

Федеральное государственное автономное образовательное
учреждение высшего образования
«Уральский федеральный университет
имени первого Президента России Б.Н. Ельцина»
Институт естественных наук и математики

На правах рукописи

Лю Гоцзюнь

**ПАТОФИЗИОЛОГИЧЕСКИЕ ГЕНОЗАВИСИМЫЕ
МЕХАНИЗМЫ ОТДЕЛЬНЫХ ТИПОВ
ИММУНОПОСРЕДОВАННОЙ ПАТОЛОГИИ**

14.03.03 – Патологическая физиология

ДИССЕРТАЦИЯ
на соискание ученой степени
кандидата биологических наук

Научные руководители:
Тузанкина Ирина Александровна,
доктор медицинских наук, заслуженный деятель
науки Российской Федерации, профессор

Черешнев Валерий Александрович,
академик РАН, доктор медицинских наук,
профессор

Екатеринбург – 2019

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
ГЛАВА 1 – ОБЗОР ЛИТЕРАТУРЫ.....	19
1.1 – Распространенность, прогноз, лечение, молекулярные подтипы рака мочевого пузыря	19
1.2 – Классификация, распространенность, диагностика, лечение и профилактика первичных иммунодефицитов	28
1.3 – Проблемы исследований первичных иммунодефицитов	35
1.4 – Моногенные и мультигенные дефекты при общем вариабельном иммунодефиците	39
1.5 – Фенотип и генотип синдрома Хеннекама	43
ГЛАВА 2 – МАТЕРИАЛЫ И МЕТОДЫ, ИСПОЛЬЗУЕМЫЕ В РАБОТЕ	45
2.1 – Сбор данных	45
2.2 – Оценка уровней инфильтрирующих иммунных клеток у пациентов с мышечно-инвазивным раком мочевого пузыря	50
2.3 – Кластерный анализ данных пациентов с мышечно-инвазивным раком мочевого пузыря	51
2.4 – Анализ выживаемости пациентов с мышечно-инвазивным раком мочевого пузыря	52
2.5 – Анализ обогащения по функциональной принадлежности генов, характерных для разных субтипов мышечно-инвазивного рака мочевого пузыря	52
2.6 – Скрининг репрезентативных генов между подтипами с использованием подходов машинного обучения	55
2.7 – Определения различий в экспрессии генов между опухолевыми и нормальными образцами тканей мочевого пузыря.....	57
2.8 – Конструирование подтип-зависимой сети взаимодействий мРНК-мкРНК-днРНК при мышечно-инвазивным раке мочевого пузыря	58
2.9 – Оценка плотности генной сети и биологического расстояния для генов общего вариабельного иммунодефицита и других первичных иммунодефицитов.....	59
2.10 – Прогнозирование генов-кандидатов общего вариабельного иммунодефицита.....	61

2.11 – Определение новых генов-кандидатов общего вариабельного иммунодефицита.....	62
2.12 – Визуализация сетей молекулярного взаимодействия с помощью программы Cytoscape	63
2.13 – Полногеномное секвенирование и поиск гена-кандидата для пациента с синдромом Хеннекама	64
ГЛАВА 3 – ИСПОЛЬЗОВАНИЕ МОЛЕКУЛ MIR-141-5P, MIR-141-3P, MIR-200C-3P, AC010326.3, AC073335.2, MIR100HG, CLIC4, GATA3, PALLD В КАЧЕСТВЕ ПРОГНОСТИЧЕСКИХ БИОМАРКЕРОВ У ПАЦИЕНТОВ С МЫШЕЧНО-ИНВАЗИВНЫМ РАКОМ МОЧЕВОГО ПУЗЫРЯ.....	66
3.1 – Влияние уровня инфильтрирующих иммунных клеток на прогрессирование и выживаемость при мышечно-инвазивном раке мочевого пузыря	66
3.2 – Кластерный анализ данных экспрессии РНК пациентов с мышечно-инвазивным раком мочевого пузыря	67
3.3 – Анализ обогащения по функциональной принадлежности данных пациентов с мышечно-инвазивным раком мочевого пузыря.....	75
3.4 – Поиск дифференциально экспрессируемых генов при мышечно-инвазивном раке мочевого пузыря традиционными методами и с помощью машинного обучения	78
3.5 – Конструирование подтип-зависимой мРНК-мкРНК-днРНК сети для мышечно-инвазивного рака мочевого пузыря.....	82
3.6 – Анализ выживаемости при мышечно-инвазивном раке мочевого пузыря	84
ГЛАВА 4 – ИДЕНТИФИКАЦИЯ ГЕНОВ-КАНДИДАТОВ ОБЩЕГО ВАРИАБЕЛЬНОГО ИММУНОДЕФИЦИТА	89
4.1 – Исследование мультигенной природы общего вариабельного иммунодефицита.....	89
4.2 – Анализ взаимосвязей генов нозологических синдромов группы общего вариабельного иммунодефицита	89
4.3 – Оценка выявленных генов-кандидатов общего вариабельного иммунодефицита.....	94
ГЛАВА 5 – МЕТОД ИДЕНТИФИКАЦИИ МУТАЦИЙ ГЕНОВ КАК ВЕРОЯТНО ПАТОГЕННЫХ (FAT4, RAG1, PIK3CD и CSF3R), ПРИ ФЕНОТИПЕ СИНДРОМА ХЕННЕКАМА.....	101
5.1 – Фенотипические проявления синдрома Хеннекама	101

5.2 – Поиск вероятно-патогенных мутаций для верификации синдрома Хеннекама.....	104
5.3 – Оценка вероятной патогенности выявленных в исследовании мутаций	104
ЗАКЛЮЧЕНИЕ	112
ВЫВОДЫ.....	118
ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ.....	119
СПИСОК СОКРАЩЕНИЙ	121
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	124

ВВЕДЕНИЕ

Актуальность темы исследования

Современная патофизиология предполагает системный подход при анализе любого процесса в организме. Привычные методы исследования, такие как оценка клеточных и гуморальных параметров иммунитета, не позволяют всеобъемлюще охватить все задействованные в физиологии или патологии интегральные механизмы нервной, эндокринной и иммунной систем. Системный анализ требует привлечения и методов системной биологии, в том числе биоинформационных. Благодаря развитию генетики появилась возможность выявлять общие закономерности сетевых взаимодействий внутри генома, которые позволяют объяснить многочисленные разнообразные процессы организма, возникающие в ответ на конкретные точечные воздействия. При этом механизмы обеспечения врожденного и адаптивного иммунитета являются главными регуляторами и исполнителями процессов поддержания гомеостаза.

Примером иммуноопосредованной патологии являются первичные иммунодефициты, аутоиммунные, аутовоспалительные, аллергические болезни, а также давно изучаемая онкологическая патология.

Причинами развития различной патологии могут быть как серьезные дефекты отдельных генов, так и генома в целом, нарушения их взаимодействий и взаимовлияний, что в конечном итоге приведёт к неэффективности известных терапевтических методов, применяемых при фенотипических проявлениях патологии. В результате – развитие осложненных форм заболеваний, отсутствие эффекта проводимой терапии, снижение качества жизни и инвалидизация пациентов.

Использование новых знаний о патогенетических механизмах развития иммуноопосредованной патологии позволит своевременно применять методы определения маркеров болезней для их ранней диагностики,

патогенетического обоснования новых терапевтических технологий, предупреждения осложнений медицинских манипуляций и развитие необратимых изменений органов и систем, прогнозирование течения болезни и мониторинга эффективности терапии. Кроме того, систематическая идентификация значимых для болезней генов позволяет раскрыть патогенетические механизмы их формирования, помочь в разработке эффективных персонализированных методов лечения и новых лекарственных средств. Поэтому исследование молекулярно-генетических механизмов формирования патологии и прогнозирования течения болезней всегда было темой глобальных фундаментальных исследований и актуальной проблемой для решения задач в области персонифицированной медицины.

Идентификация генов, связанных с болезнями, применяя методы биологических экспериментов, является сложной и трудоемкой задачей. Развитие методов секвенирования нового поколения (NGS), при котором повышается точность, скорость секвенирования, а затраты на его проведение снижаются, немало способствовало прогрессу в этой области знаний. Применение методов биоинформатики при трактовке результатов секвенирования стало важным средством обнаружения новых генов, предопределяющих развитие патологии.

Стремительное развитие наук о жизни, являющихся интегративными и объединяющими для многих областей знаний, свидетельствует о том, что люди вступили в эру «больших данных». В эту эпоху требования к клиническим медицинским исследованиям непрерывно повышаются, а возникающие проблемы становятся все более сложными. Использование только традиционных методов исследования и оценки отдельных функций человека являются недостаточными для того, чтобы трактовать знания в требованиях современной парадигмы клинических медицинских исследований. Междисциплинарная модель (биоинформатика,

патофизиология, генетика и иммунология) постепенно становится основной тенденцией развития клинических исследований.

Изменяются представления о давно известных фактах. Так, заболевания, трактуемые до недавнего времени как моногенные, в настоящее время расцениваются как полигенные, что определяет огромное фенотипическое разнообразие известных болезней. Открываются значимые перспективы для разработки персонализированной терапии известных болезней с учётом особенностей конкретного индивида, расширяется спектр эффективных методов лечения ранее прогностически неблагоприятной патологии.

Для полигенных заболеваний основными методами выявления причинно-значимых генов являются анализ ко-экспрессии, белок-белковых взаимодействий (protein-protein interaction) и метод обогащения значимости сигнальных путей (enrichment analysis). При выявлении генов-кандидатов при различных видах патологии требуется их верификация, для чего необходимо использование как экспериментальных, так и клинических данных, анализ независимых выборок для определения их роли в процессе развития заболевания. Для моногенного заболевания патогенетическая значимость мутации прогнозируется, в первую очередь, на основе данных полногеномного или полноэкзомного секвенирования генетического материала пациента и других членов его семьи.

Важнейшими проявлениями иммунозависимой патологии являются нарушения пролиферации, репарации и течения воспаления. В нашем исследовании в качестве моделей иммунозависимой патологии отобраны заболевания, протекающие с теми или иными нарушениями пролиферации – рак мочевого пузыря, группа первичных иммунодефицитов, объединённых под термином «общий переменный иммунодефицит» и первичный иммунодефицит из группы комбинированных первичных иммунодефицитов, ассоциированных с синдромальными проявлениями – синдром Хеннекама.

Степень разработанности темы исследования

Работа основывается на исследованиях различных иммунозависимых заболеваний – рака мочевого пузыря и первичных иммунодефицитов, которые принципиально объединены тем, что являются иммуноопосредованными процессами, а патогенез каждого из них во многом определяется экспрессией тех или иных молекул иммунной системы.

Рак мочевого пузыря (РМП) занимает девятое место среди наиболее распространенных видов рака во всем мире с ежегодной регистрацией – около 430 000 новых случаев [60]. Согласно данным M. Wong et al., 2019, по мере старения населения и его роста, абсолютная заболеваемость раком мочевого пузыря в европейских странах может возрасти [181]. Ранее было отмечено, что молекулы микроРНК miRNA-141-5p, miRNA-141-3p, miRNA-200c-3p и ген GATA3 являются наиболее важными маркерами РМП [43]. Кроме того, предыдущие исследования показали, что низкий уровень экспрессии miRNA-200c и miRNA-141 связан с повышенной экспрессией гена ZEB1 [48, 61, 62]. Также низкий уровень экспрессии miRNA-200c сочетается с пониженной экспрессией BMI-1 и E2F3 [124], что играет важную роль в процессах инвазии, миграции и эпителиально-мезенхимального перехода опухолевых клеток рака мочевого пузыря. В последнее время исследованию регуляторной сети мкРНК семейства miR-200 и их целевых генов уделяют пристальное внимание со стороны иммунологов. Известно, что мкРНК семейства miR-200 (miR-200a, -200b, miR-200c, 141 и -429) критически важны в патогенезе рака и метастазировании, а также необходимы для формирования CD8⁺ Т-клеток памяти [202].

Первичные иммунодефициты (ПИД) представляют собой гетерогенную группу генетических нарушений, обусловленных пороками развития или дисфункцией компонентов иммунной системы [154]. Новые данные охарактеризовывают ПИД как широко распространенную патологию, которая серьезно влияет на выживаемость, качество жизни и увеличивает

экономическое бремя семьи [3]. В данной работе уделяется внимание одной из наиболее распространенных форм ПИД – общей вариабельной иммунной недостаточности (ОВИН) и одной из самых редких форм – синдрому Хеннекама.

Использование привычных иммунологических методов исследования, оценивающих клеточные и гуморальные параметры, не позволяют проникнуть в суть патогенеза заболевания, отражая лишь косвенные данные о патологии. Поэтому всё большую перспективность принимают молекулярные методы исследования, результаты которых требуют применения биоинформационных технологий.

Itan Y. и Casanova J.L. в 2015 году предсказали более 3000 генов-кандидатов ПИД, которые имеют тесные взаимосвязи с уже известными генами ПИД, на основе данных о межбелковых взаимодействиях (PPI), полученных с использованием данных высокопроизводительного секвенирования и их математических моделей, которые называют коннектомом генома человека (HGC) [103]. Shivakumar и et al., 2009, предсказали 1442 гена-кандидата ПИД, используя Киотскую энциклопедию генов и геномов (KEGG) и алгоритм опорных векторов [150]. Ortutay et al., 2008, предсказали 26 генов-кандидатов ПИД на основе данных аннотаций белок-белкового взаимодействия и обогащенной онтологии генов (GO) [141]. В модели прогнозирования генов-кандидатов необходимо рассматривать факторы ко-экспрессии, поскольку гены с аналогичными функциями, участвующими в том же сигнальном пути, как правило, более ко-экспрессируются [27, 37, 40, 68, 94, 121, 194, 199, 201]. Поэтому полное использование профилей экспрессии генов большого количества образцов ОВИН позволило нам более эффективно прогнозировать гены-кандидаты для ОВИН.

Синдром Хеннекама представляет собой аутосомно-рецессивное расстройство, характеризующееся ассиметричным расширением

лимфатических сосудов, как периферических, так и внутренних (в том числе кишечника), лицевым дисморфизмом и легкими или умеренными уровнями когнитивных нарушений, возникающих в результате мальформации лимфатической системы [18]. В Соединенных Штатах Национальный институт здравоохранения (NIH) сообщает, по крайней мере, о 50 пациентах с синдромом Хеннекама в 2015 году [160]. В России статистики по синдрому Хеннекама в настоящее время нет. Тем не менее, этот синдром является уникальной природной моделью для изучения механизмов формирования и функционирования лимфатической системы и её взаимосвязи с иммунитетом. Ранее сообщалось, что мутации в *CCBE1*, *FAT4* и *ADAMTS3* могут вызывать синдром Хеннекама (соответственно 1, 2 и 3 типа), но их функциональные исследования не проводились [18, 77, 115]. Относительно низкая заболеваемость и ограниченные информационные ресурсы делают синдром Хеннекама сложным заболеванием для клиницистов и исследователей. Это обосновывает необходимость выявления генов заболевания с использованием NGS, биоинформатики, базы данных генетических вариаций человека. Полное использование этих ресурсов может значительно снизить трудность диагностики таких заболеваний и повысить куративную эффективность.

Цель исследования: выявление значимых патофизиологических механизмов формирования отдельных видов иммунозависимой патологии на основе биоинформационного анализа.

Задачи исследования:

1. Проанализировать патофизиологические механизмы гетерогенности мышечно-инвазивного рака мочевого пузыря.
2. Выявить прогностические биомаркеры двух основных подтипов мышечно-инвазивного рака мочевого пузыря.

3. Провести оценку экспрессии молекул ингибирующих иммунные функции (PD-1, PD-L1, CTLA-4, HAVCR-2 и LAG-3) у пациентов с различными подтипами мышечно-инвазивного рака мочевого пузыря.

4. Проанализировать механизмы, лежащие в основе мультигенной природы общей вариабельной иммунной недостаточности.

5. Выявить новые кандидатные гены в группе болезней «общая вариабельная иммунная недостаточность».

6. Выявить мутации генов, приводящих к развитию первичного иммунодефицита «синдром Хеннекама».

Методология и методы исследования. Работа выполнена на кафедре медицинской биохимии и биофизики Института естественных наук и математики Уральского федерального университета.

Клинические данные пациентов с мышечно-инвазивным раком мочевого пузыря (МИРМП), а также данные секвенса мРНК, длинных некодирующих РНК (днРНК) были получены из публично доступного сервера «Атласа генома рака» (<https://cancergenome.nih.gov>), данные секвенса микро РНК (мкРНК) – из публично доступного сервера «Broad GDAC Firehose» (<https://gdac.broadinstitute.org>).

Для выполнения исследований, связанных с мышечно-инвазивным раком мочевого пузыря использовались необработанные данные секвенса РНК 403 пациентов и 19 образцов для контроля из нормальной ткани, полученные из портала данных TCGA (<https://cancergenome.nih.gov>).

Данные экспрессии генов, полученных с использованием микрочипов ДНК 139 пациентов с мышечно-инвазивным раком мочевого пузыря, которые были получены из базы данных Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov>).

Данные экспрессии генов, полученных с использованием микрочипов ДНК 111 пациентов с ОВИН и 56 относительно здоровых контрольных образцов были получены из базы данных GEO.

Геномная ДНК пациента с клиническим диагнозом «синдром Хеннекама», была выделена из моноклеарных клеток периферической крови с использованием набора ДНК FlexiGene (Qiagen) в соответствии с инструкциями производителя. Для пробоподготовки использовалась методика фрагментации ДНК PCR free технологии пробоподготовки для анализа. Секвенирование всего генома (WGS) осуществлялось в компании «Геномед» (Москва) по технологии секвенирования нового поколения методом парно-концевого чтения. Анализ покрывает 98,5 % всего генома. Среднее покрытие составляет не менее 30 прочтений.

Анализ результатов полногеномного секвенирования и выявление мутаций и однонуклеотидных полиморфизмов (SNP) осуществлялся на суперкомпьютере, предоставленном центром коллективного пользования ИММ УрО РАН «Суперкомпьютерный центр ИММ УрО РАН», (Екатеринбург). От родителей пациента было получено информированное согласие на использование обезличенных результатов исследований.

Статистический анализ, разработка биоинформационных инструментов и математические модели были выполнены с использованием Python версии 3.7.1 (<https://www.python.org/>), R версии 3.4.3 (<https://www.r-project.org/>) на операционной системе Linux.

Степень достоверности и апробация результатов. Достоверность результатов исследования обеспечивается теоретической достоверностью, достаточным количеством образцов, воспроизводимостью исследований, статистическим анализом данных, независимым набором данных тестирования. Достоверность результатов исследования подтверждена комиссией по проверке первичной документации №33.15-03/163 (акт проверки первичных документов от 02.02.2019 г.). Основные положения диссертации представлены на III Всероссийском конгрессе «Аутоиммунные и иммунодефицитные заболевания» (Москва, 16-17 ноября 2018 г.); X Конгрессе НОДГО «Актуальные проблемы и перспективы развития детской

гематологии-онкологии в Российской Федерации» (Сочи, 27-29 апреля 2019 г.), IV Объединенном иммунологическом форуме в Новосибирске (24-29 июня 2019 г.), 17 Международном конгрессе по иммунологии (Пекин, 19-23 октября 2019), на II Санкт-Петербургском лимфологическом форуме «Лимфология XXI века: новые подходы и актуальные исследования (Санкт-Петербург, 10-11 октября 2019 г.), Международном Евро-Азиатском Конгрессе по вопросам биоэтики, молекулярной и персонализированной медицины «Biomed-inn-2019» (Пермь, 5-8 ноября 2019 г.).

Личное участие автора. Соискатель непосредственно участвовал на всех этапах исследования, в том числе в создании основной идеи, планировании научной работы, формулировании рабочей гипотезы, задач, рабочих заданий, определении методологии исследования. Интерпретация и анализ результатов диссертации проведена совместно с научным руководителем – заслуженным деятелем науки России, профессором, доктором медицинских наук Тузанкиной Ириной Александровной и академиком РАН, профессором, доктором медицинских наук Черешневым Валерием Александровичем.

В обсуждении результатов работы участвовали заведующая кафедрой медицинской биохимии и биофизики, доктор биологических наук Данилова Ирина Георгиевна и старший научный сотрудник лаборатории иммунологии воспаления Института иммунологии и физиологии Уральского отделения РАН кандидат медицинских наук Болков Михаил Артемович. Отдельные биоинформационные исследования были проведены в сотрудничестве с профессором Гочин Лю, ключевой лабораторией функциональной геномной биоинформатики университета науки и техники Внутренней Монголии (Баотоу, Китай).

Положения, выносимые на защиту:

1. В качестве диагностических маркеров иммунозависимой патологии – мышечно-инвазивного рака мочевого пузыря, а также для

прогноза и дифференциальной диагностики базального и люминального подтипа необходимо использовать данные об экспрессии молекул мкРНК miR-141-5p, miR-141-3p, miR-200C-3P, днРНК AC010326.3, AC073335.2, MIR100HG и мРНК CLIC4, GATA3, PALLD.

2. Разница в экспрессии иммуносупрессивных молекул PD-1, PD-L1, CTLA-4, HAVCR-2 и LAG-3 позволяет различать базальные и люминальные подтипы мышечно-инвазивного рака мочевого пузыря - базальные опухоли демонстрируют более высокие уровни экспрессии иммуносупрессивных молекул, чем опухоли люминального подтипа.

3. Молекулярно-генетическую этиологию группы нозологических синдромов первичных иммунодефицитов «Общая вариабельная иммунная недостаточность» представляет расширенный спектр из 172 кандидатных генов, выделенных в настоящем исследовании.

Научная новизна исследования. В результате проведенных научных исследований впервые в Российской Федерации выполнено следующее.

Впервые проведен систематический анализ сигнальных путей, значимых для иммунитета, приводящих к гетерогенности базального и люминального подтипов мышечно-инвазивного рака мочевого пузыря.

Впервые при мышечно-инвазивном раке мочевого пузыря, при разных его подтипах, выявлены различия результатов молекулярных взаимодействий молекул мкРНК, днРНК и мРНК.

Впервые представлена новая перспектива понимания мультигенной природы группы синдромов первичных иммунодефицитов «Общая вариабельная иммунная недостаточность», благодаря полученным результатам анализа взаимосвязей генов.

Разработан новый метод прогнозирования генов-кандидатов группы синдромов первичных иммунодефицитов «Общая вариабельная иммунная недостаточность» путем включения в клинико-биоинформационный анализ

факторов ко-экспрессии, межбелкового взаимодействия и данных сигнальных путей.

Представлен перечень вероятно-патогенных мутаций в генах, комбинация которых может приводить к развитию первичного иммунодефицита «синдром Хеннекама».

Теоретическая и практическая значимость работы. Молекулы микро РНК miR-141-5p, miR-141-3p, miR-200C-3p, длинных некодирующих РНК (днРНК) AC010326.3 и AC073335.2 (контролирующие GATA3), MIR100HG (ингибирующая CLIC4 и PALLD), мРНК молекул CLIC4 (хлоридный внутриклеточный канал 4), GATA3 (транскрипционный фактор), PALLD (Palladin, белок, ассоциированный с цитоскелетом) могут иметь большое значение в онкогенезе и прогрессировании опухоли. Эти молекулы могут служить новыми прогностическими параметрами, а также маркерами классификации подтипов мышечно-инвазивного рака мочевого пузыря. Выявлен факт повышенной экспрессии генов, относящихся к процессу эпителиально-мезенхимального перехода, метастазирования и функционирования иммунной системы, при базальном подтипе рака мочевого пузыря. Люминальный подтип характеризуется относительно повышенной экспрессией генов, отвечающих за метаболические процессы. При базальном подтипе рака выявлены более высокие значения экспрессии иммуносупрессорных рецепторов (PD-1, PD-L1, CTLA-4, HAVCR-2 и LAG-3), чем при люминальном подтипе. Эти данные обеспечивают более глубокое понимание иммунологической гетерогенности рака мочевого пузыря.

Выводы о том, что гены ОВИН по своим функциональным взаимосвязям более похожи друг на друга и тесно взаимодействуют друг с другом по сравнению с остальными генами ПИД, помогут лучше понять мультигенную природу ОВИН. Предложенные 172 гена-кандидата ОВИН помогут осуществить раннюю диагностику ОВИН и помочь в создании таргетной терапии.

Гомозиготная мутация (g.125452634G>A) в гене FAT4 (член семейства кадгерина 4 - супрессор опухоли гомолог 4, функция неизвестна, распространён повсеместно), гетерозиготная мутация (g.36575963G>A) в гене RAG1 (ген активации рекомбиназы 1, функционирует при формировании антигенраспознающих молекул Т- и В-лимфоцитов), гетерозиготная мутация (g.9715914T>A) в гене PIK3CD (специфичная для лейкоцитов субъединица фермента фосфатидилинозитол-3-киназы) и гетерозиготной мутации (g.36471505C>T) в CSF3R (рецептор для колониестимулирующего фактора 3, необходимого для созревания лейкоцитов) могут иметь патогенетическую значимость и позволят расширить диагностические возможности, с учетом понимания патогенеза синдрома Хеннекама.

Внедрение результатов исследования в практику. Полученные результаты внедрены в практику ГАУЗ Свердловской области «Областная детская клиническая больница», были использованы в работе Региональной комиссии Министерства здравоохранения Свердловской области по младенческой смертности.

Список генов-кандидатов ОВИН внедрен для использования в работе лаборатории иммунологии воспаления Института иммунологии и физиологии УрО РАН в преподавательской деятельности и консультативной медицинской практике.

Публикации. По результатам диссертации опубликовано 7 работ, индексируемых в международных электронных базах Web of Science и Scopus, а также в изданиях, рецензируемых ВАК.

Объем и структура диссертации. Диссертация изложена на 144 страницах машинописного текста, состоит из введения, обзора литературы, главы «Материалы и методы», 3 глав с результатами собственных исследований, выводов, практических рекомендаций, списка сокращений и списка использованных литературных источников (204 источника, включая 5

отечественных и 199 зарубежных). Работа иллюстрирована 12 таблицами, 30 рисунками и 3 формулами.

Благодарности. В первую очередь, выражаю особую благодарность моим научным руководителям, профессору Тузанкиной Ирине Александровне и академику Черешневу Валерию Александровичу за их поддержку и руководство моей диссертацией. Они обогатили мои академические и профессиональные навыки с помощью плодотворных научных дискуссий, блестящих идей и бесконечной мотивации.

Моя глубокая признательность заведующей кафедрой медицинской биохимии и биофизики, доценту Даниловой Ирине Георгиевне за помощь в проведении этого исследования и огромную доброту. Она помогала мне на протяжении всего периода аспирантуры.

Я благодарен старшему научному сотруднику лаборатории иммунологии воспаления Института иммунологии и физиологии УрО РАН, кандидату медицинских наук Болкову Михаилу Артемовичу за его эффективные комментарии и обсуждение работы.

Я благодарен заведующему лабораторией иммунологии воспаления Института иммунологии и физиологии УрО РАН, профессору Евгению Юрьевичу Гусеву за системный взгляд на проблемы, поднятые в работе и важные замечания относительно общих молекулярных механизмов и сигнальных путей.

Я благодарен ЗДН РФ, профессору Юшкову Борису Германовичу и доктору медицинских наук, старшему научному сотруднику Института иммунологии и физиологии УрО РАН, Сарапульцеву Алексею Петровичу за корректные и точные замечания, которые позволили скорректировать работу в нужном направлении.

Моя искренняя благодарность всем друзьям и коллегам на кафедре медицинской биохимии и биофизики за рабочую атмосферу, невероятные прекрасные воспоминания, которые я всегда буду помнить, и их честные

отзывы и критические замечания, которые помогли мне улучшить мою работу.

Последнее и самое главное, я хотел бы высказать особую благодарность моим родителям, подруге и младшей сестре, за их заботы и большую любовь ко мне. Их постоянная поддержка бесценна для меня и никакие слова не могут выразить мои чувства по отношению к ним.

ГЛАВА 1 – ОБЗОР ЛИТЕРАТУРЫ

Онкологические заболевания и врожденные дефекты иммунной системы представляют огромную медико-социальную проблему в виду высокой распространенности, сложности диагностики, наличия клинической маскировки болезней, длительного характера течения указанной патологии, а также развитием осложнений и риском летального исхода или инвалидизации, существенно нарушающей качество жизни пациентов. Поэтому проблема иммуноопосредованной патологии активно изучается во всем мире.

1.1 – Распространенность, прогноз, лечение, молекулярные подтипы рака мочевого пузыря

Как известно, злокачественные заболевания занимают ведущее место среди всех болезней, приводящих к летальному исходу. Этот класс болезней имеет полигенную природу с синергичным действием нескольких генов. Реализация развития опухоли может происходить в границах одного органа, различных органах одной системы или нескольких системах организма. По крайней мере, около 33% случаев рака имеют подтверждения генетической природы [64].

Рак мочевого пузыря (РМП) является одним из наиболее распространенных и смертельных заболеваний во всем мире. Согласно сообщению Китайского национального онкологического центра, в течение 2015 года РМП был выявлен у 80 500 китайских граждан и явился причиной 32 900 случаев смертей [26]. Рак мочевого пузыря (РМП) является четвертой наиболее распространенной формой рака у мужчин и 11-й у женщин, ежегодно в США регистрируется 76 960 новых случаев заболевания и 16 390 смертей по его причине [171]. По данным из MEDLINE (Medical Literature

Analysis and Retrieval System Online) частота рецидивов РМП достигает почти 80 %, и этот тип онкологии является самым дорогим для лечения в пересчете на одного пациента. Высокая частота рецидивов инициирует множество проблем, влияющих на качество жизни из-за его устойчивости [81, 179]. РМП обычно классифицируется на мышечно-инвазивный (МИРМП) и мышечно-неинвазивный рак мочевого пузыря, в зависимости от того, ограничены ли раковые клетки собственной пластинкой слизистой оболочки (*lamina propria*) или проникают в *muscularis propria* [15, 22, 75]. Четверть пациентов с РМП страдают от мышечно-инвазивной формы с более высоким риском метастазирования, при котором клетки опухоли могут распространяться в региональные лимфатические узлы малого таза и / или висцеральной области, в результате чего болезнь становится инкурабельной [102].

Большим числом исследователей указывалось, что в зависимости от общих паттернов экспрессии РНК или специфических геномных изменений, МИРМП можно дополнительно разделить на два основных подтипа, а именно базальный и люминальный подтип МИРМП [8, 43, 44, 85, 100, 101, 152], которые поразительно похожи на молекулярные подтипы, впервые описанные при раке молочной железы [129, 145]. Базальный подтип привлек большое внимание исследователей, поскольку связан с более агрессивным фенотипом опухоли и имеет более высокий риск отдаленных метастазов, чем люминальный [43, 85]. Молекулярные биомаркеры и сигнальные пути, вовлеченные в разные подтипы мышечно-инвазивного рака являются ключевыми шагами к пониманию их различий. Первое из отличий двух подтипов в том, что этиологически они развиваются по разным сигнальным путям. Для базального субтипа характерно повышение экспрессии генов, вовлеченных в работу иммунной системы и механизмы эпителиально-мезенхимального перехода [85].

Базальный подтип МИРПМ по своей природе более агрессивен и связан с более сильным метастазированием, тогда как люминальный принято считать более мягким. Для базального подтипа МИРМП описаны следующие биологические маркеры (гены с повышенной экспрессией): CD44, CDH3, KRT14, KRT16, KRT5, KRT6A, KRT6B, KRT6C, STAT3. Для люминального подтипа МИРМП более характерно относительное повышение экспрессии таких генов, как CYP2J2, ERBB2, ERBB3, FGFR3, FOXA1, GATA3, GPX2, KRT18, KRT19, KRT20, KRT7, KRT8, PPARG, TRIM24 [120].

Для понимания патогенеза необходимо знать различия в дифференцировки клеток и их развитие, что ведет к различиям в степени тяжести течения болезни, продолжительности, локализации и чувствительности или толерантности к лекарственным средствам, что, в свою очередь, влияет на разницу в диагностике, методах прогнозирования и лечении рака, которая в настоящее время еще остаётся неудовлетворительной.

Традиционные прогностические критерии, которые используются врачами для расчета прогноза выживаемости, включают в себя гистологические данные и стадию опухоли по системе определения стадии TNM, в которой учитывается количество или размер опухоли, и количество метастазов в лимфатических узлах [156, 176, 190]. В соответствии с этими традиционными показателями, прогноз пациентов с раком может сильно отличаться, даже если они находились на одной и той же стадии и получали традиционное лечение. Поэтому особенно важно найти различно экспрессированные гены или определенные мутации, которые тесно соотносятся с генетическим типом рака для проведения специфического лечения. Высоко производительные методы секвенирования, наряду с улучшением точности секвенирования и снижением затрат, сильно повлияли на использование медицинских биомаркеров в ранней диагностике, прогнозе рака, прогнозировании рецидива, мониторинге реакций на лекарственные

средства, а также разработку таргетных методов лечения [30, 65, 84, 133, 164, 200]. Все это значительно улучшает выживаемость пациентов.

Sanguedolce F., 2015, были опубликованы сведения о том, что модели прогнозирования с использованием молекулярных маркеров являются более точными, чем традиционные критерии прогноза. Несмотря на то, что в последние годы был достигнут значительный прогресс как в сфере выявления молекулярных маркеров, так и в создании многофакторных инструментов для оценки прогноза РМП, в настоящее время доступны всего лишь несколько квалифицированных биомаркеров для использования в прогнозе течения РМП [128, 200]. Таким образом, необходимо определение новых молекулярных биомаркеров и включение их в многофакторные прогнозирующие инструменты.

В течение многих десятилетий в традиционной стратегии лечения рака основными были хирургия, химиотерапия, лучевая терапия [153, 163]. Радикальная цистэктомия с неоадьювантной химиотерапией на основе цисплатина является стандартом первой линии мультимодального лечения больных с МИРМП. Тем не менее, около 60 % пациентов с МИРМП не проявляют существенного ответа на лечение из-за остаточных злокачественных клеток и/или метастазирования [89]. К тому же, для многих пациентов традиционно применяемая терапия цисплатина противопоказана из-за высокой токсичности препарата [168].

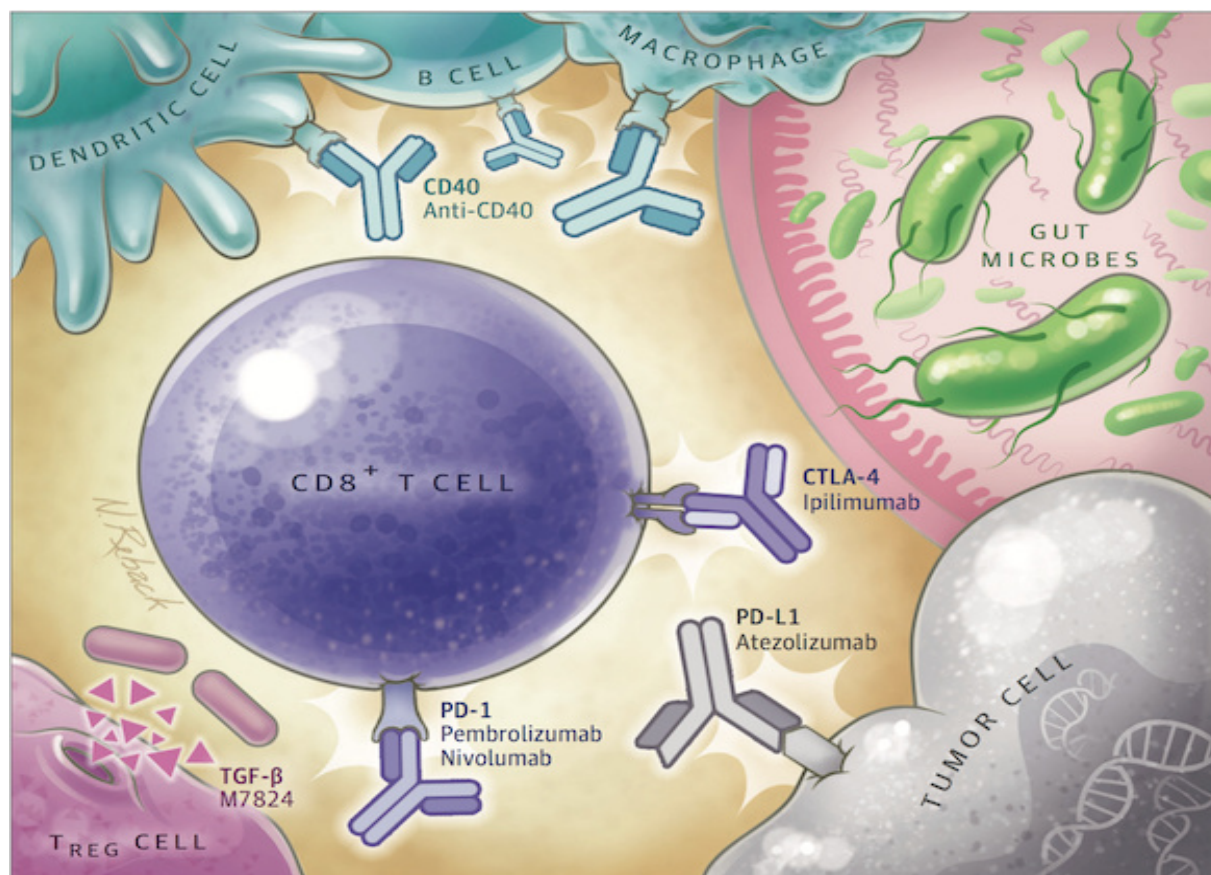
Иммунотерапия становится привлекательной и подкупающей доверие стратегией среди различных терапевтических возможностей. За последние годы этот метод уже успел зарекомендоваться как эффективный в лечении злокачественных опухолей, но при этом необходимо учитывать обеспечение баланса между стимулирующими и ингибирующими сигналами, которые имеют решающее значение для предотвращения неконтролируемых иммунных реакций, приводящих к развитию осложнений. Эффекторные клетки иммунной системы могут быть блокированы различными

ингибиторами, называемыми «иммунными контрольными точками». Рецепторы на поверхности лимфоцитов, такие как PD-1 и CTLA-4, являются иммунными контрольными точками, подавляющими сигналы активации Т-клетки для защиты нормальных тканей от аутоиммунитета. Однако при раке эти механизмы часто активируются для подавления противоопухолевого иммунного ответа [66, 196] (рисунок 1). В 2013 году журнал Science назвал иммунотерапию рака «открытием года», а в 2018 году Нобелевская премия по физиологии и медицине была присуждена двум исследователям, а именно, James P. Allison и Tasuku Honjo, за исследования, которые, в конечном счете, привели к использованию ингибиторов иммунных контрольных точек для лечения рака. В настоящее время ингибиторы PD-1/PD-L1, такие как пембролизумаб, атезолизумаб, дурвалумаб, ниволумаб и авелумаб одобрены Управлением по контролю за продуктами и лекарствами США (FDA) для использования в клинической практике для пациентов с прогрессирующим МИРМП, у которых платиносодержащие режимы химиотерапии оказались неэффективными [17].

В клинических испытаниях, ниволумаб – человеческий иммуноглобулин субкласса IgG4 к рецептору PD1 сильный иммунный ответ наблюдался менее чем у 20 % пациентов с местнораспространенным или метастатическим уротелиальным раком [153]. Многие исследования направлены на разработку методов использования ингибиторов иммунных контрольных точек, чтобы повысить их эффективность, минимизировать их токсичность и расширить их использование для более широкого диапазона видов онкологии.

Мутационная нагрузка опухоли (TMB, TML) – это количество мутаций, содержащихся в 1 мегабазе кодирующего генома опухолевых клеток. TMB анализируется с помощью полноэкзомного или полногеномного секвенирования биопсийного или парафинизированного материала, полученного от пациентов. TMB представляет собой количественный

биомаркер, который используется для прогнозирования чувствительности солидных опухолей к терапии ингибиторов иммунных контрольных точек и не зависит от экспрессии PD-L1, но коррелирует с частотой ответа на иммунотерпию [135].



**Рисунок 1 – Лечение рака путем ингибирования PD1 и CTLA4
(Friedrich M.J., 2019) [66]**

Инфильтрирующие опухоль лимфоциты (TILs) - благоприятный прогностический фактор у пациентов со злокачественными опухолями [34, 191]. Инфильтрация опухоли происходит цитотоксическими Т-лимфоцитами, макрофагами, естественными киллерами, дендритными клетками, которые могут оказывать как стимулирующее, так и ингибирующее влияние на опухолевый рост, а также имеют различное прогностическое значение [157, 189].

Комбинация методов оценки мутационной нагрузки опухоли (ТМВ) и инфильтрирующих опухоль лимфоцитов (TILs) помогает лучше определять восприимчивость пациентов с некоторыми злокачественными новообразованиями к иммунотерапии [142]. Выявление и изучение молекулярных подтипов рака поможет создать независимую оценку прогноза и получить дополнительную информацию для прогнозирования реакции на иммунотерапию [153].

Что касается молекулярных драйверов развития рака, принято считать, что онкогенные мутации и нижестоящие сигнальные пути в предраковой или раковой клетке играют решающую роль в формировании и прогрессировании опухоли. Кроме того, недавние исследования показали, что метаболическое перепрограммирование играет гораздо более важную роль, чем считалось ранее [25]. Вполне возможно, что большое количество геномных мутаций, обнаруженных при раке, обеспечивает селективное преимущество для раковых клеток в стрессовом микроокружении опухоли путем перепрограммирования процессов клеточного метаболизма [25]. Независимо от того, что является основной причиной развития рака, становится ясно, что и онкогенные сигнальные пути, и перепрограммирование метаболических путей затрагивают множество генов, согласованно работающих в сложной сети. Следовательно, представление регуляции данных процессов в виде геной сети дает более глубокое понимание причин развития рака.

МикроРНК (мкРНК или miRNA), как правило, могут регулировать экспрессию генов или огромные и сложные регуляторные сети генов для широкого спектра клеточных процессов, таких как пролиферация клеток, обмена веществ, апоптоза, противовирусного иммунитета [39, 47, 92, 95, 104, 122, 174, 184, 203]. Известно, что микроРНК не только обладают ингибирующим действием на экспрессию генов, но, в некоторых случаях, также способствуют экспрессии генов [125, 167]. Длинные некодирующие РНК (днРНК или lncRNA) также оказывают стимулирующие или

подавляющие эффекты на генные сети во время развития опухоли [159, 165]. Кроме того, отклонение от нормы или возмущение в микроРНК-опосредованных уровнях экспрессии мРНК и lncRNA имеет значительную корреляцию с серьезными клиническими последствиями, в том числе в виде заболеваний различного происхождения и злокачественных новообразований [6, 7, 96, 111, 140, 183]. В предшествующих публикациях различные исследователи проводили оценку взаимосвязей МИРМП-связанных микроРНК и их генов-мишеней, однако оценка генетической гетерогенности подтипов МИРМП не учитывалась [113, 114, 204]. Поэтому одной из задач нашего исследования стало выявление характерных для подтипов молекулярных сигнальных путей и идентификации новых биомаркеров, тем более, что базальный и люминальный подтипы существенно друг от друга отличаются.

Атлас ракового генома (TCGA) представляет собой проект, целью которого является систематизация данных о генетических мутациях, приводящих к возникновению рака. Этот проект обеспечивает комплексную геномную характеристику различной онкопатологии, что ускоряет наше понимание биологии рака и ставит новые научные задачи и терапевтические цели. Проект TCGA, с бюджетом более чем в \$300 миллионов долларов, является вторым самым крупным проектом после «Проекта Аполлон». Проект TCGA был запущен в 2005 году совместно с Национальным институтом рака и Национальным научно-исследовательским институтом генома человека (США), и ставит перед собой трудную задачу победить рак с помощью исследователей со всего мира [188]. Сегодня TCGA включает в себя омиксные данные более чем 30 видов рака. Омиксными называют технологии, основанные на достижениях геномики, транскриптомики, протеомики и метаболомики, таким образом омиксные технологии являются одним из главных инструментов геномной и постгеномной медицины. Для каждого из наиболее распространенных видов рака (легких, молочной

железы, предстательной железы, мочевого пузыря и других) TCGA аккумулировала геномные и транскриптомные портреты, а также данные по метилированию более чем 300-500 образцов. Клинические данные в базе данных TCGA включают в себя более 34 000 образцов с привязкой к клиническому состоянию на момент анализа, диагноз-специфической, диагностической информацией, первоначальных протоколов лечения, а также другую информацию (рисунок 2).

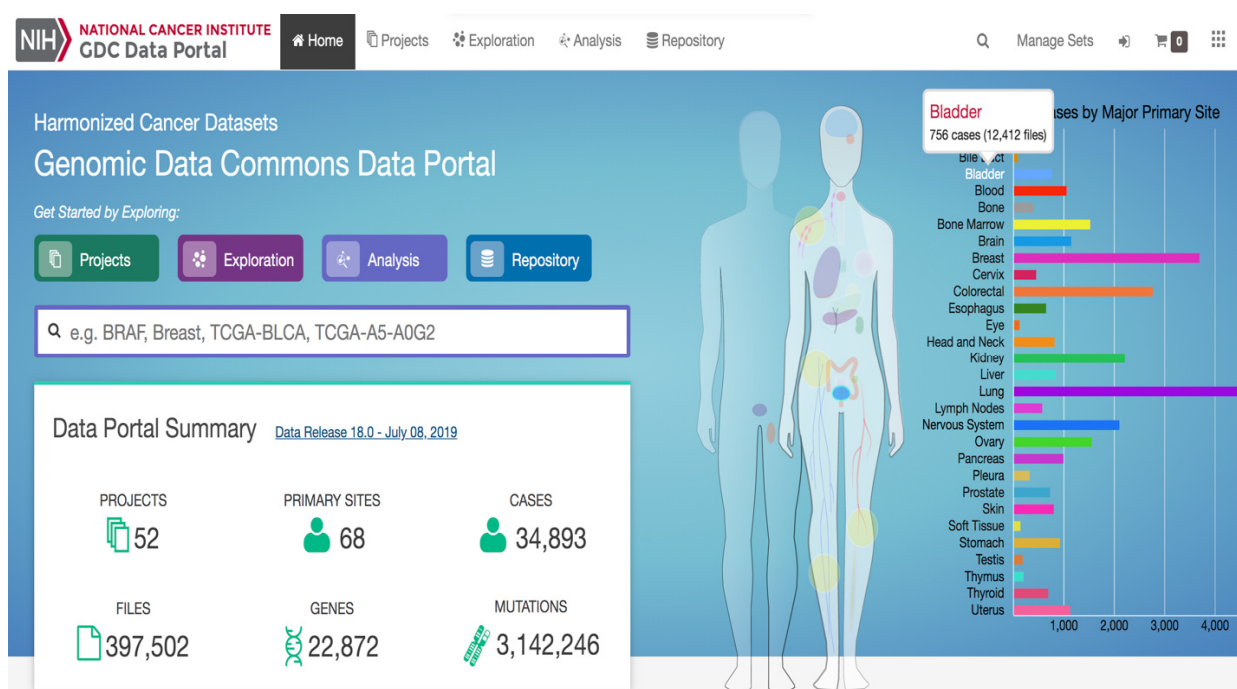


Рисунок 2 – Снимок первой страницы базы данных TCGA

Примечание: на изображении показан фрагмент первой страницы веб-сайта <https://portal.gdc.cancer.gov>, при этом во всплывающей подсказке показано число клинических случаев рака мочевого пузыря, которые были загружены в базу (756 случаев, 12412 файлов). Слева можно производить поиск по ключевым словам, дана общая статистическая картина базы.

Рост числа наборов данных образцов опухолей в проекте TCGA помогает повысить статистическую мощность и способность обнаруживать молекулярные дефекты при раковых заболеваниях. Это делает TCGA полезным источником информации для оценки экспрессий генов [106], молекулярной классификации подтипа опухоли [41, 42, 45] открытий

изменений в генах-драйверах [86], определения прогностических маркеров [88, 162] и других прикладных направлений. Кроме того, последний прогресс в комплексном мульти-омиксном анализе пролил свет на новое представление о геноме злокачественных новообразований [9, 67, 87, 93]. С накоплением богатства многомерных данных, появляется большая потребность в методах, направленных в сторону интегративного анализа множественных геномных источников данных.

1.2 – Классификация, распространенность, диагностика, лечение и профилактика первичных иммунодефицитов

Первичные иммунодефициты (ПИД) представляют собой гетерогенную группу наследственных заболеваний, связанных с дефектами развития или дисфункцией компонентов иммунной системы [154]. Анализ 10 000 американских семей показал, что, за исключением селективного дефицита иммуноглобулина А (IgA), распространенность ПИД приближается к 1 на 1200 населения. По данным Jeffrey Modell Foundation (JMF) разными типами ПИД в той или иной степени может быть поражено 1-2% популяции всех людей на планете [73]. При этом для детей с ПИД общий риск развития малигнизации оценивается в 4-25% [55]. В настоящее время во всем мире, в том числе и в Российской Федерации, в течение первого года жизни умирают дети из-за отсутствия своевременной диагностики и патогенетически обоснованной терапии [1, 2, 3, 5].

Чаще всего ПИД дебютируют в раннем возрасте, оказывая влияние на здоровье новорожденных и детей ясельного возраста [109], однако новые данные свидетельствуют о том, что большинство пациентов выживают и взрослеют, при этом ПИД-патология маскируется под другими диагнозами и формирует серьезную нагрузку на систему здравоохранения, проявляясь хроническими трудно излечимыми болезнями [4, 119].

Одна из первых попыток классифицировать ПИД сделана в 1970 году комитетом Всемирной Организации Здравоохранения. Классификация включала 16 нозологий. К 1998 году их число достигло 50. В настоящее время их число приближается к 460 [99].

Согласно докладу Международного союза иммунологических обществ, существует 9 групп ПИД [99]:

1. Комбинированные иммунодефициты (клеточные и гуморальные дефекты, в том числе варианты тяжелого комбинированного иммунного дефицита (ТКИН) и менее тяжелые, чем ТКИН, комбинированные дефекты).
2. Комбинированные ПИД, ассоциированные с синдромальными проявлениями (в том числе дефекты репарации ДНК и др.).
3. Преимущественно антительные дефекты (при отсутствии, снижении или нормальном количестве В-лимфоцитов).
4. ПИД с иммунной дисрегуляцией (наследственные гемофагоцитарные синдромы, дефекты регуляторных Т-клеток, аутоиммунный лимфопролиферативный синдром и др.).
5. Врожденные дефекты числа и функций фагоцитов (врожденные нейтропении, дефекты подвижности лейкоцитов, дефекты респираторного взрыва).
6. Дефекты врожденного иммунитета (Менделевская восприимчивость к микобактериям, эпидермоплазии, восприимчивость к вирусным инфекциям, к инвазивным грибковым инфекциям и т.д.).
7. Аутовоспалительные синдромы (в том числе идеопатические лихорадки и асептический остеомиелит).
8. Дефекты комплемента.
9. Фенокопии ПИД (вызванные соматическими мутациями или вследствие аутоиммунитета к продуктам генов иммунной системы).

Различные формы ПИД имеют различные патогенетические механизмы и могут быть абсолютно друг на друга не похожи. Более того, каждый из

пациентов с одним и тем же генетически подтвержденным диагнозом может иметь различные варианты дебюта и фенотипические проявления болезни, от крайне тяжелых до бессимптомных форм. Описаны многочисленные клинические случаи выявления родственников с той же мутацией, что и у пробанда, но без клинических проявлений [4].

Вторичные иммунодефициты вызываются различными причинами, не связанными с врожденными генетическими дефектами. ПИД принципиально отличается от вторичных иммунодефицитов, которые могут возникнуть в результате таких причин, как дельтаретровирусные инфекции (Т-лимфотропный вирус человека, HTLV) или бактериальные инфекции, недостаточность питания, потери белков (в том числе ожоговая болезнь) и иммуноглобулина, злокачественные опухоли или использование иммуносупрессивных препаратов [155].

Несмотря на данные о том, что распространенность диагностированного ПИД в США приближается к 1:1200, это число не учитывает людей с легкими дефектами иммунной системы, которые не получили официального диагноза в виду отсутствия генетической диагностики. Дефицит IgA является наиболее распространенным из ПИД, и обнаруживается в частоте от 1:300 до 1:500 человек [23]. Другие наиболее частые формы ПИД включают в себя общий переменный иммунодефицит (ОВИН), тяжелый комбинированный иммунодефицит (ТКИН), агаммаглобулинемию (болезнь Брутона), синдром Вискотт-Олдрича, синдром Ди Джорджи, атаксию телеангиэктазию (синдром Луи-Барр), семейную средиземноморскую лихорадку, синдром Маршалла, наследственный ангионевротический отек и др. [4, 118].

Если коротко охарактеризовать главные различия между ПИД, то они зависят от уровня генетического дефекта и соответствующих дефектов рецепторов или протеинов. Функционально иммунная система разделяется на два главных компонента – врожденный и адаптивный иммунный ответ, и

в зависимости от этого уже можно выделить условно две крупные группы иммунодефицитов, в зависимости от того, какой компонент иммунного ответа в первую очередь нарушен. Однако, фактически, представленная Международным союзом иммунологических обществ (International Union of Immunological Societies – IUIS) классификация наиболее полно отражает возможные классы иммунодефицитов, а дефекты врожденного иммунитета занимают лишь 1/9 часть возможного разнообразия [99]. Одним из первых описанных первичных иммунодефицитов была Х-сцепленная агаммаглобулинемия, описанная в 1952 году доктором Огденом Брутоном. На простом электрофорезе белков плазмы он отметил у мальчика с симптомами повторных бактериальных инфекций практически полное отсутствие гамма-глобулинов. Сегодня известно, что В-клетки после их трансформации в плазматические клетки, в большей степени, ответственны за гуморальный иммунитет и являются основными продуцентами гамма-глобулинов, то есть антител. Дефекты, возникающие на любом этапе развития и созревания В-клеток, приводят к тем или иным формам гипо- или агаммаглобулинемии. Нарушения созревания эффекторных или хелперных Т-клеток приводят к различным иммунодефицитам с дефектом Т-клеточного звена. Но эти зависимости условные, так как продукция антител В-клетками требует комплекса взаимодействий различных Т-клеток и моноцитов, а большинство дефектов Т-клеток приводит к комбинированной (Т- и В-клеточной) недостаточности [56]. Однако Т- и В-клеточные механизмы – это прерогатива адаптивной иммунной системы, тогда как врожденные иммунные реакции представляют собой первую линию защиты от потенциальных патогенов. Соответствующие дефекты рецепторов врожденного иммунитета (например, TLR) и любого из компонентов каскада активации врожденных механизмов приводит к столь же серьезным иммунодефицитам, как и дефект Т- или В-клеток. Огромный спектр ПИД включает в себя все большее число синдромов, которые напрямую связаны с

аутоиммунитетом и иммунной дисрегуляцией, тогда как признаки инфекционной патологии отходят на второй план и не являются главными [4, 169].

Конкретные симптомы первичного иммунодефицита зависят от типа дефекта. Характер инфекций, а также дополнительные симптомы, могут дать ключ относительно точной природы иммунного дефекта [109]. Однако, в диагностике иммунодефицитов большую значимость имеет не столько наличие инфекционной патологии, сколько другие симптомы, связанные с патогенезом ПИД.

Наиболее частыми проявлениями ПИД на первом году жизни являются:

- лимфопения ($<2,5 \cdot 10^9/\text{л}$), панцитопения;
- гипоплазия тимуса по результатам УЗИ-сканирования;
- тяжелые поражения кожи и/или диарея;
- пороки развития – сердца, органов пищеварения, костной системы и др.;
- нарушения репарации, которые могут приводить к развитию осложнений послеоперационного периода после проведенных оперативных вмешательств и других медицинских манипуляций;
- неадекватность развития воспалительного процесса:
 - отсутствие терапевтического эффекта,
 - отсутствие доказательств инфекционного воспаления,
 - сепсис-подобный синдром и гиперцитокинемия при лимфогистиоцитозе;
- микозы, кожно-слизистые кандидозы;
- гипогаммаглобулинемия – экстремальное снижение концентрации иммуноглобулина G в сыворотке крови;
- БЦЖит и другие вакцинальные инфекции;
- стойкий сахарный диабет и другие аутоиммунные болезни;
- позднее отпадение пуповины;

- семейный анамнез: случаи регрессирующей беременности или ранней детской смертности [4].

Среди прочих признаков ПИД в более зрелом возрасте является обширный список разнообразных симптомов, среди которых врожденные дефекты различных органов, в том числе сердца и костной системы, болезни кожи и соединительной ткани, аутоиммунные заболевания, когда иммунная система атакует собственные ткани организма, или опухоли (иногда специфические формы рака, такие как лимфома), аутовоспалительные болезни, когда нарушен каскад воспалительных реакций и происходит разрушение тканей за счет спонтанного нерегулируемого воспаления. Пациенты с иммунодефицитом имеют повышенный риск развития злокачественной онкотрансформации тканей вследствие различных врожденных дефектов иммунитета, связанных с распознаванием раковых клеток, сопротивления им, а также при многочисленных дефектах репарации ДНК. Более того, благодаря успехам медицины, продолжительность жизни больных с первичным и вторичным иммунодефицитами возрастает, что ведет к повышенному риску развития злокачественных новообразований у пациентов, которые выжили вопреки наличию иммунодефицита.

Диагноз ПИД может быть разделен на три типа в зависимости от степени доказательности. Критерии для постановки диагноза были согласованы в 1999 году [46]. Выделяли три типа диагноза: возможный, вероятный и окончательный.

Окончательный диагноз ставится в том случае, когда пациент имеет 98 % вероятность того же самого диагноза спустя 20 лет; такой уровень диагноза может быть установлен при наличии генетического подтверждения или при очень точных и конкретных обстоятельствах и фенотипических проявлениях.

Вероятный диагноз устанавливается, когда генетическая диагностика отсутствует, но у пациента есть все другие характеристики конкретного

заболевания; вероятность того, что тот же диагноз будет поставлен через 20 лет, оценивается в 85–97%.

Возможный диагноз устанавливается, когда у пациента есть определенные признаки заболевания, но не все [46].

Современные методы лечения ПИД включают симптоматическую поддержку, таргетную терапию, заместительную терапию и два вида радикальных операций: трансплантацию гемопоэтических стволовых клеток (ТГСК) и генную терапию. Следует отметить, что генная терапия все еще находится в экспериментальной стадии исследования. Для некоторых детей с ПИД ТГСК является наиболее важным и даже единственным средством для лечения болезни и восстановления функций иммунной системы. При этом отсутствие генетического подтверждения не является противопоказанием для трансплантации гемопоэтических стволовых клеток.

Среди, так называемых, профилактических методов для ПИД доступен неонатальный скрининг. Задача скрининга состоит в том, чтобы установить предварительный диагноз у новорожденного до появления симптомов заболевания, когда ещё возможно «окно возможностей» для своевременного применения куративных мер. В настоящее время в скрининг новорожденных в США включены, по меньшей мере, 54 заболевания [137]. При этом скрининг новорожденных на дефекты Т- и В-клеточного иммунитета путем оценки в сухих пятнах крови молекул TREC и KREC уже производится во всех штатах, во многих странах Европы и Ближнего Востока, Юго-Восточной Азии. В России этот метод применяется в качестве пилотных исследований.

1.3 – Проблемы исследований первичных иммунодефицитов

Следующие факты делают первичные иммунодефициты сложной группой заболеваний как для практикующего врача, так и для научного исследователя [147].

Во-первых, низкая осведомленность врачей о стремительно развивающейся сфере медицины. Эта проблема стоит перед врачами во всем мире. Сравнительно недавно ПИД считались редкими болезнями, поэтому многие врачи общей практики, а также узкие специалисты имеют низкую настороженность или осведомленность о ПИД. Многие исследователи общей вариабельной иммунной недостаточности (ОВИН) отмечают, что временная задержка между проявлением симптомов и началом заместительной терапии иммуноглобулином составляет от 5 до 10 лет [50, 53, 110, 178].

Вторым камнем преткновения является то, что фенотипические изменения, связанные с ПИД, как правило, очень разнообразны. Так, например, в случае пациентов с дефектами белка WASP (синдрома Вискотт-Олдрича), точный характер дефекта гена (например, миссенс или нонсенс-мутация, точное расположение аномалий сплайс-сайта) может значительно изменить фенотип синдрома. Проявления мутаций этого гена могут быть от очень тяжелых, до мягких: X-сцепленная тромбоцитопения, В-клеточная лимфома, частые бактериальные и грибковые инфекции, экземы, небольшие размеры тромбоцитов или нейтропения. Диагностика ПИД, как правило, требует глубокого анализа симптомов вкупе с оценкой анамнеза и генеалогии.

Последняя проблема исследователей и клиницистов связана с трудностью поиска информации. В открытом доступе базы данных содержат мало образцов, разнообразие которых, в сравнении с данными по онкологии, например, представлены также ограничено. Существует очень немного информационных ресурсов, содержащих взаимосвязь клинических описаний

и функциональных геномных данных, межбелковых взаимодействий (protein-protein interactions) и сигнальных путей (pathways). Примеры специализированных баз данных включают UniProt, IntAct, STRING и KEGG.

Для диагностики ПИД на современном этапе у врачей и исследователей существует множество инструментов, которые Crina Samarghitean et al., 2009, разделили на семь категорий (рисунок 3).

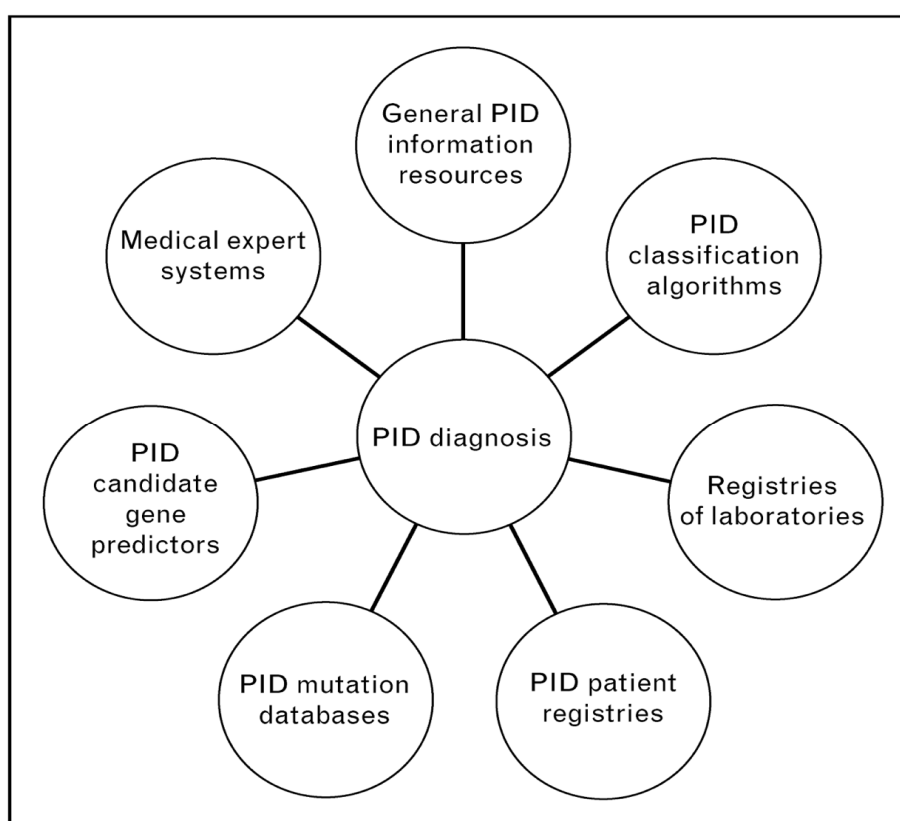


Рисунок 3 – Схематическая группировка биоинформатических ресурсов и инструментов, обеспечивающих получение информации о первичных иммунодефицитах (Samarghitean C., 2009) [166]

Во-первых, основные ПИД-ресурсы приводят много информации, ранжированной по уровням от геной до белковых структур, от моделей болезни до определенных групп диагнозов, и так далее.

Во-вторых, существуют классификации ПИД, содержащие клинические признаки.

В-третьих, доступны лабораторные критерии и соответствующие инструменты.

В-четвертых, существуют национальные и международные регистры пациентов ПИД, которые дополняются базами данных мутаций (пятая категория), из которых информация может быть использована для сравнения рассматриваемого случая с ранее описанными.

В-шестых, для прогнозирования или определения приоритетности новых генов-кандидатов ПИД доступны инструменты биоинформатики, которые также используются для принятия решений (седьмой категории) в ПИД-диагностике [166].

Большое количество разнообразных ПИД создает трудности для диагностики, в том числе и на клиническом уровне. Кроме того, многие заболевания настолько редки, что невозможно найти достаточное количество семей для анализа сцепления. Скрининг и раннее выявление ПИД является серьезной проблемой для врачей. Больше число известных в настоящий момент генетических дефектов получили в последние несколько лет благодаря высокопроизводительному секвенированию. Выявление новых генов-кандидатов для ПИД позволит проводить дальнейшую приоритизацию генов при подтверждении у больных с ПИД, при которых точный причинный ген пока не определен.

Некоторое время назад, Keerthikumar et al., 2009, использовали метод опорных векторов для классификации всех человеческих генов на ПИД-гены и не-ПИД гены. В основе принципа этой классификации был расчёт показателя достоверности для каждого гена-кандидата ПИД на основе 69 особенностей, наблюдаемых для 148 известных ПИД генов [150]. На основании литературных поисков, мы обнаружили, что внимание научных исследователей, участвующих в идентификации генов ПИД, были также направлены на интеграцию функциональных аннотаций из генной онтологии (GO) и составления наборов данных сети межбелковых взаимодействий. Ortutay et al., 2008, предсказал 26 генов-кандидатов ПИД путем анализа свойств межбелковых взаимодействий (PPI) сети всех известных на тот

момент генов иммунной системы человека и их онтологии (GO). Также на основании расчета, так называемого, биологического расстояния (имеется ввиду функциональная взаимозависимость) предсказано 3 110 генов-кандидатов ПИД [28, 103].

Исследователи, желающее принять участие в исследованиях по проблеме ПИД, каждый раз сталкиваются с вопросом: являются ли ПИД мультигенными или моногенными? Первоначально ПИД считали врожденными и моногенными расстройствами, которые следуют принципам менделевского наследования [24, 161]. Моногенные заболевания являются результатом изменений в одном гене, происходящих во всех клетках организма. Однако прогресс в области секвенирования ДНК способствовал выявлению мультигенных и соматических причин ПИД, констатирована широкая фенотипическая изменчивость этих заболеваний [74, 192].

Понимание того, что большинстве ПИД имеют мультигенный характер, является первым шагом в понимании патогенеза всех заболеваний. В мультигенной концепции ПИД являются результатами сложных взаимодействий между генами. Исходя из этого, ученые попытались найти «биологическое расстояние» между ПИД-генами и другими белок-кодирующими генами человека, и они обнаружили, что гены ПИД, по сравнению с другими генами человека, располагаются, как правило, в центральной ступице сети генома человека и более тесно взаимодействуют друг с другом (рисунок 4) [103]. Кроме того, гены ПИД образуют несколько тесно связанных суб-кластеров, и большинство из них имеют, по крайней мере, одного функционально близкого соседа среди широкого спектра биологических механизмов [28, 103, 182]. Раскрытие этих взаимосвязей может обеспечить более глубокое понимание разнообразия генетических путей, лежащих в основе ПИД, что, в свою очередь, может способствовать открытию новых возможностей для создания лекарственных препаратов и терапевтических подходов.

1.4 – Моногенные и мультигенные дефекты при общем переменном иммунодефиците

Общий переменный иммунодефицит (ОВИН) – наиболее распространенная форма ПИД среди взрослых пациентов, характеризующийся низким уровнем сывороточных иммуноглобулинов и дефицитом специфических антител при исключении других причин гипогаммаглобулинемии. Предполагаемая распространенность ОВИН от 1:50 000 до 1:25 000 населения [32]. У большинства пациентов с ОВИН иммунодефицит выявляется во втором, третьем или четвертом десятилетии жизни после нескольких пневмоний, однако дебют диагноза также возможен у детей и пожилых. Основной симптоматикой могут быть бактериальные, вирусные, грибковые и паразитарные инфекции. Однако у большей части пациентов неинфекционные проявления являются лидирующими и пациенты наблюдаются у ревматологов, гастроэнтерологов и онкологов под другими диагнозами. Примерно у 50% пациентов наблюдаются аутоиммунные проявления. Существует повышенный риск малигнизации.

Большинство ОВИН являются мультигенными, их причиной могут быть различные мутации множества генов, результатом которых становятся одинаковые или различные фенотипы ОВИН. Моногенные причины ОВИН были найдены в около 10 % случаев [32].

Моногенные причины ОВИН включают в себя следующие мутации:

I. Рecessивно наследуемые мутации в IL21, IL21R, LRBA, ICOS, PRKCD, CD19, CD20, CD21, CD27, CD81, RAC2;

II. Доминантно наследуемые мутации в CTLA4, TNFSF12, NFKB1, PLCG2, NFKB2, PIK3CD, PIK3R1, VAV1, BLK, IKZF1, IRF2BP2;

III. Моноаллельные или биаллельные мутации в TNFRSF13B и TNFRSF13C [71].

Например, комплекс В-клеточного ко-рецептора состоит из CD19, CD21, CD81 и CD225, которые вместе снижают порог для активации В-клеток связыванием антигена с В-клеточным рецептором. В недавних исследованиях были выявлены мутации в генах рецепторов CD19, CD81 и CD21 при аутосомно-рецессивных формах ОВИН [12].

В-лимфоидная тирозиновая протеинкиназа (BLK) играет важную роль в передаче сигналов с В-клеточного рецептора (BCR) и в осуществлении взаимосвязи с Т-клетками. Ранее гетерозиготная LOF мутация (мутация с потерей функции гена) в гене BLK была обнаружена у двух родственных пациентов с ОВИН [59]. NFkB1 кодирует зрелую субъединицу P52 и его предшественник P105 фактора транскрипции NF-кВ в В-лимфоцитах. И для этого белка также были найдены соответствующие клинические случаи, когда мутации этого гена в спорадических или родственных случаях приводили к появлению ОВИН [76, 144]. Мутация гена TNFRSF13B приводит к ухудшению уже нарушенной передачи сигналов Toll-подобных рецепторов (TLR) или вызывает дефекты передачи сигналов от TLR [134], что также было обнаружено у многих пациентов с ОВИН [71].

Несмотря на то, что все случаи ОВИН, как полагают, являются результатом генетических дефектов, в подавляющем большинстве точная причина заболевания неизвестна [98].

Существуют многочисленные примеры мультигенных (или полигенных) причин ОВИН, при которых варианты мутаций в нескольких генах могут внести вклад в формирование одних и тех же или различных фенотипов этого заболевания [16, 63, 105].

Почему же ОВИН - более мультигенная когорта, по сравнению с другими ПИД? Одной из возможных причин является тот факт, что мультигенная природа ОВИН является результатом сложных взаимодействий между генами.

Понимание механизма их взаимодействия должно помочь врачам и исследователям получить представление о патофизиологии ПИД, что повысит диагностику и изменит подходы к лечению. Itan Y. и Casanova J.L. в 2015 году путем расчета биологического функционального расстояния для всех белок-кодирующих генов показали, что при визуализации сети генома человека гены ОВИН, как правило, располагаются в центральной ступице сети [103] (рисунок 4).

Таким образом, было выявлено более тесное взаимодействие друг с другом генов ОВИН, по сравнению с другими генами ПИД. Это исследование частично помогает лучше понять мультигенную природу ОВИН и других заболеваний.

Основные механизмы, лежащие в основе сети межбелковых взаимодействий (PPI) генов ОВИН остаются неясными, отсутствует систематический уровень интерпретации. С недавним накоплением новых данных об экспрессии генов при ОВИН [16, 71], надежных данных молекулярных взаимодействий [21, 79, 83], и самых современных методов вычислений [20, 21, 149], существует настоятельная необходимость выявления генов кандидатов ОВИН.

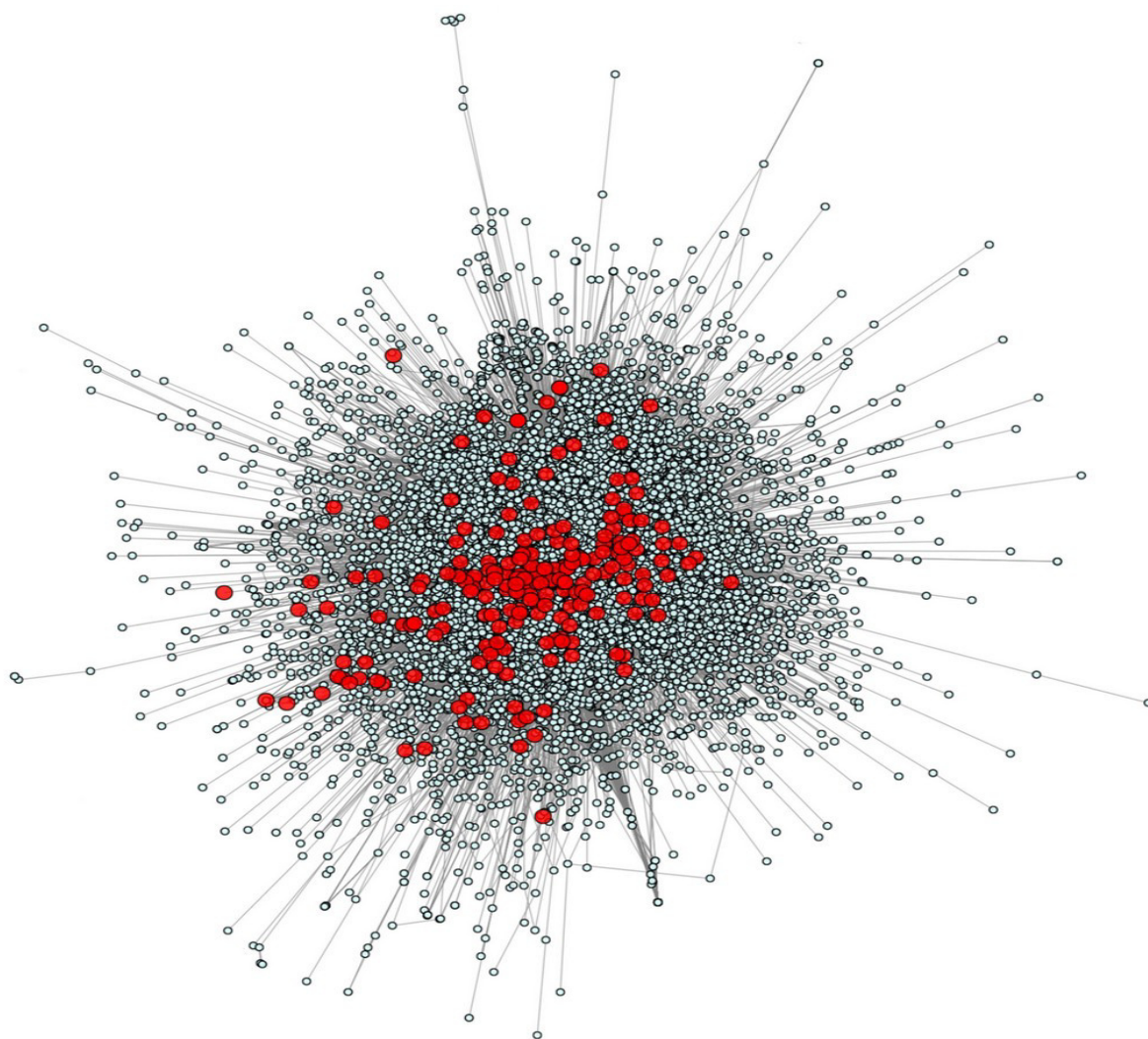


Рисунок 4 – Сеть генома человека и генов ПИД

Примечание: показано 229 известных на момент исследования ПИД-генов (красным) и 14131 белок-кодирующих генов, для которых была доступна информация о т.н. «биологической дистанции» (Ithan Y., Casanova J.L., 2015) [103].

Использование методов системной биологии и биоинформатики ускорит и повысит точность обнаружения новых генов ОВИН, что позволит глубже понять патогенетические механизмы данной патологии. К тому же это недорогой и быстрый способ, который помогает клиническому специалисту в диагностике пациентов с фенотипом ОВИН и неизвестной генетической причиной.

1.5 – Фенотип и генотип синдрома Хеннекама

Синдром Хеннекама (Hennekam syndrome) является аутосомно-рецессивным заболеванием, и это одна из самых редких форм первичных иммунодефицитов. Фактически синдром Хеннекама относится к мальформациям лимфатической системы.

Аномалии лимфатических сосудов блокируют циркуляцию лимфы, а накопление жидкости влияет на различные области тела, в том числе лицо, конечности, а также внутренние органы. Черты лица характеризуются уплощенными внешним видом, широкой переносицей, гипертелоризмом, эпикантом и другими аномалиями [18].

Синдром Хеннекама разделяется на три основных типа в зависимости от генетической причины, влияющей на VEGF-C/VEGFR-3 сигнальные пути [116, 132, 177]. Синдром Хеннекама 1 типа, также известный как CCBE1-ассоциированный синдром Хеннекама, первоначально был описан голландским врачом Раулем Хеннекамом в 1989 году [18].

Основным молекулярным механизмом лимфедемы при синдроме Хеннекама 1 типа принято считать уменьшение возможности мутированного белка CCBE1 (коллаген- и кальций-связывающий протеин 1, содержащий белок, подобный эпидермальному фактору роста) ускорять и концентрировать активацию первичного лимфангиогенного фактора роста VEGF-C [132].

Для синдрома Хеннекама 2 типа причиной является гомозиготная или компаунд-гетерозиготные мутации в гене FAT4 на хромосоме 4q28. Интересно, что мутация в гене FAT4 может вызвать синдром Ван Мальдергема (VMLDS2), другое заболевание, при котором некоторые симптомы перекрываются с фенотипом синдрома Хеннекама [77].

В исследовании 2017 года коллектив авторов во главе с P. Brouillard выявили синдром Хеннекама 3 типа, при котором выявлена гетерозиготная

мутация в гене ADAMTS3 на хромосоме 4q13. Что еще более важно, исследователи подчеркнули тесную функциональную взаимосвязь между белками ADAMTS3 и CCBE1 в активации молекулы VEGFR3, которая является краеугольным камнем для дифференциации и функционирования лимфатических эндотелиальных клеток [115].

Тем не менее, мутации в этих генах обнаруживаются только у части пациентов, и генетическая этиология большинства пациентов с синдромом Хеннекама остается неизвестной в значительной степени из-за того, что синдром генетически гетерогенен.

ГЛАВА 2 – МАТЕРИАЛЫ И МЕТОДЫ, ИСПОЛЬЗУЕМЫЕ В РАБОТЕ

2.1 – Сбор данных

Клинические данные пациентов с мышечно-инвазивным раком мочевого пузыря (МИРМП) и нормализованные методом FPKM (fragments per kilobase of exon model per million mapped reads) данные секвенса мРНК и днРНК были получены из публично доступного сервера Атласа генома рака (TCGA) с веб-адреса <https://cancergenome.nih.gov>.

Нормализованные методом RPM (reads per million) данные секвенса мкРНК были загружены из публично доступного сервера «Broad GDAC Firehose» (<https://gdac.broadinstitute.org>). Для клинических данных были отобраны образцы, обозначенные как мышечно-инвазивная уротелиальная карцинома в стадии pT2 или выше и исключили из исследования образцы, у которых отсутствовали данные долгосрочного наблюдения, поскольку было необходимо оценить выживаемость пациентов. Стилль предоставления клинических данных представлен в виде примера в таблице 1.

Таблица 1 – Стилль предоставления клинических данных пациентов с мышечно-инвазивным раком мочевого пузыря

Sample\Clinical	Gender	Stage		Status	Time
TCGA-UY-A78P	FEMALE	Stage II		Alive	398
TCGA-LT-A8JT	FEMALE	Stage II		Alive	1522
TCGA-CF-A7I0	MALE	Stage II		Alive	641
.....					
.....					
.....					

В колонках были представлены идентификаторы образца тканей опухоли, диагноз, ткань, тип опухоли, стадия, метастазирование, раса пациента, возраст, пол, курит пациент или нет, а также был ли жив пациент на момент исследования или нет.

Набор данных мРНК секвенса пациентов представлял из себя таблицу значений уровней экспрессии каждой из 19 181 мРНК для 422 образцов тканей (403 опухолевых и 19 образцов нормальных тканей).

Стиль предоставления данных экспрессии мРНК пациентов представлен в таблице 2. В колонках показаны идентификаторы образцов и уровень экспрессии для каждой из мРНК.

Таблица 2 – Стиль представления мРНК данных пациентов с мышечно-инвазивным раком мочевого пузыря

mRNA\Sample	TCGA-DK-AA6R	TCGA-DK-AA76		TCGA-DK-AA6P
SAMD11	0.90622308	0.47413391		0.98372
NOC2L	15.914834	42.7130759		23.0476527154
KLHL17	2.3565348	3.85048108		2.32950289
.....				
.....				

Набор данных днРНК (lncRNA) пациентов с МИРПП состоял из значений экспрессии 14 376 различных днРНК и 422 образцов (403 опухолевых и 19 образцов нормальных тканей). Стиль предоставления данных экспрессии днРНК пациентов представлен в таблице 3. В колонках показаны идентификаторы образцов и уровень экспрессии для каждой из днРНК.

Набор данных мкРНК пациентов с МИРМП содержал значения экспрессии 2 588 различных мкРНК и 422 образцов (403 опухолевых и 19 образцов нормальных тканей). Стиль предоставления данных экспрессии

мкРНК пациентов представлен в таблице 4. В колонках показаны идентификаторы образцов и уровень экспрессии для каждой из мкРНК.

Таблица 3 – Стиль представления данных экспрессии днРНК пациентов с мышечно-инвазивным раком мочевого пузыря

lncRNA\Sample	TCGA-DK-AA6R	TCGA-DK-AA76		TCGA-DK-AA6P
LINC01128	0.47886052	0.72885094		0.39689252
LINC00115	0.39191886	0.90750898		0.29728013
FAM41C	0.01870541	0		0
.....				

Таблица 4 – Стиль представления данных экспрессии мкРНК пациентов с мышечно-инвазивным раком мочевого пузыря

miRNA\Sample	TCGA-DK-AA6R	TCGA-DK-AA76		TCGA-DK-AA6P
hsa-let-7a-3p	54.639048	48.253177		37.221532
hsa-let-7a-2-3p	9.446938	7.659235		2.788584
.....				
.....				
hsa-miR-32-5p	68.937116	43.274676		23.884825
hsa-miR-32-3p	1.787259	1.148885		0.242486

Наборы данных ДНК-микрочипирования пациентов с МИРМП и ОВИН были получены из исследований, опубликованных в базе данных GEO (Gene Expression Omnibus, <https://www.ncbi.nlm.nih.gov>).

Для исследований генов, связанных с МИРМП, были загружены наборы данных GSE13507 и GSE31684. После отборов образцов, отмеченных как мышечно-инвазивный рак мочевого пузыря (стадия pT2 или выше), набор данных GSE31684 содержал 78 образцов МИРМП, GSE13507 - 61

образец МИРМП, которые в дальнейшем были использованы для прогнозирования на основе уровней экспрессии генов, связанных с болезнью.

Для исследований генов, связанных с ОБИН, были загружены наборы данных GSE72625 и GSE51406. Набор данных GSE72625 содержал сведения о 20 пациентах с ОБИН и 17 относительно здоровых взрослых, набор данных GSE51406 содержал сведения о 91 пациенте с ОБИН и 39 относительно здоровых взрослых. Процедура предварительной обработки для GSE72625 и GSE51406 состояла в удалении генов, содержащих более 70–75% пропущенных значений, и $\log_2$ -преобразовании значений экспрессии гена.

Список известных на момент проведения исследования 351 гена ПИД и, в том числе, 39 генов ОБИН, был получен с сайта Европейского общества иммунодефицитов (<https://esid.org>).

Ниже представлен список 351 гена ПИД, используемый в нашем исследовании:

ACP5, ACTB, ADA, ADAM17, ADAR1, AICDA, AIRE, AK2, AP1S3, AP3B1, AP3D1, APOL1, APOLLO, ARPC1B, ATM, ATP6AP1, B2M, BACH2, BAFF-R, BCL10, BCL11B, BLM, BLNK, BTK, C1QA, C1QB, C1QC, C1R, C1S, C2, C2BP1, C3, C4A, C4B, C5, C6, C7, C8A, C8B, C8G, C9, CARD11, CARD11, CARD14, CARD15, CARD9, CASP10, CASP8, CCBE1, CD19, CD247, CD27, CD3D, CD3E, CD3G, CD40, CD40LG, CD46, CD55, CD59, CD70, CD79A, CD79B, CD81, CD8A, CDCA7, CEBPE, CECR1, CFB, CFD, CFH, CFHR1-5, CFI, CFP, CFTR, CHD7, CIAS1, CIITA, CLCN7, CLPB, COPA, CORO1A, CR2, CSF2RA, CSF2RB, CSF3R, CTC1, CTLA4, CTPS1, CTSC, CXCR4, CYBA, CYBB, DCLRE1B, DCLRE1C, DKC1, DNAJC21, DNASE1L3, DNASE2, DNMT3B, DOCK2, DOCK8, ELANE, EPG5, ERCC6L2, EXTL3, FAAP24, FADD, FASLG, FAT4, FCGR3A, FCN3, FERMT3, FOXN1, FOXP3, FPR1, G6PC3, G6PD, G6PT1, GATA2, GCS1, GFI1, GINS1, GOF, HAX1, HELLS, HMOX, HOIL1, HOIP1, HYOU1, ICOS, IFIH1, IFNAR2, IFNGR1, IFNGR2, IGHM, IGKC, IGLL1, IKBA, IKBKB, IKBKG, IKZF1, IL10, IL10RA, IL10RB,

IL12B, IL12RB1, IL17F, IL17RA, IL17RC, IL1RN, IL21, IL21R, IL2RA, IL2RG, IL36RN, IL7R, INO80, IRAK1, IRAK4, IRF2BP2, IRF3, IRF7, IRF8, ISG15, ITCH, ITGB2, ITK, JAGN1, JAK1, JAK3, KDM6A, KMT2D, LAMTOR2, LAT, LCK, LIG1, LIG4, LPIN2, LRBA, LYST, MAGT1, MALT1, MAP3K14, MASP2, MCM4, MEFV, MKL1, MLL2, MOGS, MS4A1, MSH6, MSN, MTHFD1, MVK, MYD88, MYSM1, NBAS, NBS1, NCF1, NCF2, NCF4, NCSTN, NEMO, NFAT5, NFKB1, NFKB2, NFKBIA, NHEJ1, NHP2, NLRC4, NLRP1, NLRP12, NLRP3, NOD2, NOP10, NSMCE3, ORAI1, OSTM1, OTULIN, PARN, PEPD, PGM3, PIK3CD, PIK3R1, PLCG2, PLEKHM1, PMS2, PNP, POLA1, POLE, POLE2, PRF1, PRKCD, PRKDC, PSEN, PSENEN, PSMB8, PSTPIP1, PTEN, PTPRC, PYPAF1, RAB27A, RAC2, RAG1, RAG2, RANBP2, RASGRP1, RBCK1, RECQL3, RELB, RFX5, RFXANK, RFXAP, RHOH, RLTPR, RMRP, RNASEH2A, RNASEH2B, RNASEH2C, RNF168, RNF31, RNU4ATAC, RORC, RPSA, RTEL1, SAMD9, SAMD9L, SAMHD1, SBDS, SEMA3E, SERPING1, SH2D1A, SH3BP2, SLC29A3, SLC35C1, SLC46A1, SMARCAL1, SMARCD2, SNM1, SNX10, SP110, SPINK5, SRP54, STAT1, STAT1, STAT2, STAT3, STAT5B, STIM1, STK4, STN1, STX11, STXBP2, TACI, TAP1, TAP2, TAPBP, TAZ, TBK1, TBX1, TCF3, TCIRG1, TCN2, TERC, TERT, TFRC, THBD, TICAM1, TINF2, TIRAP, TLR3, TMC6, TMC8, TMEM173, TNFAIP3, TNFRSF11A, TNFRSF13B, TNFRSF13C, TNFRSF1A, TNFRSF4, TNFRSF5, TNFRSF6, TNFSF11, TNFSF12, TNFSF5, TNFSF7, TPP1, TPP2, TRAC, TRAF3, TRAF3IP2, TREX1, TRNT1, TTC37, TTC7A, TYK2, UNC13D, UNC93B1, UNG, USB1, USP18, VPS13B, VPS45, WAS, WDR1, WIPF1, WRAP53, XIAP, ZAP70, ZBTB24.

Список из 39 генов, связанных с ОБИИ, используемый в этом исследовании:

ACVR2A, BCL2L1, BTK, CD19, CD81, CLEC16A, CLU, CORO1B, CTLA4, FAS, FCGR2A, FUS, GATA2, IKZF1, IL10RA, IL21R, IL2RG, ITGAM, ITPKB, KCNN4, LRBA, MS4A1, NFKB1, NFKB2, NLRP12, PIK3CD, PIK3R1,

PLCG2, PRKCD, PTPRCAP, RAC2, RAG1, RPS6KB2, STAT1, TENM3, TNFAIP3, TNFRSF13B, TNFSF12, VAV1.

Образец ДНК пациента с диагнозом «синдром Хеннекама» экстрагировали из мононуклеарных клеток периферической крови с использованием набора для выделения ДНК FlexiGen (Qiagen) в соответствии с инструкциями изготовителя. Использовалась методика фрагментации ДНК PCR free технологии пробоподготовки для анализа. Полногеномное секвенирование выполнено компанией Геномед (Москва). Анализ ДНК проводился по технологии секвенирования нового поколения методом парно-концевого чтения. Анализ покрывает 98,5% всего генома. Среднее покрытие составляло не менее 30 прочтений.

Протокол исследования был одобрен комитетом по этике Института естественных наук и математики ФГАОУ ВО "Уральский Федеральный Университет имени первого Президента России Б.Н. Ельцина", от родителей было получено информированное согласие.

2.2 – Оценка уровней инфильтрирующих иммунных клеток у пациентов с мышечно-инвазивным раком мочевого пузыря

Относительные уровни 22 инфильтрирующих опухоль иммунных клеток (включая эозинофилы, нейтрофилы, моноциты, плазматические клетки, наивные и памяти В-клетки, покоящиеся и активированные дендритные клетки, покоящиеся и активированные тучные клетки, покоящиеся и активированные естественные клетки-киллеры, макрофаги M0, макрофаги M1, Макрофаги M2, CD8+ Т-клетки, регуляторные Т-клетки, наивные CD4+ Т-клетки, фолликулярные хелперные Т-клетки, $\gamma\delta$ Т-клетки, покоящиеся CD4+ Т-клетки памяти, активированные CD4+ Т-клетки) для каждого образца оценивали с использованием CIBERSORT, с матрицей подписи LM22 и 1000 перестановками (<https://cibersort.stanford.edu>). Сумма

всех фракций оцениваемых типов инфильтрирующих иммунных клеток равна единице для каждого образца опухоли, и все оценки, следовательно, относятся к общему содержанию лейкоцитов.

2.3 – Кластерный анализ данных пациентов с мышечно-инвазивным раком мочевого пузыря

Перед кластерным анализом мы провели отбор по вариабельным генам и провели логарифмическую нормализацию. Для наборов данных мРНК, мкРНК и днРНК пациентов с мышечно-инвазивным раком мочевого пузыря (МИРМП) вначале было удалено 25% генов, которые имели низкое среднее арифметическое значение данного гена в образцах, а затем сохранили 3000 мРНК, 300 микроРНК и 3000 lncRNA с высокой дисперсией. Впоследствии все значения экспрессии генов были подвергнуты нормализации $\log_2$ (значение экспрессии + 1).

В предыдущих исследованиях, комбинация методов кластеризации кластеров (Cluster of Cluster, CoC) и консенсус-кластерного анализа показали себя как надежный способ кластеризации данных для исследований рака [97, 112, 131]. Консенсусный кластерный анализ относится к ситуации, в которой различные (входные) кластеризации были получены при анализе разных наборов данных, но необходимо найти одну (консенсусную) кластеризацию. Анализ «Кластеризации кластеров» - это метод интеграции результатов первичной кластеризации в окончательные назначения кластеров. Вначале был проведен консенсусный кластерный анализ на основе отдельных данных мРНК, мкРНК и днРНК, затем провели анализ кластеризации кластеров, для определения подтипов базального и люминального рака мочевого пузыря с учетом всех трех наборов данных.

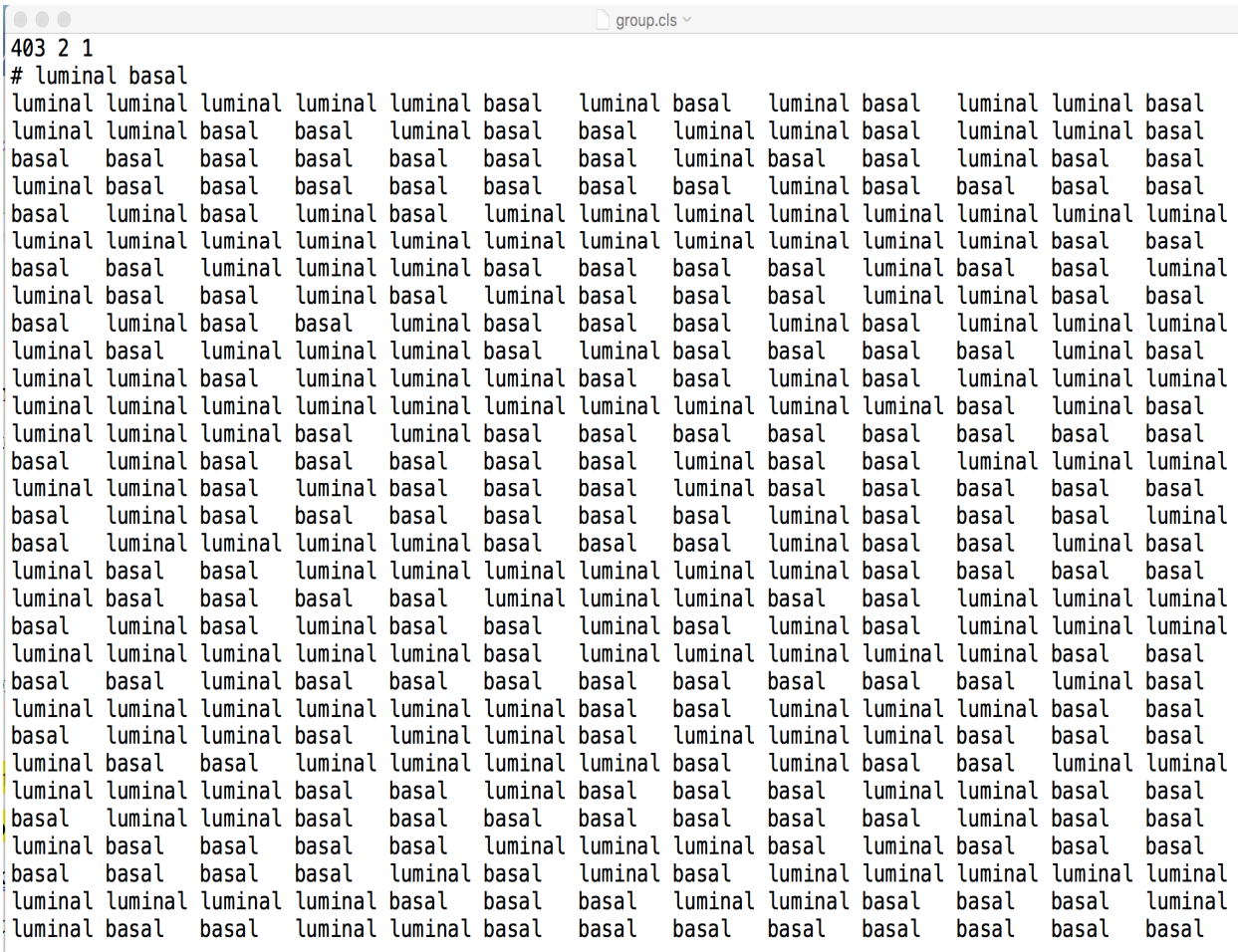
2.4 – Анализ выживаемости пациентов с мышечно-инвазивным раком мочевого пузыря

Анализ выживаемости проводился на основе кривой Каплана-Мейера (К-М) с использованием R-пакета (программы) «выживание» [187]. Логарифмический ранговый тест (логранговый критерий, log-rank test) использовался для сравнения времени выживания между двумя группами. Учитывались результаты с коэффициентом $p < 0,05$.

2.5 – Анализ обогащения по функциональной принадлежности генов, характерных для разных субтипов мышечно-инвазивного рака мочевого пузыря

Анализ обогащения по функциональной принадлежности (GSEA) — набор методов для ассоциации набора генов с изменением фенотипа [70]. Для формализации существующих данных о фенотипе такие методы часто используют базы данных предварительно аннотированных наборов генов (например, термины проекта Gene Ontology (GO): молекулярные функции, биологические процессы или клеточные компоненты [130]). Результатом применения метода (выдачей программы) в этом случае является множество преаннотированных наборов, которые помогают определить, зависит ли упорядоченный список генов от фенотипа или они просто распределены случайным образом [69]. Такие преаннотированные наборы называют перепредставленными (если частота выше фоновой) или недопредставленными (если частота ниже фоновой). Показатель обогащения (ES) представляет собой статистику, подобную Колмогорову-Смирнову, представляющую количество, до которого гены перепредставлены в верхней или недопредставлены в нижней части ранжированного списка генов.

Анализ обогащения был выполнен с использованием программы GSEA (версия 3.0), на основе Gene Ontology Biological Process (GO-BP), Kyoto Encyclopedia of Genes and Genomes (KEGG), и Hallmark annotation data set чтобы соотнести подтипы МИРМП с биологическими процессами, данные о которых представлены в этих базах. В качестве входных данных были использованы три набора генетических данных (“h.all.v5.1.symbols.gmt”, “c2.cp.kegg.v5.1.symbols.gmt” и “c5.bp.v5.1.symbols.gmt”). Набор данных экспрессии мРНК состоял из 19181 мРНК (строк) и 403 образцов опухоли (колонок). Файл меток фенотипов (cls) определяет исследуемые фенотипы и связывает каждый образец в наборе данных с одним из этих фенотипов. Полученный файл показан на рисунке 5.



```

403 2 1
# luminal basal
luminal luminal luminal luminal luminal basal luminal basal luminal basal luminal luminal basal
luminal luminal basal basal luminal basal basal luminal luminal basal luminal luminal basal
basal basal basal basal basal basal basal luminal basal basal luminal basal basal
luminal basal basal basal basal basal basal basal basal luminal basal basal basal basal
basal luminal basal luminal luminal basal luminal luminal luminal luminal luminal luminal luminal
luminal luminal luminal luminal luminal luminal basal basal basal basal luminal basal basal luminal
basal basal luminal luminal luminal basal basal basal basal basal luminal basal basal luminal
luminal basal basal luminal luminal luminal basal luminal basal basal basal basal luminal basal
luminal luminal basal luminal luminal luminal basal basal basal luminal luminal luminal luminal
luminal luminal luminal basal luminal luminal luminal luminal luminal luminal luminal luminal
luminal luminal luminal basal basal luminal basal basal basal basal basal basal basal basal
basal luminal basal basal luminal luminal basal basal basal basal basal basal basal basal basal
luminal basal basal luminal luminal luminal luminal basal luminal basal basal luminal luminal
luminal luminal luminal basal basal luminal basal basal basal basal basal basal basal basal
luminal basal basal basal basal luminal basal luminal basal luminal luminal luminal luminal
luminal luminal luminal luminal basal basal basal luminal luminal basal basal basal luminal
luminal basal basal luminal luminal basal basal basal basal basal basal basal basal basal

```

Рисунок 5 – Пример файла меток фенотипов (cls)

Примечание: на рисунке представлен фрагмент файла, который был получен после анализа (клинического и выживания) различных образцов тканей. В результате все образцы были разделены либо на люминальный, либо на базальный субтип, что в последующем использовалось в вычислениях.

Для алгоритма в GSEA были установлены следующие параметры: *permutations=1000*, *permutation type=gene_set*, *enrichment statistic=weighted*, *metric for ranking genes=Signal2Noise* (рисунок 6).

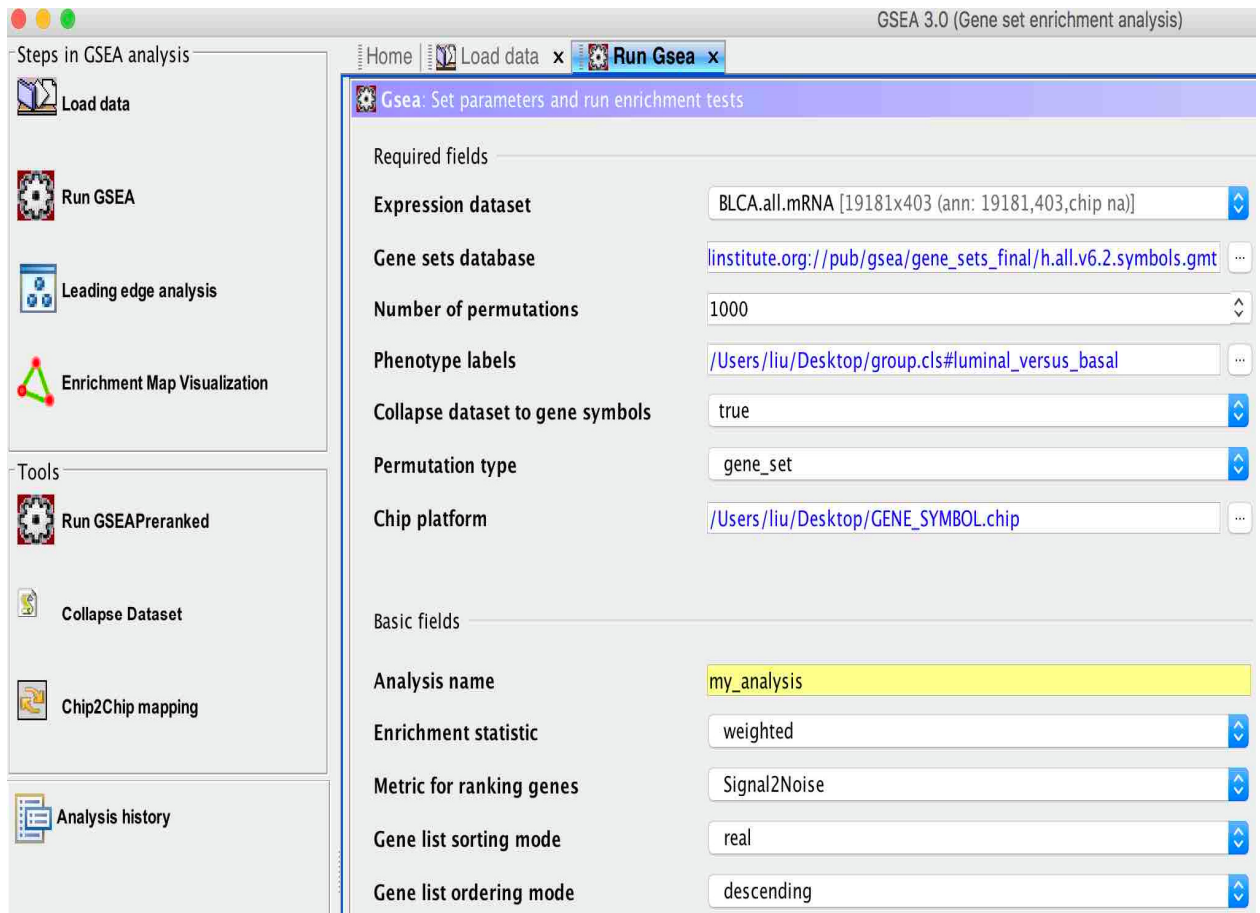


Рисунок 6 – Снимок экрана, показывающий установку параметров для программы GSEA для данного исследования

Примечание: установка параметров в программе GSEA в соответствии с необходимыми критериями повторов, выборки, метода анализа обогащения и др.

2.6 – Скрининг репрезентативных генов между подтипами с использованием подходов машинного обучения

Многие методы машинного обучения широко применяются во многих областях биологии, таких как предсказание областей, обогащенных нуклеосомами и деплецированных регионов, классификация семейств генов, предсказание гепатотоксичности, предсказание РНК метилизации, предсказание и классификация онкогенов [31, 54, 57, 117, 126, 151, 195]. Как предлагалось в предыдущих исследованиях, две модели машинного обучения - Random Forest (RF) и eXtreme Gradient Boosting (XGBoost or XG) являются мощными моделями для классификации рака [13, 38, 80].

Для алгоритмов машинного обучения RF и XGBoost были использованы наборы данных мРНК, (3000 мРНК и 403 образца опухолей), набор данных мкРНК (300 мкРНК и 403 образца опухоли) и набор данных днРНК (3000 днРНК и 403 образца опухоли). Работа выполнялась с использованием соответствующих R-пакетов «randomForest» и «XGBoost» [29, 108]. Случайным образом 403 образца МИРМП были разделены на 60% для обучающего набора и 40% тестируемого набора данных.

Точная настройка прогностической модели машинного обучения является важным первым шагом для повышения точности результатов классификации. Качество классификации оценивалось ROC-анализом. ROC-кривая (англ. receiver operating characteristic) — график, позволяющий оценить качество бинарной классификации, отображает соотношение между долей объектов от общего количества носителей признака, верно классифицированных как несущих признак, (англ. true positive rate, TPR) и долей объектов от общего количества объектов, не несущих признака,

ошибочно классифицированных как несущие признак (англ. false positive rate, FPR) при варьировании порога решающего правила. Количественную интерпретацию ROC даёт показатель AUC (англ. area under ROC curve). Чем больше площадь под кривой, тем качественнее классификатор, при этом значение $AUC=0,5$ демонстрирует непригодность выбранного метода классификации (соответствует случайному гаданию). Значение менее 0,5 свидетельствует о том, что классификатор действует с точностью до наоборот: если положительные назвать отрицательными и наоборот, классификатор будет работать лучше.

Затем мы использовали обученные модели для прогнозирования классификации подтипов МИРМП на основе тестовых наборов данных, а AUC использовался для оценки их производительности. После прогнозирования подтипов с помощью обученных моделей мы хотели выяснить, какие гены являются более репрезентативными в модели. В модели RF мы использовали $Mean\ Decrease\ Accuracy > 0$ в качестве предельного значения для скрининга репрезентативных генов. В модели XGBoost мы использовали $Gain > 0$ в качестве значения cut-off для скрининга репрезентативных генов.

Модель для машинного обучения Random Forest (в переводе с английского – «Случайный лес») основан на «деревьях решений». Понятие дерева решений опирается на понятие дерева из теории графов и понятие обучающей выборки из распознавания образов.

В основе модели машинного обучения XGBoost лежит алгоритм градиентного бустинга деревьев решений, то есть такая техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений.

2.7 – Определения различий в экспрессии генов между опухолевыми и нормальными образцами тканей мочевого пузыря

Для определения различий экспрессии генов в наборе данных пациентов с мышечно-инвазивным раком мочевого пузыря (МИРМП) был применен традиционный метод DEG (Differentially expressed genes), с использованием программного пакета «Ballgown»

Разница в экспрессии генов выявлялась между опухолью и нормальными образцами ткани мочевого пузыря. В качестве параметров частота ложных открытий FDR учитывалась при $p < 0,05$ и $|\log_2 fold change| > 0,57$ [19]. Частота ложных открытий (FDR) - это метод концептуализации уровня ошибок I типа при проверке нулевых гипотез при проведении множественных сравнений. *Log₂ fold change* – это значение cut-off (отсекания параметров), имеющее значение для расчета разницы между уровнями экспрессии, используется по модулю (без учета знака).

Затем выявленные традиционным методом определения разницы экспрессии гены сравнивались с теми, которые отличались по экспрессии с репрезентативными генами, которые были выявлены методами машинного обучения Random Forest и XGBoost. Таким образом, появилась возможность перекрытия данных для выявления генов, которые у пациентов с мышечно-инвазивным раком мочевого пузыря всегда имеют отличие в экспрессии в сравнении с относительно здоровыми контролями, что было использовано для конструирования сети взаимодействия мРНК-мкРНК-днРНК.

2.8 – Конструирование подтип-зависимой сети взаимодействий мРНК-мкРНК-днРНК при мышечно-инвазивным раке мочевого пузыря

Коэффициент корреляции Пирсона (РСС), также называемый Pearson's r – это статистическая формула, измеряющая силу корреляции между двумя переменными x и y . Коэффициент корреляции Пирсона вычислялся по формуле

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}} \quad (1)$$

где n – число образцов (пациентов), а x и y – значения экспрессии генов x и y . Коэффициент r может принимать значения между +1 и -1, где 1 – общая положительная линейная корреляция, 0 – отсутствие линейной корреляции, а -1 – общая негативная линейная корреляция. Анализ РСС выполняется для генов с различным уровнем экспрессии. Для оценки были определены пары днРНК и мкРНК, мкРНК и мРНК, днРНК и мРНК с модулем $|r| \geq 0,4$ и $p < 0,05$. Если коэкспрессия наблюдалась между днРНК и мРНК, и одновременно была коэкспрессия с мкРНК, то это определялось как мкРНК-зависимая днРНК-мРНК коэкспрессия.

Программа miRWalk2.0 – это интеграция шести самых используемых баз данных (miRWalk, miRanda, miRDB, miRNAMap, RNA22, и Targetscan), которая предоставляет самую большую из доступных коллекций, предсказанных и экспериментально подтвержденных взаимодействий мкРНК-мишеней [127].

Наши предполагаемые ко-экспрессированные взаимодействия, включая взаимодействия мРНК-мкРНК и днРНК-мкРНК, сравнивались с взаимодействиями, происходящими по данным базы miRWalk2.0. По критериям этой базы, мРНК считается истинной мишенью для мкРНК, если

их взаимодействие происходит по крайней мере в четырех базах данных среди miRWalk, miRanda, miRDB, miRNAmap, RNA22 и Targetscan. По тем же критериям днРНК принято считать истинной мишенью для мкРНК, если их взаимодействие поддерживается хотя бы в одной базе данных среди miRWalk, miRanda и Targetscan.

2.9 – Оценка плотности генной сети и биологического расстояния для генов общего вариабельного иммунодефицита и других первичных иммунодефицитов

Данные межбелковых взаимодействий (PPI) для генов ПИД и ОВИН были получены из базы данных STRING (<https://string-db.org/>). База STRING не содержит информации о механизме белок-белковых взаимодействий, а также о том, в какое время клеточного цикла может иметь место данное взаимодействие, как оно зависит от внешних условий и насколько оно тканеспецифично. Однако STRING содержит информацию обо всех возможных межбелковых взаимодействиях в данном организме, в том числе информацию, предсказанную с определённой достоверностью, что делает STRING наиболее полным ресурсом о межбелковых взаимодействиях, доступным на сегодняшний день, и особенно полезным для поиска информации о белках, не изученных экспериментально [175]. В базе мы делали выборку по критерию Homo Sapiens и установили минимальный требуемый показатель взаимодействия 0,4. Сеть данных PPI была визуализирована с использованием программного обеспечения Cytoscape (версия 3.5.1) [52]. Плотность сети ($D_{network}$) является наиболее широко используемой концепцией в исследованиях сетей регуляции генов и сетей PPI, и ее можно использовать для определения, является ли сеть плотной или нет.

Плотность сети ($D_{network}$) определяется по формуле [82]

$$D_{network} = \frac{\sum_{i=1}^n \sum_{j \neq i} a_{ij}}{n(n-1)} \quad (2)$$

где a_{ij} – это, так называемая, «парная смежность», иначе говоря наличие взаимосвязи ($a_{ij} = 0$ если между парой генов нет прямой связи или $a_{ij} = 1$, если есть), $\sum_{j \neq i} a_{ij}$ представляет «связность» i -го гена (невзвешенное сетевое соединение равно количеству генов, которые непосредственно связаны с геном i), а n - количество генов в сети. Данные РРІ для группы ОВИН (39 генов ОВИН) и 10 случайных групп генов (каждая из групп содержала 39 генов ПИД) были соответственно преобразованы в симметричную матрицу смежности (a_{ij} , $i, j = 1, \dots, n$) с помощью пакета “igraph” программы R [49]. Значение плотности сети использовалась для сравнения сплоченности генной сети. Чем больше была плотность сети в группе, тем теснее взаимодействие генов в группе.

Понятие биологической дистанции (B_{ij}) впервые было введено Itan Y. et al., 2013. Под биологическим расстоянием исследователи функциональных взаимосвязей генов в сети геномных взаимодействий подразумевают не фактическое расстояние между генами в молекуле ДНК или на хромосоме, а функциональную близость между парами генов и внутри группы генов [182]. Используя значение биологического расстояния, Itan Y. et al. показали, что ПИД-гены, как правило, располагаются в центре геномной сети человека и образуют несколько тесно взаимосвязанных подгрупп по различным биологическим механизмам [103].

Биологическая дистанция ($B_{i,j}$) определяется по формуле

$$B_{i,j} = \begin{cases} \frac{C}{S_{i,j}} & \text{if } C = 1 \\ \frac{C}{S_{i,1} + S_{1,2} + S_{2,3} + \dots + S_{C-2,C-1} + S_{C-1,j}} & \text{if } C > 1 \end{cases} \quad (3)$$

где $S_{i,j}$ представляет собой объединенную оценку между геном i и геном j , предоставленную базой данных STRING, а C является числом прямых связей между геном i и геном j . Чем меньше биологическое расстояние в группе, тем теснее функциональная биологическая взаимосвязь между генами в группе. Мы рассчитали биологическое расстояние в группе генов из 39 генов ОВИН и двух случайных групп (в каждой по 39 генов ПИД) с помощью пакета, предоставленным Itan et al., «Human gene connectome, HGC» на программном языке Python [182].

2.10 – Прогнозирование генов-кандидатов общего варибельного иммунодефицита

Были предприняты следующие три шага, чтобы предсказать гены-кандидаты ОВИН:

I. Расчет корреляционного анализа Пирсона проводился на значениях экспрессии 39 генов ОВИН и всех белок-кодирующих генов человека (или генов-кандидатов) на основе наборов данных GSE51406 и GSE72625. В параметрах установлены $|r| > 0,9$ и $p < 0,05$ в качестве пороговых значений, гены-кандидаты были выбраны из каждого из двух наборов данных. Перекрывающиеся гены-кандидаты, полученные из двух наборов данных, были использованы для последующего анализа.

II. Данные межбелкового взаимодействия (PPI) для всех белок-кодирующих генов человека были получены из публикации Cheng F. и соавт [136], в том числе 217 160 взаимодействий, предоставляемых

одиннадцатью базами данных (BioGRID [173], HI-II-14_Net [10], HPRD [83], Instruct [91], InnateDB [90], IntAct [79], MINT [123], PINA [148], SignaLink2.0 [172], KinomeNetworkX [158], PhosphositePlus [146]). Если в данных PPI имело место взаимодействие между геном ОВИН и геном-кандидатом, полученным на предыдущем этапе, ген-кандидат затем сохранялся.

III. Чтобы установить биологическую роль выявленных генов ОВИН, на основе данных KEGG (Kyoto Gene and Genomic Encyclopedia) был проведен анализ с использованием пакета R "clusterProfiler" [36].

2.11 – Определение новых генов-кандидатов общего варибельного иммунодефицита

Чтобы определить, подходит ли наш метод для предсказания генов-кандидатов ОВИН, мы вычислили биологические расстояния ($B_{i,j}$) предсказанных генов-кандидатов ОВИН и сравнили их с 39 из известных генов ОВИН.

В дальнейшем было проведено «функциональное геномное выравнивание» (functional genomics alignment, FGA) и филогенетический кластерный анализ. Эти этапы были выполнены с использованием пакета «APE», доступного в R, для оценки биологической корреляции между генами-кандидатами ОВИН и генами ОВИН [143, 182]. В частности, мы сначала создали матрицу биологической дистанции между генами ОВИН и генами-кандидатами ОВИН, а затем применили алгоритм объединения соседей (функция nj), чтобы сгенерировать филогенетическое веерообразное дерево, показывающее иерархическую кластеризацию генов-кандидатов ОВИН и известных генов ОВИН. Если гены-кандидаты ОВИН были равномерно распределены по всему диапазону известных генов ОВИН, это означает, что эти гены-кандидаты ОВИН были тесно связаны с известными

генами ОВИН. Если кандидат ОВИН и известные гены ОВИН были разделены на две или несколько групп, это означает обратное.

2.12 – Визуализация сетей молекулярного взаимодействия с помощью программы Cytoscape

Cytoscape - это программная платформа с открытым исходным кодом для биоинформатики, предназначенная для визуализации сетей молекулярного взаимодействия и интеграции с профилями экспрессии генов и другими данными. Программное обеспечение Cytoscape (версия 3.5.1) использовалось для визуализации мкРНК-зависимой сети днРНК-мРНК пациентов с мышечно-инвазивным раком мочевого пузыря и сети межбелковых взаимодействий (PPI) генов первичных иммунодефицитов (рисунок 7).

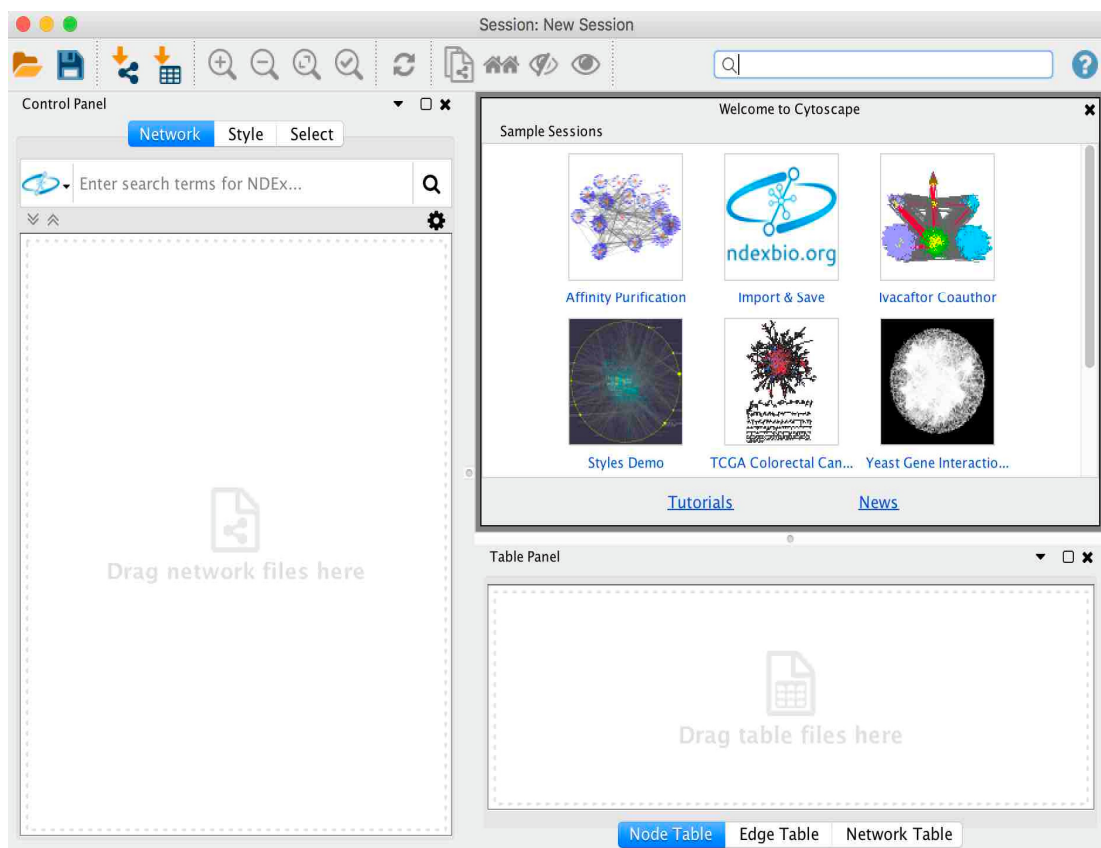


Рисунок 7 – Скриншот, демонстрирующий программное обеспечение Cytoscape, которое используется в этом исследовании

2.13 – Полногеномное секвенирование и поиск и гена-кандидата для пациента с синдромом Хеннекама

Результаты секвенса были выровнены в соответствие со стандартной последовательностью человеческого генома hg38 при помощи программного обеспечения Burrows-Wheeler Aligner (BWA) [107]. Затем для изменения порядка, сортировки и преобразования файлов SAM, использовалась программа «SAMtools» [185].

Варианты однонуклеотидный (SNVs) и вставки / делеции (инсерции) вариантные вызовы были идентифицированы с анализом генома инструментария версией 4.1.2.0 (GATK4, <http://www.broadinstitute.org/gatk/>) [180].

Чтобы уменьшить некачественные результаты, возникающие из-за ошибок картирования, использовались варианты в экзонной области с глубиной чтения $> 10\times$ и минимальным показателем качества отображения 30 с помощью программного обеспечения VCFtools [186]. Затем мы отбросили все синонимичные SNV, вставки и удаления без сдвига рамки считывания (InDels) и варианты, экзонная функция которых помечена как «NA» или «неизвестно». Кандидаты SNVs и инсерции, полученные на предыдущих стадиях, были дополнительно отфильтрованы по связи с первичным иммунодефицитом.

Кандидаты SNV (или InDels) затем разделялись на редкие или широко распространенные в зависимости от частоты минорного аллеля - MAF (minor allele frequency), которая должна была быть меньше, чем 0,01 во всех данных международного консорциума ExAC (Exome Aggregation Consortium, содержит данные более 60 тыс. экзоменов), 1000G («Тысяча геномов»), и базы gnomAD (The Genome Aggregation Database).

SNV (или Indels) далее считались вредными, если SNV (или Indels) были определены как вредные, по крайней мере, в одной из моделей среди

перечисленных ниже: Functional Analysis through Hidden Markov Models (FATHMM), Protein Variation Effect Analyzer (PROVEAN), and Combined Annotation Dependent Depletion (CADD). FATHMM и PROVEAN выполнялись с помощью программного обеспечения ANNOVAR [193], а CADD – с использованием онлайн-сервера (<https://cadd.gs.washington.edu/snv>, версия: GRCh38-v1.5).

Все потенциально патогенные SNVs и Indels были вручную проверены с использованием базы данных однонуклеотидных полиморфизмов (Single Nucleotide Polymorphism Database, dbSNP) (<https://www.ncbi.nlm.nih.gov/snp/>) и программы Интегративного просмотра генома (Integrative Genome Viewer, IGV) версии 2.4.5. Если два MAF для SNV (или INDEL), полученные из ANNOVAR и dbSNP были неоднозначны, то MAF, полученный в dbSNP рассматривался в качестве приоритетной (истинной). Мутация-кандидат определялась как истинная, если мутация, идентифицированная в GATK4, была также обнаружена в IGV.

ГЛАВА 3 – ИСПОЛЬЗОВАНИЕ МОЛЕКУЛ MIR-141-5P, MIR-141-3P, MIR-200C-3P, AC010326.3, AC073335.2, MIR100HG, CLIC4, GATA3, PALLD В КАЧЕСТВЕ ПРОГНОСТИЧЕСКИХ БИОМАРКЕРОВ У ПАЦИЕНТОВ С МЫШЕЧНО-ИНВАЗИВНЫМ РАКОМ МОЧЕВОГО ПУЗЫРЯ

3.1 – Влияние уровня инфильтрирующих иммунных клеток на прогрессирование и выживаемость при мышечно-инвазивном раке мочевого пузыря

Уровни 22 типов инфильтрирующих опухоль иммунных клеток были оценены в программе CIBERSORT. Мы также оценили, коррелируют ли уровни иммунных клеток с общей выживаемостью у пациентов с мышечно-инвазивным раком мочевого пузыря (МИРМП). Мы использовали медианный уровень в качестве отсечения, чтобы разделить образцы пациентов на группы высокого и низкого уровня, и сравнили время выживания двух групп, используя log-rank test тест. Затем использовали критерий Уилкоксона для сравнения различий в показателях иммунных клеток между двумя группами. Результаты показали, что из всех иммунологических параметров уровни опухолевых макрофагов (подтип M2) значительно коррелировали с выживаемостью пациентов, злокачественностью и патологической стадией ($p < 0,01$) (рисунок 8).

Кроме того, в нашем исследовании было подтверждено, что для более тяжелого течения МИРМП также характерно повышение параметра мутационной опухолевой нагрузки (tumor mutation burden – TMB). Так же как и повышенная экспрессия молекул, характерных для макрофагов M2 (в данном случае – опухолевых макрофагов), так и повышение TMB в целом соотносится с типическим патологическим процессом опухолевого роста,

поскольку известно, что эти параметры коррелируют с тяжестью опухолевого процесса. Это объясняется как функционально, так и подтверждается морфологически.

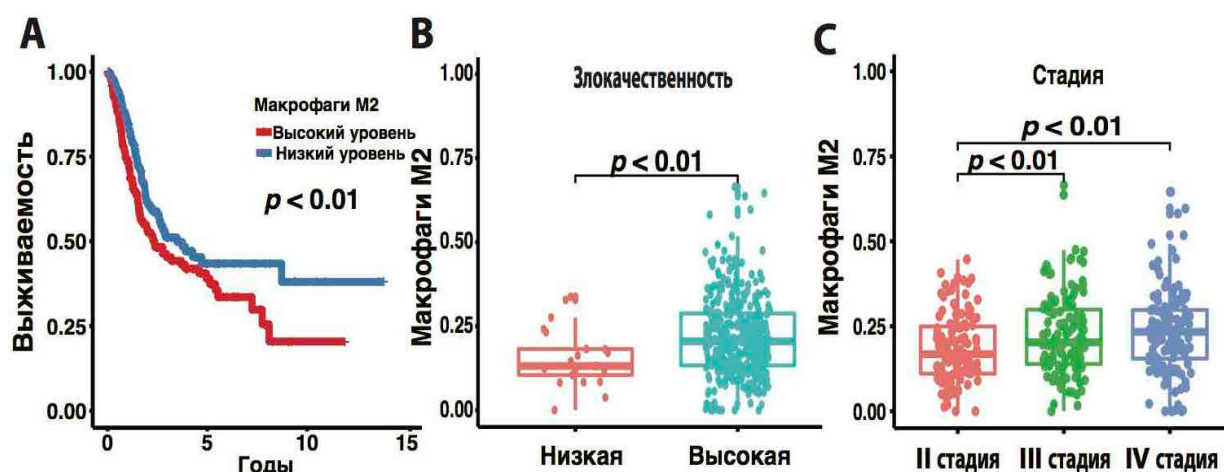


Рисунок 8 – Графики кривой Каплана-Мейера для M2-макрофагов (А) и уровень M2 макрофагов у пациентов с мышечно-инвазивным раком мочевого пузыря (В, С)

3.2 – Кластерный анализ данных экспрессии РНК пациентов с мышечно-инвазивным раком мочевого пузыря

Для выделения молекулярных подтипов МИРМП был применен кластерный анализ.

Чтобы определить соответствующие кластеры, был выполнен консенсус-кластерный анализ (Consensus Clustering (CC) analysis) для наборов данных мРНК, мкРНК и днРНК с использованием программного пакета языка R “ConsensusClusterPlus”, в котором параметры были настроены следующим образом: *reps=1000*, *innerLinkage=complete*, *clusterAlg=hc*, *K=6*, *distance=pearson* [198]. В качестве примера набор данных кластеризации показан в таблице 5. В колонках кроме идентификатора образца указывалась группа кластера к которой относился образец при разбиении на два, три или более разнообразных кластеров.

Алгоритмы оценки средней ширины силуэта (ASW) были выполнены для оценки оптимальных кластеров (K) с использованием пакета R «fpc» [170]. Обычно мы можем вывести оптимальный K на основе наибольшего значения ASW [78]. Оптимальный K , предложенный ASW в наборах данных мРНК, мкРНК и днРНК, равен 2 (рисунок 9).

Таблица 5 – Кластеры, сгенерированные с помощью консенсус-кластерного анализа

Sample\Cluster	Cluster=2	Cluster=3		Cluster=6
TCGA-2F-A9KO	2	2		1
TCGA-2F-A9KP	2	3		2
.....				
.....				
TCGA-4Z-AA7Y	2	1		3
TCGA-4Z-AA80	2	2		6

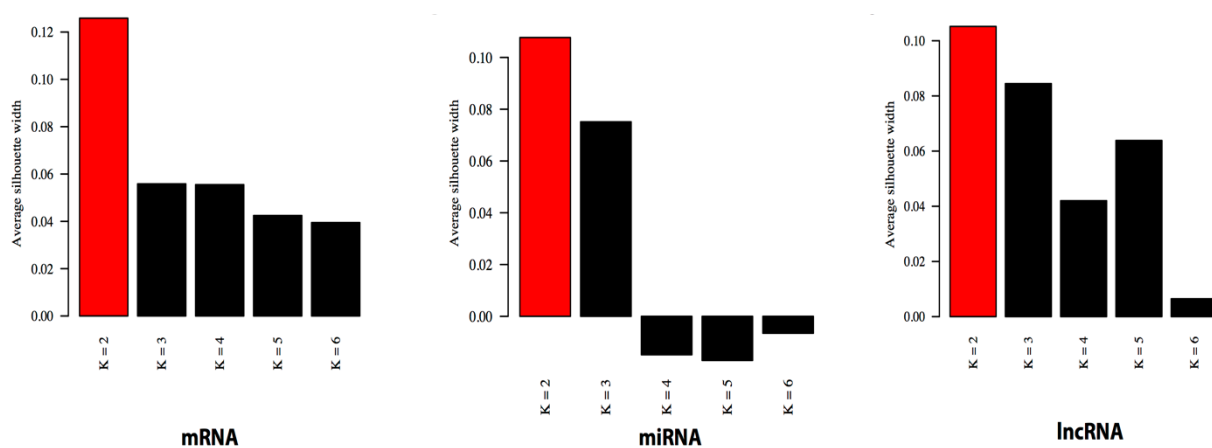


Рисунок 9 – Результат оценки алгоритмом ASW оптимального количества кластеров в данном наборе данных

Примечание: красным представлен столбец $K=2$, который означает, что для соответствующих наборов данных оптимальное число кластеров равно двум.

Анализ кластеризации кластеров (Cluster of Cluster (CoC)) использует кластерные назначения, полученные при анализе отдельных молекулярных наборов данных, для получения единого решения, таким образом объединяя результаты анализа наборов данных мРНК, мкРНК и днРНК. Сначала мы подготавливали набор двоичных данных на основе результатов кластеризации, когда $K=2$ (таблица 6).

Таблица 6 – Набор двоичных данных, используемый для анализа кластеризации кластеров

Sample\Dataset	mRNA1	mRNA2	miRNA1	miRNA2	lncRNA1	lncRNA2
TCGA-2F-A9KO	0	1	0	1	1	0
TCGA-2F-A9KP	0	1	0	1	0	1
TCGA-2F-A9KQ	1	0	1	0	0	1
.....						
.....						
TCGA-4Z-AA7Y	0	1	1	0	1	0
TCGA-4Z-AA80	1	0	1	0	1	0

Примечание: 0 – порядковый номер первого кластера, 1 – второго.

Затем мы выполнили консенс-кластеризацию для двоичного набора данных при условии, что $K=2$, чтобы сгенерировать окончательное назначение кластера (таблица 7).

Таблица 7 – Конечные кластеры, полученные в результате анализа кластеризации кластеров

Sample	Cluster
TCGA-2F-A9KO	1
TCGA-2F-A9KP	2
TCGA-2F-A9KQ	1
.....	
.....	
.....	
TCGA-4Z-AA7Y	2
TCGA-4Z-AA80	1

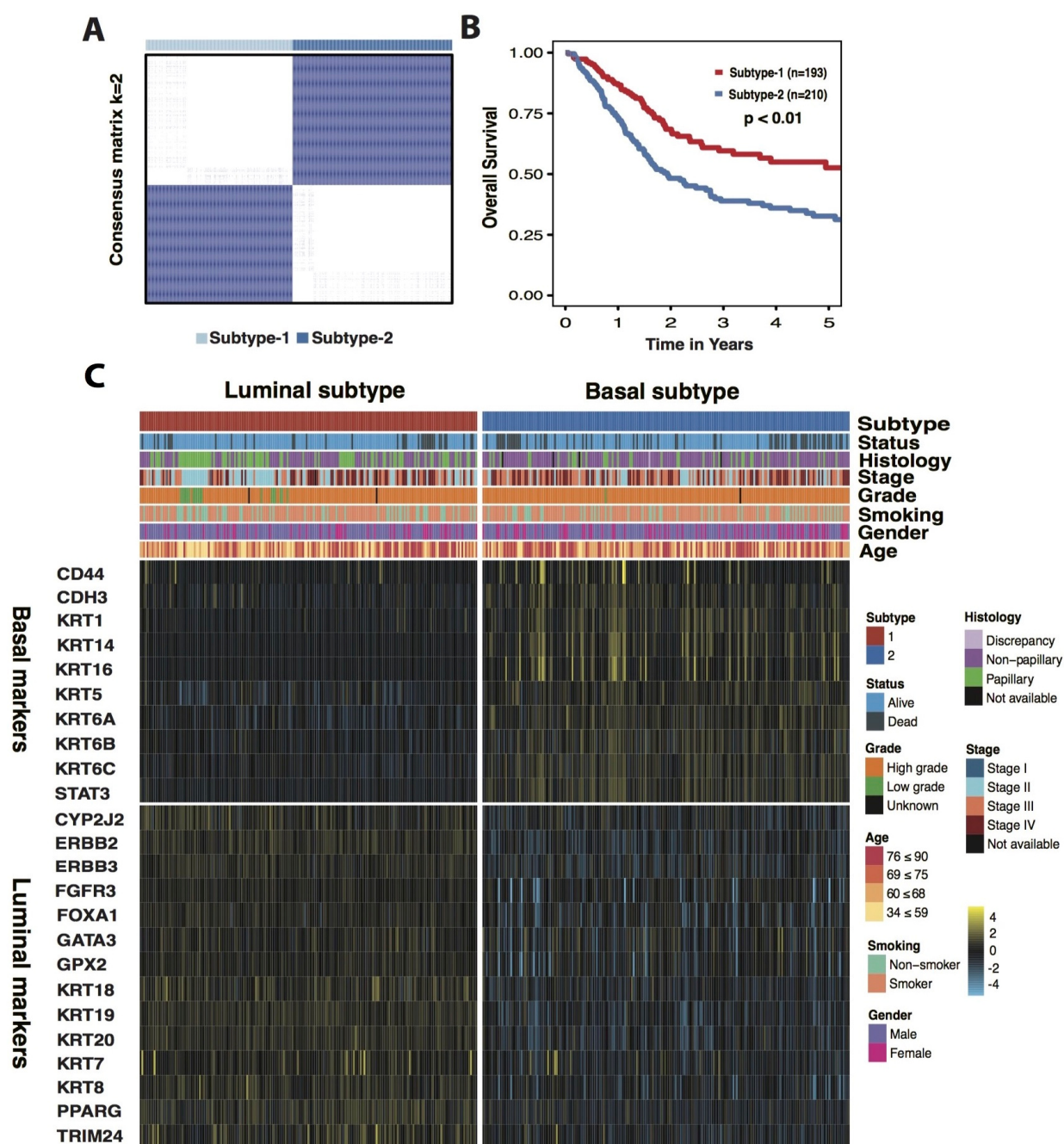


Рисунок 10 – Разделение мышечно-инвазивного рака мочевого пузыря на подтип-1 и подтип-2

Примечание: (A) иерархически кластеризованная тепловая карта для набора данных CoC ($K=2$). (B) график К-М для общей 5-летней выживаемости подтипа-1 и подтипа-2 (подтип-1=193, подтип-2=210, log-rank test, $p<0,05$). (C) тепловая карта изображает профили экспрессии базальных (вверх) и люминальных (низ) биомаркеров в подтипе-1 (слева) и подтипе-2 (справа). Аннотация и соответствующие цвета показывают некоторые клинические особенности. Желтый и бирюзовый цвета соответствуют высокому и низкому относительному уровню экспрессии. Исходное значение экспрессии было log2-преобразовано.

В результате кластерного анализа 403 образцов опухоли МИРМП были разделены на два стабильных подтипа. Иерархически кластеризованная «тепловая» карта подтипа-1 и подтипа-2 показана на рисунке 10А.

Анализ выживаемости по методу Каплан-Мейера (КМ) показал, что общая 5-летняя выживаемость для подтипа-1 (кривая красного цвета) составляла 55%, а для подтипа-2 (кривая голубого цвета) – 30 %, что свидетельствует о значительном различии в клиническом прогнозе (log-rank test, $p < 0,05$, рисунок 10В).

На рисунке 10С представлена подробная карта, где каждый из 403 образцов был разделен на два подкласса, при этом для базального и люминального подтипа определены маркеры, клинические показатели. Подтип-1 характеризовался высокой экспрессией люминальных маркеров, таких как CYP2J2, ERBB2 и KRT18, в то время как подтип 2 характеризовался высокой экспрессией базальных маркеров, таких как CD44, CDH3 и KRT1.

Ген KRT1 кодирует протеин керотин 1, который характерен для цитоплазмы в дифференцирующихся клетках сквамозного эпителия. Обнаружение этого маркера в тканях внутренних органов как правило свидетельствует о трансформации эпителиальной ткани и онкотрансформации. Повышенная экспрессия мРНК KRT1 при базальном субтипе – яркий маркер, свидетельствующий о тяжелом злокачественном процессе, что в целом соответствует базальным типам опухолей и может подтверждать правильность методики, благодаря которой были выявлены другие маркеры.

Повышенная экспрессия мРНК CDH3 при базальном субтипе МИРМП означает повышенную экспрессию Р-кадгерина, который является классической молекулой межклеточной адгезии с гомеостатическим влиянием во многих тканях в норме. Однако его поведение в условиях злокачественного роста заметно зависит от клеточного контекста. В

некоторых моделях опухолей, таких как меланома и плоскоклеточный рак полости рта, Р-кадгерин действует как супрессор опухолей, поскольку его отсутствие связано с более агрессивным фенотипом раковых клеток. В то же время, избыточная экспрессия этой молекулы связана со значительными эффектами, стимулирующими опухоль, в новообразованиях молочной железы, яичников, предстательной железы, эндометрия, кожи, желудка, поджелудочной железы и толстой кишки. Повышенная экспрессия CDH3, выявленная при базальном субтипе МИРМП, свидетельствует о том, что его уровень в данном случае коррелирует с более тяжелым течением рака мочевого пузыря, так же как при раке молочной железы или простаты.

CD44 — интегральный клеточный гликопротеин, играющий важную роль в межклеточных взаимодействиях, клеточной адгезии и миграции.

CYP2J2 — эпоксигеназа цитохрома P450, которая превращает арахидоновую кислоту в четыре региоизомерные эпоксиэйкозатриеновые кислоты. Известно, что CYP2J2 сверхэкспрессируется в раковых тканях человека и линиях раковых клеток. Считается, что эпоксиэйкозатриеновые кислоты усиливают рост опухоли, увеличивают пролиферацию карциномы и предотвращают апоптоз раковых клеток.

Ген KRT18 кодирует цепочки промежуточных филаментов кератина 18 первого типа, который входит в состав большинства тканей. Известно, что его экспрессия усиливается в опухолевой ткани, но какую-либо специфичность исследователям определить не удастся.

Ген ERBB2 кодирует ERBB2 рецептор тирозинкиназ. Обычно ген ERBB2 усиливается или сверхэкспрессируется приблизительно в 30% случаев рака молочной железы человека и во многих других типах рака, включая рак яичников, желудка, мочевого пузыря, слюнных желез и при раке легких.

Обнаружение известных маркеров, характерных для той или иной тяжести онкологического процесса, указывает на то, что метод,

используемый для типирования рака на два подтипа - базальный и люминальный, используемый в данном исследовании, – работоспособен и его можно использовать для идентификации других неизвестных маркеров разных субтипов схожих видов рака.

Таблица 8 – Соотношение между клиническими параметрами и подтипами мышечно-инвазивного рака мочевого пузыря (Критерий χ^2)

Переменные	Пациентов	p-значение
Подтип		
базальный	210	
люминал	193	
Стадия		<0,001
I	2	
II	129	
III	138	
IV	132	
недоступно	2	
Возраст		0,4498
34 ≤ 59	87	
60 ≤ 68	112	
69 ≤ 75	95	
76 ≤ 90	109	
Пол		0,0612
мужчина	297	
женский	106	
Курение		0,3345
курильщик	294	
некурящий	109	
Статус		0,0173
в живых	298	
мертв	105	
Гистология		<0,001
папиллярный	128	
плоский	270	
нет данных	5	
Степень злокачественности		<0,001
высокая	380	
низкая	20	
нет данных	3	

В этом исследовании критерий χ^2 использовался для проверки независимости выборок информации о подтипах и клинических показателями пациентов с МИРМП. Гистология, стадия рака, степень злокачественности и выживаемость значительно различались между двумя подтипами ($p < 0,05$) (таблица 8).

Подтип-1 и подтип-2 напоминают люминальный и базальный подтип соответственно, как по кривым выживаемости, так и по биомаркерам и клиническим показателям [85]. Таким образом, подтип-1 и подтип-2 были повторно определены как люминальный и базальный подтипы соответственно.

Исследована взаимосвязь между подтипами и маркерами иммуносупрессивных молекул, которые важны для определения эффективности иммунотерапии. Было обнаружено, что паттерны экспрессии молекул контрольных точек иммунного ответа (например, PD-1, PD-L1, CTLA-4, HAVCR-2 и LAG-3) позволяют различать базальные и люминальные подтипы рака мочевого пузыря, и базальные опухоли демонстрируют более высокие уровни экспрессии молекул контрольных точек иммунного ответа, чем люминальный подтип (Т-критерий Вилкоксона, $p < 0,01$) (рисунок 11).

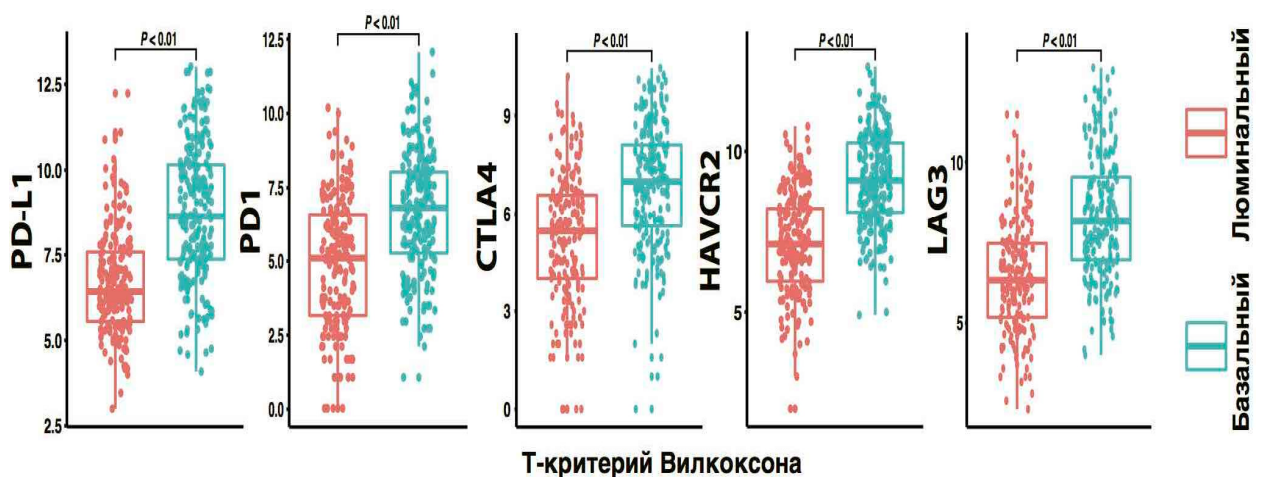


Рисунок 11 – Уровня экспрессии иммуносупрессивных молекул при базальном и люминальном подтипах

Примечание: исходное значение экспрессии было $\log_2$ -преобразовано.

3.3 – Анализ обогащения по функциональной принадлежности данных пациентов с мышечно-инвазивным раком мочевого пузыря

Анализ обогащения по функциональной принадлежности (GSEA – gene set enrichment analysis) показал, что высоко экспрессируемые сигнальные пути в люминальном подтипе, в основном, участвуют в метаболизме (например, метаболизм ретинола, метаболизм линолевой кислоты, метаболизм желчной кислоты, окислительное фосфорилирование, метаболизм препарата цитохром P450, катаболизм монокарбоновых кислот, катаболизм жирных кислот, бетта-окисление жирных кислот, цепи переноса электронов, катаболизм органических кислот, сборка комплекса дыхательной цепочки в митохондриях, окисление липидов, рибосомы, метаболизм альфа-линоленовой кислоты, метаболизм ксенобиотиков цитохрома P450, деградация валин лейцина и изолейцина, окислительное фосфорилирование, пероксисомы, липогенез, и метаболизм жирных кислот) (таблица 9).

В то же время высоко экспрессируемые пути в базальном подтипе в основном связаны с процессами в иммунной системе (например, аутоиммунный тиреоидит, организация внеклеточной структуры, отторжение аллотрансплантата, передача сигналов mTORC1 и передача сигналов TNF- α через NF-kB, обработку антигена и представление экзогенного пептидного антигена с помощью класса I MHC, обработка антигена и презентации, обработка антигена и представление пептидного антигена, обработку антигена и представление пептидного антигена с помощью класса I MHC, регуляция врожденного иммунного ответа, и положительного регулирования защитной реакции), метастазирование (фокальная адгезия и взаимодействие цитокин-цитокиновый рецептор, и цитокина опосредованного сигнального пути) и эпителиально-мезенхимальный переход ($|NES| > 1$ and $FDR\ q\text{-val} < 0,25$) (таблица 10).

Таблица 9 – Рейтинг значений из баз данных GO-BP, KEGG и наборы генов Hallmark для люминального подтипа

Имя Gene Set	Размер	NES	FDR q
GO-BP			
GO катаболизм монокарбоновых кислот	95	2,5151	0
GO окислительное фосфорилирование	73	2,4333	0
GO катаболизм жирных кислот	80	2,4284	0
GO бета-окисление жирных кислот	51	2,3211	0
GO цепи переноса электронов	78	2,3116	0
GO катаболизм органических кислот	92	2,2781	2.26E-04
GO сборка комплекса дыхательной цепочки в митохондриях	42	2,1788	0,0014
GO окисление липидов	63	2,1586	0,0015
GO биогенез комплекса I дыхательной цепочки в митохондриях	199	2,1449	0,0018
GO создание белковой локализации в эндоплазматическом ретикулуме	70	2,1285	0,0022
KEGG			
KEGG рибосомы	87	2,3289	0
KEGG метаболизм альфа-линоленовой кислоты	19	2,077	5.29E-04
KEGG метаболизм ксенобиотиков цитохрома P450	68	2,0424	5.36E-04
KEGG деградация валин лейцина и изолейцина	44	1,9767	0,00137
KEGG метаболизм препарата цитохром P450	70	1,9727	0,0012
KEGG окислительное фосфорилирование	116	1,9372	0,00209
KEGG пероксисомы	78	1,932	0,00214
KEGG метаболизм жирных кислот	42	1,9184	0,00229
KEGG метаболизм ретинола	63	1,8697	0,00391
KEGG метаболизм линолевой кислоты	29	1,8393	0,00482
Наборы генов Hallmark			
Hallmark окислительное фосфорилирование	198	1,5145	0,0583
Hallmark метаболизм желчной кислоты	112	1,411	0,07668
Hallmark пероксисомы	103	1,4095	0,05174
Hallmark липогенез	191	1,3794	0,05125
Hallmark метаболизм жирных кислот	156	1,2596	0,11892

Примечание: NES=Нормализованные значения обогащения; GO-BP = биологические процессы из Онтологии генов; KEGG = Киотская энциклопедия генов и геномов; Размер - число генов в наборе генов. Положительное NES означает, что гены перепредставленные гены в наборе данных люминального подтипа с повышенной экспрессией.

Таблица 10 – Топ-рейтинг категории GO-BP, KEGG и Hallmark наборов генов для базального подтипа

Имя Gene Set	Размер	NES	FDR q
GO-BP			
GO внеклеточной структура организации	297	-2,8256	0
GO обработку антигена и представление экзогенного пептидного антигена с помощью класса I MHC	65	-2,7258	0
GO обработка антигена и презентации	206	-2,6334	0
GO обработка антигена и представление пептидного антигена	170	-2,6246	0
GO обработку антигена и представление пептидного антигена с помощью класса I MHC	90	-2,6134	0
GO хондроитин сульфат процесс биосинтеза	25	-2,6008	0
GO коллагена организации фибрилл	36	-2,5958	0
GO регуляция врожденного иммунного ответа	349	-2,5825	0
GO положительного регулирования защитной реакции	360	-2,5802	0
GO цитокина опосредованного сигнального пути	440	-2,5675	0
KEGG			
KEGG фокальная адгезия	197	-2,6862	0
KEGG цитокин-цитокиновое рецепторное взаимодействие	257	-2,5127	0
KEGG взаимодействие с рецептором ECM	84	-2,512	0
KEGG протеосома	43	-2,4802	0
KEGG лейшманиоз	69	-2,4718	0
KEGG вирусный миокардит	68	-2,4178	0
KEGG линия гемопоэтических клеток	85	-2,4134	0
KEGG регуляция актина цитоскелета	211	-2,3911	0
KEGG отторжение аллотрансплантата	35	-2,3902	0
KEGG аутоиммунный тиреоидит	50	-2,3778	0
Наборы генов Hallmark			
Hallmark эпителиально-мезенхимальный переход	197	-3,2473	0
Hallmark воспалительный ответ	197	-3,019	0
Hallmark интерферон гамма ответ	197	-2,9964	0
Hallmark интерферон альфа ответ	94	-2,9491	0
Hallmark отторжение аллотрансплантата	199	-2,901	0
Hallmark G2M	194	-2,6389	0
Hallmark E2F	196	-2,6177	0
Hallmark передача сигналов TNF-а через NF-kB	198	-2,5512	0
Hallmark комплемент	195	-2,5475	0
Hallmark передача сигналов mTORC1	198	-2,441	0

Примечание: NES= Нормализованные значения обогащения; GO-BP = биологические процессы из Онтологии генов; KEGG = Киотская энциклопедия генов и геномов; Размер - число генов в наборе генов. Отрицательное NES означает, что гены перепредставленные гены в наборе данных базального подтипа с повышенной экспрессией.

3.4 – Поиск дифференциально экспрессируемых генов при мышечно-инвазивном раке мочевого пузыря традиционными методами и с помощью машинного обучения

Анализ дифференциальной экспрессии традиционным методом с использованием программного пакета «Ballgown» показал, что 4167 мРНК (2314 с повышенной экспрессией и 1853 с пониженной), 208 мкРНК (148 с повышенной экспрессией и 60 с пониженной) и 2488 днРНК (1402 с повышенной экспрессией и 1086 с пониженной) были дифференциально экспрессированы между опухолевыми и нормальными образцами тканей мочевого пузыря (рисунок 12).

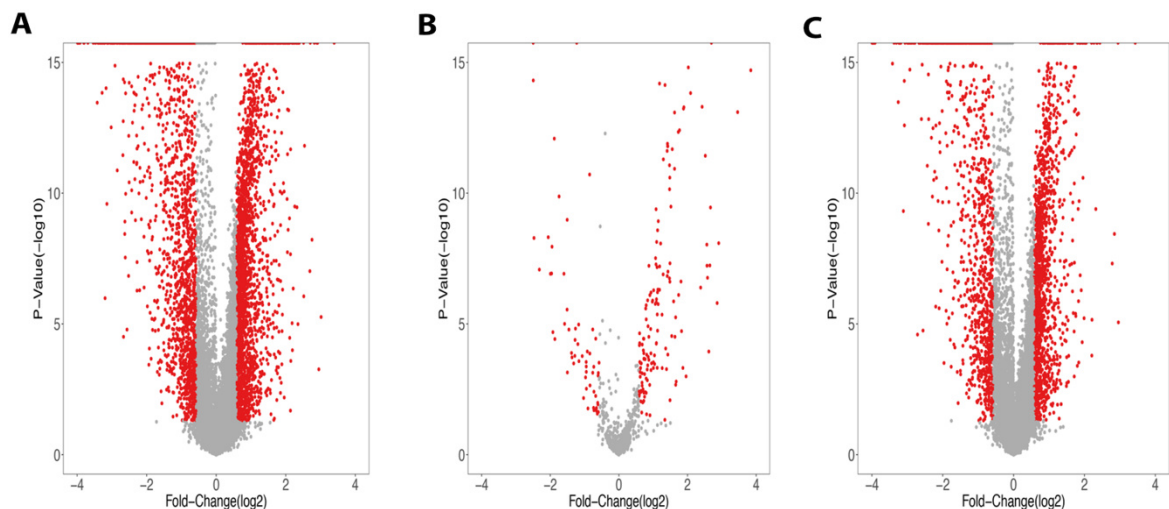


Рисунок 12 – Результат анализа разницы экспрессии генов между пациентами с мышечно-инвазивным раком мочевого пузыря и здоровыми контролями, в том числе 4167 мРНК (А), 208 мкРНК (В) и 2488 днРНК (С)

Примечание: скорректированное значение $p < 0,05$ и $|\log_2 \text{fold change}| > 0,57$.

Машинное обучение – наиболее эффективный метод для определения критически важных генов и классификации их на подтипы в зависимости от паттернов экспрессии мРНК, мкРНК, и днРНК. Первым шагом в машинном

обучении является тренировка имеющихся моделей машинного обучения. Мы использовали две модели – Random Forest (RF) и XGBoost (XG).

Модель для машинного обучения Random Forest (в переводе с английского – «Случайный лес») основан на «деревьях решений». Понятие дерева решений опирается на понятие дерева из теории графов и понятие обучающей выборки из распознавания образов. Идея алгоритма RF состоит в разбиении множества переменных на непересекающиеся множества и подгонку простой модели для каждого такого множества. Для RF мы настроили оптимальное значение «*mtry*» (число признаков) в соответствии с большой площадью под значением ROC-кривой, а затем настроили значение «*ntree*» (количество деревьев решений) в соответствии с небольшими значениями ошибки классификации «*Oob Error Rates*» (*Oob* – метод оценки вероятности ошибочной классификации алгоритмом Random Forest).

В основе модели машинного обучения XGBoost лежит алгоритм градиентного бустинга деревьев решений, то есть такая техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений. Для модели машинного обучения XGBoost параметр «*iter*» (количество итераций, то есть повторов) был настроен в соответствии с методом 10-кратной перекрестной проверки. Параметры “*Mean Decrease Accuracy*” RF и “*Gain*” модели машинного обучения XGBoost используются для идентификации репрезентативных генов. Кроме того, для XGBoost использовались следующие важные параметры: *gamma*=1, *min_child_weight*=1, *max_depth*=14, *nrounds*=2000.

В наборе данных мРНК в качестве оптимальных параметров мы получили *mtry*=9 и *ntree*=300 для модели RF и *iter*=2000 в качестве оптимального параметра для модели XGBoost (рисунок 13).

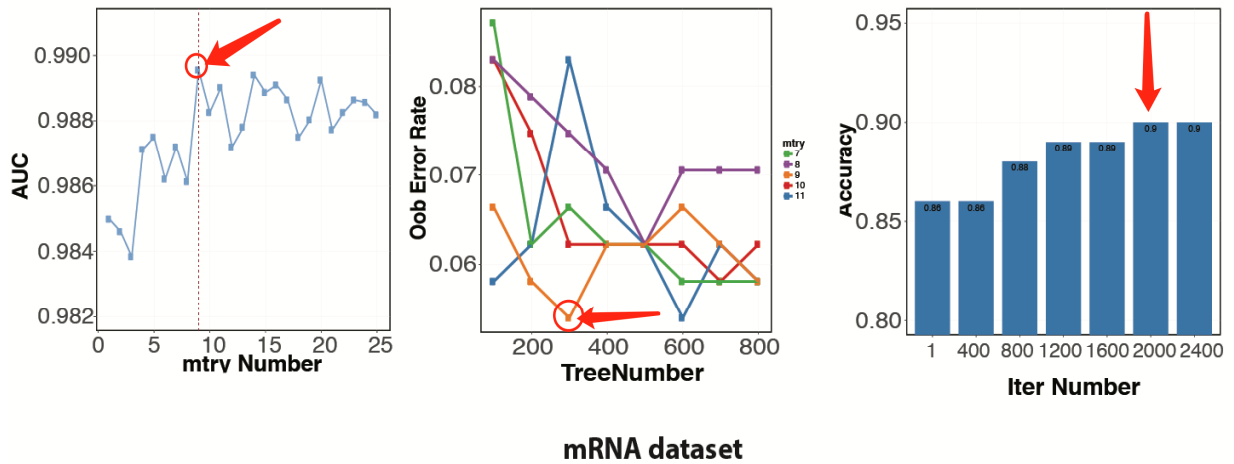


Рисунок 13 – Выбор параметров Random Forest и XGBoost для набора данных мРНК

Примечание: указано определение оптимальных параметров $mtry$ и $ntree$ для настройки моделей машинного обучения Random Forest и $iter$ для моделей машинного обучения XGBoost

В наборе данных мкРНК мы получили $mtry=7$, $ntree=300$ в качестве оптимальных параметров модели RF и $iter=2000$ в качестве оптимального параметра модели XGBoost (рисунок 14).

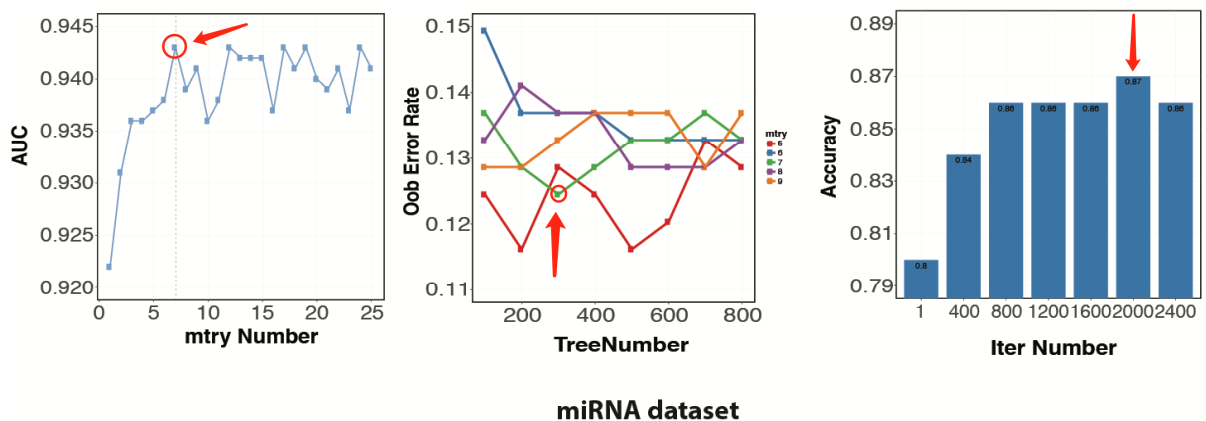


Рисунок 14 – Выбор параметров Random Forest и XGBoost в наборе данных мкРНК

Примечание: указано определение оптимальных параметров $mtry$ и $ntree$ для настройки моделей машинного обучения Random Forest и $iter$ для моделей машинного обучения XGBoost.

В наборе данных днРНК мы получили $mtry=16$, $ntree=200$ в качестве оптимальных параметров для RF-модели и $iter=1600$ в качестве оптимального параметра для модели XGBoost (рисунок 15).

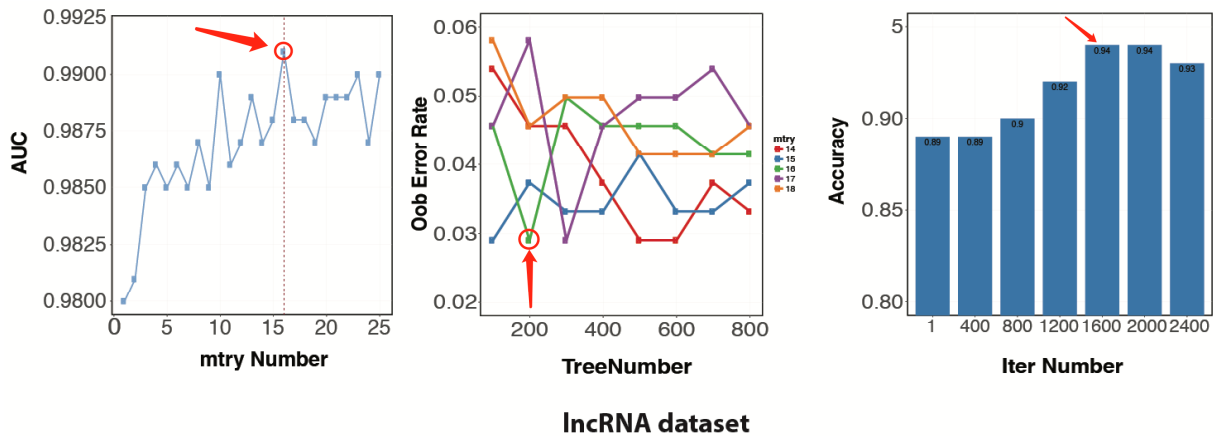


Рисунок 15 – Выбор параметров Random Forest и XGBoost в наборе данных днРНК

Примечание: указано определение оптимальных параметров mtry и ntree для настройки моделей машинного обучения Random Forest и iter для моделей машинного обучения XGBoost.

ROC-кривая позволяет оценить качество бинарной классификации и отображает соотношение между долей объектов от общего количества носителей признака, верно классифицированных как несущие признак, и долей объектов от общего количества объектов, не несущих признака, ошибочно классифицированных как несущие признак.

Площадь под ROC-кривой показала, что модели RF и XGBoost имели значения AUC более 90% в трех наборах данных, что свидетельствует о высокой качественности выбранной классификации (рисунок 16).

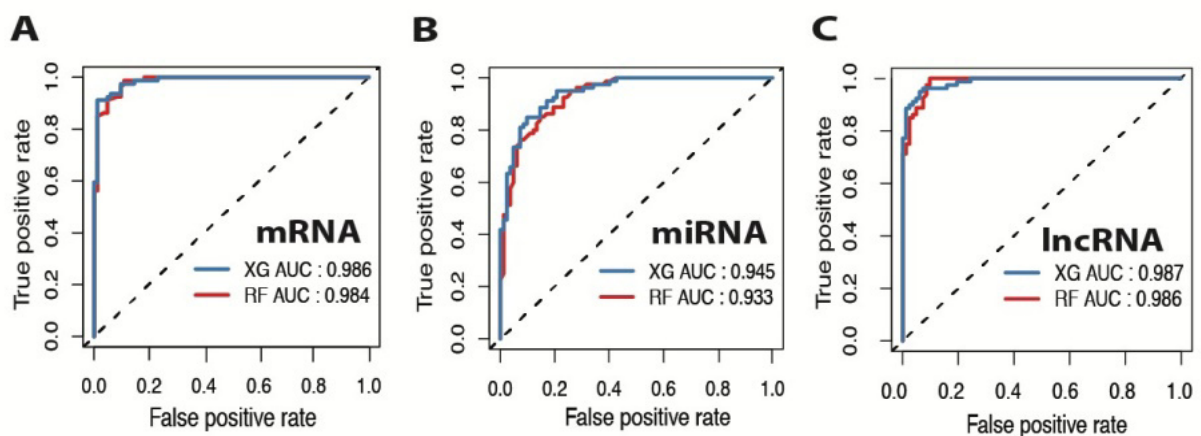


Рисунок 16 – Оценка качества дифференциальной экспрессии генов моделями машинного обучения Random Forest и XGBoost в наборах данных мРНК (А), мкРНК (В) и днРНК (С)

Примечание: голубой линией обозначена ROC-кривая для модели машинного обучения XGBoost, красной для модели машинного обучения Random Forest.

Перекрытие результатов традиционного анализа оценки разницы в экспрессии генов с использованием программного пакета «Ballgown» и репрезентативными генами, полученными на выходе из двух типов машинного обучения включали 57 днРНК, 120 мкРНК и 278 мРНК, которые использовались для последующего анализа (рисунок 17).

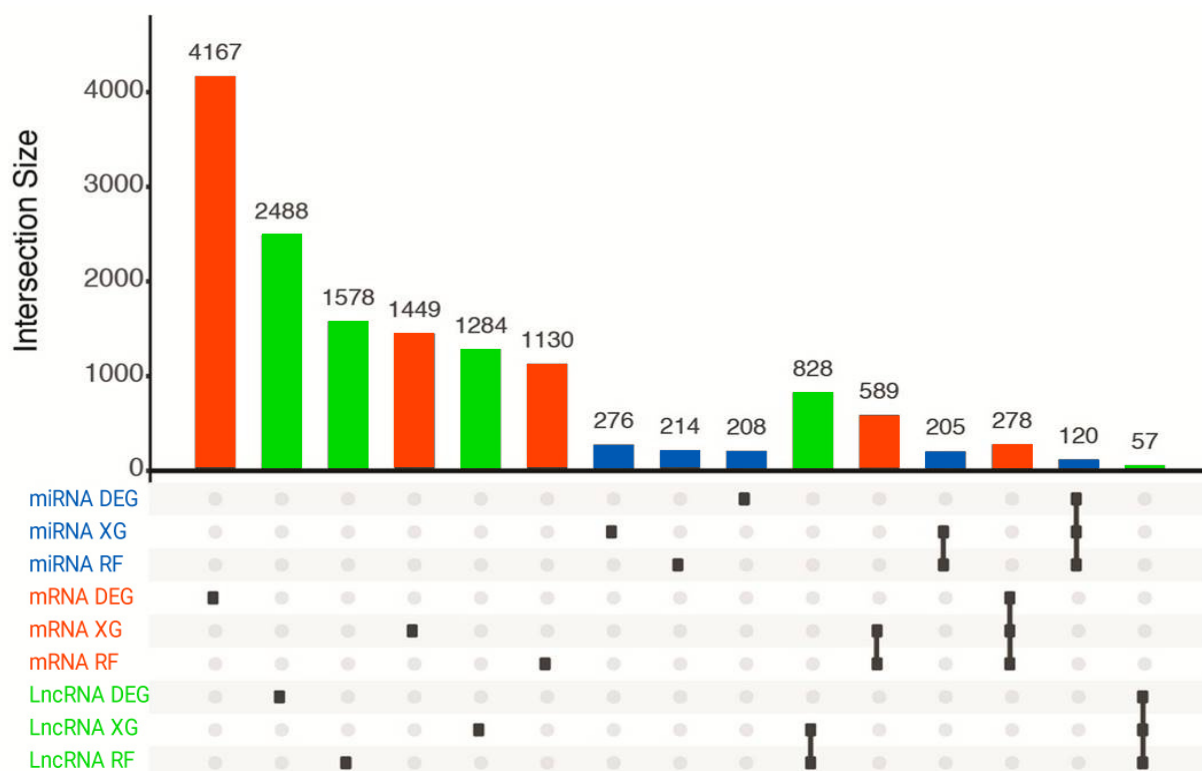


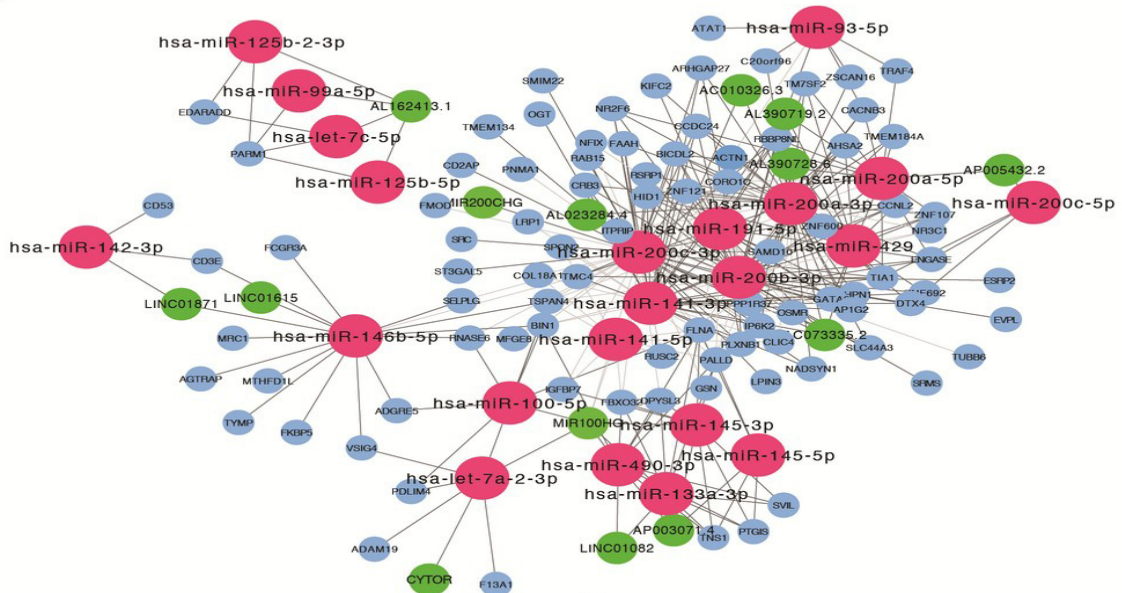
Рисунок 17 – Диаграмма, показывающая перекрытие данных между традиционным анализом различий экспрессии генов, характерной для мышечно-инвазивного рака мочевого пузыря, и результатами, полученными с использованием моделей машинного обучения Random Forest и XGBoost

3.5 – Конструирование подтип-зависимой мРНК-мкРНК-днРНК сети для мышечно-инвазивного рака мочевого пузыря

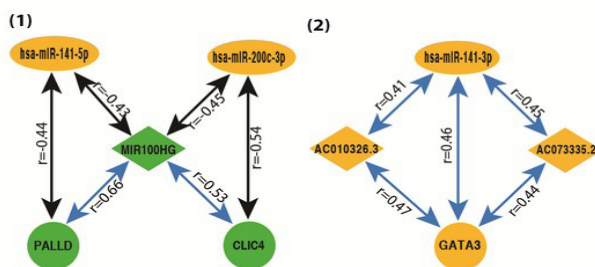
Коэффициент корреляции Пирсона ($|r| \geq 0,4$ и $p < 0,05$) и программное обеспечение Cytoscape позволили сконструировать мкРНК-зависимую сеть ко-экспрессии мРНК-днРНК, которая состояла из 90 мРНК, 22 мкРНК и 14 днРНК (рисунок 18А). С использованием базы данных miRWalk мы

обнаружили 4 пересекающиеся (перекрестные) взаимозависимости между мРНК, мкРНК и днРНК.

A



B



C

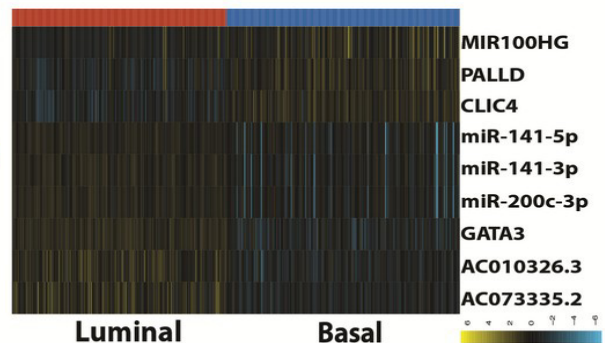


Рисунок 18 – Характерные для подтипов рака перекрестные взаимозависимости мРНК, мкРНК и днРНК

Примечание: (A) сеть ко-экспрессии для мкРНК-опосредованных взаимодействий мРНК-днРНК. Узлы малинового цвета - мкРНК, зеленые узлы - днРНК, а небесно-голубые узлы - мРНК. (B) Перекрестные взаимозависимости молекул мРНК, мкРНК и днРНК, подтвержденные программой mirWalk 2.0. Зеленым цветом обозначены молекулы с пониженной экспрессией в опухоли по сравнению с нормальным образцом, тогда как желтый цвет соответствует повышенной экспрессии в опухоли. Синие линии представляют положительные корреляции, а черные - отрицательные. (C) Тепловая карта, отображающая уровень экспрессии девяти описанных РНК в базальном и люминальном подтипах. Желтый и бирюзовый цвета соответствуют высокому и низкому относительному уровню экспрессии соответственно. Исходное значение экспрессии было $\log_2$ -преобразовано.

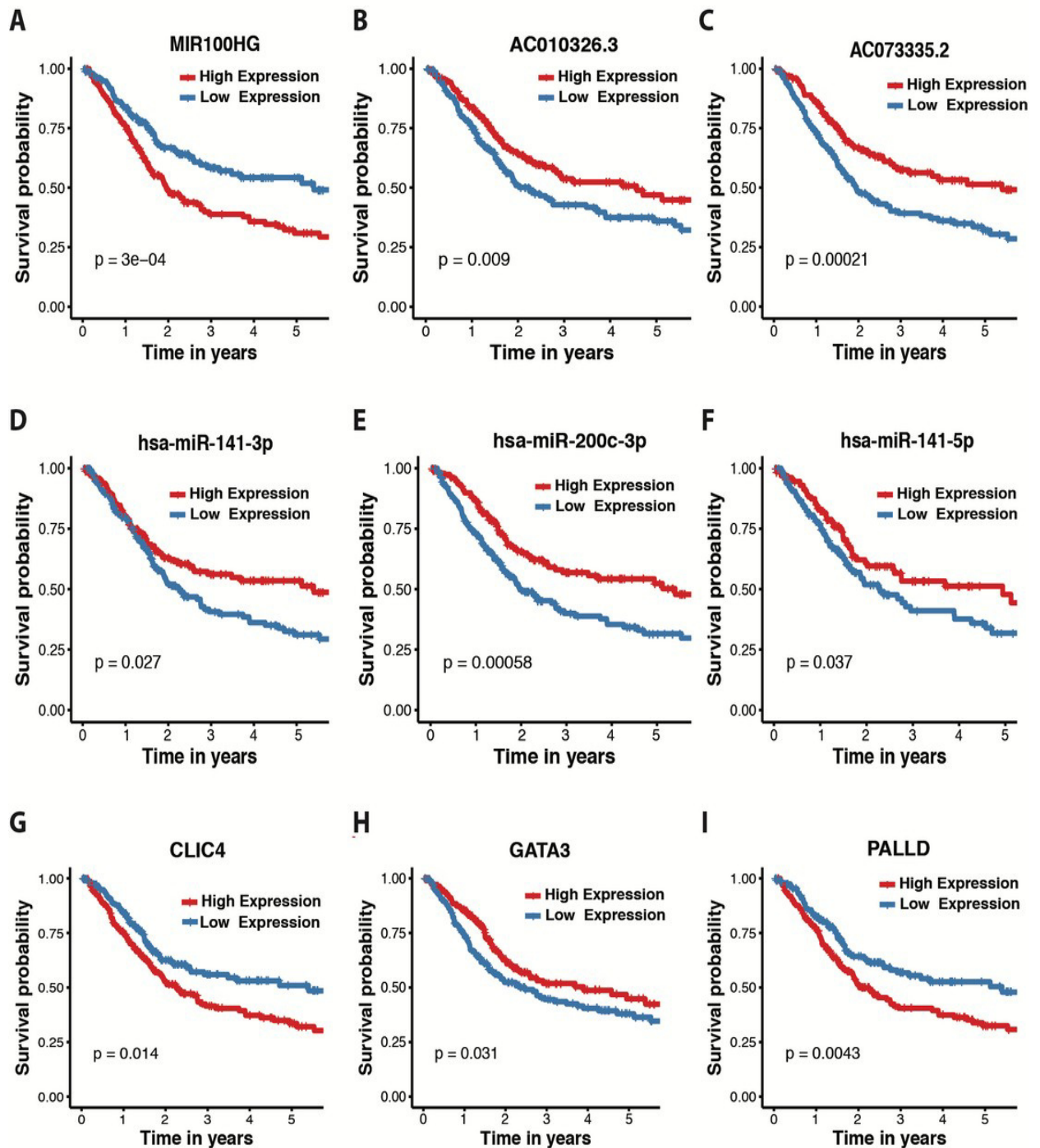
В частности, две ко-экспрессируемые пары днРНК-мРНК, а именно AC010326.3-GATA3 и AC073335.2-GATA3, позитивно регулируется мкРНК

miR-141-3p. Пары днРНК-мРНК, такие как MIR100HG–CLIC4 и MIR100HG–PALLD, негативно регулируется miR-200c-3p и miR-141-5p, соответственно (рисунок 18B). График типа «тепловой карты» показал, что эти девять генов значительно различаются по своей экспрессии между базальным и люминальным подтипами (рисунок 18C).

3.6 – Анализ выживаемости при мышечно-инвазивном раке мочевого пузыря

Мы также оценили, коррелирует ли экспрессия вовлеченных генов с общей выживаемостью пациентов с МИРМП. Для сравнения времени выживания между двумя группами, образцы пациентов были разделены на группы с высокой и низкой экспрессией с медианными уровнями экспрессии гена в качестве порогового значения (cut-off), использовался логарифмический ранговый критерий. Анализ выживаемости показал, что каждая из выявленных молекул РНК, вовлеченных в выявленную перекрестную взаимозависимость, тесно связана с прогнозом выживаемости. В частности, более высокий уровень экспрессии miR-141-5p, miR-141-3p, AC010326.3, AC073335.2, miR-200c-3p и GATA3 предсказывает благоприятный прогноз, указывая на то, что они могут функционировать как ингибиторы опухоли (рисунок 19 B, C, D, E, F и H). В отличие от этого, более высокий уровень экспрессии MIR100HG, PALLD и CLIC4 связан с неблагоприятным прогнозом, что позволяет предположить, что они могут играть онкогенную роль (рисунок 19 A, G и I).

Чтобы проверить прогностические эффекты трех мРНК (CLIC4, PALLD и GATA3), задействованных в перекрестных взаимозависимостях, зависящих от miR-200c и miR141, мы использовали два набора данных микрочипов, GSE13507 (61 пациентов с МИРМП) и GSE31684 (78 пациентов с МИРМП). Для подтверждения прогностической роли днРНК и мкРНК необходим дальнейший подбор данных.



**Рисунок 19 – Графики кривой Каплан-Мейера для девяти
взаимозависимых РНК из набора данных TCGA (A-I)**

Примечание: красные линии представляют высокую экспрессию РНК, в то время как синие линии представляют соответствующую низкую экспрессию. Р-значение рассчитывали с использованием логарифмического рангового критерия, где $p < 0,05$ представляет статистическую значимость.

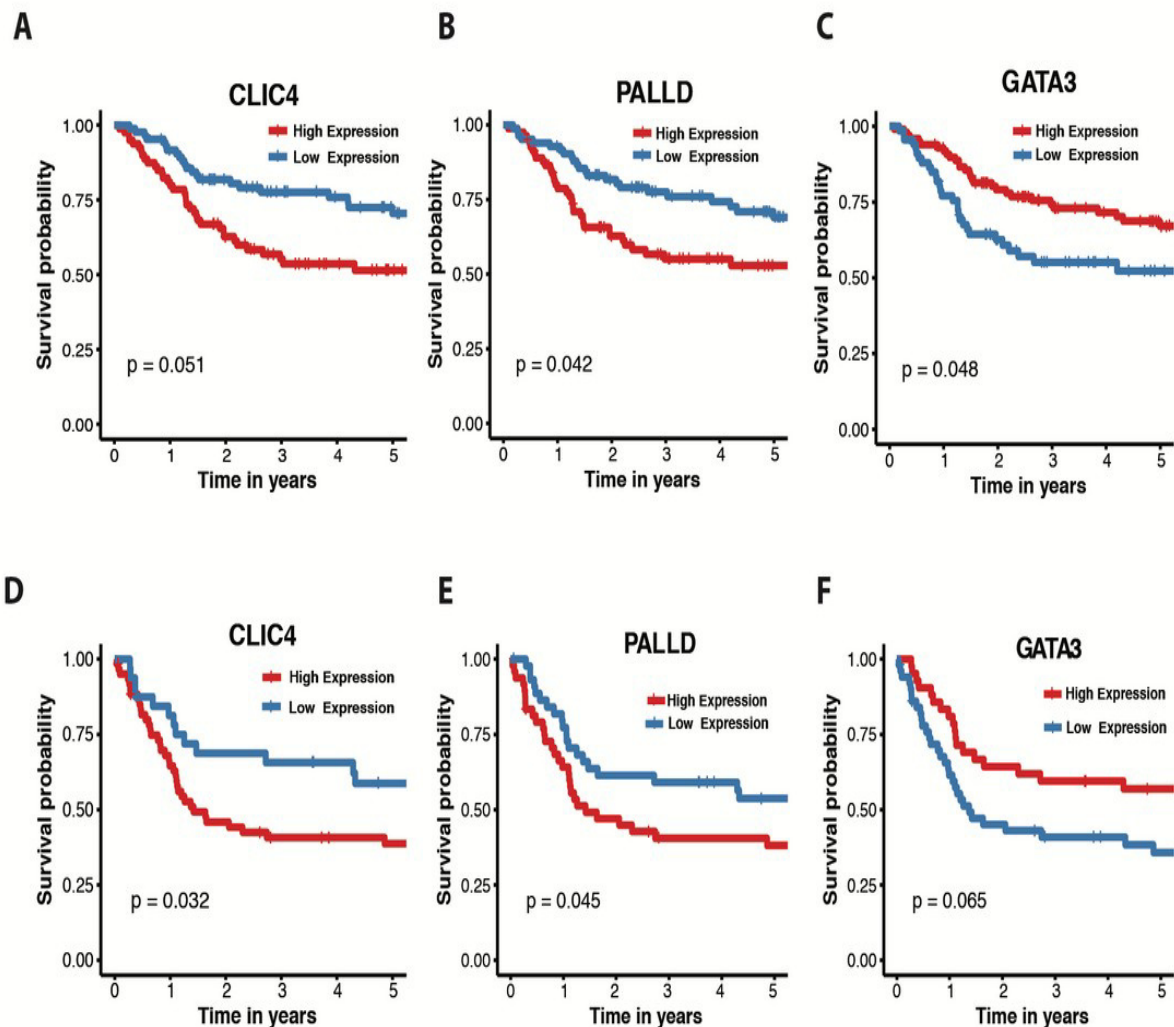


Рисунок 20 – Графики Каплана-Мейера для CLIC4, PALLD и GATA3 в наборе данных GSE13507 и GSE31684

Примечание: (A-C) кривые выживания К-М, показывают общую выживаемость в соответствии с высокой экспрессией и низкой экспрессией CLIC4, PALLD и GATA3 в наборе данных GSE13507. (D-F) Кривые выживаемости К-М, показывают общую выживаемость в соответствии с высокой экспрессией и низкой экспрессией CLIC4, PALLD, и GATA3 в наборе данных GSE31684.

Результаты наших исследований показали, что GATA3 можно использовать в качестве маркера супрессоров опухолей, а CLIC4 и PALLD можно использовать в качестве онкогенов. Благодаря перекрестной взаимозависимости в качестве эффективных клинических биомаркеров для прогноза МИРМП можно использовать miR-200c и miR-141 (рисунок 20 A-F).

Резюме

В этой главе мы использовали данные секвенса РНК, имеющиеся в базе данных TCGA, для изучения гетерогенности опухоли при мышечно-инвазивном раке мочевого пузыря (МИРМП), а также для прогнозирования прогностических биомаркеров. Мы классифицировали 403 образца МИРМП на два молекулярных типа, которые соответственно напоминают ранее идентифицированные базальные и люминальные подтипы.

Выявлен факт того, что базальные подтипы рака мочевого пузыря характеризуются повышенной экспрессией генов, относящихся к процессу эпителиально-мезенхимального перехода, метастазирования и функционированию иммунной системы. Люминальные подтипы характеризуются относительно повышенной экспрессией генов, отвечающих за метаболические процессы. Эти результаты обеспечивают более глубокое понимание гетерогенности рака мочевого пузыря.

Что еще более важно, мы идентифицировали мкРНК-зависимое перекрестное взаимовлияние мРНК и днРНК, которая состоит из GATA3, CLIC4, PALLD, MIR-200c-3p, miRNA-141-3p, miRNA-141-5p, AC010326.3, AC073335.2 и MIR100HG. Выявлено, что уровень экспрессии генов в этой перекрестной взаимозависимости значительно различался между базальным и люминальным подтипами. Эти результаты указывают на то, что перекрестное взаимовлияние мРНК и днРНК, обусловленное miR-200c и miR-141, может играть значительную роль в онкогенезе и прогрессировании опухоли. GATA3, CLIC4, PALLD, MIR-200c-3p, miRNA-141-3p, miRNA-141-5p, AC010326.3, AC073335.2 и MIR100HG могут служить прогностическими биомаркерами мышечно-инвазивного рака мочевого пузыря, а также использоваться как маркеры для классификации подтипов.

Исследована взаимосвязь между подтипами и маркерами иммунотерапии. Было обнаружено, что паттерны экспрессии молекул контрольных точек иммунного ответа (например, PD-1, PD-L1, CTLA-4,

HAVCR-2 и LAG-3) позволяют различать базальные и люминальные подтипы рака мочевого пузыря. При базальном подтипе МИРМП выявлены более высокие значения экспрессии молекул контрольных точек иммунного ответа, чем при люминальном подтипе, что необходимо учитывать при назначении иммунотерапии опухолей.

Список работ, опубликованных по 3-й главе

1. Identification of miR-200c and miR141-mediated lncRNA-mRNA crosstalks in muscle-invasive bladder cancer subtypes / Guojun Liu, Z. Chen, I.G. Danilova, M.A. Bolkov, I.A. Tuzankina, Guoqing Liu // *Frontiers in genetics*. – 2018. – Vol. 9. – P. 422.
2. A co-expression network for differentially expressed genes in bladder cancer and a risk score model for predicting survival / Z. Chen, Guojun Liu, I. Tuzankina, M. Bolkov, Guoqing Liu, W. Tang // *Hereditas*. – 2019. – Vol. 156, № 1. – P. 24.

ГЛАВА 4 – ИДЕНТИФИКАЦИЯ ГЕНОВ-КАНДИДАТОВ ОБЩЕГО ВАРИАБЕЛЬНОГО ИММУНОДЕФИЦИТА

4.1 – Исследование мультигенной природы общего вариабельного иммунодефицита

Исследование межбелковых взаимодействий (PPI) генов общего вариабельного иммунодефицита (ОВИН) в геноме человека является ключевым для понимания мультигенной природы ОВИН и дальнейшей идентификации идеального гена-кандидата ОВИН. Данные PPI для генов ПИД были получены из базы данных STRING и визуализированы с помощью программного обеспечения Cytoscape. Было обнаружено, что гены ОВИН имеют тенденцию к централизации в сети (образно говоря, в ступице). Это даёт основания предполагать, что гены ОВИН могут взаимодействовать между собой чаще, чем другие гены ПИД (рисунок 21).

4.2 – Анализ взаимосвязей генов нозологических синдромов группы общего вариабельного иммунодефицита

Для дальнейшего изучения феномена сложных взаимодействий генов ОВИН, данные PPI группы ОВИН, состоящей из 39 генов ОВИН, сравнили с PPI десяти случайных групп (каждая группа состояла из 39 случайно отобранных генов ПИД).

Для измерения и сравнения когезии и плотности сетей PPI каждой группы рассчитывалась плотность сети ($D_{network}$). Принято считать, что чем больше плотность сети для группы, тем ближе взаимодействие генов внутри группы.

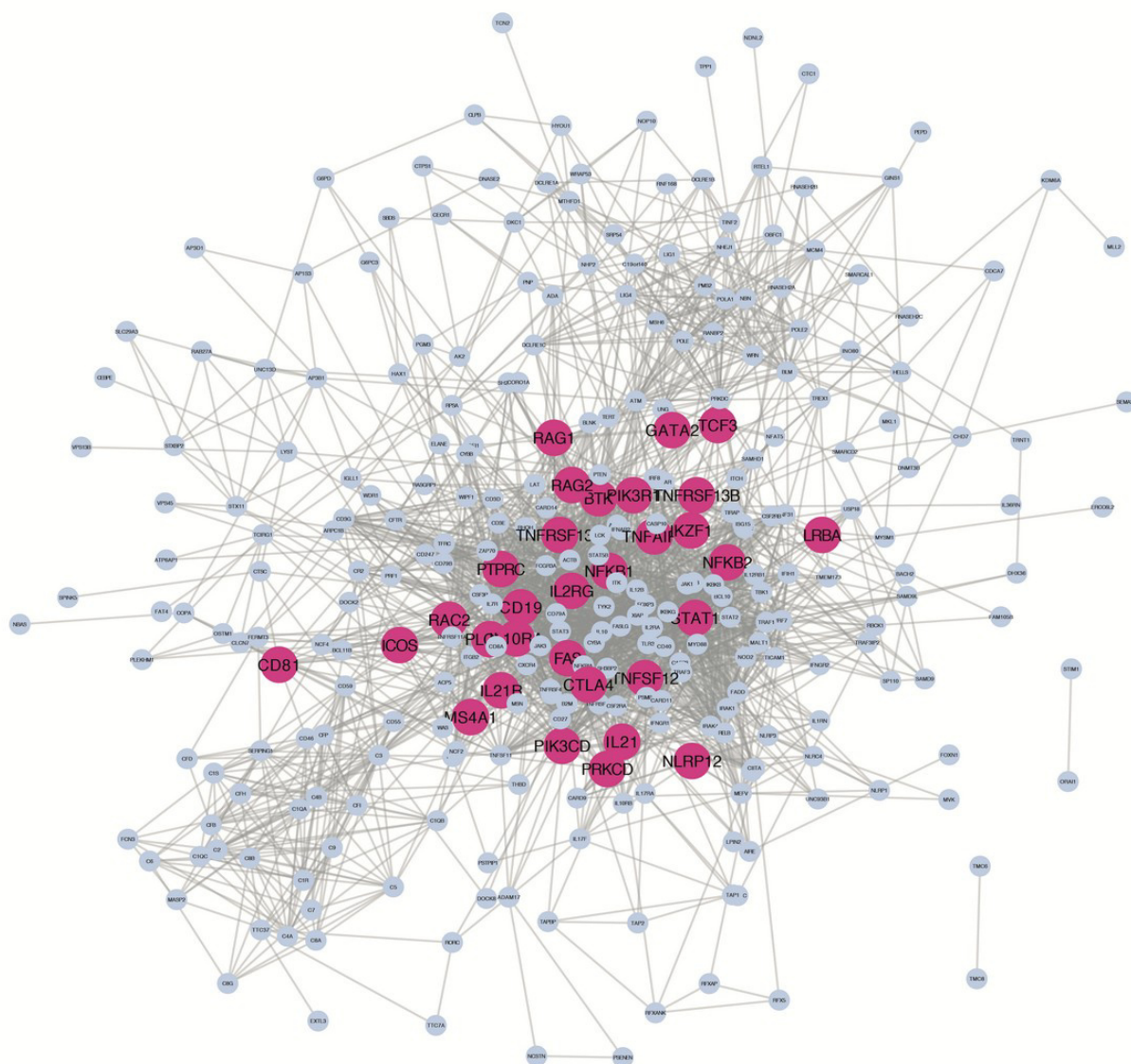


Рисунок 21 – Сеть межбелкового взаимодействия генов первичных иммунодефицитов

Примечание: данные межбелкового взаимодействия (PPI) для генов первичных иммунодефицитов, полученные из базы данных STRING, визуализировались с помощью программного обеспечения Cytoscape. Красные узлы представляют гены общей варибельной иммунной недостаточности, светло-голубые узлы представляют другие гены первичных иммунодефицитов.

Результаты показали, что группа ОВИН обладает более высокой плотностью сети, чем десять случайных групп. Это указывает на то, что гены ОВИН действительно более тесно взаимодействуют друг с другом, чем любые другие гены ПИД (рисунок 22А). Было рассчитано биологическое расстояние ($B_{i,j}$) группы ОВИН и двух случайных групп (каждая группа

состояла из 39 генов ПИД) и проведено сравнение распределения плотности биологического расстояния. При трактовке результатов мы имели ввиду, что чем меньше биологическое расстояние для группы, тем ближе функциональная значимость между генами в группе. Результаты показали, что распределение плотности группы ОВИН имело медианное значение 8,8, в то время как медианные значения случайных групп были около 14, что указывало на более тесную функциональную взаимосвязь между генами ОВИН (рисунок 22В).

Полученные результаты свидетельствуют о том, что гены ОВИН более тесно взаимодействуют друг с другом и имеют более тесную биологическую взаимосвязь, чем другие гены ПИД.

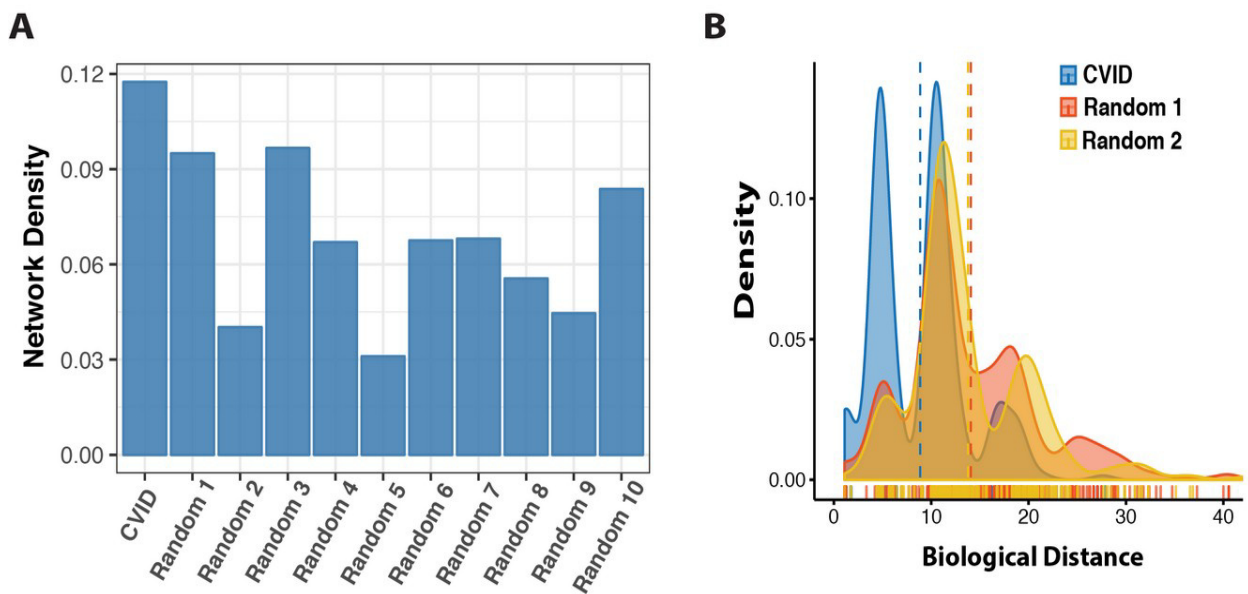


Рисунок 22 – Плотности сети и биологические расстояния группы общей варибельной иммунной недостаточности и случайных групп

Примечание: (А) график плотности сети группы ОВИН и десяти случайных групп. (В) График плотности биологических расстояний группы ОВИН и двух случайных групп.

На основе корреляционного анализа Пирсона (PCC) и данных межбелковых взаимодействий, предоставленных Cheng F. et al., 2018 [136], были получены 2751 ОВИН-специфических взаимодействий, включая 1716

генов-кандидатов. Используя данные KEGG, нами был выполнен так называемый анализ функционального обогащения известных генов ОВИН (по сути – привязка генов из списка к их биологическим функциям) и обнаружено, что в общей сложности 15 сигнальных путей, описанных в базе KEGG (например, инфекция вируса Эпштейна-Барр, взаимодействие цитокин-цитокиновых рецепторов и сигнальный путь рецептора В-клеток), были статистически значимыми ($p < 0,05$, рисунок 23).

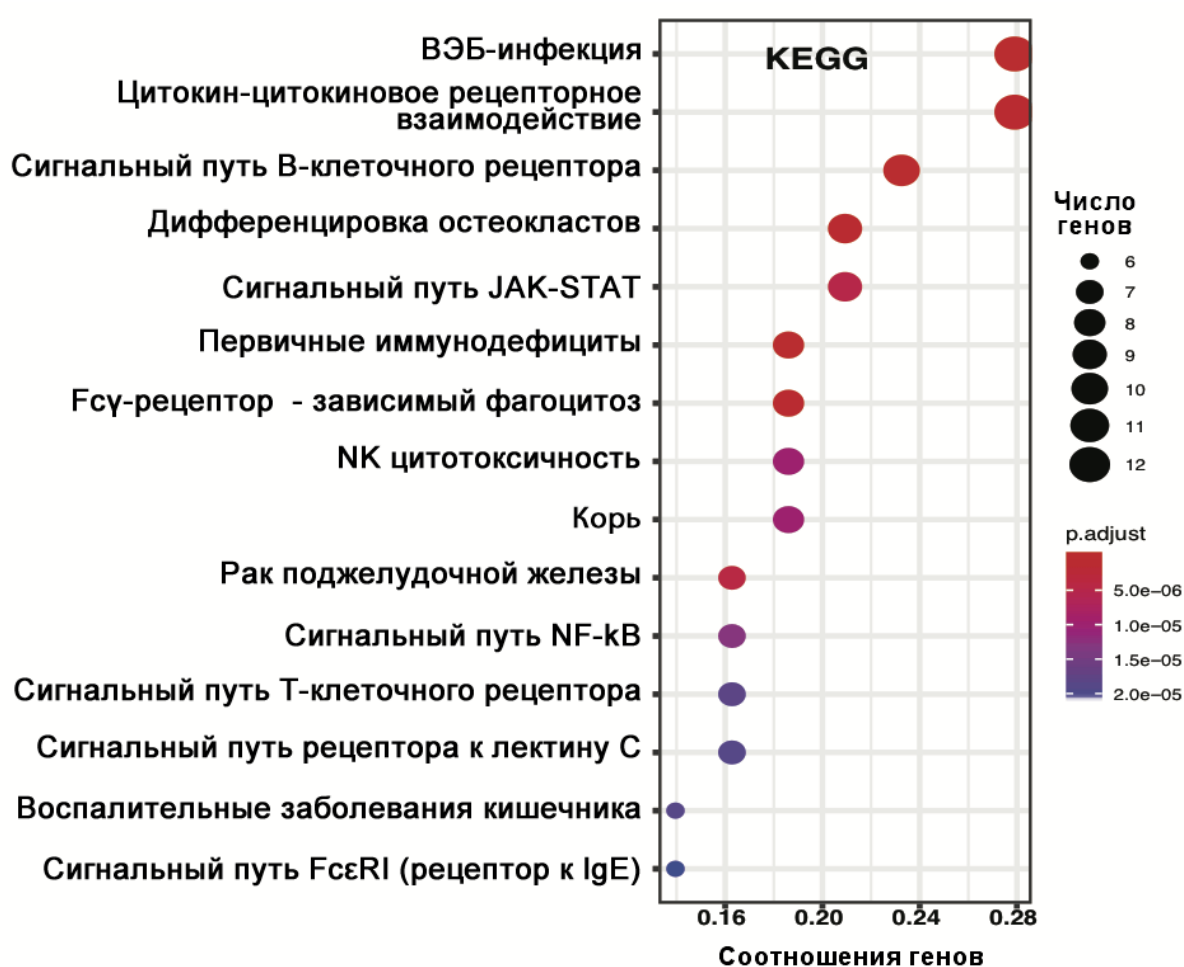
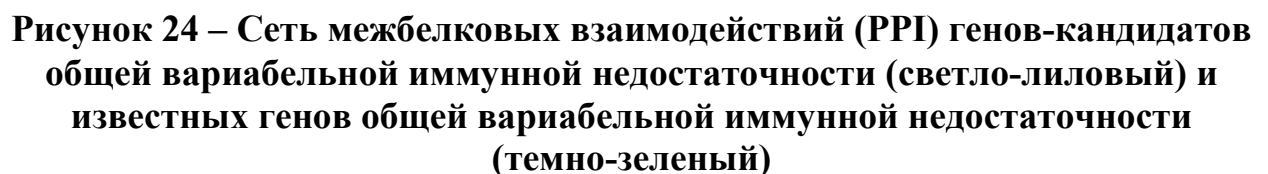


Рисунок 23 – Функциональное обогащение генов общей варибельной иммунной недостаточности биологическим значением из базы KEGG

Примечание: размер пузырька соответствует количеству генов, а цвет пузырька соответствует скорректированному р-значению.

Чтобы определить гены-кандидаты, функционально сходные с известными генами ОВИН, дополнительно был проведен скрининг на



Список выявленных генов-кандидатов общего варибельного иммунодефицита: LMNA, BAD, PRKAB2, CRK, CRKL, CASP8, CTNNB1, STAT3, HRAS, PTK2, JAK1, BCL2, PDGFRB, CHUK, JAK2, GRB2, SRC, RELA, NFKBIA, EGFR, CREBBP, EP300, MAPK1, IKBKB, PIK3CA, SHC1,

PIK3CB, IL10, ERBB2, PPP1CA, TRAF6, INHBA, GNAQ, VAV2, TGFBR1, TGFBR2, BCAR1, NOS2, FYN, SOCS1, SMAD4, TAB2, TNFRSF1A, TNF, HLA-DRA, FLNA, ITGB1, ACTN1, AKT1, SP1, PIK3CG, BRAF, INHBC, CD47, MET, PIP5K1C, SOS1, GP6, JUN, ICAM1, GP1BA, PRKACA, PRKCA, PRKAG1, TYK2, KDR, ACTB, MAPK3, PLCB1, RAC1, ADCY7, RAF1, SDC4, IGF1R, ITGA6, ITGB4, ITGB7, PAK1, AKT3, CDC42, SMAD1, CCR5, CASP3, ZYX, ITGB2, RAP1A, PDGFRA, MAPK8, VAV3, CASP9, GSK3B, IRAK1, MYD88, TLR9, PRKCB, BMP7, RHOA, NFKBIB, MAPK14, BMP6, INHBB, GDF5, TLR4, CAV1, CD40, ITGA4, C8B, C9, IL10RB, PECAM1, PPP2CA, CUL1, IL1R1, CD14, HSPA1B, BIRC3, HSPA1L, IRAK4, BMP2, SOS2, MYLK, PIK3R3, RAPGEF1, MAPK13, GNAS, HSPA2, MAPK9, ELK1, MYC, PPP1CB, CCND3, MYLK2, HSPA6, BIRC2, DAG1, FLNB, FLNC, ID4, XIAP, ACTN4, SMAD3, MAPK12, PPP2R1A, PRKACB, MAPK10, MAP2K1, MAP2K6, MAP3K7, BMPR1A, HSPB1, PRKCG, ACVR1, PPP1CC, RPS6KB1, HLA-DRB5, SERPINB9, FN1, COL6A2, VCAM1, RAB7A, ACTG1, MYL12A, PPP2R1B, SKP1, ALOX5, RAB5A, PAK2, RBX1, SMURF1, EMD, TPM2, IFNGR.

4.3 – Оценка выявленных генов-кандидатов общего варибельного иммунодефицита

Чтобы оценить наши прогнозы, было рассчитано биологическое расстояние 172 генов-кандидатов ОВИН и проведено его сравнение с 39 известными генами ОВИН. В результате среднее биологическое расстояние между генами-кандидатами ОВИН составило 6,01, что аналогично расстоянию между известными генами ОВИН, и указывает на сильную биологическую связь между генами-кандидатами ОВИН и генами ОВИН (рисунок 25).

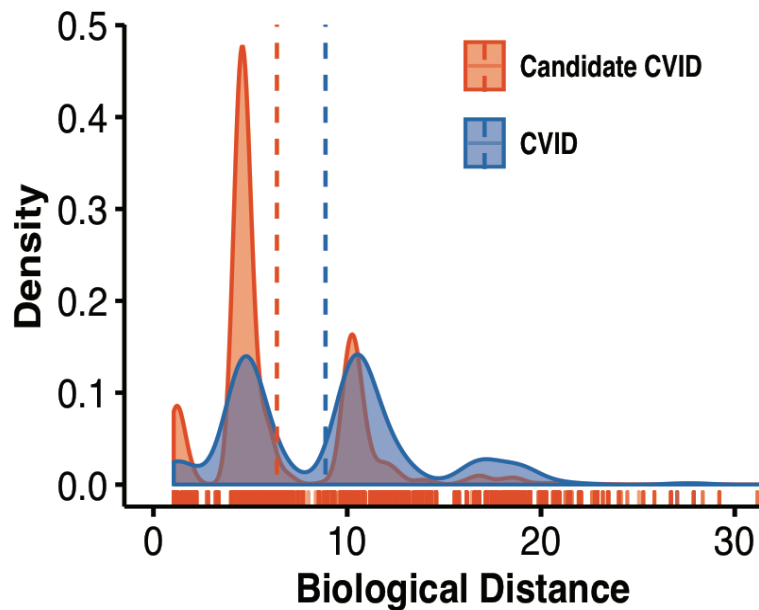


Рисунок 25 – График плотности биологических расстояний известных генов ОВИН и прогнозируемых генов-кандидатов ОВИН

Затем гены-кандидаты ОВИН смешивали с генами ОВИН и рассчитывали биологическое расстояние смешанных генов. Это было сделано для того, чтобы выполнить филогенетический иерархический кластерный анализ. Полученное веерообразное филогенетическое дерево биологических расстояний, генерируемое функциональным геномным выравниванием (FGA), позволяет оценить корреляцию между генами ОВИН и генами-кандидатами ОВИН. Результаты показали, что гены-кандидаты ОВИН были равномерно распределены по всему диапазону известных генов ОВИН, что еще раз свидетельствует о том, что эти гены-кандидаты были тесно связаны с известными генами (рисунок 26).

Выявив новые гены-кандидаты, мы провели исследование качества прогноза, проведя обзор некоторых исследований. В частности, восемь генов-кандидатов ОВИН (AKT1, AKT3, RELA, SOCS1, STAT3, XIAP, CD40 и CASP8), которые не были включены в наш первоначальный список генов ОВИН, полученный из ESID, но были нами предсказаны, оказались выявленными в клинических случаях общего вариабельного

иммунодефицита. Это демонстрирует значимость выявленных генов-кандидатов ОВИН (таблица 11).

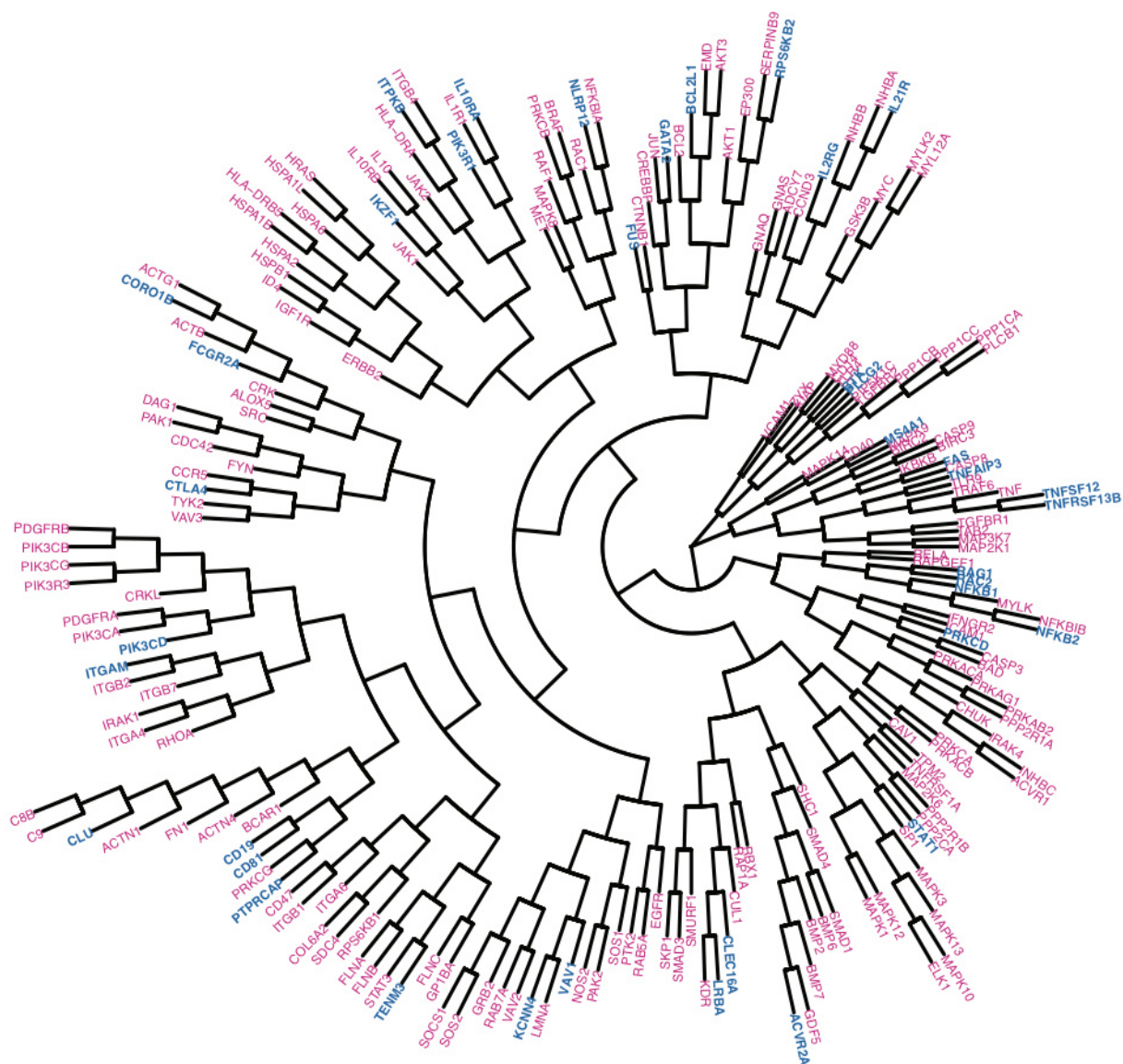


Рисунок 26 – Филогенетический иерархический кластерный анализ генов-кандидатов общей варибельной иммунной недостаточности (ОВИН) и известных генов ОВИН, показывающий иерархическую кластеризацию всех известных генов ОВИН (синий) и прогнозируемых генов ОВИН (фиолетово-красный)

Примечание: длина ветви указывает на силу разделения между субъектами.

Таблица 11 – Список опубликованных генов-кандидатов общей вариабельной иммунной недостаточности (ОВИН) с выявленной связью с ОВИН, не включенных в список генов ОВИН на сайте Европейского общества иммунодефицитов (ESID) и не использованных в исследовании в качестве исходных генов с известной связью с ОВИН

Gene symbol	Description	Aliases	References
AKT1	AKT serine/threonine kinase 1	AKT, CWS6, PKB, PKB-ALPHA, PRKBA, RAC, RAC-ALPHA	PMID: 27664934
AKT3	AKT serine/threonine kinase 3	MPPH, MPPH2, PKB-GAMMA, PKBG, PRKBG, RAC-PK-gamma, RAC-gamma, STK-2	PMID: 26081581
RELA	RELA proto-oncogene, NF-kB subunit	NFKB3, P65	PIMD: 27461466
SOCS1	Suppressor of cytokine signaling 1	CIS1, CISH1, JAB, SOCS-1, SSI-1, SSI1, TIP-3, TIP3	PMID: 29618830
STAT3	Signal transducer and activator of transcription 3	ADMIO, ADMIO1, APRF, HIES	PMID: 26360251 PMID: 27379089
XIAP	X-linked inhibitor of apoptosis	API3, BIRC4, IAP-3, ILP1, MIHA, XLP2, hIAP-3, hIAP3	PMID: 27492372
CD40	CD40 molecule	Bp50, CDW40, TNFRSF5, P50	PMID: 23305827 PMID: 30464201 PMID: 28756897
CASP8	Caspase 8	ALPS2B, CAP4, Casp-8, FLICE, MACH, MCH5	PMID: 28756897 PMID: 25814141

Резюме

Общий переменный иммунодефицит представляет собой такой тип ПИД, который обладает глубокой фенотипической и генотипической гетерогенностью и вызывается как моногенными, так и сложными мультигенными причинами. В этой главе путем анализа плотности сети и плотности биологического расстояния мы обнаружили, что гены ОВИН функционально более похожи друг на друга и тесно взаимодействуют друг с другом по сравнению с другими генами ПИД. Это соответствует представлению о том, что ОВИН более мультигенная, по сравнению с большинством ПИД когорты заболеваний. С другой стороны, это ещё раз подтверждает идею о том, что ОВИН – скорее диагноз исключения и объединяет огромную когорту разнообразных нозологий с более-менее общими фенотипическими проявлениями, при которых ещё остаются невыявленными генетические предпосылки развития патологии и не определены соответствующие таргетные терапевтические воздействия.

Кроме того, мы разработали методы, позволяющие предсказать причинно значимые гены для развития ОВИН. Методы объединяют данные ко-экспрессии и межгенных взаимодействий, а также данные, характеризующие функции генов. Мы определили 172 новых гена-кандидата ОВИН, которые находятся во взаимозависимости с известными генами ОВИН в одних и тех же биологических путях и показывают высокую биологическую корреляцию с известными генами ОВИН.

Кроме того, восемь генов-кандидатов ОВИН, которые не были включены в наш первоначальный список генов ОВИН, полученный из Европейского общества иммунодефицитов (ESID), но были нами предсказаны, оказались выявленными другими исследователями в клинических исследованиях общего переменного иммунодефицита. Это демонстрирует значимость полученных в нашем исследовании выявленных генов-кандидатов ОВИН.

В заключение можно сказать, что это исследование выявило глубокие взаимодействия между молекулами, дефект которых ведет к манифестации ОВИН и помогает понять возможную причину мультигенной природы заболеваний.

Многие из предсказанных и доказанных в клинических исследованиях генов ОВИН обладают широкими функциями и полное выключение функций некоторых генов может привести к гибели человека в эмбриональном периоде.

Однако известно, что ОВИН довольно часто дебютирует у лиц старше 20-40 лет. Причины позднего дебюта генетических болезней обычно объясняются окружением и комбинацией внутренних и внешних причин. В зависимости от патогенности мутации и их влияния на функцию белка, мутации могут как оставаться совершенно не проявленными или проявляться только в определенных условиях, так и иметь строгие генотип-фенотипические корреляции, но последнее является редкостью. Мультигенная природа ОВИН объясняется в частности тем, что для дебюта болезни необходимо сочетание множества факторов для того, чтобы функционально слабые белки в определенных условиях сформировали клиническое проявление болезни.

Список работ, опубликованных по 4-й главе

1. Постмортальный анализ иммунозависимой патологии у детей раннего возраста / Г. Лю, М.А. Болков, С.С. Дерябина, Я.М. Крохалева, И.А. Тузанкина // Аутоиммунные и иммунодефицитные заболевания: тез. докл. III всерос. конгресса. – 2018. – Т. 56, № 3. – С. 5.

2. Функциональные взаимоотношения между генами общей вариабельной иммунной недостаточности в геноме человека / М.А. Болков, Г. Лю, И.А. Тузанкина, И.Г. Данилова // Российский журнал детской гематологии и онкологии. – 2019. – Т. 6, № 2. – С. 47.

3. Identification of candidate disease genes in patients with common variable immunodeficiency / Guojun Liu, M. Bolkov, I. Tuzankina, I. Danilova // Quantitative Biology. – 2019. – Vol. 7, № 3. – P. 190-201.

ГЛАВА 5 – МЕТОД ИДЕНТИФИКАЦИИ МУТАЦИЙ ГЕНОВ КАК ВЕРОЯТНО ПАТОГЕННЫХ (FAT4, RAG1, PIK3CD и CSF3R), ПРИ ФЕНОТИПЕ СИНДРОМА ХЕННЕКАМА

Данное исследование было проведено ввиду того, что под нашим наблюдением оказалась пациентка с фенотипическими проявлениями синдрома Хеннекама, но с клиническими особенностями, которые позволяли усомниться в правильности этого диагноза. После проведения полногеномного секвенирования в коммерческой лаборатории диагноз не подтвердился. В виду этого было принято решение применить новые методы для выявления мутантных генов, связанных с фенотипом пациента и подтвердить диагноз.

Данный раздел исследования позволил показать эффективность и широкие возможности прикладного использования методов, примененных для фундаментальных исследований. Кроме того, в данном случае получилось выявить мутации в патогенетически значимых генах, что дало возможность обосновать применение на практике новых терапевтических технологий, апробированных за рубежом, но не включенных для применения в России.

5.1 – Фенотипические проявления синдрома Хеннекама

Пациентом с диагнозом «синдром Хеннекама» была пятилетняя девочка. Её клинические проявления включали следующее:

- ассиметричная лимфедема с преимущественным поражением правой стороны, впервые выявленная ещё в перинатальном периоде на 34 неделе гестации (рисунок 27А);
- бородавчатые образования на коже ладони, указательного и большого пальцев правой кисти (рисунок 27В);

- лимфоангиоэктазии, которые осложнились присоединением кожного патологического процесса, предположительно папилломовирусной этиологии;
- хилезный асцит, выявленный с пренатального периода, который сохранялся до настоящего времени (рисунок 27В);
- кистозные дефекты нижней челюсти с вовлечением костной ткани (МРТ от 11.10.2017);
- правосторонний гидроторакс;
- внутренняя бивентрикулярная открытая гидроцефалия;
- лимфаденопатия всех групп шейных узлов (больше справа), надключичных, подключичных, парастернальных, а также грудных лимфоузлов, единичный лимфоузел правой поясничной области в диаметре до 5мм, множественные, увеличенные до 17 мм лимфоузлы левой подвздошной и паховых групп, в настоящее время - лимфоузлы мелкие до 0,3 см, подвижные, пальпируются преимущественно в аксиллярных группах;
- гирсутизм, выраженный на верхних конечностях (рисунок 26В);
- повышение мочевины, гипоплазия почки.

Кроме того, **в наследственности** –

- наличие регрессирующей беременности, предшествующей настоящей, прервавшейся в 12 недель гестации;
- наличие в нескольких поколениях по материнской линии первичной вегето-сосудистой дистонии на фоне соединительнотканной дисплазии, протекавшей с выраженными сосудистыми изменениями в виде варикозной болезни, геморроя, субкомпенсированного тонзиллита, у мамы конституционально обусловленной (первичной) вегето-сосудистой дистонии, дисфункция сопровождала всю беременность – гипотония и вазомоторный ринит, применение сосудосуживающих капель всю беременность, многоводие;

- рождение крупных плодов – сибс 13 лет родился с массой 4150г, а настоящая девочка – 4750г с длиной 57см;
- эндокринная патология (не исключен аутоиммунный процесс) у мамы – гипотиреоз и слабость родовой деятельности при I родах, послужившая основанием для оперативных родов кесаревым сечением;
- по отцовской линии наличие патологии органов пищеварения в виде панкреатита и калькулезного холецистита, потребовавшего оперативного вмешательства.

А**В**

Рисунок 27 – Фотография пациента с синдромом Хеннекама

Примечание: (А) лимфедема правой половины тела (в частности, нижней конечности) и (В) бородавчатые образования на правой ладони, асцит, лимфедема правой руки, гирсутизм (на правой руке).

5.2 – Поиск вероятно-патогенных мутаций для верификации синдрома Хеннекама

Анализ полногеномных данных пациента проводился с использованием программного обеспечения BWA, GATK4, VCFtools. Анализ так называемой нисходящей фильтрации результатов был выполнен с помощью программного обеспечения ANNOVAR и базы данных CADD, его результаты схематически представлены на рисунке 28. Первым этапом фильтрации было удаление всех синонимичных SNV, не связанных с фреймами InDels и вариантов, помеченных как «NA» или «неизвестно».

В общей сложности было идентифицировано 10302 вариантов SNV или InDel. Затем выявленные варианты были отфильтрованы через наложение на известные 351 ген ПИД и 2 известных гена синдрома Хеннекама, не включенных в тот список (FAT4 и CSF3R). Отобранные 111 вариантов были сохранены для поиска более вероятных вариантов. После исключения распространенных вариантов, MAF>0,01 по ExAC, 1000g и gnomAD, в общей сложности осталось 6 редких вариантов. Для выбора вредоносных мутаций были использованы модели CADD, FATHMM, и PROVEAN, и, наконец, были предсказаны четыре патогенные мутации, которые являются вероятно приводящими к развитию заболевания в исследуемом случае.

5.3 – Оценка вероятной патогенности выявленных в исследовании мутаций

В частности, была идентифицирована гомозиготная мутация (g.125452634G>A) в гене FAT4, о котором сообщалось, что обе биаллельные и моноаллельные мутации связаны с синдромом Хеннекама. Кроме того, у пациента были выявлены следующие мутации: гетерозиготная мутация (g.36575963G>A) в RAG1, гетерозиготная мутация (g.9715914T>A) в PIK3CD, гетерозиготная мутация (g.36471505C>T) в CSF3R (таблица 12).

Таблица 12 – Сводка результатов SNV, выявленных у пациента и вероятно связанных с заболеванием, включая местоположение мутаций, прогнозы патогенности и популяционные данные

	dbSNP ID	Характерный описанный фенотип	Локализация на хромосоме	Гомо или гетерозигота	Мутация	Аминокислотная замена	Распространенность в популяции по сведениям баз данных (1000g/gnomAD/ExAC)	Вероятность патогенности по сведениям баз данных (FATHMM/PROVEAN/CADD)
FAT4	rs12650153	Синдром Ханнекама	Chr4(q28.1)	Hom	g.125452634G>A	p.Ser3875Asn	0.009/0.007/0.002	T/N/20.1
RAG1	rs4151034	Разные ПИД	Chr11(p12)	Het	g.36575963G>A	p.Asp887Asn	0.001/0.006/0.004	D/N/23.5
PIK3CD	rs142285826	ОВИН	Chr1(p36.22)	Het	g.9715914T>A	p.Phe146Ile	0.001/0.003/0.002	T/N/23.4
CSF3R	rs3918019	Врожденная нейтропения	Chr1(p34.3)	Het	g.36471505C>T	p.Glu405Lys	0.005/0.002/0.006	T/N/17.7

Примечание: 1000g, gnomAD, ExAC (при <0,01 считаются редкими); FATHMM (Т - толерантная мутация; D - вредная); PROVEAN (N, нейтральная мутация; D - вредная); CADD (при>15 вероятно вредная); Het - гетерозиготная, Hom – гомозиготная мутация.

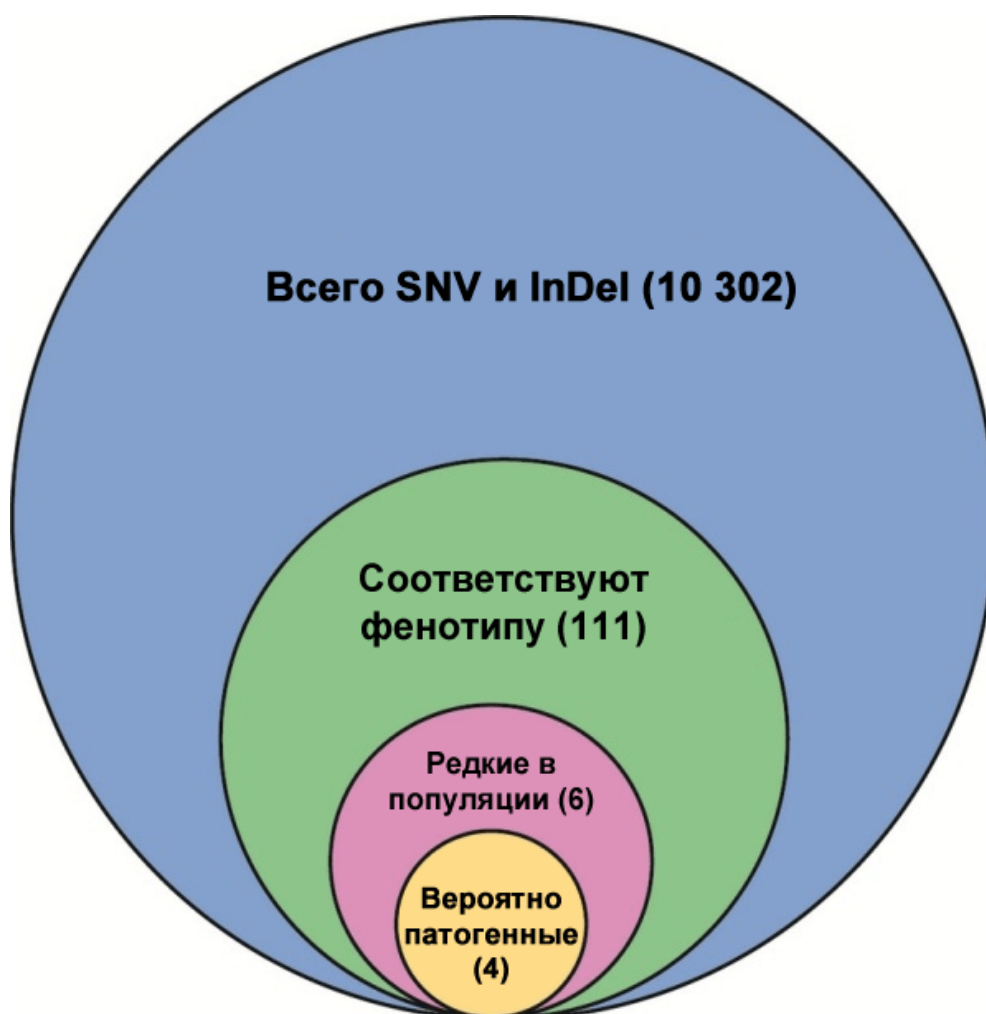


Рисунок 28 – Основные этапы фильтрации результатов для идентификации потенциально патогенных мутаций

Примечание: количество отфильтрованных генов размещено в скобках.

Все мутации-кандидаты были подтверждены путем изучения данных в программном обеспечении «Integrated genomics viewer (IGV)» (рисунок 29).

Относительно гомозиготной мутации g.125452634G>A гена FAT4 встречаются противоречивые сведения. По данным базы dbSNP Национального центра биотехнологической информации (NCBI), США, мутация является редкой и доброкачественной, не связанной с каким-либо фенотипом болезни (данные от 8 января 2016).

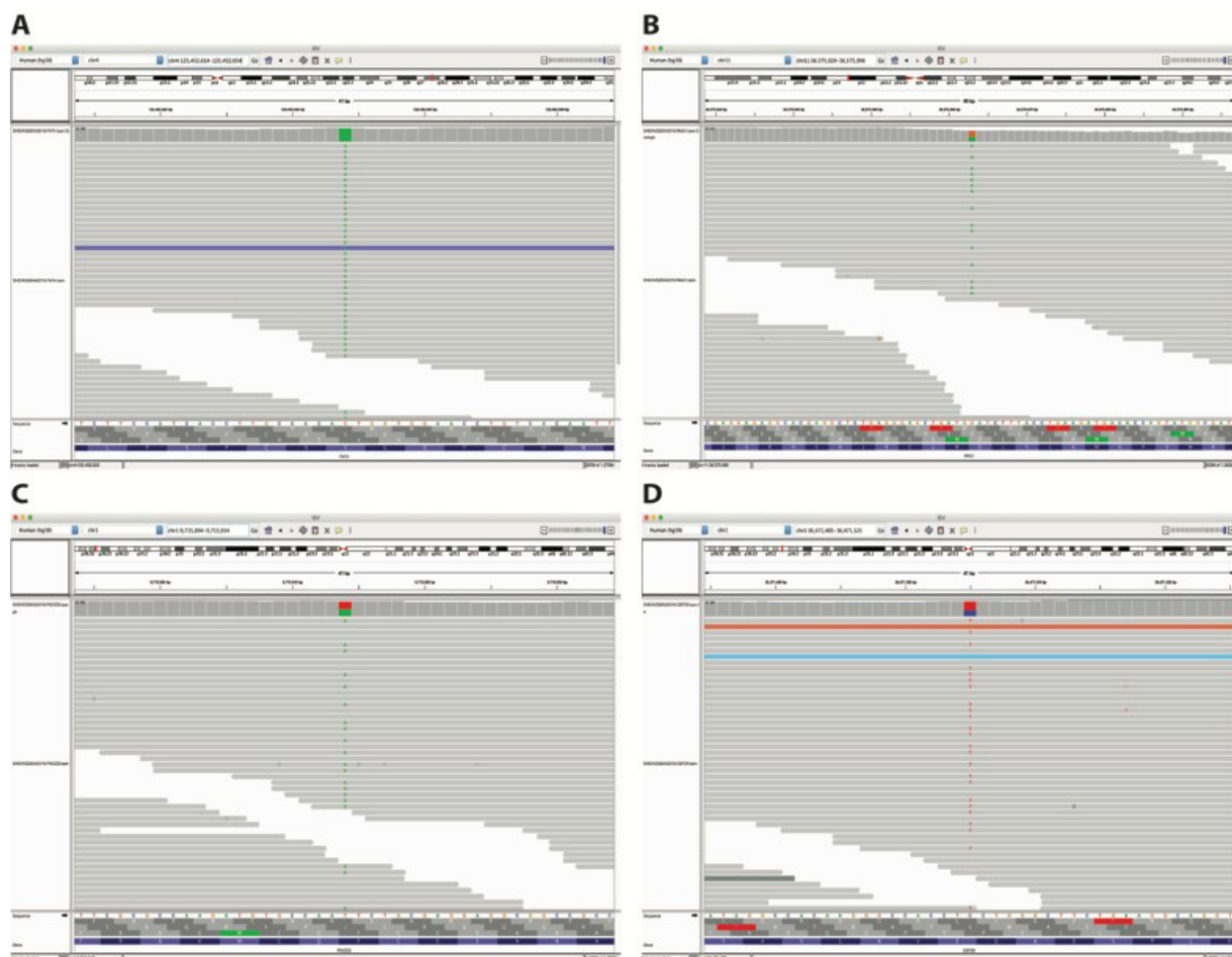


Рисунок 29 – Визуализация четырех мутаций-кандидатов в Integrative genomics viewer (IGV)

Примечание: (A) G>A замена на Chr4:125452634 в FAT4. (B) G>A замена на Chr11:125452634 в RAG1. (C) T>A замена на Chr1:9715914 в PIK3CD. (D) замена C>T на Chr1:36471505 в CSF3R. Одна ячейка на дорожке покрытия представляет гетерозиготную мутацию, а две ячейки на дорожке покрытия представляют гомозиготную мутацию.

По данным последней версии базы CADD, использующей референсный геном человека Ch38 (GRCh38-v1.5) мутация потенциально патогенная. Учитывая, что мутации в FAT4 ранее были определены, как значимые для синдрома Ханнекама, эту мутацию в данном исследовании оценили как потенциально значимую и её этиологическую значимость полностью отрицать было нельзя. Атипичный кадгерин 4 (FAT4) участвует в сигнальных путях Hippo и Warts, а бородавчатое образование ладонной поверхности правой кисти не имело инфекционной природы и могло

объясняться воздействиями нарушенных сигнальных путей, связанных с FAT4. Интересно, что кроме синдрома Хеннекама аллельная мутация в локусе FAT4 приводит к развитию синдрома Ван Мальдергема, который проявляется некоторыми из признаков синдрома Хеннекама, такими как дисморфизм лица и когнитивные нарушения, но при котором не бывает мальформаций лимфатической системы с симптомами лимфедемы и лимфангиэктазии [77].

Функциональные исследования вариантов гена FAT4 в рассматриваемом нами случае ранее выполнены не были. Проверка патогенности этой мутации для данного пациента была проведена методом Сенгера. При сопоставлении ДНК родителей, сибса и пробанда выяснилось, что все родственники являются носителями той же самой гомозиготной мутации FAT4 (g.125452634G>A), что и пациент, тогда как фенотип синдрома Хеннекама наблюдался только у пробанда. Было резюмировано, что G>A замена на Chr4:125452634 в FAT4 может быть не патогенная, и является полиморфизмом, вариантом нормы (рисунок 30).

Однако, идентичные мутации в одном и том же гене могут приводить к различным симптомам наследственного заболевания человека за счет снижения пенетрантности и переменной экспрессивности, которые вызываются такими факторами, как взаимодействия генов, факторы окружающей среды, аллельные вариации и стохастичность [197]. Следовательно, вопрос о том, могут ли гомозиготные мутации (g.125452634G> A) в гене FAT4 приводить к различным симптомам у разных людей, заслуживают дальнейшего анализа. Кроме того, комплексные генетические факторы и факторы окружающей среды, которые привели к снижению пенетрантности и переменной экспрессивности, необходимо изучить путем проведения когортных исследований.

Мутации в генах RAG1, PIK3CD и CSF3R также потенциально важны для формирования фенотипического комплекса признаков патологии в

исследуемом нами случае. Обычно ген активации рекомбинации 1 (RAG1) играет значительную роль в формировании комплекса белков для процесса V (D) J рекомбинации в лимфоцитах. Тем не менее, известно, что дефекты RAG1 или RAG2 приводят к различным первичным иммунодефицитам, в том числе Т-В-ТКИН, повышенной восприимчивости к цитомегаловирусной инфекции, аутоиммунным заболеваниям, синдрому Оменн и др. [138], поэтому игнорировать значимость данной мутации преждевременно.

Мутации с усилением функции гена PIK3CD, который кодирует каталитические субъединицы P110 δ и P85 фосфоинозитид-3-киназы-дельта, крайне интересны в данном случае, так как затрагивают основной белок mTOR пути, который участвует в многочисленных процессах, связанных с клеточным стрессом и влияет на выживание и пролиферацию клеток.

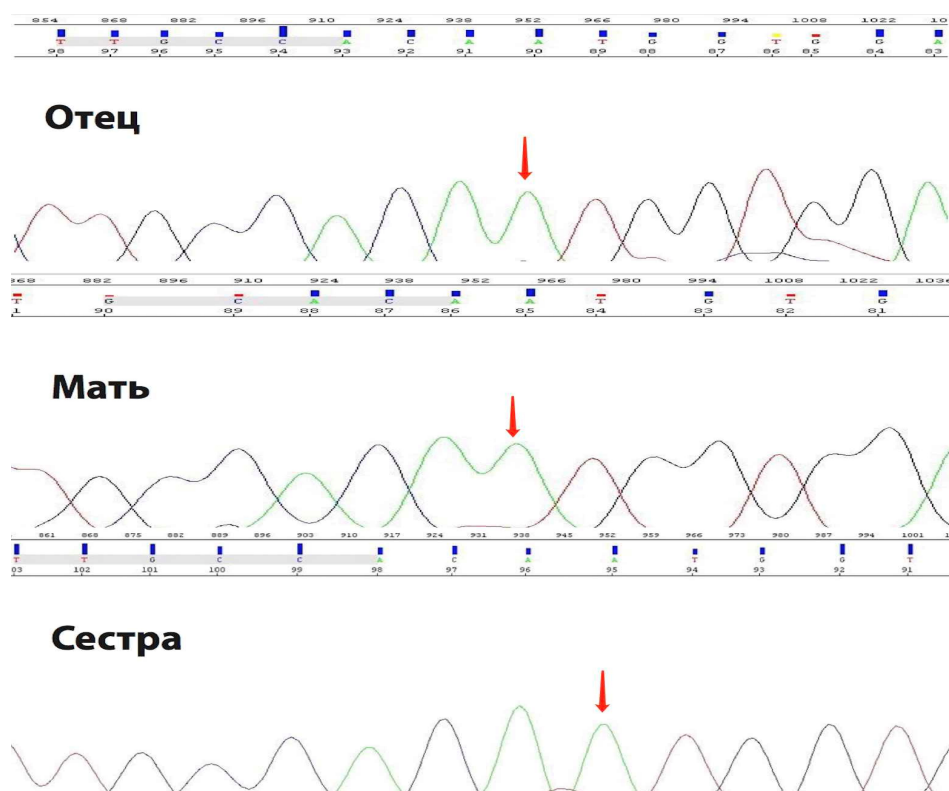


Рисунок 30 – Анализ нуклеотидной последовательности гена FAT4 у сибса и родителей пробанда с синдромом Хеннекама

Примечание: сплошная линия со стрелкой указывает на замену g.125452634G>A.

Среди первичных иммунодефицитов известен синдром активации фосфоинозид 3-киназы δ (ранее считавшийся одним из ОВИН), в фенотипе которого одним из лейтмотивов является лимфопролиферация, которая может приводить к В-лимфомам. Кроме того, для этого синдрома характерны гипогаммаглобулинемия, рецидивирующие вирусные инфекции, бронхоэктазы, герпесвирусные инфекции [11, 35, 51, 139]. При лечении синдрома активации фосфоинозид 3-киназы δ наряду с заместительной терапией иммуноглобулинами, используются препараты, ингибирующие mTOR – например, рапамицин (сиролимус, эверолимус). Ранее этот препарат использовался для профилактики отторжения аллотрансплантата при трансплантации почки и печени. В клинической иммунологии этот препарат стал применяться в последние 10 лет именно при лечении синдрома активации фосфоинозид 3-киназы δ . Однако последние сообщения исследователей указывают на то, что ингибиторы mTOR пути могут успешно использоваться при мальформациях лимфатических сосудов [14]. Данной пациентке был назначен эверолимус, который также показал себя эффективным, элиминировал бородавчатое разрастание на коже и сократил скорость увеличения лимфатических отеков, что позволило уйти от травмирующих и болезненных манипуляций лимфодренажа. Эффект препарата объясняется участием mTOR пути и гена PIK3CD в развитии лимфатических сосудов человека.

Мутации в гене CSF3R имеют значение для интрацитоплазматического домена G-CSFR, которые были открыты у больных с врожденной нейтропенией [58]. Также известно, что эти мутации могут проявляться как соматические мутации, связанные с прогрессированием лейкемии при врожденной нейтропении [33, 58, 72].

Резюме

В исследуемом нами случае пациентка имела проявления патологии в виде ассиметричной лимфангиэктазии и лимфедемы, бородавчатых образований, гирсутизма, гидроторакса и лимфатического асцита. Для лучшего понимания патогенеза и предоставления таргетных вариантов лечения, было выполнено полногеномное секвенирование. С использованием инструментов биоинформатики было выявлено несколько мутаций, которые могут иметь значение для формирования патологии в анализируемом нами случае первичного иммунодефицита у девочки 5 лет.

Таким образом, последовательное применения методов биоинформационного анализа позволило выявить четыре мутации в генах, известных, как патогенные при ПИД, включая синдром Хеннекама. Результаты, полученные в проведенном исследовании демонстрируют высочайшую актуальность применения новых фундаментальных междисциплинарных разработок для решения практических задач по верификации новой, не известной ранее патологии. Подобный междисциплинарный подход позволит получить практическое обоснование для применения новых нерегламентированных терапевтических технологий.

Список работ, опубликованных по 5-й главе

1. Молекулярно-генетическая природа синдрома Хеннекама / Лю Г., Тузанкина И.А., Болков М.А., Долгих М.А. // Трансляционная медицина. II Санкт-Петербургский лимфологический форум. Лимфология XXI века: новые подходы и актуальные исследования. – 2019. – прил. 2, окт. – С. 51.

2. Liu, G. A novel mutation in FAT4 caused a mild form of Hennekam syndrome with warty formations / G. Liu, M. Bolkov, I. Tuzankina // European journal of immunology. Abstracts of IUIS 2019 Beijing 17th International Congress of Immunology. – 2019. – Vol. 49, suppl. 3. – P. 1228-1229.

ЗАКЛЮЧЕНИЕ

От «Проекта генома человека» к «Инициативе точной медицины» научное и медицинское сообщество шаг за шагом движется в соответствии с установленными целями к «индивидуализированной медицине». С разработкой высокопроизводительной технологии секвенирования нового поколения началась генерация и накопление огромных объемов геномных данных, что обуславливает необходимость их интерпретации и решение более сложных технических проблем. Разумная интерпретация этих крупномасштабных и сложных многомерных данных становится ключевым шагом в современных исследованиях различной патологии человека, включая иммунозависимые заболевания, в том числе онкопатологию, аутовоспалительные, аутоиммунные болезни и первичные иммунодефициты.

Исследованы патофизиологические генозависимые механизмы отдельных типов иммуноопосредованной патологии – мышечно-инвазивного рака мочевого пузыря и ряда первичных иммунодефицитов.

Проведенные исследования позволили взглянуть на процессы системной биологии с патофизиологической точки зрения и найти общие закономерности. Как первичные иммунодефициты, так и онкопатология могут быть представлены в единой схеме на уровне общих типовых патологических процессов, таких как тканевой стресс и его участие в общих сигнальных и метаболических путях.

Выявленные в исследованиях гены с повышенной экспрессией кодируют молекулы, активность которых характерна для эпителиально-мезенхимального перехода, межклеточной адгезии, регуляции окислительного фосфорилирования, активации транскрипции и блокирования апоптоза. Все молекулы являются физиологическими и работают при нормальных условиях в клетках во многих тканях организма. Однако, они меняют своё значение в контексте опухолевого роста и являются свидетелями типичного патологического процесса.

Наши результаты исследований 403 случаев мышечно-инвазивного рака мочевого пузыря (МИРМП) позволили классифицировать МИРМП на базальный и люминальный подтип. При базальном подтипе МИРМП выявлены более высокие значения экспрессии иммуносупрессивных молекул («контрольных точек иммунного ответа»), чем при люминальном подтипе, что необходимо учитывать при иммунотерапии опухолей. Базальные подтипы рака мочевого пузыря, по данным наших исследований, характеризовались повышенной экспрессией генов, относящихся к процессу эпителиально-мезенхимального перехода, метастазирования и функционированию иммунной системы. Люминальные подтипы характеризовались относительно повышенной экспрессией генов, отвечающих за метаболические процессы. Эти результаты обеспечивают более глубокое понимание гетерогенности рака мочевого пузыря.

На основании анализа выживания пациентов с мышечно-инвазивным раком мочевого пузыря (МИРМП), различной экспрессии иммунных контрольных точек в тканях при МИРМП, а также разницы в экспрессии различных РНК, находящихся в условиях перекрестного взаимовлияния, мы смогли глубже понять гетерогенную природу рака мочевого пузыря. Этот факт оказывается значительным в решении вопросов прогнозирования, точности диагностики рака мочевого пузыря и определении терапевтической программы, эффективной в купировании патологического процесса и повышении выживаемости пациентов.

Общая клиническая картина формируется при участии огромной сети взаимодействий регуляторных механизмов, включая такие некодирующие РНК, как мкРНК и днРНК. Мы идентифицировали перекрестное взаимовлияние мРНК, мкРНК и днРНК, включающее мРНК генов GATA3, CLIC4, PALLD, мкРНК MIR-200c-3p, miRNA-141-3p, miRNA-141-5p, и днРНК AC010326.3, AC073335.2 и MIR100HG. Уровень экспрессии генов в этой перекрестной взаимозависимости значительно различался между

базальным и люминальным подтипами и имел тесную связь с прогнозом МИРМП. Эти результаты свидетельствуют о том, что перекрестное взаимовлияние некодирующих РНК может играть значительную роль в онкогенезе и прогрессировании опухоли, а соответствующие молекулы могут служить прогностическими биомаркерами мышечно-инвазивного рака мочевого пузыря, используемыми для классификации подтипов патологии.

Исследуя плотности сети и так называемое биологическое (а по сути функциональное) расстояние между генами, значимыми для формирования первичных иммунодефицитов (ПИД) и, в частности, общего варибельного иммунодефицита (ОВИН), было обнаружено, что гены ОВИН функционально более похожи друг на друга и тесно взаимодействуют друг с другом по сравнению с другими генами ПИД. Это соответствует представлению о том, что ОВИН является мультигенной группой нозологических синдромов, чем большинство заболеваний ПИД. С другой стороны, это ещё раз подтверждает мнение о том, что ОВИН является скорее диагнозом исключения, объединяющим огромную когорту разнообразных нозологий с более или менее общими фенотипическими проявлениями, для которых ещё не определена генетическая причина и соответствующая таргетная терапия.

Кроме того, нами были разработаны методы, которые объединяют данные ко-экспрессии, межгенных взаимодействий, а также данные, характеризующие функции генов, позволяющие предсказать причинно значимые гены при развитии ОВИН. Были определены 172 новых гена-кандидата ОВИН, которые находятся во взаимозависимостях с известными генами ОВИН в одних и тех же биологических путях и демонстрируют высокую биологическую корреляцию с известными генами ОВИН. Несколько методов проверки анализа и подтверждение нескольких предсказанных в нашем исследовании генов в клинических исследованиях

других исследований демонстрирует важность прогнозируемых генов-кандидатов.

Данное исследование способствовало лучшему пониманию молекулярных механизмов генов ОВИН, расширив спектр кандидатных генов.

Используя апробированные на теоретических исследованиях методы биоинформационного анализа, был проанализирован конкретный случай с редким сочетанием фенотипических признаков в одном симптомокомплексе - лимфангиэктазия-лимфедема, бородавчатые образования, гирсутизм, гидроторакс и лимфатический асцит. Для лучшего понимания патогенеза и обоснования необходимых терапевтических вмешательств, было выполнено полногеномное секвенирование. В результате биоинформационного анализа выявлено четыре мутации в генах, известных, как патогенные при ПИД, включая синдром Хеннекама. Выявленные мутации помогут приблизиться к более точному диагнозу и обосновать патогенетическое лечение пациента, а также позволят определить происхождение этих мутаций и исследовать молекулярные механизмы формирования болезни.

Кроме того, исследования редких случаев патологии человека позволяют решить общепатологические вопросы формирования болезней, обогатив патофизиологическую науку знаниями о закономерностях функционирования организма человека, о молекулярных механизмах формирования патологии, став объективными обоснованиями для разработки и применения прицельной терапевтической тактики, что открывает перспективу создания новых таргетных лекарственных средств.

Перспективы дальнейшей разработки темы

В нашем исследовании выявлена статистически значимая повышенная экспрессия молекул сигнальных путей цитокин-цитокиновых рецепторных взаимодействий, как при базальном типе МИРМП, так и при ОВИН.

Продолжение исследований в данной области поможет раскрыть общие механизмы развития такой иммунопатологии, как рак, в том числе развившейся как осложнение первичных иммунодефицитов.

Подход, примененный для выявления механизмов иммунопатогенеза группы синдромов «Общая переменная иммунная недостаточность» может быть экстраполирован для исследования формирования фенотипа других синдромов первичных иммунодефицитов.

Выявление конкретных генов синдромов ОВИН и определение молекулярных основ иммунопатологии позволят изучить патогенез, дифференцировать нозологические формы из обширной неоднородной группы ОВИН и приблизить создание специфической таргетной терапии, гораздо более эффективной, чем традиционная заместительная терапия внутривенными или подкожными иммуноглобулинами. Это позволит нивелировать клинические проявления патологии, улучшить прогноз и качество жизни пациентов.

Процесс верификации генов первичных иммунодефицитов может быть усовершенствован разработкой программного обеспечения для прогнозирования генов-кандидатов различной иммунопатологии, обеспечивающего возможность его эффективного применения в клинических исследованиях.

Выявленные закономерности молекулярно-генетических основ иммунопатологии при формировании разных подтипов мышечно-инвазивного рака мочевого пузыря в дальнейшем могут быть продолжены экспериментальными исследованиями. Взаимодействия микро-РНК miR-200c и miR-141, обнаруженные с помощью вычислительного анализа, также должны быть подтверждены биологическими экспериментальными исследованиями в дальнейшем.

Примененный комплексный подход выявления закономерностей формирования иммунопатологии может быть использован для исследований любых других нарушений иммунитета.

Относительно выявления причин развития синдромов мальформации лимфатических сосудов и наследственной лимфедемы при трактовке полученных результатов необходимо иметь в виду, что возможны другие формы передачи наследственной информации, в том числе аутосомно-доминантные, ко-доминантные и ко-экспрессивные. Это определяет необходимость продолжения поиска патогенетических основ формирования иммунопатологии с учетом новых данных о механизмах патогенеза и вовлеченных сигнальных путях, используя биоинформационный анализ и накопленный клинический опыт выявления редкой патологии, связанной с врожденными ошибками иммунитета. Кроме того, согласно недавним исследованиям, вариации ДНК вне экзонов также могут привести к генетическим нарушениям, что побуждает проведение полномасштабных исследований генома человека на международном уровне.

Таким образом, проведенные нами исследования позволили сделать следующие выводы.

ВЫВОДЫ

1. Опухолевые макрофаги, относящиеся к семейству M2, значительно коррелировали с выживаемостью пациентов, злокачественностью, стадией опухоли мышечно-инвазивного рака мочевого пузыря и являются маркерами опухолевого процесса при мышечно-инвазивном раке мочевого пузыря.

2. Перекрестное взаимовлияние различных кодирующих (GATA3, PALLD, CLIC4) и некодирующих РНК (мкРНК miR-200c-3p, miR-141-3p, miR-141-5p, днРНК AC010326.3, AC073335.2 и MIR100HG) играет значительную роль в онкогенезе и прогрессировании опухоли, а выявленные биомаркеры важны при определении подтипов мышечно-инвазивного рака мочевого пузыря и прогнозировании развития патологии.

3. Экспрессии иммуносупрессивных молекул PD-1, PD-L1, CTLA-4, HAVCR-2 и LAG-3 позволяют различать базальные и люминальные подтипы мышечно-инвазивного рака мочевого пузыря - базальные опухоли демонстрируют более высокие уровни экспрессии иммуносупрессивных молекул, чем опухоли люминального подтипа.

4. В основе мультигенной природы общей варибельной иммунной недостаточности лежит функциональная общность генов и их более тесное взаимодействие между собой на уровне сигнальных путей, по сравнению с другими генами первичных иммунодефицитов.

5. Прогнозирование генов-кандидатов группы первичных иммунодефицитов «Общая варибельная иммунная недостаточность», обеспеченное комплексным анализом развития иммунопатологии, данных ко-экспрессии и взаимодействий генов, сведениями об их биологической роли, позволило идентифицировать 172 гена-кандидата.

6. Идентифицированы «вероятно-патогенные» мутации в генах FAT4, RAG1, PIK3CD и CSF3R, которые могут иметь клиническое значение и приводить к фенотипу первичного иммунодефицита синдрома Хеннекама.

ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ

1. Перед началом иммунотерапии при мышечно-инвазивном раке мочевого пузыря следует определить подтип рака (базальный или люминальный) методом определения биомаркеров, что позволит сократить стоимость лечения и предупредить развитие побочных эффектов.

2. Для определения подтипа мышечно-инвазивного рака мочевого пузыря необходимо исследование следующих биомаркеров заболевания: GATA3, PALLD, CLIC4, miR-200c-3p, miR-141-3p, miR-141-5p, AC010326.3, AC073335.2 и MIR100HG.

3. После определения клинического фенотипа первичного иммунодефицита при проведении исследований по прогнозированию генов-кандидатов общего вариабельного иммунодефицита необходимо включать в анализ факторы ко-экспрессии, межбелкового взаимодействия и анализ сигнальных путей.

4. Для дифференциальной диагностики синдромов группы первичных иммунодефицитов «Общая вариабельная иммунная недостаточность», необходимо дополнительно включить в перечень тестируемых генов 172 гена-кандидата, выявленных в нашем исследовании: LMNA, BAD, PRKAB2, CRK, CRKL, CASP8, CTNNB1, STAT3, HRAS, PTK2, JAK1, BCL2, PDGFRB, CHUK, JAK2, GRB2, SRC, RELA, NFKBIA, EGFR, CREBBP, EP300, MAPK1, IKBKB, PIK3CA, SHC1, PIK3CB, IL10, ERBB2, PPP1CA, TRAF6, INHBA, GNAQ, VAV2, TGFBR1, TGFBR2, BCAR1, NOS2, FYN, SOCS1, SMAD4, TAB2, TNFRSF1A, TNF, HLA-DRA, FLNA, ITGB1, ACTN1, AKT1, SP1, PIK3CG, BRAF, INHBC, CD47, MET, PIP5K1C, SOS1, GP6, JUN, ICAM1, GP1BA, PRKACA, PRKCA, PRKAG1, TYK2, KDR, ACTB, MAPK3, PLCB1, RAC1, ADCY7, RAF1, SDC4, IGF1R, ITGA6, ITGB4, ITGB7, PAK1, AKT3, CDC42, SMAD1, CCR5, CASP3, ZYX, ITGB2, RAP1A, PDGFRA, MAPK8, VAV3, CASP9, GSK3B, IRAK1, MYD88, TLR9, PRKCB, BMP7, RHOA,

NFKBIB, MAPK14, BMP6, INHBB, GDF5, TLR4, CAV1, CD40, ITGA4, C8B, C9, IL10RB, PECAM1, PPP2CA, CUL1, IL1R1, CD14, HSPA1B, BIRC3, HSPA1L, IRAK4, BMP2, SOS2, MYLK, PIK3R3, RAPGEF1, MAPK13, GNAS, HSPA2, MAPK9, ELK1, MYC, PPP1CB, CCND3, MYLK2, HSPA6, BIRC2, DAG1, FLNB, FLNC, ID4, XIAP, ACTN4, SMAD3, MAPK12, PPP2R1A, PRKACB, MAPK10, MAP2K1, MAP2K6, MAP3K7, BMPR1A, HSPB1, PRKCG, ACVR1, PPP1CC, RPS6KB1, HLA-DRB5, SERPINB9, FN1, COL6A2, VCAM1, RAB7A, ACTG1, MYL12A, PPP2R1B, SKP1, ALOX5, RAB5A, PAK2, RBX1, SMURF1, EMD, TPM2, IFNGR.

СПИСОК СОКРАЩЕНИЙ

ВОЗ	Всемирная Организация Здравоохранения
днРНК	длинные некодирующие рибонуклеиновые кислоты
МИРМП	Мышечно-инвазивный рак мочевого пузыря
мкРНК	микро рибонуклеиновые кислоты
мРНК	Матричные (ядерные) рибонуклеиновые кислоты
ОВИН	Общий вариабельный иммунодефицит
ПИД	Первичные иммунодефициты
РМП	Рак мочевого пузыря
ТКИН (SCID)	Тяжелый комбинированный иммунодефицит
AUC	Площадь под кривой ROC
BWA	Burrows-Wheeler Aligner – выравнивание Барроуза-Уилера
DEG	Метод оценки дифференциально экспрессируемых генов
ESID	Европейское общество иммунодефицитов
FGA	functional genomics alignmen – функциональные геномное выравнивание относительно стандартного генома
GATK4	The Genome Analysis Toolkit – Программа для качественного определения вариаций в геноме из данных секвенирования
GEO	Gene Expression Omnibus – База данных экспрессии генов
GSEA	gene set enrichment analysis – Анализ обогащения по функциональной принадлежности, анализ дополнения генетической информации биологическим значением
IGV	integrated genomics viewer – Программа для просмотра секвенированной информации
InDels	insertions/deletions – варианты инсерций/делеций
JMF	Jeffrey Modell Foundation – это общественная благотворительная организация, занимающаяся ранней точной диагностикой и лечением посредством клинических и фундаментальных исследований, обучения врачей, поддержки пациентов, защиты интересов общественности и просвещения

KEGG	Киотская энциклопедия генов и геномов
K-M	Кривая выживаемости Каплана-Мейера
LOF	Мутации с потерей функции (lost of function)
NGS	Высокопроизводительное секвенирование
NIH	Национальный институт здравоохранения США
PCC	Pearson correlation coefficient – Коэффициент корреляции Пирсона
PPI	Protein-protein interaction – Белок-белковые взаимодействия или межбелковые взаимодействия
RF	Алгоритм машинного обучения Random forest
ROC	Receiver operating characteristic рабочая характеристика входных данных или кривая зависимости доли верных положительных классификаций от доли ложных положительных классификаций при варьировании порога решающего правила
SAM	Sequence Alignment Map – карта выравнивания секвенированной последовательности
SNP	single nucleotide polymorphism – однонуклеотидный полиморфизм
SNVs	single nucleotide variants – однонуклеотидные замены
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins – база данных и веб-ресурс для поиска информации об известных и предсказанных белок-белковых взаимодействиях
TCGA	The Cancer Genome Atlas – Атлас акового генома
XG/XGBoost	Алгоритм машинного обучения «eXtreme Gradient Boosting»

Сокращения генов:

LMNA, BAD, PRKAB2, CRK, CRKL, CASP8, CTNNB1, STAT3, HRAS, PTK2, JAK1, BCL2, PDGFRB, CHUK, JAK2, GRB2, SRC, RELA, NFKBIA, EGFR, CREBBP, EP300, MAPK1, IKBKB, PIK3CA, SHC1, PIK3CB, IL10, ERBB2, PPP1CA, TRAF6, INHBA, GNAQ, VAV2, TGFBR1, TGFBR2, BCAR1, NOS2, FYN, SOCS1, SMAD4, TAB2, TNFRSF1A, TNF, HLA-DRA, FLNA, ITGB1,

ACTN1, AKT1, SP1, PIK3CG, BRAF, INHBC, CD47, MET, PIP5K1C, SOS1, GP6, JUN, ICAM1, GP1BA, PRKACA, PRKCA, PRKAG1, TYK2, KDR, ACTB, MAPK3, PLCB1, RAC1, ADCY7, RAF1, SDC4, IGF1R, ITGA6, ITGB4, ITGB7, PAK1, AKT3, CDC42, SMAD1, CCR5, CASP3, ZYX, ITGB2, RAP1A, PDGFRA, MAPK8, VAV3, CASP9, GSK3B, IRAK1, MYD88, TLR9, PRKCB, BMP7, RHOA, NFKBIB, MAPK14, BMP6, INHBB, GDF5, TLR4, CAV1, CD40, ITGA4, C8B, C9, IL10RB, PECAM1, PPP2CA, CUL1, IL1R1, CD14, HSPA1B, BIRC3, HSPA1L, IRAK4, BMP2, SOS2, MYLK, PIK3R3, RAPGEF1, MAPK13, GNAS, HSPA2, MAPK9, ELK1, MYC, PPP1CB, CCND3, MYLK2, HSPA6, BIRC2, DAG1, FLNB, FLNC, ID4, XIAP, ACTN4, SMAD3, MAPK12, PPP2R1A, PRKACB, MAPK10, MAP2K1, MAP2K6, MAP3K7, BMPR1A, HSPB1, PRKCG, ACVR1, PPP1CC, RPS6KB1, HLA-DRB5, SERPINB9, FN1, COL6A2, VCAM1, RAB7A, ACTG1, MYL12A, PPP2R1B, SKP1, ALOX5, RAB5A, PAK2, RBX1, SMURF1, EMD, TPM2, IFNGR, CCBE1, FAT4, ADAMTS3, GATA3, PALLD, CLIC4 – названия генов приведены в соответствии с номенклатурой HUGO Gene Nomenclature Committee (HGNC).

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Латышева, Е.А. Первичные иммунодефициты: состояние проблемы на сегодняшний день. JMF-центры в России / Е.А. Латышева // Вопросы современной педиатрии. – 2013. – Т. 12, № 6. – С. 73-77.
2. Современные возможности скрининга и диагностики первичных иммунодефицитных состояний в педиатрии / Е.В. Чернышова, [и др.] // Педиатрия. – 2016. – Т. 95, № 3. – С. 124-128.
3. Тузанкина, И.А. К вопросу диагностики иммунопатологии / И.А. Тузанкина // Медицинская иммунология. – 2010. – Т. 12, № 6. – С. 485-496.
4. Тузанкина, И.А. Первичные иммунодефициты в раннем возрасте / И.А. Тузанкина, С.С. Дерябина, М.А. Болков. – Москва, 2018. – 176 с.
5. Щербина, А.Ю. Первичные иммунодефициты-реалии XXI века / А.Ю. Щербина // Вопросы гематологии/онкологии и иммунопатологии в педиатрии. – 2016. – Т. 15, № 1. – С. 8-9.
6. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? / L. Salmena, [et al.] // Cell. – 2011. – Vol. 146, № 3. – P. 353-358.
7. A long noncoding RNA activated by TGF- β promotes the invasion-metastasis cascade in hepatocellular carcinoma / J. Yuan, [et al.] // Cancer cell. – 2014. – Vol. 25, № 5. – P. 666-681.
8. A molecular taxonomy for urothelial carcinoma / G. Sjödaahl, [et al.] // Clinical Cancer Research. – 2012. – Vol. 18, № 12. – P. 3377-3386.
9. A multivariate approach to the integration of multi-omics datasets / C. Meng, [et al.] // BMC Bioinformatics. – 2014. – Vol. 15, № 1.
10. A proteome-scale map of the human interactome network / T. Rolland, [et al.] // Cell. – 2014. – Vol. 159, № 5. – P. 1212-1226.
11. Activating PI3K δ mutations in a cohort of 669 patients with primary immunodeficiency / M. Elgizouli, [et al.] // Clinical and Experimental Immunology. – 2016. – Vol. 183, № 2. – P. 221-229.
12. An antibody-deficiency syndrome due to mutations in the CD19 gene

/ M.C. Van Zelm, [et al.] // New England Journal of Medicine. – 2006. – Vol. 354, № 18. – P. 1901-1912.

13. An extensive comparison of recent classification tools applied to microarray data / J.W. Lee, [et al.] // Computational Statistics and Data Analysis. – 2005. – Vol. 48, № 4. – P. 869-885.

14. An updated review on activated PI3 kinase delta syndrome (APDS) / A. Singh, [et al.]. // Genes & Diseases. – 2019.

15. Anastasiadis A. Best practice in the treatment of nonmuscle invasive bladder cancer / A. Anastasiadis, T.M. de Reijke // Therapeutic Advances in Urology. – 2012. – Vol. 4, № 1. – P. 13-32.

16. Application of whole genome and RNA sequencing to investigate the genomic landscape of common variable immunodeficiency disorders / P.A. van Schouwenburg, [et al.] Clinical Immunology. – 2015. – Vol. 160, № 2. – P. 301-314.

17. Approved checkpoint inhibitors in bladder cancer: which drug should be used when? / P. Ghatalia, [et al.] // Therapeutic advances in medical oncology. – 2018. – Vol. 10. – P. 1758835918788310.

18. Autosomal recessive intestinal lymphangiectasia and lymphedema, with facial anomalies and mental retardation / R.C.M. Hennekam, [et al.] // American Journal of Medical Genetics. – 1989. – Vol. 34, № 4. – P. 593-600.

19. Ballgown bridges the gap between transcriptome assembly and expression analysis / A.C. Frazee [et al.]. // Nature biotechnology. – 2015. – Vol. 33, № 3. – P. 243.

20. Bioinformatics for cancer immunology and immunotherapy / P. Charoentong, [et al.] // Cancer Immunology, Immunotherapy. – 2012. – Vol. 61, № 11. – P. 1885-1903.

21. Bioinformatics tools and databases for analysis of next-generation sequence data / H.C. Lee, [et al.] // Briefings in Functional Genomics. – 2012. – Vol. 11, № 1. – P. 12-24.

22. Bladder cancer / A.M. Kamat, [et al.] // *Lancet*. – 2016. – Vol. 6736, № 16. – P. 1-15.
23. Boyle J.M. Population Prevalence of Diagnosed Primary Immunodeficiency Diseases in the United States / J.M. Boyle, R.H. Buckley // *Journal of Clinical Immunology*. – 2007. – Vol. 27, № 5. – P. 497-502.
24. Brief report Novel HAX1 mutations in patients with severe congenital neutropenia reveal isoform-dependent genotype-phenotype associations / M. Germeshausen, [et al.] // *Blood*. – 2008. – Vol. 111, № 10. – P. 4954-4957.
25. Cairns R.A. Regulation of cancer cell metabolism / R.A. Cairns, I.S. Harris, T.W. Mak // *Nature Reviews Cancer*. – 2011. – Vol. 11, № 2. – P. 85.
26. Cancer statistics in China, 2015 / W. Chen, [et al.] // *CA: A Cancer Journal for Clinicians*. – 2016. – Vol. 66, № 2. – P. 115-132.
27. Care M.A. Parsimonious Gene Correlation Network Analysis (PGCNA): a tool to define modular gene co-expression for refined molecular stratification in cancer / M.A. Care, D.R. Westhead, R.M. Tooze // *NPJ systems biology and applications*. – 2019. – Vol. 5, № 1. – P. 13.
28. CDG: An online server for detecting biologically closest disease-causing genes and its application to primary immunodeficiency / D. Requena, [et al.] // *Frontiers in Immunology*. – 2018. – Vol. 9, № JUN. – P. 1340.
29. Chen T. Xgboost: a scalable tree boosting system In: *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining* / T. Chen, C. Guestrin // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. – 2016. – P. 785–794.
30. Circular RNA is enriched and stable in exosomes: A promising biomarker for cancer diagnosis / Y. Li, [et al.] // *Cell research*. – 2015. – Vol. 25, № 8. – P. 981.
31. Classification of Small GTPases with Hybrid Protein Features and Advanced Machine Learning Techniques / Z. Liao, [et al.] // *Current*

Bioinformatics. – 2017. – Vol. 13, № 5. – P. 492-500.

32. Clinical picture and treatment of 2212 patients with common variable immunodeficiency / B. Gathmann, [et al.] // Journal of Allergy and Clinical Immunology. – 2014. – Vol. 134, № 1. – P. 116-126.

33. Clinical Relevance of Point Mutations in the Cytoplasmic Domain of the Granulocyte Colony-Stimulating Factor Receptor Gene in Patients With Severe Congenital Neutropenia / N. Tidow, [et al.] // Blood. – 1997. – Vol. 89, № 7. – P. 2369-2375.

34. Clinical significance of immune cell infiltration within gallbladder cancer / Y. Nakakubo, [et al.] // British Journal of Cancer. – 2003. – Vol. 89, № 9. – P. 1736-1742.

35. Clinical spectrum and features of activated phosphoinositide 3-kinase δ syndrome: A large patient cohort study / T.I. Coulter, [et al.] // Journal of Allergy and Clinical Immunology. – 2017. – Vol. 139, № 2. – P. 597-606.

36. ClusterProfiler: An R package for comparing biological themes among gene clusters / G. Yu, [et al.] // OMICS A Journal of Integrative Biology. – 2012. – Vol. 16, № 5. – P. 284-287.

37. Coexpresion analysis of human genes across many microarray data sets / H.K. Lee, [et al.] // Genome Research. – 2004. – Vol. 14, № 6. – P. 1085-1094.

38. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data / B. Wu, [et al.] // Bioinformatics. – 2003. – Vol. 19, № 13. – P. 1636-1643.

39. Comprehensive analysis of differentially expressed profiles of lncRNAs and circRNAs with associated co-expression and ceRNA networks in bladder carcinoma / M. Huang, [et al.] // Oncotarget. – 2016. – Vol. 7, № 30. – P. 47186-47200.

40. Comprehensive bioinformatics analysis of the characterization and determination underlying mechanisms of over-expression and co-expression of

genes residing on 20q in colorectal cancer / D. Li, [et al.] // *Oncotarget*. – 2017. – Vol. 8, № 45. – P. 78642-78659.

41. Comprehensive molecular characterization of gastric adenocarcinoma / A.J. Bass, [et al.] // *Nature*. – 2014. – Vol. 513, № 7517. – P. 202-209.

42. Comprehensive molecular characterization of human colon and rectal cancer / D.M. Muzny, [et al.] // *Nature*. – 2012. – Vol. 487, № 7407. – P. 330-337.

43. Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer / A.G. Robertson, [et al.] // *Cell*. – 2017. – Vol. 171, № 3. – P. 540-556.

44. Comprehensive molecular characterization of urothelial bladder carcinoma / J.N. Weinstein, [et al.] // *Nature*. – 2014. – Vol. 507, № 7492. – P. 315-322.

45. Comprehensive molecular portraits of human breast tumours / D.C. Koboldt, [et al.] // *Nature*. – 2012. – Vol. 490, № 7418. – P. 61-70.

46. Conley, M.E. Diagnostic Criteria for Primary Immunodeficiencies / M.E. Conley, L.D. Notarangelo, A. Etzioni // *Clinical Immunology*. – 1999. – Vol. 93, № 3. – P. 190-197.

47. Construction of differential mRNA-lncRNA crosstalk networks based on ceRNA hypothesis uncover key roles of lncRNAs implicated in esophageal squamous cell carcinoma / S. Yang, [et al.] // *Oncotarget*. – 2016. – Vol. 7, № 52. – P. 85728-85740.

48. Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer / E.D. Wiklund, [et al.] // *International Journal of Cancer*. – 2011. – Vol. 128, № 6. – P. 1327-1334.

49. Csardi, G. The igraph software package for complex network research / G. Csardi, T. Nepusz // *InterJournal, Complex Systems*. – 2006. – Vol. 1695, № 5. – P. 1-9.

50. Cunningham-Rundles, C. How I treat common variable immune deficiency / C. Cunningham-Rundles // *Blood*. – 2010. – Vol. 116, № 1. – P. 7-15.

51. CVID-associated tumors: Czech nationwide study focused on epidemiology, immunology, and genetic background in a cohort of patients with CVID / P. Kralickova, [et al.] // *Frontiers in Immunology*. – 2019. – Vol. 10, № JAN. – P. 3135.

52. Cytoscape: A software Environment for integrated models of biomolecular interaction networks / P. Shannon, [et al.] // *Genome Research*. – 2003. – Vol. 13, № 11. – P. 2498-2504.

53. Delayed diagnosis and complications of predominantly antibody deficiencies in a cohort of Australian adults / C.A. Slade, [et al.] // *Frontiers in Immunology*. – 2018. – Vol. 9. – P. 694.

54. Developing a Multi-Dose Computational Model for Drug-induced Hepatotoxicity Prediction based on Toxicogenomics Data / R. Su, [et al.] // *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. – 2018. – Vol. 16, № 4.– P. 1.

55. Development of cancer in patients with primary immunodeficiencies / K. Salavoura, [et al.] // *Anticancer Research*. – 2008. – Vol. 28, № 2B. – P. 1263-1269.

56. Disturbed canonical nuclear factor of κ light chain signaling in B cells of patients with common variable immunodeficiency / B. Keller, [et al.] // *Journal of Allergy and Clinical Immunology*. – 2017. – Vol. 139, № 1. – P. 220-231.

57. DNA physical properties outperform sequence compositional information in classifying nucleosome-enriched and -depleted regions / G. Liu, [et al.] // *Genomics*. – 2019. – Vol. 111, № 5. – P. 1167-1175.

58. Dong F. Identification of a nonsense mutation in the granulocyte-colony-stimulating factor receptor in severe congenital neutropenia / F. Dong, L. Hoefsloot // *Proceedings of the National Academy of Sciences*. – 1994. – Vol. 91, № 10. – P. 4480-4484.

59. Dysfunctional BLK in common variable immunodeficiency perturbs B-cell proliferation and ability to elicit antigen-specific CD4⁺ T-cell help /

E.B. Compeer, [et al.] // *Oncotarget*. – 2015. – Vol. 6, № 13. – P. 10759-10771.

60. Epidemiology of bladder cancer: a systematic review and contemporary update of risk factors in 2018 / M. Cumberbatch, [et al.] // *European urology*. – 2018. – Vol. 74, № 6. – P. 784-795.

61. Epithelial-mesenchymal transition, a novel target of sulforaphane via COX-2/MMP2, 9/Snail, ZEB1 and miR-200c/ZEB1 pathways in human bladder cancer cells / Y. Shan, [et al.] // *Journal of Nutritional Biochemistry*. – 2013. – Vol. 24, № 6. – P. 1062-1069.

62. Evaluation of miR-141, miR-200c, miR-30b expression and clinicopathological features of bladder cancer / A. Mahdavinezhad, [et al.] // *International journal of molecular and cellular medicine*. – 2015. – Vol. 4, № 1. – P. 32.

63. Exome Sequencing Analysis Reveals Variants in Primary Immunodeficiency Genes in Patients with Very Early Onset Inflammatory Bowel Disease / J.R. Kelsen, [et al.] // *Gastroenterology*. – 2015. – Vol. 149, № 6. – P. 1415-1424.

64. Familial risk and heritability of cancer among twins in Nordic countries / L. Mucci, [et al.] // *Jama*. – 2016. – Vol. 315, № 1. – P.68-76.

65. Floyd, E. Development and Use of Biomarkers in Oncology Drug Development / E. Floyd, T.M. Mcshane // *Toxicologic Pathology*. – 2004. – Vol. 32, № suppl_1. – P. 106-115.

66. Friedrich, M.J. Immunotherapy 2.0: Improving the Response to Checkpoint Inhibitors / M.J. Friedrich // *Jama*. – 2019. – Vol. 321, № 2. – P. 131-133.

67. Frost, H.R. A multi-omics approach for identifying important pathways and genes in human cancer / H.R. Frost, C.I. Amos // *BMC Bioinformatics*. – 2018. – Vol. 19, № 1. – P. 479.

68. Gene co-expression analysis for functional classification and gene-disease predictions / S. van Dam, [et al.] // *Briefings in bioinformatics*. – 2018. –

Vol. 19, № 4. – P. 575-592.

69. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles / A. Subramanian, [et al.] // Proceedings of the National Academy of Sciences of the United States of America. – 2005. – Vol. 102, № 43. – P. 15545-15550.

70. Gene set enrichment analysis: Performance evaluation and usage guidelines / J.H. Hung, [et al.] // Briefings in Bioinformatics. – 2012. – Vol. 13, № 3. – P. 281-291.

71. Genes associated with common variable immunodeficiency: One diagnosis to rule them all? / D.J.A. Bogaert, [et al.] // Journal of Medical Genetics. – 2016. – Vol. 53, № 9. – P. 575-590.

72. Germeshausen M. Implications of Mutations in Hematopoietic Growth Factor Receptor Genes in Congenital Cytopenias / M. Germeshausen, M. Ballmaier, K. Welte // Annals of the New York Academy of Sciences. – 2001. – Vol. 938, № 1. – P. 305-321.

73. Global overview of primary immunodeficiencies: a report from Jeffrey Modell Centers worldwide focused on diagnosis, treatment, and discovery / V. Modell, [et al.] // Immunologic Research. – 2014. – Vol. 60, № 1. – P. 132-144.

74. Guidelines for genetic studies in single patients: Lessons from primary immunodeficiencies / J.L. Casanova, [et al.] // Journal of Experimental Medicine. – 2014. – Vol. 211, № 11. – P. 2137-2149.

75. Guidelines on non-muscle-invasive (TA, T1, CIS) bladder cancer / Babjuk, M. [et al.] // Eur Urol. – 2011. – Vol. 59. – № 4. – P. 584-594.

76. Haploinsufficiency of the NF- κ B1 Subunit p50 in Common Variable Immunodeficiency / M. Fliegauf, [et al.] // American Journal of Human Genetics. – 2015. – Vol. 97, № 3. – P. 389-403.

77. Hennekam syndrome can be caused by FAT4 mutations and be allelic to Van Maldergem syndrome / M. Alders, [et al.] // Human Genetics. – 2014. – Vol. 133, № 9. – P. 1161-1167.

78. Hennig, C. fpc: Flexible Procedures for Clustering / C. Hennig // R package version. – 2010. – Vol. 2, № 2. – P. 0-3.
79. Hermjakob, H. IntAct: an open source molecular interaction database / H. Hermjakob // Nucleic Acids Research. – 2004. – Vol. 32, № suppl_1. – P. 452D - 455.
80. High-dimensional variable selection for survival data / H. Ishwaran, [et al.] // Journal of the American Statistical Association. – 2010. – Vol. 105, № 489. – P. 205-217.
81. Hong, Y.M. Economic impact of tumor markers in bladder cancer surveillance / Y.M. Hong, K.R. Loughlin // Urology. – 2008. – Vol. 71, № 1. – P. 131-135.
82. Horvath, S. Geometric interpretation of gene coexpression network analysis / S. Horvath, J. Dong // PLoS Computational Biology. – 2008. – Vol. 4, № 8. – P. e1000117.
83. Human Protein Reference Database - 2009 update / T.S. Keshava Prasad, [et al.] // Nucleic Acids Research. – 2009. – Vol. 37, № suppl_1. – P. 767-772.
84. Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer / F. Crea, [et al.] // Oncotarget. – 2014. – Vol. 5, № 3. – P. 764-774.
85. Identification of Distinct Basal and Luminal Subtypes of Muscle-Invasive Bladder Cancer with Different Sensitivities to Frontline Chemotherapy / W. Choi, [et al.] // Cancer Cell. – 2014. – Vol. 25, № 2. – P. 152-165.
86. Identification of druggable cancer driver genes amplified across TCGA datasets / Y. Chen, [et al.] // PLoS ONE. – 2014. – Vol. 9, № 5. – P. e98293.
87. Identification of ovarian cancer driver genes by using module network integration of multi-omics data / O. Gevaert, [et al.] // Interface Focus. – 2013. – Vol. 3, № 4. – P. 20130013.
88. Identification of prognostic gene signatures of glioblastoma: A study

based on TCGA data analysis / Y.W. Kim, [et al.] // *Neuro-Oncology*. – 2013. – Vol. 15, № 7. – P. 829-839.

89. Impact of Molecular Subtypes in Muscle-invasive Bladder Cancer on Predicting Response and Survival after Neoadjuvant Chemotherapy / R. Seiler, [et al.] // *European Urology*. – 2017. – Vol. 72, № 4. – P. 544-554.

90. InnateDB: Systems biology of innate immunity and beyond - Recent updates and continuing curation / K. Breuer, [et al.] // *Nucleic Acids Research*. – 2013. – Vol. 41, № D1. – P. D1228-D1233.

91. Instruct: A database of high-quality 3D structurally resolved protein interactome networks / M.J. Meyer, [et al.] // *Bioinformatics*. – 2013. – Vol. 29, № 12. – P. 1577-1579.

92. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma / B.J. Raphael, [et al.] // *Cancer Cell*. – 2017. – Vol. 32, № 2. – P. 185-203.

93. Integrated multi-omics analysis of oligodendroglial tumours identifies three subgroups of 1p/19q co-deleted gliomas / A. Kamoun, [et al.] // *Nature Communications*. – 2016. – Vol. 19, № 7. – P. 11263.

94. Integrating coexpression networks with GWAS to prioritize causal genes in maize / R.J. Schaefer, [et al.] // *Plant Cell*. – 2018. – Vol. 30, № 12. – P. 2922-2942.

95. Integrative analysis of lncRNAs and miRNAs with coding RNAs associated with ceRNA crosstalk network in triple negative breast cancer / N. Yuan, [et al.] // *OncoTargets and therapy*. – 2017. – Vol. 10. – P. 5883–5897.

96. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks / X. Zeng, [et al.] // *Briefings in bioinformatics*. – 2015. – Vol. 17, № 2. – P. 193-203.

97. Integrative clustering reveals a novel split in the luminal A subtype of breast cancer with impact on outcome / M.R. Aure, [et al.] // *Breast Cancer*

Research. – 2017. – Vol. 19, № 1. – P. 1-18.

98. International Consensus Document (ICON): Common Variable Immunodeficiency Disorders / F.A. Bonilla, [et al.] // The Journal of Allergy and Clinical Immunology: In Practice. – 2016. – Vol. 4, № 1. – P. 38-59.

99. International Union of Immunological Societies: 2017 Primary Immunodeficiency Diseases Committee Report on Inborn Errors of Immunity / C. Picard, [et al.] // Journal of Clinical Immunology. – 2018. – Vol. 38, № 1. – P. 96-128.

100. Intrinsic basal and luminal subtypes of muscle-invasive bladder cancer / W. Choi, [et al.] // Nature Reviews Urology. – 2014. – Vol. 11, № 7. – P. 400.

101. Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology / J.S. Damrauer, [et al.] // Proceedings of the National Academy of Sciences of the United States of America. – 2014. – Vol. 111, № 8. – P. 3110-3115.

102. Invasive bladder cancer: genomic insights and therapeutic promise / J. Kim, [et al.] // Clin Cancer Res. – 2015. – Vol. 21, № 20. – P. 4514-4524.

103. Itan, Y. Novel primary immunodeficiency candidate genes predicted by the human gene connectome / Y. Itan, J.L. Casanova // Frontiers in Immunology. – 2015. – Vol. 6. – P. 142.

104. Johnston, R.J. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans* / R.J. Johnston, O. Hobert // Nature. – 2003. – Vol. 426, № 6968. – P. 845-849.

105. Kienzler, A.K. The role of genomics in common variable immunodeficiency disorders / A.K. Kienzler, C.E. Hargreaves, S.Y. Patel // Clinical and Experimental Immunology. – 2017. – Vol. 188, № 3. – P. 326-332.

106. Large scale comparison of gene expression levels by microarrays and rnaseq using TCGA data / Y. Guo, [et al.] // PLoS ONE. – 2013. – Vol. 8, № 8. – P. e71462.

107. Li, H. Fast and accurate short read alignment with Burrows-Wheeler transform / H. Li, R. Durbin // *Bioinformatics*. – 2009. – Vol. 25, № 14. – P. 1754-1760.
108. Liaw, A. Classification and regression by randomForest / A. Liaw, M. Wiener // *R news*. – 2002. – Vol. 2, № 3. – P. 18-22.
109. Lim M.S. The molecular pathology of primary immunodeficiencies / M.S. Lim, K.S.J. Elenitoba-Johnson // *The Journal of molecular diagnostics*. – 2004. – Vol. 6, № 2. – P. 59.
110. Lin, L.J. Clinical and immunological features of common variable immunodeficiency in China / L.J. Lin, Y.C. Wang, X.M. Liu // *Chinese Medical Journal*. – 2015. – Vol. 128, № 3. – P. 310-315.
111. LincRNAs: Systemic Computational Identification and Functional Exploration / H. Hu, [et al.] // *Current Bioinformatics*. – 2017. – Vol. 12, № 1. – P. 34-42.
112. Liu, G. Integrated multiple “-omics” data reveal subtypes of hepatocellular carcinoma / G. Liu, C. Dong, L. Liu // *PLoS ONE*. – 2016. – Vol. 11, № 11. – P. 1-15.
113. Long non-coding RNA urothelial cancer-associated 1 promotes bladder cancer cell migration and invasion by way of the hsa-miR-145-ZEB1/2-FSCN1 pathway / M. Xue, [et al.] // *Cancer Science*. – 2016. – Vol. 107, № 1. – P. 18-27.
114. Long noncoding RNA in prostate, bladder, and kidney cancer / E. Martens-Uzunova, [et al.] // *European urology*. – 2014. – Vol. 65, № 6. – P. 1140-1151.
115. Loss of ADAMTS3 activity causes Hennekam lymphangiectasia-lymphedema syndrome 3 / P. Brouillard, [et al.] // *Human molecular genetics*. – 2017. – Vol. 26, № 21. – P. 4095-4104.
116. Lymphedema-lymphangiectasia-mental retardation (Hennekam) syndrome: A review / I.D.C. Van Balkom, [et al.] // *American journal of medical*

genetics. – 2002. – Vol. 112, № 4. – P. 412-421.

117. Machine learning applications in cancer prognosis and prediction / K. Kourou, [et al.] // Computational and structural biotechnology journal. – 2015. – Vol. 13. – P. 8-17.

118. McCusker C. Primary immunodeficiency / C. McCusker, J. Upton, R. Warrington // Allergy, Asthma & Clinical Immunology. – 2018. – Vol. 14, № S2. – P. 61.

119. Mccusker, C. IMMUNOLOGY Primary immunodeficiency / C. Mccusker, R. Warrington // Allergy, Asthma & Clinical Immunology. – 2011. – Vol. 7, № suppl_1. – P. S11.

120. Meta-Analysis of the Luminal and Basal Subtypes of Bladder Cancer and the Identification of Signature Immunohistochemical Markers for Clinical Use / V. Dadhania, [et al.] // EBioMedicine. – 2016. – Vol. 12. – P. 105-117.

121. Michalak, P. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes / P. Michalak // Genomics. – 2008. – Vol. 91, № 3. – P. 243-248.

122. MicroRNAs: A new paradigm in the treatment and diagnosis of heart failure? / V. Oliveira-Carvalho, [et al.] // Arquivos Brasileiros de Cardiologia. – 2012. – Vol. 98, № 4. – P. 362-369.

123. MINT: The Molecular INTeraction database / A. Chatr-aryamontri, [et al.] // Nucleic Acids Research. – 2006. – Vol. 35, № suppl_1. – P. D572-D574.

124. MiR-200c Inhibits invasion, migration and proliferation of bladder cancer cells through down-regulation of BMI-1 and E2F3 / L. Liu, [et al.] // Journal of Translational Medicine. – 2014. – Vol. 12, № 1. – P. 305.

125. miR-34/449 miRNAs are required for motile ciliogenesis by repressing cp110 / R. Song, [et al.] // Nature. – 2014. – Vol. 510. – № 7503. – P. 115.

126. miRClassify: An advanced web server for miRNA family classification and annotation / Q. Zou, [et al.] // Computers in Biology and

Medicine. – 2014. – Vol. 45, № 1. – P. 157-160.

127. MiRWalk - Database: Prediction of possible miRNA binding sites by “walking” the genes of three genomes / H. Dweep, [et al.] // Journal of Biomedical Informatics. – 2011. – Vol. 44, № 5. – P. 839-847.

128. Molecular markers in bladder cancer: Novel research frontiers / F. Sanguedolce, [et al.] // Critical Reviews in Clinical Laboratory Sciences. – 2015. – Vol. 52, № 5. – P. 242-255.

129. Molecular portraits of human breast tumours / C.M. Perou, [et al.] // Nature. – 2000. – Vol. 406, № 6797. – P. 747-752.

130. Mooney, M.A. Gene set analysis: A step-by-step guide / M.A. Mooney, B. Wilmot // American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics. – 2015. – Vol. 168, № 7. – P. 517-527.

131. Multilevel Genomics-Based Taxonomy of Renal Cell Carcinoma / F. Chen, [et al.] // Cell Reports. – 2016. – Vol. 14, № 10. – P. 2476-2489.

132. Mutations in CCBE1 cause generalized lymph vessel dysplasia in humans / M. Alders, [et al.] // Nature Genetics. – 2009. – Vol. 41, № 12. – P. 1272-1274.

133. Nagata, M. Molecular biomarkers in bladder cancer: novel potential indicators of prognosis and treatment outcomes / M. Nagata, S. Muto, S. Horie // Disease Markers. – 2016. – Vol. 2016. – P. 1-5.

134. Naturally occurring mutation affecting the MyD88-binding site of TNFRSF13B impairs triggering of class switch recombination / M.B. Almejun, [et al.] // European Journal of Immunology. – 2013. – Vol. 43, № 3. – P. 805-814.

135. Neoadjuvant PD-1 blockade in resectable lung cancer / P.M. Forde, [et al.] // New England Journal of Medicine. – 2018. – Vol. 378, № 21. – P. 1976-1986.

136. Network-based approach to prediction and population-based validation of in silico drug repurposing / F. Cheng, [et al.] // Nature Communications. – 2018. – Vol. 9, № 1. – P. 1-12.

137. Newborn screening: Toward a uniform screening panel and system - Executive summary / M.S. Watson, [et al.] // *Pediatrics*. – 2006. – Vol. 117, № 5. – P. S296-S307.

138. Novel RAG1 mutation in a case of severe combined immunodeficiency / J. Zhang, [et al.] // *Pediatrics*. – 2005. – Vol. 116, № 3. – P. e445-e449.

139. Ochs, H.D. Common variable immunodeficiency (CVID): New genetic insight and unanswered questions / H.D. Ochs // *Clinical and Experimental Immunology*. – 2014. – Vol. 178, № S1. – P. 5-6.

140. Orang, A.V. Mechanisms of miRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation / A.V. Orang, R. Safaralizadeh, M. Kazemzadeh-Bavili // *International journal of genomics*. – 2014. – Vol. 2014.

141. Ortutay, C. Identification of candidate disease genes by integrating Gene Ontologies and protein-interaction networks: case study of primary immunodeficiencies / C. Ortutay, M. Vihinen // *Nucleic acids research*. – 2008. – Vol. 37, № 2. – P. 622-628.

142. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy / R. Cristescu, [et al.] // *Science*. – 2018. – Vol. 362, № 6411. – P. eaar3593.

143. Paradis, E. APE: Analyses of phylogenetics and evolution in R language / E. Paradis, J. Claude, K. Strimmer // *Bioinformatics*. – 2004. – Vol. 20, № 2. – P. 289-290.

144. Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy / B. Lo, [et al.] // *Science*. – 2015. – Vol. 349, № 6246. – P. 436-440.

145. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer / A. Prat, [et al.] // *Breast Cancer Research*. – 2010. – Vol. 12, № 5. – P. R68.

146. PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse / P.V. Hornbeck, [et al.] // *Nucleic Acids Research*. – 2012. – Vol. 40, № D1. – P. D261-D270.

147. PIDO: the primary immunodeficiency disease ontology / N. Adams, [et al.] // *Bioinformatics*. – 2011. – Vol. 27, № 22. – P. 3193-3199.

148. PINA v2.0: Mining interactome modules / M.J. Cowley, [et al.] // *Nucleic Acids Research*. – 2012. – Vol. 40, № D1. – P. D862-D865.

149. PMTDS: a computational method based on genetic interaction networks for Precision Medicine Target-Drug Selection in cancer / V. Vasudevaraja, [et al.] // *Quantitative Biology*. – 2017. – Vol. 5, № 4. – P. 380-394.

150. Prediction of candidate primary immunodeficiency disease genes using a support vector machine learning approach. / S. Keerthikumar, [et al.] // *DNA research*. – 2009. – Vol. 16, № 6. – P. 345-351.

151. Prediction of central nervous system embryonal tumour outcome based on gene expression / S.L. Pomeroy, [et al.] // *Nature*. – 2002. – Vol. 415, № 6870. – P.436.

152. Prevalence and co-occurrence of actionable genomic alterations in high-grade bladder cancer / G. Iyer, [et al.] // *Journal of Clinical Oncology*. – 2013. – Vol. 31, № 25. – P. 3133-3140.

153. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy / P. Sharma, [et al.] // *Cell*. – 2017. – Vol. 168, № 4. – P. 707-723.

154. Primary Immunodeficiency Diseases: an Update on the Classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015 / C. Picard, [et al.] // *Journal of Clinical Immunology*. – 2015. – Vol. 35, № 8. – P. 696-726.

155. Primary vs. secondary antibody deficiency: Clinical features and infection outcomes of immunoglobulin replacement / S.S. Duraisingham, [et al.] //

PLoS ONE. – 2014. – Vol. 9, № 6. – P. e100324.

156. Prognostic factors in survival of patients with stage Ta and T1 bladder urothelial tumors: the role of G1-S modulators (p53, p21Waf1, p27Kip1, cyclin D1, and cyclin D3), proliferation index, and clinicopathologic parameters / A. Lopez-Beltran, [et al.] // American journal of clinical pathology. – 2004. – Vol. 122, № 3. – P. 444-452.

157. Prognostic value of tumor infiltrating lymphocytes in the vertical growth phase of primary cutaneous melanoma / C.G. Clemente, [et al.] // Cancer. – 1996. – Vol. 77, № 7. – P. 1303-1310.

158. Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy / F. Cheng, [et al.] // Oncotarget. – 2014. – Vol. 5, № 11. – P. 3697-3710.

159. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus / Q. Sun, [et al.] // Science. – 2013. – Vol. 340, № 6132. – P. 619-621.

160. Rao, B.S. Hennekam lymphangiectasia syndrome / B.S. Rao, M. Sree Vani, B.S. Kanth // International Journal of Research in Medical Sciences. – 2015. – Vol. 3, № 2. – P. 516-519.

161. Revisiting human primary immunodeficiencies / J.L. Casanova, [et al.] // Journal of internal medicine. – 2008. – Vol. 264, № 2. – P. 115-127.

162. Ricketts, C.J. Tumor-specific hypermethylation of epigenetic biomarkers, including SFRP1, predicts for poorer survival in patients from the TCGA kidney renal clear cell carcinoma (KIRC) project / C.J. Ricketts, V.K. Hill, W.M. Linehan // PLoS ONE. – 2014. – Vol. 9, № 1. – P. e85621.

163. Rius, M. Epigenetic cancer therapy: Rationales, targets and drugs. Vol. 31 / M. Rius, F. Lyko // Oncogene. – 2012. – Vol. 31, № 39. – P. 4257.

164. RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics / M.G. Best, [et al.] // Cancer Cell. – 2015. – Vol. 28, № 5. – P. 666-676.

165. RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation / C. Gontan, [et al.] // *Nature*. – 2012. – Vol. 485, № 7398. – P. 386.
166. Samarghitean, C. Bioinformatics services related to diagnosis of primary immunodeficiencies / C. Samarghitean, M. Vihinen // *Current Opinion in Allergy and Clinical Immunology*. – 2009. – Vol. 9, № 6. – P. 531-536.
167. Sayed, D. Micornas in development and disease / D. Sayed, M. Abdellatif // *Physiological Reviews*. – 2011. – Vol. 91, № 3. – P. 827-887.
168. Second-line single-agent versus doublet chemotherapy as salvage therapy for metastatic urothelial cancer: A systematic review and meta-analysis / D. Raggi, [et al.] // *Annals of Oncology*. – 2016. – Vol. 27, № 1. – P. 49-61.
169. Seidel, M.G. Autoimmune and other cytopenias in primary immunodeficiencies: Pathomechanisms, novel differential diagnoses, and treatment / M.G. Seidel // *Blood*. – 2014. – Vol. 124, № 15. – P. 2337-2344.
170. Şenbabaoğlu, Y. Critical limitations of consensus clustering in class discovery / Y. Şenbabaoğlu, G. Michailidis, J.Z. Li // *Scientific Reports*. – 2014. Vol. 4. – P. 6207.
171. Siegel, R.L. Cancer statistics, 2016. - PubMed - NCBI / R.L. Siegel, K.D. Miller, A. Jemal // *CA: A Cancer Journal for Clinicians*. – 2016. – Vol. 66, № 1. – P. 7-30.
172. Signalink 2 - a signaling pathway resource with multi-layered regulatory networks / D. Fazekas, [et al.] // *BMC systems biology*. – 2013. – Vol. 7. – P. 7.
173. Stark, C. BioGRID: a general repository for interaction datasets / C. Stark // *Nucleic Acids Research*. – 2006. – Vol. 34, № 90001. – P. D535-D539.
174. Stem cell division is regulated by the microRNA pathway / S.D. Hatfield, [et al.] // *Nature*. – 2005. – Vol. 435, № 7044. – P. 974.
175. STRING 8—a global view on proteins and their functional interactions in 630 organisms 2008 / L.J. Jensen, [et al.] // *Nucleic acids research*. – 2008. – Vol. 37, № suppl_1. – P. 412-416.

176. Superficial bladder cancer: an update on etiology, molecular development, classification, and natural history,. / E. Pasin, [et al.] // Reviews in urology. – 2008. – Vol. 10, № 1. – P. 31-43.

177. Syndrome de Hennekam / T. Erkan, [et al.] // Archives de Pediatrie. – 1998. – Vol. 5, № 12. – P. 1344-1346.

178. Takahashi, N. Common variable immunodeficiency / N. Takahashi, T. Morio // Japanese Journal of Clinical Immunology. – 2008. – Vol. 31, № 1. – P. 9-16.

179. The economics of bladder cancer: Costs and considerations of caring for this disease / R.S. Svatek, [et al.] // European urology. – 2014. – Vol. 66, № 2. – P. 253-262.

180. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data / A. McKenna, [et al.] // Genome Research. – 2010. – Vol. 20, № 9. – P. 1297-1303.

181. The global epidemiology of bladder cancer: a joinpoint regression analysis of its incidence and mortality trends and projection / M. Wong, [et al.] // Scientific reports. – 2018. – Vol. 8, № 1. – P. 1129.

182. The human gene connectome as a map of short cuts for morbid allele discovery / Y. Itan, [et al.] // Proceedings of the National Academy of Sciences of the United States of America. – 2013. – Vol. 110, № 14. – P. 5558-5563.

183. The multilayered complexity of ceRNA crosstalk and competition / Y. Tay, [et al.] // Nature. – 2014. – Vol. 505, № 7483. – P. 344.

184. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation / J. Chen, [et al.] // Nature genetics. – 2006. – Vol. 38, № 2. – P. 228.

185. The Sequence Alignment/Map format and SAMtools / H. Li, [et al.] // Bioinformatics. – 2009. – Vol. 25, № 16. – P. 2078-2079.

186. The variant call format and VCFtools / P. Danecek, [et al.] // Bioinformatics. – 2011. – Vol. 27, № 15. – P. 2156-2158.

187. Therneau, T. A Package for Survival Analysis in S. version 2.38 / T. Therneau // R package. – 2015.
188. Tomczak, K. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge / K. Tomczak, P. Czerwińska, M. Wiznerowicz // Contemporary oncology. – 2015. – Vol. 19, № 1A. – P. A68.
189. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: Results from the FinHER trial / S. Loi, [et al.] // Annals of Oncology. – 2014. – Vol. 25, № 8. – P. 1544-1550.
190. Tumor size predicts the survival of patients with pathologic stage t2 bladder carcinoma / L. Cheng, [et al.] // Cancer. – 1999. – Vol. 85, № 12. – P. 2638-2647.
191. Tumour-infiltrating lymphocytes are correlated with higher expression levels of PD-1 and PD-L1 in early breast cancer / A. Kitano, [et al.] // ESMO Open. – 2017. – Vol. 2, № 2. – P. e000150.
192. Uses of next-generation sequencing technologies for the diagnosis of primary immunodeficiencies / M. Seleman, [et al.] // Frontiers in immunology. – 2017. – Vol. 8. – P. 847.
193. Wang, K. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data / K. Wang, M. Li, H. Hakonarson // Nucleic Acids Research. – 2010. – Vol. 38, № 16. – P. e164-e164.
194. Wang, K. A brief procedure for big data analysis of gene expression / K. Wang, W. Wang, M. Li // Animal Models and Experimental Medicine. – 2018. – Vol. 1, № 3. – P. 189-193.
195. Wei, L. M6APred-EL: A Sequence-Based Predictor for Identifying N6-methyladenosine Sites Using Ensemble Learning / L. Wei, H. Chen, R. Su // Molecular Therapy - Nucleic Acids. – 2018. – Vol. 12. – P. 635-644.
196. Wei, S.C. Fundamental mechanisms of immune checkpoint blockade therapy / S.C. Wei, C.R. Duffy, J.P. Allison // Cancer discovery. – 2018. – Vol. 8,

№ 9. – P. 1069-1086.

197. Where genotype is not predictive of phenotype: Towards an understanding of the molecular basis of reduced penetrance in human inherited disease / D.N. Cooper, [et al.] // *Human genetics*. – 2013. – Vol. 132, № 10. – P. 1077-1130.

198. Wilkerson, M.D. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking / M.D. Wilkerson, D.N. Hayes // *Bioinformatics*. – 2010. – Vol. 26, № 12. – P. 1572-1573.

199. Williams E.J.B. Coexpression of neighboring genes in the genome of *Arabidopsis thaliana* / E.J.B. Williams, D.J. Bowles // *Genome Research*. – 2004. – Vol. 14, № 6. – P. 1060-1067.

200. Williams S.G. Molecular markers for diagnosis, staging, and prognosis of bladder cancer / S.G. Williams, M. Buscarini, J.P. Stein // *Oncology*. – 2001. – Vol. 15, № 11. – P. 1461.

201. Yuan Y. CNNC: Convolutional neural networks for co-expression analysis / Y. Yuan, Z. Bar-Joseph // *BioRxiv*. – 2018. – P. 365007.

202. ZEB1, ZEB2, and the miR-200 family form a counterregulatory network to regulate CD8⁺ T cell fates / T. Guan, [et al.] // *Journal of Experimental Medicine*. – 2018. – Vol. 215, № 4. – P. 1153-1168.

203. Zhao Y. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis / Y. Zhao, E. Samal, D. Srivastava // *Nature*. – 2005. – Vol. 436, № 7048. – P. 214.

204. Zhong Z. Screening differential circular RNA expression profiles reveals the regulatory role of circTCF25-miR-103a-3p/miR-107-CDK6 pathway in bladder carcinoma / Z. Zhong, M. Lv, J. Chen // *Scientific Reports*. – 2016. – Vol. 6. – P. 30919.