный мета-анализ без дополнительных данных менее точен и эффективен по сравнению с пропорциональным мета-анализом, поэтому мы сравнивали медианы только с помощью критерия Манна-Уитни. Тем не менее, мы выявили значительные отличия между персонализированными и неперсонализированными исследованиями, что может указывать на большую важность генетического тестирования для выбора эффективной стратегии лечения.

Schwaederle и соавторы (2015) провели аналогичный мета-анализ эффективности персонализированного подхода к лекарственному лечению онкологических заболеваний [308]. Авторы использовали данные 2010-2012 годов и продемонстрировали значительно лучшие показатели ЧОО, ВБП и ОВ в исследованиях, в которых лечение было персонализировано. Schwaederle и соавторы рассматривали только работы, в которых применялся один

лекарственный препарат, а персонализацию рассматривали с общей точки зрения. Применяемые секвенирование нового поколения, другие тесты на биомаркеры и комплексный персональный подход к отбору пациентов были оценены, чтобы отличить исследования с персонализацией и без персонализации [308]. Мы провели более ограниченный поиск с уточненными терминами в базе данных PubMed, но включили группы, получавшие терапию одновременно несколькими препаратами и иммунотерапию. В проведенном нами мета-анализе были получены аналогичные результаты, однако в качестве персонализации рассматривалось только генетическое тестирование, а в анализ были включены данные, полученные в 2012-2021 гг.; следовательно, полученные результаты подтверждают данные предыдущих публикаций [308].

Таргетная и иммунотерапия разрабатываются для конкретного фенотипа опухоли, связанного с определенными мутациями [119, 353, 393]. Назначение таких типов лечения без предварительного генетического обследования может приводить к худшим исходам [353]. Цитотоксическая химиотерапия является системным противоопухолевым лечением, которое в большинстве случаев назначается без учета генетических особенностей опухоли. Тем не менее, результаты опубликованных исследований свидетельствуют о том, что выявление определенных типов генетических вариантов может быть ассоциировано с чувствительностью или резистентностью к различным типам химиопрепаратов. Так, в публикации Wu и соавторов было показано, что клетки рака мочевого пузыря с мутацией в гене TP53 проявляют значительную чувствительность к митомицину - C, доксорубину и гемцитабину [379]. Вместе с тем, Stires и соавторы продемонстрировали, что наличие мутаций в гене *GRM1* является предиктором резистентности к доксициклину при РМЖ за счет изменения активности MAP-киназного сигнального пути [27, 334].

Согласно результатам мета-анализа, проведенного в рамках настоящей работы, все виды медикаментозного лечения рака могут привести к лучшим исходам при назначении с учетом генетического профилирования. Следует отметить, что в некоторых случаях клинические рекомендации не содержат возможности изменить стратегию лечения с учетом мутаций. В то же время можно предположить, что целесообразно проводить генетическое профилирование как часть диагностики в онкологической практике во всех случаях, где это возможно. Данное предложение было описано во множестве предыдущих обзоров и оригинальных статей. Кроме того, обсуждается создание фармакогеномных баз данных, которые могут помочь в разработке персонализированного подхода к лекарственному лечению рака, улучшающего исходы лечения [294].

В настоящее время секвенирование нового поколения и другие технологии анализа генетического материала включаются в стандарты диагностики онкологических заболеваний [108, 294]. Это особенно важно в контексте того, что неперсонализированная таргетная или

иммуноterapia, в том числе проводимая в сочетании с цитотоксической химиотерапией, приводит к худшим результатам. Вместе с тем необходимо отметить, что доля исследований неперсонализированной таргетной терапии снизилась за последнее десятилетие с 95% до 34,2% [308].

Несмотря на наличие клинических исследований, подтверждающих улучшение исходов при применении персонализированного подхода к лечению онкологических заболеваний, существуют различные точки зрения относительно целесообразности широкого внедрения персонализации лечения, требующего дополнительных затрат на проведение генетических и иных исследований. Критика персонализированного подхода в основном связана с потенциально более высокой финансовой нагрузкой на системы здравоохранения [305]. Вместе с тем, опубликован ряд авторитетных экономических исследований, говорящих о том, что персонализированная терапия онкологических заболеваний способна привести к благоприятным экономическим исходам за счет оптимизации схем лечения [201].

В мета-анализ, осуществленный в рамках настоящей работы, было включено 49 клинических исследований, включающих 101 группу пациентов. Ограничением проведенного мета-анализа является включение только клинических исследований II фазы. Проведение дальнейших мета-анализов, включающих испытания III фазы, а также пострегистрационные протоколы, способно далее расширить понимание преимуществ персонализированного лечения онкологических заболеваний.

На основе результатов, полученных в реализованном мета-анализе, были сделаны выводы о перспективности внедрения персонализированного подхода в онкологическую практику в Российской Федерации. Положительный опыт персонализации лечения, полученный в рамках различных клинических исследований, а также результаты проведенного мета-анализа подтвердили тезис об актуальности разработки фармакогеномной платформы и СППВР для подбора персонализированного лекарственного лечения онкологических больных.

Согласно литературным данным, использование СППВР в клинической практике позволяет улучшить результаты лечения в сравнении со стандартными протоколами. Особенностью онкологических заболеваний является их многообразие и гетерогенность, при этом их профилирование и характеристика связаны с обработкой “больших данных”, что требует использования не узкоспециализированных, а платформенных решений. В связи с этим проблемы разработки и внедрения фармакогеномных платформ для поддержки принятия персонализированных врачебных решений в медицине, особенно в онкологии, являются актуальной прикладной научной тематикой.

В рамках реализации настоящей работы была создана оригинальная отечественная фармакогеномная платформа, структурно состоящая из трёх элементов: озера данных,

биоинформатических конвейеров, функционирующих с применением технологий машинного обучения, и блока управления и анализа. Озеро данных в составе фармакогеномной платформы позволяет аккумулировать большие объёмы собственных данных полногеномного, полноэкзомного и панельного секвенирования биологического материала больных, а также массив клинико-фенотипической и иной информации. Озеро данных также включает большие объёмы информации из открытых международных баз, позволяющей осуществлять аннотирование найденных геномных вариантов, а также способствовать подбору эффективного персонализированного таргетного, иммунотерапевтического и химиотерапевтического лечения СППВР с учетом биомаркеров больных. Чертой разработанной фармакогеномной платформы является то, что интеграция информации из различных баз данных позволяет осуществлять подбор персонализированной терапии в отсутствие подключения к сети Интернет.

В современной онкологической практике применяется широкий спектр лечебных стратегий, вместе с тем, фармакогеномный подход является одним из наиболее перспективных [10–12, 15]. Разработка фармакогеномных платформ, объединяющих большие массивы информации и служащих фундаментом для анализа, осуществляется многими престижными медицинскими и научными учреждениями [43]. Фармакогеномный портал SYNERGxDB агрегирует информацию о синергическом эффекте различных противоопухолевых препаратов [310]. SYNERGxDB в настоящее время аккумулирует результаты различных исследований в одной из крупнейших в данной области баз данных с включением девяти коллекций фармакологических и молекулярных профилей доклинических модельных систем [310]. Фармакологические коллекции портала охватывают 477 839 комбинаций препаратов, протестированных на 151 клеточной линии, что позволяет эффективно исследовать синергизм лекарственных препаратов в более чем 500 тысячах комбинаций. SYNERGxDB также включает в себя данные о метаболомике, экспрессии генов и мутационные профили клеточных линий. Каждый препарат и клеточная линия имеют уникальный идентификатор [310]. Также необходимо отметить фармакогеномные платформы, агрегирующие различные типы омиксных данных. Так, Integrated Pharmacogenomic Platform of Cancer Cell Lines and Tissues (Интегративная фармакогеномная платформа раковых клеточных линий и тканей) представляет собой фармакогеномную платформу, позволяющую аккумулировать информацию и находить генетические особенности клеточных линий, чувствительных к тем или иным противоопухолевым препаратам [320]. Необходимо отметить, что данная платформа агрегирует в своем составе общеизвестные открытые базы данных. Такой же подход по интеграции открытых баз данных был использован при разработке фармакогеномной платформы в настоящем исследовании.

Вместе с тем, важно отметить следующие особенности разработанной в настоящем исследовании фармакогеномной платформы. Упомянутые выше платформы SYNERGxDB и Интегративная фармакогеномная платформа раковых клеточных линий и тканей аккумулируют информацию о влиянии противоопухолевых препаратов на клеточные линии. Вместе с тем, в современной онкологии высокой значимостью обладают первичные данные, полученные на высокопроизводительных аналитических машинах, ассоциированные с оказанием медицинской помощи больным (англ. real world data). В данном контексте необходимо отметить, что разработанная в настоящем исследовании фармакогеномная платформа ориентирована на интеграцию широкого перечня данных о больных ЗНО. Также одним из важных отличий разработанной фармакогеномной платформы является возможность аккумулирования информации о больных из российской популяции. Важность агрегирования больших массивов данных о больных из определенных популяций обоснована возможностью выявления геномных, гистологических и иных характеристик ЗНО.

Одним из отличий фармакогеномной платформы, разработанной в настоящем исследовании, является применение инструментов на основе технологий машинного обучения в биоинформатических конвейерах, функционирующих в её составе. Описанная в работе Yu и соавторов фармакогеномная платформа, использующая клетки, полученные от больных, для выявления основных механизмов резистентности у пациентов с распространенным или рефрактерным раком легкого и подбора персонализированного лечения, не включает алгоритмы машинного обучения при выявлении соматических геномных вариантов [390].

Важно отметить, что в отличие от многих зарубежных аналогов, разработанная в настоящем исследовании фармакогеномная платформа позволяет аккумулировать и хранить не только геномные данные больных и сведения об ассоциациях «биомаркер - лекарственный препарат», но и массив клинико-фенотипической информации. Аккумулирование не только омиксных, но и клинических данных является важным этапом для дальнейшего анализа и поиска факторов, способных определять ответ на проводимое противоопухолевое лечение. В эпоху развития технологий искусственного интеллекта появляются все более совершенные инструменты анализа больших массивов данных. Справедливо предположить, что при условии накопления больших массивов различных данных, включая геномные и клинико-фенотипические, в перспективе станет возможным выявление новых взаимосвязей между теми или иными параметрами и эффективностью различных лекарственных противоопухолевых средств. Гибкость разработанной фармакогеномной платформы сможет позволить интегрировать в ее состав также дополнительные аналитические инструменты, которые будут созданы в будущем, для установления новых взаимосвязей между накопленными геномными и клинико-

фенотипическими данными и эффективностью определённых противоопухолевых лекарственных препаратов.

Одним из ключевых элементов разработанной фармакогеномной платформы, обеспечивающим аккумуляцию всего массива релевантной информации, включая информацию из открытых мировых баз данных, является разработанное озеро данных. Современные подходы к лечению онкологических заболеваний подразумевают широкое использование молекулярно-генетических методов для идентификации геномных вариантов, значимых для подбора персонализированного лечения. Вместе с тем, интеграция методов геномной медицины неразрывно связана с задачей хранения больших объёмов данных, получаемых по итогам проводимого секвенирования, а также по итогам биоинформатической обработки данных секвенирования. При разработке научной фармакогеномной платформы было создано озеро данных на основе инструмента управления озером данных Dataiku. Инструменты управления озерами данных используются в медицинских исследованиях, аккумулирующих большие объёмы данных. Так, Dataiku использовался для выявления пациентов с высоким риском иммунодефицита [270], а также в масштабном исследовании географической доступности медицинских услуг для онкологических пациентов [125]. В последнем случае авторы использовали данный инструмент для извлечения и загрузки данных в дополнение к языкам R и Python. В исследовании с использованием нейронных сетей и машинного обучения для улучшения диагностики рака кожи применялся инструмент управления озером данных KNIME, точность модели, при этом, составила более 77% [245].

Изучение молекулярно-генетических характеристик ЗНО обладает высокой значимостью, поскольку позволяет углубить понимание особенностей различных типов рака и продвигаться в изыскании новых лечебных стратегий [8, 192, 391]. На настоящий момент, в сферах фундаментальной и клинической онкологической науки накоплены большие объёмы данных, в том числе о различных генетических характеристиках опухолей, ответах на различные типы лечения и т.д. Накопление указанных крайне обширных объёмов данных обусловило создание и функционирование различных хранилищ, аккумулирующих информацию и обеспечивающую возможность работы с ней. В мировой практике можно выделить устоявшиеся и широко применяемые базы данных, такие как Clinvar, Cosmic, GnomAd, TCGA и т.д. [80, 149, 329, 363]. Вместе с тем необходимо отметить постоянную разработку новых баз данных различными организациями и учреждениями. Создаваемые базы данных могут быть посвящены отдельным аспектам изучения онкологических заболеваний, отдельным аспектам лечения, а также отдельным нозологиям. В созданное в настоящем исследовании озеро данных была осуществлена интеграция информации из различных мировых баз данных, что позволило нам сформировать собственную объёмную базу данных для аннотирования геномных вариантов и

подбора эффективного персонализированного лекарственного лечения онкологических заболеваний. Одной из важных характеристик, отличающих разработанную в настоящем исследовании платформу, является настройка механизмов автоматизированной актуализации озера данных, что подразумевает возможность добавления новых данных высокопроизводительного секвенирования и интеграцию новой информации из открытых мировых баз данных. Названный подход обеспечивает соответствие содержащейся в озере данных информации последним научным сведениям.

Биоинформатические конвейеры для обработки данных высокопроизводительного секвенирования также являются структурным элементом созданной фармакогеномной платформы. Разработка собственных биоинформатических конвейеров была основана на анализе уже существующих. Были сформулированы требования, предъявляемые к биоинформатическим конвейерам, и по итогам анализа было заключено, что ни один из существующих конвейеров не отвечал всем требованиям. С учётом недостатков проанализированных аналогов была осуществлена сборка собственного биоинформатического конвейера, отвечающего всем сформулированным требованиям. Одной из ключевых черт разработанных биоинформатических конвейеров стала интеграция инструментов на основе технологий машинного обучения на этапе выявления геномных вариантов в исследуемых образцах. В том числе была использована программа по выявлению герминальных геномных вариантов DeepVariant, основанная на применении технологий искусственного интеллекта, в частности - машинного обучения [282]. Для выявления соматических вариантов в разработанный биоинформатический конвейер были интегрированы программы Rfcaller и Varnet, также функционирующие с применением технологий машинного обучения.

Методы машинного обучения находят широкое применение в различных областях фундаментальной медицинской науки и в клинической практике [132]. С учётом эффективности DeepVariant при выявлении геномных вариантов, продемонстрированной в опубликованных работах, было принято решение об интеграции данной программы на основе машинного обучения в разработанный биоинформатический конвейер [17]. DeepVariant функционирует на основе сверточной нейронной сети, эффективно реализующей функции по выявлению генетических вариантов по итогам проведённого секвенирования. Разработанные в настоящем исследовании биоинформатические конвейеры объединяют современные инструменты и методологии для обработки данных секвенирования полного экзона, идентификации вариантов как в герминальных, так в моделях опухоли-нормы образцах, и интерпретации их потенциальной биологической значимости в контексте исследования. Создание собственных алгоритмов машинного обучения не входила в задачи настоящего исследования; был сделан выбор в пользу интеграции готовых инструментов, использующих такие технологии. Однако разработка

собственных моделей машинного обучения является перспективным направлением для будущих исследований.

Мировая практика показывает, что применение генетического тестирования для персонализации лечения онкологических заболеваний может быть реализовано различными способами. В настоящем исследовании был выбран подход, в рамках которого осуществлялось полноэкзомное секвенирование. Вместе с тем, вопрос выбора наиболее оптимального объема секвенирования для реализации масштабных фундаментальных и прикладных проектов по персонализации лечения онкологических заболеваний остается дискуссионным. Так, одним из наиболее известных подходов является применение СППВР FoundationOne®CDx (Foundation Medicine, Inc., Cambridge, MA, USA), одобренной FDA. Платформа FoundationOne®CDx (Foundation Medicine, Inc., Cambridge, MA, USA) представляет собой панельный тест, в рамках которого исследуются различные гены, значимые для подбора персонализированной терапии. Несомненной значимостью обладают полногеномные исследования, позволяющие выявлять геномные варианты во всех кодирующих и некодирующих последовательностях. Проведение полногеномного секвенирования является наиболее дорогостоящей опцией, в то время как полноэкзомное секвенирование и панельные тесты позволяют снизить затраты.

Вопрос о предпочтительности панельного либо полноэкзомного генетического тестирования является сложным [7, 285]. Среди плюсов панельных генетических тестов выделяется более низкая стоимость исследования [129, 284]. Однако в контексте сравнительного анализа стоимости различных подходов к секвенированию, важно отметить, что на настоящий момент разработаны подходы, позволяющие существенно сократить расходы на пробоподготовку и тем самым приблизить стоимость полноэкзомного секвенирования к стоимости панельных тестов [273]. Кроме того, полноэкзомное исследование биологического материала онкологических больных позволяет выявлять геномные варианты во всех кодирующих регионах и, соответственно, может способствовать подбору более широкого спектра препаратов для персонализированного лечения онкологических заболеваний [117]. В настоящей работе по итогам проведенного полноэкзомного секвенирования биологического материала 30 больных ТНРМЖ были выявлены варианты в генах *KMT2C*, *CDKN2A*, *FANCI*, *NOTCH1*, *STK11*, *GNA11*, *SLX4*, *EZH2*, *RAD52*, *NOTCH2*, *PIK3CB*, *WT1*, *SETBP1*, исследование которых не входит в панель AVENIO (Roche, США). Также был выявлен вариант в гене *TSC1*, не все регионы которого исследуются в названной панели.

Одним из существенных преимуществ полноэкзомного секвенирования также является возможность выявлять особенности геномного ландшафта у онкологических больных, принадлежащих к различным популяциям. Данный подход может способствовать не только выявлению новых геномных вариантов, но и формированию общей картины геномного профиля

той или иной популяции. Важность выявления ландшафта геномных вариантов в пределах популяции обосновывается тем, что полученные данные о превалировании тех или иных геномных вариантов у онкологических больных могут лечь в основу рекомендаций по целесообразности генетического исследования того или иного гена.

В настоящей работе нами было принято решение о проведении полноэкзомного секвенирования и об интеграции большого объема полученных геномных данных в разработанную научную фармакогеномную платформу. Такое решение было продиктовано целесообразностью выполнения не только задачи по подбору персонализированного лекарственного лечения ЗНО, но и фундаментальной ценностью получения и публикации полноэкзомных данных больных РПЖ и ТНРМЖ из российской популяции.

Полноэкзомное секвенирование биологического материала больных РПЖ и ТНРМЖ из российской популяции было реализовано впервые в российской научной практике и помогло выявить ландшафт геномных вариантов при указанных онкологических заболеваниях. Был сформирован ландшафт геномных вариантов для 40 больных РПЖ и подсчитано количество геномных вариантов в различных генах. Кроме того, мы построили ландшафт геномных вариантов в виде тепловой карты (хитмэп), и визуализировали гены с наибольшим количеством геномных вариантов, в число которых вошли гены *APC*, *BRCA2*, *MUC16*, *MUC19*. Полученные данные соответствуют результатам зарубежных исследований о распространенности данных геномных вариантов у больных РПЖ, что подтверждает важность скрининга этих генов и открывает возможности для создания таргетной терапии. Так, достоверная связь геномных вариантов в генах *BRCA2* и *PALB2* с высоким риском РПЖ была подтверждена в исследовании с участием более чем 1600 пациентов [173]. Результаты NGS анализа российской популяции больных РПЖ отличаются от других популяций в части высокой мутационной нагрузки в указанных генах [314]. Необходимо также отметить, что выявленные нами геномные варианты в генах *MUC16* и *TTN* часто наблюдаются у пациентов с различными злокачественными заболеваниями, но далеко не всегда ассоциированы с клинически значимыми изменениями, что позволяет считать такие геномные варианты т.н. “пассажирскими” мутациями, не оказывающими существенного влияния на процессы канцерогенеза. Так, например, *TTN* является крупнейшим геном, что увеличивает вероятность случайных вариантов [259]. Перспективным направлением является дальнейший анализ различий между геномными профилями, описанными в литературе, и нашими результатами на более объемных выборках больных.

Проведенное высокопроизводительное секвенирование биологического материала 100 больных ТНРМЖ и биоинформатическая обработка данных разработанными биоинформатическими конвейерами позволили также выявить ландшафт геномных вариантов больных данным ЗНО в российской популяции. Результаты высокопроизводительного

секвенирования больных РМЖ из российской популяции уже были представлены в некоторых исследованиях, однако публикация дополнительных результатов NGS исследований имеет большое значение для увеличения объемов информации. Solodskikh и соавторы проанализировали группу больных РМЖ из российской популяции и показали, что варианты rs1799967 и rs4987117 в генах *BRCA1* и *BRCA2*, соответственно, связаны с риском РМЖ [326]. Важно, что *BRCA1* и *BRCA2* относятся к генам, выполняющим ключевые регуляторные функции в клетке. Так, *BRCA2* принято относить к опухолевым супрессорам, участвующим в процессе репарации двухцепочечных разрывов ДНК. Примечательно, что среди выявленных геномных вариантов в когорте больных ТНРМЖ настоящего исследования мы не выявили упомянутые варианты rs1799967 и rs4987117 [7, 165]. Различие в результатах, полученных в нашем исследовании и в предыдущих работах, подчеркивает необходимость дальнейших исследований с применением полноэкзомного секвенирования для более глубокого изучения ландшафта геномных вариантов в российской популяции больных ТНРМЖ. Исследование мутаций *BRCA* имеет доказанное значение для подбора противоопухолевой терапии: опухоли, имеющие данные мутации, более чувствительными к ингибиторам PARP [141, 224, 293].

В настоящем исследовании была разработана СППВР для подбора эффективного персонализированного лекарственного лечения ЗНО, функционирующая с использованием больших объемов данных, интегрированных в созданную фармакогеномную платформу. Компьютеризированные СППВР являются перспективным инструментом в работе практикующего клинициста в условиях непрерывного роста информации, подлежащей учету при подборе эффективного лечения [316]. Ji и соавторы указали на наличие СППВР, использующих технологии искусственного интеллекта, в 38 больницах в 6 провинциях Китая в 2021 году [182]. В мировой практике описаны различные СППВР, функционирующие на разных принципах. Так, некоторые СППВР направлены на поддержку в диагностике заболеваний, другие типы СППВР имеют своей целью формулирование рекомендаций по лечению [341]. Важно отметить, что не все СППВР используют в своей работе индивидуальные геномные данные больных для подбора лечения, подобно СППВР, разработанной в настоящем исследовании.

Широко известными примерами СППВР для подбора персонализированного лечения ЗНО, учитывающими геномные характеристики, являются система FoundationOne[®] CDx (Foundation Medicine Inc, США), одобренная FDA, а также отечественные системы Oncobox и Atlas Solo Комплекс Плюс. Кроме того, необходимо отметить СППВР Cancer Seek (Caris Life Sciences), получившую одобрение FDA в 2024 году [406]. Предварительное изучение характеристик названных СППВР позволило заключить, что ни одна из них не способна в полной мере реализовывать задачи по подбору широкого спектра таргетных и иммунотерапевтических препаратов для персонализированного лечения ЗНО на основе данных полноэкзомного

секвенирования. В связи с этим, была разработана собственная СППВР, отличающаяся от названных отечественных и мировых аналогов. В рамках системы FoundationOne®CDx проводится секвенирование панели 324 генов, в Atlas Solo Комплекс Плюс - панели из 160 генов. Хотя в рамках системы MI Cancer Seek осуществляется исследование полного экзона для выявления уровня ТМВ, поиск однонуклеотидных вариантов и вставок/делеций производится в 228 генах [406]. Разработанная в настоящем исследовании СППВР способна подбирать лечение на основе результатов секвенирования полного экзона, что позволяет расширить перечень лекарственных препаратов для персонализированного лечения.

Еще одним отличием разработанной СППВР является более широкий перечень лекарственных препаратов, анализируемых при выявленных у больного биомаркерах. В системе Oncobox производится подбор персонализированного лечения из списка, состоящего из 160 препаратов, в системах FoundationOne®CDx и MI Cancer Seek перечень препаратов, исследуемых для подбора персонализированной терапии также ограничен. Разработанная нами СППВР функционирует с использованием данных, содержащихся в фармакогеномной платформе, что позволяет существенно расширить спектр анализируемых противоопухолевых препаратов. Для подбора персонализированного лекарственного лечения СППВР получает доступ к большим объемам информации из 18 открытых мировых баз данных, содержащих информацию о клинической значимости геномных вариантов и об ассоциациях биомаркеров с чувствительностью к противоопухолевым препаратам.

В условиях непрерывного увеличения объема информации о молекулярно-генетических характеристиках различных онкологических заболеваний, условием эффективного функционирования современных СППВР является их взаимодействие с фармакогеномными платформами, аккумулирующими большие объемы релевантной информации. В фармакогеномной платформе предусмотрены механизмы добавления новой информации, что позволяет СППВР опираться на актуальные мировые данные при подборе персонализированных рекомендаций. Разработанная в настоящем исследовании СППВР не интегрирована в состав фармакогеномной платформы и функционирует как отдельный элемент, получающий информацию для анализа из озера данных фармакогеномной платформы. Подобный подход был применен для обеспечения возможности приложения СППВР к иным озёрам/базам данных при возникновении такой необходимости. Более того, необходимо отметить, что гибкий дизайн разработанной СППВР также может позволить интегрировать информацию из открытых мировых баз данных в состав самой СППВР для автономной работы.

Созданная СППВР является связующим звеном между клиницистом и разработанной фармакогеномной платформой. СППВР включает аналитический блок, черпающий информацию из озера данных в составе фармакогеномной платформы и осуществляющий подбор

эффективного персонализированного лекарственного лечения на основе геномного варианта/ГМВ. СППВР также включает блок определения мутационного профиля опухоли и блок визуализации, позволяющий формировать отчеты о подборе персонализированного лечения ЗНО. Создание российской СППВР для подбора лекарственного лечения на основе геномных особенностей больных будет способствовать более широкому применению персонализированного подхода в отечественной онкологической практике.

Необходимо упомянуть сформулированные в литературе плюсы и минусы различных СППВР. Так, в качестве плюса выделяют снижение риска неправильного диагноза и лечения. Вместе с тем, среди потенциальных недостатков СППВР обозначают риск снижения автономии воли медицинского специалиста при принятии клинических решений [248]. В данном контексте необходимо отметить, что СППВР, созданная в настоящем исследовании, осуществляет подбор персонализированного лекарственного лечения ЗНО и формулирует результаты своей работы в виде рекомендаций. Сформулированные в отчете рекомендации СППВР не являются обязывающими для клиницистов, и решение о назначении окончательного лечения принимается сугубо лечащим врачом с учётом всех клинических данных, связанных с конкретным больным. Соответственно, разработанная СППВР оказывает поддержку клиницисту в принятии решения, но не подменяет его функцию по принятию решений. Разработанная СППВР предоставляет рекомендацию того или иного препарата с указанием в отчёте различных уровней доказательности, данных о регистрации/одобрении препарата и дат проведения исследований, что также учитывается лечащим врачом при назначении терапии.

Разработанная СППВР формирует рекомендации лекарственных препаратов в том случае, если у больного выявлен биомаркер, информация о клинической аннотации которого содержится в созданном озере данных. Вместе с тем, в тестовом режиме были опробованы механизмы предсказания влияния на чувствительность к различным лекарственным препаратам геномных вариантов, информация о клинической аннотации которых не содержалась в озере данных. Если информация о влиянии геномного варианта на чувствительность к препаратам не содержалась в озере данных, при этом любые иные геномные варианты в рассматриваемом гене всегда приводили к рекомендации одного и того же препарата, то такой вариант в тестовом режиме связывали с указанным лекарственным препаратом с пометкой о переназначении. Переназначение вариантов использовалось для того, чтобы в тестовом режиме учесть варианты, которые, вероятно, могут иметь значение для подбора лекарственного лечения. Вместе с тем, на настоящем этапе мы не формировали отчеты о рекомендации персонализированного лекарственного лечения онкологических заболеваний на основе таких геномных вариантов. Усовершенствование алгоритмов предсказания влияния геномных вариантов на

чувствительность к различным противоопухолевым лекарственным препаратам является актуальным направлением дальнейших исследований.

В настоящей работе образцы биологического материала больных РПЖ ($n=40$) из российской популяции были проанализированы с использованием разработанных фармакогеномной платформы и СППВР. По итогам анализа, СППВР сформировала рекомендации персонализированного противоопухолевого лекарственного лечения для 36 случаев (90%) РПЖ.

Также было установлено, что наличие соматического геномного варианта 12:g.25380276T>C (p.Gln61Arg) в гене *KRAS* при РПЖ обуславливает рекомендации наиболее широкого спектра различных препаратов и их комбинаций. При наличии названного геномного варианта СППВР было рекомендовано 16 различных препаратов/комбинаций препаратов: 1 - Abemaciclib, 2 - Atezolizumab, Nivolumab, 3 - Binimetinib, 4 - Docetaxel, 5 - Docetaxel, Selumetinib, 6 - Erlotinib, Docetaxel, 7 - Ингибитор MEK RO4987655, 8 - Pemetrexed, Trametinib, Docetaxel, 9 - Salirasib, 10 - SCH772984, Ингибитор PI3Kbeta AZD8186, 11 - Selumetinib, Cetuximab, 12 - Selumetinib, Erlotinib, 13 - Sorafenib, Refametinib, 14 - Temsirolimus, Ridaforolimus, 15 - Teprotumumab, Erlotinib, 16 - Trametinib.

По итогам анализа геномных вариантов больных ТНРМЖ ($n=100$), СППВР сформировала рекомендации персонализированного противоопухолевого лекарственного лечения для 83 случаев ТНРМЖ (83%). Наибольшее количество рекомендаций персонализированного лекарственного лечения было сформировано при наличии геномного варианта 17:g.37880220T>C (p.Leu755Ser) в гене *ERBB2*: 8 различных препаратов (1 - Neratinib, 2 - Ado-Trastuzumab Emtansine, 3 - Lapatinib, 4 - Margetuximab, 5 - Trastuzumab, 6 - Pertuzumab, 7 - Tucatinib, 8 - Trastuzumab Deruxtecane).

Таким образом, было протестировано функционирование разработанной СППВР на образцах опухолевой ткани больных РПЖ и ТНРМЖ. В данном исследовании не стояла задача проанализировать реальную клиническую эффективность рекомендаций системы. Вместе с тем, оценка эффективности таких рекомендаций является значимым направлением для будущих исследований, задел для которых был сформирован в настоящей работе.

В настоящее время в мировой клинической практике не существует решения в области СППВР, признанного в качестве “золотого стандарта”. Несовершенство существующих решений подчеркивается исследованием согласованности предложений по лечению пациентов в трех инструментах принятия решений (CureMatch, NAVIFY и QCI) на базе одной и той же панели из 77 генов у больных метастатическим РМЖ и раком легкого. Было выявлено наличие существенных расхождений в стратегии определения мишеней для лекарственных препаратов и

низкая частота согласованных событий. Соответственно, резко различались предлагаемые системами рекомендации [276]. Данное наблюдение подчеркивает потребность в продолжении исследований для разработки новых фармакогеномных платформ и СППВР.

В тестовом режиме нами было произведено сравнение рекомендаций по подбору персонализированного лечения при колоректальном раке, сформулированных разработанной СППВР и системой FoundationOne®CDx (Foundation Medicine, Inc., Cambridge, MA, USA), одобренной FDA.

На рисунке 67 показано, что совпадение предсказаний резистентности к препаратам Cetuximab и Panitumumab как системой FoundationOne®CDx (Foundation Medicine, Inc., Cambridge, MA, USA), так и разработанной в настоящем исследовании СППВР при колоректальном раке и наличии геномных вариантов в гене KRAS. Кроме того, разработанная в настоящем исследовании СППВР также предсказала резистентность/отсутствие ответа к комбинации препаратов Cetuximab и Panitumumab. Корректное предсказание резистентности к определенным противоопухолевым препаратам обладает крайне важным значением, поскольку позволяет принять решение о подборе иных более эффективных опций.

FOUNDATIONONE®CDx PATIENT TUMOR TYPE: Colon adenocarcinoma (CRC) REPORT DATE: 13 Apr 2022

Genomic Signatures
 Microsatellite status - MS-Stable
 Tumor Mutational Burden - 4 Muts/Mb
Gene Alterations
 KRAS G13D
 NRAS wildtype
 APC N965fs*14, R1450*
 PIK3CA H1047R, R88Q
 FAM123B R497*
 SMAD4 G386C
 2 Disease relevant genes with no reportable alterations: BRAF, NRAS

Report Highlights
 Targeted therapies with potential resistance based on this patient's genomic findings: **Cetuximab** (p. 8), **Panitumumab** (p. 8)
 Evidence-matched clinical trial options based on this patient's genomic findings: (p. 9)

GENOMIC SIGNATURES	THERAPY AND CLINICAL TRIAL IMPLICATIONS	
Microsatellite status - MS-Stable	No therapies or clinical trials. see Genomic Signatures section	
Tumor Mutational Burden - 4 Muts/Mb	No therapies or clinical trials. see Genomic Signatures section	
GENE ALTERATIONS	THERAPIES WITH CLINICAL RELEVANCE (IN PATIENT'S TUMOR TYPE)	THERAPIES WITH CLINICAL RELEVANCE (IN OTHER TUMOR TYPE)
KRAS - G13D 10 Trials see p. 10	Cetuximab ✖ Panitumumab ✖	none
NRAS - wildtype 0 Trials	Cetuximab ✖ Panitumumab ✖	none
APC - N965fs*14, R1450* 4 Trials see p. 9	none	none
PIK3CA - H1047R, R88Q 10 Trials see p. 12	none	none

✖ Extensive evidence showing variant(s) in this sample may confer resistance to this therapy

А

Эффективность таргетных и иммунотерапевтических препаратов

Пациент: Дима: Колоректальный рак Дата отчета: 21.12.2025

Эффективность таргетных и иммунотерапевтических препаратов

У пациента выявлены следующие соматические геномные варианты, имеющие терапевтическое значение:

Ген	Вариант	Область	Доля мутантного аллеля	Величина популяционного варианта
KRAS	ENST00000256078.4:c.35G>T ENSP00000256078.4:p.Gly12Val 17:g.25398284c>a p.G12V	Exon #2	0.336662	701x

Для данного типа злокачественного образования
 Препараты, зарегистрированные в Российской Федерации или одобренные FDA/EMA для лечения при данном типе злокачественного образования (ссылка на сайт):

Ген/Вариант	Тип рака	Уровень доказательств	Клиническая значимость	Ссылка на исследование	Рейтинг
Cetuximab, Panitumumab					
KRAS p.G12V	Колоректальный рак	A: Утвержденная рекомендация	Свидетельство за: Резистентность/отсутствие ответа	Benson et al., 2017, 3 N Engl J Med, 366, 1711-1720	★★★★

Для других показаний
 Препараты одобренные/зарегистрированные для лечения по другим показаниям (ссылка на сайт):

Ген/Вариант	Тип рака	Уровень доказательств	Клиническая значимость	Ссылка на исследование	Рейтинг
Panitumumab					
KRAS p.G12V	Колоректальный рак	B: Клиническое доказательство	Свидетельство за: Резистентность/отсутствие ответа	Amado et al., 2008, 3 Clin Oncol	★★★
Cetuximab					
KRAS p.G12V	Колоректальный рак	B: Клиническое доказательство	Свидетельство за: Резистентность/отсутствие ответа	Soupiras et al., 2005, W J Cancer	★★★

Представленная информация не имеет самостоятельного клинического (диагностического) значения.

Б

Рисунок 67. А) Фрагмент отчета системы FoundationOne®CDx (Foundation Medicine, Inc., Cambridge, MA, USA) по подбору персонализированного лечения при колоректальном раке: на основе геномного варианта в гене KRAS предсказана резистентность к препаратам Cetuximab и Panitumumab. Б) Фрагмент отчета разработанной СППВР по подбору персонализированного

лечения при колоректальном раке: на основе геномного варианта в гене *KRAS* предсказана резистентность к препаратам Cetuximab и Panitumumab, а также к комбинации данных препаратов

Наши предварительные результаты сравнения разработанной СППВР с референсной системой FoundationOne®CDx (Foundation Medicine, Inc., Cambridge, MA, USA) демонстрируют высокий уровень соответствия. Сравнение результатов рекомендаций персонализированного противоопухолевого лекарственного лечения, формируемых разработанной СППВР и мировыми признанными аналогами на больших выборках биологического материала, является перспективным направлением для дальнейших исследований.

Трансляция научных достижений последних лет создала предпосылки к широкому внедрению технологий персонализированной медицины в онкологическую практику. В числе персонализированных методов лечения - эффективные, но наиболее дорогостоящие биотехнологические и высокотехнологичные препараты: CAR-T клеточная терапия, дендритно-клеточные вакцины, моноклональные антитела, онколитические вирусы и другие. Сегодня данные препараты лечения активно выходят за пределы поля клинических исследований и занимают свое место в системе оказания помощи онкологическим пациентам. Большинство данных препаратов отличает техническая сложность и чрезвычайно высокая стоимость производства. Некоторые из них изготавливаются на “нерутинной” основе по индивидуальному назначению для конкретного пациента.

Согласно Решению Совета Евразийской экономической комиссии (ЕЭК) от 3 ноября 2016 г. N 78 "О Правилах регистрации и экспертизы лекарственных средств для медицинского применения" (в ред. от 20.10.2023 N 114) генотерапевтические лекарственные препараты, а также лекарственные препараты на основе соматических клеток и тканеинженерные лекарственные препараты относятся к Биотехнологическим лекарственным препаратам (БТЛП). В соответствии с п 5.1. Решения Совета ЕЭК: "5.1. В случаях и порядке, предусмотренных законодательством государств-членов, допускается предоставление пациентам и применение в отношении их незарегистрированных лекарственных препаратов. К таким лекарственным препаратам относятся в том числе: высокотехнологичные лекарственные препараты, изготавливаемые на нестандартизированной (нерутинной) основе и применяемые на территории того же государства-члена в стационаре в целях исполнения индивидуального медицинского назначения лекарственного препарата, специально произведенного для отдельного пациента."

В соответствии с Федеральным законом от 12 апреля 2010 года N 61-ФЗ, ред от 30 января 2024 государственной регистрации не подлежат: биотехнологические лекарственные препараты, предназначенные для применения в соответствии с индивидуальным медицинским назначением и имеющие в своем составе соединения, синтезированные по результатам генетических

исследований материала, полученного от пациента, и могут быть изготовлены и применены в стационаре.

Новое правовое регулирование отражает потребности существующей онкологической практики в персонализированном лечении и создает условия для широкого внедрения СППВР и фармакогеномики в клиническую практику для прогнозирования эффективности данных препаратов и определения показаний к их назначению.

Обеспечение высокого уровня медицинской помощи больным ЗНО требует внедрения новейших и наиболее перспективных подходов в клиническую практику. Эффективность персонализации лекарственного лечения онкологических заболеваний с учетом индивидуального геномного профиля подтверждена многочисленными клиническими и аналитическими исследованиями. Вместе с тем, грамотная реализация фармакогеномного подхода к лечению онкологических заболеваний является комплексной задачей, требующей разработки вычислительных, аналитических, биоинформатических и иных инструментов, позволяющих обрабатывать и хранить большие объемы справочной, геномной и клинико-фенотипической информации. В рамках настоящего исследования был разработан отечественный комплекс “фармакогеномная платформа - СППВР”, объединяющий различные инструменты, в том числе основанные на технологиях машинного обучения.

Разработанные фармакогеномная платформа и СППВР были применены для анализа биологического материала больных РПЖ и ТНРМЖ из российской популяции и позволили выявить ландшафт геномных вариантов при данных типах ЗНО, а также сформировать рекомендации по подбору персонализированного лекарственного лечения. С учетом полученных результатов представляется возможным заключить, что применение разработанной фармакогеномной платформы и СППВР обладает высоким потенциалом для дальнейшего совершенствования медицинской помощи при онкологических заболеваниях.

ЗАКЛЮЧЕНИЕ

Повышение эффективности лечения онкологических заболеваний является одной из приоритетных задач в мировой и отечественной клинической практике. Некоторые типы ЗНО характеризуются агрессивным течением и плохим прогнозом ввиду ограниченного перечня вариантов действенного лечения. В связи с этим, разработка подходов, способных привести к улучшению клинических исходов представляет особую значимость.

На сегодняшний день широкое признание получили подходы, обеспечивающие персонализацию противоопухолевого лечения. Индивидуальные геномные характеристики больных и опухолевых клеток являются эффективными рычагами воздействия на ЗНО и основой для разработки различных таргетных и иммунотерапевтических препаратов [53]. Проведенные знаковые исследования подчеркнули преимущества подхода, основанного на персонализации с учетом геномных особенностей, перед традиционными тактиками химиотерапевтического лечения [293]. В эпоху геномной медицины, объем знаний о потенциальных мишенях для терапии, результатах исследований различных уровней, разработках новых препаратов, ассоциациях геномных биомаркеров с эффективностью противоопухолевых лекарственных средств, растет с высокой скоростью. На сегодняшний день практикующему онкологу необходимы специализированные инструменты, позволяющие быстро и эффективно ориентироваться, и применять существующий массив данных. Таким инструментом являются фармакогеномные платформы, способные аккумулировать и хранить большие объемы различных данных, релевантных для сферы онкологии, и обеспечивать эффективную работу с ними.

Перспективность разработки и применения фармакогеномных платформ подтверждается практикой. Существуют различные платформы, среди которых фармакогеномный портал SYNERGxDB, аккумулирующий данные о синергетическом эффекте разных противоопухолевых препаратов [310], а также Интегративная фармакогеномная платформа раковых клеточных линий и тканей, обеспечивающая поиск геномных характеристик клеточных линий, чувствительных к противоопухолевым лекарственным средствам [320]. Кроме того, следует отметить платформы, разработанные в авторитетных зарубежных медицинских учреждениях, например, в Национальном онкологическом центре Республики Корея [271]. Разработка отечественной платформы представляет особый интерес, поскольку такой инструмент сможет обеспечить не только интеграцию знаний об ассоциациях биомаркеров с эффективностью противоопухолевых препаратов, но и аккумулирование и защищенное хранение больших массивов геномных, клинических и иных данных о больных из российской популяции. Такая информация, в свою очередь, сможет стать солидной основой для проведения фундаментальных исследований,

расширяющих карту знаний. Кроме того, гибкий дизайн платформ позволяет интегрировать в их структуру конвейеры для биоинформатической обработки данных высокопроизводительного секвенирования. В этой связи необходимо отметить актуальность внедрения инструментов, на основе технологий машинного обучения, обеспечивающих полноту и точность выявления геномных вариантов [282].

Для применения массива фармакогеномных данных в реальной клинической практике необходим инструмент, обеспечивающий взаимодействие практикующего специалиста с платформой, а также генерирующий наглядные отчеты с результатами подбора персонализированного лечения. Такой функционал реализуют СППВР. Актуальность создания СППВР подтверждается наличием признанных зарубежных аналогов. Так, система FoundationOne® CDx (Foundation Medicine, США), одобренная FDA, активно применяется для подбора персонализированного лечения на основе геномного профиля опухолей [238]. Тем не менее, зарубежные решения часто труднодоступны или слишком затратны для широкого внедрения в клиническую практику в России. Отечественные аналоги пока немногочисленны: существуют система Oncobox [330], а также генетический тест Atlas Solo Комплекс Плюс [403], однако указанные решения обладают рядом ограничений, например, по количеству анализируемых лекарственных препаратов, либо числу исследуемых генов. В связи с изложенным выше, приобретает актуальность разработка новой СППВР, не обладающей названными ограничениями и позволяющей формировать персонализированные рекомендации по подбору противоопухолевых лекарственных средств из наиболее широкого перечня препаратов.

Разработка отечественной фармакогеномной платформы и СППВР будет способствовать внедрению персонализированного подхода к лечению ЗНО в клиническую практику, а также получению и накоплению уникального массива геномных, клинических и иных данных о больных ЗНО из российской популяции, тем самым расширяя существующий объем медико-биологических знаний.

Цель настоящего исследования была сформулирована следующим образом: разработать фармакогеномную платформу и систему поддержки принятия врачебных решений для подбора персонализированного лекарственного лечения онкологических заболеваний на основе анализа данных высокопроизводительного полноэкзомного секвенирования биологического материала больных с применением методов машинного обучения.

Для достижения поставленной цели были поставлены и решены задачи: 1. Провести мета-анализ клинических исследований медикаментозного лечения различных злокачественных новообразований, реализованных с применением персонализации на основе индивидуальных геномных характеристик больных либо без персонализации, для определения клинической

эффективности указанных подходов. 2. Разработать адаптированный протокол пробоподготовки и секвенирования экзомных ДНК-библиотек и получить сырые данные высокопроизводительного секвенирования из парных образцов крови и опухолевого материала больных раком поджелудочной железы и трижды негативным раком молочной железы из российской популяции. 3. Разработать биоинформатические конвейеры обработки данных высокопроизводительного секвенирования, включающие инструменты на основе технологий машинного обучения, для выявления геномных вариантов, показателей опухолевой мутационной нагрузки и микросателлитной нестабильности. 4. Обработать данные полноэкзомного секвенирования с применением разработанных биоинформатических конвейеров и описать ландшафт геномных вариантов больных раком поджелудочной железы и трижды негативным раком молочной железы для расширения карты знаний по российской популяции. 5. Создать единое локальное пополняемое “озеро данных” (data lake) с включением результатов выявления геномных вариантов, клинико-фенотипической информации, информации из открытых мировых баз данных, необходимой для аннотирования геномных вариантов и формирования рекомендаций персонализированного лечения, и иных сведений. 6. Спроектировать и разработать фармакогеномную платформу на основе озера данных и биоинформатических конвейеров и провести оценку её эффективности в сравнении с эталонными решениями на примере трижды негативного рака молочной железы путем сопоставительного анализа результатов выявления геномных вариантов. 7. Разработать систему поддержки принятия врачебных решений, обеспечивающую взаимодействие врача с фармакогеномной платформой и формирующую рекомендации персонализированного лекарственного лечения онкологических заболеваний, и оценить её пригодность на примере данных полноэкзомного секвенирования образцов рака поджелудочной железы и трижды негативного рака молочной железы.

В рамках проведения мета-анализа был осуществлен поиск клинических исследований 2-й фазы, посвященных лечению ЗНО и реализованных либо с персонализацией на основе индивидуальных геномных характеристик больных, либо без персонализации. На различных этапах мета-анализа руководствовались рекомендациями PRISMA (англ. Preferred reporting items for systematic reviews and meta-analyses. Предпочтительные элементы отчетности для систематических обзоров и мета-анализов) [268], и применяли необходимые методы статической обработки.

После проведения мета-анализа решали практические задачи. Настоящая работа является биоинформатическим и клиническим ретроспективным исследованием. В рамках диссертационной работы были созданы фармакогеномная платформа и СППВР. Для апробации данных инструментов был использован биологический материал (образцы опухолевой ткани и периферической крови) больных РПЖ (n=40) и ТНРМЖ (n=100), проходивших лечение в ФГБУ

«НМИЦ Радиологии» Минздрава России в период с 2016 по 2023 гг. По обозначенным больным также был собран массив клинико-фенотипических данных для описания когорт и интеграции в фармакогеномную платформу.

При разработке адаптированного протокола пробоподготовки и секвенирования производили модификацию существующего протокола обогащения экзомных библиотек с использованием реактивов Roche (США) и Agilent (США) для его адаптации к секвенатору MGISEQ-G400 (MGI, Китай). Было произведено модифицирование (полная замена) блокирующих олигонуклеотидов для обеспечения работы зондов Roche и Agilent на библиотеках MGI. В рамках модифицирования протокола обогащения библиотек были разработаны и синтезированы блокирующие олигонуклеотиды Block 1 и Block 2 с LNA-модификацией (добавлением замкнутых нуклеотидов).

Был произведен анализ существующих биоинформатических конвейеров Sarek [130], Vcbio-Nextgen [150], SpeedSeq [82], DNAr [75], и DNAscan2 [230], по итогам которого было выявлено их несоответствие сформулированным нами требованиям. Это обусловило задачу по разработке собственных биоинформатических конвейеров для определения герминального и соматического геномного профиля. Разработка собственных конвейеров основывалась на лучших практиках построения биоинформатических конвейеров GATK. Обработку данных производили на языках Bash, Python, R, многие плагины были написаны на языках Perl и C++.

Выявление ландшафта геномных вариантов при РПЖ и ТНPMЖ на основе данных парного полноэкзомного секвенирования производили с использованием разработанных биоинформатических конвейеров и адаптированного протокола пробоподготовки и секвенирования. Результаты были визуализированы в виде графиков типа “Хит-мэп” (*англ. heatmap*).

Была разработана архитектура фармакогеномной платформы, предполагающая включение в её состав озера данных наряду с биоинформатическими конвейерами и техническим блоком. Для создания озера данных был использован современный инструмент управления Dataiku. Методология наполнения озера предполагала интеграцию широкого спектра различных собственных данных (геномных, клинико-фенотипических, лабораторных и других), а также опубликованной информации из открытых мировых источников, необходимой для аннотирования геномных вариантов и аккумуляции сведений об ассоциациях биомаркеров с чувствительностью к противоопухолевым лекарственным препаратам.

Оценка эффективности работы созданной фармакогеномной платформы была осуществлена путем сравнения геномных вариантов, выявленных разработанными биоинформатическими конвейерами, и референсными наборами и программами для биоинформатической обработки AVENIO (Roche, США).

При создании СППВР использовали языки программирования JavaScript, NodeJS, R, Python, Bash и системы управления пакетами conda, npm; Linux (Debian 10) применяли в качестве основной операционной системы для разработки и выполнения программ. Также использовали открытые программные пакеты NodeJS, Playwright, TAPES и Chromium.

В ходе выполнения настоящей работы получены важные научно-практические результаты, которые в совокупности представляют собой значимый вклад в развитие персонализированной медицины в онкологии. Проведённый сравнительный мета-анализ клинических исследований убедительно продемонстрировал преимущество персонализации на основе геномных характеристик больных при лекарственной терапии ЗНО. Обобщение данных 49 исследований (7012 пациентов) показало, что при назначении лекарственного лечения с учетом результатов генетического тестирования частота объективного ответа повышается на 18% ($p=0.009$), выживаемость без прогрессирования на 9,94 мес ($p=0.018$), а общая выживаемость - на 3,2 мес ($p=0.032$). Полученные данные согласуются с более ранними публикациями [308], и, дополняя их свежими сведениям, укрепляют доказательную базу персонализированной терапии. Результаты мета-анализа свидетельствуют о том, что геномное профилирование пациента перед лечением способно привести к улучшению клинических исходов, что подчеркивает фундаментальную целесообразность разработки отечественной фармакогеномной платформы и СППВР.

Была создана отечественная фармакогеномная платформа для интеграции и анализа больших биомедицинских данных в сфере онкологии. В рамках проведенных работ в озеро платформы добавлен уникальный массив собственных данных о 3686 пациентах, проходивших лечение и/или обследование в филиалах ФГБУ «НМИЦ радиологии» Минздрава России. Больные, сведения о которых были интегрированы в озеро данных, наиболее часто страдали раком молочной железы (2214), РПЖ (331), раком яичников (253). В озере был аккумулирован широкий перечень собственных данных о больных ЗНО из российской популяции: 1. Регистрационные и клинико-фенотипические данные. 2. Лабораторные метрики пробоподготовки и секвенирования биологического материала. 3. Нуклеотидная последовательность ДНК больных (необработанные данные секвенирования в формате FASTQ и их метрики качества). 4. Герминальные и соматические геномные варианты, показатели опухолевой мутационной нагрузки, микросателлитной нестабильности и тд., выявленные по итогам работы биоинформатических конвейеров. 5. Метаданные, метрики качества работы герминального и соматического биоинформатических конвейеров. 6. Результаты подбора персонализированного лекарственного лечения онкологических заболеваний, сформированные СППВР. Агрегирование в озере данных фармакогеномной платформы большого объема знаний, характеризующих ЗНО в российской популяции, является важным результатом, на основе

которого возможно проведение фундаментальных исследований в онкологии. В частности, суммарный объем интегрированных в озеро необработанных данных секвенирования в формате FASTQ составляет 26,2 терабайта. После биоинформатической обработки, в озеро интегрированы сведения о 65 438 000 герминальных и о 42 493 соматических геномных вариантах.

Также в озеро добавлено 1132 Гб информации из 18 открытых мировых баз данных: GENCODE, dbNSFP, COSMIC, ClinVar, 1000 Genome project, ExAC, dbSNP, OMIM, gnomAD, Pfam, UniProtKB, DoCM, CancerHotspots, CancerMine, OpenTargetsPlatform, ChEMBL, CIViC, CGI. Интеграция информации из указанных источников позволила нам создать собственную базу данных, обеспечивающую реализацию 2 важнейших функций: 1. аннотирование геномных вариантов разработанными биоинформатическими конвейерами; 2. подбор персонализированного лекарственного лечения СППВР на основе аккумулированных сведений об ассоциациях биомаркеров с чувствительностью к противоопухолевым лекарственным препаратам. Благодаря проведенной интеграции информации, аннотирование и подбор персонализированного лечения ЗНО могут производиться без подключения к сети Интернет. Важнейшей особенностью разработанного решения является возможность его автономного функционирования в защищенном контуре медицинской организации: все необходимые данные для принятия решений хранятся локально, и их пополнение происходит пакетно по мере необходимости.

Вторым важнейшим компонентом созданной фармакогеномной платформы являются разработанные оригинальные биоинформатические конвейеры, включающие инструменты на основе технологий машинного обучения. Внедрение таких инструментов в процесс обработки данных позволило с высокой точностью выявлять герминальные и соматические геномные варианты. Проверка работы конвейеров в составе фармакогеномной платформы на реальных данных полноэкзомного секвенирования биологического материала больных ТНРМЖ продемонстрировала их высокую эффективность: все клинически значимые геномные варианты обнаруженные референсными протоколами и программами для биоинформатической обработки AVENIO (Roche, США), были выявлены разработанными биоинформатическими конвейерами.

Созданные в настоящей работе адаптированный протокол пробоподготовки и секвенирования, а также биоинформатические конвейеры были успешно апробированы на образцах биологического материала больных РПЖ и ТНРМЖ, что позволило описать геномный ландшафт при данных заболеваниях в российской популяции. Было выявлено, что герминальные геномные варианты при РПЖ содержались в 16240 генах. При этом в генах *ATM*, *BRCA2*, *COG4*, *DNMT1*, *PCNT*, *SLC12A6*, *SYNE2*, *WNK1* геномные варианты были у всех больных из исследуемой группы. Наибольшее число геномных вариантов отмечено в генах *APC*, *MUC19*, *MUC16*, *BRCA2*,

BRCA1, MUC4, OBSCN, PKDIL2, TTN, AHNAK2. В свою очередь детальное изучение ландшафта значимых патогенных и вероятно патогенных герминальных геномных вариантов при РПЖ показало наличие таких вариантов у 4 (10%) больных из когорты РПЖ: у 2 больных в гене *ATM*, у 1 больного в гене *ERCC3*, в *FANCM* - также у 1 больного.

Соматические геномные варианты при РПЖ были выявлены в 3223 генах. Детальное исследование значимых соматических вариантов 1-го и 2-го классов патогенности продемонстрировало их наличие у 35 больных (87,5%) РПЖ, причем у 25 было одновременно нескольких таких геномных вариантов. У наибольшего количества больных РПЖ (n=33) были выявлены значимые соматические варианты в гене *KRAS*, при этом у 2 больных наблюдалось одновременно 2 различных варианта. У 24 больных были обнаружены значимые соматические варианты в гене *TP53*. Также были выявлены значимые соматические варианты в генах *CDKN2A, ATM, SMAD4, RB1, EGFR, KMT2C, PDGFRA, SETBP1* и *SF3B1* [314]. Уровень ТМВ был в диапазоне от 0,02 до 14,3 мут/Мб (среднее $2,55 \pm 2,32$; медиана 2,23). Высокая ТМВ (≥ 10 мут/Мб) была выявлена у 1 больного (2,5%) РПЖ. У всех больных РПЖ был стабильный микросателлитный статус. Полученные данные согласуются с известной картиной мутационного профиля РПЖ: как и в глобальных исследованиях, в российской выборке подтверждено преобладание мутаций в генах *KRAS* и *TP53*.

Применение разработанного протокола пробоподготовки и секвенирования, а также биоинформатических конвейеров позволило описать геномный ландшафт при ТНПМЖ. Герминальные геномные варианты были выявлены в 19240 генах. Наибольшее количество вариантов было в генах *MUC19, MUC16, PWRN1, TTN, OBSCN, PKDIL2, MUC4, AHNAK2* и *LAMA5*. Вместе с тем, анализ значимых для подбора персонализированного лечения патогенных и вероятно патогенных герминальных геномных вариантов показал их присутствие у 28 больных (28%) ТНПМЖ. В наибольшем количестве случаев ТНПМЖ (n=18) патогенные герминальные варианты были детектированы в гене *BRCA1*. У 4 больных такие варианты были в гене *BRCA2*. Также в когорте больных ТНПМЖ были выявлены значимые герминальные варианты в генах *BARD1, FANCA, FANCI, FANCL, FANCM, PMS2, RAD52* и в *SLX4*. При этом у 2 больных было обнаружено одновременно 2 значимых герминальных геномных варианта (1- *BRCA1* и *FANCI*, 2 - *BRCA1* и *SLX4*).

Соматические геномные варианты при ТНПМЖ были выявлены в 14328 генах. При этом варианты 1-го и 2-го классов патогенности, значимые для подбора персонализированного лечения, были обнаружены у 84 больных (84%) ТНПМЖ. У 32 больных наблюдалось наличие одновременно нескольких таких соматических геномных вариантов в различных генах либо в одном гене. У наибольшего числа больных ТНПМЖ (n=64) значимые соматические варианты были выявлены в гене *TP53*, при этом у 3 больных было выявлено одновременно 2 различных

варианта в этом гене. У 16 больных - в гене *PIK3CA*, при этом у одной больной было выявлено одновременно 2 варианта). Также значимые соматические геномные варианты были выявлены в генах *PTEN*, *SETBP1*, *RB1*, *KMT2C*, *ASXL1*, *NOTCH1*, *KRAS*, *BRCA1*, *FBXW7*, *STK11*, *PTCH1*, *ESR1*, *NOTCH2*, *WT1*, *AKT1*, *VAP1*, *CDKN2A*, *BIRC3*, *ERBB2*, *EZH2*, *CTNNB1*, *GNA11*, *PRDM1*, *NRAS*, *PIK3CB*, *SMARCA4*, *PTPRB*, *SF3B1*, *SMARCB1*, *TERT*, *TSC1*. Уровни ТМВ были в диапазоне от 1,3721 мут/Мб до 98,8837 мут/Мб с медианой 5,66 мут/Мб и средним значением $10,02 \pm 14,68$. Высокая ТМВ (≥ 10 мут/Мб) выявлена у 19 больных (19%) ТНРМЖ. У всех больных был стабильный микросателлитный статус. Таким образом, платформа дала возможность расширить научные знания о генетических характеристиках РПЖ и ТНРМЖ в выборке больных из российской популяции. Накопление подобных сведений способствует углублению фундаментального понимания природы данных опухолей и может послужить основой для поиска новых прогностических и предиктивных факторов.

В настоящей работе была разработана оригинальная СППВР, являющаяся связующим звеном между врачом и фармакогеномной платформой. СППВР реализует функции по подбору персонализированного лекарственного лечения ЗНО, а также предоставляет отчеты о сформированных рекомендациях. СППВР является отдельным инструментом, не входящим в состав фармакогеномной платформы, но опирающимся на информацию, содержащуюся в её озере данных. В сформированном отчёте указываются обнаруженные у пациента значимые молекулярно-генетические изменения, и на их основе даются рекомендации по назначению противоопухолевых лекарственных препаратов. При этом учитывается широкий спектр доступных терапевтических опций: таргетные и иммунотерапевтические препараты, химиотерапия. Рекомендации сопровождаются обоснованием: СППВР автоматически подтягивает из озера данных справочные сведения (клинические исследования, одобренные показания), подтверждающие эффективность предложенного подхода. Таким образом, врач получает научно обоснованную поддержку в режиме реального времени, что особенно важно при разборе сложных случаев и выборе из множества возможных опций лечения. Разработанная СППВР была успешно применена к данным полноэкзомного секвенирования образцов РПЖ и ТНРМЖ и продемонстрировала свою эффективность: были сформированы рекомендации персонализированного лекарственного лечения для 36 больных РПЖ (90%) и для 83 больных ТНРМЖ (83%).

Таким образом, в настоящей работе были созданы оригинальные инструменты, способствующие развитию как фундаментальной науки, так и практической персонализированной медицины в онкологии. Разработанная отечественная фармакогеномная платформа позволяет аккумулировать большой массив геномных, клинико-фенотипических и других данных о различных аспектах ЗНО у больных из российской популяции. С учетом

большого количества пациентов, проходящих лечение в ФГБУ «НМИЦ радиологии» Минздрава России, накопленная информация представляет собой ценный ресурс для проведения онкологических исследований.

Успешное функционирование фармакогеномной платформы, в частности - биоинформатических конвейеров в её составе, было продемонстрировано на биологическом материале больных крайне агрессивными ЗНО, такими как РПЖ и ТНРМЖ. По итогам были получены ценные данные о геномном профиле при данных заболеваниях в российской популяции. Полноэкзомное секвенирование позволило расширить спектр исследуемых регионов ДНК в целях аккумуляции ценной геномной информации в озере данных.

Разработанная фармакогеномная платформа также непосредственно служит цели развития персонализированной медицины, поскольку она обеспечивает эффективную работу СППВР. Интеграция справочной информации из открытых мировых баз данных в озеро платформы обеспечила независимость работы созданных инструментов от внешних источников, что важно для бесперебойной клинической практики. Разработка СППВР являлась актуальной задачей, поскольку на настоящий момент передовые стандарты оказания медицинской помощи в сфере онкологии подразумевают способность клинициста оперировать большими объемами данных по молекулярной биологии опухолей и сведениями о персонализированных терапевтических опциях.

Важным итогом стала возможность подбора широкого спектра таргетных, иммунотерапевтических и химиотерапевтических лекарственных препаратов, без ограничения количества анализируемых вариантов. Создание отечественного комплекса «фармакогеномная платформа - СППВР» вносит весомый вклад в формирование современной парадигмы лечения ЗНО, где решения принимаются с опорой на геномную информацию пациента, что в конечном счете способствует повышению эффективности и адресности противоопухолевой терапии.

ВЫВОДЫ

1. Из результатов проведенного мета-анализа значительного массива данных (7012 больных) следует, что подбор персонализированного лекарственного лечения онкологических заболеваний на основе индивидуальных геномных характеристик больных способен улучшить клинические исходы, в том числе повысить частоту объективного ответа на 18% ($p=0.009$), выживаемость без прогрессирования на 9,94 мес ($p=0.018$) и общую выживаемость на 3,2 мес ($p=0.032$) за счет повышения эффективности лекарственного лечения.
2. Разработанный протокол адаптации реагентов производства компании Roche (США) и Agilent Technologies (США) к секвенатору MGISEQ-G400 производства компании MGI (Китай) позволяет перейти с платформы Illumina (США) на платформу MGI Tech (Китай) без потери качества парного полноэкзомного секвенирования при выполнении молекулярно-генетического анализа в онкологии.
3. Разработаны биоинформатические конвейеры, включающие инструменты на основе технологий машинного обучения. Интеграция таких инструментов позволяет с высокой точностью определять соматические и герминальные варианты при возможной кросс-валидации, что уменьшает ошибку вызова единичных нуклеотидных полиморфизмов и инделей.
4. Обработка данных парного полноэкзомного секвенирования с использованием разработанных биоинформатических конвейеров позволила выявить спектр геномных вариантов, характерных для больных раком поджелудочной железы и трижды негативным раком молочной железы. Полученные данные расширили существующие представления и дополнили карту знаний о генетических вариациях в российской популяции, что, в перспективе, может способствовать формированию персонализированных подходов к диагностике и лечению онкологических заболеваний.
5. Созданное озеро данных содержит обширный объем регулярно обновляемой информации, включая 1132 Гб структурированных данных из международных референсных источников (генетических, фармакогеномных и клинических баз), а также геномные, клиничко-фенотипические и иные сведения о 3686 пациентах из российской популяции. Централизованное хранение этих данных в озере обеспечивает автономную работу платформы и системы поддержки принятия врачебных решений даже при отсутствии подключения к сети Интернет.

6. Созданная фармакогеномная платформа, основанная на интеграции озера данных и биоинформатических конвейеров, продемонстрировала сопоставимую эффективность в сравнении с эталонными решениями на примере трижды негативного рака молочной железы. Это подтверждает применимость платформы для задач клинической геномики и её потенциал в контексте персонализированной медицины.
7. Разработанная система поддержки принятия врачебных решений, реализующая функционал анализа больших объемов геномной и справочной информации для подбора персонализированного лекарственного лечения различных типов злокачественных новообразований, продемонстрировала свою пригодность в результате апробации на данных полноэкзомного секвенирования парных образцов рака поджелудочной железы и трижды негативного рака молочной железы.
8. Созданная фармакогеномная платформа также является научной платформой, агрегирующей большие объёмы геномной, клинико-фенотипической и иной информации о больных злокачественными новообразованиями из российской популяции. Накопленный массив данных платформы не только является неотъемлемым условием эффективного функционирования разработанной системы поддержки принятия врачебных решений, но и открывает новые возможности для проведения дальнейших крупных аналитических исследований в онкологии.

ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ

1. На основе результатов клинических исследований лекарственного лечения ЗНО, персонализированного с учётом индивидуальных геномных характеристик и результатов собственного мета-анализа, рекомендовано придерживаться подхода, учитывающего при назначении лечения геномные характеристики каждого из конкретных больных ЗНО.

2. Для подбора персонализированного лекарственного лечения ЗНО рекомендовано осуществлять молекулярно-генетическое исследование с использованием технологии полноэкзомного высокопроизводительного секвенирования. Рекомендовано следовать протоколам пробоподготовки и секвенирования, подтвердившим свою эффективность, в том числе - адаптированному протоколу пробоподготовки и секвенирования, разработанному в настоящем исследовании.

3. При проведении биоинформатического анализа данных высокопроизводительного секвенирования следует использовать биоинформатические конвейеры, включающие инструменты на основе технологий машинного обучения для эффективного выявления геномных вариантов, в том числе - биоинформатические конвейеры, разработанные в настоящем исследовании.

4. Для подбора персонализированного лекарственного лечения ЗНО рекомендовано использовать разработанную фармакогеномную платформу, аккумулирующую геномную, справочную и иную информацию, а также СППВР анализирующую биомаркеры, выявленные у больного.

5. При составлении плана лечения врачу следует ознакомиться с отчётом СППВР о рекомендациях персонализированного лекарственного лечения, сформированным с учётом выявленных геномных характеристик больного, и оценить предложения по их использованию в качестве вспомогательного инструмента, при этом оставляя окончательный выбор лекарственного лечения непосредственно за врачом.

6. Необходимо интегрировать информацию о геномных и клинико-фенотипических характеристиках больных ЗНО из российской популяции в озеро данных фармакогеномной платформы для накопления и последующего анализа больших объёмов информации о различных аспектах онкологических заболеваний.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

АПХТ - адъювантная полихимиотерапия

БД - база данных

БП1 - базальноподобный 1 подтип трижды негативного рака молочной железы

БП2 - базальноподобный 2 подтип трижды негативного рака молочной железы

БПИА - базальноподобный иммуноактивный подтип трижды негативного рака молочной железы

БПИС - базальноподобный иммуносупрессивный подтип трижды негативного рака молочной железы

БПРМЖ - базальноподобный рак молочной железы

ВБП - выживаемость без прогрессирования

ВПМН - внутрисекторные папиллярные муцинозные новообразования

ГР - гомологичная рекомбинация

ДИ - доверительный интервал

ЕС - Европейский Союз

ЗНО - злокачественное новообразование

ИГХ - иммуногистохимическое исследование

ИИ - искусственный интеллект

ИМ - иммуномодулирующий подтип трижды негативного рака молочной железы

КТ - компьютерная томография

ЛАР - люминальный с рецептором андрогенов подтип трижды негативного рака молочной железы

МП - мезенхимоподобный подтип трижды негативного рака молочной железы

МРТ - магнитно-резонансная терапия

МСП - мезенхимальный стволоподобный подтип трижды негативного рака молочной железы

НАПХТ - неадъювантная полихимиотерапия

ОВ - общая выживаемость

ОР - отношение рисков

ПАПЖ - протоковая аденокарцинома поджелудочной железы

ПО - программное обеспечение

ПЦР - полимеразная цепная реакция

РА - рецептор андрогенов

РМЖ - рак молочной железы

РПЖ - рак поджелудочной железы

РП - рецепторы прогестерона

РЭ - рецепторы эстрогенов

СПЕ - синдром Пейтца-Егерса

СППВР - система поддержки принятия врачебных решений

СУБД - система управления базами данных

США - Соединенные Штаты Америки

ТНРМЖ - трижды негативный рак молочной железы

ТРК - тропомиозин-рецепторная киназа

ТТ - таргетная терапия

ХТ - химиотерапия

ЧОО - частота объективного ответа

ЭВМ - электронная вычислительная машина

AJCC (англ. American Joint Committee on Cancer) - Американская объединенная комиссия по злокачественным новообразованиям

API (англ. application programming interface) - программный интерфейс приложения

ASCAT (англ. allele-specific copy number analysis of tumors) - анализ аллель-специфического числа копий в опухолях

ASCO (англ. American Society of Clinical Oncology) - Американское общество клинической онкологии

BAM (англ. binary alignment map) - карта двоичного выравнивания

CAPS (англ. Cancer of the Pancreas Screening Study) - Международный Консорциум по скринингу рака поджелудочной железы

CNV (англ. copy number variation) - вариация числа копий

CPS (англ. combined positive score) - комбинированный положительный балл

CPSR (англ. Cancer predisposition sequencing reporter) - программа оценки предрасположенности к онкологии на основе секвенирования

dMMR (англ. deficient mismatch repair) - дефицит репарации ошибочно спаренных нуклеотидов

DNB (англ. DNA nanoball) - ДНК наносферы

EGFR (англ. epidermal growth factor receptor)- рецептор эпидермального фактора роста

ETL процессы (англ. extract, transform, load) - процессы извлечения, преобразования, загрузки

FDA (англ. U.S. Food and Drug Administration) - Управление по санитарному надзору за качеством пищевых продуктов и медикаментов США

FTP - (англ. File Transfer Protocol) - протокол передачи файлов

FFPE - (англ. Formalin-Fixed Paraffin-Embedded) - фиксированные в формалине парафинизированные образцы

GATK (англ. Genome Analysis Toolkit) - набор инструментов для анализа генома

GWAS (англ. Genome-wide association study) - полногеномный поиск ассоциаций

ICGC - (англ. International Cancer Genome Consortium) - Международный консорциум по геному рака

ID - идентификатор

INDEL (англ. insertions/deletions) - вставки/делеции

LDAP (англ. lightweight directory access protocol) - облегчённый протокол доступа к каталогам

miRNA (англ. microRNA, miRNA)- МикроРНК

MSI (англ. microsatellite instability) - микросателлитная нестабильность

MSS (англ. microsatellite stable) - стабильный микросателлитный статус

NGS (англ. next generation sequencing) - секвенирование нового поколения

PanIN (англ. pancreatic intraepithelial neoplasia) - панкреатическая интраэпителиальная неоплазия

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

PCGR (англ. Personal Cancer Genome Reporter) - программа Персональный репортёр генома рака

pMMR (англ. proficient mismatch repair) - высокая степень репарации неспаренных оснований

SNP (англ. single nucleotide polymorphism) - однонуклеотидный полиморфизм

SNV (англ. single nucleotide variant) - однонуклеотидный вариант

TGF- β (англ. transforming growth factor beta) - трансформирующий фактор роста- β

TCGA - (англ. The Cancer Genome Atlas) - Атлас генома рака

TMB (англ. tumor mutational burden) - опухолевая мутационная нагрузка

VCF (англ. variant call format) - формат вызова вариантов

WGS (англ. WGS, whole genome sequencing) - полногеномное секвенирование

СПИСОК ЛИТЕРАТУРЫ

1. Гордиев М. Г., Бровкина О. И., Еникеев Р. Ф. Руководство по интерпретации клинически значимых соматических мутаций при солидных опухолях, выявленных методом секвенирования следующего поколения (NGS), с целью их клинического использования / М. Г. Гордиев, О. И. Бровкина, Р. Ф. Еникеев, Общероссийское изд., Москва:, 2020.
2. Иванов С. А. [и др.]. Способ прогноза прогрессирования заболевания у больных раком желудочно-кишечного тракта после проведенного лечения // Патент на изобретение № 2758079, 2021 г.
3. Каприн А. Д., Старинский В. В., Шахзадова А. О. Состояние онкологической помощи населению России в 2022 году. / А. Д. Каприн, В. В. Старинский, А. О. Шахзадова, МНИОИ им.-е изд., М.:, 2023. 239 с.
4. Кубышкин В. А., Вишневский В. А. Рак поджелудочной железы. / В. А. Кубышкин, В. А. Вишневский, Москва:, 2003.
5. Попович А. Ю. Рак поджелудочной железы // Вестник клуба панкреатологов. 2009. № 3. С. 31–39.
6. Потиевский М. Б., Шегай П. В., Каприн А. Д. Перспективы применения методов эволюционной биологии в онкологии. // Журнал эволюционной биохимии и физиологии. 2022. № 2 (58). С. 84–95.
7. Романова М. Э., [и др.]. Оценка частоты встречаемости молекулярных нарушений в опухолевых клетках при злокачественных новообразованиях молочной железы методом высокопроизводительного секвенирования // Онкология. Журнал им. П.А. Герцена. А Герцена. 2021. № 4 (10). С. 23–29.
8. Русаков И. Г., [и др.]. Роль молекулярно-генетических нарушений гена VHL в развитии светлоклеточного рака почки и оптимизации таргетной терапии сунитинибом // Российский онкологический журнал. 2009. № 5. С. 15–18.
9. Семиглазов В. Ф., [и др.]. Неoadъювантное и постнеoadъювантное лечение трижды негативного рака молочной железы // Фарматека. // Фарматека. 2020. № 11 (27). С. 8–13.
10. Фалалеева Н. А., [и др.]. Иммуноterapia: неoadъювант и/или адъювант? Обзор литературы. // Практическая онкология. 2023. № 3 (24). С. 337–357.
11. Федорова И. К., [и др.]. Лечение молодых пациентов с раком полости рта: текущее состояние и перспективы (консорциум “Этиология и патогенез рака полости рта у пациентов молодого возраста”). Часть 1 // Голова и шея. 2022. № 4 (10). С. 83–91.
12. Федорова И. К., [и др.]. Лечение молодых пациентов с раком полости рта: текущее состояние и перспективы (консорциум «Этиология и патогенез рака полости рта у пациентов

- молодого возраста»). Часть 2. // Голова и шея. 2023. № 1 (11). С. 74–84.
13. Циммерман Я. С. Рак поджелудочной железы: terra incognita современной гастроэнтерологии // Клиническая медицина. 2015. № 10 (93). С. 5–13.
14. Шаталов П. А., Веселовский, Е.М. Райгородская М. П., [и др.]. Интеграция NGS-тестирования со стандартными методами молекулярно-генетических исследований в онкологии // Онкология. Журнал им. П.А. Герцена. 2024. № 6 (13). С. 84-90.
15. Шаталов П. А., [и др.]. Роль фармакогеномики в персонифицированном лечении онкологических заболеваний // Онкология. Журнал им. П.А. Герцена. 2023. Т. 12. № 6. С. 75-81.
16. Aaltonen L. A. [и др.]. Pan-cancer analysis of whole genomes // Nature. 2020. № 7793 (578). С. 82–93.
17. Abdelwahab O., Belzile F., Torkamaneh D. Performance analysis of conventional and AI-based variant callers using short and long reads // BMC Bioinformatics. 2023. № 1 (24). С. 472.
18. Abraham S. C. [и др.]. Distinctive Molecular Genetic Alterations in Sporadic and Familial Adenomatous Polyposis-Associated Pancreatoblastomas // The American Journal of Pathology. 2001. № 5 (159). С. 1619–1627.
19. Abraham S. C. [и др.]. Solid-Pseudopapillary Tumors of the Pancreas Are Genetically Distinct from Pancreatic Ductal Adenocarcinomas and Almost Always Harbor β -catenin Mutations // The American Journal of Pathology. 2002. № 4 (160). С. 1361–1369.
20. Abreu F. B. de [и др.]. Effective quality management practices in routine clinical next-generation sequencing // Clinical Chemistry and Laboratory Medicine (CCLM). 2016. № 5 (54).
21. Adams S. [и др.]. Pembrolizumab monotherapy for previously treated metastatic triple-negative breast cancer: cohort A of the phase II KEYNOTE-086 study // Annals of Oncology. 2019. № 3 (30). С. 397–404.
22. Afghahi A., Kurian A. W. The Changing Landscape of Genetic Testing for Inherited Breast Cancer Predisposition // Current Treatment Options in Oncology. 2017. № 5 (18). С. 27.
23. Agaimy A. [и др.]. Pancreatic undifferentiated rhabdoid carcinoma: KRAS alterations and SMARCB1 expression status define two subtypes // Modern Pathology. 2015. № 2 (28). С. 248–260.
24. Aguirre A. J. [и др.]. Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma // Genes & Development. 2003. № 24 (17). С. 3112–3126.
25. Ahn E. R. [и др.]. Olaparib (O) in patients (pts) with pancreatic cancer with BRCA1/2 inactivating mutations: Results from the Targeted Agent and Profiling Utilization Registry (TAPUR) study. // Journal of Clinical Oncology. 2020. № 15_suppl (38). С. 4637–4637.
26. Ahn E. R. [и др.]. Olaparib in Patients With Pancreatic Cancer With BRCA1 / 2 Mutations: Results From the Targeted Agent and Profiling Utilization Registry Study // JCO Precision Oncology. 2024. № 8.

27. Ajith A. K. [и др.]. Chemotherapeutic Resistance Genes of Breast Cancer Patients – An Overview // *Advanced Pharmaceutical Bulletin*. 2021.
28. Alföldi J., Lindblad-Toh K. Comparative genomics as a tool to understand evolution and disease // *Genome Research*. 2013. № 7 (23). С. 1063–1068.
29. Alhmoud J. F. [и др.]. DNA Damage/Repair Management in Cancers // *Cancers*. 2020. № 4 (12). С. 1050.
30. Ali S. [и др.]. Cell Cycle and Factors Involved in Inhibition or Progression of Breast Cancer // *IntechOpen*, 2020.
31. Alsop K. [и др.]. BRCA Mutation Frequency and Patterns of Treatment Response in BRCA Mutation–Positive Women With Ovarian Cancer: A Report From the Australian Ovarian Cancer Study Group // *Journal of Clinical Oncology*. 2012. № 21 (30). С. 2654–2663.
32. Amato E. [и др.]. Targeted next-generation sequencing of cancer genes dissects the molecular profiles of intraductal papillary neoplasms of the pancreas // *The Journal of Pathology*. 2014. № 3 (233). С. 217–227.
33. Anderson K. [и др.]. Genetic variegation of clonal architecture and propagating cells in leukaemia // *Nature*. 2011. № 7330 (469). С. 356–361.
34. Andor N. [и др.]. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity // *Nature Medicine*. 2016. № 1 (22). С. 105–113.
35. André F. [и др.]. Alpelisib for PIK3CA -Mutated, Hormone Receptor–Positive Advanced Breast Cancer // *New England Journal of Medicine*. 2019. № 20 (380). С. 1929–1940.
36. Andrés R. [и др.]. Association of BRCA1 germline mutations in young onset triple-negative breast cancer (TNBC) // *Clinical and Translational Oncology*. 2014. № 3 (16). С. 280–284.
37. Andrew Clayton T. [и др.]. Pharmaco-metabonomic phenotyping and personalized drug treatment // *Nature*. 2006. № 7087 (440). С. 1073–1077.
38. Anestis A. [и др.]. Androgen Receptor in Breast Cancer—Clinical and Preclinical Research Insights // *Molecules*. 2020. № 2 (25). С. 358.
39. Antoniou A. [и др.]. Average Risks of Breast and Ovarian Cancer Associated with BRCA1 or BRCA2 Mutations Detected in Case Series Unselected for Family History: A Combined Analysis of 22 Studies // *The American Journal of Human Genetics*. 2003. № 5 (72). С. 1117–1130.
40. Artcibasova A. V. [и др.]. MiRImpact, a new bioinformatic method using complete microRNA expression profiles to assess their overall influence on the activity of intracellular molecular pathways // *Cell Cycle*. 2016. № 5 (15). С. 689–698.
41. Aslanian H. R., Lee J. H., Canto M. I. AGA Clinical Practice Update on Pancreas Cancer Screening in High-Risk Individuals: Expert Review // *Gastroenterology*. 2020. № 1 (159). С. 358–362.
42. Atchley D. P. [и др.]. Clinical and Pathologic Characteristics of Patients With BRCA -Positive and

- BRCA -Negative Breast Cancer // *Journal of Clinical Oncology*. 2008. № 26 (26). С. 4282–4288.
43. Auwerx C. [и др.]. From pharmacogenetics to pharmaco-omics: Milestones and future directions // *Human Genetics and Genomics Advances*. 2022. № 2 (3). С. 100100.
44. Babayan A. [и др.]. Failure of immune checkpoint inhibitors for microsatellite instability-positive pancreatic adenocarcinoma with atypical pattern of short tandem repeat mutation // *Immunotherapy*. 2024. № 13 (16). С. 853–858.
45. Baghdadi T. Al [и др.]. Palbociclib in Patients With Pancreatic and Biliary Cancer With CDKN2A Alterations: Results From the Targeted Agent and Profiling Utilization Registry Study // *JCO Precision Oncology*. 2019. № 3. С. 1–8.
46. Bailey P. [и др.]. Genomic analyses identify molecular subtypes of pancreatic cancer // *Nature*. 2016. № 7592 (531). С. 47–52.
47. Baker W. L. [и др.]. Understanding heterogeneity in meta-analysis: the role of meta-regression // *International Journal of Clinical Practice*. 2009. № 10 (63). С. 1426–1434.
48. Bardia A. [и др.]. Sacituzumab Govitecan-hziy in Refractory Metastatic Triple-Negative Breast Cancer // *New England Journal of Medicine*. 2019. № 8 (380). С. 741–751.
49. Barker T. H. [и др.]. Conducting proportional meta-analysis in different types of systematic reviews: a guide for synthesisers of evidence // *BMC Medical Research Methodology*. 2021. № 1 (21). С. 189.
50. Basturk O. [и др.]. A Revised Classification System and Recommendations From the Baltimore Consensus Meeting for Neoplastic Precursor Lesions in the Pancreas // *American Journal of Surgical Pathology*. 2015. № 12 (39). С. 1730–1741.
51. Belova V. [и др.]. Comparative evaluation of four exome enrichment solutions in 2024: Agilent, Roche, Vazyme and Nanodigmbio // *BMC Genomics*. 2025. № 1 (26). С. 76.
52. Berraondo P., Umansky V., Melero I. Changing the Tumor Microenvironment: New Strategies for Immunotherapy // *Cancer Research*. 2012. № 20 (72). С. 5159–5164.
53. Bhat G. R. [и др.]. Cancer cell plasticity: from cellular, molecular, and genetic mechanisms to tumor heterogeneity and drug resistance. // *Cancer metastasis reviews*. 2024. № 1 (43). С. 197–228.
54. Biankin A. V. [и др.]. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes // *Nature*. 2012. № 7424 (491). С. 399–405.
55. Birkbak N. J. [и др.]. Paradoxical Relationship between Chromosomal Instability and Survival Outcome in Cancer // *Cancer Research*. 2011. № 10 (71). С. 3447–3452.
56. Blackford A. [и др.]. SMAD4 Gene Mutations Are Associated with Poor Prognosis in Pancreatic Cancer // *Clinical Cancer Research*. 2009. № 14 (15). С. 4674–4679.
57. Bloomston M. [и др.]. MicroRNA Expression Patterns to Differentiate Pancreatic Adenocarcinoma From Normal Pancreas and Chronic Pancreatitis // *JAMA*. 2007. № 17 (297). С. 1901.

58. Boddy A. M., Harrison T. M., Abegglen L. M. Comparative Oncology: New Insights into an Ancient Disease // *iScience*. 2020. № 8 (23). С. 101373.
59. Bonavia R. [и др.]. Heterogeneity Maintenance in Glioblastoma: A Social Network // *Cancer Research*. 2011. № 12 (71). С. 4055–4060.
60. Borazanci E. [и др.]. An Analysis of Patients with DNA Repair Pathway Mutations Treated with a PARP Inhibitor // *The Oncologist*. 2020. № 1 (25). С. e60–e67.
61. Bousman C. A., Zierhut H., Müller D. J. Navigating the Labyrinth of Pharmacogenetic Testing: A Guide to Test Selection // *Clinical Pharmacology & Therapeutics*. 2019. № 2 (106). С. 309–312.
62. Bozic I. [и др.]. Evolutionary dynamics of cancer in response to targeted combination therapy // *eLife*. 2013. (2).
63. Branzei D., Foiani M. Regulation of DNA repair throughout the cell cycle // *Nature Reviews Molecular Cell Biology*. 2008. № 4 (9). С. 297–308.
64. Burg S. H. van der [и др.]. Vaccines for established cancer: overcoming the challenges posed by immune evasion // *Nature Reviews Cancer*. 2016. № 4 (16). С. 219–233.
65. Burgess S. [и др.]. Combining multiple imputation and meta-analysis with individual participant data // *Statistics in Medicine*. 2013. № 26 (32). С. 4499–4514.
66. Burstein H. J. [и др.]. Estimating the benefits of therapy for early-stage breast cancer: the St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019 // *Annals of Oncology*. 2019. № 10 (30). С. 1541–1557.
67. Burstein M. D. [и др.]. Comprehensive Genomic Analysis Identifies Novel Subtypes and Targets of Triple-Negative Breast Cancer // *Clinical Cancer Research*. 2015. № 7 (21). С. 1688–1698.
68. Byrski T. [и др.]. Response to neo-adjuvant chemotherapy in women with BRCA1-positive breast cancers // *Breast Cancer Research and Treatment*. 2008. № 2 (108). С. 289–296.
69. Cagan A. [и др.]. Somatic mutation rates scale with lifespan across mammals // *Nature*. 2022. № 7906 (604). С. 517–524.
70. Caldas C. [и др.]. Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma // *Nature Genetics*. 1994. № 1 (8). С. 27–32.
71. Cameron D. [и др.]. Adjuvant bevacizumab-containing therapy in triple-negative breast cancer (BEATRICE): primary results of a randomised, phase 3 trial // *The Lancet Oncology*. 2013. № 10 (14). С. 933–942.
72. Canto M. I. [и др.]. Frequent Detection of Pancreatic Lesions in Asymptomatic High-Risk Individuals // *Gastroenterology*. 2012. № 4 (142). С. 796–804.
73. Carey L. A. [и др.]. Race, Breast Cancer Subtypes, and Survival in the Carolina Breast Cancer Study // *JAMA*. 2006. № 21 (295). С. 2492.
74. CAREY M. [и др.]. Financial and social impact of supporting a haematological cancer survivor //

- European Journal of Cancer Care. 2012. № 2 (21). С. 169–176.
75. Causey J. L. [и др.]. DNAP: A Pipeline for DNA-seq Data Analysis // *Scientific Reports*. 2018. № 1 (8). С. 6793.
76. Cecchin E., Stocco G. Pharmacogenomics and Personalized Medicine // *Genes*. 2020. № 6 (11). С. 679.
77. Chalmers Z. R. [и др.]. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden // *Genome Medicine*. 2017. № 1 (9). С. 34.
78. Chang T.-C. [и др.]. Transactivation of miR-34a by p53 Broadly Influences Gene Expression and Promotes Apoptosis // *Molecular Cell*. 2007. № 5 (26). С. 745–752.
79. Chang T.-C. [и др.]. Somatic and Germline Variant Calling from Next-Generation Sequencing Data 2022.С. 37–54.
80. Chang Y. [и др.]. Cancer carrier screening in the general population using whole-genome sequencing // *Cancer Medicine*. 2023. № 2 (12). С. 1972–1983.
81. Chen S., Parmigiani G. Meta-Analysis of BRCA1 and BRCA2 Penetrance // *Journal of Clinical Oncology*. 2007. № 11 (25). С. 1329–1333.
82. Chiang C. [и др.]. SpeedSeq: ultra-fast personal genome analysis and interpretation // *Nature Methods*. 2015. № 10 (12). С. 966–968.
83. Choi M., Kipps T., Kurzrock R. ATM Mutations in Cancer: Therapeutic Implications // *Molecular Cancer Therapeutics*. 2016. № 8 (15). С. 1781–1791.
84. Chun S.-M. [и др.]. Next-Generation Sequencing Using S1 Nuclease for Poor-Quality Formalin-Fixed, Paraffin-Embedded Tumor Specimens // *The Journal of Molecular Diagnostics*. 2018. № 6 (20). С. 802–811.
85. Cibulskis K. [и др.]. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples // *Nature Biotechnology*. 2013. № 3 (31). С. 213–219.
86. Cives M., Strosberg J. R. Gastroenteropancreatic Neuroendocrine Tumors // *CA: A Cancer Journal for Clinicians*. 2018. № 6 (68). С. 471–487.
87. Cohen J. D. [и др.]. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers // *Proceedings of the National Academy of Sciences*. 2017. № 38 (114). С. 10202–10207.
88. Coignard J. [и др.]. A case-only study to identify genetic modifiers of breast cancer risk for BRCA1/BRCA2 mutation carriers // *Nature Communications*. 2021. № 1 (12). С. 1078.
89. Collins K. K. [и др.]. Effects of breast cancer surgery and surgical side effects on body image over time // *Breast Cancer Research and Treatment*. 2011. № 1 (126). С. 167–176.
90. Collins M. A. [и др.]. Oncogenic Kras is required for both the initiation and maintenance of pancreatic cancer in mice // *Journal of Clinical Investigation*. 2012. № 2 (122). С. 639–653.

91. Comen E. [и др.]. Relative contributions of BRCA1 and BRCA2 mutations to “triple-negative” breast cancer in Ashkenazi Women // *Breast Cancer Research and Treatment*. 2011. № 1 (129). С. 185–190.
92. Cortes-Ciriano I. [и др.]. A molecular portrait of microsatellite instability across multiple cancers // *Nature Communications*. 2017. № 1 (8). С. 15180.
93. Cortes J. [и др.]. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial // *The Lancet*. 2020. № 10265 (396). С. 1817–1828.
94. Couch F. J. [и др.]. Inherited Mutations in 17 Breast Cancer Susceptibility Genes Among a Large Triple-Negative Breast Cancer Cohort Unselected for Family History of Breast Cancer // *Journal of Clinical Oncology*. 2015. № 4 (33). С. 304–311.
95. Couch M. E. [и др.]. Cancer cachexia update in head and neck cancer: Definitions and diagnostic features // *Head & Neck*. 2015. № 4 (37). С. 594–604.
96. Daniele A. [и др.]. Can harmful lifestyle, obesity and weight changes increase the risk of breast cancer in BRCA 1 and BRCA 2 mutation carriers? A Mini review // *Hereditary Cancer in Clinical Practice*. 2021. № 1 (19). С. 45.
97. Dent R. [и др.]. Triple-Negative Breast Cancer: Clinical Features and Patterns of Recurrence // *Clinical Cancer Research*. 2007. № 15 (13). С. 4429–4434.
98. Dent R. [и др.]. Pattern of metastatic spread in triple-negative breast cancer // *Breast Cancer Research and Treatment*. 2009. № 2 (115). С. 423–428.
99. Dere W. H., Suto T. S. The role of pharmacogenetics and pharmacogenomics in improving translational medicine. // *Clinical cases in mineral and bone metabolism : the official journal of the Italian Society of Osteoporosis, Mineral Metabolism, and Skeletal Diseases*. 2009. № 1 (6). С. 13–6.
100. DerSimonian R., Laird N. Meta-analysis in clinical trials // *Controlled Clinical Trials*. 1986. № 3 (7). С. 177–188.
101. Derynck R., Zhang Y. E. Smad-dependent and Smad-independent pathways in TGF- β family signalling // *Nature*. 2003. № 6958 (425). С. 577–584.
102. Díaz-Navarro A. [и др.]. Rfcaller: a machine learning approach combined with read-level features to detect somatic mutations // *NAR Genomics and Bioinformatics*. 2023. № 2 (5).
103. Dienstmann R. [и др.]. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer // *Nature Reviews Cancer*. 2017. № 2 (17). С. 79–92.
104. Dinov I. D. [и др.]. Applications of the pipeline environment for visual informatics and genomics computations // *BMC Bioinformatics*. 2011. № 1 (12). С. 304.
105. Dlamini Z. [и др.]. Artificial intelligence (AI) and big data in cancer and precision oncology //

Computational and Structural Biotechnology Journal. 2020. (18). С. 2300–2311.

106. Doebele R. C. [и др.]. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1–2 trials // *The Lancet Oncology*. 2020. № 2 (21). С. 271–282.

107. Dogan B. E., Turnbull L. W. Imaging of triple-negative breast cancer // *Annals of Oncology*. 2012. (23). С. vi23-vi29.

108. Doherty G. J. [и др.]. Cancer Treatment in the Genomic Era // *Annual Review of Biochemistry*. 2019. № 1 (88). С. 247–280.

109. Doi T. [и др.]. Safety, pharmacokinetics, and antitumour activity of trastuzumab deruxtecan (DS-8201), a HER2-targeting antibody–drug conjugate, in patients with advanced breast and gastric or gastro-oesophageal tumours: a phase 1 dose-escalation study // *The Lancet Oncology*. 2017. № 11 (18). С. 1512–1522.

110. Doyle M. A. [и др.]. Studying Cancer Genomics Through Next-Generation DNA Sequencing and Bioinformatics 2014.С. 83–98.

111. Drilon A. [и др.]. Efficacy of Larotrectinib in TRK Fusion–Positive Cancers in Adults and Children // *New England Journal of Medicine*. 2018. № 8 (378). С. 731–739.

112. Easton D. F. [и др.]. Gene-Panel Sequencing and the Prediction of Breast-Cancer Risk // *New England Journal of Medicine*. 2015. № 23 (372). С. 2243–2257.

113. Egger S. J. [и др.]. Platinum-containing regimens for triple-negative metastatic breast cancer // *Cochrane Database of Systematic Reviews*. 2020. № 12 (2020).

114. El-Deiry W. S. [и др.]. The current state of molecular testing in the treatment of patients with solid tumors, 2019 // *CA: A Cancer Journal for Clinicians*. 2019. № 4 (69). С. 305–343.

115. Emens L. A. [и др.]. First-line atezolizumab plus nab-paclitaxel for unresectable, locally advanced, or metastatic triple-negative breast cancer: IMpassion130 final overall survival analysis // *Annals of Oncology*. 2021. № 8 (32). С. 983–993.

116. Epelbaum R. [и др.]. Molecular Profiling-Selected Therapy for Treatment of Advanced Pancreaticobiliary Cancer: A Retrospective Multicenter Study // *BioMed Research International*. 2015. (2015). С. 1–9.

117. Esplin E. D., Oei L., Snyder M. P. Personalized Sequencing and the Future of Medicine: Discovery, Diagnosis and Defeat of Disease // *Pharmacogenomics*. 2014. № 14 (15). С. 1771–1790.

118. Falzone L., Salomone S., Libra M. Evolution of Cancer Pharmacological Treatments at the Turn of the Third Millennium // *Frontiers in Pharmacology*. 2018. (9).

119. Farkona S., Diamandis E. P., Blasutig I. M. Cancer immunotherapy: the beginning of the end of cancer? // *BMC Medicine*. 2016. № 1 (14). С. 73.

120. Farmer H. [и др.]. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic

strategy // *Nature*. 2005. № 7035 (434). С. 917–921.

121. Favero F. [и др.]. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data // *Annals of Oncology*. 2015. № 1 (26). С. 64–70.

122. Ferlay J. [и др.]. Global Cancer Observatory: Cancer Today. // Lyon, France: International Agency for Research on Cancer. [Электронный ресурс]. URL: <https://gco.iarc.who.int/today>.

123. Finelli L. A., Narasimhan V. Leading a Digital Transformation in the Pharmaceutical Industry: Reimagining the Way We Work in Global Drug Development // *Clinical Pharmacology & Therapeutics*. 2020. № 4 (108). С. 756–761.

124. Flaherty K. T. [и др.]. Inhibition of Mutated, Activated BRAF in Metastatic Melanoma // *New England Journal of Medicine*. 2010. № 9 (363). С. 809–819.

125. Fonseca B. de P. [и др.]. Geographic accessibility to cancer treatment in Brazil: A network analysis // *The Lancet Regional Health - Americas*. 2022. (7). С. 100153.

126. Ford D. [и др.]. Genetic Heterogeneity and Penetrance Analysis of the BRCA1 and BRCA2 Genes in Breast Cancer Families // *The American Journal of Human Genetics*. 1998. № 3 (62). С. 676–689.

127. Frank T. S. [и др.]. Genomic profiling guides the choice of molecular targeted therapy of pancreatic cancer // *Cancer Letters*. 2015. № 1 (363). С. 1–6.

128. Fussenegger M., Bailey J. E. Molecular Regulation of Cell-Cycle Progression and Apoptosis in Mammalian Cells: Implications for Biotechnology // *Biotechnology Progress*. 1998. № 6 (14). С. 807–833.

129. Gamble C. R. [и др.]. Precision medicine testing in ovarian cancer: The growing inequity between patients with commercial vs medicaid insurance // *Gynecologic Oncology*. 2021. № 1 (162). С. 18–23.

130. Garcia M. [и др.]. Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants // *F1000Research*. 2020. (9). С. 63.

131. Gareth E. D. [и др.]. MRI breast screening in high-risk women: cancer detection and survival analysis // *Breast Cancer Research and Treatment*. 2014. № 3 (145). С. 663–672.

132. Garg A., Mago V. Role of machine learning in medical research: A survey // *Computer Science Review*. 2021. (40). С. 100370.

133. Garraway L. A., Lander E. S. Lessons from the Cancer Genome // *Cell*. 2013. № 1 (153). С. 17–37.

134. Gatenby R. A., Brown J. S. Integrating evolutionary dynamics into cancer therapy // *Nature Reviews Clinical Oncology*. 2020. № 11 (17). С. 675–686.

135. Gerlinger M. [и др.]. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing // *Nature Genetics*. 2014. № 3 (46). С. 225–233.

136. Ghadirian P. [и др.]. Breast cancer risk in relation to the joint effect of BRCA mutations and diet diversity // *Breast Cancer Research and Treatment*. 2009. № 2 (117). С. 417–422.
137. Giardiello F. M. [и др.]. Very high risk of cancer in familial Peutz–Jeghers syndrome // *Gastroenterology*. 2000. № 6 (119). С. 1447–1453.
138. Gluckman P. [и др.]. *Principles of Evolutionary Medicine* / P. Gluckman, A. Beedle, T. Buklijas, F. Low, M. Hanson, Oxford University Press, 2016.
139. Goggins M. [и др.]. Genetic alterations of the transforming growth factor beta receptor genes in pancreatic and biliary adenocarcinomas. // *Cancer research*. 1998. № 23 (58). С. 5329–32.
140. Goggins M. [и др.]. Management of patients with increased risk for familial pancreatic cancer: updated recommendations from the International Cancer of the Pancreas Screening (CAPS) Consortium // *Gut*. 2020. № 1 (69). С. 7–17.
141. Golan T. [и др.]. Maintenance Olaparib for Germline BRCA -Mutated Metastatic Pancreatic Cancer // *New England Journal of Medicine*. 2019. № 4 (381). С. 317–327.
142. Gonzalez-Angulo A. M. [и др.]. Incidence and Outcome of BRCA Mutations in Unselected Patients with Triple Receptor-Negative Breast Cancer // *Clinical Cancer Research*. 2011. № 5 (17). С. 1082–1089.
143. Gopal P. [и др.]. Clonal selection confers distinct evolutionary trajectories in BRAF-driven cancers // *Nature Communications*. 2019. № 1 (10). С. 5143.
144. Grant R. C. [и др.]. Prevalence of Germline Mutations in Cancer Predisposition Genes in Patients With Pancreatic Cancer // *Gastroenterology*. 2015. № 3 (148). С. 556–564.
145. Greaves M., Maley C. C. Clonal evolution in cancer // *Nature*. 2012. № 7381 (481). С. 306–313.
146. Grossberg A. J. [и др.]. Multidisciplinary standards of care and recent progress in pancreatic ductal adenocarcinoma // *CA: A Cancer Journal for Clinicians*. 2020. № 5 (70). С. 375–403.
147. Grunspan D. Z. [и др.]. Core principles of evolutionary medicine // *Evolution, Medicine, and Public Health*. 2018. № 1 (2018). С. 13–23.
148. Gucalp A. [и др.]. Phase II Trial of Bicalutamide in Patients with Androgen Receptor–Positive, Estrogen Receptor–Negative Metastatic Breast Cancer // *Clinical Cancer Research*. 2013. № 19 (19). С. 5505–5512.
149. Gudmundsson S. [и др.]. Variant interpretation using population databases: Lessons from gnomAD // *Human Mutation*. 2022. № 8 (43). С. 1012–1030.
150. Guimera R. V. bcbio-nextgen: Automated, distributed next-gen sequencing pipeline // *EMBnet.journal*. 2012. № B (17). С. 30.
151. Guolo A., Varin C. Random-effects meta-analysis: the number of studies matters // *Statistical Methods in Medical Research*. 2017. № 3 (26). С. 1500–1518.
152. Habbe N. [и др.]. MicroRNA miR-155 is a biomarker of early pancreatic neoplasia // *Cancer*

Biology & Therapy. 2009. № 4 (8). С. 340–346.

153. Harinck F. [и др.]. A multicentre comparative prospective blinded analysis of EUS and MRI for screening of pancreatic cancer in high-risk individuals // *Gut*. 2016. № 9 (65). С. 1505–1513.
154. Hartman A. [и др.]. Prevalence of BRCA mutations in an unselected population of triple-negative breast cancer // *Cancer*. 2012. № 11 (118). С. 2787–2795.
155. Hayashi A. [и др.]. A unifying paradigm for transcriptional heterogeneity and squamous features in pancreatic ductal adenocarcinoma // *Nature Cancer*. 2020. № 1 (1). С. 59–74.
156. He Y. [и др.]. Clinical and genetic characteristics in pancreatic cancer from Chinese patients revealed by whole exome sequencing // *Frontiers in Oncology*. 2023. (13).
157. Heijden M. S. van der [и др.]. In vivo Therapeutic Responses Contingent on Fanconi Anemia/BRCA2 Status of the Tumor // *Clinical Cancer Research*. 2005. № 20 (11). С. 7508–7515.
158. Heijnsdijk E. A. M. [и др.]. Differences in Natural History between Breast Cancers in BRCA1 and BRCA2 Mutation Carriers and Effects of MRI Screening-MRISC, MARIBS, and Canadian Studies Combined // *Cancer Epidemiology, Biomarkers & Prevention*. 2012. № 9 (21). С. 1458–1468.
159. Higgins J. P. T., López-López J. A., Aloe A. M. *Meta-Regression First edition*. | Boca Raton : Taylor and Francis, [2020] | Series: Chapman & Hall/CRC handbooks of modern statistical methods: Chapman and Hall/CRC, 2020.С. 129–150.
160. Higgins J. P. T., Thompson S. G. Quantifying heterogeneity in a meta-analysis // *Statistics in Medicine*. 2002. № 11 (21). С. 1539–1558.
161. Hingorani S. R. [и др.]. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse // *Cancer Cell*. 2003. № 6 (4). С. 437–450.
162. Hingorani S. R. [и др.]. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice // *Cancer Cell*. 2005. № 5 (7). С. 469–483.
163. Holter S. [и др.]. Germline BRCA Mutations in a Large Clinic-Based Cohort of Patients With Pancreatic Adenocarcinoma // *Journal of Clinical Oncology*. 2015. № 28 (33). С. 3124–3129.
164. Hong S.-M. [и др.]. Multiple genes are hypermethylated in intraductal papillary mucinous neoplasms of the pancreas // *Modern Pathology*. 2008. № 12 (21). С. 1499–1507.
165. Honrado E. [и др.]. Immunohistochemical Expression of DNA Repair Proteins in Familial Breast Cancer Differentiate BRCA2 -Associated Tumors // *Journal of Clinical Oncology*. 2005. № 30 (23). С. 7503–7511.
166. Hosoda W. [и др.]. Genetic analyses of isolated high-grade pancreatic intraepithelial neoplasia (HG-PanIN) reveal paucity of alterations in TP53 and SMAD4 // *The Journal of Pathology*. 2017. № 1 (242). С. 16–23.
167. Housman D., Ledley F. D. Why pharmacogenomics? Why now? // *Nature Biotechnology*. 1998.

№ 6 (16). С. 492–493.

168. Howlader N. [и др.]. The Effect of Advances in Lung-Cancer Treatment on Population Mortality // *New England Journal of Medicine*. 2020. № 7 (383). С. 640–649.

169. Hruban R. H. [и др.]. K-ras oncogene activation in adenocarcinoma of the human pancreas. A study of 82 carcinomas using a combination of mutant-enriched polymerase chain reaction analysis and allele-specific oligonucleotide hybridization. // *The American journal of pathology*. 1993. № 2 (143). С. 545–54.

170. Hruban R. H. [и др.]. Progression model for pancreatic cancer. // *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2000. № 8 (6). С. 2969–72.

171. Hruban R. H., Pitman M. B., Klimstra D. S. *Tumors of the Pancreas* / R. H. Hruban, M. B. Pitman, D. S. Klimstra, American Registry of Pathology Washington, DC, 2007.

172. Hu C. [и др.]. Association Between Inherited Germline Mutations in Cancer Predisposition Genes and Risk of Pancreatic Cancer // *JAMA*. 2018. № 23 (319). С. 2401.

173. Hu C. [и др.]. Multigene Hereditary Cancer Panels Reveal High-Risk Pancreatic Cancer Susceptibility Genes // *JCO Precision Oncology*. 2018. № 2. С. 1–28.

174. Huang L. [и др.]. KRAS mutation: from undruggable to druggable in cancer // *Signal Transduction and Targeted Therapy*. 2021. № 1 (6). С. 386.

175. Huang R., Zhou P.-K. DNA damage repair: historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy // *Signal Transduction and Targeted Therapy*. 2021. № 1 (6). С. 254.

176. Hudis C. A., Gianni L. Triple-Negative Breast Cancer: An Unmet Medical Need // *The Oncologist*. 2011. № S1 (16). С. 1–11.

177. Humphris J. L. [и др.]. Hypermutation In Pancreatic Cancer // *Gastroenterology*. 2017. № 1 (152). С. 68–74.e2.

178. Hyman D. M. [и др.]. HER kinase inhibition in patients with HER2- and HER3-mutant cancers // *Nature*. 2018. № 7691 (554). С. 189–194.

179. Iacobuzio-Donahue C. A. [и др.]. DPC4 Gene Status of the Primary Carcinoma Correlates With Patterns of Failure in Patients With Pancreatic Cancer // *Journal of Clinical Oncology*. 2009. № 11 (27). С. 1806–1813.

180. Isakoff S. J. Triple-Negative Breast Cancer // *The Cancer Journal*. 2010. № 1 (16). С. 53–61.

181. Jackson D., Bowden J. Confidence intervals for the between-study variance in random-effects meta-analysis using generalised heterogeneity statistics: should we use unequal tails? // *BMC Medical Research Methodology*. 2016. № 1 (16). С. 118.

182. Ji M. [и др.]. Status of AI-Enabled Clinical Decision Support Systems Implementations in China // *Methods of Information in Medicine*. 2021. № 05/06 (60). С. 123–132.

183. Jimenez R. E. [и др.]. Sequential Accumulation of K-ras Mutations and p53 Overexpression in the Progression of Pancreatic Mucinous Cystic Neoplasms to Malignancy // *Annals of Surgery*. 1999. № 4 (230). С. 501.
184. Johansen J. S. [и др.]. The potential diagnostic value of serum microRNA signature in patients with pancreatic cancer // *International Journal of Cancer*. 2016. № 10 (139). С. 2312–2324.
185. Jones J. B. [и др.]. Detection of mitochondrial DNA mutations in pancreatic cancer offers a «mass»-ive advantage over detection of nuclear DNA mutations. // *Cancer research*. 2001. № 4 (61). С. 1299–304.
186. Jones S. [и др.]. Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses // *Science*. 2008. № 5897 (321). С. 1801–1806.
187. Kanda M. [и др.]. Mutant GNAS detected in duodenal collections of secretin-stimulated pancreatic juice indicates the presence or emergence of pancreatic cysts // *Gut*. 2013. № 7 (62). С. 1024–1033.
188. Kastrinos F. Risk of Pancreatic Cancer in Families With Lynch Syndrome // *JAMA*. 2009. № 16 (302). С. 1790.
189. Kaufman B. [и др.]. Olaparib Monotherapy in Patients With Advanced Cancer and a Germline BRCA1/2 Mutation // *Journal of Clinical Oncology*. 2015. № 3 (33). С. 244–250.
190. Kaur J. [и др.]. Whole-Exome Sequencing Reveals High Mutational Concordance between Primary and Matched Recurrent Triple-Negative Breast Cancers // *Genes*. 2023. № 9 (14). С. 1690.
191. Kaur R. [и др.]. Metabolomics-Driven Biomarker Discovery for Breast Cancer Prognosis and Diagnosis // *Cells*. 2024. № 1 (14). С. 5.
192. Kekeeva T. V. [и др.]. Loss of heterozygosity and microsatellite instability in tumor-associated stromal cells and tumor epithelium of prostate cancer // *Molecular Biology*. 2008. № 1 (42). С. 86–90.
193. Khalaf D. J. [и др.]. Optimal sequencing of enzalutamide and abiraterone acetate plus prednisone in metastatic castration-resistant prostate cancer: a multicentre, randomised, open-label, phase 2, crossover trial // *The Lancet Oncology*. 2019. № 12 (20). С. 1730–1739.
194. Kimmelman J., Tannock I. The paradox of precision medicine // *Nature Reviews Clinical Oncology*. 2018. № 6 (15). С. 341–342.
195. King M.-C., Marks J. H., Mandell J. B. Breast and Ovarian Cancer Risks Due to Inherited Mutations in BRCA1 and BRCA2 // *Science*. 2003. № 5645 (302). С. 643–646.
196. Klug F. [и др.]. Low-Dose Irradiation Programs Macrophage Differentiation to an iNOS+/M1 Phenotype that Orchestrates Effective T Cell Immunotherapy // *Cancer Cell*. 2013. № 5 (24). С. 589–602.
197. Kremontsov N. Off with your heads: isolated organs in early Soviet science and fiction. // *Studies in history and philosophy of biological and biomedical sciences*. 2009. № 2 (40). С. 87–100.

198. Krishnamachari K. [и др.]. Accurate somatic variant detection using weakly supervised deep learning // *Nature Communications*. 2022. № 1 (13). С. 4248.
199. Krop I. E. [и др.]. Trastuzumab emtansine versus treatment of physician's choice for pretreated HER2-positive advanced breast cancer (TH3RESA): a randomised, open-label, phase 3 trial // *The Lancet Oncology*. 2014. № 7 (15). С. 689–699.
200. Kryukov G. V. [и др.]. MTAP deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells // *Science*. 2016. № 6278 (351). С. 1214–1218.
201. Krzyszczyk P. [и др.]. The growing role of precision and personalized medicine for cancer treatment // *TECHNOLOGY*. 2018. № 03n04 (06). С. 79–100.
202. Kuchenbaecker K. B. [и др.]. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers // *JAMA*. 2017. № 23 (317). С. 2402.
203. Kurian A. W., Antoniou A. C., Domchek S. M. Refining Breast Cancer Risk Stratification: Additional Genes, Additional Information // *American Society of Clinical Oncology Educational Book*. 2016. № 36. С. 44–56.
204. Kurian A. W., Kingham K. E., Ford J. M. Next-generation sequencing for hereditary breast and gynecologic cancer risk assessment // *Current Opinion in Obstetrics & Gynecology*. 2015. № 1 (27). С. 23–33.
205. Kyrochristos I. D., Roukos D. H. Comprehensive intra-individual genomic and transcriptional heterogeneity: Evidence-based Colorectal Cancer Precision Medicine // *Cancer Treatment Reviews*. 2019. (80). С. 101894.
206. Landau D. A. [и др.]. Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia // *Cell*. 2013. № 4 (152). С. 714–726.
207. Larghi A. [и др.]. Screening for Pancreatic Cancer in High-Risk Individuals: A Call for Endoscopic Ultrasound // *Clinical Cancer Research*. 2009. № 6 (15). С. 1907–1914.
208. Latham A. [и др.]. Microsatellite Instability Is Associated With the Presence of Lynch Syndrome Pan-Cancer // *Journal of Clinical Oncology*. 2019. № 4 (37). С. 286–295.
209. Le D. T. [и др.]. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency // *New England Journal of Medicine*. 2015. № 26 (372). С. 2509–2520.
210. Le D. T. [и др.]. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade // *Science*. 2017. № 6349 (357). С. 409–413.
211. Lehmann B. D. [и др.]. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies // *Journal of Clinical Investigation*. 2011. № 7 (121). С. 2750–2767.
212. Lehmann B. D., Pietenpol J. A. Identification and use of biomarkers in treatment strategies for triple-negative breast cancer subtypes // *The Journal of Pathology*. 2014. № 2 (232). С. 142–150.

213. Lennon A. M. [и др.]. The Early Detection of Pancreatic Cancer: What Will It Take to Diagnose and Treat Curable Pancreatic Neoplasia? // *Cancer Research*. 2014. № 13 (74). С. 3381–3389.
214. León A., Pastor Ó. Enhancing Precision Medicine: A Big Data-Driven Approach for the Management of Genomic Data // *Big Data Research*. 2021. (26). С. 100253.
215. Li M. M. [и др.]. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer // *The Journal of Molecular Diagnostics*. 2017. № 1 (19). С. 4–23.
216. Li X., Heyer W.-D. Homologous recombination in DNA repair and DNA damage tolerance // *Cell Research*. 2008. № 1 (18). С. 99–113.
217. Lih C.-J. [и др.]. Analytical Validation of the Next-Generation Sequencing Assay for a Nationwide Signal-Finding Clinical Trial // *The Journal of Molecular Diagnostics*. 2017. № 2 (19). С. 313–327.
218. Lin N. U. [и др.]. Sites of distant recurrence and clinical outcomes in patients with metastatic triple-negative breast cancer // *Cancer*. 2008. № 10 (113). С. 2638–2645.
219. Lin S. [и др.]. Synthesis of novel dual target inhibitors of PARP and HSP90 and their antitumor activities // *Bioorganic & Medicinal Chemistry*. 2020. № 9 (28). С. 115434.
220. Liu Y.-R. [и др.]. Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer // *Breast Cancer Research*. 2016. № 1 (18). С. 33.
221. Livasy C. A. [и др.]. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma // *Modern Pathology*. 2006. № 2 (19). С. 264–271.
222. Lord C. J., Ashworth A. BRCAness revisited // *Nature Reviews Cancer*. 2016. № 2 (16). С. 110–120.
223. Lowery M. A. [и др.]. An Emerging Entity: Pancreatic Adenocarcinoma Associated with a Known BRCA Mutation: Clinical Descriptors, Treatment Implications, and Future Directions // *The Oncologist*. 2011. № 10 (16). С. 1397–1402.
224. Lowery M. A. Genotype–phenotype correlation in BRCA1/2 mutation-associated pancreatic cancer // *British Journal of Cancer*. 2020. № 3 (122). С. 293–294.
225. Luchini C. [и др.]. Pancreatic undifferentiated carcinoma with osteoclast-like giant cells is genetically similar to, but clinically distinct from, conventional ductal adenocarcinoma // *The Journal of Pathology*. 2017. № 2 (243). С. 148–154.
226. Mainiero M. B. [и др.]. American College of Radiology ACR Appropriateness Criteria: Breast cancer screening. // *Journal of the American College of Radiology*. 2013. № 1 (10). С. 11–14.
227. Maitra A. [и др.]. The Human MitoChip: A High-Throughput Sequencing Microarray for Mitochondrial Mutation Detection // *Genome Research*. 2004. № 5 (14). С. 812–819.
228. Maley C. C. [и др.]. Genetic clonal diversity predicts progression to esophageal adenocarcinoma

// Nature Genetics. 2006. № 4 (38). С. 468–473.

229. Marabelle A. [и др.]. Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair–Deficient Cancer: Results From the Phase II KEYNOTE-158 Study // Journal of Clinical Oncology. 2020. № 1 (38). С. 1–10.
230. Marriott H. [и др.]. DNAscan2: a versatile, scalable, and user-friendly analysis pipeline for human next-generation sequencing data // Bioinformatics. 2023. № 4 (39).
231. Masuda H. [и др.]. Differential Response to Neoadjuvant Chemotherapy Among 7 Triple-Negative Breast Cancer Molecular Subtypes // Clinical Cancer Research. 2013. № 19 (19). С. 5533–5540.
232. Mavaddat N. [и др.]. Pathology of Breast and Ovarian Cancers among BRCA1 and BRCA2 Mutation Carriers: Results from the Consortium of Investigators of Modifiers of BRCA1 / 2 (CIMBA) // Cancer Epidemiology, Biomarkers & Prevention. 2012. № 1 (21). С. 134–147.
233. Mavrakis K. J. [и др.]. Disordered methionine metabolism in MTAP/CDKN2A-deleted cancers leads to dependence on PRMT5 // Science. 2016. № 6278 (351). С. 1208–1213.
234. Merlo L. M. F. [и др.]. A Comprehensive Survey of Clonal Diversity Measures in Barrett’s Esophagus as Biomarkers of Progression to Esophageal Adenocarcinoma // Cancer Prevention Research. 2010. № 11 (3). С. 1388–1397.
235. Mertins P. [и др.]. Proteogenomics connects somatic mutations to signalling in breast cancer // Nature. 2016. № 7605 (534). С. 55–62.
236. Middha S. [и др.]. Reliable Pan-Cancer Microsatellite Instability Assessment by Using Targeted Next-Generation Sequencing Data // JCO Precision Oncology. 2017. № 1. С. 1–17.
237. Miki Y. [и др.]. A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1 // Science. 1994. № 5182 (266). С. 66–71.
238. Milbury C. A. [и др.]. Clinical and analytical validation of FoundationOne®CDx, a comprehensive genomic profiling assay for solid tumors. // PloS one. 2022. № 3 (17). С. e0264138.
239. Miles D. [и др.]. Bevacizumab plus paclitaxel versus placebo plus paclitaxel as first-line therapy for HER2-negative metastatic breast cancer (MERiDiAN): A double-blind placebo-controlled randomised phase III trial with prospective biomarker evaluation // European Journal of Cancer. 2017. (70). С. 146–155.
240. Miles D. [и др.]. Primary results from IMpassion131, a double-blind, placebo-controlled, randomised phase III trial of first-line paclitaxel with or without atezolizumab for unresectable locally advanced/metastatic triple-negative breast cancer // Annals of Oncology. 2021. № 8 (32). С. 994–1004.
241. Mina A., Yoder R., Sharma P. Targeting the androgen receptor in triple-negative breast cancer: current perspectives // OncoTargets and Therapy. 2017. (Volume 10). С. 4675–4685.

242. Mir M. A. [и др.]. Targeting Different Pathways Using Novel Combination Therapy in Triple Negative Breast Cancer // *Current Cancer Drug Targets*. 2020. № 8 (20). С. 586–602.
243. Misra B. B. [и др.]. Integrated omics: tools, advances and future approaches // *Journal of Molecular Endocrinology*. 2019. № 1 (62). С. R21–R45.
244. Moffitt R. A. [и др.]. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma // *Nature Genetics*. 2015. № 10 (47). С. 1168–1178.
245. Mohamad A., Ismail E.-F. Detection and Classification of Skin Cancer Using Deep Convolutional Neural Networks (CNN) via KNIME Analytics Platform Software // *Surman Journal of Science and Technology*. 2024. № 1 (6). С. 054–086.
246. Morris G. J. [и др.]. Differences in breast carcinoma characteristics in newly diagnosed African–American and Caucasian patients // *Cancer*. 2007. № 4 (110). С. 876–884.
247. Moskaluk C. A., Hruban R. H., Kern S. E. p16 and K-ras gene mutations in the intraductal precursors of human pancreatic adenocarcinoma. // *Cancer research*. 1997. № 11 (57). С. 2140–3.
248. Muhiyaddin R. [и др.]. The Impact of Clinical Decision Support Systems (CDSS) on Physicians: A Scoping Review. // *Studies in health technology and informatics*. 2020. (272). С. 470–473.
249. Murphy S. J. [и др.]. Genetic Alterations Associated With Progression From Pancreatic Intraepithelial Neoplasia to Invasive Pancreatic Tumor // *Gastroenterology*. 2013. № 5 (145). С. 1098–1109.e1.
250. Musolino A. [и др.]. BRCA mutations, molecular markers, and clinical variables in early-onset breast cancer: A population-based study // *The Breast*. 2007. № 3 (16). С. 280–292.
251. Nakata B. [и др.]. Prognostic value of microsatellite instability in resectable pancreatic cancer. // *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2002. № 8 (8). С. 2536–40.
252. Nakken S. MSI classification from somatic mutation profiles [Электронный ресурс]. URL: https://rpubs.com/sigven/msi_classification_v3.
253. Nanda R. [и др.]. Pembrolizumab plus standard neoadjuvant therapy for high-risk breast cancer (BC): Results from I-SPY 2. // *Journal of Clinical Oncology*. 2017. № 15_suppl (35). С. 506–506.
254. Nazarian R. [и др.]. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation // *Nature*. 2010. № 7326 (468). С. 973–977.
255. Nichols H. B. [и др.]. Declining Incidence of Contralateral Breast Cancer in the United States From 1975 to 2006 // *Journal of Clinical Oncology*. 2011. № 12 (29). С. 1564–1569.
256. Nielsen T. O. [и др.]. Immunohistochemical and Clinical Characterization of the Basal-Like Subtype of Invasive Breast Carcinoma // *Clinical Cancer Research*. 2004. № 16 (10). С. 5367–5374.
257. Nowell P. C. The Clonal Evolution of Tumor Cell Populations // *Science*. 1976. № 4260 (194). С. 23–28.

258. O'Reilly E. M. [и др.]. Durvalumab With or Without Tremelimumab for Patients With Metastatic Pancreatic Ductal Adenocarcinoma // *JAMA Oncology*. 2019. № 10 (5). С. 1431.
259. Oh J.-H. [и др.]. Spontaneous mutations in the single TTN gene represent high tumor mutation burden // *npj Genomic Medicine*. 2020. № 1 (5). С. 33.
260. Okamura R. [и др.]. Analysis of NTRK Alterations in Pan-Cancer Adult and Pediatric Malignancies: Implications for NTRK-Targeted Therapeutics // *JCO Precision Oncology*. 2018. № 2. С. 1–20.
261. Omura N. [и др.]. Genome-wide profiling at methylated promoters in pancreatic adenocarcinoma // *Cancer Biology & Therapy*. 2008. № 7 (7). С. 1146–1156.
262. Orsi G. [и др.]. Chemotherapy toxicity and activity in patients with pancreatic ductal adenocarcinoma and germline BRCA1-2 pathogenic variants (gBRCA1-2pv): a multicenter survey // *ESMO Open*. 2021. № 5 (6). С. 100238.
263. Osheroff J. A. Improving Medication Use and Outcomes with Clinical Decision Support: A Step by Step Guide. / J. A. Osheroff, HIMSS Publ-е изд., 2009.
264. Ott P. A. [и др.]. T-Cell–Inflamed Gene-Expression Profile, Programmed Death Ligand 1 Expression, and Tumor Mutational Burden Predict Efficacy in Patients Treated With Pembrolizumab Across 20 Cancers: KEYNOTE-028 // *Journal of Clinical Oncology*. 2019. № 4 (37). С. 318–327.
265. Owens D. K. [и др.]. Screening for Pancreatic Cancer // *JAMA*. 2019. № 5 (322). С. 438.
266. Ozanne E. M. [и др.]. Which Risk Model to Use? Clinical Implications of the ACS MRI Screening Guidelines // *Cancer Epidemiology, Biomarkers & Prevention*. 2013. № 1 (22). С. 146–149.
267. Paez J. G. [и др.]. EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy // *Science*. 2004. № 5676 (304). С. 1497–1500.
268. Page M. J. [и др.]. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews // *BMJ*. 2021. С. n71.
269. Pao W., Chmielecki J. Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer // *Nature Reviews Cancer*. 2010. № 11 (10). С. 760–774.
270. Papanastasiou G. [и др.]. Large-scale deep learning analysis to identify adult patients at risk for combined and common variable immunodeficiencies // *Communications Medicine*. 2023. № 1 (3). С. 189.
271. Park C. [и др.]. Abstract 2805: Pharmacogenomic platform using patient-derived cells to delineate tumor heterogeneity and therapeutic response of refractory lung cancer // *Cancer Research*. 2023. № 7_Supplement (83). С. 2805–2805.
272. Park W. [и др.]. Genomic Methods Identify Homologous Recombination Deficiency in Pancreas Adenocarcinoma and Optimize Treatment Selection // *Clinical Cancer Research*. 2020. № 13 (26). С. 3239–3247.

273. Parobek C. M. [и др.]. Clinical exome sequencing uncovers genetic disorders in neonates with suspected hypoxic–ischemic encephalopathy: A retrospective analysis // *Clinical Genetics*. 2024. № 1 (106). С. 95–101.
274. Parsi M. A. [и др.]. DNA Methylation Alterations In Endoscopic Retrograde Cholangiopancreatography Brush Samples of Patients With Suspected Pancreaticobiliary Disease // *Clinical Gastroenterology and Hepatology*. 2008. № 11 (6). С. 1270–1278.
275. Pearce O. M. T. [и др.]. Deconstruction of a Metastatic Tumor Microenvironment Reveals a Common Matrix Response in Human Cancers // *Cancer Discovery*. 2018. № 3 (8). С. 304–319.
276. Perakis S. O. [и др.]. Comparison of three commercial decision support platforms for matching of next-generation sequencing results with therapies in patients with cancer. // *ESMO open*. 2020. № 5 (5). С. e000872.
277. Pereira R., Oliveira J., Sousa M. Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics // *Journal of Clinical Medicine*. 2020. № 1 (9). С. 132.
278. Perou C. M. [и др.]. Molecular portraits of human breast tumours // *Nature*. 2000. № 6797 (406). С. 747–752.
279. Phallen J. [и др.]. Direct detection of early-stage cancers using circulating tumor DNA // *Science Translational Medicine*. 2017. № 403 (9).
280. Phillips E. J. [и др.]. Clinical Pharmacogenetics Implementation Consortium Guideline for HLA Genotype and Use of Carbamazepine and Oxcarbazepine: 2017 Update // *Clinical Pharmacology & Therapeutics*. 2018. № 4 (103). С. 574–581.
281. Pishvaian M. J. [и др.]. Overall survival in patients with pancreatic cancer receiving matched therapies following molecular profiling: a retrospective analysis of the Know Your Tumor registry trial // *The Lancet Oncology*. 2020. № 4 (21). С. 508–518.
282. Poplin R. [и др.]. A universal SNP and small-indel variant caller using deep neural networks // *Nature Biotechnology*. 2018. № 10 (36). С. 983–987.
283. Prat A. [и др.]. Clinical implications of the intrinsic molecular subtypes of breast cancer // *The Breast*. 2015. (24). С. S26–S35.
284. Ramdzan A. R. [и др.]. Cost-Effectiveness of Colorectal Cancer Genetic Testing // *International Journal of Environmental Research and Public Health*. 2021. № 16 (18). С. 8330.
285. Réda M. [и др.]. Implementation and use of whole exome sequencing for metastatic solid cancer // *EBioMedicine*. 2020. (51). С. 102624.
286. Reiner A. S. [и др.]. Breast Cancer Family History and Contralateral Breast Cancer Risk in Young Women: An Update From the Women’s Environmental Cancer and Radiation Epidemiology Study // *Journal of Clinical Oncology*. 2018. № 15 (36). С. 1513–1520.
287. Reis-Filho J. S., Tutt A. N. J. Triple negative tumours: a critical review // *Histopathology*. 2008.

№ 1 (52). С. 108–118.

288. Ribic C. M. [и др.]. Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer // *New England Journal of Medicine*. 2003. № 3 (349). С. 247–257.

289. Robert C. [и др.]. Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): post-hoc 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study // *The Lancet Oncology*. 2019. № 9 (20). С. 1239–1251.

290. Roberts N. J. [и др.]. Whole Genome Sequencing Defines the Genetic Heterogeneity of Familial Pancreatic Cancer // *Cancer Discovery*. 2016. № 2 (6). С. 166–175.

291. Robertson L. [и др.]. BRCA1 testing should be offered to individuals with triple-negative breast cancer diagnosed below 50 years // *British Journal of Cancer*. 2012. № 6 (106). С. 1234–1238.

292. Robinson S. [и др.]. Differences in Molecular Pathologic Characteristics of Pancreatic Adenocarcinoma between Egyptian and Moroccan Patients. // *Middle East journal of cancer*. 2010. № 1 (1). С. 27–36.

293. Robson M. [и др.]. Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation // *New England Journal of Medicine*. 2017. № 6 (377). С. 523–533.

294. Roden D. M. [и др.]. Pharmacogenomics // *The Lancet*. 2019. № 10197 (394). С. 521–532.

295. Rodon J. [и др.]. Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial. // *Nature medicine*. 2019. № 5 (25). С. 751–758.

296. Roylance R. [и др.]. Relationship of Extreme Chromosomal Instability with Long-term Survival in a Retrospective Analysis of Primary Breast Cancer // *Cancer Epidemiology, Biomarkers & Prevention*. 2011. № 10 (20). С. 2183–2194.

297. Ryu J. K. [и др.]. Aberrant MicroRNA-155 Expression Is an Early Event in the Multistep Progression of Pancreatic Adenocarcinoma // *Pancreatology*. 2010. № 1 (10). С. 66–73.

298. Sá P. H. C. G. de [и др.]. Next-Generation Sequencing and Data Analysis Elsevier, 2018. С. 191–207.

299. Salameh J.-P. [и др.]. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist // *BMJ*. 2020. С. m2632.

300. Sarnecka A. K. [и др.]. DNA extraction from FFPE tissue samples – a comparison of three procedures // *Współczesna Onkologia*. 2019. № 1 (23). С. 52–58.

301. Saslow D. [и др.]. American Cancer Society Guidelines for Breast Screening with MRI as an Adjunct to Mammography // *CA: A Cancer Journal for Clinicians*. 2007. № 2 (57). С. 75–89.

302. Sato N. [и др.]. STK11/LKB1 Peutz-Jeghers Gene Inactivation in Intraductal Papillary-Mucinous Neoplasms of the Pancreas // *The American Journal of Pathology*. 2001. № 6 (159). С. 2017–2022.

303. Sato N. [и др.]. Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays. // *Cancer research*. 2003. № 13 (63). С. 3735–42.
304. Sato N. [и др.]. Frequent hypomethylation of multiple genes overexpressed in pancreatic ductal adenocarcinoma. // *Cancer research*. 2003. № 14 (63). С. 4158–66.
305. Savard J. Personalised Medicine: A Critique on the Future of Health Care // *Journal of Bioethical Inquiry*. 2013. № 2 (10). С. 197–203.
306. Schmid P. [и др.]. Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer // *New England Journal of Medicine*. 2018. № 22 (379). С. 2108–2121.
307. Schutte M. [и др.]. Abrogation of the Rb/p16 tumor-suppressive pathway in virtually all pancreatic carcinomas. // *Cancer research*. 1997. № 15 (57). С. 3126–30.
308. Schwaederle M. [и др.]. Impact of Precision Medicine in Diverse Cancers: A Meta-Analysis of Phase II Clinical Trials // *Journal of Clinical Oncology*. 2015. № 32 (33). С. 3817–3825.
309. Sedgwick P. Meta-analyses: what is heterogeneity? // *BMJ*. 2015. № mar16 1 (350). С. h1435–h1435.
310. Seo H. [и др.]. SYNERGxDB: an integrative pharmacogenomic portal to identify synergistic drug combinations for precision oncology // *Nucleic Acids Research*. 2020. № W1 (48). С. W494–W501.
311. Shaffer S. M. [и др.]. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance // *Nature*. 2017. № 7658 (546). С. 431–435.
312. Sharma P. [и др.]. Germline BRCA mutation evaluation in a prospective triple-negative breast cancer registry: implications for hereditary breast and/or ovarian cancer syndrome testing // *Breast Cancer Research and Treatment*. 2014. № 3 (145). С. 707–714.
313. Sharma P. Biology and Management of Patients With Triple-Negative Breast Cancer // *The Oncologist*. 2016. № 9 (21). С. 1050–1062.
314. Shatalov P. A. [и др.]. Genetic and therapeutic landscapes in cohort of pancreatic adenocarcinomas: next-generation sequencing and machine learning for full tumor exome analysis // *Oncotarget*. 2024. № 1 (15). С. 91–103.
315. Shatalov P. A. [и др.]. Adaptation of PCR-based library preparation for MGI platform for cancer mutation testing in clinical setting. // *PloS one*. 2025. № 6 (20). С. e0323685.
316. Shegai P. V. [и др.]. Challenges Faced by Clinicians in the Personalized Treatment Planning: A Literature Review and the First Results of the Russian National Cancer Program // *Critical Care Research and Practice*. 2021. (2021). С. 1–7.
317. Shegay P. V. [и др.]. Evolutionary View on Lactate-Dependent Mechanisms of Maintaining Cancer Cell Stemness and Reprimitivization // *Cancers*. 2022. № 19 (14). С. 4552.
318. Shendure J., Ji H. Next-generation DNA sequencing // *Nature Biotechnology*. 2008. № 10 (26). С. 1135–1145.

319. Shi C., Daniels J. A., Hruban R. H. Molecular Characterization of Pancreatic Neoplasms // *Advances in Anatomic Pathology*. 2008. № 4 (15). С. 185–195.
320. Shoaib M. [и др.]. IPCT: Integrated Pharmacogenomic Platform of Human Cancer Cell Lines and Tissues // *Genes*. 2019. № 2 (10). С. 171.
321. Sicklick J. K. [и др.]. Molecular profiling of cancer patients enables personalized combination therapy: the I-PREDICT study. // *Nature medicine*. 2019. № 5 (25). С. 744–750.
322. Singhi A. D. [и др.]. Recurrent Rearrangements in PRKACA and PRKACB in Intraductal Oncocytic Papillary Neoplasms of the Pancreas and Bile Duct // *Gastroenterology*. 2020. № 3 (158). С. 573–582.e2.
323. Slater E. [и др.]. PALB2 mutations in European familial pancreatic cancer families // *Clinical Genetics*. 2010. № 5 (78). С. 490–494.
324. Smyth L. M. [и др.]. Efficacy and Determinants of Response to HER Kinase Inhibition in HER2 - Mutant Metastatic Breast Cancer // *Cancer Discovery*. 2020. № 2 (10). С. 198–213.
325. Soda M. [и др.]. Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer // *Nature*. 2007. № 7153 (448). С. 561–566.
326. Solodskikh S. A. [и др.]. Targeted sequencing to discover germline variants in the BRCA1 and BRCA2 genes in a Russian population and their association with breast cancer risk // *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2019. (813). С. 51–57.
327. Solomon B. J. [и др.]. First-Line Crizotinib versus Chemotherapy in ALK -Positive Lung Cancer // *New England Journal of Medicine*. 2014. № 23 (371). С. 2167–2177.
328. Solomon J. P. [и др.]. NTRK fusion detection across multiple assays and 33,997 cases: diagnostic implications and pitfalls // *Modern Pathology*. 2020. № 1 (33). С. 38–46.
329. Sondka Z. [и др.]. COSMIC: a curated database of somatic variants and clinical data for cancer // *Nucleic Acids Research*. 2024. № D1 (52). С. D1210–D1217.
330. Sorokin M. [и др.]. Oncobox Bioinformatical Platform for Selecting Potentially Effective Combinations of Target Cancer Drugs Using High-Throughput Gene Expression Data. // *Cancers*. 2018. № 10 (10). С. 365.
331. Springer S. [и др.]. A Combination of Molecular Markers and Clinical Features Improve the Classification of Pancreatic Cysts // *Gastroenterology*. 2015. № 6 (149). С. 1501–1510.
332. Sprouffske K. [и др.]. Cancer in Light of Experimental Evolution // *Current Biology*. 2012. № 17 (22). С. R762–R771.
333. Stewart L. A. [и др.]. Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data // *JAMA*. 2015. № 16 (313). С. 1657.
334. Stires H. [и др.]. Integrated molecular analysis of Tamoxifen-resistant invasive lobular breast cancer cells identifies MAPK and GRM/mGluR signaling as therapeutic vulnerabilities // *Molecular*

and Cellular Endocrinology. 2018. (471). С. 105–117.

335. Stockley T. L. [и др.]. Molecular profiling of advanced solid tumors and patient outcomes with genotype-matched clinical trials: the Princess Margaret IMPACT/COMPACT trial // *Genome Medicine*. 2016. № 1 (8). С. 109.

336. Stoffel E. M. [и др.]. Evaluating Susceptibility to Pancreatic Cancer: ASCO Provisional Clinical Opinion // *Journal of Clinical Oncology*. 2019. № 2 (37). С. 153–164.

337. Stracker T. H. [и др.]. The ATM signaling network in development and disease // *Frontiers in Genetics*. 2013. (4).

338. Subhi A. L. [и др.]. Loss of Methylthioadenosine Phosphorylase and Elevated Ornithine Decarboxylase Is Common in Pancreatic Cancer // *Clinical Cancer Research*. 2004. № 21 (10). С. 7290–7296.

339. Sullivan R. J., Flaherty K. T. Resistance to BRAF-targeted therapy in melanoma // *European Journal of Cancer*. 2013. № 6 (49). С. 1297–1304.

340. Supernat A. [и др.]. Comparison of three variant callers for human whole genome sequencing // *Scientific Reports*. 2018. № 1 (8). С. 17851.

341. Sutton R. T. [и др.]. An overview of clinical decision support systems: benefits, risks, and strategies for success // *npj Digital Medicine*. 2020. № 1 (3). С. 17.

342. Szafranska A. E. [и др.]. MicroRNA expression alterations are linked to tumorigenesis and non-neoplastic processes in pancreatic ductal adenocarcinoma // *Oncogene*. 2007. № 30 (26). С. 4442–4452.

343. Talevich E. [и др.]. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing // *PLOS Computational Biology*. 2016. № 4 (12). С. e1004873.

344. Tanaka Y. [и др.]. Frequent beta-catenin mutation and cytoplasmic/nuclear accumulation in pancreatic solid-pseudopapillary neoplasm. // *Cancer research*. 2001. № 23 (61). С. 8401–4.

345. Telli M. L. [и др.]. Asian ethnicity and breast cancer subtypes: a study from the California Cancer Registry // *Breast Cancer Research and Treatment*. 2011. № 2 (127). С. 471–478.

346. Tentori L., Graziani G. Chemopotential by PARP inhibitors in cancer therapy // *Pharmacological Research*. 2005. № 1 (52). С. 25–33.

347. The Cancer Genome Atlas Network Comprehensive molecular portraits of human breast tumours // *Nature*. 2012. № 7418 (490). С. 61–70.

348. Thompson K. J. [и др.]. Luminal androgen receptor breast cancer subtype and investigation of the microenvironment and neoadjuvant chemotherapy response // *NAR Cancer*. 2022. № 2 (4).

349. Timiliotis J. [и др.]. A Novel Diagnostic Decision Support System for Medical Professionals: Prospective Feasibility Study // *JMIR Formative Research*. 2022. № 3 (6). С. e29943.

350. Tollis M., Schiffman J. D., Boddy A. M. Evolution of cancer suppression as revealed by

- mammalian comparative genomics // *Current Opinion in Genetics & Development*. 2017. (42). С. 40–47.
351. Traina T. A. [и др.]. Enzalutamide for the Treatment of Androgen Receptor–Expressing Triple-Negative Breast Cancer // *Journal of Clinical Oncology*. 2018. № 9 (36). С. 884–890.
352. Trivers K. F. [и др.]. The epidemiology of triple-negative breast cancer, including race // *Cancer Causes & Control*. 2009. № 7 (20). С. 1071–1082.
353. Tsimberidou A.-M. Targeted therapy in cancer // *Cancer Chemotherapy and Pharmacology*. 2015. № 6 (76). С. 1113–1132.
354. Tung N. [и др.]. Counselling framework for moderate-penetrance cancer-susceptibility mutations // *Nature Reviews Clinical Oncology*. 2016. № 9 (13). С. 581–588.
355. Tung N. M. [и др.]. TBCRC 048: Phase II Study of Olaparib for Metastatic Breast Cancer and Mutations in Homologous Recombination-Related Genes // *Journal of Clinical Oncology*. 2020. № 36 (38). С. 4274–4282.
356. Tutt A. [и др.]. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial // *The Lancet*. 2010. № 9737 (376). С. 235–244.
357. Tutt A. [и др.]. Carboplatin in BRCA1/2-mutated and triple-negative breast cancer BRCAness subgroups: the TNT Trial // *Nature Medicine*. 2018. № 5 (24). С. 628–637.
358. Verma S. [и др.]. Trastuzumab Emtansine for HER2-Positive Advanced Breast Cancer // *New England Journal of Medicine*. 2012. № 19 (367). С. 1783–1791.
359. Villarroel M. C. [и др.]. Personalizing Cancer Treatment in the Age of Global Genomic Analyses: PALB2 Gene Mutations and the Response to DNA Damaging Agents in Pancreatic Cancer // *Molecular Cancer Therapeutics*. 2011. № 1 (10). С. 3–8.
360. Vo T. T. [и др.]. Pharmacogenomics Implementation: Considerations for Selecting a Reference Laboratory // *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*. 2017. № 9 (37). С. 1014–1022.
361. Waddell N. N. [и др.]. Whole genomes redefine the mutational landscape of pancreatic cancer // *Nature*. 2015. № 7540 (518). С. 495–501.
362. Wang H. [и др.]. A case report of a dramatic response to olaparib in a patient with metastatic pancreatic cancer harboring a germline BRCA2 mutation // *Medicine*. 2019. № 40 (98). С. e17443.
363. Wang Z., Jensen M. A., Zenklusen J. C. A Practical Guide to The Cancer Genome Atlas (TCGA) 2016. С. 111–141.
364. Warton K. [и др.]. Comparison of 4 commercial kits for the extraction of circulating DNA from plasma // *Cancer Genetics*. 2018. (228–229). С. 143–150.
365. Waters A. M., Der C. J. KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer

// Cold Spring Harbor Perspectives in Medicine. 2018. № 9 (8). C. a031435.

366. Watkins J. A. [и др.]. Genomic scars as biomarkers of homologous recombination deficiency and drug response in breast and ovarian cancers // *Breast Cancer Research*. 2014. № 3 (16). C. 211.
367. Weidenbusch B. [и др.]. Transcriptome based individualized therapy of refractory pediatric sarcomas: feasibility, tolerability and efficacy // *Oncotarget*. 2018. № 29 (9). C. 20747–20760.
368. Weinstein J. N. [и др.]. The Cancer Genome Atlas Pan-Cancer analysis project // *Nature Genetics*. 2013. № 10 (45). C. 1113–1120.
369. Weischer M. [и др.]. CHEK2 *1100delC Heterozygosity in Women With Breast Cancer Associated With Early Death, Breast Cancer–Specific Death, and Increased Risk of a Second Breast Cancer // *Journal of Clinical Oncology*. 2012. № 35 (30). C. 4308–4316.
370. Welsh S. J. [и др.]. Resistance to combination BRAF and MEK inhibition in metastatic melanoma: Where to next? // *European Journal of Cancer*. 2016. (62). C. 76–85.
371. Westra W. H. [и др.]. K-ras Oncogene Mutations in Osteoclast-like Giant Cell Tumors of the Pancreas and Liver // *The American Journal of Surgical Pathology*. 1998. № 10 (22). C. 1247–1254.
372. Wilcken N. *Treating Metastatic Breast Cancer: The Evidence for Targeted Therapy* под ред. D. Tovey, Chichester, UK: John Wiley & Sons, Ltd, 2014.
373. Wilentz R. E. [и др.]. Loss of expression of Dpc4 in pancreatic intraepithelial neoplasia: evidence that DPC4 inactivation occurs late in neoplastic progression. // *Cancer research*. 2000. № 7 (60). C. 2002–6.
374. Wilentz R. E. [и др.]. Genetic, Immunohistochemical, and Clinical Features of Medullary Carcinoma of the Pancreas // *The American Journal of Pathology*. 2000. № 5 (156). C. 1641–1651.
375. Williamson C. T. [и др.]. Enhanced cytotoxicity of PARP inhibition in mantle cell lymphoma harbouring mutations in both ATM and p53 // *EMBO Molecular Medicine*. 2012. № 6 (4). C. 515–527.
376. Win A. K. [и др.]. Colorectal and Other Cancer Risks for Carriers and Noncarriers From Families With a DNA Mismatch Repair Gene Mutation: A Prospective Cohort Study // *Journal of Clinical Oncology*. 2012. № 9 (30). C. 958–964.
377. Winter J. M. [и др.]. Absence of E-Cadherin Expression Distinguishes Noncohesive from Cohesive Pancreatic Cancer // *Clinical Cancer Research*. 2008. № 2 (14). C. 412–418.
378. Wooster R. [и др.]. Localization of a Breast Cancer Susceptibility Gene, BRCA2, to Chromosome 13q12-13 // *Science*. 1994. № 5181 (265). C. 2088–2090.
379. Wu G. [и др.]. Significance of TP53 mutation in bladder cancer disease progression and drug selection // *PeerJ*. 2019. (7). C. e8261.
380. Wu J. [и др.]. Recurrent GNAS Mutations Define an Unexpected Pathway for Pancreatic Cyst Development // *Science Translational Medicine*. 2011. № 92 (3).

381. Wu J. [и др.]. Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways // *Proceedings of the National Academy of Sciences*. 2011. № 52 (108). С. 21188–21193.
382. Xia X. [и др.]. Bile acid interactions with cholangiocytes. // *World journal of gastroenterology*. 2006. № 22 (12). С. 3553–63.
383. Yalamanchili H. K., Wan Y., Liu Z. Data Analysis Pipeline for RNA-seq Experiments: From Differential Expression to Cryptic Splicing // *Current Protocols in Bioinformatics*. 2017. № 1 (59).
384. Yang X. [и др.]. Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families // *Journal of Clinical Oncology*. 2020. № 7 (38). С. 674–685.
385. Yates L. R. [и др.]. The European Society for Medical Oncology (ESMO) Precision Medicine Glossary // *Annals of Oncology*. 2018. № 1 (29). С. 30–35.
386. Yates L. R., Campbell P. J. Evolution of the cancer genome // *Nature Reviews Genetics*. 2012. № 11 (13). С. 795–806.
387. Yin L. [и др.]. Triple-negative breast cancer molecular subtyping and treatment progress // *Breast Cancer Research*. 2020. № 1 (22). С. 61.
388. Yoo B. C. [и др.]. Clinical multi-omics strategies for the effective cancer management // *Journal of Proteomics*. 2018. (188). С. 97–106.
389. Yu J. [и др.]. Digital next-generation sequencing identifies low-abundance mutations in pancreatic juice samples collected from the duodenum of patients with pancreatic cancer and intraductal papillary mucinous neoplasms // *Gut*. 2017. № 9 (66). С. 1677–1687.
390. Yu N. [и др.]. Patient-derived cell-based pharmacogenomic assessment to unveil underlying resistance mechanisms and novel therapeutics for advanced lung cancer // *Journal of Experimental & Clinical Cancer Research*. 2023. № 1 (42). С. 37.
391. Zabolotneva A. A. [и др.]. A systematic experimental evaluation of microRNA markers of human bladder cancer // *Frontiers in Genetics*. 2013. (4).
392. Zazo S. [и др.]. Actionable mutational profiling in solid tumors using hybrid-capture-based next-generation sequencing in a real-world setting in Spain // *Cancer Medicine*. 2024. № 3 (13).
393. Zhang J. [и др.]. Characterization of cancer genomic heterogeneity by next-generation sequencing advances precision medicine in cancer treatment // *Precision Clinical Medicine*. 2018. № 1 (1). С. 29–48.
394. Zhang L. [и др.]. Androgen Receptor, EGFR, and BRCA1 as Biomarkers in Triple-Negative Breast Cancer: A Meta-Analysis // *BioMed Research International*. 2015. (2015). С. 1–12.
395. Zhao S. [и др.]. Accuracy and efficiency of germline variant calling pipelines for human genome data // *Scientific Reports*. 2020. № 1 (10). С. 20222.
396. Zhao Y. [и др.]. Empirical comparisons of heterogeneity magnitudes of the risk difference,

- relative risk, and odds ratio // *Systematic Reviews*. 2022. № 1 (11). С. 26.
397. Zhen D. B. [и др.]. BRCA1, BRCA2, PALB2, and CDKN2A mutations in familial pancreatic cancer: a PACGENE study // *Genetics in Medicine*. 2015. № 7 (17). С. 569–577.
398. Zhen Y. [и др.]. Protective effect of bicyclol against bile duct ligation-induced hepatic fibrosis in rats. // *World journal of gastroenterology*. 2015. № 23 (21). С. 7155–64.
399. Zhu L. [и др.]. Current status of biomarker and targeted nanoparticle development: The precision oncology approach for pancreatic cancer therapy // *Cancer Letters*. 2017. (388). С. 139–148.
400. Zolotovskaia M. A. [и др.]. Disparity between Inter-Patient Molecular Heterogeneity and Repertoires of Target Drugs Used for Different Types of Cancer in Clinical Oncology // *International Journal of Molecular Sciences*. 2020. № 5 (21). С. 1580.
401. Zolotovskaia M. A. [и др.]. Pan-cancer antagonistic inhibition pattern of ATM-driven G2/M checkpoint pathway vs other DNA repair pathways // *DNA Repair*. 2023. (123). С. 103448.
402. Zubarik R. [и др.]. Screening for pancreatic cancer in a high-risk population with serum CA 19-9 and targeted EUS: a feasibility study // *Gastrointestinal Endoscopy*. 2011. № 1 (74). С. 87–95.
403. Атлас Соло (Atlas Solo) Комплекс Плюс. [Электронный ресурс]. URL: <https://oncoatlas.ru/therapy/solo-complex/> (дата обращения: 08.11.2024).
404. GATK Best Practices [Электронный ресурс]. URL: <https://gatk.broadinstitute.org/hc/en-us/articles/360035894711-About-the-GATK-Best-Practices>.
405. BBmap Guide [Электронный ресурс]. URL: <https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbmap-guide/>.
406. MI Cancer Seek | Caris Life Sciences [Электронный ресурс]. URL: <https://www.carislifesciences.com/physicians/physician-tests/mi-cancer-seek/> (дата обращения: 28.01.2025).

ПРИЛОЖЕНИЯ

Приложение 1. Клинико-фенотипические сведения о больных трижды-негативным раком молочной железы, содержащиеся в своде

№	Клиническая информация	№	Клиническая информация
1	Дата рождения	44	Биопсия сторожевого лимфатического узла после НАПХТ
2	Возраст	45	Биопсия сторожевого л/у
3	Диагноз	46	Лимфаденэктомия - Дата операции
4	ПМЗНО	47	Лимфаденэктомия отсроч. - pN отсроч.
5	Стадия	48	Лимфаденэктомия отсроч. - Дата операции 2
6	cT cN cM	49	Замена экспандера
7	ycT ycN ycM	50	Эндопротез 2
8	ypT ypN ypM	51	Дата симметризирующей операции
9	pT pN M	52	Симметризирующая операция
10	Локализация T	53	Эндопротез. симметризирующая операция
11	Размер T УЗИ	54	Дата операции
12	Размер T МГ	55	Профилактическая мастэктомия
13	Размер ПЭТ	56	Эндопротез
14	Размер T МРТ	57	Липофилинг
15	Grade	58	Осложнения до ЛТ
16	Ki67%	59	Дата осложнения до ЛТ
17	Her2/neu Fish	60	Операцию по осложнению до ЛТ
18	C/K 5/6%	61	Лучевая терапия на ГС
19	p63	62	Лучевая терапия на МЖ
20	АПХТ	63	Лучевая терапия на РегЗон.
21	C-II НАПХТ	64	Дата ЛТ
22	C-II Токсик ХТ	65	Осложнения после ЛТ
23	Размер pT после операции	66	Дата осложнения после ЛТ
24	pN после операции	67	Операция по осложнению
25	Grade после операции	68	Рецидив
26	Her2/neu после операции	69	Гистологическое исследование рецидива
27	Ki67 после операции	70	Лечение рецидива
28	LV	71	Прогрессирование процесса
29	Pn	72	Дата прогрессирования
30	Лечебный патоморфоз МЖ по Лавниковой	73	Биопсия очага прогресса
31	RCB	74	Гистологическое исследование прогресса

32	Индекс ОП	75	ИГХ прогрессирувание
33	ЛП ЛУ по Лавниковой	76	Ki67% прогресс
34	ЛП CR	77	PDL
35	PR	78	SP142
36	SD	79	PDL 22c3
37	PD	80	Her2/neu прогресс
38	Дата операции 1	81	Лечение прогресс
39	Вид операции	82	Лучевая прогресс
40	Экспандер	83	Менструальная функция
41	Эндопротез	84	Гистологический тип (инвазивная дольковая карцинома, инвазивная протоковая карцинома и т.д.)
42	Сетчатый имплантат	85	Маркер пролиферации KI 67 и другие параметры
43	Органосохранное лечение	86	Ответ на лечение по рекомендациям СППВР

Приложение 2. Метрики качества FASTQ файлов

№	Параметр
1	sequencing
2	before_filtering__total_reads
3	before_filtering__total_bases
4	before_filtering__q20_bases
5	before_filtering__q30_bases
6	before_filtering__q20_rate
7	before_filtering__q30_rate
8	before_filtering__read1_mean_length
9	before_filtering__read2_mean_length
10	before_filtering__gc_content
11	after_filtering__total_reads
12	after_filtering__total_bases
13	after_filtering__q20_bases
14	after_filtering__q30_bases
15	after_filtering__q20_rate
16	after_filtering__q30_rate
17	after_filtering__read1_mean_length
18	after_filtering__read2_mean_length
19	after_filtering__gc_content
20	filtering_result__passed_filter_reads
21	filtering_result__corrected_reads
22	filtering_result__corrected_bases
23	filtering_result__low_quality_reads
24	filtering_result__too_many_n_reads
25	filtering_result__too_short_reads
26	filtering_result__too_long_reads
27	duplication__rate
28	insert_size__peak
29	insert_size__unknown
30	adapter_cutting__adapter_trimmed_reads
31	adapter_cutting__adapter_trimmed_bases
32	adapter_cutting__read1_adapter_sequence
33	adapter_cutting__read2_adapter_sequence

34	read1_before_filtering_total_reads
35	read1_before_filtering_total_bases
36	read1_before_filtering_q20_bases
37	read1_before_filtering_q30_bases
38	read1_before_filtering_total_cycles
39	read2_before_filtering_total_reads
40	read2_before_filtering_total_bases
41	read2_before_filtering_q20_bases
42	read2_before_filtering_q30_bases
43	read2_before_filtering_total_cycles
44	read1_after_filtering_total_reads
45	read1_after_filtering_total_bases
46	read1_after_filtering_q20_bases
47	read1_after_filtering_q30_bases
48	read1_after_filtering_total_cycles
49	read2_after_filtering_total_reads
50	read2_after_filtering_total_bases
51	read2_after_filtering_q20_bases
52	read2_after_filtering_q30_bases
53	read2_after_filtering_total_cycles
54	pipeversion
55	pipetype

Приложение 3. Собираемые метрики качества работы биоинформатического конвейера

№	Тип метрики качества
1	time_stamp
2	id
3	link_to_folder
4	file_name
5	bait_intervals
6	target_intervals
7	input
8	output
9	reference_sequence
10	metric_accumulation_level
11	near_distance
12	minimum_mapping_quality
13	minimum_base_quality
14	clip_overlapping_reads
15	include_indels
16	coverage_cap
17	sample_size
18	allele_fraction
19	verbosity
20	quiet
21	validation_stringency
22	compression_level
23	max_records_in_ram
24	bait_set
25	create_index
26	create_md5_file
27	ga4gh_client_secrets
28	use_jdk_deflater
29	use_jdk_inflater
30	bait_territory
31	bait_design_efficiency
32	on_bait_bases
33	near_bait_bases
34	off_bait_bases

35	pct_selected_bases
36	pct_off_bait
37	on_bait_vs_selected
38	mean_bait_coverage
39	pct_usable_bases_on_bait
40	pct_usable_bases_on_target
41	fold_enrichment
42	hs_library_size
43	hs_penalty_10x
44	hs_penalty_20x
45	hs_penalty_30x
46	hs_penalty_40x
47	hs_penalty_50x
48	hs_penalty_100x
49	target_territory
50	genome_size
51	total_reads
52	pf_reads
53	pf_bases
54	pf_unique_reads
55	pf_uq_reads_aligned
56	pf_bases_aligned
57	pf_uq_bases_aligned
58	on_target_bases
59	pct_pf_reads
60	pct_pf_uq_reads
61	pct_pf_uq_reads_aligned
62	mean_target_coverage
63	median_target_coverage
64	max_target_coverage
65	min_target_coverage
66	zero_cvg_targets_pct
67	pct_exc_dupe
68	pct_exc_adapter
69	pct_exc_mapq
70	pct_exc_baseq
71	pct_exc_overlap

72	pct_exc_off_target
73	fold_80_base_penalty
74	pct_target_bases_1x
75	pct_target_bases_2x
76	pct_target_bases_10x
77	pct_target_bases_20x
78	pct_target_bases_30x
79	pct_target_bases_40x
80	pct_target_bases_50x
81	pct_target_bases_100x
82	pct_target_bases_250x
83	pct_target_bases_500x
84	pct_target_bases_1000x
85	pct_target_bases_2500x
86	pct_target_bases_5000x
87	pct_target_bases_10000x
88	pct_target_bases_25000x
89	pct_target_bases_50000x
90	pct_target_bases_100000x
91	at_dropout
92	gc_dropout
93	het_snp_sensitivity
94	het_snp_q
95	sample
96	library
97	read_group
98	pipetype
99	pipeversion
100	patient_id

Приложение 4. Набор параметров, собирающих информацию о патогенных и вероятно-патогенных вариантах, выявленных в результате работы герминального биоинформатического конвейера

№	Параметр
1	Count of records
2	general_id
3	patient_id
4	Nomer_zapuska_MGI
5	symbol
6	dbsnprsid
7	clinvar_classification
8	hgvs
9	gt
10	#chrom
11	pos
12	pipeversion
13	pipetype
14	id
15	ref
16	alt
17	vaf
18	dp
19	exon
20	cdna_position
21	cds_position
22	protein_position
23	amino_acids
24	codons
25	ncer_percentile
26	gerp_score
27	qua
28	ad
29	filter
30	pl
31	csq
32	vep_all_csq
33	variant_class
34	consequence
35	impact
36	gene

37	feature_type
38	feature
39	ensembl_transcript_id
40	ensembl_gene_id
41	ensembl_protein_id
42	entrez_id
43	uniprot_id
44	uniprot_acc
45	refseq_mrna
46	refseq_peptide
47	principal_isoform_flag
48	transcript_mane_select
49	gencode_tag
50	gencode_transcript_type
51	intogen_driver
52	cancer_predisposition_source
53	dbnsfp
54	biotype
55	intron
56	existing_variation
57	cosmic_mutation_id
58	allele_num
59	strand
60	pick
61	symbol_source
62	hgnc_id
63	canonical
64	ccds
65	ensp
66	swissprot
67	uniparc
68	refseq
69	civic_id
70	transcript_mane_plus_clinical
71	hgvs_offset
72	lof
73	lof_info
74	mutation_hotspot
75	mutation_hotspot_transcript
76	mutation_hotspot_cancertype

77	clinvar_pmid_somatic
78	coding_status
79	exonic_status
80	intron_position
81	exon_position
82	hgvsp_short
83	amino_acid_start
84	amino_acid_end
85	uniprot_feature
86	tumor_suppressor_evidence
87	oncogene_evidence
88	cancer_syndrome_cui
89	cancer_predisposition_moi
90	cancer_predisposition_mod
91	gwas_hit
92	hgvsp
93	effect_predictions
94	flags
95	timestamp
96	none
97	Otkuda_material
98	Data_postupleniia
99	Tip_materiala
100	Chei_patsient
101	Nomer_istorii_bolezni
102	Familiia
103	Imia_Otchestvo
104	Status
105	Pol
106	Komentarii
107	Diagnoz
108	Konsultatsiia
109	Proekt
110	Napravliaiushchii_vrach
111	Oplachennoe_issledovanie
112	Nukleinovaia_kislota
113	Chto_delaetsia_tekhniches
114	Chem_vydelena_NK
115	Kontsentratsiia_NK_ng_mkl
116	Vziato_v_probu_NK_mkl_mak

117	AE_Bufer_v_probu
118	Tip_fragmentatsii
119	Tip_ochistki_posle_fragme
120	Kit_dlia_prigotovleniia_b
121	Adapter
122	Kol_vo_tsiklov_PTsR
123	Kontsentratsiia_bibliotek
124	Dlina_fragmentov_v_biblio
125	Obem_obraztsa_vziaty_v_p
126	Nomer_pula
127	Nomer_pula_v_zapuske
128	Dlina_fragmentov_v_pule
129	Kontsentratsiia_obogashch
130	Obem_pula_vziaty_v_tsirk
131	Obem_LOW_TE_bufera_dlia_d
132	Kit_pri_obogashchenii
133	Kontsentratsiia_pula_posl
134	Kontsentratsiia_pula_posl_1
135	Obem_pula_vziaty_v_DNB_m
136	Obem_LOW_TE_bufera_dlia_d_1
137	Kontsentratsiia_DNB_ng_mk
138	Kto_provodil_analiz
139	Kto_delaet_bioinformatiku
140	Dedlain_polucheniia_rezul
141	Data_otveta
142	Vremia_vypolnenie_analiza
143	Rezultat
144	Diagnoz_gruppirovanyi
145	Tip_raka_PCGR
146	Data_rozhdeniia
147	Nomer_bloka
148	Fraction
149	Vnishnii_ID
150	Data_peredachi_External
151	Nomer_vneshnego_rana

Приложение 5. Колонки в аннотации герминального VCF файла

№	Параметр/колонка в аннотации
1	timestamp
2	QUAL
3	POS
4	ALT
5	FILTER
6	GT
7	GQ
8	DP
9	AD
10	VAF
11	PL
12	CSQ
13	NCER_PERCENTILE
14	GERP_SCORE
15	DBNSFP
16	PCGR_ONCO_XREF
17	NON_CANCER_AC_GLOBAL
18	NON_CANCER_AN_GLOBAL
19	NON_CANCER_AF_GLOBAL
20	NON_CANCER_AC_ASJ
21	NON_CANCER_AN_ASJ
22	NON_CANCER_AF_ASJ
23	NON_CANCER_NHOMALT_ASJ
24	NON_CANCER_AC_NFE
25	NON_CANCER_AN_NFE
26	NON_CANCER_AF_NFE
27	NON_CANCER_NHOMALT_NFE
28	NON_CANCER_AC_SAS
29	NON_CANCER_AF_SAS
30	NON_CANCER_NHOMALT_SAS
31	NON_CANCER_AN_FIN
32	NON_CANCER_AF_FIN
33	NON_CANCER_NHOMALT_FIN
34	NON_CANCER_AC_EAS
35	NON_CANCER_AN_EAS
36	NON_CANCER_NHOMALT_EAS
37	NON_CANCER_AC_AMR
38	NON_CANCER_AN_AMR

39	NON_CANCER_AF_AMR
40	NON_CANCER_NHOMALT_AMR
41	NON_CANCER_AC_AFR
42	NON_CANCER_AN_AFR
43	NON_CANCER_AF_AFR
44	NON_CANCER_NHOMALT_AFR
45	NON_CANCER_AC_OTH
46	NON_CANCER_AN_OTH
47	NON_CANCER_NHOMALT_OTH
48	REGULATORY_ANNOTATION
49	VEP_ALL_CSQ
50	Consequence
51	IMPACT
52	SYMBOL
53	Gene
54	Feature_type
55	Feature
56	ENSEMBL_TRANSCRIPT_ID
57	ENSEMBL_GENE_ID
58	ENSEMBL_PROTEIN_ID
59	ENTREZ_ID
60	UNIPROT_ID
61	UNIPROT_ACC
62	REFSEQ_MRNA
63	REFSEQ_PEPTIDE
64	PRINCIPAL_ISOFORM_FLAG
65	TRANSCRIPT_MANE_SELECT
66	GENCODE_TAG
67	GENCODE_TRANSCRIPT_TYPE
68	INTOGEN_DRIVER
69	CANCER_PREDISPOSITION_SOURCE
70	OPENTARGETS_DISEASE ASSOCS
71	OPENTARGETS_TRACTABILITY_COMPOUND
72	OPENTARGETS_TRACTABILITY_ANTIBODY
73	ESSENTIAL_GENE_CRISPR
74	ESSENTIAL_GENE_CRISPR2
75	PROB_GNOMAD_LOF_INTOLERANT
76	PROB_GNOMAD_LOF_INTOLERANT_HOM
77	PROB_GNOMAD_LOF_TOLERANT_NULL
78	PROB_EXAC_LOF_INTOLERANT

79	PROB_EXAC_LOF_INTOLERANT_HOM
80	PROB_EXAC_LOF_TOLERANT_NULL
81	PROB_EXAC_NONTCGA_LOF_INTOLERANT
82	PROB_EXAC_NONTCGA_LOF_INTOLERANT_HOM
83	PROB_EXAC_NONTCGA_LOF_TOLERANT_NULL
84	BIOTYPE
85	INTRON
86	HGVSc
87	Existing_variation
88	COSMIC_MUTATION_ID
89	DBSNPRSID
90	ALLELE_NUM
91	STRAND
92	PICK
93	VARIANT_CLASS
94	SYMBOL_SOURCE
95	HGNC_ID
96	CANONICAL
97	CCDS
98	ENSP
99	SWISSPROT
100	UNIPARC
101	RefSeq
102	GLOBAL_AF_1KG
103	AFR_AF_1KG
104	AMR_AF_1KG
105	EAS_AF_1KG
106	EUR_AF_1KG
107	SAS_AF_1KG
108	GLOBAL_AF_GNOMAD
109	AFR_AF_GNOMAD
110	AMR_AF_GNOMAD
111	ASJ_AF_GNOMAD
112	EAS_AF_GNOMAD
113	FIN_AF_GNOMAD
114	NFE_AF_GNOMAD
115	OTH_AF_GNOMAD
116	SAS_AF_GNOMAD
117	PHENO, CODING_STATUS
118	EXONIC_STATUS

119	INTRON_POSITION
120	EXON_POSITION
121	HGVSp_short
122	CDS_CHANGE
123	CLINVAR_MSID
124	CLINVAR_PMID
125	CLINVAR_CLNSIG
126	CLINVAR_VARIANT_ORIGIN
127	CLINVAR_CONFLICTED
128	CLINVAR_UMLS_CUI
129	CLINVAR_HGVSP
130	CLINVAR_CLNSIG_SOMATIC
131	CLINVAR_ALLELE_ID
132	CLINVAR_REVIEW_STATUS_STARS
133	CLINVAR_CLASSIFICATION
134	CLINVAR_ENTREZGENE
135	EXON
136	cDNA_position
137	CDS_position
138	Protein_position
139	Amino_acids
140	Codons
141	TREMBL
142	PFAM_DOMAIN
143	NearestExonJB
144	AMINO_ACID_START
145	AMINO_ACID_END
146	UNIPROT_FEATURE
147	TUMOR_SUPPRESSOR_EVIDENCE
148	ONCOGENE_EVIDENCE
149	CANCER_SYNDROME_CUI
150	CANCER_PREDISPOSITION_MOI
151	CANCER_PREDISPOSITION_MOD
152	GWAS_HIT
153	CORUM_ID
154	DBMTS
155	TCGA_DRIVER
156	SIGNALING_PATHWAY
157	CLINVAR_MOLECULAR_EFFECT
158	DISTANCE

159	HGVSp
160	EFFECT_PREDICTIONS
161	DBNSFP_SIFT
162	DBNSFP_MUTATIONTASTER
163	DBNSFP_MUTATIONASSESSOR
164	DBNSFP_FATHMM
165	DBNSFP_FATHMM_MKL
166	DBNSFP_PROVEAN
167	DBNSFP_M_CAP
168	DBNSFP_MUTPRED
169	DBNSFP_GERP
170	DBNSFP_PRIMATEAI
171	DBNSFP_DEOGEN2
172	DBNSFP_LIST_S2
173	DBNSFP_BAYESDEL_ADDAF
174	DBNSFP_ALOFTPRED
175	DBNSFP_META_RNN
176	DBNSFP_SPLICE_SITE_ADA
177	DBNSFP_SPLICE_SITE_RF
178	RMSK_HIT
179	CIVIC_ID_SEGMENT
180	CHEMBL_COMPOUND_ID
181	CANCER_SUSCEPTIBILITY_CUI
182	CLINVAR_UMLS_CUI_SOMATIC
183	OLD_MULTIALLELIC
184	CIVIC_ID
185	TRANSCRIPT_MANE_PLUS_CLINICAL
186	HGVS_OFFSET
187	LoF
188	LoF_info
189	MUTATION_HOTSPOT
190	MUTATION_HOTSPOT_TRANSCRIPT
191	MUTATION_HOTSPOT_CANCERTYPE
192	CLINVAR_P MID_SOMATIC
193	FLAGS
194	sampleid
195	pipeversion
196	#CHROM
197	pipetype
198	ID
199	REF

Приложение 6. Колонки в аннотации парного VCF файла

1	TUMOR_SUPPRESSOR
2	SIMPLEREPEATS_HIT
3	WINMASKER_HIT
4	OPENTARGETS_RANK
5	index
6	ENTREZ_ID
7	GLOBAL_AF_GNOMAD
8	GLOBAL_AF_1KG
9	NCER_PERCENTILE
10	CALL_CONFIDENCE
11	AF_TUMOR
12	DP_CONTROL
13	AF_CONTROL
14	TCGA_PANCANCER_COUNT
15	PUTATIVE_DRIVER_MUTATION
16	ONCOGENE
17	POS
18	ONCOGENE_EVIDENCE
19	TUMOR_SUPPRESSOR_EVIDENCE
20	CONSEQUENCE
21	PROTEIN_CHANGE
22	PROTEIN_DOMAIN
23	CODING_STATUS
24	EXONIC_STATUS
25	CDS_CHANGE
26	HGVSp
27	HGVSc
28	EFFECT_PREDICTIONS
29	MUTATION_HOTSPOT
30	MUTATION_HOTSPOT_TRANSCRIPT
31	MUTATION_HOTSPOT_CANCERTYPE
32	CHASMPLUS_DRIVER
33	CHASMPLUS_TTYPE
34	VEP_ALL_CSQ
35	COSMIC_MUTATION_ID
36	TCGA_FREQUENCY
37	ICGC_PCAWG_OCCURRENCE
38	CHEMBL_COMPOUND_ID
39	CHEMBL_COMPOUND_TERMS

40	CLINVAR
41	CLINVAR_CLNSIG
42	TIER
43	TIER_DESCRIPTION
44	pipetype
45	pipeversion
46	DBSNPRSID
47	CHROM,
48	REF
49	ALT
50	GENOMIC_CHANGE
51	GENOME_VERSION
52	VCF_SAMPLE_ID
53	VARIANT_CLASS
54	SYMBOL
55	GENE_NAME
56	CCDS
57	CANONICAL
58	UNIPROT_ID
59	ENSEMBL_TRANSCRIPT_ID
60	ENSEMBL_GENE_ID
61	REFSEQ_MRNA
62	REFSEQ_PEPTIDE
63	TRANSCRIPT_MANE_SELECT

Приложение 7. Учебно-методическое пособие “Молекулярно-генетические исследования в онкологии: учебно-методическое пособие” (г. Обнинск, 2023 г.)

Учебно-методическое пособие “Молекулярно-генетические исследования в онкологии: учебно-методическое пособие” (г. Обнинск, 2023 г.) издано ФГБУ «НМИЦ радиологии» Минздрава России, 2023. 82 с. Серия «Библиотека врача-онколога».



Глубокоуважаемые коллеги!

Мы подготовили для вас серию публикаций, которые содержат актуальную информацию о новых методических подходах в диагностике и лечении онкологических заболеваний. Эти материалы охватывают все области практической онкологии, включая инструментальную и лабораторную диагностику, хирургическое лечение, химио- и лучевую терапию, иммунотерапию, мониторинг и учет онкологических больных, паллиативную помощь. Кроме того, в серии будут представлены обзоры мировой литературы по наиболее актуальным проблемам онкологии.

Надеемся, что каждый из вас найдет в этих изданиях полезную информацию, которая поможет вам в работе и позволит внедрить новые технологии в повседневную практику.

Главный внештатный специалист-онколог Минздрава России,
Генеральный директор ФГБУ «НМИЦ радиологии»
Минздрава России, академик РАН

Андрей Дмитриевич КАПРИН



Библиотека врача-онколога



Молекулярно-генетические исследования в онкологии

Под редакцией академика РАН А.Д. Каприна

Android



iOs



Материалы издания
доступны на платформах
App Store и Google Play



Обнинск – Москва
2023

Приложение 8. Патенты и свидетельства РИД

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ПАТЕНТ

НА ИЗОБРЕТЕНИЕ
№ 2758079

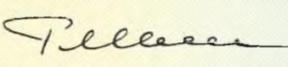
**Способ прогноза прогрессирования заболевания у
больных раком желудочно-кишечного тракта после
проведенного лечения**

Патентообладатель: *Федеральное государственное бюджетное учреждение "Национальный медицинский исследовательский центр радиологии" Министерства здравоохранения Российской Федерации (ФГБУ "НМИЦ радиологии" Минздрава России) (RU)*

Авторы: *Михайлова Галина Фёдоровна (RU), Цепенко Виктория Викторовна (RU), Шкаврова Татьяна Геннадьевна (RU), Черкесов Владимир Николаевич (RU), Шегай Петр Викторович (RU), Иванов Сергей Анатольевич (RU), Каприн Андрей Дмитриевич (RU)*

Заявка № **2021114150**
 Приоритет изобретения **19 мая 2021 г.**
 Дата государственной регистрации
 в Государственном реестре изобретений
 Российской Федерации **26 октября 2021 г.**
 Срок действия исключительного права
 на изобретение истекает **19 мая 2041 г.**

*Руководитель Федеральной службы
по интеллектуальной собственности*


Г.П. Ивлиев



РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2020666699

Программа автоматизированного биоинформатического
анализа геномных данных для подбора
персонализированной терапии

Правообладатель: *Шегай Петр Викторович (RU)*

Авторы: *Каприн Андрей Дмитриевич (RU), Шегай Петр
Викторович (RU), Шаталов Петр Алексеевич (RU), Горбачев
Алексей Юрьевич (RU), Коростин Дмитрий Олегович (RU)*

Заявка № 2020665123

Дата поступления 18 ноября 2020 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 14 декабря 2020 г.



Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2024684693

Программа автоматизированного биоинформатического анализа омиксных данных для подбора противоопухолевой лекарственной терапии, в том числе мРНК вакцин, и других биотехнологических лекарственных препаратов, производимых по индивидуальному назначению на основании результатов генетических исследований

Правообладатель: *Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр радиологии» Министерства здравоохранения Российской Федерации (RU)*

Авторы: *Каприн Андрей Дмитриевич (RU), Шегай Петр Викторович (RU), Шаталов Петр Алексеевич (RU), Траснов Алексей Александрович (RU), Райгородская Мария Павловна (RU), Веселовский Егор Михайлович (RU)*



Заявка № 2024683232

Дата поступления 08 октября 2024 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 21 октября 2024 г.

*Руководитель Федеральной службы
по интеллектуальной собственности*

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 0492e7c1a2300c4542401670bca2026
Владелец **Зубов Юрий Сергеевич**
Действителен с 10.07.2024 по 03.10.2025

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2020620626

**«База данных оборудования и расходных материалов,
использующихся для комплексной молекулярно-генетической
диагностики в онкогинекологии, включая методы
высокопроизводительного секвенирования»**

Правообладатель: **Федеральное государственное бюджетное
учреждение «Национальный медицинский исследовательский
центр радиологии» Министерства здравоохранения Российской
Федерации (ФГБУ «НМИЦ радиологии» Минздрава России) (RU)**

Авторы: **см. на обороте**

Заявка № **2020620436**

Дата поступления **18 марта 2020 г.**

Дата государственной регистрации
в Реестре баз данных **30 марта 2020 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

 **Г.П. Ивлиев**



РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2020622407

База данных, содержащая информацию о фармакогеномных подходах при различных вариантах опухолей различной этиологии, для персонализации терапии

Правообладатель: *Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр радиологии» Министерства здравоохранения Российской Федерации (ФГБУ «НМИЦ радиологии» Минздрава России) (RU)*

Авторы: *см. на обороте*

Заявка № 2020622289

Дата поступления 18 ноября 2020 г.

Дата государственной регистрации

в Реестре баз данных 26 ноября 2020 г.

Руководитель Федеральной службы
по интеллектуальной собственности

 Г.П. Ивлиев



РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2020620919

**«База данных оборудования и расходных материалов,
использующихся для комплексной
молекулярно-генетической диагностики в гематологии,
включая методы высокопроизводительного
секвенирования»**

Правообладатель: *Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр радиологии» Министерства здравоохранения Российской Федерации (ФГБУ «НМИЦ радиологии» Минздрава России) (RU)*

Авторы: *см. на обороте*

Заявка № **2020620439**

Дата поступления **18 марта 2020 г.**

Дата государственной регистрации

в Реестре баз данных **04 июня 2020 г.**

*Руководитель Федеральной службы
по интеллектуальной собственности*

 **Г.П. Ильев**



РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2020620920

**«База данных оборудования и расходных материалов,
использующихся для комплексной молекулярно-генетической
диагностики в нейроонкологии, включая методы
высокопроизводительного секвенирования»**

Правообладатель: **Федеральное государственное бюджетное
учреждение «Национальный медицинский исследовательский
центр радиологии» Министерства здравоохранения Российской
Федерации (ФГБУ «НМИЦ радиологии» Минздрава России) (RU)**

Авторы: *см. на обороте*

Заявка № 2020620440

Дата поступления 18 марта 2020 г.

Дата государственной регистрации

в Реестре баз данных 04 июня 2020 г.

Руководитель Федеральной службы
по интеллектуальной собственности

 Г.П. Ильев



РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2020620921

**«База данных оборудования и расходных материалов,
использующихся для комплексной
молекулярно-генетической диагностики в пульмонологии,
включая методы высокопроизводительного
секвенирования»**

Правообладатель: *Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр радиологии» Министерства здравоохранения Российской Федерации (ФГБУ «НМИЦ радиологии» Минздрава России) (RU)*

Авторы: *см. на обороте*

Заявка № 2020620441

Дата поступления 18 марта 2020 г.

Дата государственной регистрации

в Реестре баз данных 04 июня 2020 г.

Руководитель Федеральной службы
по интеллектуальной собственности

 Г.П. Ивлиев



РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2020620922

«База данных оборудования и расходных материалов,
использующихся для комплексной
молекулярно-генетической диагностики в онкоурологии,
включая методы высокопроизводительного
секвенирования»

Правообладатель: *Федеральное государственное бюджетное
учреждение «Национальный медицинский исследовательский
центр радиологии» Министерства здравоохранения Российской
Федерации (ФГБУ «НМИЦ радиологии» Минздрава России) (RU)*

Авторы: *см. на обороте*

Заявка № 2020620442

Дата поступления 18 марта 2020 г.

Дата государственной регистрации

в Реестре баз данных 04 июня 2020 г.



Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ишев

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2020621080

**База данных оборудования и расходных материалов,
использующихся для комплексной молекулярно-генетической
диагностики в онкоэндокринологии, включая методы
высокопроизводительного секвенирования**

Правообладатель: **Федеральное государственное бюджетное
учреждение «Национальный медицинский исследовательский
центр радиологии» Министерства здравоохранения Российской
Федерации (ФГБУ «НМИЦ радиологии» Минздрава России) (RU)**

Авторы: **Шаталов Петр Алексеевич (RU), Шегай Петр
Викторович (RU), Шинкаркина Анна Петровна (RU)**

Заявка № 2020620894

Дата поступления 17 июня 2020 г.

Дата государственной регистрации

в Реестре баз данных 30 июня 2020 г.

Руководитель Федеральной службы
по интеллектуальной собственности

 Г.П. Ильев



РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2020621116

База данных оборудования и расходных материалов,
использующихся для комплексной молекулярно-генетической
диагностики в онкодерматологии, включая методы
высокопроизводительного секвенирования

Правообладатель: *Федеральное государственное бюджетное
учреждение «Национальный медицинский исследовательский
центр радиологии» Министерства здравоохранения Российской
Федерации (ФГБУ «НМИЦ радиологии» Минздрава России) (RU)*

Авторы: *Шаталов Петр Алексеевич (RU), Шегай Петр
Викторович (RU), Шинкаркина Анна Петровна (RU)*

Заявка № 2020620886

Дата поступления 17 июня 2020 г.

Дата государственной регистрации

в Реестре баз данных 02 июля 2020 г.

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев



РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2020620627

**«База данных оборудования и расходных материалов,
использующихся для комплексной молекулярно-генетической
диагностики в онкогастроэнтерологии и колопроктологии,
включая методы высокопроизводительного секвенирования»**

Правообладатель: *Федеральное государственное бюджетное
учреждение «Национальный медицинский исследовательский
центр радиологии» Министерства здравоохранения Российской
Федерации (ФГБУ «НМИЦ радиологии» Минздрава России) (RU)*

Авторы: *см. на обороте*

Заявка № 2020620437

Дата поступления 18 марта 2020 г.

Дата государственной регистрации

в Реестре баз данных 30 марта 2020 г.

Руководитель Федеральной службы
по интеллектуальной собственности

 Г.П. Ильев



РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2024625034

**Цикл презентаций научно-популярных лекций по
дисциплине «онкология» на тему:
«Высокопроизводительное NGS секвенирование в
практической онкологии»**

Правообладатель: *Федеральное государственное бюджетное учреждение
«Национальный медицинский исследовательский центр радиологии»
Министерства здравоохранения Российской Федерации (RU)*

Авторы: *Шаталов Петр Алексеевич (RU), Каприн Андрей Дмитриевич
(RU), Шегай Петр Викторович (RU), Траснов Алексей Александрович
(RU), Райгородская Мария Павловна (RU), Веселовский Егор
Михайлович (RU), Дорошенко Юлия Артуровна (RU), Жарова Елена
Петровна (RU), Климович Маргарита Сергеевна (RU)*

Заявка № 2024624712

Дата поступления 25 октября 2024 г.

Дата государственной регистрации

в Реестре баз данных 11 ноября 2024 г.



*Руководитель Федеральной службы
по интеллектуальной собственности*

документ подписан электронной подписью
Сертификат 0692e7c1a6300c15412401670bcca2026
Владелец **Зубов Юрий Сергеевич**
Действителен с 10.07.2024 по 03.10.2025

Ю.С. Зубов