

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
УЧРЕЖДЕНИЕ НАУКИ
ИНСТИТУТ ОБЩЕЙ ГЕНЕТИКИ им. Н.И. ВАВИЛОВА
РОССИЙСКОЙ АКАДЕМИИ НАУК

На правах рукописи

Кашатникова Дарья Алексеевна
**ГЕНЕТИЧЕСКИЙ ЛАНДШАФТ И ИММУННЫЙ СТАТУС ПРИ
COVID-19 И ПНЕВМОНИИ**

специальность – 1.5.7. – генетика

Диссертация на соискание учёной степени
кандидата биологических наук

Научный руководитель:
доктор биологических наук,
Сальникова Л.Е.

Москва, 2024

СОДЕРЖАНИЕ

Список сокращений.....	4
ВВЕДЕНИЕ.....	5
ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ.....	13
1. Редкие и распространенные заболевания.....	13
2. ПИД.....	16
3. Генетика инфекционных заболеваний.....	23
3.1 COVID-19.....	23
3.2 Пневмония.....	27
4. Генетика иммунитета. Образование вариабельных Т- и В-лимфоцитов.....	30
5. Аннотация генов и вариантов.....	35
ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ.....	46
1. Биологические образцы.....	46
1.1 Демографические и клинические данные пациентов.....	46
1.1.1 Пациенты с COVID-19.....	46
1.1.2 Пациенты с пневмонией и контрольная группа.....	47
2. Молекулярно-генетические методы.....	48
2.1 Выделение ДНК пациентов с COVID-19 и секвенирование экзомов.....	48
2.2 Подготовка списков генов и вариантов.....	49
2.3 Аннотация вариантов.....	49
3. Количественное измерение показателей TREC/KREC.....	51
3.1 Создание генетических конструкций TREC, KREC и человеческого альбумина.....	51
3.2 Выделение ДНК и количественная мультиплексная полимеразная цепная реакция (ПЦР) в режиме реального времени.....	51
4. Базы данных и ресурсы.....	53
4.1 Базы данных «генотип-фенотип» ClinVar и HGMD®.....	53
4.2 Глобальный ресурс генетических вариантов человека dbSNP.....	54

4.3 Транскриптомные данные для нормальных тканей и клеток крови человека.....	54
5. Статистическая обработка данных.....	55
ГЛАВА 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ.....	56
ЗАКЛЮЧЕНИЕ.....	101
ВЫВОДЫ.....	102
БЛАГОДАРНОСТИ.....	103
СПИСОК ЛИТЕРАТУРЫ.....	104
ПРИЛОЖЕНИЕ.....	128

Список сокращений

ДНК – дезоксирибонуклеиновая кислота

ПЦР – полимеразная цепная реакция

ПИД – первичный иммунодефицит

NGS – последнее поколение секвенирования

WGS – полногеномное секвенирование

WES – полноэкзомное секвенирование

ВП – внебольничная пневмония

НП – нозокомиальная пневмония

ОРДС – острый респираторный дистресс-синдром

TREC – сигнальное эксцизионное кольцо рекомбинации рецептора Т-лимфоцита

KREC – сигнальное эксцизионное кольцо рекомбинации рецептора В-лимфоцита

AF – аллельная частота

HI (high impact) – варианты с потенциально патогенным воздействием на белок

ВВЕДЕНИЕ

Актуальность и степень разработанности темы исследования

Изучение генетических аспектов иммунитета является одним из актуальных направлений науки. Большая часть иммунных функций человека находится под контролем генетики, и только лишь малая часть зависит от факторов окружающей среды (Mangino M. et al., 2017). Нахождение причин многих распространенных иммунных заболеваний и вызываемых ими осложнений затрудняет наличие генетической гетерогенности, при которой клиническое проявление заболевания обусловлено различными генетическими дефектами. Особенно наглядно это показала пандемия инфекции SARS-CoV-2 в 2019 году, распространившаяся по всему миру (Khare K. et al., 2022) и ставшая глобальным кризисом здравоохранения нашего времени. Различия в клиническом проявлении среди схожих по факторам риска групп сравнения дали основания полагать, что генетика оказывает влияние на ход развития COVID-19 (The COVID-19 Host Genetics Initiative, 2020). Инфекции нижних дыхательных путей, включая бактериальную пневмонию, которая также является иммуноопосредованным многофакторным заболеванием, по данным Всемирной организации здравоохранения в 2019 занимали четвертое место среди основных причин смертности в мире (<https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates>). С этиологической точки зрения, прогноз и течение пневмонии определяются особенностями возбудителя, факторами окружающей среды и генетическими особенностями человека (Karnaushkina M.A. et al. 2022). Недавние исследования также акцентируют внимание на роли генетики в предрасположенности к инфекционным заболеваниям, тяжести их течения (Lowe A.L. et al., 2021) и вариабельности ответа на иммунизацию (Quach H. et al., 2020).

Большинство болезней, связанных со значительным нарушением работы иммунной системы, возникают в раннем возрасте и требуют

целенаправленного неотложного лечения. Однако даже при изучении врожденных моногенных иммунологических нарушений могут возникать трудности, так как большинство болезней имеют разнообразные клинические проявления (Notarangelo L.D. et al., 2020). Одними из таких сложных генетических заболеваний являются первичные иммунодефициты (ПИД). Это гетерогенная группа редких врожденных заболеваний, включающая в себя более 450 фенотипов, вызванных нарушением одного или сразу нескольких механизмов иммунной защиты. Информация о генетических и фенотипических особенностях ПИД постоянно обновляется (Tangye S.G. et al. 2022). Варианты лечения таких пациентов отличаются, но основой лечения тяжелых форм иммунодефицитов является донорская трансплантация гемопоэтических стволовых клеток (ТГСК) (Castagnoli R. et al., 2019). Несмотря на успешность лечения ряда иммунодефицитов, пожизненного приживания ауто-гемопоэтических стволовых клеток не происходит (Panchal N. et al., 2021), что является довольно сильным ограничением и необходимостью в пожизненной терапии таких пациентов.

Выявление ключевых генов и вариантов, ассоциированных с таким многофакторным заболеванием, как COVID-19, затруднено тем, что иммунная защита организма от инфекций специфична и происходит на нескольких уровнях (Casanova J.L. et al., 2020). Недавно представленная в журнале Cell омнигенная модель, предполагает, что заболевание может быть связано с большим количеством вариантов, находящихся в основных генах, а также периферических генах, которые связаны с основными генами генными сетями (Boyle E. A. et al., 2017; Liu X. et al., 2019). Гены, ассоциированные с заболеванием, как правило, экспрессируются в тех тканях, с заболеваниями которых они связаны. Эти гены также демонстрируют тесную связь в генных сетях (Kolobkov D.S. et al., 2022) и особенно это характерно для генов, ответственных за иммунный ответ (Delavari S. et al., 2021). На основе широкогеномных исследований уже удалось охарактеризовать 23 локуса у

людей с COVID-19, которые были ассоциированы с восприимчивостью к инфекции SARS-CoV-2 и с критическим течением заболевания (COVID-19 Host Genetics Initiative, 2022). Особенно подвержены неблагоприятным исходам от инфекции SARS-CoV-2 люди с имеющимися иммунными нарушениями. Например, пациенты с ПИД могут иметь более тяжелое течение COVID-19, по сравнению со здоровыми (Drzymalla E. et al., 2022), однако это характерно не для всех пациентов с ПИД и зависит от конкретного генетического повреждения или формы и стадии ПИД у пациента. Также следует учитывать сопутствующие заболевания, демографические факторы и возраст (Seyed Alinaghi S. et al., 2022). Было показано, что тяжелая форма течения COVID-19 наблюдалась у большого процента здоровых людей с бессимптомным ПИД, который у них не был подтвержден в течение жизни (Gray P.E. et al., 2022). У этих людей были обнаружены генетические дефекты, нарушающие врожденные иммунные функции, опосредованные интерфероном I типа.

Дополнительные исследования в области геномных данных открывают новые перспективы в разработке инновационных методов в диагностике, профилактике и лечении иммунных нарушений и инфекционных заболеваний.

Цель исследования – выявить генетические особенности иммуноопосредованных болезней с различными сопутствующими состояниями с помощью секвенирования, молекулярно-генетической диагностики и биоинформатического подхода.

Задачи исследования:

1. Провести сбор образцов ДНК пациентов с инфекционными заболеваниями (пневмония и COVID-19).
2. Провести анализ данных секвенирования у пациентов с тяжелой и легкой/умеренной формой течения COVID-19 для оценки эффектов редких потенциально патогенных генетических вариантов (HI (high impact))

вариантов) на уровне всего экзоста и в различных наборах генов, включая гены первичных иммунодефицитов (ПИД).

3. Охарактеризовать патогенный ландшафт генов, ассоциированных с ПИД, с помощью «генотип-фенотип» баз данных ClinVar и HGMD[®] и ресурса dbSNP, для сравнения представленности НИ вариантов между базами данных и экзомными данными пациентов с COVID-19.

4. Провести *in silico* сравнение профилей экспрессии генов ПИД в субпопуляциях клеток крови человека по отношению к другим генам.

5. Рассмотреть гены ПИД и их функциональные гены-партнеры в качестве генов, ассоциированных с развитием тяжелой формы течения COVID-19, используя для анализа результаты полноэкзомного секвенирования молодых пациентов с легкой/умеренной и тяжелой формой COVID-19.

6. Охарактеризовать иммунный статус молодых пациентов с COVID-19 с помощью тестирования косвенных маркеров иммунного ответа, основанного на Т- и В-клеточной рекомбинации (TREC и KREC), и соотнести уровни TREC и KREC с числом редких потенциально патогенных вариантов на уровне всего экзоста и в генах ПИД у пациентов с тяжелой и легкой/умеренной формами течения COVID-19.

7. Оценить диагностический потенциал TREC и KREC в других группах пациентов, включая пациентов с бактериальной пневмонией.

Научная новизна

Оценен совокупный вклад редких потенциально патогенных вариантов (НИ вариантов) в развитие тяжелого течения COVID-19. Впервые гены, ассоциирующиеся с ПИД, были выявлены ключевыми генами при тяжелом течении COVID-19. Определены наибольшие эффекты совокупного вклада редких потенциально патогенных вариантов (НИ вариантов) в основных генах

и в совместно связанных функциональных генах-партнерах, ассоциированных с ПИД. У пациентов с тяжелой формой течения COVID-19 обнаружены сниженные уровни TREC при повышенном числе редких HI (high impact) вариантов (потенциально патогенные варианты). Показана высокая диагностическая эффективность учета уровней TREC/KREC у пациентов с тяжелой формой COVID-19 и бактериальной пневмонией.

Теоретическая значимость

Результаты исследования подтверждают омнигенную гипотезу, согласно которой развитие признаков определяется сетью взаимосвязанных генов, которые вносят вклад в развитие признака в зависимости от их близости к рассматриваемому фенотипу. В соответствии с данной гипотезой, впервые была показана значимость для тяжелого COVID-19 генов ПИД как коровых генов и их ближайших генов-партнеров из сетей межбелковых взаимодействий как околочоровых периферических генов.

Практическая значимость

В данной работе впервые продемонстрирован высокий диагностический потенциал измерения TREC и KREC у пациентов с тяжелыми вирусными и бактериальными инфекциями. Представленный в работе количественный метод определения уровней TREC и KREC может быть использован для прогнозирования групп риска по тяжелому течению и неблагоприятному исходу у пациентов, поступающих в клинику с внебольничной инфекцией, или у пациентов с риском развития внутрибольничной инфекции.

Основные положения, выносимые на защиту:

1. Генетический груз редких потенциально патогенных вариантов (HI вариантов) может быть фактором риска развития более тяжелой формы COVID-19.

2. У молодых пациентов с тяжелой формой COVID-19 наблюдается избыток редких потенциально патогенных вариантов в генах ПИД и их функциональных генах-партнерах, по сравнению с пациентами с легкой/умеренной формой COVID-19.

3. Диагностика иммунного статуса у молодых пациентов с инфекционными заболеваниями (COVID-19 и бактериальная пневмония) с помощью метода TREC и KREC эффективна при оценке вероятности осложненного течения заболевания.

Личный вклад соискателя

Соискателем проанализирована научная литература по исследуемой проблеме, проведены экспериментальные исследования по иммунному статусу у пациентов с COVID-19 и пневмонией, проведена загрузка и анализ данных из генетических онлайн-ресурсов, выполнен анализ и подготовка к публикации экспериментальных данных и данных из биоинформатических ресурсов.

Степень достоверности и апробации результатов

Достоверность результатов работы подтверждается использованием современных экспериментальных и вычислительных методов, используемых при молекулярно-генетических исследованиях и функциональной аннотации геномов. Обсуждение результатов проведено с учетом современных исследований, опубликованных в области исследований генетических аспектов иммунных особенностей человека, а также особенностей интерпретации функций генов и генетических вариантов. Научные положения и выводы, изложенные в диссертации, обоснованы и подтверждены фактическим материалом.

Основные результаты диссертационной работы были представлены в виде устных докладов на конференциях: конференция с международным участием «Жизнеобеспечение при критических состояниях» (10-13 ноября

2021-2023 г., Москва), научно-практическая конференция «Медицинская наука в век цифровой трансформации» (10 декабря 2021 г., Курск), научная конференции с международным участием, посвященная 50-летию лаборатории популяционной генетики им. Ю. П. Алтухова ИОГен (11-14 октября 2022 г., Москва), итоговая научно-практическая конференция ФГБУ ФНКЦ ФХМ им. Ю. М. Лопухина ФМБА (20-21 декабря 2022 г., Москва). Стендовые доклады были представлены на международной конференции «The 20th Biennial Meeting of the European Society for Immunodeficiency (ESID)» (12-15 октября 2022 г., Gothenburg) и на итоговой научно-практической конференции ФГБУ ФНКЦ ФХМ им. Ю. М. Лопухина ФМБА (19-20 декабря 2023 г., Москва).

Публикации по результатам исследования

Результаты исследования представлены в 14 научных публикациях, в том числе в 6 статьях в ведущих научных журналах, индексируемых в базах данных Scopus и Web of Science и рекомендованных ВАК для защиты диссертаций:

1. Salnikova L.E., Kolobkov D.S., Sviridova D.A. (**Kashatnikova D.A.**), Abilev S.K. An overview of germline variations in genes of primary immunodeficiencies through integrative analysis of ClinVar, HGMD® and dbSNP databases // Human Genetics. 2021. №140. С.1379–1393.
2. Khadzhieva M.B., Kalinina E.V., Larin S.S., Sviridova D.A. (**Kashatnikova D.A.**), Gracheva A.S., Chursinova J.V., Stepanov V.A., Redkin I.V., Avdeikina L.S., Rumyantsev A.G., Kuzovlev A.N., Salnikova L.E. TREC/KREC levels in COVID-19 young patients // Diagnostics. 2021. №11(8). С. 1486.
3. Kolobkov D.S.; Sviridova D.A. (**Kashatnikova D.A.**); Abilev S.K.; Kuzovlev A.N.; Salnikova L.E. Genes and Diseases: Insights from Transcriptomics Studies // Genes. 2022. №13. С. 1168.

4. **Kashatnikova D.A.**; Khadzhieva M.B.; Kolobkov D.S.; Belopolskaya O.B.; Smelaya T.V.; Gracheva A.S.; Kalinina E.V.; Larin S.S.; Kuzovlev A.N.; Salnikova L.E. Pneumonia and Related Conditions in Critically Ill Patients—Insights from Basic and Experimental Studies // *Int. J.Mol. Sci.* 2022. №23. С. 9896.

5. Khadzhieva M.B., Gracheva A.S., Belopolskaya O.B., Kolobkov D.S., **Kashatnikova D.A.**, Redkin I.V., Kuzovlev A.N., Grechko A.V. and Salnikova L.E. COVID-19 severity: does the genetic landscape of rare variants matter? // *Front. Genet.* 2023. №14. С. 1152768.

6. Khadzhieva M.B.; Kolobkov D.S.; **Kashatnikova D.A.**; Gracheva A.S.; Redkin I.V.; Kuzovlev A.N.; Salnikova L.E. Rare Variants in Primary Immunodeficiency Genes and Their Functional Partners in Severe COVID-19 // *Biomolecules.* 2023. №13. С. 1380.

Объем и структура диссертации

Диссертационная работа изложена на 131 странице машинописного текста и включает следующие разделы: введение, обзор литературы, материалы и методы, результаты и обсуждение, заключение, выводы, благодарности, список литературы и приложение. Иллюстративный материал диссертации представлен в виде 13 таблиц и 28 рисунков. Список литературы включает 260 литературных источников, из которых 255 источников зарубежной литературы.

ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

1. Распространенные и редкие заболевания

Редкие заболевания характеризуются широким спектром симптомов и признаков, которые могут различаться у людей с одинаковыми патологиями. Известно, что около 350 миллионов человек страдают от редких заболеваний (<https://rarediseases.info.nih.gov/diseases/pages/31/faqs-about-rarediseases>), 75% из них являются педиатрическими заболеваниями, а 30% заболевших детей не доживают до 5 летнего возраста. По оценкам последнего отчета RARE-X (<https://rare-x.org/case-studies/the-power-of-being-counted/>) насчитывается около 10000 генетических и негенетических редких заболеваний. Но, несмотря на рост числа скрининговых методов, позволяющих выявлять наследственные болезни, установить точное количество редких заболеваний довольно трудно, из-за различных способов определения редких заболеваний во всем мире, а также отличающихся стандартов и мер, применяемых в различных базах данных. Большинство редких заболеваний (80%), по отчету RARE-X, теоретически поддаются диагностике, но оставшиеся 20% настолько плохо определены, что могут оказаться неразрешимой задачей в клинической практике. Даже при нынешних ограниченных знаниях о генетической основе многих редких заболеваний известно, что разные патогенные варианты одного и того же гена могут иметь разнообразные последствия, которые часто не регистрируются должным образом (Haendel M. et al., 2020). Диагностика редких заболеваний усложняется схожими симптомами, совпадающими со многими распространенными заболеваниями, а также многие редкие заболевания трудно обнаружить из-за низкой частоты встречаемости или неполной пенетрантности (Васичкина Е. С. и др., 2022). В большинстве случаев такие расстройства имеют полигенную основу, то есть ассоциированы с несколькими генами, и в зависимости от того, насколько

много имеется у человека дефектных аллелей, могут быть возможны различные степени тяжести проявления заболевания. При этом существующая генетическая адаптация может изменять восприимчивость к редким заболеваниям, что может быть частично обусловлено различной историей селективного давления на гены, участвующих в развитии заболевания (Di Rienzo A., 2006; Hancock A.M. et al., 2008).

Генетическая основа распространенных болезней еще более неясна. Большинство распространенных заболеваний наследуется на 30-60%, при этом важную роль также играют факторы окружающей среды и образ жизни человека (Price A.L., et al., 2015). Поэтому такие заболевания в еще большей степени имеют многофакторные причины возникновения, по сравнению с редкими. Очевидно, что полностью предсказать риски возникновения какого-либо распространенного заболевания не получится, так как преобладающее число выявляемых вариантов, ассоциированных с заболеванием, с помощью GWAS, оказывают относительно небольшое влияние на предрасположенность к заболеванию. Были предприняты попытки в выявлении рисков распространенных заболеваний, которые показали влияние редких вариантов. Так, исландская группа ученых выявила значимую и воспроизводимую связь между низкочастотными вариантами с сильным эффектом, которые ассоциировались с болезнью Альцгеймера, остеопорозом и диабетом 2 типа (Jonsson T. et al., 2013; Jonsson T. et al., 2012; Styrkarsdottir U. et al., 2013; Helgason H. et al., 2013; Steinthorsdottir V. et al., 2014).

Влияние редких вариантов является предметом серьезных дискуссий. Они могут вносить серьезный вклад в возникновение и развитие заболевания, а также повышать восприимчивость к распространенным заболеваниям (Pritchard J.K., 2001; Tennessen J.A. Et al., 2012). Особенно сильное воздействие могут оказывать редкие варианты, специфичные для конкретной популяции, по сравнению с распространенными вариантами

(Peischl S. et al., 2013). Исследования популяционной генетики показывают, что большинство вариантов с сильными функциональными эффектами имеют тенденцию быть редкими и частными, и лишь небольшая часть вариантов с сильным влиянием является общими для разных популяций (Quintana-Murci L., 2016). Причинные варианты, ассоциированные с распространенными болезнями, могут иметь частоты как редкие, так и распространенные. При этом генетическая парадигма о том, что распространенные болезни ассоциируются с распространенными вариантами с низкой пенетрантностью, полностью не утратила своей актуальности (Reich D.E. et al., 2001). В последние годы для определения рисков развития распространенных заболеваний используют показатели полигенности, которые рассчитываются по количеству вариантов риска и размеру их эффекта (Visscher P.M. et al., 2021). Такие показатели можно использовать как дополнительный инструмент для подтверждения рисков возникновения распространенных заболеваний. Однако темпы выявления ассоциаций новых генетических вариантов с болезнями значительно отстают от темпов открытий патологических молекулярных механизмов. Один и тот же аллель может увеличивать риск заболевания не у всех людей, у которых данный аллель присутствует, следовательно, определять различные функциональные последствия на молекулярном и физиологическом уровне при наличии этого аллеля довольно трудно (Price A.L. et al., 2015). С помощью полногеномного секвенирования также стало ясно, что изменения в регуляции генов в большой степени ассоциированы с вариантами в межгенных и интронных областях, и при этом данные варианты могут влиять на экспрессию генов в разных тканях и органах (Hirschhorn J.N. et al., 2005). Из этого следует, что для обычного заболевания может существовать огромное количество вариантов, каждый из которых по отдельности имеет очень малое воздействие, что и объясняет совокупный эффект действия этих вариантов. Таким образом, модель развития как редкого, так и распространенного заболевания ассоциируется с наличием у

индивида варианта/вариантов с большим эффектом, и последствия этих вариантов могут быть усилены или ослаблены полигенным фоном, а также и другими факторами (Visscher P.M. et al. 2021).

Для изучения генетической основы болезней необходим анализ частоты возникновения и распространения вредных мутаций в популяции. На сегодняшний день мы имеем довольно низкий уровень накопления и использования генетических данных пациентов в клинической практике. Это негативно сказывается на здоровье и жизни пациентов тем, что откладывает своевременное лечение. Поэтому использование генетических данных для лечения пациентов с широким спектром заболеваний открывает возможности для улучшения диагностики и понимания патогенеза как редких, так и распространенных заболеваний.

2. Первичные иммунодефициты (ПИД)

Первичные иммунодефициты (ПИД) — это гетерогенная группа редких наследственных заболеваний с разнообразными клиническими фенотипами, которые связаны как с различными повторяющимися инфекциями (вирусные, бактериальные, грибковые, паразитарные), так и с тяжелыми комбинированными иммунодефицитами, с возможными летальными исходами (Кузьменко Н. Б. и др., 2017; Casanova J.L. et al., 2007). Группа заболеваний, относящихся к ПИД, одна из самых сложных в определении, так как фенотипическое проявление и генетика ПИД часто совпадают с распространенными инфекционными заболеваниями. С использованием технологий WGS и WES в настоящее время удалось обнаружить более 400 генов (Bousfiha A. et al., 2019), ассоциированных с ПИД. Группа генов ПИД, по текущей версии Международного союза иммунологических сообществ (International Union of Immunologic Societies

– IUIS) (Tangye S.G. et al., 2020), имеет свои особенности. В частности, в этом списке содержатся как гены, связанные с относительно частыми заболеваниями (анемия Фанкони, наследственная предрасположенность к раку или муковисцидоз), так и гены, ассоциированные с редкими болезнями, которые имеют недостаточно изученный патогенный потенциал. При этом количество выявляемых ассоциированных генов и фенотипов ПИД постоянно растет (Bousfiha A. et al., 2018.; Tangye S.G. et al., 2020). Международный союз иммунологических сообществ обновляет свои данные для ПИД каждые два года и подразделяет ПИД на 10 основных групп (Tangye S.G. et al., 2022):

1. Дефекты клеточного и гуморального звеньев иммунитета (синдром Ди Джорджи, хронический кандидоз кожи и слизистых оболочек), включая тяжелую комбинированную иммунную недостаточность – ТКИН и др. В данную группу входят 5-10% ПИД, приводящих к тяжелым инфекциям, возникающих у детей первых месяцев жизни. Причиной являются серьезные генетические факторы, приводящие к снижению количества и функционирования Т- и В-клеток.

2. Комбинированные ПИД, ассоциированные с такими синдромами как Вискотт-Олдрич, Ниймеген, Луи-Бар (атаксия-телеангиэктазия) и др. Данная группа ПИД характеризуется не только пониженной выработкой Т- и В-лимфоцитов, но также задержкой в развитии, склонностью к злокачественным новообразованиям, неврологическим проблемам.

3. Преимущественно дефекты гуморального звена иммунитета (селективный дефицит иммуноглобулина А, Х-сцепленная агаммаглобулинемия (ХЛА), ОВИН и др.). В данную группу входит около 60% всех ПИД. Она разнообразна по фенотипическим и клиническим проявлениям – от полного отсутствия иммунных клеток, таких как В-лимфоциты и иммуноглобулины, до недоразвитых лимфоидных

периферических органов (XLA) и выраженного семейного анамнеза (ОВИН).

4. ПИД с иммунной дисрегуляцией (наследственные гемофагоцитарные синдромы, аутоиммунные лимфопролиферативные синдромы и др.). Очень редкая группа ПИД, с мутациями в нескольких генах. Такие ПИД проявляют себя с первых дней жизни, и их зачастую несвоевременная или неправильная диагностика приводит к летальному исходу. На сегодняшний день лечения для этих синдромов нет, кроме пересадки костного мозга для временной поддержки организма.

5. Количественные и качественные дефекты фагоцитов (врожденные нейтропении, хроническая гранулематозная болезнь и др.). Данная группа ассоциирована в основном с дефектом поглощения вредных патогенов. Дефектные фагоциты неспособны разрушать поглощенные бактерии, грибы и т.д.

6. Дефекты врожденного иммунитета. Данная группа очень обширна и включает в основном предрасположенности к тяжелым вирусным инфекциям, хроническим кожным грибковым заболеваниям, герпесам и т.д.

7. Аутовоспалительные заболевания. Группа включает в себя только заболевания связанные или не связанные с дефектом инфламмосомы - многобелкового комплекса, отвечающего за активацию воспалительного ответа, что является частью врожденного иммунитета. Сюда входят такие заболевания как сахарный диабет 1 типа, воспалительные болезни кишечника, витилиго, рассеянный склероз, артрит и многие хронические воспалительные заболевания.

8. Дефекты системы комплемента (рецидивирующие инфекции, системная красная волчанка, гломерулонефрит). Более редкая группа ПИД (менее 2%

всех ПИД) с наследственными или приобретенными нарушениями отдельных нарушений компонентов комплемента и ингибиторов.

9. Недостаточность костного мозга (апластическая анемия, при которой поврежденные стволовые клетки приводят к снижению числа эритроцитов, лейкоцитов или тромбоцитов). В данную группу входят как наследственные заболевания (анемия Фанкони, синдром Швахмана-Даймонда, врожденный дискератоз), так и идиопатические заболевания.

10. Фенокопии ПИД, вызванные соматическими мутациями, которые возникают под влиянием неблагоприятных факторов окружающей среды.

Изначально ПИД относили к менделевским заболеваниям, но из-за немоногенного наследования, неполной пенетрантности, генетических и эпигенетических модификаций, а также аутосомно-доминантного и аутосомно-рецессивного типов наследования, влияния окружающей среды, ПИД не в полной мере соответствуют менделевскому типу наследования. У подавляющего большинства пациентов с ПИД нарушена функциональность Т- и В-клеток. Многие ПИД возникают в результате антиген-специфической реактивности адаптивного иммунитета, что затрагивает Т- или В-клетки. Тяжелое течение инфекций, которые возникают из-за слабопатогенных микроорганизмов, также иллюстрирует ключевую роль адаптивного иммунитета. Многие ПИД нарушают также и врожденные свойства иммунной системы (Picard C. et al., 2011; Zhang S.Y. et al., 2019), например, изменяя развитие или функционирование субпопуляций антигенпрезентирующих миелоидных клеток. Это согласуется с разнообразием и повторяемостью различных инфекций, поскольку нарушается как способность распознавать широкий спектр микробных антигенов, так и иммунологическая память. Такие нарушения впоследствии могут приводить к комбинированным иммунодефицитам. Пациентам с тяжелым комбинированным иммунодефицитом необходима

своевременная трансплантация стволовых клеток для исключения летального исхода. Некоторые ПИД ассоциированы с несколькими вариантами в разных генах (например, гемафагоцитарные или лимфопролиферативные синдромы), которые, в свою очередь, имеют разные проявления (Redmond M.T. et al., 2022). Также мутации в одном гене могут повлиять на один или несколько иммунологических компонентов человека (Notarangelo L. D., 2010). С другой стороны, при выявленном множестве генетических дефектов у многих пациентов с первичным дефицитом антител, для подавляющего большинства точный диагноз не был поставлен (van Zelm M.C. et al., 2011). Более того, некоторые генетические варианты известных генов, вызывающих заболевания, могут обуславливать широкий клинический спектр фенотипов (Schuetz C. et al., 2008; Conley M. E. et al., 2009). Например, известно, что дефекты 5'- и 3'-сайта сплайсинга гена *Artemis* приводят к тяжелому комбинированному иммунодефициту (ТКИД) или синдрому Оменна, однако, были описаны неродственные пациенты с мутацией в 3'-сайте сплайсинга гена *Artemis* с относительно легким клиническим фенотипом ТКИД (тяжелый комбинированный иммунодефицит) (JSpeert H. et al., 2011).

Учитывая всю сложность при определении ПИД, необходимы более точные и массовые методы для быстрого определения диагноза и своевременного лечения, так как неправильный диагноз или задержка в лечении может привести к сильному ухудшению состояния пациента. Молекулярные механизмы, связывающие иммунные функции с развитием ПИД, остаются плохо изученными (Щербина А. Ю., 2016). Несмотря на повышенную диагностическую эффективность NGS по сравнению с традиционным генетическим тестированием для пациентов с ПИД, около 60% пациентов остаются без диагноза (Stray-Pedersen A. et al., 2017). Этот факт указывает на то, что имеются некоторые дефициты в технологиях, анализе данных или нашем понимании ПИД. Основным ограничением

геномной диагностики заболеваний является отсутствие функциональных доказательств. Однако разносторонние функциональные анализы необходимы для демонстрации биологического эффекта варианта. Используя базу данных STRING, предоставляющую межбелковые взаимодействия среди генов, гены ПИД были рассмотрены в контексте их наибольшей представленности среди биологических путей и молекулярных функций (рис. 1, табл.1).

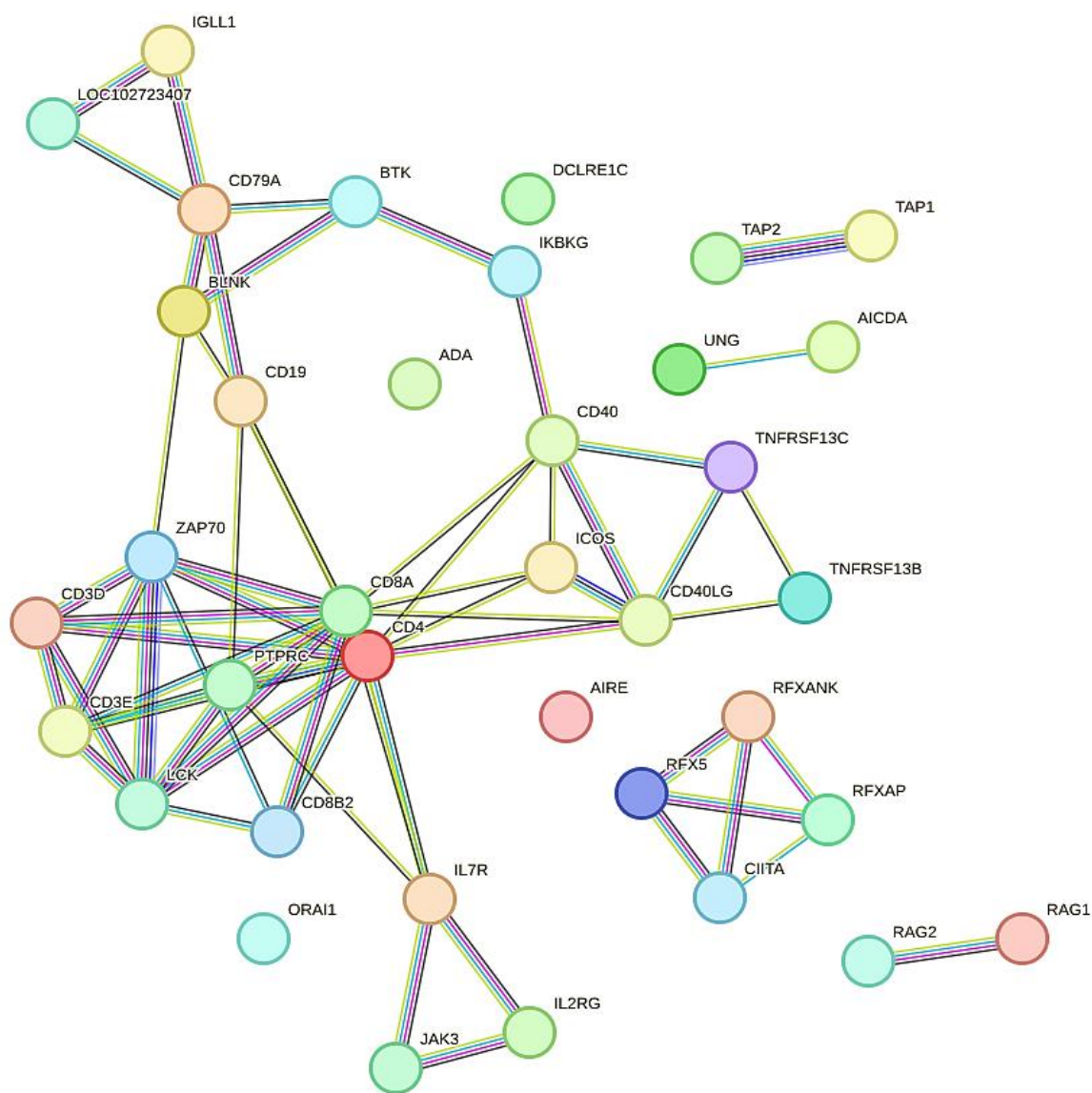


Рис. 1. Функциональная связь между белками генов ПИД

Таблица 1. Топ 10 ассоциаций для биологических путей и молекулярных функций ПИД генов

№	Биологические процессы	Молекулярные функции	Пути KEGG
1	Сигнальный путь, опосредованный интерлейкином-7	Активность пептидного антигенного транспортера ABC-типа	Первичный иммунодефицит
2	Аллельное исключение в ранних про-В-клетках	Связывание с TAP1	Иммунная связь с кишечником для продукции IgA
3	Индукция толерантности Т-клеток	Связывание белков с МНС Ib класса	Обработка и представление антигенов Т-лимфоцитам
4	Сигнальный путь, опосредованный интерлейкином-4	Связывание белков с МНС I класса	Сигнальный путь Т-клеточного рецептора
5	Транспорт из цитозоля в эндоплазматический ретикулум	Связывание белков с МНС	Астма
6	Дифференцировка DN2 Т-клеток	Связывание фосфолипазы	Дифференцировка Th1 и Th2 клеток
7	Сигнальный путь, опосредованный интерлейкином-2	Немембранная тирозинкиназная активность белков	Сигнальный путь NF-κB
8	Ответ на интерлейкин-15	Связывание с антигеном	Линии дифференцировки гемопоэтических клеток
9	Позитивная селекция про-Т-клеток	Активность корцептора	Реакция отторжения трансплантата
10	Негативная селекция про-Т-клеток	Связывание с сигнальными рецепторами	Дифференцировка клеток Th17

Как видно, гены ПИД участвуют в разнообразных биологических и молекулярных процессах, поэтому определение ранее не зарегистрированных мутаций даже в хорошо описанных генах требует междисциплинарных знаний в области молекулярной и клеточной биологии, биохимии и иммунологии. Таким образом, NGS является лишь одним из нескольких этапов диагностики. Определение последствий мутации на фенотипическом уровне требует также биоинформатического анализа, включающего разнообразные типы вариантов (синонимические, несинонимические, потенциально патогенные, неопределенного значения и др.), транскриптомного анализа, а также изучение мутации на клеточных и животных моделях (Seleman M. et al., 2017).

3. Генетика инфекционных заболеваний

3.1 COVID-19

В 2019 году SARS-CoV-2 стала одной из самых опасных вирусных инфекций, которая имела разнообразные последствия – от абсолютно бессимптомных до заболевания COVID-19 с развитием жизнеугрожающих состояний (Buitrago-Garcia D. et al, 2020; Khare K. et al., 2022). Несмотря на то, что в эпоху современной медицины существует множество способов, с помощью которых возможно добиваться улучшения состояния пациентов в процессе лечения, COVID-19 по-прежнему остается глобальной проблемой: зарегистрировано более 667 миллионов случаев заболевания COVID-19 и более 6,8 миллионов случаев смерти (<https://coronavirus.jhu.edu>, по данным на 3 ноября 2023 г.).

Известно, что коронавирусы, чье название произошло от внешнего вида оболочки вируса, поражают не только животных, но и человека (Cheever F. S. et al., 1949). Патогенез и молекулярные механизмы коронавирусов изучаются с 1960-х годов (Tyrrel D. A. J. et al., 1968). Стоит отметить, что в 2003 году в Китае уже происходила вспышка

респираторной инфекции с развивающейся на этом фоне атипичной пневмонией (Parry J., 2003), которая распространилась на многие страны, как и в период пандемии COVID-19 во всем мире в 2019 году. Несмотря на то, что основным симптомом является самокупирующаяся инфекция верхних дыхательных путей, у значимой части пациентов развивались ОРДС (острый респираторный дистресс-синдром), полиорганная недостаточность и, в конечном итоге, смерть (Deng Y. et al., 2020; Zhou F. et al., 2020). Сейчас уже стало известно, что основной причиной тяжелого течения COVID-19 является чрезмерный иммунный ответ хозяина и наличие у пациента некоторых сопутствующих заболеваний (Wu F. et al., 2020; Huang C. et al., 2020). Немаловажную роль в развитии и течении COVID-19 играет генетическая предрасположенность, которой изначально придавалось меньше значения, чем клиническим и эпидемиологическим характеристикам. Несмотря на то, что сопутствующие хронические заболевания пациента коррелируют с тяжестью заболевания, они не дают объяснений всей изменчивости самой тяжести заболевания, которая наблюдается у разных людей.

Молекулярные механизмы, лежащие в основе критических заболеваний у пациентов с COVID-19, становится возможным изучать с помощью полногеномных исследований. Всемирная инициатива по генетике хозяина COVID-19 при помощи GWAS мета-анализа обнаружила 23 локуса с частотами аллелей в диапазоне от 0,003 до 0,66, из которых 7 локусов влияли на восприимчивость к инфекции SARS-CoV-2, и 16 локусов, включая локусы, участвующие в воспалении или контроле врожденного иммунитета (например, гены *OAS1/OAS2/OAS3*, *IFNAR2*, *DPP9*, *TYK2*, *SFTPD* и *MUC5B*) были ассоциированы с критическим течением заболевания (The COVID-19 Host Genetics Initiative, 2022). Другая группа ученых у пациентов с COVID-19 выявила набор генов, состоящий из *SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6* и *XCRI*, на участке хромосомы

3p21.31, ассоциирующийся с дыхательной недостаточностью (The Severe Covid-19 GWAS Group-2020; Shelton J. F. et al., 2020). Несмотря на то, что, клинические показатели у пациентов с COVID-19 сильно различаются, генетические данные пациента являются важным фактором, определяющим то, как он справится с вирусом (Butler-Laporte G. et al., 2022). Например, потеря функции гена *TLR7* у людей молодого возраста, который влияет на основные иммунные процессы при инфицировании клеток и индуцирует провоспалительные цитокины и интерфероны, может привести к тяжелому заболеванию (Plenge R. M. 2020), что подтверждается на пациентах с тяжелой формой течения COVID-19 (Kosmicki J. A. et al., 2021). Также некоторые редкие варианты с потерей функции в генах, участвующих в передаче сигналов интерферона I типа, могут быть причиной тяжелого течения COVID-19 (Zhang Q. et al., 2020). Система комплемента, которая является важной частью как врожденного, так и адаптивного иммунитета, играет важную роль в устранении патогенов. Нарушения в активации комплемента могут быть связаны с патогенезом острого повреждения легких, индуцируемым вирусом. Например, фактор комплемента C5 α и его рецептор CD88 играют ключевую роль в инициации и поддержании воспалительных реакций, активируя нейтрофилы и моноциты (Wang R. et al., 2015). Иногда высокий уровень C5 α обостряет воспалительную реакцию и может наносить вред пациенту, способствуя развитию «цитокинового шторма». У пациентов с COVID-19 уровень C5 α пропорционально увеличивался с тяжестью течения инфекции SARS-CoV-2, а в крови и в легочных миелоидных клетках наблюдался высокий уровень экспрессированного CD88 (Carvelli J. et al., 2020). Это подтверждает роль C5 α -CD88 в возникновении ОРДС. Кроме того, активация C5 α считается одним из маркеров сепсиса (Angus D. C. et al., 2013).

С помощью полногеномных ассоциативных исследований было обнаружено множество генетических локусов, влияющих на

восприимчивость к инфекционным заболеваниям и тяжесть их течения (Kwok A.J. et al., 2021), что также передается по наследству (Sørensen T. I. et al., 1988). Такие исследования строятся на основе полигенной модели, но это также предполагает не совсем точное выявление сложных взаимоотношений между патогеном и хозяином. Иногда «недостающая наследственность», которая была зафиксирована с помощью GWAS, объясняется тем, что генетическая предрасположенность зависит от количества ассоциированных генетических вариантов, при этом каждый из них по отдельности вносит незначительный вклад в развитие заболевания (Claussnitzer M. et al., 2020). В последнее время полигенную модель определения многих причинных генов и вариантов, ответственных за болезнь, объединяют с омнигенной моделью. Омнигенная гипотеза предполагает, что на развитие заболевания влияет совокупный эффект генетических вариантов, расположенных в основных (коровых) и периферийных генах, объединенных межбелковыми связями. При этом по отдельности генетические варианты имеют очень низкий эффект воздействия на конкретный признак (Boyle E. A. et al., 2017; Liu X. et al., 2019). То есть, если один генетический вариант внесет изменения в работу одного из генов — это может повлиять на работу всей сети. Это фактически означает, что изменение в экспрессии любого периферического гена может сказаться на работе основных генов, ассоциированных с заболеванием.

Изучение роли генома в сочетании с клиническими характеристиками COVID-19 у пациентов могло бы в более короткие сроки выявлять лиц с высоким или низким риском восприимчивости к заболеванию и его тяжелого течения, назначать своевременное лечение, а также вносить вклад в изучение заболевания COVID-19 (The COVID-19 Host Genetics Initiative, 2020).

3.2 Пневмония

Пневмония – это острая респираторная инфекция, которая поражает нижние отделы дыхательных путей человека, и является четвертой по частоте встречаемости причиной смерти в мире, уступая только ишемической болезни сердца, инсультам и хронической обструктивной болезни легких (ХОБЛ) (<https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>).

Заболеваемость пневмонией крайне высока, а смертность от нее затрагивает все возрастные группы по всему миру, в особенности детей первых лет жизни и взрослых старше 70 лет (рис. 2). Около 450 миллионов человек в год заболевают пневмонией во всем мире (GBD 2016 Lower Respiratory Infections Collaborators, 2018). Это ложится значительным бременем на ресурсы здравоохранения и общество.

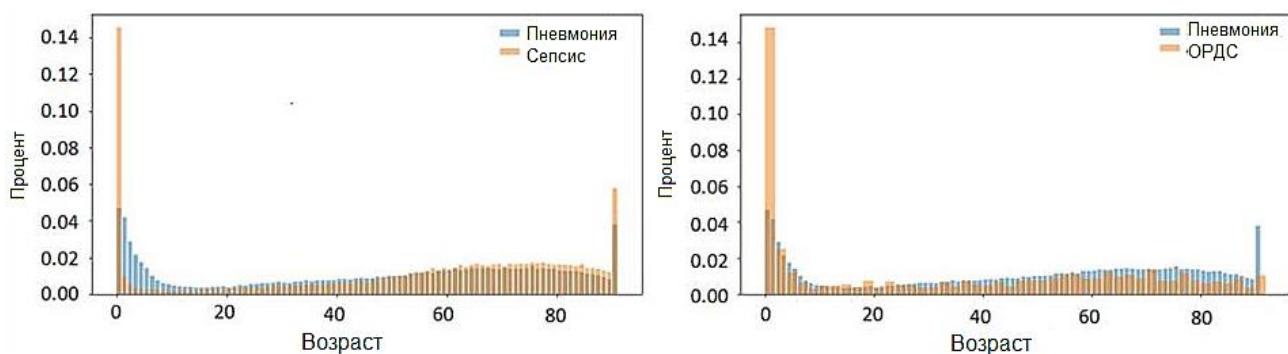


Рис. 2. Распределение по возрасту заболеваемости пневмонией

Пневмонию делят на внебольничную (ВП), которая приобретается вне стационара и нозокомальную (НП), которая развивается в стационарах (Mandell L. A. et al., 2019). Развитие пневмонии в значительной степени зависит от иммунного ответа хозяина, при этом предполагается, что возбудители играют менее заметную роль (Torres A. et al., 2021). Одним из частых и серьезных легочных осложнений при тяжелом течении пневмонии является острый респираторный дистресс-синдром (ОРДС), который характеризуется острым некардиогенным отеком легких, гипоксемией и необходимостью в искусственной вентиляции легких (Ferguson N.D. et al., 2012). Смертность от ОРДС варьирует в достаточно широких пределах от 11

до 87% (Máca J. et al., 2017). Также тяжелое течение пневмонии может сочетаться с сепсисом – опасной для жизни органной дисфункцией, которая вызывается неконтролируемым ответом хозяина на инфекцию (Evans L. et al., 2021). Смертность от сепсиса также высока: от 15 до 56% (Bauer M. et al., 2020). Хронические сопутствующие заболевания также могут оказывать влияние на тяжесть течения пневмонии (Torres A. et al., 2021). Известно, что пациенты с ослабленным иммунитетом имеют более высокий риск пневмонии, чем население в целом (Di Pasquale M.F. et al., 2019). Следует отметить, что как среди пациентов с ВП, так и с НП неоднократно обнаруживались пациенты с нарушенными иммунными функциями (Furman C.D. et al., 2004; Cillóniz C. et al., 2013). Тем не менее реальная распространенность иммунодефицита среди больных пневмонией до сих пор неизвестна. Клинические рекомендации при обнаружении у пациента ВП и НП не касались этой темы, более того, такие пациенты специально исключались из выборки, если у них подтверждался диагноз иммунодефицита (Di Pasquale M.F. et al., 2019). Поэтому оценить истинное количество пациентов с пневмонией с дополнительными иммунными дефектами не предоставляется возможным.

Врожденные и адаптивные иммунные реакции, защищающие организм от патогенов, тесно связаны между собой и зависят от иммунных возможностей хозяина. Лимфоциты, включая Т- и В-клетки, являются основными клетками адаптивной иммунной системы. В частности, Т-лимфоциты не обладают цитотоксической активностью и не способны самостоятельно уничтожить патогены, но способны управлять иммунным ответом и передавать сигналы другим иммунным клеткам для этих задач (Nakajima T. et al., 2010). Сепсис вызывает сильное истощение лимфоцитов, что приводит к неспособности хозяина противостоять инфекциям и располагает к вторичным оппортунистическим инфекциям (Hotchkiss R.S. et al., 2001). Исследования на животных показывают, что иммунные дефекты при сепсисе имеют решающее значение при патогенезе и смертности (Oberholzer C. et al., 2001; Hotchkiss R.

S. et al., 2000). В развитии ОРДС, который также может возникать при инфекции, Т-клетки играют важную роль в защите хозяина. Исследования показывают, что субпопуляции регуляторных Т-клеток (Treg) способствуют восстановлению легких после острого повреждения, вызванного ОРДС (D'Alessio F.R. et al., 2009).

Известно, что гены, контролирующие иммунные процессы, экспрессируются практически во всех тканях (Kolobkov D.S. et al., 2022). Легочная ткань демонстрирует взаимосвязанную сеть резидентных иммунных клеток. Например, альвеолярные макрофаги (АМ), располагающиеся на нижней поверхности дыхательных путей, играют ключевую роль в развитии воспаления легких, продуцируя провоспалительные цитокины (интерлейкины) при взаимодействии с вирусами и бактериями (Preston J. A. et al., 2019). Было показано, что защитная роль АМ при пневмонии часто нарушена (Quinton L. J. et al., 2007). Недавно открытый тип лимфоцитов - врожденные лимфоидные клетки (ILCs) - участвуют в поддержании гомеостаза и иммунной защиты на поверхности слизистой оболочки, за счет быстрой продукции цитокинов, а также являются важными посредниками между микробиотой и адаптивной иммунной системой человека (Panda S. K. et al., 2019). Нарушение данного механизма защиты приводит к тому, что не происходит взаимодействия между НК-клетками и макрофагами для уничтожения патогенов. Опасность пневмонии иногда также заключается в ее бессимптомном течении, при котором отсутствует своевременное лечение или оно недостаточно, ввиду схожих симптомов с гриппом или ОРВИ. Поэтому изучение пациентов с пневмонией, с учетом генетически обусловленных иммунных процессов, способствовало бы обнаружению новых патологических механизмов и оказанию своевременной помощи таким пациентам.

перестройки, позволяющие им соединять одни сегменты с другими (Schuetz C. et al., 2008).

Предшественники Т-клеток изначально начинают перестройку своего клеточного рецептора D и TCRG в тимусе. После того, как сформируется функциональный рецептор, Т-клетка превращается в TCR $\gamma\delta^+$ и выходит из тимуса. Преобладающее большинство клеток сразу начинают перестройку генов TCRB и TCRA без образования функционального рецептора. Далее такие клетки выходят в виде $\alpha\beta$ TcR. Локусы TCRD и TCRB перестройку генов D в J индуцируют посредством перестройки V в DJ. В локусах TCRA и TCRG отсутствуют сегменты D гена, поэтому сегмент V напрямую связывается с сегментом J.

В-клеточный рецептор имеет отличное от Т-рецептора строение в виде тяжелой и легкой цепи. В предшественниках В-клеток перестройки генов начинаются с D в J в локусе тяжелой цепи (IGH), а затем переходят к перестройкам генов V в DJ. Далее процесс переходит к перестройкам в локусе IGK легкой цепи: V гены напрямую связывается с J генами. Если перестройки в легкой цепи IGK не увенчались успехом, то в другом локусе легкой цепи – IGL, V гены начинают перестраивать в J. В таком виде В-клетки уже выходят из костного мозга и являются зрелыми (Schuetz C. et al., 2008).

При генных перестройках в рецепторах Т- и В-клеток белки RAG1 и RAG2 при разрывах двухцепочечной ДНК активируют рекомбинацию таким образом, что на границах образуются рекомбинационные сигнальные последовательности. Эти концы лигируются, образуется кодирующий участок. Промежуточная ДНК вырезается, ее лигирование происходит посредством слияния тупых сигнальных концов. Такие вырезаемые (эксцизионные) рекомбинационные кольца при формировании Т- и В-клеток не имеют специфических функций. Эти кольца стабильны, но не способны к саморепликации и не участвуют в дупликации при митозе. Кодирующая же часть, соответствующая перестройка геномной ДНК в рецепторах,

наследуется всеми дочерними клетками (Livak F. et al., 1996; Breit T. M. et al., 1997; van Zelm M. C. et al., 2007). После завершения перестроек Т- и В-клетки начинают пролиферацию, при которой только одна из дочерних клеток несет в себе сигнальное эксцизионное кольцо. Рекомбинация V(D)J локусов на тяжелой цепи (IGH) и TCR образована таким образом, что уникальные комбинации V, D, и J создают большой репертуар антигенных рецепторов. Таким образом, сформировавшиеся Т- и В-клетки не могут быть мишенями для изучения пролиферации. Но у большинства клеток происходит только одна генная перестройка для образования одного конкретного нефункционального локуса. Так, например, перестройка δ REC- ψ J α удаляет весь локус TCRD в Т-клетках, а перестройка интрона рекомбинационной сигнальной последовательности – Kde удаляет конститутивную область IGK в В-клетках (рис. 4).

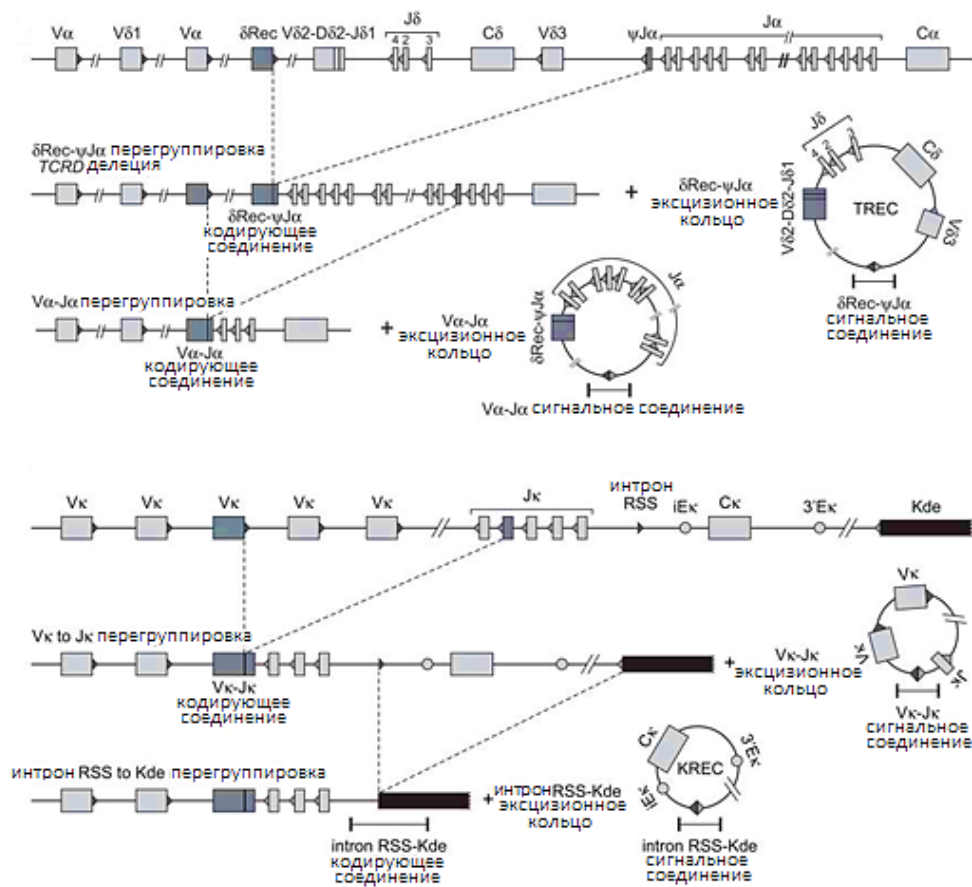


Рис. 4. Рекомбинация V(D)J и образование сигнальных эксцизионных колец (Schuetz C. et al., 2008)

Перегруппировка δ REC- ψ J α происходит в зрелых $\alpha\beta$ +T-клетках на аллели TCRD после неудавшейся генерации $\gamma\delta$ TcR (Verschuren M. C. et al., 1997). Практически все новообразованные $\alpha\beta$ +T-клетки имеют сигнальные эксцизионные кольца T-клеточных рецепторов (TREC) с сигнальным соединением δ REC- ψ J α (рис.4). Так как кодирующая часть этой перестройки стабильно в геноме не присутствует, а локус TCRD расположен внутри локуса TCRA, то последующее перестроение V α -J α удаляет кодирующую часть δ REC- ψ J α и помещает ее в следующую вырезаемую часть (Breit T. M. et al., 1997, Hazenberg M. D. et al., 2000).

В локусе IGK B-клеток перегруппировка интрона рекомбинационной сигнальной последовательности (RSS) Kde образует стабильную кодирующую часть, которая не удаляется последующими перестройками генов. Более того, эта перестройка происходит еще до миграции B-клеток примерно у 35-40% аллелей IGK. Она является самой последней, поэтому не сопровождается пролиферацией в костном мозге (van Zelm M. C. et al., 2005). Таким образом, преобладающее большинство B-клеток будут иметь каппа-делеционное рекомбинационное эксцизионное кольцо (KREC) с сигнальным интроном рекомбинационной последовательности Kde.

Эксцизионные кольца TREC и KREC образуются посредством слияния тупых сигнальных концов после расщепления RAG. Место сцепления двух сигнальных последовательностей является идеальной мишенью для праймеров в ПЦР в реальном времени на основе TaqMan, так как позволяет количественно определять эксцизионные кольца TREC и KREC. Чтобы точно оценивать количество, необходимы дополнительные условия. Во-первых, требуется стандартная кривая для количественной оценки. Обычно это делается при последовательных разведениях конструкции сигнального эксцизионного кольца, клонированного в бактериальную плазмиду. Во-вторых, точное количество уровней TREC и KREC может быть получено, когда количественное определение выполняется относительно контрольного

гена, такого как альбумин или постоянная область TCRA (Zubakov D. et al., 2010). Это требует параллельной ПЦР на том же образце и количественной оценки с использованием серии разбавлений бактериальной плазмиды, содержащей ген-мишень.

Таким образом, количественные измерения TREC и KREC используют в клинической практике для оценки степени дефектов Т- и/или В-клеток. Изначально данный метод использовался для скрининга новорожденных на Т- или В-клеточную лимфопению, такую как тяжелый комбинированный иммунодефицит и X-сцепленную агаммаглобулинемию (XLA) (Dasouki M. et al., 2020), так как количественное определение δ TREC- ψ Ja TREC было информативным для оценки неогенеза Т-клеток тимуса (Douek D. C. et al., 1998; Hazenberg M.D. et al., 2000). Также измерение количественных оценок TREC/KREC применяют и для взрослых людей. Показано, что с возрастом происходит линейное снижение уровней TREC/KREC в цельной крови человека (Kwok J.S.Y. et al., 2020), что может быть использовано не только в медицинской практике, но и для судебной экспертизы (Ou X.L. et al., 2012). Особенно интересно отметить, что количественный анализ TREC является необходимым после трансплантации гемопоэтических стволовых клеток (ТГСК), что помогает оценить успешность донорской пересадки костного мозга и нарастающую положительную динамику, связанную с восстановлением Т-клеток, для пациентов любого возраста (Söderström A. et al., 2022; Gaballa A. et al., 2016; Valotti M. et al., 2014). Однако уровни TREC изучаются и при других заболеваниях: ВИЧ (Cobos Jiménez V. et al., 2016; Touloumi G. et al., 2004; Franco J. M. et al., 2002), красная волчанка (Daitaro K., 2007), atopический дерматит и псориаз (Just H.L. et al., 2008), рассеянный склероз (Chiarini M. et al., 2015; Paghera S. et al., 2020; Hug A. et al., 2003), аксиальный спондилоартрит (Fessler J. et al., 2016), сахарный диабет 2 типа (Dworacki G. et al., 2015; Iskakova S. et al., 2015), болезнь Грейвса (Strawa K. et

al., 2014), миастения Гравис (Sempowski G. D. et al., 2001) и хроническая сердечная недостаточность (Tang T.T. et al., 2011).

5. Аннотация генов и генетических вариантов

Секвенирование всего генома и отдельно экзона служит ценным инструментом для диагностики генетических заболеваний человека. На сложность функционального аннотирования генов влияет множество факторов - от наличия некодирующих областей в геноме до экспрессии генов ограниченным образом, например, в определенных типах клеток или на определенных стадиях развития (The ENCODE, 2004). Одного генетического подхода уже недостаточно ввиду того, что для аннотирования необходимо три составляющих: генетика, статистика и биология (Alyousfi D. et al., 2019). В последние годы сочетание экспериментальных методов с биоинформатическими упрощает способы поиска генов, ассоциированных с болезнями.

Функциональное аннотирование генов человека помогает в изучении различных наборов генов, имеющих ряд особенностей. Существуют гены, потеря функции которых ставит под угрозу жизнеспособность организма: приводит к глубокой потере приспособленности или к летальному исходу. Это эссенциальные и консервативные гены (Chen H. et al., 2020). Они считаются повсеместно экспрессируемыми и необходимыми для нормальной жизнедеятельности клетки, а также для выживания и размножения (Tu Z. et al., 2006). Эти гены имеют более короткую длину транскрипта, в отличие от генов, ответственных за некоторые заболевания (Eisenberg E. et al., 2006). Гены, участвующие в одном и том же заболевании, как было обнаружено, образуют функционально связанные подсети (Gandhi T.K. et al., 2006). Эта особенность используется при определении основных механизмов, влияющих на заболевание и его классификацию (Feldman I. et

al., 2008). Картирование неправильно собранных областей белков от генов, ассоциированных с болезнями, показало их связь с раком (Iakoucheva L.M. et al., 2002), диабетом (Uversky V.N., 2008), сердечно-сосудистыми (Cheng Y. et al., 2006) и нейродегенеративными (Uversky V.N., 2009) заболеваниями, а также другими (Uversky V.N. et al., 2008; Uversky V.N., 2008). Отдельно разработанные биоинформатические программы рассматривают ряд биологических сетей, связанных с фосфорилированием, передачей сигналов, межбелковым взаимодействием, а также регуляторными генетическими сетями (Khurana E. et al., 2013), что помогает в изучении молекулярных основ многих биологических процессов. В целом, для идентификации свойств генов используют различные вычислительные инструменты (Таблица 2).

Таблица 2. Примеры наиболее используемых ресурсов для функциональной аннотации генов

Название	Краткое описание
Protein interaction (PPINT)	Используя различия в сетевых взаимодействиях белков, показывают тесную связь генов болезней с другими генами той же болезни
Gene function (gene ontology)	Исходя из генной онтологии (GO), функциональные свойства гена будут схожи с другими генами болезни. Сравнение идет по трем направлениям: среди биологических процессов, клеточных компонентов и молекулярной функции
Pathway (PATH)	Сравнение общих путей среди генов болезней, с использованием KEGG, Reactome, BioCarta, BioCyc, GenMAPP, MSigDB
Gene expression (GEXP)	Использование экспрессии генов болезней в сочетании с GO, PPINT и другими функциями для определения коэкспрессирующихся генов

Основным ограничением данных методов, которые также могут применяться в комбинациях друг с другом, является качество используемых экспериментальных данных и, вследствие этого, полученная на их основе информация о картировании заболеваний, используемая в дальнейшем для обучения биоинформатических методов аннотации генов. Создание достоверных онтологий для сложных процессов и болезней, а также совершенствование методов по аннотации генов и их одновременное использование из разных источников – это очень трудные задачи, которые могут серьезно ограничивать прогресс поиска приоритетных генов. Списки генов, ассоциированных с болезнями, на сегодняшний день возможно изучать с помощью баз данных. Одной из наиболее полных и известных баз является OMIM (Amberger J.S. et al., 2015), содержащая ассоциации между менделевскими болезнями и генами, а также базы по ассоциациям между вариантами/генами и заболеваниями: ClinVar (Landrum M.J. et al., 2016), HGMD (Stenson P.D. et al., 2012), DisGeNet (Piñero J. et al., 2016), MalaCards (Rappaport N. et al., 2016) и др. С помощью данных ресурсов можно выявлять заболевания, развитие которых связано с несколькими генами.

Функциональное аннотирование отдельных генетических вариантов, имеющих отношение к заболеванию, также необходимо для диагностики, прогнозирования и лечения заболеваний. Однако способность обнаруживать генетические варианты намного опережает способность интерпретировать их функциональное влияние. При этом подавляющее большинство выявленных вариантов – не имеют патогенного воздействия и, следовательно, не связаны с заболеваниями (MacArthur D. G. et al., 2012). Установление точного генетического диагноза во многих случаях до сих пор остается трудной задачей, а существующие инструменты для определения клинической значимости вариантов также сильно отличаются друг от друга (Jamuar S. S. et al., 2015; MacArthur D. G. et al., 2012; Castellán S. et al., 2013; Dong C. et al., 2015; Frousios K. et al., 2013; Grimm D. G. et al., 2015; Martelotto L. G. et al.,

2014). Для прогнозирования функций генетических вариантов существует множество методов. В таблице 3 приведены самые последние и используемые программы.

Таблица 3. Примеры инструментов и методов, используемых для функциональной аннотации генетических вариантов

Название	Краткое описание	Ссылка
Snpsift	Аннотирует варианты, используя разницу между двумя и более геномами. Обучающие данные – геномы двух независимо полученных мутантных штаммов <i>Drosophila melanogaster</i> .	Cingolani P. et al., 2012a
Snpeff	Аннотирует варианты на основе их геномного местоположения и прогнозирует эффекты кодирования. Может предсказывать эффекты кодирования, такие как синонимичные или несинонимичные замены аминокислот, добавление или потерю стартовых кодонов, стоп-кодонов или сдвига рамки считывания. Обучающие данные – геномы двух независимо полученных мутантных штаммов <i>Drosophila melanogaster</i> .	Cingolani P. et al., 2012b
1000G	Всестороннее описание генетических вариантов человека с помощью полногеномного секвенирования и генотипирования людей из разных популяций.	1000 Genomes Project Consortium, 2015
VEP	Аннотирует варианты со специфическими и четко определенными изменениями и крупные структурные варианты. Каждая версия VEP привязана к определенному выпуску Ensembl, что гарантирует стабильность результатов.	McLaren W. et al., 2016
TopMed	Широкомасштабный анализ полногеномного секвенирования (WGS) людей с разнообразными фенотипическими данными. Предоставляет уникальные возможности для изучения вклада редких и некодирующих вариантов	Taliun D. et al., 2021

FAVOR	Предоставляет функционально аннотированные варианты путем интеграции данных из множества различных источников, таких как CADD, GENCODE, Annovar, ClinVar, SnpEff, 1000 Genome, dbSNP, gnomAD и др.	Zhou H. et al., 2023
-------	--	----------------------

Известно, что большая часть генетических вариантов представлена однонуклеотидными заменами (SNP). Разнообразие SNP в геноме обуславливает фенотипические различия между индивидами, но сложность заключается в выявлении конкретных SNP, которые имеют отношение к определенным фенотипам. В кодирующих белок регионах наибольший интерес представляют SNP, приводящие к несинонимичным заменам – замещающим одну аминокислоту на другую на уровне белка (миссенс варианты). Биологическая функция белка, образующегося с такими точечными мутациями, может быть частично или полностью нарушенной (Thusberg J. et al., 2011). Если замена произошла на участках белка, имеющих решающее значение для его функции, например, в каталитических сайтах или местах, ответственных за связывание лигандов, то такой структурно измененный белок потеряет свои функции (Karchin R. et al., 2005). Но, с другой стороны, замена аминокислоты на другую не всегда влияет на функциональные особенности белка. Было обнаружено, что некоторые миссенс варианты, ассоциированные с болезнями, полностью не изменяют каталитическую функцию аминокислот, а лишь дестабилизируют ее, что затрудняет их масштабное исследование и определение (Wang Z. et al., 2001). Известно, что точность предсказательных алгоритмов сильно зависит от варианта последовательности и имеет высокую значимость, когда доступны различные ортологи (гомологичные белки от разных организмов). С помощью количественной оценки физико-химических характеристик аминокислот, основанных на наблюдаемых эволюционных вариантах, можно получить довольно высокую прогностическую оценку. Избыточность таких миссенс-мутаций на ген или наборы генов можно

оценивать с помощью Z-оценок (Samocha K.E. et al., 2014). Однако некоторые варианты, которые приводят к нарушению функции белка и ассоциированы с заболеванием человека, могут быть неправильно аннотированы (Stone E. A. et al., 2005). Например, полученные выводы о серьезном нарушении функции белка в гене муковисцидоза *CFTR* часто не коррелируют с частотой и тяжестью заболевания (Dorfman R. et al., 2010), или функциональное заключение о тяжелых последствиях мутации в гене-супрессоре опухоли *TP53* существенно не влияет на клинические исходы у пациентов (Masica D.L. et al., 2015).

Многие мутации в кодирующих последовательностях могут никак не влиять на структуру белка, так как вырожденность генетического кода позволяет кодировать аминокислоты не одним триплетом. Раньше считалось, что такие синонимичные мутации никакого воздействия на последовательность генома не оказывают, однако, многочисленные исследования подтверждают их функциональное влияние (Xia X., 1996; Vinogradov A.E., 2003; Chamary J.V. et al., 2005; Parmley J.L. et al., 2006). Например, наличие более редкого кодона, чем дикий тип, в большей степени повлияло на эффективное действие гликопротеина Р, кодирующегося геном *MDR1*, который отвечает за множественную лекарственную устойчивость (Goumer P., 2007). Использование редких кодонов, по-видимому, влияет на скорость трансляции, что, в свою очередь, влияет на сворачивание белка (Xie T. et al., 1998). Редкие кодоны могут быть использованы, когда некоторые виды тРНК истощаются (Kimchi-Sarfaty S. et al., 2006). Другим примером, более ярко показывающим влияние синонимичных замен, является конформация ионных каналов – мембранных белковых комплексов – которые при помощи контроля градиента напряжения играют жизненно важную роль в различных клеточных процессах: сокращение сердца, активация Т-клеток, секреция гормонов и т.д. (Thorneloe K.S. et al., 2005; Krasznai Z., 2005; Mason W.T. et

al., 1988). Даже единичная мутация, затрагивающая данные последовательности, может привести к изменению проводимости канала (Schwartz J.L. et al., 1997).

Когда невозможно точно определять функции генов и вариантов, вводятся различные оценки, позволяющие рассчитать вероятностную степень воздействия или проявления гена/варианта. Значительный интерес в последнее время представляют LoF (loss-of-function) варианты, которые вызывают потерю функции кодирующих генов (Lek M. et al., 2017). Вероятностная оценка степени толерантности или нетерпимости гена к мутациям LoF – pLI из ресурса GnomAD (https://storage.googleapis.com/gnomadpublic/release/2.1.1/constraint/gnomad.v2.1.1.lof_metrics.by_gene.txt.bgz) предполагает, что гены, имеющие высокий показатель pLI ($\geq 0,9$), больше остальных нетерпимы к LoF вариантам, а гены с низким показателем pLI ($\leq 0,001$) – толерантны к LoF вариантам. При этом не все мутации с потерей функции вредны, так как было показано, что геном здорового человека содержит около 100 вариантов с потерей функции (MacArthur D. G. et al., 2012), а недавние исследования насчитывают 200-800 LoF вариантов на человека (1000 Genomes Project Consortium, 2015). Большое количество обнаруживаемых LoF вариантов может никак не влиять на здоровье человека, а только, вероятно, показывать устойчивость человеческого генома к мутациям. С другой стороны, нетерпимость генов к LoF вариантам зависит от того, где эти варианты располагаются в геноме здорового человека. Они могут быть: гетерозиготными вариантами при тяжелом рецессивном заболевании; гомозиготными вариантами у здоровых людей; нейтральными вариантами; ошибками секвенирования и аннотирования. Например, около 15% предсказанных LoF вариантов были исправлены в результате ручной аннотации (Rausell A. et al., 2020). Различные попытки проверить LoF варианты продолжаются и сегодня, так как обнаружили некоторые гетеро- и гомозиготные LoF варианты,

способные вызывать тяжелые менделевские болезни (Chong J.X. et al., 2015). При этом другие LoF варианты, вызывающие преждевременную остановку, сдвиг рамки считывания или ошибки в местах слияния экзонов оказывались нейтральными даже в случаях гомозиготности (Karczewski K.J. et al., 2020).

Наиболее очевидным патогенетическим механизмом реализации гетерозиготных мутаций с потерей функции является гаплонедостаточность. В такой ситуации одного нормального аллеля бывает недостаточно для компенсации эффекта мутантного аллеля, что приводит к нарушению функции гена (Huang N. et al., 2010). Механизм гаплонедостаточности заключается в дефекте фактора транскрипции, будь то делеция, вставка или точечная мутация, которая может преждевременно прекратить трансляцию. Часто такая инактивация одного аллеля может наполовину снизить уровень кодируемого белка и этого будет недостаточно для нормального функционирования многих факторов транскрипции, что, в свою очередь, приведет к разнообразным клиническим проявлениям (Seidman J.G. et al., 2002). Так как LoF варианты распространены среди здоровых людей, возникающая по их причине гаплонедостаточность генов является одним из возможных биологических механизмов, связывающим LoF варианты с их патогенным проявлением (Veitia R.A et al., 2010). Ресурс DECIPHER (<https://www.deciphergenomics.org/about/downloads/data>) предоставляет показатели гаплонедостаточности, которые определяются на основе прогнозируемых вероятностей. Баллы в диапазоне 0–10%, по версии ресурса DECIPHER, означают более высокую вероятность того, что ген будет гаплонедостаточным, а баллы в диапазоне 90–100% означают, что ген вряд ли будет гаплонедостаточным. Поэтому важность прогностических методов, позволяющих правильно интерпретировать гаплонедостаточные гены у человека, сложно переоценить. Для определения гаплонедостаточных генов в основном используют сети межбелковых

взаимодействий, но такие алгоритмы, как правило, хорошо работают на достаточно изученных генах. В недавних исследованиях гаплонедостаточности генов межбелковые сети заменяют коэкспрессией и машинным обучением, которое объединяет в себе геномные и эволюционные особенности с функциональным аннотированием (Steinberg J. et al., 2015; Shihab H.A. et al., 2017). Но часто бывает, что гаплонедостаточные гены вовлечены как в моногенные, так и в полигенные заболевания, так как часто экспрессируются во многих тканях и не являются тканеспецифичными, что и затрудняет их определение на молекулярном уровне.

Очевидно, что функционально аннотировать легче те варианты, которые кодируют белок, и, соответственно большая часть публикаций сосредоточена именно на таких вариантах. Однако, также известно, что мутации в некодирующей части генома, далекой от кодирующей последовательности, могут влиять на экспрессию генов. Данные GWAS показывают, что около 88% вариантов со слабым эффектом находятся в некодирующих областях (Choi M. et al., 2009), что довольно трудно подтвердить на молекулярном уровне. Одним из таких глобальных и общедоступных инструментов, который выполняет аннотацию и анализ большинства типов вариантов в кодирующих и некодирующих областях генома, является Ensembl Variant Effect Predictor (VEP) (McLaren W. et al., 2016). Автоматизируя аннотирование стандартным способом и сокращая время, необходимое для ручного просмотра, VEP помогает справиться со многими распространенными проблемами, связанными с анализом SNV (однонуклотидная замена в геноме одного вида или между гомологичными участками хромосом), короткими вставками или делециями, вариантами с большим количеством копий и структурными вариантами. VEP аннотирует варианты, используя широкий спектр справочных данных, включая транскрипты, регуляторные области, аллельную частоту вариантов,

информацию о клинической значимости и прогнозы о функциональных последствиях вариантов. Особый интерес для исследований представляют HI (high impact) варианты, выдаваемые Ensembl VEP, с предположительно высоким разрушающим действием на белок. Тестирование на наличие HI вариантов проводится не только на уровне генов, но и на больших специфических наборах генов. Например, значительный избыток редких HI вариантов наблюдается в генах, нетерпимых к потере функции, или среди миссенс-вариантов, у людей с психическими расстройствами (Ganna A. et al., 2018; Feng Y.C.A. et al., 2019; Zoghbi A.W. et al., 2021). Наличие редких HI вариантов в наборе генов, связанных с иммунными заболеваниями, было обнаружено среди пациентов с псориазом (Xu H. et al., 2021). В отношении инфекционных болезней влияние HI вариантов никогда не изучалось, однако, такой подход может быть интересен, поскольку восстановление после острых инфекций требует скоординированной работы многих генов (Talla A. et al., 2021).

Нахождение конкретных генов и вариантов, ответственных за определенные болезни, усложняет генетическая гетерогенность, которая является неотъемлемой частью биологических систем и способствует генетическому разнообразию (Altschuler S.J. et al., 2010). Поэтому выявленные генетические дефекты не всегда совпадают с клиническими проявлениями заболевания, так как реализация генетической информации до фенотипического проявления подвергается сложной иерархической системе, в которую также входят экологические факторы (рис.5).



Рис. 5. Реализация генетической информации до фенотипического уровня

В последнее время для связи количеств генетических вариантов со сложными признаками используют две основные гипотезы: полигенную и омнигенную. В полигенной модели варианты, связанные с заболеванием, объединяются в ключевые пути, которые определяют проявление и прогрессирование заболевания (Boyle E. A. et al., 2017). Согласно омнигенной модели, заболевание может быть ассоциировано с огромным количеством вариантов, находящихся в основных и периферийных генах, поскольку регуляторные генные сети тесно взаимосвязаны. Понятие омнигенной модели более обширно и не противоречит полигенной гипотезе. Омнигенная модель была разработана на основе общегеномных данных для распространенных генетических вариантов, но некоторые работы показали ее справедливость и для редких вариантов (Pullabhatla V. et al., 2018; Scelsi M.A. et al., 2021). Предполагают, что редкие варианты могут вносить более серьезный вклад в развитие распространенных заболеваний (Chen W. et al., 2022; Wang Q. et al., 2021).

ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ

1. Биологические образцы

1.1 Демографические и клинические данные пациентов

1.1.1 Пациенты с COVID-19

В исследование были включены пациенты с диагнозом COVID-19, находившиеся на стационарном лечении в Московском областном научно-исследовательском клиническом институте им. М.Ф. Владимирского (МОНИКИ), Московском клиническом центре инфекционных болезней (МКЦИБ) «Вороновское» и Городской клинической больнице Департамента здравоохранения Москвы имени В.П. Демикова с 27 апреля по 28 ноября 2020 года.

В исследование включались не вакцинированные пациенты обоих полов в возрасте от 18 до 96 лет, подписавшие информированное согласие на участие в исследовании. Критериями исключения были: терминальные неизлечимые заболевания; первичный и/или вторичный иммунодефицит; беременность; длительный прием кортикостероидов, алкоголизм, наркомания, ВИЧ. При поступлении всем пациентам был проведен ПЦР-тест на вирус SARS-CoV-2 из мазков носоглотки. Диагноз и тяжесть течения заболевания были установлены в соответствии с международными рекомендациями по профилактике, диагностике и лечению новой коронавирусной инфекции (COVID-19) (World Health Organization, 2020) и «Временными клиническими рекомендациями по профилактике, диагностике и лечению COVID-19 – 2020» Министерства здравоохранения Российской Федерации (BMP_COVID-19_V17.pdf (minzdrav.gov.ru)). Диагноз сепсиса был определен в соответствии с версией рекомендаций Surviving Sepsis Campaign 2021 года (Evans L. et al., 2021). Наличие ОРДС и определение степени его тяжести проводили в соответствии с Берлинским определением (ARDS Definition Task Force et al., 2012).

1.1.2 Пациенты с пневмонией и контрольная группа

Молодые пациенты (18–40 лет) с ВП и НП были набраны из числа госпитализированных в клинические стационары Федерального научно-клинического центра реаниматологии и реабилитологии в Москве с января 2008 г. по декабрь 2019 г. Критериями исключения для групп ВП и НП были: отсутствие информированного согласия; состояния, связанные с иммунной системой (инфекции, анафилактические состояния, аутоиммунные заболевания); хронические респираторные заболевания; злокачественные новообразования; декомпенсированные хронические состояния; обширные кровотечения и беременность. Контрольную группу составили лица соответствующего возраста, заявившие, что на момент участия в исследовании у них не было хронических или острых заболеваний. Диагноз ВП устанавливался на основании наличия острых симптомов инфекции нижних дыхательных путей, подтвержденных клиническими данными (кашель, выделение мокроты, аускультативные данные, соответствующие пневмонии, температура $>38\text{ }^{\circ}\text{C}$ или $<35\text{ }^{\circ}\text{C}$), рентгенологическими данными (инфильтрат на рентгенограмме грудной клетки) и лабораторными (количество лейкоцитов $>10\times 10^9$ клеток/л или $<4\times 10^9$ клеток/л) данными. Группу НП составили пострадавшие от несчастных случаев с физической травмой и пациенты, нуждающиеся в хирургическом вмешательстве, у которых НП развилась во время пребывания в стационаре. Диагноз НП был поставлен на основании нового инфильтрата на рентгенограмме грудной клетки при наличии клинико-лабораторных данных.

Тяжесть состояния пациента оценивали с использованием шкал оценки острого состояния физиологии и хронического состояния здоровья (APACHE) (Knaus W.A. et al., 1985), которые рассчитывали в течение 24 ч после поступления пациента. Диагноз сепсис был определен в соответствии с версией рекомендаций Surviving Sepsis Campaign 2021 года (Evans L. et al., 2021). Наличие ОРДС было основано на берлинском определении (ARDS

Definition Task Force et al., 2012). Тяжесть пневмонии оценивали с использованием индекса тяжести пневмонии (PSI) (Fine M.J. et al., 1997).

2. Молекулярно-генетические методы

2.1 Выделение ДНК пациентов с COVID-19 и секвенирование экзонов

ДНК выделяли из крови с использованием набора DNA blood mini (Qiagen). 50 образцов были секвенированы в компании Геномед (Москва, Россия). ДНК фрагментировали и штрих-кодировали с использованием набора Swift 2S[®] Turbo DNA Library Kit. Обогащение проводили с использованием Twist Human Core Exome (<https://www.twistbioscience.com/products/ngs/fixed-panels/human-core-exome>). Секвенирование осуществляли на платформе Illumina HiSeq X Ten с чтением парных концов длиной 150 п.н. 36 образцов были секвенированы в «Биобанк Центре», научно-исследовательском центре Санкт-Петербургского государственного университета (Санкт-Петербург, Россия). Образцы были подготовлены для секвенирования с помощью набора Illumina TruSeq DNA Exome (<https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-exome.html>) и секвенированы на Illumina HiSeq2500 и Illumina HiSeq4000 - платформы с чтением парных концов длиной 90 п.н.

Данные секвенирования 86 образцов анализировали в формате FASTQ. Показания были приведены в соответствии с GRCh38.p13 с использованием инструмента выравнивания Burrows-Wheeler Alignment (BWA MEM версии 0.6 <https://bio-bwa.sourceforge.net/bwa.shtml>). Контроль качества результатов секвенирования, выполненный в программе FastQC показал высокое качество показаний и наличие незначительного количества адаптерных последовательностей. Генетические варианты должны были иметь как минимум 10-кратное покрытие. Формирование списков вариантов

производилось с помощью HaplotypeCaller пакета GATK. Фильтрацию вариантов осуществляли с помощью программы VariantFiltration пакета GATK.

2.2 Подготовка наборов генов

Гены, связанные с иммунной системой, имеют некоторые особенности, отличающие их от генов других систем организма. Группа генов, ассоциированных с первичными иммунодефицитами (ПИД), разнообразна и включает гены, связанные как с частыми заболеваниями, так и с редкими, которые недостаточно изучены. Список генов ПИД был сформирован из обновленной классификации Комитета экспертов Международного союза иммунологических сообществ (IUIS) 2022 года (Tangye S.G. et al., 2022). Наборы генов-партнеров для ПИД генов были сформированы с помощью базы данных STRING 11.5 версии. База данных GENCODE была использована для отбора переаннотированных и обновленных генов, ассоциирующихся с COVID-19. Базы данных OMIM, ORPHANET, DDG2P, DisGeNet, MalaCards использовались для создания наборов генов, ассоциированных с заболеваниями человека. Иммунные тканеспецифические гены взяты из работы (Kolobkov D.S. et al., 2022), которые получали с помощью анализа обогащения наборов генов, используя пакет TissueEnrich R (Jain A. et al., 2019). Наборы генов, связанных с заболеваниями человека, поражающими определенные органы и системы, были взяты из работы Kolobkov D.S. et al. (2022), набор эссенциальных генов из Dickinson M. E., et al. (2016), гены цитокинов из Salnikova L. E. et al. (2020).

2.3 Аннотация генов и вариантов

Аннотацию вариантов проводили по программам SnpSift (Cingolani P. et al., 2012a), SnpEff (Cingolani P. et al., 2012b) и FAVOR (Zhou H. et al., 2023) с использованием баз данных dbSNP, ClinVar и популяционных баз данных GnomAD (Karczewski K.J. et al., 2020), 1000G (Консорциум проекта 1000

геномов, 2015) и TopMed (Taliun D. et al., 2021). Для идентификации генов с бременем мутаций, которые статистически различались у пациентов с тяжелым и легким/умеренным течением COVID-19, генетические варианты были сгруппированы. Варианты с предположительно разрушающим воздействием на белок были классифицированы как HI (high impact). HI варианты были определены при помощи (https://grch37.ensembl.org/info/genome/variation/prediction/predicted_data.html) и включали следующие варианты: потенциально патогенные варианты, вызывающие сдвиг рамки считывания, преждевременное образование стоп-кодонов, мутации акцепторного и донорного сайтов сплайсинга, мутации иницирующих и терминальных кодонов. Данные Haploinsufficiency Predictions Version 3 из Ресурса DECIPHER (<https://www.deciphergenomics.org/about/downloads/data>) использовались для оценки вероятностной гаплонедостаточности генов. Баллы в диапазоне 0-10% указывают на более высокую вероятность гаплонедостаточности гена. Для создания наборов генов, нетерпимых к потере функции и миссенс-вариантам, использовались рекомендуемые метрики $pLI > 0,9$ и $Z \text{ score} > 3,09$, соответственно, которые были определены на основе метрик из gnomAD v2.1 (https://storage.googleapis.com/gnomadpublic/release/2.1.1/constraint/gnomad.v2.1.1.lof_metrics.by_gene.txt.bgz). Частоты аллелей были загружены с веб-сайта GnomAD (<https://gnomad.broadinstitute.org/>) версии v.2.1.1 (hg19) и v.3.1 (hg38), а также ExAC (hg19). Были включены только те варианты GnomAD, которые прошли проверку качества. Преобразование версии генома человека hg19 в hg38 было реализовано с помощью Lift Genome Annotations (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>). Миссенс-варианты были определены как потенциально патогенные с использованием инструмента изучения ансамбля редких экзомных вариантов (REVEL) с рекомендуемым порогом $> 0,5$ для классификации варианта как «вредного» (Ioannidis N.M. et al., 2016). Анализ был сосредоточен на редких вариантах, которые были отфильтрованы в соответствии с частотами альтернативных аллелей (AF) в

базе данных геномов (GnomAD), базах данных 1000G и TopMed. Варианты были разделены по AF на три основных бина: с $AF \leq 0,001$ до $\leq 0,01$, с $AF < 0,001$ и без данных AF в популяционных ресурсах, т.е. в любой из рассматриваемых баз данных.

3. Количественное измерение показателей TREC/KREC

3.1 Создание генетических конструкций TREC, KREC и человеческого альбумина

Ген альбумина человека, последовательность T-клеточного рецептора δ TREC- ψ Ja и каппа-вырезаемую последовательность (делецию) перестройки B-клеточного рецептора амплифицировали методом ПЦР из ДНК, полученной из периферической крови здорового донора, и отдельно клонировали в вектор pBlueScript SK(II-) (Addgene). Последовательность T-клеточного рецептора (TREC) лигировали с сайтом рестрикции SpeI, делецию каппа- перестройки B-клеточного рецептора (KREC) лигировали по сайтам рестрикции HindIII и SpeI. Полученные плазмиды ДНК трансформировали в компетентные клетки *E. coli*. Эффективность тестировали на серии разведений плазмидной конструкции.

3.2 Выделение ДНК и количественная мультиплексная полимеразная цепная реакция (ПЦР) в режиме реального времени

Образцы крови, взятые при поступлении пациента, использовали для измерения уровней TREC и KREC. ДНК выделяли из периферической крови с последующей экстракцией ДНК изопропанолом. Образец смеси (20 мкл) для количественной ПЦР содержал 1,5 мкл праймеров для плазмиды TREC, 3 мкл праймеров для плазмиды KREC, 1,5 мкл праймеров для гена альбумина, 0,25 мкл FAM-меченого зонда TaqMan для sj-TREC и Cy5-меченого зонда TaqMan для гена альбумина, 0,5 мкл HEX-меченого зонда TaqMan для KREC, 1x буфер для ДНК-полимеразы HS Taq (Евроген, Россия), 0,2 mM dNTP, 5 mM MgCl₂,

0,5 ед ДНК-полимеразы HS Taq (Евроген, Россия), 150 нг ДНК и 0,4 нг бычьего сывороточного альбумина (БСА). ПЦР проводили в термоциклере CFX96 (Bio-Rad, Hercules, CA, USA). Амплификацию альбумина проводили в том же планшете. Последовательности праймеров и зондов приведены в таблице 4. Стандартный протокол амплификации состоял из начального нагревания при 95° С в течение 7 минут, за которыми следовали 45 циклов денатурации при 93 °С в течение 30 секунд и комбинированный отжиг праймеров/зондов при 59 °С в течение 1 минуты. TREC, KREC и копии гена альбумина рассчитывали путем экстраполяции соответствующих количеств образца из стандартной кривой, полученной из серийных разведений (10⁵, 10⁴, 10³, 10² и 10 (50 для KREC)) плазмидной ДНК, содержащей δREC–ψJα T-клеточный рецептор, каппа-делетирующую вставку, альбумин человека, которые амплифицировали в одном ПЦР-планшете. Количество TREC/KREC на 100000 клеток рассчитывали по следующей формуле:

$$\frac{\text{Количество копий } TREC(KREC)}{(\text{Количество копий гена альбумина})/2} \times 100000$$

Таблица 4. Последовательность праймеров и зондов

Праймер	Последовательность
TREC_F	CACATCCCTTTCAACCATGCT
TREC_R	TGCAGGTGCCTATGCATCA
TREC-TM Borte	(6-FAM) – ACACCTCTGGTTTTTGTAAAGGTGCCCACT – (BHQ-1)
KREC_F	TAAGCTTTCAGCGCCCATTA
KREC_R	CTCCAGGAGCCAGCTCTTAC
KREC_TM2	(HEX) – CTGTCTCTTCCCTTAGTGGCATTATTTG – (BHQ-1)
ALB_F	TGAAACATACGTTCCCAAAGAGTTT
ALB_R	CTCTCCTTCTCAGAAAGTGTGCATAT
TM_ALB	(Cy5) – TGCTGAAACATTACCTTCCATGCAGA – (BHQ-1)

4. Базы данных и ресурсы

4.1 Базы данных «генотип-фенотип» ClinVar и HGMD®

Клиническая значимость генетических вариантов представляет собой сложную проблему, зависящую от геноспецифического и вариант-специфического «генотип-фенотип» взаимодействий. Для детального и более глубокого изучения патогенности генов ПИД, данные по этим генам, представленные в ClinVar и HGMD®, были объединены на основе герминальных вариантов.

ClinVar — это свободный общедоступный ресурс ассоциаций вариантов с фенотипами человека, поддерживаемый Национальным институтом здравоохранения (NIH). Он включает варианты из OMIM, GeneReviews, некоторых локус-специфичных баз данных (LSDB), участвующих испытательных лабораторий и других отправителей (Landrum M.G. et al., 2020). База данных ClinVar была загружена с ftp-сайта (<ftp.ncbi.nlm.nih.gov/pub/clinvar/>) (по состоянию на 14 июля 2020 г.). После получения данных, связанных с генами ПИД (Tangye S.G. et al., 2020), все варианты, которые не имели четкой идентификации, были отфильтрованы. База данных мутаций генов человека (HGMD®) содержит данные о связанных с заболеванием генетических вариантах, опубликованных в рецензируемой литературе (Stenson P.D. et al., 2020). Данные HGMD® загружались отдельно для каждого гена и каждого конкретного типа варианта (по состоянию на 15 июня 2020 г.). Пакетный поиск по идентификатору доступа использовался для получения имен HGVS, координат hg38 и, если было возможно - идентификатора dbSNP и ссылок ClinVar. Информация об аллелях была взята из имен HGVS или извлечена с помощью дополнительных параметров поиска. Данные HGMD®, доступные в его общедоступной версии, были идентифицированы с помощью Ensembl BioMart (Kinsella R.J. et al., 2011). ClinVar и HGMD® были объединены на основе связей HGMD® и/или

положения вариантов и аллелей. Варианты из совместной базы данных названы ClinVar/HGMD.

4.2 Глобальный ресурс генетических вариантов человека dbSNP

Все генетические варианты в генах ПИД были загружены из dbSNP 154 версии. Мультиаллельные варианты dbSNP были разделены на двуаллельные варианты, поскольку разные вариантные аллели могут различаться по своим функциональным последствиям. Варианты без информации об аллель-специфических последствиях были удалены.

4.3 Транскриптомные данные для нормальных тканей и клеток крови человека из HPA

Данные по профилям экспрессии генов в тканях были взяты с сайта Атласа белков человека (HPA) <https://www.proteinatlas.org/about/download> (по состоянию на 20 августа 2021 г.). Были загружены и проанализированы следующие наборы данных: набор данных HPA (n = 19) и Monaco (n = 30) по уровням экспрессии генов в субпопуляциях клеток крови человека. Все данные HPA соотносились с версией 21.0 и 103.38 версией Ensembl. Наборы HPA содержали информацию о 19670 генах, набор Monaco содержал 18816 генов. Все гены Monaco присутствовали в наборе генов HPA. Для проведения анализа обогащения генов, специфичных для клеток крови использовался пакет TissueEnrich R (Jain A. et al., 2019). Гены, идентифицированные с помощью TissueEnrich, были сгруппированы в соответствии со степенью проявления признака в тех или иных клетках крови. teEnrichment (с параметрами по умолчанию) применялась для расчета обогащения тканеспецифических генов в группах генов заболеваний иммунной, респираторной и нервной систем (Jain A. et al., 2019).

5. Статистическая обработка данных

Статистический анализ и визуализация данных выполнялись с использованием программного пакета R (RStudio 1.4, <https://www.rstudio.com>), по состоянию на 10 июня 2022 г. и Excel 2023. Ассоциативный анализ наборов данных полноэкзомного секвенирования был проведен с использованием двустороннего теста Кокрана-Мантела–Хензеля, который выдает комбинированное значение *p value* и отношение шансов (ОШ). Сравнительный межгрупповой анализ проводили с использованием непараметрического U-критерия Манна-Уитни. Для категориальных переменных использовался точный двусторонний критерий Фишера или отношение правдоподобия χ^2 для двух и более независимых групп сравнения. Для сравнения двух или более независимых выборок одинакового или разного размера использовали критерий Крускала-Уоллиса, а также его параметрический эквивалент – односторонний дисперсионный анализ (ANOVA). Для измерения силы и направления связи между двумя переменными использовался коэффициент непараметрической корреляции Спирмена. Уровень значимости, при котором отвергалась нулевая гипотеза об отсутствии различий между исследуемыми группами, был установлен равным 0,05. Поправка Бонферрони или метод Бенджамини-Хохберга использовались для корректировки множественных сравнений. ROC-анализ (Receiver Operating Characteristic) использовался для оценки качества моделей бинарной классификации (DeLong E.R. et al., 1988).

ГЛАВА 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

1. Пациенты с COVID-19

1.1 Демографические и клинические данные

Выборка включала 86 пациентов с COVID-19. Пациенты не проходили вакцинацию. Группа с тяжелым/крайне-тяжелым течением COVID-19 состояла из 57 пациентов, группа с легкой/умеренной формой течения COVID-19 состояла из 29 пациентов. Группа не различалась по демографическим данным, но, ожидаемо, имелись различия в количестве пациентов, поступивших в отделение интенсивной терапии, которые нуждались в искусственной вентиляции легких и имеющих более тяжелую степень поражения паренхимы легких по данным КТ (компьютерной томографии) У всех умерших пациентов COVID-19 протекал в крайне-тяжелой форме. Основные характеристики пациентов показаны в таблице 5.

Таблица 5. Характеристика пациентов с COVID-19

Параметр	Тяжелая форма COVID-19 (57 чел.)	Легкая/умеренная форма COVID-19 (29 чел.)	Значимость (p)
Демография			
Возраст, лет (± Среднеквадр.откл.)	27 – 85 (59,6 ± 13,38)	24 – 96 58,62 ± 22,86	0,537 (критерий Манна-Уитни)
Пол (мужской)	32 (56,14%)	18 (62,07%)	0,649
Сопутствующие заболевания			
Отсутствует	1 (1,75%)	5 (17,24%)	0,015
Гипертония	41 (71,93%)	17 (58,62%)	0,232
Дислипидемия	2 (3,51%)	0 (0%)	0,548
Гиперхолестеринемия	1 (1,75%)	1 (3,45%)	1,000
Диабет 2 типа	20 (35,09%)	10 (34,48%)	1,000
Ожирение	18 (31,58%)	8 (27,59%)	0,806
Сердечно-сосудистые заболевания	30 (52,63%)	12 (41,38%)	0,367
Неврологические заболевания	5 (8,77%)	7 (24,14%)	0,096

Респираторные заболевания	10 (17,54%)	7 (24,14%)	0,569
Заболевания органов пищеварения/печени	14 (24,87%)	7 (24,14%)	1,000
Рак	3 (5,26%)	2 (6,9%)	1,000
Заболевания почек	10 (17,54%)	2 (6,9%)	0,323
Клинические данные			
Пребывание в отделении интенсивной терапии	55 (96,49%)	12 (41,38%)	$1,6 \times 10^{-8}$
Вентиляция легких: инвазивная не инвазивная отсутствовала	30 (52,63%) 1 (1,75%) 26 (45,61%)	1 (3,45%) 0 (0%) 28 (96,55%)	$2,3 \times 10^{-5}$ (критерий отношения правдоподобия χ^2)
Дни госпитализации (\pm Среднеквадр.откл.)	$17,95 \pm 10,79$	$13,7 \pm 5,9$	0,265 (критерий Манна-Уитни)
Оценка КТ грудной клетки по тяжести 0 1 2 3 4 Нет данных	1 (1,75%) 4 (7,02%) 7 (12,28%) 23 (40,35%) 17 (29,82%) 5 (8,77%)	1 (3,45%) 9 (31,03%) 9 (31,03%) 10 (34,48%) 0 (0%) 0 (0%)	$6,8 \times 10^{-4}$ (критерий отношения правдоподобия χ^2 для оценок 0 – 4)
Пневмония	57 (100%)	28 (96,55%)	0,337
ОРДС	18 (31,58%)	3 (10,34%)	0,035
Сепсис/септический шок	6 (10,53%)	0 (0%)	0,093
Исход: выздоровевших умерших нет данных	23 (40,35%) 31 (54,39%) 3 (5,26%)	29 (100%) 0 (0%) 0 (0%)	$2,7 \times 10^{-8}$

Для p значений использовался точный двухпараметрический тест Фишера, если в скобках не указано иное

1.2 Результаты полноэкзомного секвенирования

При проведении полноэкзомного секвенирования 86 пациентов с COVID-19 было определено 2504482 генетических варианта. После фильтрации осталось 2419486 вариантов; количество вариантов на человека составило $28133,56 \pm 914,86$. Количество уникальных вариантов составило 116932, они

были расположены в 16816 различных генах. Не было выявлено значимых ассоциаций в масштабах всего экзома с тяжестью COVID-19 (рис. 6).

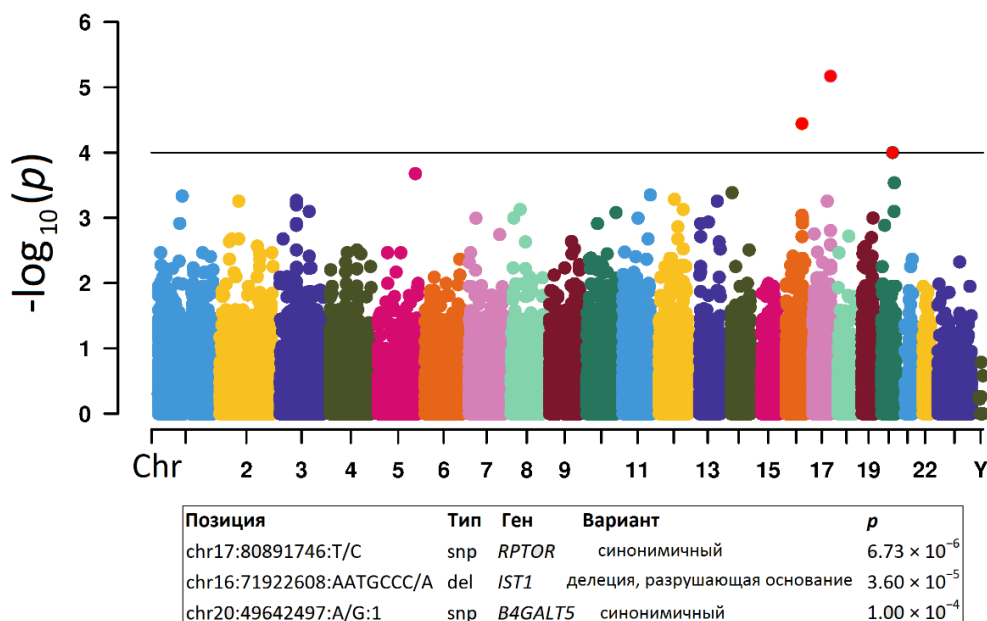


Рис. 6. Манхеттен-плот распределения частот вариантов по хромосомам. (В таблице представлены варианты, превышающие порог $1,0 \times 10^{-4}$, порог значимости с учетом поправки Бонферрони - $4,28 \times 10^{-7}$). График был построен с использованием <https://www.bioinformatics.com.cn/en>.

Генетические варианты всей выборки были распределены по бинам аллельных частот (AF), согласно популяционным данным, и различным функциональным последствиям. На рис. 7 показано, что наибольшее количество вариантов находилось в бине $AF > 0,01$ (47%), варианты в бинах $0,001 \leq AF \leq 0,01$, $AF < 0,001$ и без данных AF составили 17%, 19% и 16% соответственно. Наиболее многочисленными были миссенс варианты, что составило 51% от всех вариантов, за ними следовали синонимичные варианты (38%). Общее количество НИ вариантов составило 18896, количество уникальных вариантов НИ составило 5044. Наибольшая доля НИ вариантов была обнаружена в бине с отсутствующими данными по аллельной частоте (63%). Количество вариантов НИ на одного индивидуума составило $217,39 \pm 100,51$, что согласуется с литературными данными, согласно которым

количество НИ вариантов на одного индивидуума может варьироваться от 100 до 800 вариантов (MacArthur M.G. et al., 2012; Консорциум проекта 1000 геномов, 2015; Johnston J.J. et al., 2015; Lek M. et al., 2016).

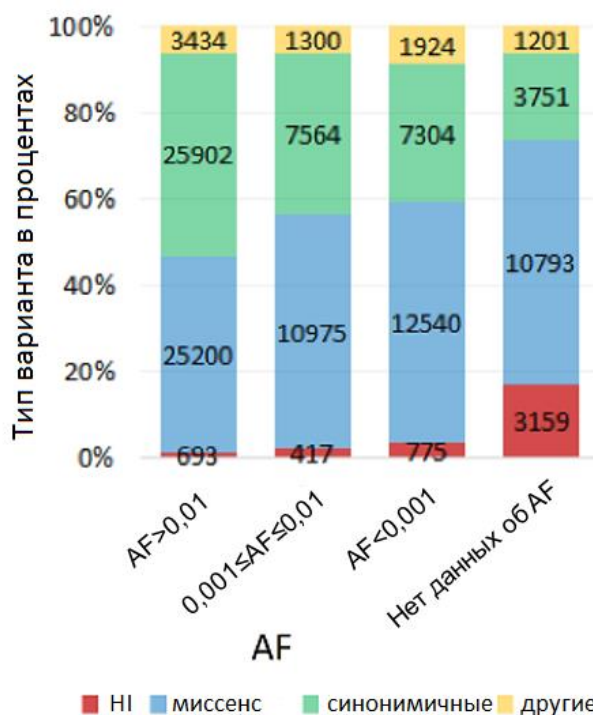


Рис. 7. Распределение вариантов по бинам аллельных частот (AF) в зависимости от функциональных последствий

Среди НИ вариантов большинство находились в гетерозиготном состоянии (81%), и более половины были представлены вариантами со сдвигом рамки считывания (67%) (Рис. 8). Общее количество вариантов указано в скобках.

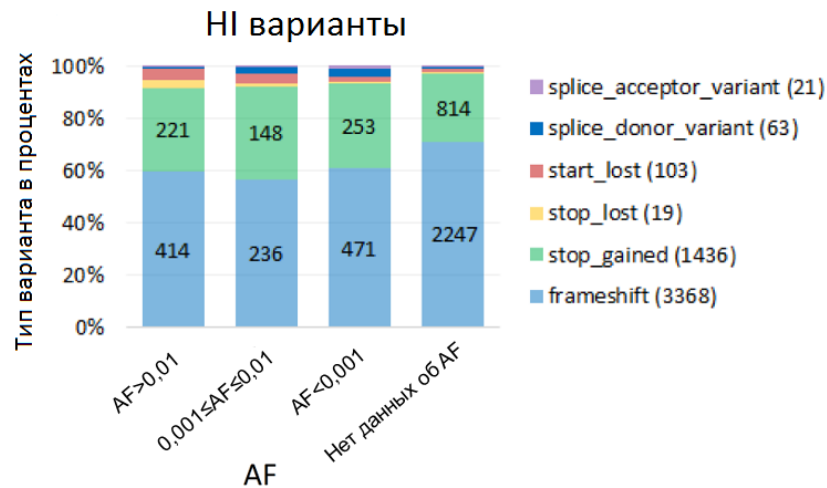
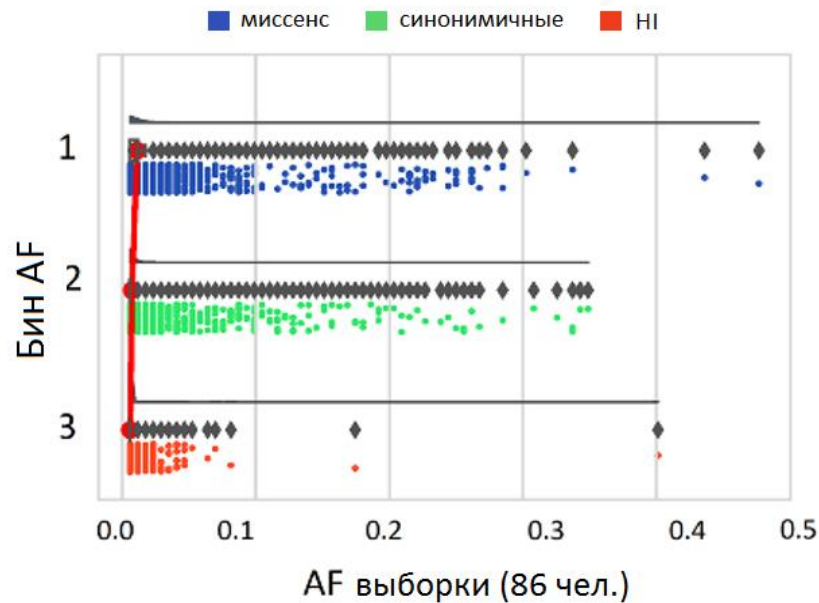


Рис. 8. Распределение НІ вариантов по бинам аллельных частот (AF) и функциональным последствиям. (Splice_аcceptor_variant – мутация акцепторного сайта сплайсинга; splice_donor_variant – мутация донорного сайта сплайсинга; start_lost – мутация иницирующего кодона; stop_lost – мутация терминального кодона; stop_gained – мутация, приводящая к преждевременному образованию стоп-кодона; frameshift – сдвиг рамки считывания)

Также было проведено сравнение аллельных частот нашей выборки с частотами из популяционных баз данных. Бины внутренних частот AF 1 ($0,001 \leq AF \leq 0,01$), AF 2 ($AF < 0,001$), AF 3 (нет данных об AF) по всему экзону (рис. 9) соответствовали популяционным данным. Бины AF 2 и 3 были схожими, но с меньшим разбросом данных в бине AF 3.



Бин AF	Бин AF	Среднее	Ср.кв.откл.
1	$0.001 \leq AF \leq 0.01$	0.0111	0.0149
2	$AF < 0.001$	0.0073	0.0110
3	no AF data	0.0060	0.0036

Рис. 9. Распределение бинов аллельных частот нашей выборки среди частот из популяционных баз данных

Принципиальный компонентный анализ (PCA) показал, что выборка из 86 пациентов была представлена европеоидами, за исключением 7 пациентов. При повторном анализе аллельных частот, исключая 7 не европеоидных пациентов, все значимые результаты сохранились (Приложение рис. 1).

1.2.1 Распределение редких патогенных/потенциально патогенных вариантов между пациентами с тяжелой (n=57) и легкой/умеренной (n=29) формой COVID-19

В данном исследовании акцент был сделан на изучение ландшафта редких генетических вариантов при различной степени тяжести заболевания с учетом омнигенной модели, при которой сложные признаки формируются при

совместном действии большого числа генетических вариантов, располагающихся по всему геному.

На первом этапе было произведено сравнение бремени таких генетических вариантов как НИ, миссенс и синонимичных у пациентов с тяжелой и легкой/умеренной формой течения COVID-19 (табл. 6).

Таблица 6. Генетический груз редких вариантов у пациентов с тяжелой и легкой/умеренной формой COVID-19

Бин аллельной частоты	ОШ (95% ДИ)	<i>P</i> value	Кол-во вариантов (генов)	Кол-во вариантов на человека (тяжелое течение COVID-19 /легкое или умеренное течение COVID-19)
НИ варианты				
0,001≤AF≤0,01	1,02 (0,88-1,18)	0,81	417 (370)	10,82 / 10,55
AF<0,001	1,38 (1,21-1,58)	2,24 ×10 ⁻⁶ *	775 (658)	14,77 / 10,93
Нет данных AF	3,69 (3,32-4,10)	7,3 ×10 ⁻¹⁵⁰ *	3159 (2413)	50,91 / 14,38
Миссенс варианты				
0,001≤AF≤0,01	0,96 (0,93-0,99)	4,32 ×10 ⁻³	10975 (6661)	239,84 / 249,86
AF<0,001	1,10 (1,06-1,14)	5,79 ×10 ⁻⁸ *	12540 (7292)	181,74 / 165,40
Нет данных AF	2,39 (2,27-2,51)	8,2 ×10 ⁻²⁷⁹ *	10793 (6629)	158,19 / 66,59
Синонимичные варианты				
0,001≤AF≤0,01	0,96 (0,93-1,00)	0,05	7564 (5254)	164,09 / 169,66
AF<0,001	1,10 (1,05-1,15)	2,52 ×10 ⁻⁵ *	7304 (5146)	110 / 100,48
Нет данных AF	2,32 (2,14-2,52)	3,22 ×10 ⁻⁹⁴ *	3751 (3011)	55,23 / 23,9

*Порог значения *p* для множественных сравнений с учетом всех протестированных сравнений составляет 5,21×10⁻⁴.

Было обнаружено, что у пациентов с тяжелой формой COVID-19 было значительно большее количество вариантов, относящихся к бину AF <0,001 и с отсутствующими данными о AF, и при этом значимого эффекта для редких вариантов в бине 0,001≤AF≤0,01 не наблюдалось. Большой эффект наблюдался среди НИ вариантов, по сравнению с миссенс-вариантами, и наименьший эффект был среди синонимичных вариантов. Максимальный

эффект наблюдался для Н1 вариантов с отсутствующими данными о АФ, которых было в 3,5 раза больше на человека при тяжелой форме по сравнению с пациентами с легкой/умеренной формой COVID-19.

1.2.2 Анализ в подгруппах пациентов, разделенных по полу и возрасту

Варианты в бинах АФ <0,001 и с отсутствующими данными о АФ были объединены в одну группу под названием «редкие варианты», так как для них наблюдались значимые эффекты. Выборка состояла из 32 мужчин и 25 женщин с тяжелым течением заболевания, из которых 13 пациентов были старше 70 лет и 44 пациента моложе 70 лет, а также из 18 мужчин и 11 женщин с легким течением заболевания, из которых 9 пациентов были старше 70 лет и 20 пациентов моложе 70 лет. Для всех групп наблюдались значимые результаты (табл. 7).

Таблица 7. Генетический груз редких вариантов в группах пациентов с тяжелой и легкой/умеренной формой COVID-19, различающихся по полу и возрасту

Сравниваемая группа	ОШ (95% ДИ)	<i>P</i> value	Кол-во вариантов (генов)	Кол-во вариантов на человека (тяжелое течение COVID-19 /легкое или умеренное течение COVID-19)
Н1 варианты				
Мужчины	1,97 (1,76-2,20)	$1,48 \times 10^{-33} *$	1595 (1311)	44,72 / 23,56
Женщины	3,41 (3,02-3,85)	$8,0 \times 10^{-99} *$	2467 (1907)	92,52 / 28,18
≥ 70 лет	5,24 (4,41-6,22)	$4,08 \times 10^{-98} *$	1200 (861)	85,92 / 17,33
< 70 лет	2,14 (1,95-2,34)	$5,88 \times 10^{-61} *$	2843 (2314)	59,7 / 28,9
Миссенс варианты				
Мужчины	1,32 (2,69-1,37)	$9,67 \times 10^{-49} *$	12584 (7444)	306,25 / 234,72

Женщины	1,7 (1,62-1,78)	$3,7 \times 10^{-123} *$	11349 (6855)	383,04 / 227,45
≥ 70 лет	1,95 (1,85-2,06)	$3,5 \times 10^{-136} *$	6662 (4598)	397,23 / 207,11
< 70 лет	1,34 (1,28-1,38)	$5,85 \times 10^{-67} *$	17180 (8958)	323 / 243,15
Синонимичные варианты				
Мужчины	1,24 (1,18-1,31)	$6,11 \times 10^{-17} *$	6225 (4533)	155,25 / 126,28
Женщины	1,49 (1,40-1,58)	$1,92 \times 10^{-36} *$	5214 (3930)	178 / 121,27
≥ 70 лет	1,47 (1,36-1,58)	$2,15 \times 10^{-24} *$	3027 (2489)	172,92 / 120,22
< 70 лет	1,3 (1,25-1,37)	$2,39 \times 10^{-29} *$	8348 (5647)	162,95 / 126,25

*Порог значения p для множественных сравнений с учетом всех протестированных сравнений составляет $5,21 \times 10^{-4}$.

Во всех подгруппах наибольшие эффекты наблюдались для НИ вариантов, за которыми следовали миссенс-варианты. В группе с тяжелым и легким/умеренным течением COVID-19 количество редких вариантов было выше у женщин, чем у мужчин, и у пациентов старше 70 лет по сравнению с пациентами моложе 70 лет. Самая высокая разница (в 5 раз) была в группе ≥70 лет среди НИ вариантов у пациентов с тяжелым течением заболеванием по отношению к более легкому течению заболевания, у которых было наименьшее количество вариантов НИ на человека, всего 17,3.

1.2.3 Редкие варианты в генах, нетолерантных к этим вариантам, при тяжелом и легком/умеренном течении COVID-19

Было проведено сравнение нагрузки редких вариантов для двух наборов генов, нетолерантных к потенциально патогенным вариантам: генов с высокой вероятностью нетолерантности к потере функции, т.е. нетолерантности к НИ вариантам (оценка $pLI > 0,9$), и генов, нетолерантных к миссенс-мутациям (оценка $Z > 3,09$) (табл. 8).

Таблица 8. Генетический груз редких вариантов в генах, нетолерантных к этим вариантам

Тип вариантов	ОШ (95% ДИ)	<i>P</i> value	Кол-во вариантов (генов)	Кол-во вариантов на человека (тяжелое течение COVID-19 /легкое или умеренное течение COVID-19)
Гены, нетолерантные к LoF мутациям (<i>pLI</i> > 0,9)				
НІ варианты	3,47 (2,84-4,23)	$1,04 \times 10^{-38} *$	808 (565)	13,32 / 3,86
Миссенс	1,64 (1,54-1,75)	$6,94 \times 10^{-53} *$	4773 (1899)	71,28 / 44,17
Синонимичные	1,31 (1,21-1,42)	$6,69 \times 10^{-11} *$	2683 (1433)	38,42 / 29,86
Гены, толерантные к LoF мутациям (<i>pLI</i> < 0,001)				
НІ варианты	2,48 (2,22-2,76)	$4,23 \times 10^{-64} *$	2062 (1531)	34,26 / 14,31
Миссенс	1,42 (1,36-1,46)	$3,43 \times 10^{-74} *$	12771 (5497)	185,16 / 132,17
Синонимичные	1,34 (1,27-1,41)	$1,11 \times 10^{-25} *$	5551 (3416)	85 / 64,21
Гены, нетолерантные к миссенс-мутациям (<i>Z</i> > 3,09)				
НІ варианты	3,28 (2,39-4,50)	$1,02 \times 10^{-14} *$	306 (212)	5,04 / 1,55
Миссенс	1,69 (1,52-1,89)	$4,06 \times 10^{-21} *$	1568 (579)	24,84 / 15,03
Синонимичные	1,44 (1,26-1,64)	$7,22 \times 10^{-8} *$	1043 (528)	15 / 10,62
Гены, толерантные к миссенс-мутациям (<i>Z</i> < 1)				
НІ варианты	2,52 (2,28-2,79)	$5,17 \times 10^{-75} *$	2322 (1748)	39,12 / 16,17
Миссенс	1,39 (1,34-1,44)	$1,07 \times 10^{-73} *$	14184 (6489)	205,11 / 149,14
Синонимичные	1,3 (1,23-1,37)	$3,37 \times 10^{-23} *$	6167 (3925)	94,3 / 73,76

*Порог значения *p* для множественных сравнений с учетом всех протестированных сравнений составляет $5,21 \times 10^{-4}$.

Суммируя эти результаты, можно сделать вывод, что чем более нетерпимы гены к вариантам, вызывающим потерю их функции, и чем более патогенными являются эти варианты (в ряду от НІ вариантов до синонимичных), тем больший эффект наблюдается при сравнении групп пациентов с различными течениями COVID-19.

1.2.4 Редкие НИ варианты у пациентов с COVID-19 в специфических наборах генов

Наборы генов, которые могут иметь биологически важное значение и быть потенциально вовлеченными в развитие и течение острой инфекции, были проверены на количество редких НИ вариантов (табл. 9). Наибольший эффект наблюдался для набора генов ПИД (n=71), за которыми следовали эссенциальные гены (n=562). В этих двух наборах генов размер эффекта был выше, чем во всем наборе генов с редкими НИ вариантами (n=2926). Два других набора генов, ассоциированных с сетью цитокинов (n=31), и обновленный список генов, связанных с инфекцией SARS-CoV-2 и/или заболеванием COVID-19 из проекта GENCODE (n=48), не показали значимого обогащения у пациентов с тяжелой формой COVID-19, возможно, из-за слишком маленького размера выборки.

Таблица 9. Генетический груз редких НИ вариантов в биологически важных наборах генов у пациентов с COVID-19

Набор генов	ОШ (95% ДИ)	<i>P</i> value	Кол-во вариантов (генов)	Кол-во вариантов на человека (тяжелое течение COVID-19 /легкое или умеренное течение COVID-19)
Все	2,7 (2,49-2,93)	$1,3 \times 10^{-137} *$	3934 (2926)	65,68 / 25,31
ПИД	4,53 (2,36-8,70)	$1,13 \times 10^{-6} *$	94 (71)	1,56 / 0,34
Эссенциальные	3,73 (3,04-4,59)	$7,57 \times 10^{-41} *$	785 (562)	13,18 / 3,55
Цитокины	2,8 (1,17-6,68)	0,02	39 (31)	0,58 / 0,21
GENCODE COVID-19	1,75 (1,03-2,98)	0,05	61 (48)	1,07 / 0,62

*Порог значения *p* для множественных сравнений с учетом всех протестированных сравнений составляет $5,21 \times 10^{-4}$.

Далее было проведено сравнение количества редких НИ вариантов в наборах генов, связанных с конкретными системными заболеваниями (табл.

10). Самый высокий размер эффекта наблюдался в наборе генов, связанных с иммунными заболеваниями (n=297), которые включали 69 генов ПИД, составляющих 23,2% от всех иммунных генов. Далее следовал набор генов, ассоциирующихся с респираторными заболеваниями (n=386), за которым следовал набор генов заболеваний нервной системы (n=752). Самый слабый эффект был обнаружен для набора генов, ассоциированных с опухолевыми заболеваниями (n=676). Редких НИ вариантов было в 2,9, 3,9 и 2,8 раза больше на человека при тяжелой форме течения COVID-19, по сравнению с пациентами с легкой/умеренной формой течения COVID-19 для наборов генов, относящихся ко всем заболеваниям, иммунным заболеваниям и опухолевым заболеваниям, соответственно.

Таблица 10. Генетический груз редких НИ вариантов в наборах генов, ассоциирующихся с системными болезнями, у пациентов с COVID-19

Набор генов	ОШ (95% ДИ)	<i>P</i> value	Кол-во вариантов (генов)	Кол-во вариантов на человека (тяжелое течение COVID-19 /легкое или умеренное течение COVID-19)
Гены болезней	3 (2,66-3,37)	$9,9 \times 10^{-82} *$	2024 (1481)	33,89 / 11,66
Иммунные	3,9 (2,90-5,24)	$4,27 \times 10^{-22} *$	390 (297)	6,67 / 1,72
Респираторные	3,53 (2,76-4,51)	$1,29 \times 10^{-26} *$	533 (386)	8,84 / 2,52
Нервные	3,48 (2,92-4,13)	$3,62 \times 10^{-50} *$	1054 (752)	17,75 / 5,41
Сердечно-сосудистые	3,21 (2,57-4,00)	$1,12 \times 10^{-27} *$	601 (440)	10,14 / 3,21
Кровеносные/лимфатические	3,12 (2,37-4,13)	$3,85 \times 10^{-17} *$	390 (295)	6,23 / 2
ЖКТ	3,12 (2,64-3,73)	$4 \times 10^{-40} *$	963 (701)	15,35 / 4,93
Урогенитальные	3,07 (2,55-3,68)	$2,73 \times 10^{-36} *$	848 (601)	14,51 / 5,03
Эндокринные/метаболические	2,87 (2,35-3,51)	$7,19 \times 10^{-27} *$	664 (488)	11,14 / 3,93
Опухолевые	2,79 (2,36-3,30)	$1,79 \times 10^{-35} *$	936 (676)	15,49 / 5,59

*Порог значения *p* для множественных сравнений с учетом всех протестированных сравнений составляет $5,21 \times 10^{-4}$.

Обсуждение. Представленная выборка имела значимую разницу в распределении редких генетических вариантов между пациентами с легким/умеренным и тяжелым течением COVID-19. Пациенты с тяжелой формой течения COVID-19 имели большее количество редких вариантов на уровне всего экзона. Также большее количество редких вариантов наблюдалось у женщин, чем у мужчин и пациентов старше 70 лет, по сравнению с более молодыми. Большее количество редких вариантов у женщин, по сравнению с мужчинами, может свидетельствовать о различии в иммунных реакциях, которые заключаются в генетических и гормональных особенностях. Например, общий лиганд *TLR7*, который в ответ на вирусную РНК начинает экспрессироваться на X-хромосоме, может определять эту разницу в эффекте дозы гена (Webb K. et al., 2019). Возраст пациентов также является серьезным фактором риска развития более тяжелых форм инфекционных заболеваний (Brodin P., 2021). Это напрямую связано с нормальным функционированием иммунной системы, так как с возрастом происходит увеличение экспрессии системных базальных медиаторов воспаления, ослабляется реакция Т-клеток, что способствуют чрезмерному воспалению, нарушению регуляторных путей и общему снижению способности иммунной системы противостоять инфекциям. Из-за нарушений в регуляторных путях избыток редких вариантов может еще более нарушать способности стареющей иммунной системы, провоцируя неконтролируемое воспаление.

Многофакторная и полигенная природа COVID-19 усложняет поиски ключевых генов, имеющих ассоциации с тяжелым течением инфекции (Casanova J.L. et al, 2020). Гены, ответственные за сложные мультифакторные заболевания, взаимодействуют друг с другом сложным образом, действующим по принципу полигенности и омнигенности (Mathieson I., 2021). При этом, существующая локусная и аллельная гетерогенность, когда один и тот же фенотип может быть обусловлен мутациями в различных генах или в

одном и том же гене, соответственно, приводит к сложностям при функциональной интерпретации вариантов. Известно, что инфекция SARS-CoV-2 в большинстве случаев переносилась легко, а большинство летальных случаев было ассоциировано с наличием у пациентов различных сопутствующих состояний (Salzberger B. et al., 2021). Количество редких вариантов среди пациентов данной выборки, тяжело перенесших инфекцию SARS-CoV-2, по всем сравниваемым группам было более высоким. Можно предполагать, что редкие варианты являются дополнительным фактором риска заболевания или изменчивости признаков и могут быть ассоциированы с различными распространенными нарушениями здоровья, однако установить точную связь с COVID-19 сложно, ввиду их низкой частоты или вообще отсутствия данных о частоте. Масштабный метаанализ по генетической предрасположенности человека к COVID-19 выявил одну связь с редкими вариантами в гене *TLR7*, ассоциирующуюся с тяжелым течением COVID-19 (Butler-Laporte G. et al., 2022). Однако существуют исследования, подтверждающие ассоциации редких вариантов со многими хроническими болезнями (Farhan S.M. et al., 2019; Xu H. et al., 2021; Ganna A. et al., 2018; Zoghbi A.W. et al., 2021). При этом тяжесть проявления заболеваний и их возникновение в более раннем возрасте дают основания полагать, что наличие меньшего числа редких вариантов у пациента будет способствовать более легкому течению инфекции и быстрому выздоровлению. В данной выборке наибольшее количество редких вариантов при тяжелом и легком/умеренном течении COVID-19 было среди НИ вариантов, за которыми следовали миссенс и синонимичные варианты. Такой эффект можно объяснить тем, что НИ варианты являются вариантами с предположительно разрушающим воздействием на белок, в то время как миссенс замена затрагивает только одну аминокислоту, а синонимичные замены, в большинстве случаев, нейтральны, но в ряде случаев могут приводить к нарушению транскрипции, сплайсинга или стабильности мРНК (Zeng Z. et al., 2019). Также более высокий уровень редких НИ вариантов наблюдался среди генов, нетолерантных к LoF и миссенс

мутациям, у пациентов с тяжелой формой течения COVID-19. Известно, что нетолерантные гены в здоровых популяциях подвергаются негативному отбору ввиду того, что они связаны со снижением плодовитости (Gardner E.J. et al., 2022).

В данной выборке количество редких вариантов в генах, связанных с заболеваниями иммунной и респираторных систем, ассоциировалось с тяжелым течением инфекции. Гены, связанные с раком и эндокринными/метаболическими заболеваниями, имели меньший размер эффекта. Примечательно, что наибольший размер эффекта редких НИ вариантов наблюдался в генах, ассоциированных с ПИД. Известно, что люди с ПИД, редкими наследственными дефектами иммунной системы, более подвержены инфекционным заболеваниям (McCusker C. et al., 2018). Популяции с ПИД могут иметь более глубокие последствия от COVID-19, по сравнению со здоровыми. Можно сказать, что гены ПИД являются основными генами, ассоциирующимися с тяжестью течения COVID-19. Недавний систематический обзор (Drzymalla E. et al., 2022) собрал данные о 459 пациентах с COVID-19, у которых также был подтвержден ПИД. Уровень смертности у таких пациентов составил 9%, госпитализации – 49%, а уровень в необходимом дополнительном кислороде при лечении – 29%. Эти данные в десятки раз выше, чем в общей популяции (<https://www.statista.com/statistics/1087466/covid19-cases-recoveries-deaths-worldwide/>). Поскольку иммунитет имеет сложную взаимосвязь действий между различными иммунными ветвями (Delavari S. et al., 2021), степень тяжести заболевания для людей с ПИД сложно предсказать, но это не противоречит тому, что часть таких пациентов будет более предрасположенной к тяжелому течению заболевания. Эссенциальные гены также показали повышенное количество редких НИ вариантов при тяжелом течении COVID-19. Данные гены имеют тенденцию к высокой экспрессии в тех тканях, с заболеваниями которой они связаны, а также они демонстрируют

тесную связь в генных сетях (Kolobkov D.S. et al., 2022). Вероятно, эссенциальные гены могут совместно регулироваться большим количеством генов, поэтому способны влиять на коровые гены, которые непосредственно ассоциированы с заболеванием.

2. Изучение генов ПИД с помощью онлайн ресурсов: «генотип-фенотип» баз данных ClinVar и HGMD[®], ресурса dbSNP и базы данных картированных белков человека HPA

Известно, что люди с ПИД, редкими наследственными дефектами иммунной системы, более подвержены инфекционным заболеваниям (McCusker C. et al., 2018). Также стало известно, что тяжелая форма течения COVID-19 наблюдалась у большого процента здоровых людей с протекающим бессимптомно ПИД, который у них не был подтвержден в течение жизни (Gray P.E. et al., 2022). У этих людей были обнаружены генетические дефекты, нарушающие врожденные иммунные функции, опосредованные интерфероном I типа. Другой обзор также подтвердил, что пациенты с COVID-19 и подтвержденным ПИД имели более высокий уровень госпитализации и смертности (Drzymalla E. et al., 2022). Так как группа заболеваний, относящихся к ПИД, одна из самых сложных в определении и характеризуется довольно широким спектром фенотипов (Bousfha A. et al., 2020; Tangye S.G. et al., 2020), было интересно рассмотреть патогенный ландшафт ассоциированных генов. Список из 403 уникальных генов (приложение рис. 2), ассоциированных с ПИД, был взят из отчета Международного союза иммунологических сообществ (IUIS) и рассмотрен в контексте «генотип-фенотип» баз данных ClinVar и HGMD[®], ресурса однонуклеотидных вариантов человека dbSNP, а также с помощью транскрипционной базы атласа белков человека (HPA).

2.1 Характеристика ландшафта патогенных вариантов в ПИД генах с помощью «генотип-фенотип» баз данных ClinVar и HGMD®

Из ClinVar было извлечено 97004 генетических варианта в генах ПИД. Половина вариантов (49,8%) имели противоречивую интерпретацию (CI) или неизвестное значение (VUS). Только 19,2% вариантов имели патогенное/вероятно патогенное значение (P/LP). Большое количество SNVs вариантов (84,2%) были среди вариантов CI/VUS, что составило 55,4% от общего количества вариантов. Основная часть патогенных вариантов была среди делеций (78,3%), инсерций (76,8%) и дупликаций (66,1%) (рис. 10).

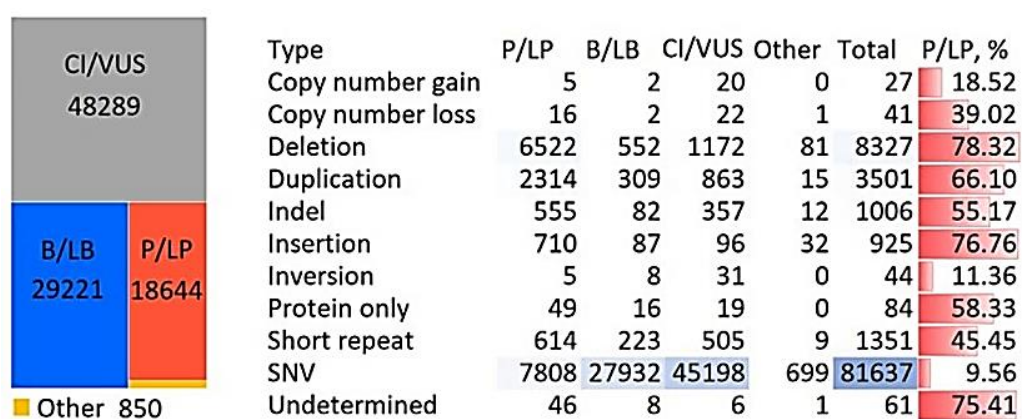


Рис. 10. Клиническая значимость и типы вариантов в базе данных ClinVar

(P/LP – патогенный/вероятно патогенный вариант; B/LB – доброкачественный/вероятно доброкачественный вариант; CI/VUS – противоречивые данные/вариант неопределенного значения; Other – другие; Total – общее количество).

В базе данных HGMD® среди всех 34743 вариантов в генах ПИД больше половины вариантов (77,2%) были охарактеризованы как мутации, вызывающие заболевания (DM). Миссенс варианты составили 39,6% от всех вариантов, небольшие делеции – 18,4%, варианты, приводящие к преждевременному образованию стоп-кодона (stop gain) – 12%. Среди всех

миссенс вариантов DM варианты составили 61,2%. Несмотря на это, кроме миссенс, регуляторных, синонимичных и повторяющихся вариантов, все остальные имели патогенность в диапазоне 85-95% (рис. 11).

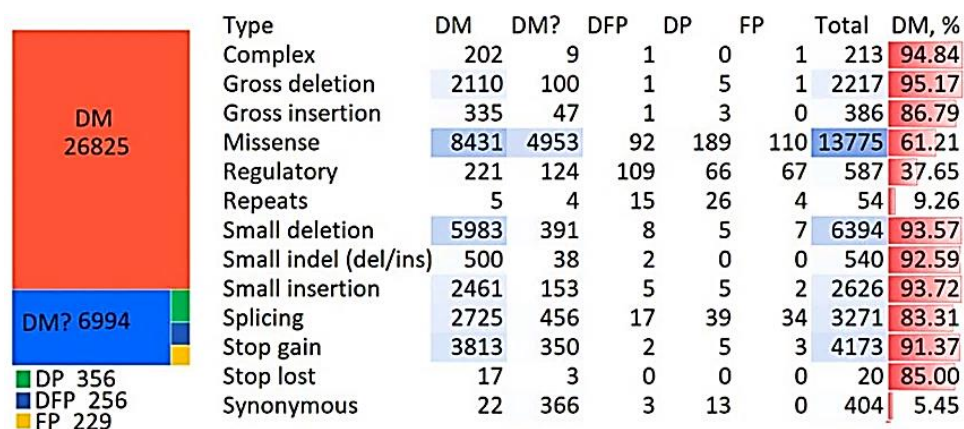


Рис. 11. Клиническая значимость и типы вариантов в базе данных HGMD®

(DM - мутация, вызывающая заболевание; DM? – мутация, вероятно вызывающая заболевание; DP – полиморфизм, ассоциированный с заболеванием; DFP - полиморфизм, ассоциированный с заболеванием и с дополнительными функциональными особенностями; FP – in vivo/in vitro функциональный полиморфизм; Total – общее количество).

Неполное соответствие вариантов в базах ClinVar и HGMD® объясняется их различным представлением данных и то, каким образом клинические данные попадают к ним. База ClinVar предоставляет данные всех фенотипов по каждому варианту, тогда как база HGMD® записывает первый фенотип по варианту, а все остальные фенотипы доступны только через дополнительные параметры. В базе ClinVar было доступно 2527 генотип-фенотип ассоциаций и 157287 вариант-фенотип ассоциаций (среди них 32641 P/LP варианта), однако после удаления записей с фенотипами, помеченными как «не предоставлено», «не указано» и «-», сохранилось 1829 генотип-фенотип ассоциаций и 111288 вариант-фенотип ассоциаций (27453 P/LP вариантов). База HGMD® включала данные о 3875 генотип-фенотип ассоциаций и 34743 вариант-фенотип ассоциациях (из них 26825 DM вариантов). Таким образом,

в базе HGMD[®] количество фенотипов для генов ПИД было в 2,1 раза больше, хотя количество вариантов было в 2,8 раза меньше, чем в базе ClinVar. Это может быть связано не только с большей представленностью фенотипов в HGMD[®], но также с разными уровнями детализации фенотипических дескрипторов в ClinVar и HGMD[®].

2.2 Объединение баз данных ClinVar и HGMD[®]

Для объединения баз были взяты варианты с известным положением, референсными и альтернативными аллелями. Получилось 80003 варианта из базы ClinVar, 17266 вариантов из базы HGMD[®] и 14671 вариант, одинаковый для обеих баз данных, всего 111940 вариантов. Классификация вариантов для объединенной базы ClinVar/HGMD изображена на рис. 12. DM варианты в базе HGMD[®] совпадали с P/LP вариантами в базе ClinVar на 79%, и наоборот, P/LP варианты в базе ClinVar совпадали с DM вариантами в базе HGMD[®] на 94,6%. Основываясь на классификациях, представленных в одной базе данных, 32452 варианта были определены как патогенные (DM/DM?, P/LP), 34643 варианта как доброкачественные (B/LB) и 44845 вариантов были отнесены к категории CI/VUS/Other.

ClinVar/HGMD	DM	DM?	Other	ND
P/LP	8281	462	13	8249
B/LB	142	730	273	28011
CI/VUS	1625	2557	81	43414
Other	440	43	24	329
ND	13857	3029	380	0

Рис. 12. Классификация вариантов в объединенной базе ClinVar/HGMD[®]

(P/LP – патогенный/вероятно патогенный вариант; B/LB – доброкачественный/вероятно доброкачественный вариант; CI/VUS – противоречивые данные/вариант неопределенного значения; DM - мутация,

вызывающая заболевание; DM? – мутация, вероятно вызывающая заболевание; Other – другие; ND – нет данных).

2.3 Распределение генов ПИД по паттерну наследования и функциональным последствиям

Информация о паттерне наследования в базе данных OMIM была доступна для 348 генов ПИД. Среди всех генов аутосомно-рецессивное (АР) наследование имеют 61,5% генов. Аутосомно-доминантный (АД), АР/АД и X-связанные (X) доминантные и рецессивные паттерны наследственности были определены для 19,5%, 13,2% и 5,75% генов, соответственно (рис. 13).

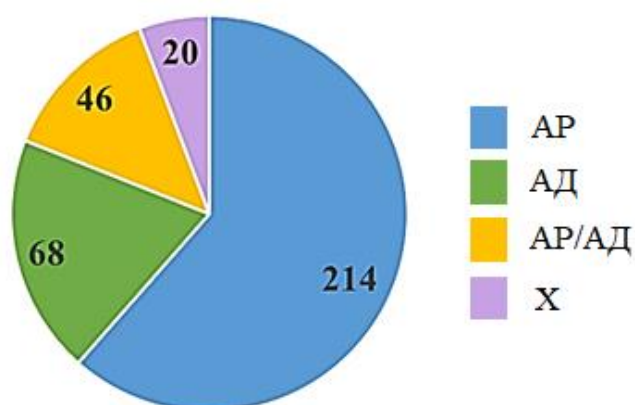


Рис. 13. Распределение генов ПИД по паттерну наследования

Далее были рассмотрены гены, нетолерантные к потере функции. Для функциональной интерпретации таких генов используется показатель вероятности потери функции – pLI, который имеет шкалу от 0 до 1, где 1 – это самая высокая оценка потери функциональности. Для 254 генов (63%) балл pLI находился в диапазоне 0-0,1 и для 89 генов (22,1%) - pLI>0,9. При этом 38,3% гаплонедостаточных генов имели оценку pLI в диапазоне от 0 до 0,1, хотя для гаплонедостаточных генов свойственен высокий показатель pLI. Процент АР генов среди низких оценок pLI составил 74,3% (рис. 14).

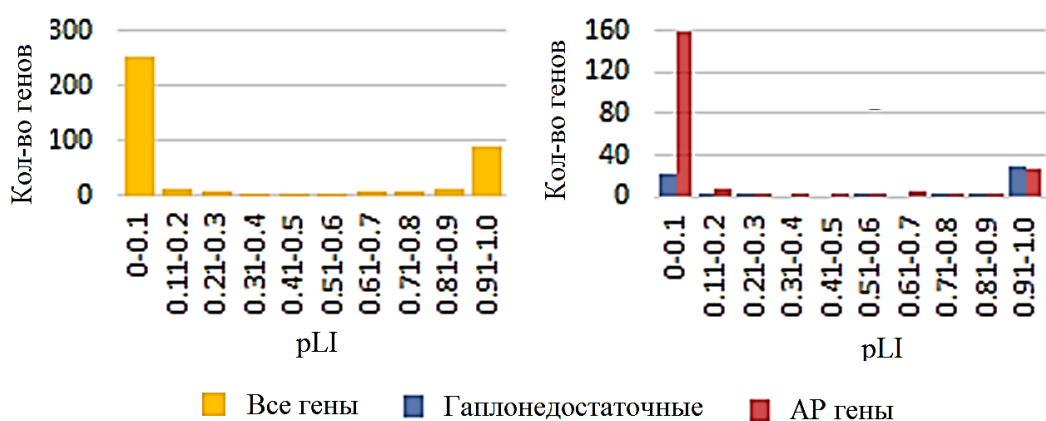


Рис. 14. Распределение генов ПИД с учетом вероятностного показателя pLI

2.4 Характеристика патогенных вариантов в генах ПИД по паттерну наследования

Доля патогенных вариантов в ClinVar/HGMD была значительно ниже в генах AP, чем в генах АД и AP/АД. Наибольшая доля патогенных вариантов ClinVar/HGMD была характерна для X-сцепленных генов, по сравнению с генами AP, АД и AP/АД (рис. 15).

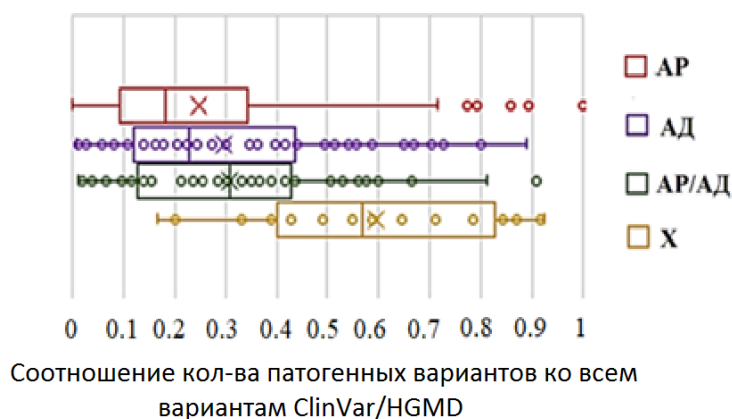


Рис. 15. Соотношение количества патогенных вариантов ко всем вариантам в генах ПИД с различным паттерном наследования

2.5 Сравнительный анализ вариантов в генах ПИД с помощью глобального ресурса генетических вариантов человека dbSNP

С помощью ресурса dbSNP было загружено 5415794 генетических вариантов генов ПИД. Среди 111940 вариантов ClinVar/HGMD с вариантами из ресурса dbSNP совпало 79807 (71,3%) вариантов. Чтобы сравнить представленность потенциально патогенных вариантов (HI вариантов) в генах ПИД в базе ClinVar/HGMD и в базе dbSNP, дополнительно использовался ресурс Ensembl для *in silico* выявления HI вариантов. Во всем пуле dbSNP вариантов был выявлен 38291 HI вариант, среди них 25500 HI вариантов (66,6%) отсутствовали в ClinVar/HGMD (рис. 16).

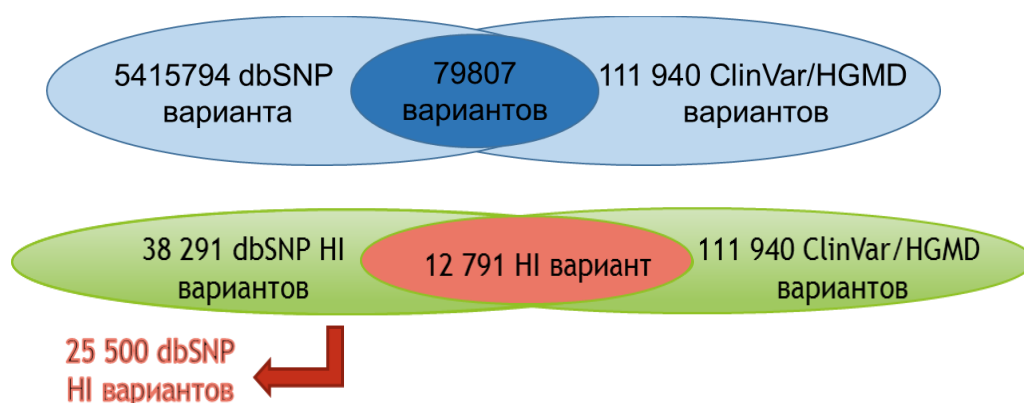


Рис. 16. Пересечение генетических вариантов между базой ClinVar/HGMD и ресурсом dbSNP

Также выявленные HI варианты в генах ПИД в базе ClinVar/HGMD были сравнены с экзомными данными по редким HI вариантам в генах ПИД у 86 пациентов с COVID-19. 7 из 105 редких HI вариантов (6,66%) у 86 пациентов с COVID-19 совпало с вариантами в объединенной базе данных ClinVar/HGMD.

2.6 Транскриптомный профиль ПИД генов в клетках крови здоровых людей

Известно, что гены, связанные с заболеванием, как правило, высоко экспрессируются в тканях того органа, с заболеванием которого они ассоциированы (Lage K. et al., 2008). Поэтому изменения в генах с высоким уровнем экспрессии, очевидно, приведут к патологии. При этом, была выдвинута гипотеза, что изменения в экспрессии генов чаще являются следствием самого заболевания, а не его причиной (Porcu E. et al., 2021). Результаты транскриптомных исследований могут быть полезны для определения генотип-фенотип связей, а также помогать в изучении молекулярных механизмов заболеваний человека. В связи с этим интересно было сравнить профили экспрессии ПИД генов среди профилей экспрессии всех генов в клетках крови человека, используя транскриптомный профиль тканей клеток крови здоровых людей из баз данных HPA (наборы данных HPA cells и MONACO). Данные по экспрессии были представлены в TPM (транскрипты на миллион). При сравнении уровней экспрессии среди клеток крови (HPA) некоторые типы клеток показали статистически значимые результаты (рис. 17). Уровни экспрессии генов ПИД в таких клетках, как неклассические моноциты и gdT-клетки были самыми высокими, по сравнению со всеми остальными. Уровни экспрессии ПИД генов среди остальных генов, экспрессирующихся в клетках крови человека (MONACO) были выше, кроме центральных CD8 Т-клеток памяти, CD4 Т-клеток памяти Th17, NK-клеток, неклассических моноцитов, наивных CD8 Т-клеток, регуляторных Т-клеток и мононуклеарных клеток периферической крови человека (рис. 18).

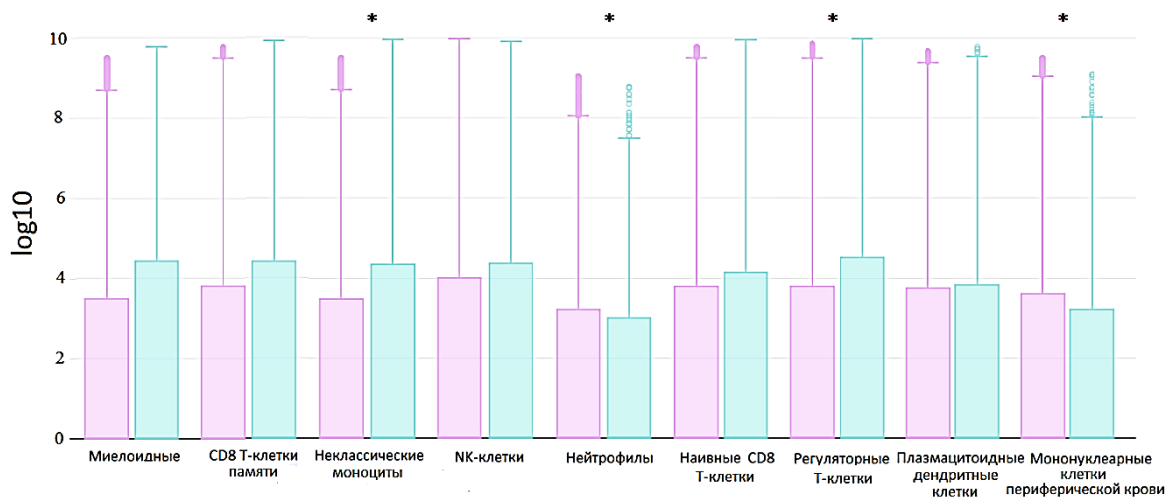
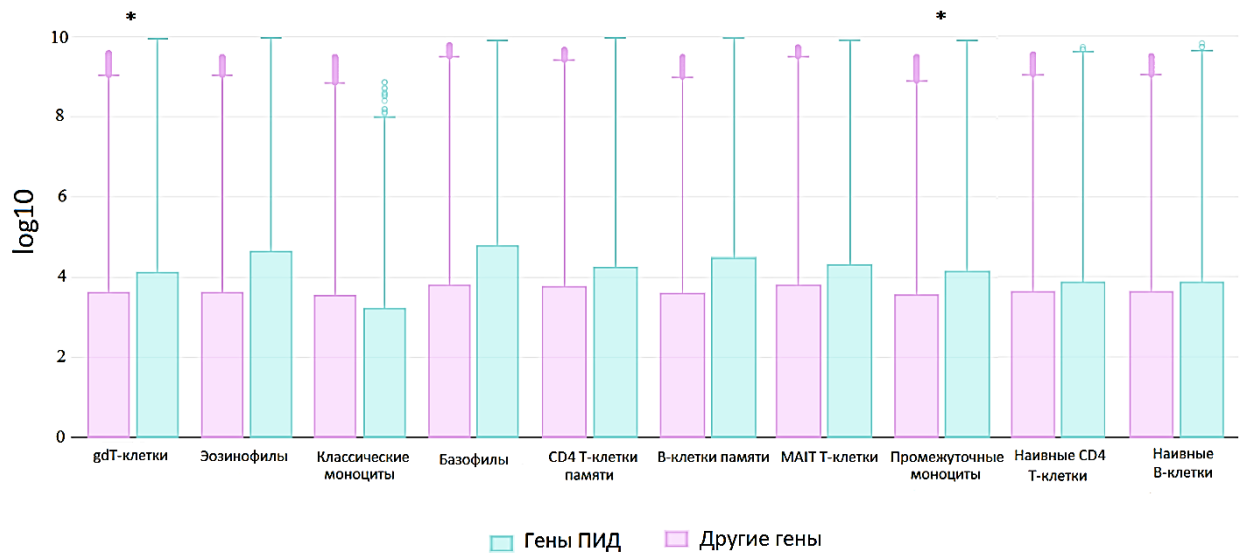


Рис. 17. Уровни экспрессии ПИД генов по сравнению со всеми остальными генами в клетках крови человека согласно базе данных HPA. (* значимый результат). Значения *p value* представлены Приложении табл. 1

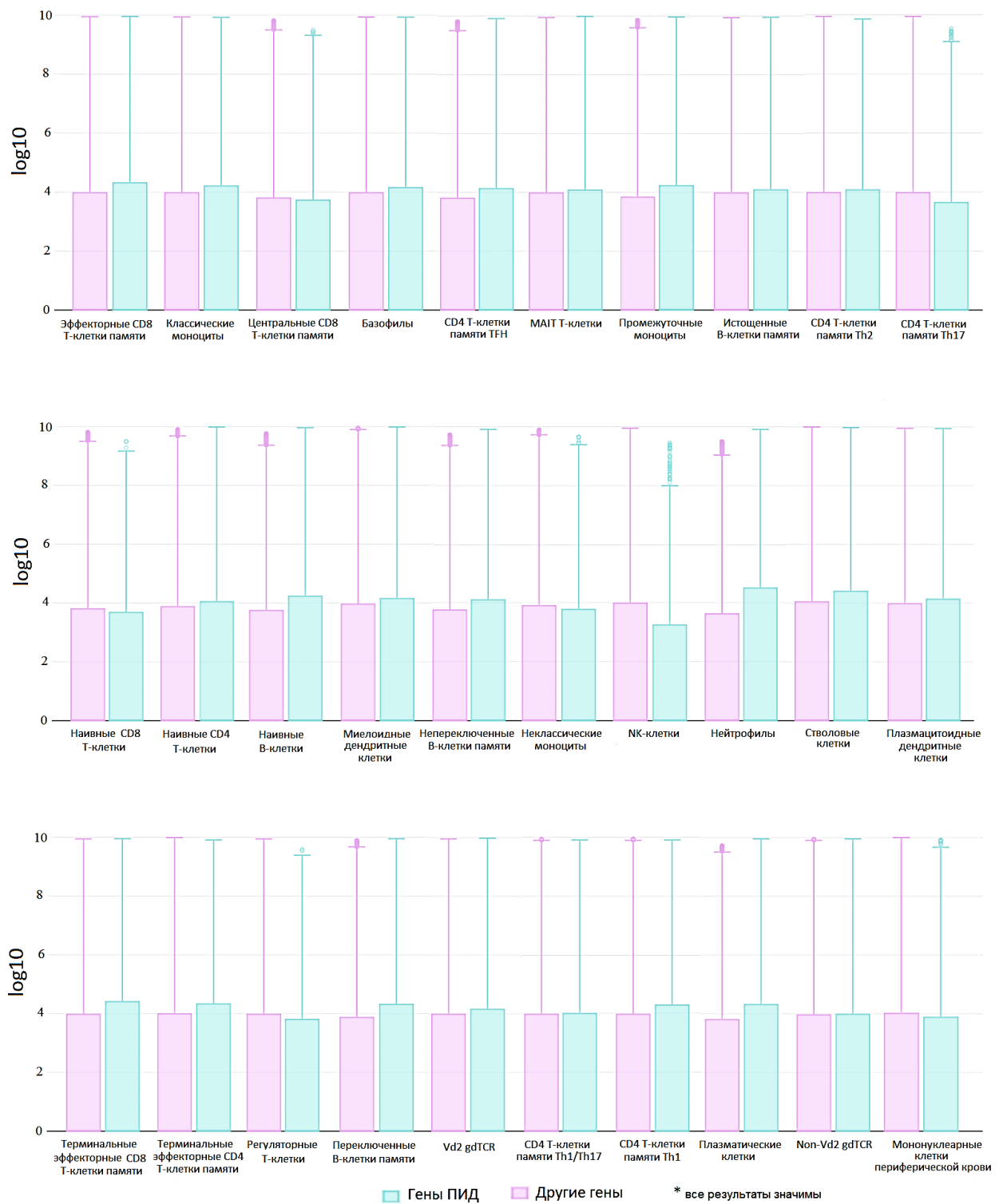


Рис. 18. Уровни экспрессии ПИД генов по сравнению со всеми остальными генами в клетках крови человека согласно набору данных MONACO. Значения *p value* представлены в Приложении табл. 2

Обсуждение. Для генов, ассоциированных с ПИД, был охарактеризован их патогенный ландшафт с использованием «генотип-фенотип» баз данных ClinVar и HGMD[®]. Стоит отметить, что представленные данные имели некоторые генетические, а также фенотипические различия. При сравнении патогенных или доброкачественных вариантов в двух базах, 94,7% патогенных и 83,7% доброкачественных вариантов совпали. При этом многие представленные фенотипы часто были идентичными или относились к схожим заболеваниям. Клиническая значимость генетических вариантов в генах ПИД является трудноразрешимой проблемой, которая зависит от специфического взаимодействия генотипа и фенотипа. Также постоянно растущее число новых выявляемых генов ПИД и их гетерогенность усложняет постановку генетического диагноза. Согласно регистру ESID, 96% пациентов с общим вариабельным иммунодефицитом (ОВИД) остались без генетического диагноза (Seidel M.G. et al., 2019). Это говорит о том, что ПИД плохо диагностируются и являются недостаточно изученными заболеваниями (Abolhassani H. et al., 2020; Поляков Д. В. и др., 2020). Бóльшее количество ассоциированных с ПИД генами имело аутосомно-рецессивный паттерн наследования и это согласуется с тем, что преобладающее большинство известных ПИД являются аутосомными или X-сцепленными рецессивными заболеваниями (Lawrence T. et al., 2005). При этом количество патогенных вариантов было самым низким среди AP генов, а также они имели самый низкий показатель нетолерантности к потере функциональности, лишь только малая часть AP генов имела высокую вероятность потери функции. Интересно, что бóльшая часть всех генов, ассоциированных с ПИД, имела низкий показатель потери функциональности, за которыми следовала группа генов с высокой вероятностью потери функции. При этом остальные гены расположились среди промежуточных оценок pLI. Потеря функциональности генов может возникать в зависимости от того, насколько патогенны генетические варианты в этих генах и какой характер заболеваний они вызывают – рецессивный или доминантный. Среди выявленных HI вариантов

66,6% вариантов отсутствовали в ClinVar/HGMD. Это явление можно объяснить индивидуальными особенностями, зависящими от сложного и многоступенчатого пути реализации генетической информации до фенотипического уровня, что отражается в разнообразии иммунологических дефектов и клинических диагнозах ПИД (Notarangelo L.D., 2020). Также могут иметь место артефакты секвенирования и аннотации (MacArthur M.G. et al., 2012; McCarthy D.J. et al., 2014).

При сравнении количества редких потенциально патогенных вариантов в ПИД генах у пациентов с COVID-19 с объединенной базой ClinVar/HGMD совпадение было менее, чем на 10%. Такое несоответствие может объясняться тем, что вклад одиночных патогенных вариантов в заболевание может компенсироваться на длительном и сложном пути реализации генетических эффектов на фенотипическом уровне (рис. 5). При этом наличие нескольких ассоциированных с заболеванием вариантов в генных сетях может значительно ослаблять компенсаторные механизмы. Не вошедшие 93,3% редких потенциально патогенных варианта, выявленных при секвенировании экзома в генах ПИД у пациентов с COVID-19, можно считать ранее не обнаруженными. Также, с внедрением генетических технологий, появляется много новой информации о геноме человека. Например, группа ученых произвела повторное WES/WGS предыдущих данных у взрослых пациентов с ПИД. Используя расширенную панель генов ПИД, ученые обнаружили новые варианты с неопределенным значением, ранее не выявленные у пациентов с ПИД (Mørup S.B. et al., 2022). Повторное секвенирование может улучшать диагностические характеристики для пациентов с подозрением на ПИД или с неподтвержденным генетическим диагнозом. Также из-за высокого уровня гетерогенности ПИД, интерпретацию редких вариантов, в особенности, ассоциированных с генами-кандидатами ПИД, и планирование функционального анализа

улучшают с помощью международного сотрудничества между пациентами, врачами и исследователями (GeneMatcher, <https://genemacher.org/>).

Гены иммунной системы богаты большим количеством функциональных вариантов (Petrovski S. et al., 2013), что предполагает частое возникновение мутаций, которое необходимо для формирования разнообразных иммунных функций (Collins A., 2015). Гены, ассоциированные с ПИД, сложно поддаются определению, так как их фенотипы часто могут имитировать менделевские или моногенные инфекции (Casanova J.L. et al., 2021). Было показано, что при полноэкзомном секвенировании семей пациентов с различными фенотипами ПИД с неизвестной этиологией, мутационное бремя в ПИД генах способствовало вариабельности заболевания и лежало в основе смешанных иммунофенотипов (Stray-Pedersen A. et al., 2017). Известно, что гены, связанные с врожденными нарушениями иммунитета, являются плеiotропными и участвуют в широком, но неслучайном спектре фенотипических нарушений. При этом Международный союз иммунологических сообществ (IUIS) классифицирует ПИД, опираясь на различия в иммунологических фенотипах и концепцию врожденного/адаптивного иммунитета. Но адаптивный иммунитет по своей природе ограничен Т- и В-клетками, поэтому такое разделение, основанное на молекулярных различиях, трудно связать с причинными генами. К тому же, иммунные гены экспрессируются не только в Т- и В-клетках, они являются повсеместно экспрессируемыми (Kolobkov D.S. et al., 2022). Гены и варианты, ассоциированные с фенотипами ПИД, влияют на многие типы иммунных клеток. Более того, генетические дефекты не гемопоэтических клеток могут серьезно нарушать развитие некоторых лейкоцитов, что видно на примере врожденных дефектов стромы тимуса, которое препятствует развитию Т-клеток (Casanova J.L. et al., 2021; Casanova J.L. et al., 2018). Исходя из полученных транскрипционных данных нормальных тканей, ПИД гены оказались более высокоэкспрессированными в клетках крови по отношению к

остальным генам. Известно, что гены с высокой экспрессией подвержены более медленной эволюции и, вследствие этого, вовлечены в различные генные сети (Gout J.F. et al., 2010; Drummond D.A. et al., 2005), поэтому любое изменение в этих генах приведет к нарушению функциональности тканей или клеток. Обнаруженная более высокая экспрессия ПИД генов в клетках крови, по сравнению с остальными генами, согласуется с существенным влиянием дефектов при дифференцировке гемопоэтических стволовых клеток в зрелые Т-лимфоциты и подтверждает сложность и многофакторность группы ПИД.

3. Гены ПИД как ключевые гены при COVID-19

Развитие сложного многофакторного заболевания связано со многими генами, действие которых регулируется генетическими и негенетическими факторами. В контексте омнигенной модели гены, ассоциированные с ПИД, и их близкие периферические гены, связанные с основными генами ПИД по сети белков (функциональные гены-партнеры), могут вносить вклад в развитие тяжелой формы течения COVID-19. Учитывая вклад НИ вариантов при тяжелой форме COVID-19, акцент был сделан именно на эти варианты во всех наборах сравнения.

3.1 Анализ генов ПИД с их функциональными генами-партнерами

Список генов ПИД был взят из отчета IUIS (Tangye S.G. et. al., 2022). Наборы генов-партнеров были сформированы с помощью базы данных STRING 11.5 версии. Было сформировано три набора: гены ПИД (1), функциональные партнеры ПИД генов с оценками комбинированных взаимодействий (combined interaction score) $\geq 0,9$ (2) и от 0,4 до 0,89 (3). Данная оценка показывает, насколько предполагаемая связь между белками является биологически значимой от 0 до 1, где 1 самая высокая взаимосвязь. Далее гены в 3-х наборах были сформированы по группам биологической значимости:

гаплонедостаточные, эссенциальные, нетолерантные к LoF-мутациям, нетолерантные к миссенс мутациям, имеющие отношение к SARS-CoV-2 инфекции и/или к COVID-19 (база данных

GENCODE), иммунные тканеспецифические, другие тканеспецифические и нетканеспецифические. Полученные наборы включали 450 генов ПИД (1), 4580 генов (2) и 6445 генов (3), которые в значительной степени пересекались (рис. 19).

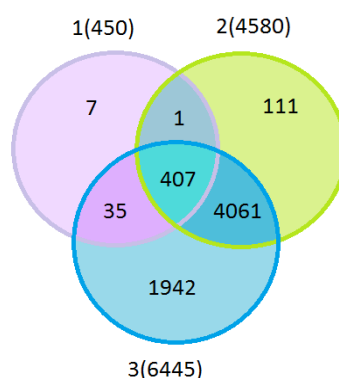


Рис. 19. Диаграмма Венна, показывающая одинаковые гены в наборах

Поскольку наборы перекрывались по генам, при последующем анализе гены ПИД были исключены из 2 набора, а также гены ПИД и гены 2 набора из 3 набора. Для сравнения эффектов редких вариантов в этих генах наборы генов были противопоставлены друг другу (табл. 11).

Таблица 11. Сравнительный анализ наборов генов по потенциальной значимости

Набор генов	ОШ (95% ДИ)	<i>P</i> value
Гаплонедостаточные ($\leq 0,1$)		
1 против 2	0,70 (0,54-0,90)	$1,0 \times 10^{-2}$
1 против 3	1,98 (1,47-2,67)	$8,9 \times 10^{-6} *$
2 против 3	2,83 (2,37-3,36)	$3,8 \times 10^{-33} *$
Нетолерантные к LoF-мутациям ($pLI > 0,9$)		
1 против 2	0,91 (0,73-1,15)	0,48
1 против 3	1,61 (1,25-2,07)	$2,4 \times 10^{-4} *$
2 против 3	1,76 (1,53-2,02)	$2,1 \times 10^{-15} *$
Нетолерантные к миссенс-мутациям ($Z > 3,09$)		

1 против 2	0,71 (0,48-1,04)	$9,7 \times 10^{-2}$
1 против 3	1,34 (0,88-2,04)	0,21
2 против 3	1,89 (1,50-2,37)	$3,7 \times 10^{-8} *$
Эссенциальные		
1 против 2	1,02 (0,82-1,25)	0,93
1 против 3	2,33 (1,84-2,93)	$6,3 \times 10^{-13} *$
2 против 3	2,29 (2,00-2,62)	$< 1,0 \times 10^{-300} *$
Из базы данных GENCODE COVID-19		
1 против 2	1,19 (0,74-1,92)	0,55
1 против 3	2,44 (1,41-4,25)	$2,0 \times 10^{-3}$
2 против 3	2,05 (1,42-2,94)	$1,0 \times 10^{-4} *$
Иммунные тканеспецифические		
1 против 2	3,10 (2,46-3,91)	$3,6 \times 10^{-23} *$
1 против 3	3,15 (2,44-4,06)	$8,6 \times 10^{-20} *$
2 против 3	1,02 (0,85-1,21)	0,89
Не иммунные тканеспецифические		
1 против 2	0,47 (0,35-0,63)	$2,0 \times 10^{-7} *$
1 против 3	0,25 (0,19-0,34)	$1,5 \times 10^{-22} *$
2 против 3	0,53 (0,47-0,60)	$1,7 \times 10^{-26} *$
Нетканеспецифические		
1 против 2	0,80 (0,65-0,97)	$2,9 \times 10^{-2}$
1 против 3	1,37 (1,11-1,68)	$3,8 \times 10^{-3} *$
2 против 3	1,72 (1,54-1,91)	$2,8 \times 10^{-22} *$

*Порог значения p для множественных сравнений с учетом всех протестированных сравнений составляет 0,0021

3.2 Анализ редких НИ вариантов в генах ПИД и их функциональных генах-партнерах у пациентов с COVID-19

В данный анализ были включены 20 молодых пациентов с COVID-19 в возрасте ≤ 45 лет. Основные демографические и клинические данные по пациентам представлены в таблице 5. В подгруппу с тяжелым течением COVID-19 вошли 9 пациентов (средний возраст \pm ср.кв.откл., $38,33 \pm 5,72$), а в подгруппу с легким или умеренным течением COVID-19 вошли 11 пациентов ($35,18 \pm 5,91$). Подгруппы не различались по демографическим характеристикам и предшествующим заболеваниям.

Так как размер выборки был небольшим и общее количество редких НИ вариантов в генах ПИД было незначительным (8 вариантов в 7 генах), анализ основывался сначала на сравнении суммарного количества редких НИ вариантов среди всех генов без разделения на конкретные группы, а затем в наборах генов (1+2, 1+2+3) между пациентами с тяжелой формой и легкой/умеренной формой течения COVID-19 (табл. 12).

Таблица 12. Генетический груз редких НИ вариантов у пациентов с тяжелой формой и легкой/умеренной формой течения COVID-19

Набор генов	ОШ (95% ДИ)	<i>P value</i>	Кол-во вариантов (генов)
Без разделения на группы генов			
Все	2,20 (1,83-2,64)	$5,9 \times 10^{-18} *$	473 (428)
1 + 2	3,11 (2,02-4,80)	$1,1 \times 10^{-7} *$	97 (88)
1 + 2 + 3	2,51 (1,80-3,52)	$4,5 \times 10^{-8} *$	147 (135)
Нетолерантные к LoF-мутациям ($pLI > 0,9$)			
Все	2,89 (1,81-4,62)	$5,4 \times 10^{-6} *$	81 (73)
1 + 2	5,23 (1,96-13,96)	$5,1 \times 10^{-4} *$	25 (22)
1 + 2 + 3	3,33 (1,61-6,88)	$1,1 \times 10^{-3}$	35 (30)
Гаплонедостаточные ($\leq 0,1$)			
Все	4,31 (2,13-8,73)	$1,9 \times 10^{-5} *$	44 (36)
1 + 2	8,35 (1,84-37,95)	$2,8 \times 10^{-3} *$	14 (13)
1 + 2 + 3	9,65 (2,15-43,23)	$8,1 \times 10^{-4} *$	16 (15)
Эссенциальные			
Все	3,00 (1,89-4,74)	$1,6 \times 10^{-6} *$	84 (73)
1 + 2	3,53 (1,65-7,53)	$9,2 \times 10^{-4} *$	35 (32)
1 + 2 + 3	3,20 (1,63-6,26)	$6,0 \times 10^{-4} *$	41 (37)
Иммунные тканеспецифические			
Все	2,44 (1,25-4,76)	0,01	38 (33)
1 + 2 = 1 + 2 + 3	2,03 (0,74-5,60)	0,25	16 (14)

*Порог значения *p* для множественных сравнений с учетом всех протестированных сравнений составляет 0,0025

Обсуждение. Данный анализ был основан на омнигенной гипотезе, с целью определения генов, ассоциированных с ПИД, как ключевых генов тяжелой формы течения COVID-19, а также роли их функциональных генов-

партнеров. При сравнении ПИД генов (1 набор) с наборами генов 2 и 3 с оценками комбинированных взаимодействий $\geq 0,9$ и от 0,4 до 0,89 было обнаружено, что они участвуют в схожих биологических процессах и метаболических путях. Было обнаружено избыточное количество редких потенциально патогенных вариантов (HI вариантов) на уровне всего экзома у пациентов с тяжелой формой течения COVID-19, что соотносится с выборкой, состоящей из 86 пациентов с COVID-19. При объединении ПИД генов с их функциональными партнерами были получены бóльшие эффекты, т.е. возрастание эффекта наблюдалось в ряду: все гены $< 1+2+3$ наборы генов $< 1+2$ наборы генов. Но более высокие ассоциативные эффекты были при рассмотрении отдельных групп генов у пациентов с тяжелой формой течения COVID-19, по сравнению с легкой/умеренной формой COVID-19, в частности, в генах, гаплонедостаточных, эссенциальных и нетолерантных к LoF-вариантам.

Омнигенная модель, по данным из литературы, определяет основные гены, как коровые гены с последовательными эффектами, а периферийные гены, как гены, которые сильно зависят от генетических факторов и факторов окружающей среды и, следовательно, обладают меж- и внутривидовой изменчивостью (Mathieson I., 2021). На то, как сильно периферический ген будет вносить вклад в общую регуляторную сеть, влияет его близкое расположение к основным генам. Чем меньше взаимосвязанных путей между периферическим и основным геном, тем сильнее ассоциативный эффект между ними. Данный анализ расширяет эту гипотезу тем, что показанная величина эффекта зависела от суммарного действия периферийных и основных генов (набор 1 + набор 2), где наблюдался более высокий эффект, по сравнению с эффектом всех остальных рассматриваемых генов. Постепенное выявление периферийных генов уже было упомянуто в литературе (Franke, B. 2017), и данный подход иллюстрирует и дополняет алгоритм выявления близких периферийных генов, находящихся рядом с основными ПИД генами.

4. Иммунный статус у пациентов с инфекционными заболеваниями

Основываясь на результатах анализа генетического ландшафта как ПИД генов, так и наборов других генов, становится ясным, что количество людей, которые потенциально могут быть более восприимчивыми к инфекционным болезням, недооценено, так как иммунный профиль хозяина зависит от его генетики (Salnikova L.E. et al., 2021). Учитывая всю сложность определения фенотипов сложных многофакторных заболеваний человека, эта проблема становится еще более актуальной, так как не всегда возможно оценить разницу между тяжелой и умеренной формами течения COVID-19. В связи с этим, дополнительным фактором, помогающим в определении высоких рисков возникновения тяжелых форм течения инфекционных заболеваний, являются различные иммунологические тесты (Mueller Y.M. et al., 2022; Sánchez-Montalvá A. et al., 2022; Khadzhieva M.B. et al., 2021; Арипова Т. У. и др., 2023). В данном разделе в качестве дополнительной оценки тяжести течения таких инфекционных заболеваний, как COVID-19 и пневмония, используется тест для количественного определения косвенных маркеров иммунного ответа – TREC и KREC, основанный на рекомбинации T- и B-клеточных рецепторов.

4.1 Пациенты с COVID-19

4.1.1 Сравнительный анализ уровней TREC/KREC между пациентами с легким и тяжелым течением COVID-19

Для группы из 20 молодых пациентов с COVID-19 (выборка описана в таблице 5) которые тяжело перенесли коронавирусную инфекцию (11 человек), и тех, кто легко перенес (9 человек) были измерены уровни TREC и KREC. Для TREC наблюдались более низкие уровни у пациентов с тяжелым течением COVID-19, по сравнению с легким течением (рис. 20).

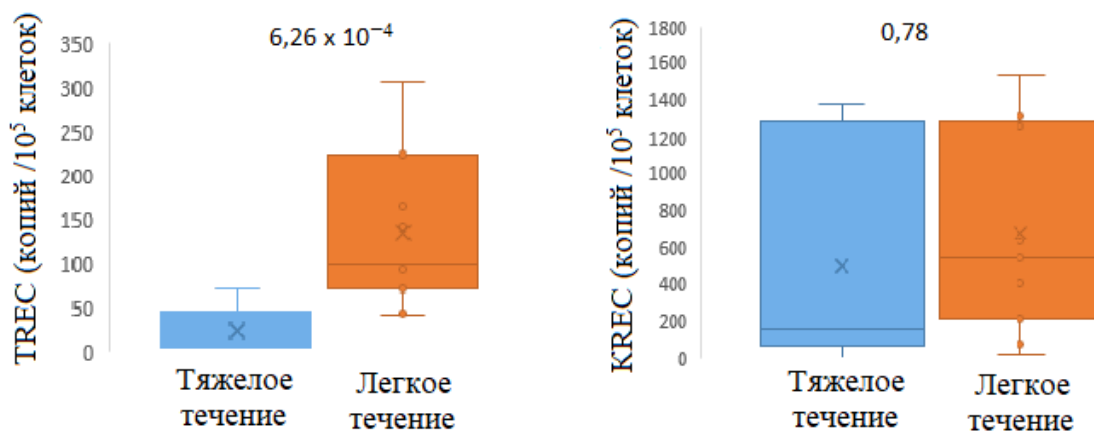


Рис. 20. Количество уровней TREC и KREC у пациентов с легким и тяжелым течением COVID-19

4.1.2 Диагностическая точность количественного анализа для TREC и KREC у пациентов с COVID-19

ROC-анализ выявил отличную диагностическую ценность TREC при тяжелом течении COVID-19 (AUC=0,96). Результаты для KREC были незначительными (рис. 21).

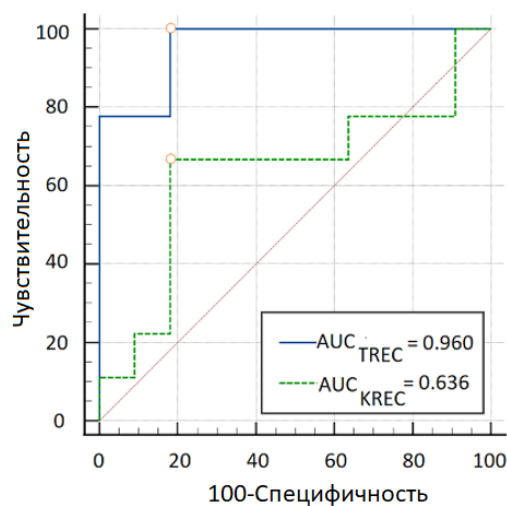


Рис. 21. ROC-анализ для количественных уровней TREC и KREC при прогнозировании тяжести COVID-19

4.1.3 Связь редких вариантов с количественными показателями TREC у пациентов с COVID-19

Далее был произведен корреляционный анализ числа редких вариантов с количественными показателями TREC. Редкие варианты в ПИД генах в бинах $0.001 \leq AF \leq 0.01$, < 0.001 и $< 0.001/c$ отсутствующими данными о AF отрицательно коррелировали с низкими значениями TREC у пациентов; $p \text{ value} = 9,1 \times 10^{-3}$, ОШ=1,27 (1,06-1,53)). Чем ниже были количественные показатели TREC у пациентов, тем больше редких вариантов было выявлено в этих генах (рис. 22). Коэффициент корреляции (r) для редких вариантов в ПИД генах составил -0,33.

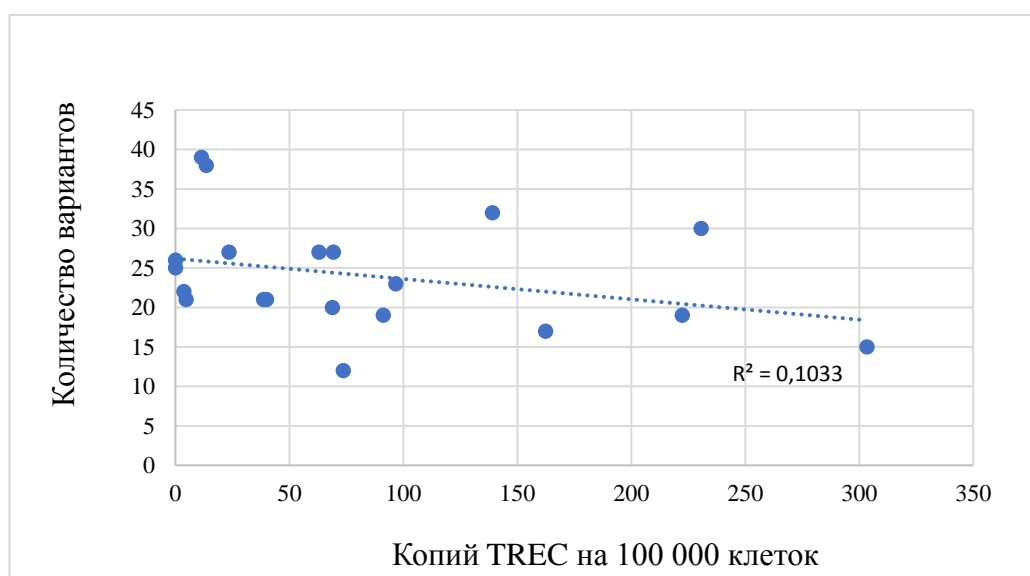


Рис. 22. Корреляция числа редких вариантов в генах ПИД с количеством TREC

Количество TREC также отрицательно коррелировало отдельно с редкими HI вариантами (в бинах $0.001 \leq AF \leq 0.01$ и < 0.001); $p \text{ value} = 3,27 \times 10^{-6}$, ОШ=1,58 (1,3-11,91)). Низкие значения TREC были ассоциированы с большим количеством наличия HI вариантов у пациентов (рис. 23). Коэффициент корреляции (r) составил -0,61.

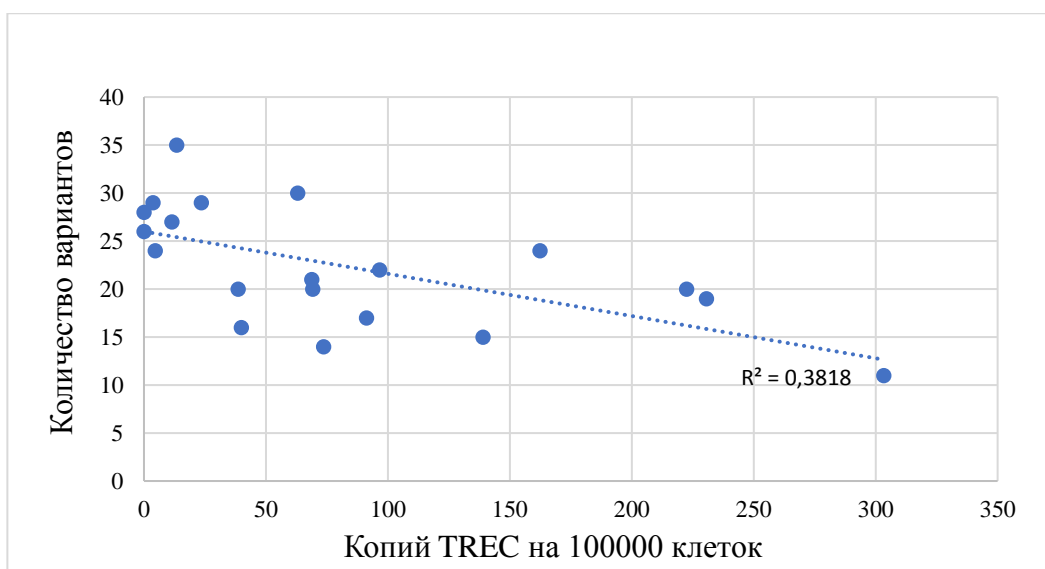


Рис. 23. Корреляция числа HI вариантов с количеством TREC

4.1.4 Сравнительный анализ уровней TREC и KREC между пациентами с ОРДС и без него

Показанная диагностическая ценность анализа TREC для дифференциации пациентов с повышенным риском тяжелого течения COVID-19 указывает на возможность его использования в качестве самостоятельного анализа для выявления пациентов с повышенным риском развития тяжелого течения COVID-19. Далее количественные уровни TREC и KREC были измерены для пациентов с инфекционными заболеваниями (COVID-19 и бактериальная пневмония) с различными сопутствующими состояниями.

Был проведен сравнительный анализ клинико-лабораторных данных и уровней TREC и KREC между пациентами с ОРДС (n =10) и без него (n =26) (выборка частично описана в таблице 5). Между этими группами не было обнаружено различий по возрасту, полу, дню манифестации болезни на момент анализа, количеству лейкоцитов, эритроцитов и тромбоцитов. Количественный анализ уровней TREC и KREC на 100000 клеток показал достоверные различия: значения TREC и KREC были ниже в группе больных с ОРДС, а также у не выживших, по сравнению с выжившими (рис. 24).

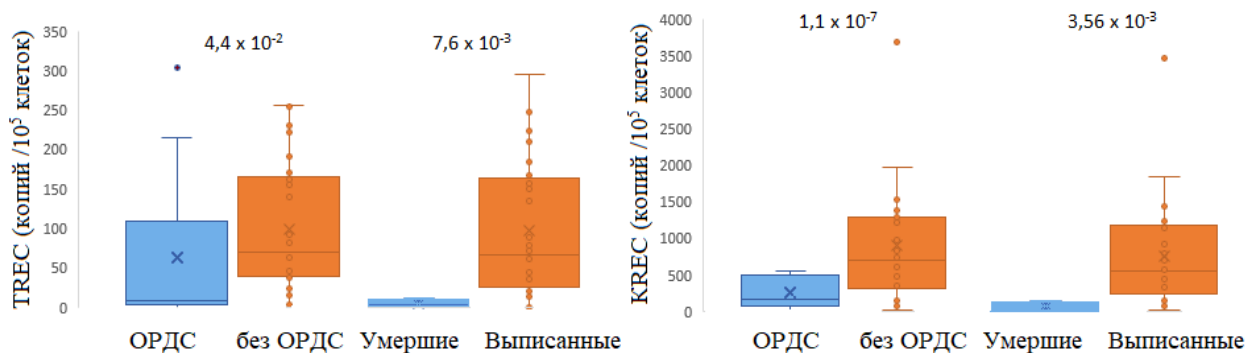


Рис. 24. Количество уровней TREC и KREC у пациентов с ОРДС/без ОРДС, умерших/выживших

4.2 Иммунный ответ хозяина, измеренный с использованием TREC и KREC у пациентов с бактериальной пневмонией

4.2.1 Пациенты с пневмонией и контрольная группа

Среди 433 пациентов, 263 пациента имели ВП или НП. Контрольную группу составили 170 пациентов (77,6% мужчины) в возрасте $23,52 \pm 5,93$ года. Контрольная группа была старше, чем пациенты с ВП (p value = $9,56 \times 10^{-4}$) и моложе, чем пациенты с НП (p value = $7,0 \times 10^{-6}$). Группа с ВП состояла преимущественно из мужчин (p value = $5,01 \times 10^{-8}$) и в ней было больше курильщиков, чем в контрольной группе (p value = $3,41 \times 10^{-6}$). Основные характеристики пациентов показаны в таблице 13.

Таблица 13. Характеристика пациентов с бактериальной пневмонией

Параметр	ВП	НП	<i>P</i> value
Общее количество	164	99	-
Возраст, лет (\pm Среднеквадр.откл.)	$21,29 \pm 3,89$	$26,68 \pm 6,68$	$4,52 \times 10^{-14}$
Пол (мужской)	159 (96,9)	93 (93,9)	0,34
Курят	126	44	0,09
Не курят	25	17	-
Травма	0	90	-
АРАСНЕ-II (\pm Среднеквадр.откл.)	$9,37 \pm 5,13$	$16,70 \pm 5,55$	$2,85 \times 10^{-19}$

Оценка PSI - 1	34 (20,7)	0 (0,0)	$3,85 \times 10^{-8}$
Оценка PSI - 2	42 (25,6)	6 (6,1)	$5,42 \times 10^{-5}$
Оценка PSI - 3	60 (36,6)	47 (47,5)	0,09
Оценка PSI - 4	24 (14,6)	32 (32,3)	$1,0 \times 10^{-3}$
Оценка PSI - 5	4 (2,4)	14 (14,1)	$5,48 \times 10^{-4}$
Двусторонняя пневмония	47 (28,7)	73 (73,7)	$8,44 \times 10^{-9}$
ОРДС	2 (1,2)	14 (14,1)	$3,45 \times 10^{-5}$
Сепсис/септический шок	0 (0,0)	29 (29,3)	$2,56 \times 10^{-14}$
Умершие	2 (1,2)	14 (14,1)	$3,50 \times 10^{-5}$

4.2.2 Уровни TREC и KREC у здоровых людей, пациентов с ВП и НП

Количественный анализ уровней TREC и KREC на 100000 клеток показал, что значения TREC в контроле были ниже, чем у пациентов с ВП, и выше, чем у пациентов с НП. Уровни KREC были выше в контроле, чем у пациентов с НП, и не отличались между контролем и пациентами с ВП (рис. 25).

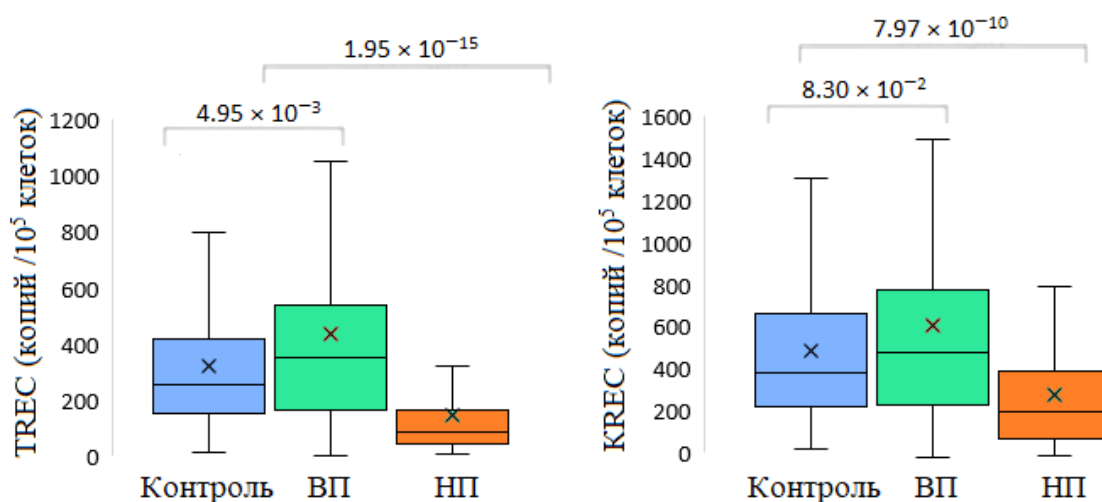


Рис. 25. Количество уровней TREC/KREC у контрольной группы и у пациентов с ВП и НП

4.2.3 Уровни TREC и KREC при пневмонии различной степени тяжести и исхода

Уровни TREC и KREC были ниже у пациентов с более тяжелым течением пневмонии по данным PSI (Pneumonia Severity Index), а также у

пациентов с двусторонней пневмонией по сравнению с односторонней (рис. 26).

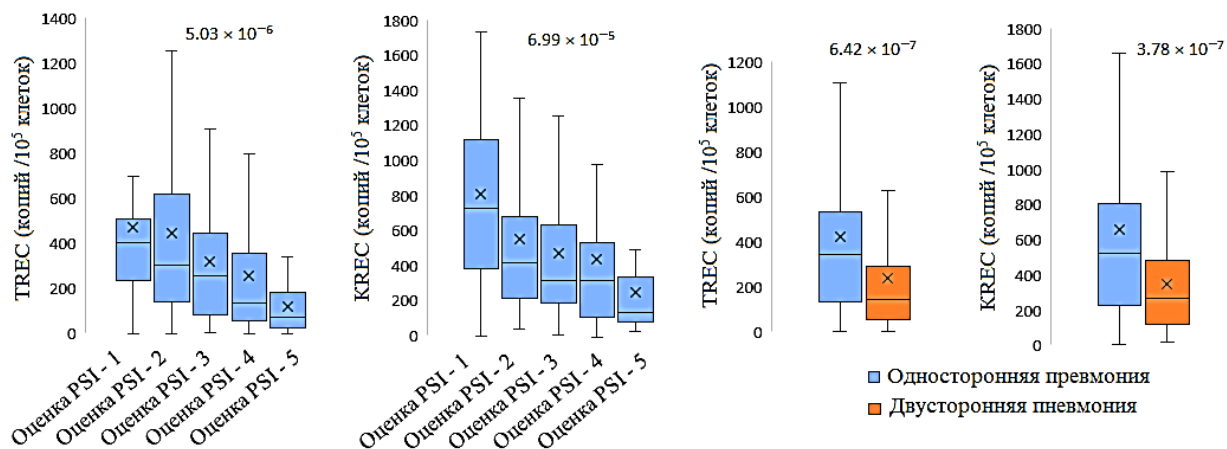


Рис. 26. Количество уровней TREC и KREC у пациентов с различной степенью тяжести пневмонии

Сравнение групп пациентов с сепсисом/без сепсиса, с ОРДС/без ОРДС и умершие/выписанные показало более низкие уровни TREC у пациентов с более тяжелым состоянием, причем самый сильный эффект был обнаружен у умерших пациентов по сравнению с выписанными. Тот же эффект наблюдался для уровней KREC, но с более низкой значимостью (рис. 27).

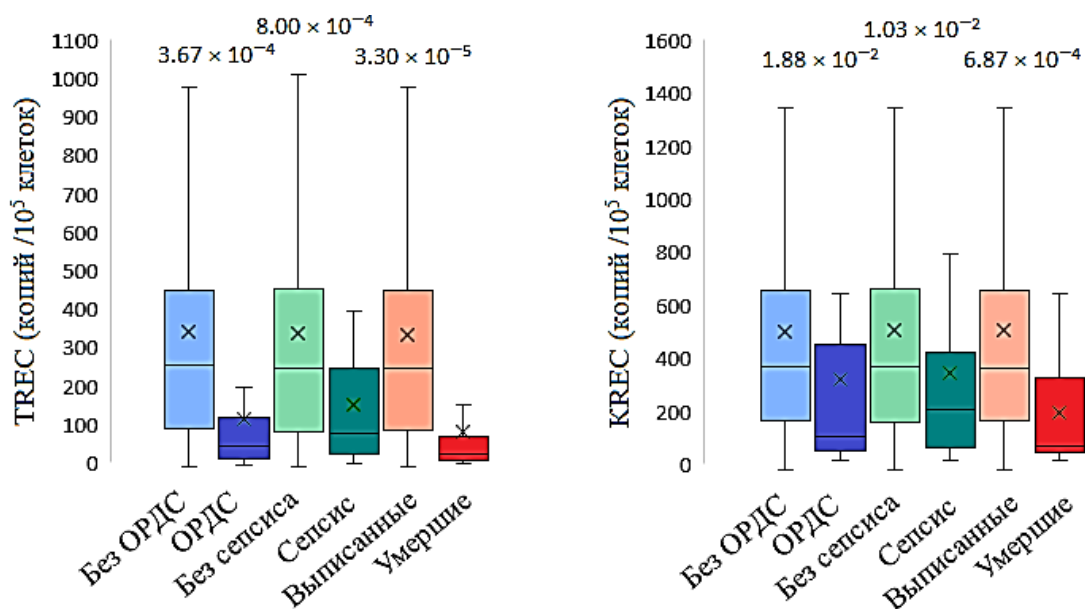


Рис. 27. Количество уровней TREC и KREC у пациентов с сепсисом, ОРДС и умерших пациентов

4.2.4 Диагностическая точность количественного анализа для TREC и KREC

Количественный анализ TREC продемонстрировал удовлетворительную диагностическую эффективность для НП, ОРДС и сепсиса/септического шока (AUC 0,7–0,8) и хорошую диагностическую эффективность (AUC 0,8–0,9) для летальных исходов. По сравнению с тестом TREC, тест KREC имел худшую диагностическую точность по отношению ко всем изученным конечным точкам (рис. 28).

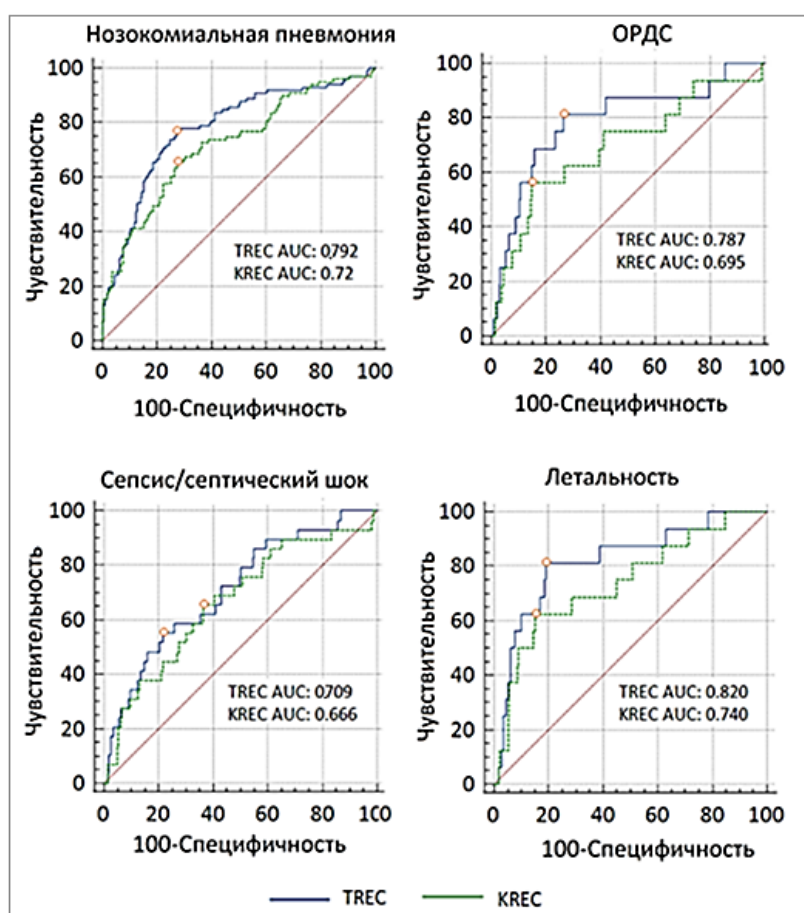


Рис. 28. ROC-анализ для количественных уровней TREC и KREC при прогнозировании тяжести течения пневмонии

Обсуждение. Во всех сравниваемых группах отягощенное течение инфекций коррелировало со сниженными маркерами иммунного ответа Т- и В-клеток (TREC и KREC). В выборке, состоящей из 20 пациентов с COVID-19, разница между уровнями TREC и KREC подтверждает различие между тяжелой и легкой/умеренной формами течения заболевания. Число редких вариантов в генах ПИД (в бинах $0.001 \leq AF \leq 0.01$, < 0.001 и $< 0.001/c$ отсутствующими данными о AF) отрицательно коррелировали с количественными данными по TREC. Группа HI вариантов среди редких вариантов генов ПИД имела еще большую отрицательную корреляцию, т.е. чем больше у пациентов было редких вариантов, тем более низким по количеству был маркер иммунного ответа – TREC. Анализ уровня TREC может быть полезным для разграничения пациентов с повышенными рисками тяжелого течения COVID-19. Учитывая вовлеченность иммунной системы во многие патологические процессы (Sattler S. 2017. Rankin L.C., 2018), наличие большего количества генетических дефектов и их более тяжелое проявление могут приводить к более раннему истощению иммунной системы, что отразилось на количественном уровне TREC.

Пациенты с COVID-19 и ОРДС имеют более выраженный иммунный дисбаланс, в виде системного воспаления и циркулирующих дисфункциональных Т- и В-лимфоцитов, что может усугублять течение инфекции (Yao C. et al., 2021). Известно, что тяжелая форма течения COVID-19 не приводит к противовирусной памяти, а вызывает нарушения в передаче иммунных сигналов, что позволяет вирусу долгое время сохраняться в организме человека, тем самым вызывая серьезные нарушения в легких, ОРДС и смерть (Giamarellos-Bourboulis E.J. et al., 2020). Несмотря на то, что механизмы, по которым коронавирусу удается уклоняться от защиты иммунитета до конца не известны, единым общеизвестным защитным путем является передача сигналов при помощи эндосомальных и цитозольных рецепторов с дальнейшей передачей сигналов через различные медиаторы,

такие как фактор некроза опухоли TNF, TRAF3, TRAF6 для активации интерферона IFN, регуляторных IRF и ядерных NF-κB факторов для вызова иммунного ответа, способного уничтожить вирус и подавить воспаление (Lester S.N. et al., 2014). Вероятно, что вирус может ингибировать многие чувствительные к нему рецепторы, инициируя тем самым нерегулируемый каскад воспалительных реакций организма человека, что и приводит к системному воспалению и ОРДС (Hu Y. et al., 2017; Li S.W. et al., 2016; Siu K.L. et al., 2009). Более того, анализы мононуклеарных клеток периферической крови пациентов с COVID-19 показывают скачки моноцитов и CD4+ Т-клеток, что дает полное основание предполагать, что нарушения в наивных Т-клетках помешают предотвратить цитокиновый шторм, тяжелое системное воспаление и в конечном итоге смерть от этих жизни угрожающих состояний.

Интересно отметить, что при инфекциях опасные для жизни состояния наблюдаются у менее чем 1 на 1000 человек в возрасте до 50 лет (Casanova J.L. et al., 2020). Известно, что пожилой возраст, сопутствующие заболевания являются основными факторами риска ослабления иммунной системы. В данной ситуации исключается такой сильно влияющий фактор как пожилой возраст, так как выборка состоит из пациентов до 45 лет, что предоставляет большую убедительность в корреляции между количеством редких генетических вариантов и косвенными маркерами образования наивных Т- и В-лимфоцитов. Самыми неисследованными когортами пациентов в данном случае остаются две группы: неинфицированные вирусом лица и ранее здоровые молодые пациенты с опасными для жизни клиническими исходами, вызванными инфекцией SARS-CoV-2. Выявленную значимую разницу как между генетическими вариантами, так и среди уровней маркеров иммунного ответа (TREC/KREC) между легкой и тяжелой формами COVID-19 среди пациентов молодого возраста, можно использовать как дополнительную оценку к уже имеющимся диагностическим тестам (Mueller Y.M. et al., 2022;

Sánchez-Montalvá A. et al., 2022), однако с существенной поправкой на маленький размер выборки.

Также количественный анализ уровней TREC/KREC, в качестве прогностической оценки, был проведен для группы молодых пациентов с ВП и НП пневмонией. Было также выявлено высокое разнообразие уровней TREC и KREC во всех сравниваемых группах. Пациенты с НП имели самые низкие уровни TREC/KREC среди всех групп сравнения, что, в целом, объясняется изначально тяжелым состоянием, способствующем развитию НП. Также было показано, что все критические состояния (сепсис, ОРДС, степень тяжести пневмонии, смерть) были ассоциированы с низкими уровнями TREC и KREC. Примечательно, что пациенты с ВП имели повышенный уровень TREC по сравнению с контролем, что, вероятно, отражает защитную реакцию тимуса как ранний ответ на инфекцию (Goggins J.A. et al., 2020). Это утверждение согласуется с данными Cuvelier P., et al. (2021), которые также показали защитную реакцию тимуса и повышение количества TREC у взрослых пациентов с COVID-19.

При тяжелом поражении легких вследствие пневмонии или сепсиса имеет место недостаток периферийных адаптивных иммунных клеток (Bermejo-Martin J.F. et al., 2017; Roth G. et al., 2003). Это связано не только с устремлением клеток в места воспаления, но также с дисфункцией и гибелью Т-лимфоцитов (Nakajima T. et al., 2010; Roth G. et al., 2003). При продолжительной антигенной стимуляции происходит истощение Т-клеток и дальнейший апоптоз, что, в свою очередь, приводит к недостаточной активации В-клеток, а также утрате иммунных функций Т- и В-клеток (De Biasi S. et al., 2020; Erickson J.J. et al., 2015). У пациентов с сепсисом наблюдается тяжелое истощение лимфоцитов в селезенке и лимфатических узлах, по сравнению с пациентами, также находящимися в критическом состоянии, но не имеющих сепсиса (Kahan S.M. et al., 2015). Потеря этих лимфоцитов существенно снижает иммунитет и приводит к неспособности

купировать инфекционный процесс. Например, люди без В-клеток в норме могут контролировать подавляющее большинство распространенных вирусов, по крайней мере, когда это возможно определить. Пациенты без аутологичных Т-клеток подвергаются гораздо большему спектру инфекций. Также следует отметить, что мутации в основных доменах *RAG1/RAG2* могут влиять на качество процесса рекомбинации генов V, D и J, необходимых для правильной работы рецепторов Т- и В-клеток (Wong S.Y. et al., 2008; Ott de Bruin L.M. et al., 2018), показывая корреляцию между средним уровнем активности мутантных белков и клиническим фенотипом (Lee Y.N, et al., 2014; Tirosh I. et al., 2019.). Считать данную корреляцию абсолютной пока что нельзя, учитывая другие генетические факторы и факторы окружающей среды, также способные изменять клинические и иммунологические фенотипы, но стоит обратить внимание на преобладающие рецидивирующие синусно-пульмональные инфекции у больных с нарушением функции *RAG* (Chou J. et al., 2012).

Основываясь на проведенных исследованиях, можно сказать, что определение уровней TREC и KREC при поступлении имеет прогностическое значение у инфекционных больных. Пациенты со сниженным уровнем TREC/KREC, выявленным при поступлении, будут нуждаться в индивидуальной поддержке для предотвращения серьезных последствий заболевания и повышения вероятности благоприятного исхода.

ЗАКЛЮЧЕНИЕ

Практически все подходы по выявлению патологий по своей природе ограничены генетической и фенотипической гетерогенностью и необходимостью множественных корректировок тестирования. Также усугубляет данную проблему недостаточность знаний о механизмах реализации генетической информации до фенотипического уровня. Без выявления связи причин со следствиями цельный путь между вариантом-кандидатом и клиническим фенотипом остается неопределенным, каким бы статистически вероятным он не был. В медицине для установления причинно-следственной связи между генотипом и фенотипом требуется тщательная проверка механизмов на молекулярном, клеточном, тканевом и целостном уровнях организма. В данной работе с использованием различных генетических подходов было показано влияние генетики человека на течение и исход инфекционного заболевания. Особое внимание было сосредоточено на омнигенной гипотезе, благодаря которой впервые на экспериментальном уровне удалось подтвердить влияние совокупного эффекта бремени редких генетических вариантов при тяжелом течении COVID-19. Также омнигенная гипотеза была расширена положением о суммарном влиянии редких потенциально патогенных вариантов в генах ПИД и в их ближайших функциональных генах-партнерах. Корреляция числа редких генетических вариантов на уровне всего экзома с количественными молекулярными данными показала, что при более тяжелом течении инфекции SARS-CoV-2 наблюдалось больше редких вариантов при меньшем количестве косвенного маркера иммунного статуса пациентов – TREC. Это указывает на то, что груз генетических повреждений в генах иммунной системы является фактором риска жизнеугрожающих состояний. Отдельно была показана диагностическая ценность молекулярного метода, основанного на T- и B-клеточной рекомбинации, для выявления пациентов с тяжелым течением бактериальной пневмонии.

Таким образом, комплексные и разносторонние генетические исследования могут помочь в нахождении причин, лежащих в основе тяжелого течения иммуноопосредованных заболеваний. В долгосрочной перспективе, полное понимание генетического патогенеза инфекционных заболеваний - лучший способ борьбы с этими болезнями.

ВЫВОДЫ

1. На основании анализа данных пациентов с тяжелой и легкой/умеренной формой течения COVID-19, впервые с помощью молекулярно-генетических и биоинформатических подходов показано влияние совокупного числа редких потенциально патогенных генетических вариантов на уровне всего экзома и в различных наборах генов, включая гены ПИД.

2. Охарактеризован патогенный генетический ландшафт генов, ассоциированных с ПИД, с помощью «генотип-фенотип» баз данных ClinVar и HGMD[®] и ресурса dbSNP. Показано, что подавляющее число редких потенциально патогенных вариантов в генах ПИД у пациентов с COVID-19 отсутствует в объединенной базе данных ClinVar/HGMD.

3. Показан более высокий уровень экспрессии генов ПИД по отношению к другим генам в субпопуляциях клеток крови человека.

4. На основе анализа результатов полноэкзомного секвенирования молодых пациентов с COVID-19, гены ПИД и их функциональные гены-партнеры были определены в качестве генов, ассоциированных с развитием тяжелой формы течения COVID-19.

5. С помощью тестирования косвенных маркеров иммунного ответа, основанного на Т- и В-клеточной рекомбинации (TREC и KREC), было показано, что иммунный статус молодых пациентов с COVID-19 был снижен при тяжелом течении заболевания. Также сниженный количественный уровень TREC коррелировал с большим числом редких

потенциально патогенных вариантов на уровне всего экзома и в генах ПИД у пациентов с тяжелой и легкой/умеренной формами течения COVID-19.

6. Показан диагностический потенциал определения уровней TREC и KREC при COVID-19 и бактериальной пневмонии с различными сопутствующими состояниями.

БЛАГОДАРНОСТИ

Автор выражает благодарность ведущему научному сотруднику ИОГена, д.б.н. Сальниковой Л.Е., под руководством которой выполнена данная диссертационная работа. Автор благодарен лаборатории молекулярной иммунологии НМИЦ ДГОИ им. Д. Рогачева за предоставление возможности проводить молекулярно-генетическое исследование, а также лаборатории молекулярной биологии за предоставление возможности к доступу профессиональной версии базы HGMD®. Автор благодарен ФНКЦ реаниматологии и реабилитологии за предоставление ценных биологических образцов пациентов.

СПИСОК ЛИТЕРАТУРЫ

1. Арипова Т. У., Шамсутдинова М. И., Холбоева Г. Б. Цитокины как предикторы тяжести заболевания у пациентов с COVID-19 // Евразийский журнал медицинских и естественных наук. 2023. Т. 3. №. 6. С. 86-95.
2. Васичкина Е. С., Костарева А. А. Редкие и неизвестные заболевания–современный тренд медицины // Российский журнал персонализированной медицины. 2022. Т. 2. №. 2. С. 72-83.
3. Кузьменко Н. Б., Щербина А. Ю. Классификация первичных иммунодефицитов как отражение современных представлений об их патогенезе и терапевтических подходах // Российский журнал детской гематологии и онкологии. 2017. №. 3. С. 51-57.
4. Поляков Д. В., Кухаренок А. Д., Мальцева И. О. Первичные иммунодефицитные состояния: недооцененная проблема // Интегративные тенденции в медицине и образовании. 2020. Т. 1. С. 70-76.
5. Щербина А. Ю. Первичные иммунодефициты-реалии XXI века // Вопросы гематологии/онкологии и иммунопатологии в педиатрии. 2016. Т. 15. №. 1. С. 8-9.
6. Abolhassani H. [и др.]. Global systematic review of primary immunodeficiency registries // Expert Rev Clin Immunol. 2020. № 16(7). С. 717–732.
7. Altschuler S.J., Wu L.F. Cellular heterogeneity: Do differences make a difference? // Cell. 2010. № 141(4). С. 559–63.
8. Alyousfi D., Baralle D., Collins A. Gene-specific metrics to facilitate identification of disease genes for molecular diagnosis in patient genomes: a systematic review // Brief Funct Genomics. 2019. № 18(1). С. 23–29.
9. Amberger J.S. [и др.]. OMIM.Org: online Mendelian inheritance in man (OMIM[®]), an online catalog of human genes and genetic disorder // Nucleic Acids Res. 2015. № 43(Database issue). С. D789–798.

10. Angus D. C., Poll T. van der. Severe sepsis and septic shock // *N. Engl. J. Med.* 2013. № 369. С. 840.
11. ARDS Definition Task Force [и др.]. Acute respiratory distress syndrome: The Berlin Definition // *JAMA.* 2012. № 307. С. 2526–2533.
12. Bailey O. T. [и др.]. A murine virus (JHM) causing disseminated encephalomyelitis with extensive destruction of myelin II // *Pathology. J. Exp. Med.* 1949. № 90. С. 195-212
13. Bauer M. [и др.]. Mortality in sepsis and septic shock in Europe, North America and Australia between 2009 and 2019—Results from a systematic review and meta-analysis // *Crit. Care.* 2020. №24. С. 239.
14. Bermejo-Martin J.F. [и др.]. NEUMONAC group Lymphopenic Community Acquired Pneumonia (L-CAP), an Immunological Phenotype Associated with Higher Risk of Mortality // *EBioMedicine.* 2017. №24. С. 231–236.
15. Bousfiha A. [и др.]. Human inborn errors of immunity: 2019 update of the IUIS phenotypical classification // *J. Clin. Immunol.* 2020 №4. С. 66–81.
16. Bousfiha A. [и др.]. The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies // *J. Clin. Immunol.* 2018. № 38. С. 129–43.
17. Boyle E. A., Li Y. I., Pritchard J. K. An expanded view of complex traits: from polygenic to omnigenic // *Cell.* 2017. № 169. С. 1177–1186.
18. Breit T. M. [и др.]. Human T cell leukemias with continuous V(D)J recombinase activity for TCR-delta gene deletion // *J. Immunol.* 1997. №159. С. 4341–4349.
19. Brodin P. Immune determinants of COVID-19 disease presentation and severity // *Nat. Med.* 2021. №27. С. 28–33.
20. Buitrago-Garcia D. [и др.]. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: a living systematic review and meta-analysis // *PLoS Med.* 2020. №17. С. e1003346.
21. Butler-Laporte G. [и др.]. Exome-wide association study to identify rare variants influencing COVID-19 outcomes: Results from the Host Genetics Initiative // *PLoS Genet.* 2022. №18. С. e1010367.

22. Carvelli J. [и др.]. Association of COVID-19 inflammation with activation of the C5a–C5aR1 axis // *Nature*. 2020. № 588. С.146–150.
23. Casanova J.L., Abel L. Human genetics of infectious diseases: Unique insights into immunological redundancy // *Semin Immunol*. 2018. №36. С. 1-12.
24. Casanova J.L., Abel L. Primary immunodeficiencies: a field in its infancy // *Science*. 2007. №317. С. 617–619.
25. Casanova J.L., Su H.C.; COVID Human Genetic Effort. A Global Effort to Define the Human Genetics of Protective Immunity to SARS-CoV-2 Infection // *Cell*. 2020. № 181(6). С. 1194-1199.
26. Casanova J.L., Abel L. Lethal Infectious Diseases as Inborn Errors of Immunity: Toward a Synthesis of the Germ and Genetic Theories // *Annual Review of Pathology: Mechanisms of Disease*. 2021. №16(1). С. 23-50.
27. Castagnoli R. [и др.]. Hematopoietic stem cell transplantation in primary immunodeficiency diseases: current status and future perspectives // *Front Pediatr*. 2019. № 7. С. 295.
28. Castellana S., Mazza T. Congruency in the prediction of pathogenic missense mutations: state-of-the-art web-based tools // *Brief Bioinform*. 2013. №14. С. 448–459.
29. Chamary J.V., Hurst L.D. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals // *Genome Biol*. 2005. № 6. С. R75.
30. Chamary J.V., Hurst L.D. Biased codon usage near intron-exon junctions: selection on splicing enhancers, splice-site recognition or something else? // *Trends Genet*. 2005. №21. С. 256–259.
31. Cheever F. S. [и др.]. A murine virus (JHM) causing disseminated encephalomyelitis with extensive destruction of myelin. I. Isolation and biological properties of the virus // *J. Exp. Med*. 1949. №90. С. 181-194.

32. Chen H. [и др.]. New insights on human essential genes based on integrated analysis and the construction of the HEGIAP web-based platform // *Brief Bioinform.* 2020. №21(4). С. 1397-1410.
33. Chen W., Coombes B.J., Larson N.B. Recent advances and challenges of rare variant association analysis in the biobank sequencing era // *Front Genet.* 2022. №13. С. 1014947.
34. Cheng Y. [и др.]. Abundance of intrinsic disorder in protein associated with cardiovascular disease // *Biochemistry.* 2006. №45. С. 10448–60.
35. Chiarini M. [и др.]. Newly produced T and B lymphocytes and T-cell receptor repertoire diversity are reduced in peripheral blood of fingolimod-treated multiple sclerosis patients // *Multiple Sclerosis Journal.* 2015. №21(6). С. 726–734.
36. Choi M. [и др.]. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing // *Proc. Natl Acad. Sci. USA.* 2009. №106. С. 19096–19101.
37. Chong J.X. [и др.]. Centers for Mendelian Genomics; Bamshad MJ. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities // *Am J Hum Genet.* 2015. №97(2). С. 199-215.
38. Chou J. [и др.]. A novel homozygous mutation in recombination activating gene 2 in 2 relatives with different clinical phenotypes: Omenn syndrome and hyper-IgM syndrome // *J Allergy Clin Immunol.* 2012. №130(6). С. 1414–1416.
39. Cillóniz C. [и др.]. Impact of age and comorbidity on cause and outcome in community-acquired pneumonia // *Chest.* 2013. №144. С. 999–1007.
40. Cingolani P. [и др.]. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3 // *Fly (Austin).* 2012. №6. С. 80–92. (b)
41. Cingolani P. [и др.]. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift // *Front. Genet.* 2012. №3. С. 35. (a)

42. Claussnitzer M. [и др.]. A brief history of human disease genetics. // *Nature*. 2020. №577. С. 179–189.
43. Cobos Jiménez V. [и др.]. T-Cell Activation Independently Associates With Immune Senescence in HIV-Infected Recipients of Long-term Antiretroviral Treatment // *J Infect Dis*. 2016. №214(2). С. 216-25.
44. Collins A. The genomic and functional characteristics of disease genes. // *Brief Bioinform*. 2015. №16(1). С. 16–23.
45. Conley M. E. [и др.]. Primary B cell immunodeficiencies: comparisons and contrasts // *Annu. Rev. Immunol*. 2009. №27. С. 199–227.
46. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19 // *Nature*. 2021. №600(7889). С. 472-477.
47. Cuvelier P. [и др.]. Protective reactive thymus hyperplasia in COVID-19 acute respiratory distress syndrome // *Crit. Care*. 2021. №25. С. 4.
48. D'Alessio F.R. [и др.]. CD4+CD25+Foxp3+ Tregs resolve experimental lung injury in mice and are present in humans with acute lung injury // *J Clin Invest*. 2009. №119(10). С. 2898-2913.
49. Daitaro K. [и др.]. Decreased numbers of signal-joint T cell receptor excision circle-containing CD4+ and CD8+ cells in systemic lupus erythematosus patients // *Mod Rheumatol*. 2007. №17(4). С. 296–300.
50. Dasouki M. [и др.]. TREC and KREC profiling as a representative of thymus and bone marrow output in patients with various inborn errors of immunity // *Clin Exp Immunol*. 2020. №202(1). С. 60-71.
51. De Biasi S. [и др.]. Marked T cell activation, senescence, exhaustion and skewing towards TH17 in patients with COVID-19 pneumonia // *Nat. Commun*. 2020. №11. С. 3434.
52. Delavari S. [и др.]. Impact of SARS-CoV-2 pandemic on patients with primary immunodeficiency // *J. Clin. Immunol*. 2021. №41. С. 345–355.
53. DeLong E.R., DeLong D.M., Clarke-Pearson D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach // *Biometrics*. 1988. №44. С. 837–845.

54. Deng Y. [и др.]. Clinical characteristics of fatal and recovered cases of coronavirus disease 2019 (COVID-19) in Wuhan, China: a retrospective study // *Chin Med J.* 2020. №133(11). С. 1261-1267.
55. Dennis A. [и др.]. Multi-organ impairment in low-risk individuals with long COVID // Preprint at medRxiv. 2020.
56. Di Pasquale M.F. [и др.]. Prevalence and Etiology of Community-acquired Pneumonia in Immunocompromised Patients // *Clin Infect Dis.* 2019. №68(9). С. 1482-1493.
57. Di Rienzo A. Population genetics models of common diseases // *Curr Opin Genet Dev.* 2006. №16 С. 630.
58. Dickinson M. E. [и др.]. High-throughput discovery of novel developmental phenotypes // *Nature.* 2016. №537. С. 508–514.
59. Dong C. [и др.]. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies // *Hum. Mol. Genet.* 2015. №24. С. 125–2137.
60. Dorfman R. [и др.]. Do common in silico tools predict the clinical consequences of amino-acid substitutions in the CFTR gene? // *Clin Genet.* 2010. №77. С. 464–473.
61. Douek D. C. [и др.]. Changes in thymic function with age and during the treatment of HIV infection // *Nature.* 1998. №396. С. 690–695.
62. Drummond D.A., [и др.]. Why highly expressed proteins evolve slowly // *Proc. Natl. Acad. Sci. USA.* 2005. №102. С. 14338–14343.
63. Drzymalla E. [и др.]. COVID-19-related health outcomes in people with primary immunodeficiency: A systematic review // *Clin. Immunol.* 2022. №243. С. 109097.
64. Dworacki G. [и др.]. Thymic emigration patterns in patients with type 2 diabetes treated with metformin // *Immunology.* 2015. №146(3). С. 456-69.
65. Eisenberg E., Levanon E.Y. Human housekeeping genes are compact // *Trends Genet.* 2003. №19(7). С. 362–5.

66. Erickson J.J. [и др.]. Acute Viral Respiratory Infection Rapidly Induces a CD8+ T Cell Exhaustion-like Phenotype // *J. Immunol.* 2015. № 95. С. 4319–4330.
67. Evans L. [и др.]. Executive Summary: Surviving Sepsis Campaign: International Guidelines for the Management of Sepsis and Septic Shock 2021 // *Crit. Care Med.* 2021. №49. С. 1974–1982.
68. Farhan S. M. [и др.]. Exome sequencing in amyotrophic lateral sclerosis implicates a novel gene, DNAJC7, encoding a heat-shock protein // *Nat. Neurosci.* 2019. №22. С. 1966–1974.
69. Feldman I., Rzhetsky A., Vitkup D. Network properties of genes harboring inherited disease mutations // *Proc Natl Acad Sci USA.* 2008. №105. С. 4323–4328.
70. Feng Y. C. A. [и др.]. Ultra-Rare Genetic Variation in the Epilepsies: A whole-exome sequencing study of 17,606 individuals // *Am. J. Hum. Genet.* 2019. №105. С. 267–282.
71. Ferguson N.D. [и др.]. The Berlin definition of ARDS: An expanded rationale, justification, and supplementary material // *Intensive Care Med.* 2012. №38. С. 1573–1582.
72. Fessler J. [и др.]. Premature senescence of T-cell subsets in axial spondyloarthritis // *Ann Rheum Dis.* 2016. №75(4). С. 748-54.
73. Fine M.J. [и др.]. A prediction rule to identify low-risk patients with community-acquired pneumonia // *N. Engl. J. Med.* 1997. №336. С. 243–250.
74. Franco J. M. [и др.]. T-cell repopulation and thymic volume in HIV-1–infected adult patients after highly active antiretroviral therapy // *Blood.* 2002. №99(10). С. 3702-3706.
75. Franke B. What’s in a name: The “omnigenic” model as a point of departure for polygenic psychiatric disorders // *J. Psychiatry Brain Sci.* 2017. №2. С. 5.

76. Frousius K. [и др.]. Predicting the functional consequences of non-synonymous DNA sequence variants—evaluation of bioinformatics tools and development of a consensus strategy // *Genomics*. 2013. №102. С. 223–228.
77. Furman C.D., Rayner A.V., Tobin E.P. Pneumonia in older residents of long-term care facilities // *Am Fam Physician*. 2004. №70. С. 1495–500.
78. Gaballa A. [и др.]. T Cell Receptor Excision Circle (TREC) Monitoring after Allogeneic Stem Cell Transplantation; a Predictive Marker for Complications and Clinical Outcome // *Int J Mol Sci*. 2016. №17(10). С. 1705.
79. Gandhi T.K. [и др.]. Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets // *Nat Genet*. 2006. №38. С. 285–293.
80. Ganna A. [и др.]. Quantifying the impact of rare and ultra-rare coding variation across the phenotypic spectrum // *Am. J. Hum. Genet*. 2018. №102. С. 1204–1211.
81. GARD. FAQs about rare diseases. 2017. [Электронный ресурс]. <https://rarediseases.info.nih.gov/diseases/pages/31/faqs-about-rarediseases>. Accessed 20 Feb 2020.
82. Gardner E. J. [и др.]. Reduced reproductive success is associated with selective constraint on human genes // *Nature*. 2022. №603. С. 858–863.
83. GBD 2016 Lower Respiratory Infections Collaborators. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016 // *Lancet Infect. Dis*. 2018. №18. С. 1191–1210.
84. Giamarellos-Bourboulis E.J. [и др.]. Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure // *Cell Host Microbe*. 2020. №27. С. 992–1000.
85. Goggins J.A., Kurtz J.R., McLachlan J.B. Control of Persistent Salmonella Infection Relies on Constant Thymic Output Despite Increased Peripheral Antigen-Specific T Cell Immunity // *Pathogens*. 2020. №9. С. 605.

86. Gout J.F. [и др.]. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution // PLoS Genet. 2010. №6. С. e1000944.
87. Goymer P. Synonymous mutations break their silence // Nature Reviews Genetics. 2007. №8(2). с. 92–92.
88. Gray P.E., Bartlett A.W., Tangye S.G. Severe COVID-19 represents an undiagnosed primary immunodeficiency in a high proportion of infected individuals // Clin Transl Immunology. 2022. №11(4). С. e1365.
89. Grimm D. G. [и др.]. The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity // Hum. Mutat. 2015. №36. С. 513–523.
90. Haendel M. [и др.]. How many rare diseases are there? // Nat Rev Drug Discov. 2020. №19(2). С. 77-78.
91. Hancock A.M. [и др.]. Adaptations to climate in candidate genes for common metabolic disorders // PLoS Genet. 2008. №4. С. e32.
92. Hazenberg M. D. [и др.]. Increased cell division but not thymic dysfunction rapidly affects the T-cell receptor excision circle content of the naive T cell population in HIV-1 infection // Nat. Med. 2000. №6. С. 1036–1042.
93. Helgason H. [и др.]. A rare nonsynonymous sequence variant in C3 is associated with high risk of age-related macular degeneration // Nat Genet. 2013. №45. С. 1371–1374.
94. Hindorff L. A. [и др.]. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits // Proc. Natl Acad. Sci. USA. 2009. №106. С. 9362–9367.
95. Hirschhorn J.N., Daly M.J. Genome-wide association studies for common diseases and complex traits // Nat. Rev. Genet. 2005. №6. С. 95–108.
96. Hotchkiss R. S. [и др.]. Caspase inhibitors improve survival in sepsis: a critical role of the lymphocyte // Nat. Immunol. 2000. №1. С. 496.

97. Hotchkiss R.S. [и др.]. Sepsis-induced apoptosis causes progressive profound depletion of B and CD4+ T lymphocytes in humans // *J Immunol*. 2001. №166. С. 6952–6963.
98. Hu Y. [и др.]. The Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Inhibits Type I Interferon Production by Interfering with TRIM25-Mediated RIG-I Ubiquitination // *J. Virol*. 2017. №91. С. e02143-16
99. Huang C. [и др.]. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China // *Lancet*. 2020. №395. С. 497–506.
100. Huang N. [и др.]. Characterising and predicting haploinsufficiency in the human genome // *PLoS Genet*. 2010. №6(10). С. 1–11.
101. Hug A. [и др.]. Thymic Export Function and T Cell Homeostasis in Patients with Relapsing Remitting Multiple Sclerosis // *J Immunol*. 2003. № 171 (1). С. 432–437.
102. Iakoucheva L.M. [и др.]. Intrinsic disorder in cell-signaling and cancer-associated proteins // *J Mol Biol*. 2002. №323. С. 573–584.
103. IJspeert H. [и др.]. Artemis splice defects cause atypical SCID and can be restored in vitro by an antisense oligonucleotide // *Genes Immun*. 2011. №12(6). С. 434-444.
104. Ioannidis N. M. [и др.]. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants // *Am. J. Hum. Genet*. 2016. №99. С. 877–885.
105. Iskakova S. [и др.]. The differences of trec (t-cell receptor excision circles) concentration in type 2 diabetic patients // *Georgian Med News*. 2015. №244-245. С. 29-36.
106. Jain A., Tuteja G. TissueEnrich: Tissue-specific gene enrichment analysis // *Bioinformatics*. 2019. №35. С. 1966–1967.
107. Jamuar S. S., Tan E. C. Clinical application of next-generation sequencing for Mendelian diseases // *Hum. Genom*. 2015. №9. С. 10.

108. Johnston J. J. [и др.]. Individualized iterative phenotyping for genome-wide analysis of loss-of-function mutations // *Am. J. Hum. Genet.* 2015. №96. С. 913–925.
109. Jonsson T. [и др.]. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline // *Nature.* 2012. №488. С. 96–99.
110. Jonsson T. [и др.]. Variant of TREM2 associated with the risk of Alzheimer's disease // *N Engl J Med.* 2013. №368. С. 107–116.
111. Just H.L. [и др.]. T-cell receptor excision circles (TREC) in CD4+ and CD8+ T-cell subpopulations in atopic dermatitis and psoriasis show major differences in the emission of recent thymic emigrants // *Acta Derm Venereol.* 2008. №88(6). С. 566-572.
112. Kahan S.M., Wherry E.J., Zajac A.J. T cell exhaustion during persistent viral infections // *Virology.* 2015. №479-480, с. 180–193.
113. Karchin R. [и др.]. LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources // *Bioinformatics.* 2005. №21(12). С. 2814–2820.
114. Karczewski K.J. [и др.]. The mutational constraint spectrum quantified from variation in 141,456 humans // *Nature.* 2020. №581(7809). С. 434–443.
115. Karnaushkina M.A. [и др.]. Genetic factors contributing to a severe course of pneumonia: a systematic review // *Bulletin of Siberian Medicine.* 2022. №21(4). С. 160–169.
116. Khadzhieva M.B. [и др.]. TREC/KREC Levels in Young COVID-19 Patients // *Diagnostics.* 2021. №11. С. 1486.
117. Khare K., Pandey R. Cellular heterogeneity in disease severity and clinical outcome: Granular understanding of immune response is key // *Front. Immunol., Sec. Viral Immunology.* 2022. №22(13). С. 973070.
118. Khurana E. [и др.]. Interpretation of genomic variants using a unified biological network approach // *PLoS Comput Biol.* 2013. №9(3). С. e1002886.
119. Kimchi-Sarfaty C. [и др.]. A 'silent' polymorphism in the MDR1 gene changes substrate specificity // *Science.* 2006. 315(5811). С. 525-528.

120. Kinsella R.J. [и др.]. Ensembl BioMarts: a hub for data retrieval across taxonomic space // Database (oxford). 2011. C. bar030.
121. Klein S. L., Jedlicka A., Pekosz A. The Xs and Y of immune responses to viral vaccines // Lancet Infect. Dis. 2010. №10, C. 338–349.
122. Knaus W.A. [и др.]. APACHE II: A severity of disease classification system // Crit. Care Med. 1985. №13. C. 818–829.
123. Kolobkov D.S. [и др.]. Genes and Diseases: Insights from Transcriptomics Studies // Genes (Basel). 2022. №13(7). C. 1168.
124. Kosmicki J. A. [и др.]. Pan-ancestry exome-wide association analyses of COVID-19 outcomes in 586,157 individuals // Am. J. Hum. Genet. 2021. №108. C. 1350–1355.
125. Krasznai Z. Ion channels in T cells: from molecular pharmacology to therapy // Arch Immunol Ther Exp (Warsz). 2005. №53. C. 127–135.
126. Kwok A. J., Mentzer A., Knight J. C. Host genetics and infectious disease: New tools, insights and translational opportunities // Nat. Rev. Genet. 2021. №22. C. 137–153.
127. Kwok A. J., Mentzer A., Knight J. C. Host genetics and infectious disease: new tools, insights and translational opportunities // Nat. Rev. Genet. 2021. №22. C. 137–153.
128. Kwok J.S.Y. [и др.]. Establishing Simultaneous T Cell Receptor Excision Circles (TREC) and K-Deleting Recombination Excision Circles (KREC) Quantification Assays and Laboratory Reference Intervals in Healthy Individuals of Different Age Groups in Hong Kong // Front Immunol. 2020. №11. C. 1411.
129. Lage K. [и др.]. A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes // Proc. Natl. Acad. Sci. USA. 2008. №105. C. 20870–20875.
130. Landrum M.J. [и др.]. ClinVar: improvements to accessing data // Nucleic Acids Res. 2020. №48(D1). C. D835–D844.

131. Landrum M.J. [и др.]. ClinVar: public archive of interpretations of clinically relevant variants // *Nucleic Acids Res.* 2016. №44(D1). С. D862–868.
132. Lawrence T. [и др.]. Autosomal-dominant primary immunodeficiencies // *Curr Opin Hematol.* 2005. №12(1). С. 22-30.
133. Lee Y.N. [и др.]. A systematic analysis of recombination activity and genotype-phenotype correlation in human recombination-activating gene 1 deficiency // *J Allergy Clin Immunol.* 2014. №133(4). С. 1099–108.
134. Lek M. [и др.]. Analysis of protein-coding genetic variation in 60,706 humans // *Nature* 2017. №536(7616). С. 285–291.
135. Lester S.N., Li K. Toll-like receptors in antiviral innate immunity // *J. Mol. Biol.* 2014. №426. С. 1246–1264.
136. Li S.W. [и др.]. SARS Coronavirus Papain-Like Protease Inhibits the TLR7 Signaling Pathway through Removing Lys63-Linked Polyubiquitination of TRAF3 and TRAF6 // *Int. J. Mol. Sci.* 2016. №17. С. 678.
137. Liu X., Li Y. I., Pritchard J. K. Trans effects on gene expression can drive omnigenic inheritance // *Cell.* 2019. №177. С. 1022–1034.
138. Livak F., Schatz D. G. T-cell receptor alpha locus V(D)J recombination by-products are abundant in thymocytes and mature T cells // *Mol. Cell. Biol.* 1996. №16, С. 609–618
139. Lowe A.L. [и др.]. Exploring the genetic susceptibility to COVID-19 // *Nature.* 2021. №597(7876). С. 493-495.
140. Máca J. [и др.]. Past and Present ARDS Mortality Rates: A Systematic Review // *Respir. Care.* 2017. №62. С. 113–122.
141. MacArthur D. G. [и др.]. A systematic survey of loss-of-function variants in human protein-coding genes // *Science.* 2012. №335. С. 823–828.
142. Mandell L. A., Niederman M. S. Aspiration pneumonia // *N. Engl. J. Med.* 2019. №380. С. 651–663
143. Mangino M. [и др.]. Innate and adaptive immune traits are differentially affected by genetic and environmental factors // *Nat Commun.* 2017. №8. С. 13850.

144. Martelotto L. G. [и др.]. Benchmarking mutation effect prediction algorithms using functionally validated cancer-related missense mutations // *Genome Biol.* 2014. №15. С. 484.
145. Masica D.L. [и др.]. Predicting survival in head and neck squamous cell carcinoma from TP53 mutation // *Hum Genet.* 2015. №134. С. 497–507.
146. Mason W.T. [и др.]. Control of secretion in anterior pituitary cells—linking ion channels, messengers and exocytosis // *J Exp Biol.* 1988. №139. С 287–316.
147. Mathieson I. The omnigenic model and polygenic prediction of complex traits // *Am. J. Hum. Genet.* 2021. №108. С. 1558–1563.
148. McCarthy D.J. [и др.]. Choice of transcripts and software has a large effect on variant annotation // *Genome Med.* 2014. №6(3). С. 26.
149. McCusker C., Upton J., Warrington R. Primary immunodeficiency // *Allergy asthma. Clin. Immunol.* 2018. №14(Suppl. 2). С. 61.
150. McLaren W. [и др.]. The Ensembl Variant Effect Predictor // *Genome Biol.* 2016. №17(1). С. 122.
151. Mørup S.B. [и др.]. Added Value of Reanalysis of Whole Exome- and Whole Genome Sequencing Data From Patients Suspected of Primary Immune Deficiency Using an Extended Gene Panel and Structural Variation Calling // *Front Immunol.* 2022. №13. С. 906328.
152. Mueller Y.M. [и др.]. Stratification of hospitalized COVID-19 patients into clinical severity progression groups by immuno-phenotyping and machine learning // *Nat. Commun.* 2022. №13. С. 915.
153. Musunuru K. [и др.]. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus // *Nature.* 2010. №466. С. 714–719
154. Nakajima T. [и др.]. T cell pathways involving CTLA4 contribute to a model of acute lung injury // *J Immunol.* 2010. №184(10). С. 5835-41.
155. Notarangelo L. D. Primary immunodeficiencies // *J. Allergy Clin. Immunol.* 2010. №125. С. S182–S194.

156. Notarangelo L.D. [и др.]. Human inborn errors of immunity: an expanding universe // *Sci Immunol*. 2020. №5. С. eabb1662.
157. Oberholzer C. [и др.]. Targeted adenovirus-induced expression of IL-10 decreases thymic apoptosis and improves survival in murine sepsis // *Proc. Natl. Acad. Sci. USA*. 2001. №98. С. 11503.
158. Ott de Bruin L.M. [и др.]. Hypomorphic Rag1 mutations alter the pre-immune repertoire at early stages of lymphoid development // *Blood*. 2018. №132(3). С. 281-292.
159. Ou X.L. [и др.]. Predicting human age with bloodstains by sjTREC quantification // *PLoS One*. 2012. №7(8). С. e42412.
160. Paghera S. [и др.]. Age-Related Lymphocyte Output During Disease-Modifying Therapies for Multiple Sclerosis // *Drugs Aging*. 2020. №37(10). С. 739–746.
161. Panchal N., Ghosh S., Booth C. T cell gene therapy to treat immunodeficiency // *Br J Haematol*. 2021. №192(3). С. 433-443.
162. Panda S. K., Colonna M. Innate lymphoid cells in mucosal immunity // *Front. Immunol*. 2019. №10. С. 861.
163. Parmley J.L., Chamary J.V., Hurst L.D. Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers // *Biol Evol*. 2006. №23. С. 301–309.
164. Parry J. WHO investigates China's fall in SARS cases // *Br. Med. J*. 2003. №326. С. 1285.
165. Peischl S. [и др.]. On the accumulation of deleterious mutations during range expansions. // *Mol Ecol*. 2013. №22. С. 5972–82.
166. Petrovski S. [и др.]. Genic intolerance to functional variation and the interpretation of personal genomes // *PLoS Genet*. 2013. №9(8). С. e1003709.
167. Picard C., Casanova J.L., Puel A. Infectious diseases in patients with IRAK-4, MyD88, NEMO, or I κ B α deficiency // *Clin. Microbiol. Rev*. 2011. №24. С. 490–97.

168. Piñero J. [и др.]. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants // *Nucl Acids Res.* 2016. №45(D1). С. D833–839.
169. Plenge R. M. Molecular underpinnings of severe coronavirus disease // *J. Am. Med. Assoc.* 2020. №324. С. 638–639.
170. Porcu E. [и др.]. Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome // *Nat. Commun.* 2021. №12. С. 5647.
171. Preston J. A. [и др.]. Alveolar macrophage apoptosis-associated bacterial killing helps prevent murine pneumonia // *Am. J. Respir. Crit. Care Med.* 2019. №200. С. 84–97.
172. Price A.L., Spencer C.C., Donnelly P. Progress and promise in understanding the genetic basis of common diseases // *Proc Biol Sci.* 2015. №282(1821). С. 20151684.
173. Pritchard J.K. Are rare variants responsible for susceptibility to complex diseases? // *Am J Hum Genet.* 2001. №69. С. 124–137.
174. Pullabhatla V. [и др.]. De novo mutations implicate novel genes in systemic lupus erythematosus // *Hum. Mol. Genet.* 2018. №27. С. 421–429.
175. Quach H. [и др.]. Epigenomic epidemiology of common diseases: prospects and challenges // *Hum Mol Genet.* 2020. №29(R2). С. R203-R210.
176. Quintana-Murci L. Understanding rare and common diseases in the context of human evolution // *Genome Biol.* 2016. №17(1). С. 225.
177. Quinton L. J. [и др.]. Functions and regulation of NF- κ B RelA during pneumococcal pneumonia // *J. Immunol.* 2007. №178. С. 1896–1903.
178. Rankin L.C., Artis D. Beyond host defense: Emerging functions of the immune system in regulating complex tissue physiology // *Cell.* 2018. №173. С. 554–567.
179. Rappaport N. [и др.]. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search // *Nucl Acids Res.* 2016. №45(D1). С. D877–887.

180. Rausell A. [и др.]. Common homozygosity for predicted loss-of-function variants reveals both redundant and advantageous effects of dispensable human genes // *Proc Natl Acad Sci U S A*. 2020. №117(24). С. 13626-13636.
181. Redmond M.T., Scherzer R., Prince B.T. Novel Genetic Discoveries in Primary Immunodeficiency Disorders // *Clin Rev Allergy Immunol*. 2022. №63(1). С. 55-74.
182. Reich D.E., Lander E.S. On the allelic spectrum of human disease // *Trends Genet*. 2001. №17. С. 502-510.
183. Risch N., Merikangas K. The future of genetic studies of complex human diseases // *Science*. 1996. №273(5281). С. 1516-1517.
184. Roth G. [и др.]. Susceptibility to programmed cell death in T-lymphocytes from septic patients: a mechanism for lymphopenia and Th2 predominance // *Biochem. Biophys. Res. Commun*. 2003. №308. С. 840–846.
185. Salnikova L. E. [и др.]. Cytokines mapping for tissue-specific expression, eQTLs and GWAS traits // *Sci. Rep*. 2020. №10. С. 14740.
186. Salnikova L.E. [и др.]. An overview of germline variations in genes of primary immunodeficiencies through integrative analysis of ClinVar, HGMD[®] and dbSNP databases // *Hum. Genet*. 2021. №140. С. 1379–1393.
187. Salzberger B. [и др.]. Epidemiology of SARS-CoV-2 // *Infection*. 2021. №49. С. 233–239.
188. Samocha K.E. [и др.]. A framework for the interpretation of de novo mutation in human disease // *Nat. Genet*. 2014. №46. С. 944–950.
189. Sánchez-Montalvá A. [и др.]. Exposing and Overcoming Limitations of Clinical Laboratory Tests in COVID-19 by Adding Immunological Parameters; A Retrospective Cohort Analysis and Pilot Study // *Front.Immunol*. 2022. №13. С. 902837.
190. Sattler S. The role of the immune system beyond the fight against infection // *Adv. Exp. Med. Biol*. 2017. №1003. С. 3–14.
191. Scelsi M. A. [и др.]. Alzheimer’s Disease Neuroimaging Initiative (ADNI) and the Alzheimer’s Disease Sequencing Project (ADSP) Network

propagation of rare variants in Alzheimer's disease reveals tissue-specific hub genes and communities // *PLoS Comput. Biol.* 2021. №7. С. e1008517.

192. Schuetz C. [и др.]. An immunodeficiency disease with RAG mutations and granulomas // *N. Engl. J. Med.* 2008. №358. С. 2030–2038.

193. Schwartz J.L. [и др.]. Single-site mutations in the conserved alternating-arginine region affect ionic channels formed by CryIAa, a *Bacillus thuringiensis* toxin // *Appl Environ Microbiol.* 1997. №63(10). С. 3978–3984.

194. Seidel M.G. [и др.]. The European Society for Immunodeficiencies (ESID) Registry Working Definitions for the clinical diagnosis of inborn errors of immunity // *J Allergy Clin Immunol Pract.* 2019. №7(6). С. 1763–1770.

195. Seidman J.G., Seidman C. Transcription factor haploinsufficiency: when half a loaf is not enough // *J Clin Invest.* 2002. №109. С. 451–455.

196. Seleman M. [и др.]. Uses of Next-Generation Sequencing Technologies for the Diagnosis of Primary Immunodeficiencies // *Front Immunol.* 2017. №24(8). С. 847.

197. Sempowski G.D. [и др.]. Effect of Thymectomy on Human Peripheral Blood T Cell Pools in Myasthenia Gravis // *J Immunol.* 2001. №166(4). С. 2808–2817.

198. SeyedAlinaghi S. [и др.]. COVID-19 mortality in patients with immunodeficiency and its predictors: a systematic review // *Eur J Med Res.* 2022. №27(1). С. 195.

199. Shelton J. F. [и др.]. Trans-ethnic analysis reveals genetic and non-genetic associations with COVID-19 susceptibility and severity 2020 // *Nat Genet.* 2021. №53(6). С. 801-808

200. Shihab H.A. [и др.]. HIPred: an integrative approach to predicting haploinsufficient genes // *Bioinformatics.* 2017. №33(12). С. 1751–1757.

201. Siu K.L. [и др.]. Severe acute respiratory syndrome coronavirus M protein inhibits type I interferon production by impeding the formation of TRAF3.TANK.TBK1/IKKepsilon complex // *J. Biol. Chem.* 2009. №284. С. 16202–16209.

202. Söderström A. [и др.]. T cell receptor excision circles are potential predictors of survival in adult allogeneic hematopoietic stem cell transplantation recipients with acute myeloid leukemia // *Front Immunol.* 2022. №13. С. 954716.
203. Sørensen T. I. [и др.]. Genetic and environmental influences on premature death in adult adoptees // *N. Engl. J. Med.* 1988. №318. С. 727–732.
204. Steinberg J. [и др.]. Haploinsufficiency predictions without study bias // *Nucleic Acids Res.* 2015. №43. С. e101.
205. Steinthorsdottir V. [и др.]. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes // *Nat Genet.* 2014. №46. С. 294–298.
206. Stenson P.D. [и др.]. The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution // *Curr Protoc Bioinformatics.* 2012. №39(1.13). С. 1.13.1–1.13.20.
207. Stenson P.D. [и др.]. The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting // *Hum Genet.* 2020. №139(10). С. 1197–1207.
208. Stone E. A., Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity // *Genome Res.* 2005. №15(7). С. 978–986
209. Strawa K. [и др.]. Increased concentration of T-cell receptor rearrangement excision circles (TREC) in peripheral blood in Graves' disease // *Clin Endocrinol.* 2014. №81. С. 769-774.
210. Stray-Pedersen A. [и др.]. Primary immunodeficiency diseases: Genomic approaches delineate heterogeneous Mendelian disorders // *J Allergy Clin Immunol.* 2017. №139(1). С. 232-245.
211. Styrkarsdottir U. [и др.]. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits // *Nature.* 2013. №497. С. 517–520.
212. Taliun D. [и др.]. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program // *Nature.* 2021. №590(7845). С. 290-299.

213. Talla A. [и др.]. Longitudinal immune dynamics of mild COVID-19 define signatures of recovery and persistence // bioRxiv [Preprint]. 2021.
214. Tang T.T. [и др.]. Impaired thymic export and apoptosis contribute to regulatory T-cell defects in patients with chronic heart failure // PLoS One. 2011. №6(9). С. e24272.
215. Tangye S.G. [и др.]. Human inborn errors of immunity: 2019 update on the classification from the International Union of Immunological Societies Expert Committee // J. Clin. Immunol. 2020. №40. С. 24–64
216. Tangye SG, [и др.]. Human Inborn Errors of Immunity: 2022 Update on the Classification from the International Union of Immunological Societies Expert Committee // J Clin Immunol. 2022. №42(7). С. 1473-1507.
217. Tennessen J.A. [и др.]. Evolution and functional impact of rare coding variation from deep sequencing of human exomes // Science. 2012. №337. С. 64–69.
218. The 1000 Genomes Project Consortium [и др.]. A global reference for human genetic variation // Nature. 2015. № 526 (7571). С. 68-74.
219. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing // Nature. 2010. №467 с. 1061–1073.
220. The COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic // Eu. J. Hum. Genet. 2020. №28. С. 715–718.
221. The ENCODE (ENCyclopedia Of DNA Elements) // Project.Science. 2004. №306. С. 636-640.
222. The Severe Covid-19 GWAS Group. Genomewide association study of severe COVID-19 with respiratory failure // N. Engl. J. Med. 2020. №383. С. 1522–1534.
223. Thorneloe K.S., Nelson M.T. Ion channels in smooth muscle: regulators of intracellular calcium and contractility // Can J Physiol Pharmacol. 2005. №83. с. 215–242.

224. Thusberg J., Olatubosun A., Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants // *Hum. Mutat.* 2011. №32. C. 358–368.
225. Thusberg J., Vihinen M. Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods // *Hum Mutat.* 2009. №30. C. 703–714.
226. Tirosh I. [и др.]. Recombination activity of human RAG2 mutations and correlation with the clinical phenotype // *J Allergy Clin Immunol.* 2019. №143(2). C. 726-735.
227. Torres A. [и др.]. Pneumonia // *Nat. Rev. Dis. Primers.* 2021. №7 с. 25
228. Touloumi G. [и др.]. Changes in T Cell Receptor Excision DNA Circle (TREC) Levels in HIV Type 1-Infected Subjects Pre- and Post-Highly Active Antiretroviral Therapy // *AIDS Research and Human Retroviruses.* 2004 №20(1). C. 47–54.
229. Tu Z., Wang L., Xu M. Further understanding human disease genes by comparing with housekeeping genes and other genes // *BMC Genomics.* 2006. №7(1). C. 31.
230. Tyrrel D. A. J. [и др.]. Coronavirus // *Nature.* 1968. №220. C. 650.
231. Uversky V.N. Intrinsic disorder in proteins associated with neurodegenerative diseases // *Front Biosci (Landmark Ed).* 2009. №14(14). C. 5188–5238.
232. Uversky V.N., Oldfield C.J., Dunker A.K. Intrinsically disordered proteins in human diseases: introducing the D2 concept // *Annu Rev Biophys.* 2008. №37. C. 215–246.
233. Uversky VN. Amyloidogenesis of natively unfolded proteins // *Curr Alzheimer Res.* 2008. №5. C. 260–287.
234. Valotti M. [и др.]. Long-lasting production of new T and B cells and T-cell repertoire diversity in patients with primary immunodeficiency who had undergone stem cell transplantation: a single-centre experience // *J Immunol Res.* 2014. C. 240453.

235. van Zelm M. C. [и др.]. Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression // *J. Immunol.* 2005. №175. С.5912–5922.
236. van Zelm M. C. [и др.]. Replication history of B lymphocytes reveals homeostatic proliferation and extensive antigen-induced B cell expansion // *J. Exp. Med.* 2007. №204. С. 645–655.
237. van Zelm M.C. [и др.]. PID comes full circle: applications of V(D)J recombination excision circles in research, diagnostics and newborn screening of primary immunodeficiency disorders // *Front Immunol.* 2011. №2. С. 12.
238. Veitia R.A., Birchler J.A. Dominance and gene dosage balance in health and disease: why levels matter! // *J. Pathol.* 2010. №220. С. 174–185.
239. Verschuren M. C. [и др.]. Preferential rearrangements of the T cell receptor-delta-deleting elements in human T cells // *J. Immunol.* 1997. №158. С. 1208–1216.
240. Vinogradov A.E. DNA helix: the importance of being GC-rich // *Nucleic Acids Res.* 2003. №31. С. 1838–1844.
241. Visscher P.M. [и др.]. Discovery and implications of polygenicity of common diseases // *Science.* 2021. №373(6562). С. 1468-1473.
242. Visscher P.M. [и др.]. Five years of GWAS discovery // *Am. J. Hum. Genet.* 2012. №90. С. 7–24.
243. Walters-Sen L. C. [и др.]. Variability in pathogenicity prediction programs: impact on clinical diagnostics. // *Mol. Genet. Genom. Med.* 2015. №3. с. 99–110.
244. Wang Q. [и др.]. Rare variant contribution to human disease in 281,104 UK Biobank exomes // *Nature.* 2021. №597(7877). С. 527-532.
245. Wang R. [и др.]. The role of C5a in acute lung injury induced by highly pathogenic viral infections // *Emerg. Microbes Infect.* 2015. №4. С. 1–7.
246. Wang Z., Moulton J. SNPs, protein structure, and disease // *Hum. Mutat.* 2001. С. 17263–17270.

247. Webb K. [и др.]. Sex and pubertal differences in the type 1 interferon pathway associate with both X chromosome number and serum sex hormone concentration // *Front Immunol.* 2019. №9, с. 3167.
248. Wong S.Y., Lu C.P., Roth D.B. A RAG1 mutation found in Omenn syndrome causes coding flank hypersensitivity: a novel mechanism for antigen receptor repertoire restriction // *J Immunol.* 2008. №181(6). С. 4124–30.
249. Wu F. [и др.]. A new coronavirus associated with human respiratory disease in China // *Nature.* 2020. №579. С. 265–269.
250. Xia X. Maximizing transcription efficiency causes codon usage bias. *Genetics.* 1996. №144. С. 1309–1320.
251. Xie T., Ding D. The relationship between synonymous codon usage and protein structure // *FEBS Lett.* 1998. №434(1-2). С. 93-96.
252. Xu H. [и др.]. Deep sequencing of 1320 genes reveals the landscape of protein-truncating variants and their contribution to psoriasis in 19,973 Chinese individuals // *Genome Res.* 2021. №31. С. 1150–1158.
253. Yao C. [и др.]. Cell-Type-Specific Immune Dysregulation in Severely Ill COVID-19 Patients // *Cell Rep.* 2021. №34. С. 108590.
254. Zeng Z., Bromberg Y. Predicting functional effects of synonymous variants: a systematic review and perspectives // *Front. Genet.* 2019. №10. С. 914.
255. Zhang Q. [и др.]. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19 // *Science.* 2020. №370. С. eabd4570.
256. Zhang S.Y. [и др.]. Human inborn errors of immunity to infection affecting cells other than leukocytes: from the immune system to the whole organism // *Curr. Opin. Immunol.* 2019. №59. С. 88–100.
257. Zhou F. [и др.]. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study // *Lancet.* 2020. №395. С. 1054–1062.
258. Zhou H. [и др.]. FAVOR: functional annotation of variants online resource and annotator for variation across the human genome // *Nucleic Acids Res.* 2023. №51(D1). С. D1300-D1311.

259. Zoghbi A. W. [и др.]. High-impact rare genetic variants in severe schizophrenia // Proc. Natl. Acad. Sci. U.S.A. 2021. №118. С. e2112560118.

260. Zubakov D. [и др.]. Estimating human age from T-cell DNA rearrangements // Curr. Biol. 2010. №20. С. R970–R971.

ПРИЛОЖЕНИЕ

Поскольку некоторые результаты могут зависеть от конкретной популяции, структура выборки из 86 пациентов с COVID-19 была сравнена с данными из 1000G. PCA основывался на пересечении вариантов между двумя наборами данных. Совместный анализ набора данных из 1000G был проведен с использованием PLINK2 (<https://doi.org/10.1186/s13742-015-0047-8>). Визуализация проводилась с использованием библиотеки Plotly на языке Python.

Принципиальный компонентный анализ (PCA) показал, что выборка из 86 пациентов с COVID-19, за исключением 7 пациентов, не являющихся европеоидами, соотносится с европейской популяцией. При повторном анализе аллельных частот, исключая 7 не европеоидных пациентов, все значимые результаты сохранились.

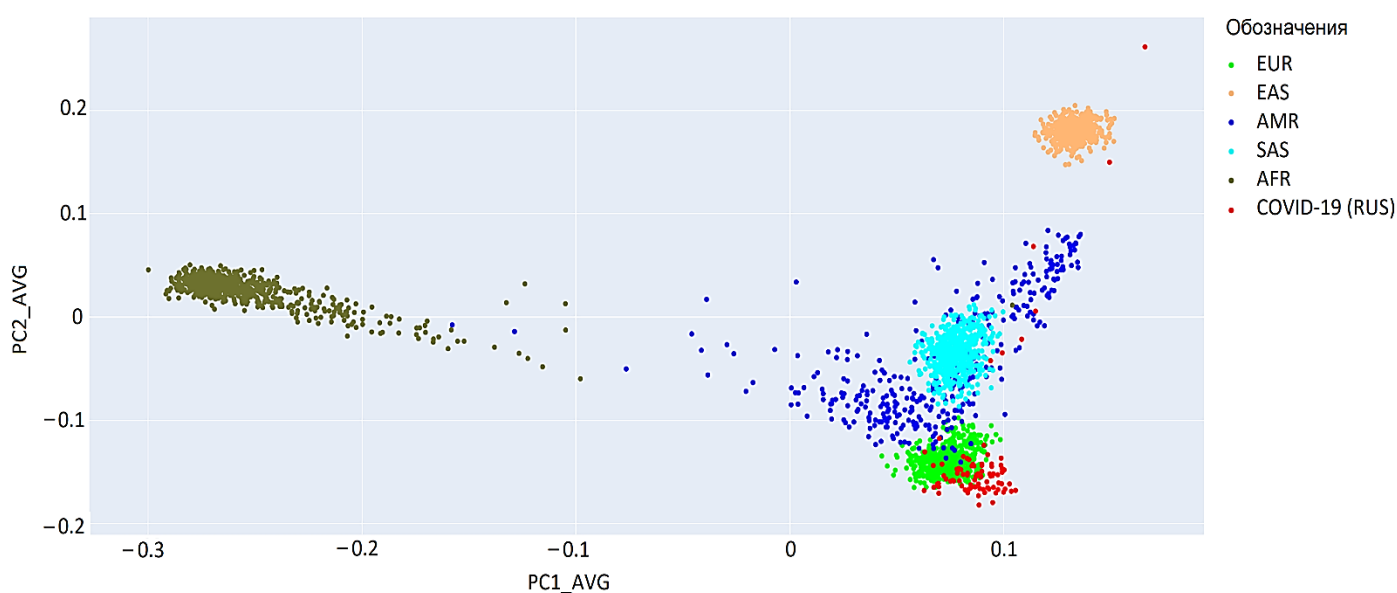
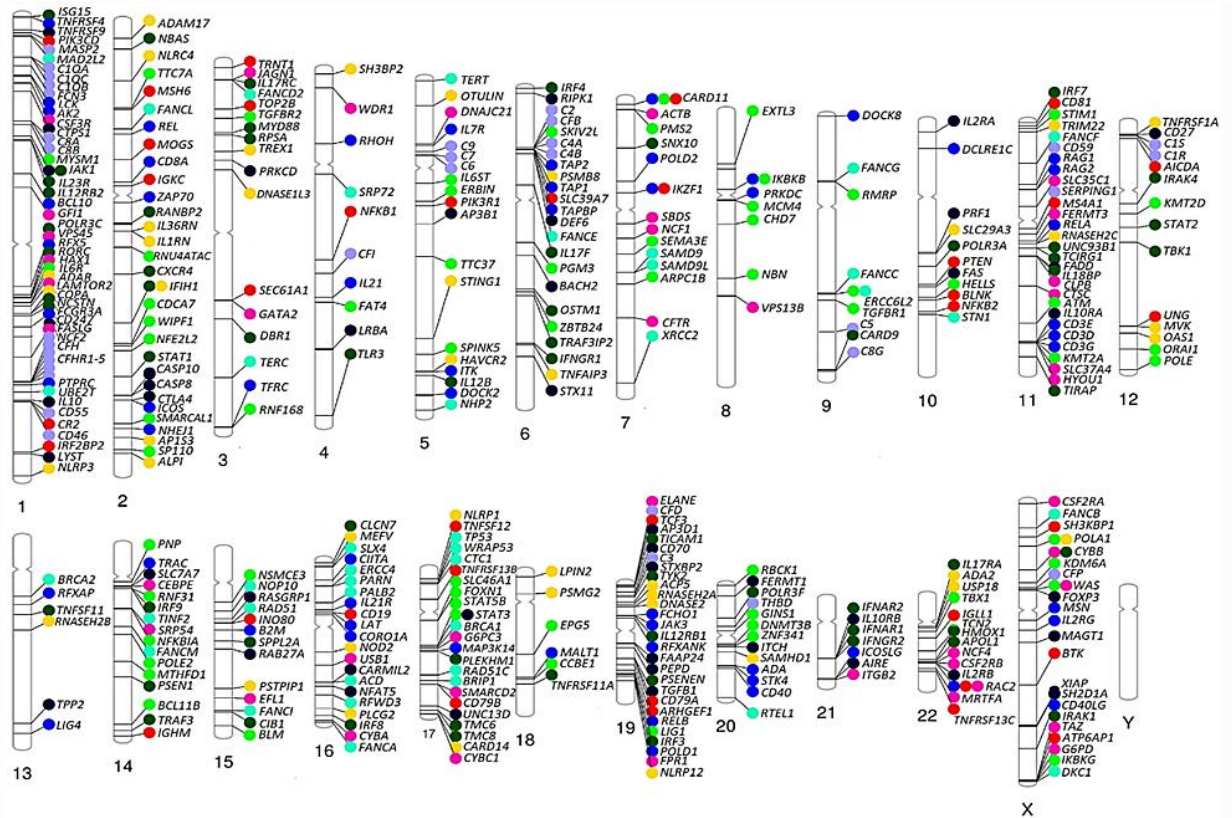


Рис. 1. Принципиальный компонентный анализ аллельных частот (PCA)



- Иммунодефициты, влияющие на клеточный и гуморальный иммунитет
- Дефекты внутреннего и врожденного иммунитета
- Комбинированные иммунодефициты с сопутствующими или синдромальными признаками
- Аутовоспалительные расстройства
- Преимущественный дефицит антител
- Недостатки комплемента
- Заболевания, связанные с нарушением иммунной регуляции
- Недостаточность костного мозга
- Врожденные дефекты количества или функции фагоцитов

Рис. 2. Цитогенетическая фенограмма расположения генов ПИД, связанных с врожденными ошибками иммунитета

Таблица 1. Значимости (*p value*) для уровней экспрессии ПИД генов по сравнению со всеми остальными генами в клетках крови человека согласно базе данных HPA

Клетки (HPA)	<i>P value</i>
gdT-клетки	$1,045 \times 10^{-4}$ *
Эозинофилы	$2,377 \times 10^{-2}$
Классические моноциты	$3,608 \times 10^{-2}$
Базофилы	0,2209
CD4 Т-клетки памяти	0,1077
В-клетки памяти	$5,319 \times 10^{-2}$
MAIT Т-клетки	$1,555 \times 10^{-2}$
Промежуточные моноциты	$4,245 \times 10^{-3}$ *
Наивные CD4 Т-клетки	0,5319
Наивные В-клетки	0,1313
Миелоидные клетки	$2,163 \times 10^{-2}$
CD8 Т-клетки памяти	0,2967
Неклассические моноциты	$6,246 \times 10^{-4}$ *
НК-клетки	0,205
Нейтрофилы	$4,929 \times 10^{-5}$ *
Наивные CD8 Т-клетки	$2,218 \times 10^{-2}$
Регуляторные Т-клетки	$5,904 \times 10^{-3}$ *
Плазмацитоидные дендритные клетки	$6,436 \times 10^{-2}$
Мононуклеарные клетки периферической крови	$4,674 \times 10^{-3}$ *

*Порог значения *p* для множественных сравнений с учетом всех протестированных сравнений составляет 0,002631

Таблица 2. Значимости (*p* value) для уровней экспрессии ПИД генов по сравнению со всеми остальными генами в клетках крови человека согласно базе данных Monaco

Клетки (MONACO)	<i>P</i> value
Эффекторные CD8 Т-клетки памяти	$1,419 \times 10^{-7}$ *
Классические моноциты	$8,172 \times 10^{-7}$ *
Центральные CD8 Т-клетки памяти	$3,009 \times 10^{-6}$ *
Базофилы	$6,16 \times 10^{-6}$ *
CD4 Т-клетки памяти TFH	$2,68 \times 10^{-5}$ *
MAIT Т-клетки	$2,969 \times 10^{-5}$ *
Промежуточные моноциты	$6,528 \times 10^{-9}$ *
Истощенные В-клетки памяти	$2,772 \times 10^{-9}$ *
CD4 Т-клетки памяти Th2	$1,681 \times 10^{-6}$ *
CD4 Т-клетки памяти Th17	$2,523 \times 10^{-8}$ *
Наивные CD8 Т-клетки	$2,36 \times 10^{-4}$ *
Наивные CD4 Т-клетки	$9,765 \times 10^{-5}$ *
Наивные В-клетки	$6,967 \times 10^{-5}$ *
Миелоидные дендритные клетки	$5,762 \times 10^{-8}$ *
Непереключенные В-клетки памяти	$6,319 \times 10^{-8}$ *
Неклассические моноциты	$2,264 \times 10^{-9}$ *
НК-клетки	$7,365 \times 10^{-6}$ *
Нейтрофилы	$2,953 \times 10^{-13}$ *
Стволовые клетки	$1,371 \times 10^{-4}$ *
Плазмацитоидные дендритные клетки	$1,614 \times 10^{-5}$ *
Терминальные эффекторные CD8 Т-клетки памяти	$5,312 \times 10^{-11}$ *
Терминальные эффекторные CD4 Т-клетки памяти	$2,551 \times 10^{-7}$ *
Регуляторные Т-клетки	$1,316 \times 10^{-5}$ *
Переключенные В-клетки памяти	$1,282 \times 10^{-9}$ *
Vd2 gdTCR	$3,6 \times 10^{-4}$ *
CD4 Т-клетки памяти Th1/Th17	$2,609 \times 10^{-6}$ *
CD4 Т-клетки памяти Th1	$1,383 \times 10^{-5}$ *
Плазматические клетки	$3,004 \times 10^{-4}$ *
Non-Vd2 gdTCR	$6,48 \times 10^{-7}$ *
Мононуклеарные клетки периферической крови	$2,317 \times 10^{-6}$ *

*Порог значения *p* для множественных сравнений с учетом всех протестированных сравнений составляет 0,001666