

**Федеральное государственное автономное образовательное учреждение  
высшего образования  
«Уральский федеральный университет  
имени первого Президента России Б.Н. Ельцина»  
Минобрнауки Российской Федерации**

**Федеральное государственное бюджетное учреждение науки  
Институт иммунологии и физиологии  
Уральского отделения Российской Федерации**

*На правах рукописи*

**ШИНВАРИ ХАЙБЕР**

**НОВЫЕ ВАРИАНТЫ ГЕНОВ В ОЦЕНКЕ ВРОЖДЕННЫХ  
ОШИБОК ИММУНИТЕТА: ДЕФИЦИТ RVSK1,  
ВРОЖДЕННАЯ НЕЙТРОПЕНИЯ, СИНДРОМ ХЕННЕКАМА**

3.2.7 – иммунология

Диссертация на соискание ученой степени  
кандидата биологических наук

**Научные руководители:**

- *Тузанкина Ирина Александровна*, д.м.н.,  
профессор, ЗДН РФ;

- *Черешнев Валерий Александрович*, академик,  
д.м.н., профессор, ЗДН РФ

Екатеринбург - 2023

**ОГЛАВЛЕНИЕ**

ВВЕДЕНИЕ.....	3
ГЛАВА 1 - ОБЗОР ЛИТЕРАТУРЫ.....	13
ГЛАВА 2 - МАТЕРИАЛЫ И МЕТОДЫ, ИСПОЛЬЗОВАННЫЕ В РАБОТЕ	56
ГЛАВА 3 – Оценка различия экспрессии генов и исследование основных сигнальных путей у пациентов с дефицитом <i>RBSK1</i> .....	94
ГЛАВА 4 – Исследование влияния выявленных несинонимичных однонуклеотидных замен в генах <i>ELANE</i> и <i>TCIRG1</i> на структуру и функцию белков <i>ELANE</i> и <i>TCIRG1</i> .....	105
ГЛАВА 5 - Идентификация новых миссенс-мутаций генов <i>CCBE1</i> , <i>FAT4</i> , <i>ADAMTS3</i> , приводящих к синдрому Хеннекама.....	150
ЗАКЛЮЧЕНИЕ .....	192
СПИСОК ЛИТЕРАТУРЫ.....	202
СПИСОК СОКРАЩЕНИЙ.....	224

## ВВЕДЕНИЕ

### Актуальность исследования

Иммунная система - это сложная биологическая система, предназначенная для борьбы с чужеродными антигенами, для распознавания чужеродных внешних и внутренних антигенов, для уничтожения инфицированных и аномально развивающихся клеток, а также для контроля толерантности к аутоантигенам и комменсальной микробиоте, выполняющая важнейшую биологическую задачу сохранения вида.

Вследствие этого врожденные ошибки иммунитета (ВОИ) или первичные иммунодефициты (ПИД) могут приводить к повышенной восприимчивости к инфекциям, аутоиммунным процессам, аутовоспалительным заболеваниям, злокачественным опухолям или аллергии. Причиной этому могут быть, в первую очередь, генетические изменения как на уровне генома, так и отдельных генов, кодирующих протеиновые молекулы, участвующие в обеспечении иммунных механизмов.

Несмотря на то, что до недавнего времени ВОИ считали редкими болезнями, а отдельные генетические нарушения могут встречаться нечасто, в совокупности они могут поражать значительное число людей. Более того, в результате улучшения диагностики и развития технологий секвенирования следующего поколения (NGS), за последние годы констатируемая распространенность первичных иммунодефицитов (ПИД) увеличилась до 40 на 100 000 населения [89, 164].

Для разработки новых методов диагностики и терапии иммунопатологии необходимо глубокое понимание реализации иммунных функций в организме. Появление высокопроизводительных биологических методов позволило получить беспрецедентное представление о молекулярных механизмах, лежащих в основе функционирования иммунной системы и её взаимосвязи с другими системами организма. Однако огромная сложность анализируемых параметров, охватывающая несколько порядков

пространственных и временных масштабов, может быть постигнута только с помощью системной вычислительной иммунологии - в частности, путем использования вычислительных подходов обработки и моделирования больших иммунологических данных.

Поэтому в нашей работе были использованы методы биоинформационного анализа при отдельных синдромах, развивающихся в результате врожденных ошибок иммунитета. Для этого были отобраны три патологии – врожденная нейтропения (одна из самых распространенных форм ПИД), синдром Хеннекама (одна из самых редких), а также дефицит *RBSK1*, относящийся к аутовоспалительным синдромам ПИД, но при этом с повышенной восприимчивостью к пиогенным инфекциям. Дефицит *RBSK1* впервые был описан в 2012 году [200], синдром Хеннекама в 1989 [25], а первые открытия в генетике врожденной нейтропении относятся к 1999 году [137, 140]. Тем не менее, до настоящего времени диагностика каждого из этих синдромов представляет сложность, продолжают выявляться новые варианты генов, которые приводят к фенотипам этих болезней, а точные механизмы патологии синдрома Хеннекама и дефицита *RBSK1* остаются предметом дискуссии.

Кроме того, важно не только выявить варианты генов, но и доказать их влияние на конечный продукт - белок, дестабилизацию структуры которого возможно оценить инструментами *in silico*. Это позволит ускорить оценку патогенности варианта гена, а также включить выявленные варианты в список причиннозначимых для ускорения диагностики, а также приблизиться к разработке методов патогенетической или генотерапии, что является конечной целью при исследовании врожденной патологии человека.

Таким образом, важность выявления причиннозначимых вариантов генов иммунопатологии, а также поиск механизмов развития патологии, которые приводят к фенотипу исследуемых синдромов, обусловили цель исследования.

**Цель исследования:** Определить роль потенциально патологических вариантов причиннозначимых генов в патогенезе врожденных ошибок иммунитета – дефиците *RBCK1*, врожденной нейтропении и синдроме Хеннекама, используя методы биоинформационного анализа.

**Задачи исследования:**

1. Провести сравнительный анализ экспрессии генов при дефиците *RBCK1* относительно здоровых детей и пациентов с синдромами CINCA/NOMID, Макл-Уэллса и дефицита мевалонаткиназы.
2. Оценить патогенность несинонимичных однонуклеотидных замен в генах *ELANE* и *TCIRG1* при врожденной нейтропении.
3. Выявить потенциально возможные новые гены-кандидаты при развитии болезней, относящихся к группе врожденных нейтропений.
4. Выявить новые варианты генов *CCBE1*, *ADAMTS3* и *FAT4*, приводящих к развитию синдрома Хеннекама.

**Методология и методы исследования.**

Работа выполнена на кафедре иммунохимии Химико-технологического института УрФУ и на базе ФГБУН Института иммунологии и физиологии УрО РАН (Екатеринбург, Россия) в соответствии с программой фундаментальных научных исследований «Иммунологические механизмы онтогенеза человека и их роль в формировании патологических состояний» (№ гос. регистрации – 01201352044).

Для решения поставленных задач были использованы различные источники данных и методы исследования.

Для исследования патогенеза дефицита *RBCK1* был проведен сравнительный анализ экспрессии генов, для чего из базы данных NCBI Gene Expression Omnibus (GEO) были загружены 2 датасета:

- 1) GSE31064, включавший данные, полученные по результатам исследования фибробластов кожи пациентов – 2 с дефицитом *RBCK1*, один с

дефицитом *MYD88*, один с синдромом *NEMO* и 3 образца здоровых людей (в качестве контроля);

2) GSE40561, включавший данные, полученные по результатам исследования цельной крови, собранной у 2 пациентов с болезнью *CINCA/NOMID*, пяти пациентов с синдромом Макл-Уэллса, двух пациентов с дефицитом мевалонаткиназы, одного пациента с дефицитом *RBSK1* и 41 здорового ребенка (в качестве контроля).

Дифференциально экспрессированные гены при заболевании могут играть ключевую роль в анализируемой болезни или состоянии и могут быть потенциальными генами-кандидатами для дальнейшего исследования.

С этой целью для поиска генов-кандидатов врожденной нейтропении был проведен анализ различий экспрессии генов двух наборов данных, загруженных из NCBI GEO (<https://www.ncbi.nlm.nih.gov>). Исследуемый материал - набор данных GSE142347 включал результаты анализа экспрессии генов 93 пациентов женского пола, 95 пациентов мужского пола, 193 контрольных результатов от здоровых, а набор данных GSE6322 – включал результаты анализа экспрессии генов 2 родителей и 2 детей с нейтропенией из одной семьи.

Данные по различным генам и однонуклеотидным заменам (SNP) при врожденной нейтропении и синдроме Хеннкама были загружены из баз данных dbSNP-NCBI (<https://www.ncbi.nlm.nih.gov/snp/>) и Ensembl (<https://www.ensembl.org/index.html>).

Для исследования однонуклеотидных полиморфизмов или замен (SNP) в генах врожденной нейтропении были загружены следующие SNP: для гена *ELANE*: 3646 SNP, из которых несинонимичных однонуклеотидных замен (nsSNP): 301; для гена *TCIRG1*, соответственно, всего 5627 SNP, из которых 811 nsSNP.

Для исследования SNP в генах синдрома Хеннекама: *CCBE1* – 73845 SNP и 407 nsSNP; *FAT4* - 68257 SNP и 3434 nsSNP; *ADAMTS3*: 70876 SNP и 911 nsSNP.

Для оценки вариантов генов были использованы другие базы данных, включая Swiss-Prot (<http://expasy.org/>) и OMIM (<https://www.omim.org/>).

Исследование вариантов генов у пациентов с врожденной нейтропенией и синдромом Хеннекама из Свердловской области выполнено благодаря имеющимся результатам секвенирования (выполненных в МГЦ Геномед), предоставленным добровольно пациентами для исследования в Институте иммунологии и физиологии УрО РАН, далее обезличенным.

Для проведения оценки вредоносности несинонимичной однонуклеотидной замены на структуру и функцию белка использовалась следующая очередность действий. В начале все выявленные в базах данных nsSNP оценивались инструментом SIFT, затем отсортированные вероятно вредоносные мутации проходили через программу PolyPhen-2, а в дальнейшем отправлялись на оценку другими инструментами биоинформатики, как программами, так и онлайн-сервисами, общим числом до 18 - PROVEAN, FATHMM, LRT, M-CAP, META SVM, METALR, Mutation Assessor, Mutation Taster, FATHMM MKLCoding, CAAD, PHD-SNP, Panter, SNP&GO, PON-P2, DANN, SNAP2 - все эти инструменты были доступны через VarCard [212] и MutPred [99]. Итоговый результат фильтрации, в котором предсказание вредоносности совпадало во всех инструментах, считался вероятно вредоносными заменами, и уже только они оценивались на предмет влияния на вторичную и третичную структуру белка, также оцениваясь в симуляции молекулярной динамики.

Для оценки влияния однонуклеотидных замен на структуру и стабильность белков использовались программы биоинформационного анализа I-Mutant и MU-PRO.

Для оценки белок-белковых взаимодействий использовался пакет программного обеспечения STRING и Cytoscape.

Для анализа функционального обогащения использовалась база KEGG.

Для анализа ко-экспрессии генов использовалась программа SemiTool.

Для построения 3D-модели структуры белков дикого и мутантного типов при оценке влияния мутаций на функцию белка были использованы программы HHPred, Alpha fold 2, Phyre2, I-Taser, Chimera UCSF Chimera и PyMOL.

Симулирование молекулярной динамики проводилось с помощью пакетов Maestro и Gromacs 4.5.3 от компании Schrödinger LLC.

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

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

**Степень достоверности и апробация результатов.** Достоверность полученных результатов исследования обеспечена обоснованностью исходных теоретических позиций, применением современных методов исследования, соблюдением правил использования соответствующих инструментов сбора данных, и высокотехнологичным современным анализом данных. Дизайн исследования был основан на обширном обзоре современных литературных данных, а гипотезы были проверены с использованием различных статистических и биоинформационных методов.

Личный вклад автора состоит в прямом участии на всех этапах диссертационного исследования. Создание основной идеи, планирование научной работы, включая формулировку рабочей гипотезы, целей, задач, определение методологии исследования, интерпретация и анализ полученных результатов были проведены диссертантом вместе с научными руководителями - заслуженным деятелем науки, доктором медицинских наук, профессором И.А. Тузанкиной и академиком РАН, доктором медицинских наук, профессором В.А. Черешневым. Для исследования была использована большая и разнообразная популяционная выборка, отобранная из онлайн-баз данных. Данные были проанализированы с помощью биоинформатических и вычислительных методов биологии. Автор совместно с биотехнологом PhD Хафизом Музамилом Рахманом и кандидатом медицинских наук, старшим научным сотрудником лаборатории иммунологии воспаления Института иммунологии и физиологии Уральского отделения Российской академии наук М.А. Болковым, провел биоинформатический анализ, который включал анализ дифференциальной экспрессии генов, анализ сигнальных путей, анализ онтологии генов, межбелковых взаимодействий, *in silico* анализ nsSNP, моделирование белков и симуляцию молекулярной динамики.

**Положения, выносимые на защиту:**

1. При дефиците *RBSK1* снижена экспрессия генов сигнальных путей иммунного ответа, воспалительного ответа и фосфорилирования белков.
2. При врожденной нейтропении выявлен перечень генов, расширяющий спектр известных генов этой группы первичных иммунодефицитов.
3. Новые несинонимичные однонуклеотидные замены в генах *TCIRG1* и *ELANE* являются дестабилизирующими для белков *TCIRG1* и *ELANE*.
4. Идентифицированные новые несинонимичные однонуклеотидные замены в генах, приводящих к развитию синдрома Хеннекама, являются

дестабилизирующими структуру и функцию белков *CCBE1*, *ADAMTS3* и *FAT4*.

### **Научная новизна исследования.**

Впервые выявлены различия экспрессии генов в мононуклеарах периферической крови у лиц с дефицитом *RBCK1* и относительно здоровых лиц.

Впервые выявленные новые патогенные варианты генов *TCIRG1* и *ELANE*, также, как и уже опубликованные варианты nsSNP, были проанализированы на предмет их влияния на соответствующие белки.

Впервые выявлены новые гены-кандидаты врожденной нейтропении, которые в будущем могут быть использованы для диагностики болезни и разработки патогенетически обоснованных методов лечения.

Впервые выявленные новые несинонимичные однонуклеотидные полиморфизмы (nsSNP) в причиннозначимых генах синдрома Хеннекама (*CCBE1*, *FAT4* и *ADMATS3*) были определены вместе с ранее опубликованными вариантами как оказывающие значительное влияние на структуру и функцию этих белков.

### **Теоретическая и практическая значимость исследования.**

Теоретическая значимость исследования заключается в получении с помощью разработанной программы последовательного использования методов биоинформационного анализа, включая симуляцию молекулярной динамики, новых данных о генетико-фенотипических взаимосвязях, составляющих патогенетическую основу формирования болезней, ассоциированных с врожденными ошибками иммунитета, а именно: дефицита *RBCK1*, врожденной нейтропении и синдрома Хеннекама. Это позволит использовать полученные сведения в дальнейших исследованиях при поиске терапевтических мишеней при данных болезнях.

Практическая значимость исследования заключается в открывающейся возможности использования предсказанных вариантов генов в дифференциально-диагностическом процессе при определении таких синдромов первичных иммунодефицитов, как дефицит *RBSK1*, врожденная нейтропения и синдром Хеннекама.

Разработанная программа последовательного использования методов биоинформационного анализа может быть использована при поиске новых генов-кандидатов болезней, ассоциированных с врожденными ошибками иммунитета.

**Внедрение результатов исследования в практику.** Результаты работы внедрены в учебный процесс аспирантуры Института иммунологии и физиологии УрО РАН и кафедры иммунохимии Химико-технологического института Уральского федерального университета имени первого Президента России Б.Н. Ельцина, а также в научно-исследовательскую практику лаборатории иммунологии воспаления Института иммунологии и физиологии УрО РАН. Полученные результаты внедрены в практику работы ГАУЗ СО «Клинико-диагностический центр «Охрана здоровья матери и ребенка».

**Публикации.** По результатам диссертации соискателем опубликовано 13 работ, в том числе в изданиях, рекомендованных ВАК категории К1 – 5, индексируемых в международных электронных базах данных Web of Science Q1 - 2, Web of Science (Q2-Q4) и Scopus – 6.

**Объем и структура диссертации.** Диссертация изложена на 230 страницах машинописного текста, состоит из введения, обзора литературы, материалов и методов, трех глав с результатами собственных исследований, выводов, практических рекомендаций, списка сокращений и списка литературы (228 источников, из них 12 отечественных и 216 иностранных). Работа иллюстрирована 20 таблицами, 79 рисунками и 2 формулами.

**Благодарности.**

Я искренне благодарен Болкову Михаилу Артемовичу, старшему научному сотруднику лаборатории иммунологии воспаления Института иммунологии и физиологии УрО РАН, за его эффективные комментарии и освещающие дискуссии.

Я благодарен профессору Гусеву Евгению Юрьевичу, заведующему лабораторией иммунологии воспаления Института иммунологии и физиологии УрО РАН, за системный подход к поднятым в работе проблемам и важные замечания по общим молекулярным механизмам и сигнальным путям, исследуемым в нашей работе.

Я благодарен доктору медицинских наук, профессору Института иммунологии и физиологии УрО РАН, член-корреспонденту РАН Юшкову Борису Германовичу и доктору медицинских наук Сарапульцеву Алексею Петровичу за точные замечания, позволившие скорректировать работу в нужном направлении.

## ГЛАВА 1 - ОБЗОР ЛИТЕРАТУРЫ

Первичные иммунодефициты (ПИД), возникающие в результате врожденных ошибок иммунитета, классификация которых включает более 485 моногенных синдромов, могут поражать до 1% населения [4, 89]. При отсутствии надлежащего выявления и лечения люди с ПИД подвержены серьезным, длительным, зачастую опасным для жизни инфекциям, аутоиммунным и аутовоспалительным процессам, репаративным нарушениям и опухолям. Несмотря на достигнутые успехи, осведомленность о ПИД остается критически важным вопросом как для медицинской общественности, так и для населения в целом, поскольку должна привести к улучшению диагностики, своевременному применению современных и эффективных терапевтических методов, использование которых позволит не только повысить качество жизни, но и сохранить ее при этих заболеваниях. Между тем, пациенты с различными проявлениями болезни не всегда имеют шансы выздоровления из-за несвоевременного выявления первопричины болезни – наличия первичных иммунодефицитов, что может наблюдаться, например, у пациентов с онкологической патологией, несмотря на известный факт повышенного риска развития злокачественных опухолей у пациентов с ПИД. То же может наблюдаться и при аутоиммунной или аутовоспалительной патологии, которая может быть единственным проявлением ПИД [149].

### **1.1 - Классификация первичных иммунодефицитов, распространенность, диагностика, лечение и профилактика**

Симптомами первичных иммунодефицитов могут быть инфекционные заболевания, причиной развития которых могут быть врожденные ошибки иммунитета [115], но часто характеризующиеся многими другими отклонениями, включая повышенную уязвимость к злокачественным опухолям, лимфопролиферативным, аутоиммунным заболеваниям и аутовоспалительным нарушениям [4, 89]. С 2017 года все первичные

иммунодефициты классифицируются как врожденные ошибки иммунитета, но в международном классификаторе болезней 11-го пересмотра по-прежнему фигурирует термин «первичный иммунодефицит» [21]. Доступ к исследованиям, относящимся к ПИД, затруднен из-за редкости расстройств и отсутствия известных причинных генетических дефектов [103]. Фенотипические проявления болезни, как и другие следствия генетической патологии, могут появиться в любом возрасте, но более тяжелые формы заболевания чаще встречаются в младенчестве или раннем детстве. Поскольку ПИД часто проявляются в виде инфекций с разнообразными симптомами и клиническими проявлениями, в практическом здравоохранении зачастую лечат инфекции, упуская основную причину [9]. Это часто приводит к повторным инфекциям, ранней хронизации, развитию тяжелых осложнений и прогрессированию заболевания, инвалидизации, необратимому повреждению органов или даже к летальному исходу. Так, например, в США в 2010х от появления первых симптомов до постановки диагноза ПИД проходит в среднем 12,4 года [228]. Это означает, что многие люди с ПИД сталкиваются с рецидивирующими инфекциями с негативными последствиями, которые могут влиять на их личную, социальную и профессиональную жизнь в течение более десяти лет. Однако после распознавания врожденных ошибок иммунитета и лечения ПИД пациенты могут вести нормальную, продуктивную жизнь при условии обеспечения патогенетически обоснованной терапии или применения радикальных куративных мер [2, 10, 51]. Для решения этой критической проблемы Фонд Джеффри Моделла (JMF) сформировал глобальную сеть специализированных центров и разработал 10 предупреждающих признаков ПИД [74]. Ознакомиться с полным оригиналом документа можно также на официальном сайте JMF. Но на данный момент в России распространен усовершенствованный вариант из 12 настораживающих признаков, в котором особое внимание уделено неинфекционным проявлениям. Они более разнообразны, а при многих синдромах

инфекционные проявления и вовсе отсутствуют или занимают второстепенное место.

Для этого в ноябре 2021 года Национальная Ассоциация Экспертов по Первичным Иммунодефицитам (НАЭПИД) совместно с благотворительным фондом «Подсолнух», оказывающим помощь детям и взрослым с нарушениями иммунитета, обновили перечень настораживающих признаков первичных иммунодефицитов для использования в здравоохранении специалистами различного профиля. Этот перечень фокусируется на признаках первичных иммунодефицитов, которые проявляются, в первую очередь, на первом году жизни и часто приводят к летальному исходу в раннем детстве (*рисунок 1*).

#### **Настораживающие признаки первичных иммунодефицитов**

1. Семейный анамнез – наличие случаев ПИД у родственников.
2. Частые бактериальные инфекции.
3. Тяжёлое течение бактериальных инфекций, требующее применения внутривенных антибиотиков.
4. Инфекции, вызванные условно-патогенными возбудителями.
5. Тяжёлые атипичные кожные проявления, отёки.
6. Воспалительные заболевания кишечника с ранним началом и/или тяжёлым течением.
7. Снижение значений показателей общего анализа крови.
8. Длительное увеличение лимфоузлов, печени, селезенки.
9. Значительное уменьшение размеров тимуса, лимфоузлов, миндалин.
10. Повторные лихорадки без очагов инфекции.
11. Сочетание нескольких аутоиммунных нарушений, включая эндокринопатии.
12. Особенности строения лица (врожденные пороки развития и малые аномалии развития).



# НАСТОРАЖИВАЮЩИЕ ПРИЗНАКИ ПЕРВИЧНЫХ ИММУНОДЕФИЦИТОВ

Разработано НАЭПИД совместно с Благотворительным Фондом «ПОДСОЛНУХ»

НАЭПИД: [noepid.ru](mailto:noepid.ru), [pidrussia@gmail.com](mailto:pidrussia@gmail.com), +7 495 221 6640  
Регистр пациентов с ПИД: [naepid-reg.ru](mailto:naepid-reg.ru)



БФ «ПОДСОЛНУХ»: 8 800 500 6335  
[fondpodsolnuh.ru](http://fondpodsolnuh.ru), [propid.ru](http://propid.ru)



Рисунок 1 – Настораживающие признаки первичных иммунодефицитов

С. Picard и соавт., 2015, определили распространенность ПИД как 1 случай на 1200, с диапазоном от 1:600 для дефицита IgA и 1:20 000 для всех иммунодефицитов, 1:50 000 для ТКИН и 1:100 000 для X-сцепленная агаммаглобулинемии [227]. Последние данные утверждают, что врожденные ошибки иммунитета встречаются у 1% популяции [79].

Считается, что иммунодефицитные заболевания более значимы для планирования работы в здравоохранении в той стране, где смерти от общих инфекций были почти полностью устранены, а дети с ПИД доживают до того возраста, чтобы быть выявленными. Поэтому Pילania et al, 2019, утверждают, что ПИД чаще всего выявляются в тех странах, в которых младенческая смертность не превышает 15/1000 рождений [57]. Эпидемиологические наблюдения ПИД в странах Азии, таких как Япония и Корея, в первых отчетах о ПИД, датируются 1950-ми годами. Первая программа опроса и регистрации пациентов с ПИД в Японии была создана в 1974 году с созданием Регистрационного центра иммунодефицитов в Японии в отделении педиатрии Университета Токио [84]. Было зарегистрировано 497 пациентов, выявленных в Японии с 1966 по 1975 годы. Среди них наиболее часто диагностируемыми ПИД были дефицит IgA, X-сцепленная агаммаглобулинемия (XLA) и атаксия-телеангиэктазия. Программа исследований при Министерстве здравоохранения, труда и социальной защиты Японии создала клиническую группу исследований с целью проведения эпидемиологических, патологических, диагностических и терапевтических исследований по ПИД, и к 2008 году была создана Сеть баз данных первичных иммунодефицитов (PIDJ) с целью расширения возможностей исследований и обслуживания пациентов. В 2011 году было зарегистрировано 1240 пациентов с ПИД, а распространенность заболевания составила 2,3 на 100 000 населения [142]. Хотя эта распространенность была выше, чем в более ранних отчетах в Японии, но она была намного ниже, чем в западных странах и на Ближнем Востоке. Было постулировано несколько причин такого несоответствия. К ним

относятся: низкий уровень кровного родства в этом регионе, смещение выборки (бессимптомный селективный дефицит IgA, транзиторная гипогаммаглобулинемия младенчества и некоторые другие ПИД не были включены в это исследование) и меньшая выявляемость ПИД у взрослых [142]. В настоящее время возможности для диагностики и лечения ПИД имеются в 66 больницах по всей Японии [209].

Самые ранние сообщения о ПИД из Китая были опубликованы в 1960-х годах. Интерес к ПИД в Китае стал более очевидным в 1980-х годах [116]. В 1981 году при Китайском педиатрическом обществе Китайской медицинской ассоциации была создана секция педиатрической иммунологии и в 1998 году была создана совместная сеть и регистр пациентов с ПИД. Самая большая когорта пациентов с ПИД была зарегистрирована в детской больнице Чунцинского медицинского университета, где диагноз был поставлен 352 пациентам в период с 2005 по 2011 год, из которых у 203 был проведен генетический анализ [219]. Большие когорты пациентов с ПИД также были выявлены в других медицинских центрах Китая, включая больницу Синьхуа Шанхайского университета Цзяотун, детскую больницу Пекина и детскую больницу Гуанчжоу.

В регионе Тайвань Институт помощи при первичных иммунодефицитах и исследований (PICAR) при Мемориальной больнице Чанг Гунг, г. Таоюань, обслуживает население численностью около 23 миллионов человек и имеет диагностические и лечебные учреждения для различных ПИД [57]. Еще один подобный центр расположен в городе Тайбэй.

Заболеваемость ПИД в регионе Тайвань составляла 2,17 на 100 000 живорождений, и Тайвань является первым регионом Юго-Восточной Азии, где в 2012 году был проведен общенациональный скрининг новорожденных на ПИД.

Университет Гонконга создал специализированную службу для детей с ПИД в 1988 году, а в 1995 году были впервые созданы условия для

молекулярной диагностики ПИД. В настоящее время Университет Гонконга проводит генетическую диагностику нескольких ПИД, используя полногеномное секвенирование.

Ввиду высокого уровня кровосмешения на Ближнем Востоке, большое количество случаев ПИД было зарегистрировано в Турции и Иране [15, 28]. В этих странах чаще встречаются аутосомно-рецессивные заболевания. Первое отделение детской иммунологии было создано в детской больнице университета Хаджеттепе в 1972 году. В 1974 году было основано Турецкое общество иммунологии. Также имеются возможности для проведения трансплантации гемопоэтических стволовых клеток (ТГСК), и до настоящего времени в Турции было зарегистрировано около 80 пациентов с ТКИН, которым была проведена ТГСК. Недавно в Турции были созданы два Центра иммунодефицитов Фонда Джеффри Моделла (JMF): отделение детской аллергии и иммунологии Университета Мармара в Стамбуле и Центр JMF при Университете Хасеттепе [208].

Первый центр клинической иммунологии и аллергии в Иране был создан в Детском медицинском центре Тегеранского университета медицинских наук в 1977 году профессором Аболхасаном Фархуди [6]. В 1999 году была создана база данных для регистрации иранских пациентов с ПИД - Иранский регистр первичных иммунодефицитов (IPIDR), который находится в Детском медицинском центре и охватывает крупные больницы по всему Ирану. К 2018 году в нем было зарегистрировано 3056 (с 1395 новыми случаями) пациентов [2]. Иранская ассоциация первичных иммунодефицитов (PIA) была основана в 1998 году и ставит своей целью улучшение диагностики, совершенствование системы управления и лечения, а также содействие исследованиям и образованию в области ПИД. В нескольких центрах также имеются возможности для трансплантации гемопоэтических стволовых клеток пациентам с ПИД.

За последние три десятилетия произошел значительный прогресс в понимании патогенеза, диагностике и лечении этих заболеваний. Однако в ряде развивающихся стран эти заболевания все еще остаются недостаточно признанными. Это связано, главным образом, с недостаточной информированностью врачей, а также с отсутствием диагностического оборудования в странах с ограниченными ресурсами.

Самые ранние сообщения о случаях ПИД из Индии относятся к концу 1960-х годов. Сначала сообщалось о случаях пациентов с синдромом Вискотта-Олдрича (WAS), агаммаглобулинемией и атаксией-телеангиэктазией [128, 131, 214]. В 2012 году Гупта и др. опубликовали исследование, в котором сравнили клинический профиль пациентов с ПИД в двух крупных педиатрических центрах (Advanced Pediatrics Centre, Postgraduate Institute of Medical Education and Research (PGIMER), Chandigarh и National Institute of Immuno-haematology (НИИ) and B.J. Wadia Children's Hospital, Mumbai) в Индии [165]. Профиль пациентов с ПИД в этих двух центрах различался. Дефицит антител был самым распространенным ПИД в Чандигархе, в то время как семейный гемофагоцитарный лимфогистиоцитоз (HLH) был самым распространенным ПИД, диагностированным в Мумбаи. Другими распространенными ПИД, диагностированными в Чандигархе, были синдром Вискотта-Олдрича (WAS), синдром гипер-IgE, атаксия-телеангиэктазия и наследственный ангионевротический отек. В Мумбаи было зарегистрировано больше случаев нейтропении, дефицита адгезии лейкоцитов (LAD), нарушений IFN $\gamma$ -IL12-пути и аутоиммунного лимфопролиферативного синдрома [165]. С ростом осведомленности в Мумбаи стало диагностироваться больше случаев ПИД по всей стране.

В других странах Юго-Восточной Азии, включая Сингапур, развиваются центры диагностики и ведения пациентов с ПИД. Так, в Сингапуре в 2003 году было зарегистрировано 39 пациентов. Данные были собраны из 3 крупных центров Сингапура, включая Детский медицинский

институт, Больницу Национального университета, Больницу Тан Ток Сенг и Женскую и детскую больницу. Наиболее распространенным ПИД был дефицит антител, за которым следовали фагоцитарные дефекты [168]. С тех пор в Сингапуре отмечается значительное увеличение числа диагностированных ПИД. Возможности для проведения ТГСК также имеются во многих центрах Сингапура.

Малайзия и Таиланд также догоняют по уровню осведомленности и диагностической базы для ПИД. Национальная инициатива по ПИД была начата в 2007 году с целью улучшения диагностической базы в различных центрах Малайзии, что привело к улучшению лечения ПИД и результатов лечения этих пациентов [43]. Малазийская сеть первичных иммунодефицитов (MyPIN) была создана в 2009 году с целью улучшения диагностической и лечебной базы для ПИД. Здесь зарегистрировано более 300 пациентов с ПИД [127]. В исследовании, опубликованном в Таиланде, сообщается о 72 пациентах с различными ПИД из клиники детской аллергологии/иммунологии/ревматологии Ramathibodi за 1991-2011 годы [167]. Терапия внутривенными иммуноглобулинами (ВВИГ) также доступна для большинства пациентов с ПИД по субсидированной ставке в Малайзии и Таиланде [43]. Структура заболеваний ПИД в Азии варьирует в разных странах Азии. Из-за высокого уровня кровосмешения в странах Ближнего Востока относительно чаще встречаются аутосомно-рецессивные (АР) заболевания [15]. В некоторых других азиатских странах чаще встречаются Х-сцепленные формы заболевания. В исследованиях, проведенных в Японии и Китае, утверждается, что Х-сцепленные формы ТКИН и хронической гранулематозной болезнью (ХГБ) встречаются чаще, чем аутосомно-рецессивные формы. [197, 219]. В результате эндогамных браков аутосомно-рецессивные заболевания также очень распространены в некоторых странах Юго-Восточной Азии, таких как Индия, Пакистан и Бангладеш [169]. Генотип может определять клинический профиль наследственных заболеваний, он

может быть изменен под влиянием множества факторов окружающей среды и определять окончательный фенотип [115]. Факторы окружающей среды влияют на микробиоту кишечника, значимыми факторами могут быть социально-экономические стандарты и спектр доступных медицинских учреждений.

Помимо различий в спектре ПИД, наблюдаемых в Азии по сравнению с остальным миром, пациенты с ПИД в Азии также имеют уникальную и отличную от других картину инфекций, которые могут способствовать заболеваемости и смертности у этих пациентов. Среди этих инфекций преобладают *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Burkholderia pseudomallei* и *Talaromyces marneffeii* [57]. Было установлено, что у пациентов с хронической гранулематозной болезнью (ХГБ) в Азии отмечается удивительно высокая распространенность туберкулезной инфекции по сравнению с пациентами с ХГБ из других стран мира [94]. Из-за более высокой эндемичности туберкулеза во многих странах Азии вакцина против туберкулеза *Bacillus Calmette Guerin* вводится при рождении. Поэтому во многих азиатских странах диссеминированная БЦЖ-инфекция является основным клиническим проявлением многих ПИД, таких как тяжелый комбинированный иммунодефицит (ТКИН), ХГБ, синдром гипер-IgM и дефекты IL12-IFN- $\gamma$ - опосредованного пути [57].

Также сообщалось о высокой частоте артрита у пациентов с ХЛА из азиатских стран [216]. Вероятно, это связано с поздней диагностикой и последующей задержкой начала заместительной иммуноглобулиновой терапии у этих пациентов. *Chromobacterium violaceum* была зарегистрирована в качестве оппортунистической инфекции при фагоцитарных дефектах (например, ХГБ) во многих азиатских странах. Первоначально о нем сообщалось у пациентов из Малайзии, затем о нем сообщалось во Вьетнаме, Тайланде, Шри-Ланке, Индии, а также в Гонконге и Тайване в Китае [144]. Сообщалось о смертности до 50% при инфицировании этим микроорганизмом

[224] Аналогично, мелиоидоз, вызванный *Burkholderia pseudomallei*, также эндемичен во многих странах и является основной проблемой среди пациентов с ПИД в Азии [98].

В последнее время выявлен ряд ПИД, связанных с предрасположенностью к эндемическим микозам (например, *Talaromyces marneffeii*, диссеминированный кокцидиоидомикоз, гистоплазмоз и паракокцидиоидомикоз) в этом регионе. Эти грибковые инфекции обычно связаны с дефектами IL-12/IFN- $\gamma$ -опосредованного пути, усилением функции STAT1 и другими заболеваниями, опосредованными Th17-лимфоцитами [113, 155].

Пероральная живая вакцина против полиомиелита все еще используется в нескольких азиатских странах и является серьезной проблемой для многих пациентов с ПИД из этих стран. Пациенты с гипогаммаглобулинемией часто получают её еще до установления диагноза иммунодефицита. Эти пациенты также могут инфицироваться вакцинным штаммом вируса при тесном контакте в семье и обществе, а дальше его очень трудно вывести из организма. Таким образом, иммунодефицит-ассоциированные полиовирусы, полученные вакцинальным путем (iVDPVs), остаются значительной проблемой для этих пациентов [75, 165]. Они также являются потенциальным резервуаром для передачи полиовируса. В международном многоцентровом исследовании выделение полиовируса изучалось у 653 пациентов с ПИД (570 имели первичный дефицит антител и 65 - комбинированный иммунодефицит). Тринадцать пациентов (2%) выделяли полиовирусы, а неполиомиелитные энтеровирусы были обнаружены у 30 пациентов. Пять из них (0,8%) были классифицированы как пациенты с иммунодефицит-связанным полиовирусом вакцинного происхождения (iVDPV) [6, 153].

В России, по данным Мухиной А.А. и соавт., 2020, минимальная общая распространенность ПИД оценивалась как 1,3:100 000 человек, со

значительными колебаниями по федеральным округам (от 0,9 до 2,8 на 100 000; 3) (рисунок 2). [166].

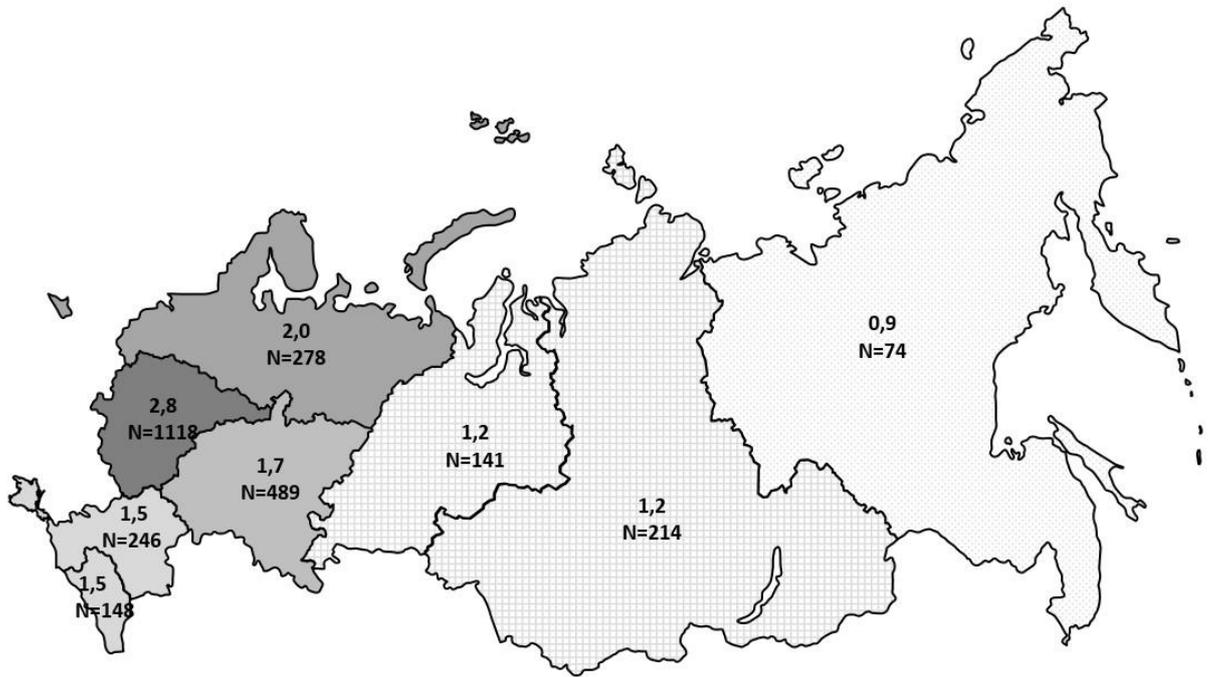


Рисунок 2 – Распространение ПИД (ВОИ) в России по округам [166].

По этим данным, в Российском национальном регистре содержалась информация о почти 3000 пациентах (60% мужчин, 40% женщин) из всех федеральных округов Российской Федерации, 68% этих пациентов были живы, среди них 77% дети, а 23% - взрослые. ПИД был впервые диагностирован до 18 лет у 88% пациентов. Наиболее распространенными группами ПИД были дефекты антител (26%) и ПИД с синдромальными особенностями (22%). Прогнозируемая рождаемость ПИД - 5,7 на 100 000 новорожденных. Медиана задержки диагностики составила 2 года; этот показатель варьировался от 4 месяцев до 11 лет в зависимости от категории ПИД [12, 166].

С 1999 года Международный союз иммунологических обществ (IUIS) классифицирует врожденные ошибки иммунитета (ВОИ), разделяя их на десять групп, в зависимости от того, какая часть иммунной системы поражена. Одна из выделенных групп, десятая, включает аутоиммунные состояния и соматические варианты, имитирующие генетически-обусловленные ВОИ. Каждая группа ВОИ связана с уникальными фенотипическими проявлениями инфекций, аутоиммунитета или воспаления. Например, пациенты с дефицитом антител обычно страдают бактериальными инфекциями дыхательных путей, а пациенты с недостаточностью конечных фракций комплемента склонны к рецидивирующему менингиту, вызванному бактериями *Neisseria*. В отчете IUIS, опубликованном в 2022 году, выделено десять групп ПИД (ВОИ), для каждой из которых описана генетическая причина [89].

### **Классификация первичных иммунодефицитов 2022.**

1. Иммунодефициты с поражением клеточного и гуморального иммунитета.
2. Комбинированные иммунодефициты, ассоциированные с синдромальными проявлениями.
3. Преимущественно антительные дефекты.
4. Болезни иммунодисрегуляции.
5. Врожденные дефекты числа и функций фагоцитов.
6. Дефекты внутреннего и врожденного иммунитета.
7. Аутовоспалительные синдромы.
8. Дефекты комплемента.
9. Костно-мозговая недостаточность.
10. Фенокопии первичных иммунодефицитов.

Если кратко описать основные различия между ПИД, то они зависят от уровня генетического дефекта и соответствующих дефектов рецепторов или

белков. Функционально иммунная система делится на два основных компонента - врожденный и адаптивный иммунный ответ, и в зависимости от того, какой компонент иммунного ответа нарушен в первую очередь, можно условно выделить две большие группы иммунодефицитов. Однако данная классификация наиболее полно отражает структуру и разнообразие врожденных ошибок иммунитета. Она включает и данные о нарушениях третьей составляющей иммунной системы – внутреннего иммунитета [112].

Современные методы лечения ПИД включают симптоматическую поддержку, целевую терапию, заместительную терапию и два вида радикальной хирургии: трансплантацию гемопоэтических стволовых клеток (ТГСК) и генную терапию. Следует отметить, что генная терапия все еще находится на экспериментальной стадии исследований, несмотря на то, что при некоторых формах генетической патологии она уже активно внедряется в клиническую практику [110]. Для некоторых детей с ПИД важнейшим и даже единственным способом лечения заболевания и восстановления функций иммунной системы является ТГСК. Более того, отсутствие генетического подтверждения не является противопоказанием для проведения ТГСК [110].

Одним из важнейших достижений современной медицины является скорость диагностики, в том числе при использовании скрининговых технологий, позволяющих выявить пациентов на доклинических стадиях развития болезни. Это позволяет начать своевременное лечение, предупреждая закрепление патологического фенотипа. Исследование молекул TREC и KREC из капли крови, нанесенной на тест-бланк – карту Гатри, проводится детям в первые сутки жизни при рождении и позволяет выявлять тяжелые комбинированные иммунодефициты и дефекты антителообразования, вызывающие угрожающие жизни заболевания [1]. С 2023 года в Российской Федерации в расширенный неонатальный скрининг включены такие тесты на количественное определение TREC и KREC для всех

новорожденных (Приказ Министерства здравоохранения Российской Федерации от 21.04.2022 274н). [7].

Однако рассматриваемые заболевания не подлежат скринингу данным методом и, как и многие другие врожденные состояния, выявляются не сразу при рождении, а спустя длительное время. При первичных иммунодефицитах скорость диагностики является критически важным фактором. Выявление SNP, которые приводят к болезням, позволяет включить их в автоматические методы оценки патологии, такие как различные диагностические тестовые панели (в том числе по технологии NGS) и биоинформационные базы данных.

## **1.2 - Первичные иммунодефициты и механизмы врожденного иммунитета**

Известно, что врожденный иммунитет включает эпителиальные и слизистые барьеры, естественные антимикробные продукты, рецепторы распознавания образов и цитокины. Он филогенетически предшествует адаптивному иммунитету и присутствует у всех многоклеточных организмов, включая растения, насекомых и животных. Хотя врожденные иммунные клетки несколько примитивны, они организуют дискретный иммунный ответ путем распознавания различных патогенов рецепторами, распознающими паттерны [108,130,157]. Нейтрофилы, макрофаги, дендритные клетки, естественные киллеры (NK) и NKT-клетки в сочетании с естественными барьерами (в основном кожа, слизистые оболочки желудочно-кишечного тракта и дыхательных путей), с антимикробными агентами, опсонинами (например, комплементом) и цитокинами являются важнейшими компонентами врожденного иммунитета.

Наследственные иммунодефициты, приводящие к повышенной восприимчивости к туберкулезу и нетуберкулезным микобактериям, в совокупности называются Менделевской восприимчивостью к микобактериальным заболеваниям (MSMD) [27,36]. Макрофаги

фагоцитируют микобактерии, что приводит к выработке интерлейкина (IL)-12 p70, гетеродимера IL-12 p40 и IL-12 p35, а также IL-23, гетеродимера IL-12 p40 и IL-12 p19. IL-12 и IL-23 стимулируют T- и NK-клетки через свои когнитивные рецепторы фосфорилировать трансдуктор сигнала и активатор транскрипции (STAT)4, что приводит к выработке интерферона (IFN)- $\gamma$ . Последний действует через свой гетеродимерный рецептор, в основном фосфорилируя STAT1 и активируя гены, реагирующие на интерферон, которые способствуют очищению от микобактерий. В последние годы было установлено, что у пациентов с MSMD обнаруживаются мутации в семи различных генах: IFNGR1, IFNGR2, STAT1, IL12B (IL-12p40), IL12RB1, TYK2 и IKBKG (NEMO), которые все вовлечены в IL-12/23-зависимый, IFN- $\gamma$ -опосредованный иммунитет. Недавно было обнаружено, что мутации в гене IRF8 также связаны с возникновением микобактериальных заболеваний (MSMD). Специфические мутации в этих локусах обуславливают различные формы наследования нарушений (аутосомно-рецессивная, аутосомно-доминантная или X-сцепленная), наличие или отсутствие экспрессии белка (миссенс- или нонсенс-мутации), тяжесть фенотипа (полный или частичный дефицит) и специфическую затронутую функцию. Эти синдромы клинически неоднородны и варьируются от локально ограниченных до угрожающих жизни, распространенных микобактериальных заболеваний. Помимо микобактерий, у пациентов с MSMD были зарегистрированы другие внутриклеточные бактерии (например, сальмонелла), вирусы (например, вирус *Varicella zoster*) и грибы (например, гистоплазмоз, кокцидиоидомикоз, паракокцидиоидомикоз) [181, 220].

Мутации в гене рецептора интерферон-гамма 1 типа (*IFNGR1*) были первыми идентифицированными как обуславливающие врожденную восприимчивость к микобактериям. Этот ген может быть изменен таким образом, что это приводит к рецессивной или доминантной передаче. Доминантная форма чаще всего характеризуется нетуберкулезным

микобактериальным остеомиелитом [45]. В то время как рецессивные полные мутации обычно не допускают экспрессии белка, доминантные мутации характеризуются избыточным накоплением на поверхности мутировавшего рецептора, который все еще связывает IFN- $\gamma$ , но в значительной степени подавляет внутриклеточный сигналинг. Дефицит *IL12RB1*, выявленный более чем у 140 пациентов по всему миру, является наиболее распространенной формой врожденной восприимчивости к микобактериям, но, по-видимому, очень чувствителен к окружающей среде. Люди, несущие биаллельные мутации, могут демонстрировать очень слабую восприимчивость к микобактериям, сальмонеллам или грибам, что соответствует частичной пенетрантности и переменной экспрессивности этого дефицита [21].

У пациентов мужского пола с мутациями в *NF- $\kappa$ B* (синдром NEMO) наблюдается широкая клиническая гетерогенность. NEMO кодирует основной модулятор ядерного фактора- $\kappa$ B, также известный как I $\kappa$ B киназа (IKK- $\gamma$ ), критический компонент комплекса IKK. Мутации в этом гене вызывают различные заболевания: аморфные аллели, приводящие к нулевым мутациям, приводят к развитию инконтиненции пигмента у женщин, но являются летальными для мужских плодов. С другой стороны, гипоморфные аллели также могут приводить к развитию потери пигментации у женщин, но у мужчин могут проявляться в виде различных комбинаций синдрома X-сцепленной ангидротической эктодермальной дисплазии и иммунодефицита. X-сцепленная ангидротическая эктодермальная дисплазия с иммунодефицитом, вероятно, является наиболее распространенным фенотипом, но в этом синдроме существует огромная гетерогенность. Ассоциации генотип-фенотип при этом заболевании удивительно неуловимы, но мутации на самом C-конце, включающие стоп-кодон, были ассоциированы с остеопетрозом и/или лимфедемой [14, 158].

Имеются описания пациентов с полным дефицитом тирозинкиназы 2 (*TYK2*) [62]. *Tyk2* является членом сигнального семейства Jak/STAT и

конститутивно связан с рецепторами для интерферонов первого типа (IFN- $\alpha$  и IFN- $\beta$ ), IL-6, IL-10, IL-12 и IL-23. Один пациент имел гомозиготную делецию в 4 пар нуклеотидов, приводящую к раннему стоп-кодону. У него наблюдался сложный клинический фенотип, характеризующийся чувствительностью к вирусам (контагиозный моллюск, простой герпес), грибам (кандидоз полости рта), бактериям (золотистый стафилококк, нетипичная сальмонелла) и микобактериям (локализованный бациллез Кальметта-Герена), а также атопическим дерматитом, умеренной эозинофилией и повышенным уровнем IgE в сыворотке крови. Недавно был зарегистрирован второй случай, в котором у пациента не было аллергических явлений. Таким образом, дефицит *Tyk2* нарушает сигналинг IFN 1 типа (восприимчивость к вирусам), сигналинг IL-12/IL-23 (восприимчивость к микобактериям и поверхностным грибкам) [177], сигналинг IL-6 (восприимчивость к *S. aureus*) [175].

Мутации в *STAT1* бывают рецессивными и доминантными, что приводит к глубокой восприимчивости к разнообразной инфекции в младенчестве (рецессивный полный дефицит) или более легкой восприимчивости к микобактериям, проявляющейся позже в детстве (частичный доминантный дефицит) [18].

Эпидермодисплазия верруциформная (ЭВ) - это редкий генодерматоз, характеризующийся избирательной восприимчивостью к кератиноцитарно-тропным инфекциям ВПЧ (подгруппа В1), обычно проявляющийся в раннем детстве [66, 117].

Синдром WHIM (MIM 193670) является редким аутосомно-доминантным нарушением с частотой примерно 1 случай на 4,3 миллиона живорождений [129]. Термин «WHIM» является аббревиатурой основных клинических проявлений, включающих бородавки, гипогаммаглобулинемию, инфекции и миелокатексис. Миелокатексис характеризуется задержкой выхода нейтрофилов из костного мозга, что приводит к снижению их

количества в крови. Это может привести к рецидивирующим бактериальным инфекциям, особенно кожи, легких и пазух носа.

Гипогаммаглобулинемия, связанная с синдромом WHIM, характеризуется снижением всех классов иммуноглобулинов, что делает пациентов уязвимыми к бактериальным и вирусным инфекциям. Бородавки, также являются распространенным признаком синдрома WHIM, могут быть устойчивыми и рецидивирующими. Пациенты с синдромом WHIM также могут иметь задержку развития костного мозга, что может привести к миелодиспластическому синдрому или острому миелоидному лейкозу.

WHIM синдром вызывается доминантными гетерозиготными патогенными вариантами GOF в гене, кодирующем хемокиновый рецептор 4 (CXCR4). Поскольку CXCR4 участвует в удержании нейтрофилов в костном мозге, зародышевые мутации GOF усугубляют этот процесс, тем самым замедляя выход нейтрофилов, что приводит к нейтропении [129].

В регуляции TLR-сигналинга важную роль играет система убиквитина. Убиквитиновая система - это система посттрансляционных изменений, которая регулирует функцию белков [184]. В некоторых ситуациях для образования цепей полиубиквитина молекула убиквитина связывается с белками-мишенями. При синтезе этих полиубиквитиновых цепей последовательная конъюгация С-концевого остатка глицина включает конъюгацию остатка глицина в одной молекуле убиквитина с одним из семи остатков лизина в другой молекуле убиквитина [111].

### **1.3 - Аутовоспалительные синдромы и дефицит *RBSK1***

Аутовоспалительные заболевания - это широкий класс патологий человека, связанных с ошибками врожденного иммунитета и ассоциированных с дефектами в механизмах воспаления. Этот класс болезней был открыт относительно недавно, но в настоящее время уже известно более 40 аутовоспалительных заболеваний. Основным отличием этих болезней является неконтролируемое аутовоспаление при отсутствии аутоантител,

поэтому ранее такие синдромы называли идиопатическими лихорадками, учитывая, что для них характерно спонтанное воспаление, сопровождающееся лихорадкой. Однако механизмы аутовоспаления были выявлены при многих давно известных заболеваниях, не имеющих прямого отношения к лихорадочному синдрому, таких как ожирение, ревматоидный артрит, болезнь Бехтерева. Они представляют собой воспаление серозных оболочек - плевры, брюшины, синовиальных оболочек суставов, глаз [161].

Так, дефект *RBCK1* (RanBP-Type And C3HC4-Type Zinc Finger-Containing Protein 1) - это аутовоспалительный синдром, отличительной чертой которого является повышенная восприимчивость к инфекциям. Кроме того, для *RBCK1*-дефекта характерно нарушение обмена гликогена, приводящее к его накоплению в мышцах (амилопектиноз). Пациенты с дефектом *RBCK1* имеют широкие и варьируемые клинические проявления, включая лихорадку, инфекционный синдром (различные кожные воспаления, рецидивирующие бактериальные инфекции, вплоть до сепсиса), а также миопатии, кардиомиопатии и энцефалопатии [161].

Известно, что белок *RBCK1*, также известный как *HOIL-1*, участвует в сборке комплекса линейной убиквитиновой цепи. Убиквитины - это белки, которые играют роль «поцелуя смерти» для белков внутри клетки, помечая их как бы черной меткой для расщепления до аминокислот в протеосоме. Комплекс сборки линейной убиквитиновой цепи (*LUBAC*) включает *RBCK1*, *RNF31* (белок с кольцевым пальцем 31, также известный как *HOIL-1*-взаимодействующий белок или *HOIP*) и *SHARPIN* (*SHANK*-ассоциированный белок с RH-доменом). Линейный комплекс сборки убиквитиновых цепей (*LUBAC*) связывается с линейными (*Met1*) убиквитиновыми цепями и направляет несколько белков в классический сигнальный путь *NF- $\kappa$ B*, предотвращая воспаление и участвуя в апоптозе [121, 189].

Исследования показали, что катализируемое *LUBAC* линейное убиквитинирование в ответ на стимуляцию *TNF* участвует в активации

канонического пути NF- $\kappa$ B и предотвращении гибели клеток [184]. RBCK1 (58 кДа) с двумя другими белками, SHANK ассоциированным RH-domain-Interacting Protein (SHARPIN) и HOIL-1 Interacting Protein (HOIP-1), наиболее известен тем, что образует комплекс ~600 кДа [32, 111].

Дефекты каждого из белков LUBAC в отдельности приводят к аутовоспалительным синдромам. Известные случаи дефицита HOIP у человека связаны со снижением количества не только HOIP, но и других белков LUBAC. Известно, что у двух пациентов с дефицитом *HOIP* аутовоспаление (особенно полиартрит мелких суставов с раннего возраста), рецидивирующие лихорадки, тяжелые бактериальные, вирусные и грибковые инфекции и патологическая реакция на пневмококковые антигены при вакцинации. Пациенты с дефицитом *RBCK1* имеют широкий спектр клинических проявлений, но причина этой индивидуальной гетерогенности неизвестна, однако все случаи в той или иной степени сопровождаются дефектом накопления гликогена [14, 095, 174].

У больных могут одновременно наблюдаться хроническое аутовоспаление и иммунодефицит, включая рецидивирующую септицемию [161]. Выявленные на сегодняшний день пациенты с мутациями *RBCK1* значительно отличаются по клиническим проявлениям (скелетные мышцы, сердечная мышца, аутовоспаление или иммунодефицит). Объяснение этой индивидуальной гетерогенности остается неясным, хотя предполагалось, что точное положение варианта в гене может быть предиктором преобладающего фенотипа, причем мутации в основном приводят к иммунологической дисфункции в N-концевой области *RBCK1*, а мутации в средней или C-концевой частях приводят к фенотипу (кардио-) миопатии [161]. M1-связанное линейное полиубиквитинирование опосредуется LUBAC, сложной модификацией, которая делает ядерный фактор- $\kappa$ B (NF- $\kappa$ B) и его плейотропную иммунную систему критической ядерной транслокацией и транскрипционным контролем.

Линейный убиквитин-сборочный комплекс (LUBAC), включающий *HOIP*, *HOIL-1* и *SHARPIN*, часто связывает линейные (*Met1*) убиквитиновые цепи в каноническом пути *NF-κB* со многими белками-мишенями [73]. Линейная убиквитин-специфическая деубиквитиназа *OTULIN* контролирует функцию LUBAC. У мышей и людей иммунная дисрегуляция отмечается при наличии дефектов в процессах линейной убиквитинизации и деубиквитинизации *K63* [13].

Каталитической субъединицей комплекса сборки цепи линейной убиквитинизации (LUBAC) является *HOIP*, который важен для сигналинга *NF-κB* и, следовательно, для обеспечения функций врожденного и адаптивного иммунитета. На сегодняшний день с такими симптомами, как иммунодефицит, системная лимфангиэктазия и аутовоспаление, дефицит *HOIP* выявлен только у одного человека [186].

Дефицит *HOIP* также проявляется лимфангиэктазиями в системном отеке, желудочно-кишечном тракте и гипоальбуминемией, которая может вызвать мальабсорбцию. Молекулярные исследования установили, что фибробласты и В-клетки пациентов, не восприимчивые к иммунным стимулам и не способные к устойчивому увеличению регуляции активности *NF-κB*, имеют иммунодефицитный фенотип, наблюдаемый у пациента. По сравнению с иммунными ответами в фибробластах, *HOIP* и *HOIL1*-дефицитные мононуклеарные клетки периферической крови (PBMCs) были высокореактивны на стимуляцию *IL-1* и экспрессировали провоспалительные цитокины *IL-6* и *MIP-1a* [186]. Дефицит белка *HOIL-1* в клетках пациентов привел к снижению уровня фосфорилирования киназы *IKK*, замедлению деградации интегрин альфа-*IIb* и снижению убиквитинирования *NEMO* в ответ на стимуляцию *TNF* или *IL-1β*, а более низкий уровень активации *NF-κB* в клетках пациентов был связан со снижением транскрипционной активности *NF-κB* [13, 193]. Линейный комплекс сборки убиквитиновых цепочек (LUBAC) имеет активность в сигнальном пути *NF-κB* и связывает

линейные полиубиквитиновые цепочки с уникальными остатками Lys белка NEMO. Фибробласты человека с дефицитом NOD1 проявляют ослабленную активацию NF-κB, что приводит к слабой транскрипции генов, контролируемых NF-κB, и развитию цитокинов в ответ на TNF и IL-1β. Эти данные согласуются с результатами исследований клеток мыши с нокаутом или нокдауном гена *RBCK1* [33].

#### **1.4 - Моногенные, мультигенные и аллогенные дефекты при врожденной нейтропении**

Нейтропения - распространенное заболевание, с которым педиатры сталкиваются регулярно, и это серьезная проблема для здоровья. При нейтропении абсолютное число полиморфных ядерных клеток снижается, что делает организм более восприимчивым к инфекциям. В результате инфекции часто бывают исключительно тяжелыми или возникают с необычно высокой частотой. Нейтрофилы являются важным компонентом врожденного иммунитета и важнейшим продуктом кроветворения. Количество нейтрофильных гранулоцитов в периферической крови используется для определения тяжести нейтропении. В большинстве случаев этиология нейтропении является ятрогенной и хорошо известна лечащему врачу. Аллогенные или аутоантитела являются второй наиболее распространенной причиной нейтропении. Хорошо известна способность некоторых вирусных инфекций индуцировать нейтропению [133, 144].

Помимо отклонений в развитии от нормального количества лейкоцитов в детском возрасте и вариаций в среднем количестве нейтрофилов у людей разных рас, абсолютное количество нейтрофилов ниже 1500/мкл принято расценивать как нейтропению. Это принимается в качестве начального симптома. Существует несколько форм нейтропении. Нейтропения классифицируется как легкая, когда абсолютное количество нейтрофилов составляет 1000-1500/мкл, умеренная, когда абсолютное количество

нейтрофилов составляет 500-1000/мкл, и тяжелая, когда абсолютное количество нейтрофилов составляет менее 500/мкл. Нейтропения является распространенным гематологическим состоянием при множественных ПИД с различными генетическими проблемами - от врожденного фагоцитарного дефицита до осложненного иммунодефицита, и может использоваться для тестирования на острые инфекции [4, 90].

Наследование врожденной нейтропии, связанной с ПИД, сложное, начиная от изолированной тяжелой врожденной нейтропии до умственной отсталости, аномалий органов, лицевого дисморфизма и депигментации кожи. Фагоцитарные врожденные аномалии делятся на две категории по классификации IUIS: дефицит количества фагоцитов (нейтропения) и дисфункция фагоцитирующих клеток [4, 89]. Хроническая или интермиттирующая нейтропения может встречаться при самых разных наследственных нарушениях иммунной системы, в частности, при различных формах дефицита антител, ретикулярной дисплазии, WHIM синдроме и других заболеваниях. Основные патофизиологические причины хронической тяжелой нейтропии у пациентов с ПИД включают аномальную дифференцировку клеток костного мозга, неправильный выброс гранулоцитов из костного мозга, повышенный апоптоз или повышенную гибель гранулоцитов периферической крови [4, 89, 144, 233]. Исследование костного мозга показало, что у большинства пациентов созревание миелопоэза останавливается на уровне промиелоцитов, что приводит к снижению количества нейтрофилов, но повышению количества атипичных промиелоцитов [223]. Такие инфекционные состояния, как отит, гингивит, кожные инфекции, пневмония, глубокие абсцессы и сепсис у таких пациентов начинаются в неонатальном периоде и, без соответствующего лечения, остаются на всю жизнь. Кроме того, у пациентов с тяжелой комбинированной иммунной недостаточностью (ТКИН) повышен риск развития лейкемии. При

этом причиной ТКИН могут являться варианты самых различных генов [90, 49, 187].

Причинами врожденной нейтропении могут быть дефекты созревания и функции нейтрофилов, синдромы иммунной дисрегуляции (различные гемофагоцитарные лимфогистиоцитозы), некоторые тяжелые комбинированные иммунодефициты (например, ретикулярный дисгенез (дефект *AK2*) и дефект активации *PAC2*), а также первичная аутоиммунная нейтропения на разных стадиях развития нейтрофилов. Обычно пациентам с врожденной нейтропенией требуется антимикробная профилактика и лечение гранулоцитарным колониестимулирующим фактором, а радикальное излечение невозможно без трансплантации гемопоэтических стволовых клеток [188]. В настоящее время существует более 30 врожденных ошибок иммунитета (или первичных иммунодефицитов), при которых наблюдается нейтропения, и хотя каждое состояние в отдельности встречается редко, общая распространенность этих состояний в популяции серьезная, и для назначения адекватной терапии необходима хорошая и своевременная диагностика [104, 227].

Известно, что мутации следующих генов ПИД приводят к развитию нейтропении и врожденной нейтропении: *ELANE*, *HAX1*, *G6PC3*, *WASP*, *JAGN1*, *GF11*, *SEC61A1*, *CSF3R*, *LYST*, *AP3P1*, *TCIRG1*, *VPS45*, *LAMTOR2*, *SBDS*, *DKC1*, *SLC37A4*, *BTK*, *CD40*, *CXCR4*, *AK2*, *GATA2*, *STK4*, *RMRP* и *VPS13B*.

Классическая врожденная нейтропения зависит от функции эластазы. Дефекты эластазы приводят к тяжелой врожденной нейтропении (SBH) типа 1 (дефицит *ELANE*), дефициту *GF11* (SBH 2), дефициту *HAX1* (SBH 3 или болезнь Костманна), дефициту *G6PC3* (SBH 4), дефициту *VPS45* (SBH 5), гликогенозу типа 1b (дефицит *G6PT1*), X-сцепленной нейтропении/миелодисплазии (мутация *WAS* GOF), дефициту *P14/LAMTOR2*, синдрому Барта (3-метилглутаконовая ацидурия, тип II)

(дефицит TAZ, X-сцепленный), синдрому Коэна (дефицит VPS13 B), синдрому Клеркуцио (дефицит USB1), дефициту JAGN1, 3-метилглутаконовой ацидурии (дефицит CLPB), дефициту рецептора G-CSF (CSF3R), дефициту SMARCD2, дефициту специфических гранул нейтрофилов (CEBPE), синдрому Швахмана-Даймонда (вызванному дефектами по крайней мере 3 генов, SBDS, DNAJC21 и EFL1), дефициту NYOU1 и дефициту SRP54 [188].

ELANE (OMIM #130130) кодирует эластазу нейтрофилов, сериновую протеазу, экспрессируемую в миеломоноцитарных клетках и их предшественниках. Нейтрофильная эластаза в основном вырабатывается на промиелоцитарной стадии созревания нейтрофилов и сохраняется в нейтрофильных гранулах азурофилов, которые участвуют в уничтожении микроорганизмов [10, 96]. Однако даже при мутации только этого белка наблюдается различная клиническая картина врожденной нейтропении, и точный патогенез каждого состояния остается невыясненным [68, 139, 137, 190].

Основные механизмы нейтропении при дефекте эластазы нейтрофилов сводятся к стрессу эндоплазматического ретикулума (ответ развернутого белка), когда неправильно сформированная эластаза накапливается в эндоплазматическом ретикулуме и приводит к активации сигналов смерти [81, 96]. Известно, что ELANE становится наиболее распространенным белком на стадии развития промиелоцитов, достигая миллимолярных концентраций в нейтрофилах, что подтверждает теорию о том, что накопление неправильно сформированного белка может вызвать дефицит шаперонных белков, который активирует сигналы смерти и апоптоз незрелых нейтрофилов. [31, 33]. Другое дело, что мутировавшая эластаза нейтрофилов блокирует дальнейшую дифференцировку, приводя к нейтропении [137]. Кроме того, полиморфизм ELANE p.G185R связан с нарушением дифференцировки нейтрофилов и снижением экспрессии генов, кодирующих критические гемопоэтические

транскрипционные факторы, белки клеточной поверхности и белки гранул нейтрофилов [96,137].

Ген Т-клеточного иммунного регулятора 1 (*TCIRG1*) кодирует субъединицу большого белкового комплекса, известного как вакуолярная Н+АТФаза (V-АТФаза). Этот белковый комплекс действует как насос для перемещения протонов через мембрану. Это перемещение протонов помогает регулировать рН клеток и окружающей их среды. Зависимое от V-АТФазы подкисление органелл необходимо для таких внутриклеточных процессов, как сортировка белков, активация зимогена и рецептор-опосредованный эндоцитоз. V-АТФаза состоит из цитозольного домена V1 и трансмембранного домена V0. В результате альтернативного сплайсинга образуется множество вариантов транскрипта. Мутации в этом гене связаны с инфантильным злокачественным остеопетрозом и тяжелой врожденной нейтропенией.

Ген *TCIRG1* у человека, прежде всего, связывают с аутосомно-рецессивным остеопетрозом. Молекулярный анализ выявил шесть новых генов (*TNFSF11*, *TNFRSF11A*, *CLCN7*, *OSTM1*, *SNX10* и *PLEKHM1*), связанных с аутосомно-рецессивным остеопетрозом у человека. Более половины всех пациентов с аутосомно-рецессивным остеопетрозом имеют мутации в гене *TCIRG1* [20, 202]. Исследования показали, что мыши с нарушением функции гена *Atp6b1* развивают тяжелый остеопетроз [23, 41]. Несмотря на огромный прогресс в понимании механизмов развития остеопорозных заболеваний, генетическая основа 30% случаев остается неясной [148]. Согласно исследованиям, мутации *TCIRG1* включают миссенс, нонсенс, небольшие делеции/инсерции, мутации сплайс-сайта, значительные геномные делеции и интронные мутации [26, 34, 60, 138]. Аутосомно-рецессивный остеопороз 1 типа вызывается мутациями в гене *TCIRG1*, что приводит к нарушению резорбции кости и аномальному скоплению плотной костной ткани. Это может приводить к переломам, недостаточности костного мозга,

неврологическим проблемам и иммунодефициту, что, в конечном итоге, может привести к преждевременной смерти. Эта проблема может быть выявлена уже в возрасте 10 дней. Наиболее распространенными признаками заболевания являются патологические переломы, недостаточность костного мозга и компрессия черепных нервов, которые вызваны нарушением структуры костной ткани, метаболизма и недостаточным расширением фораминального отверстия черепных нервов [26]. Высокая плотность костной ткани может возникнуть из-за нарушения резорбции костей, вызванного дисфункцией остеокластов, что может привести к серьезным аномалиям. Некоторые дефекты могут возникнуть на ранних стадиях развития плода, такие как микроцефалия, прогрессирующая глухота, слепота, гепатоспленомегалия и тяжелая анемия. Вторичная гипертензия черепных нервов может часто приводить к глухоте и слепоте [198].

Существуют многочисленные примеры мультигенных (или полигенных) причин врожденной нейтропении (ВН), при которых варианты мутации в нескольких генах могут способствовать формированию одинаковых или разных фенотипов этого заболевания [48]. Врожденная нейтропения является более мультигенной по сравнению с другими ПИД, и одной из возможных причин этого является то, что мультигенная природа ВН является результатом сложных взаимодействий между генами.

Понимание механизма их взаимодействия должно помочь в получении представления о патофизиологии ПИД, что позволит улучшить диагностику и изменить подходы к лечению. Однако, механизмы, лежащие в основе сети белок-белковых взаимодействий (PPI) генов ВН, остаются невыясненными; отсутствует системный уровень интерпретации патогенетических факторов. С накоплением в последнее время новых данных по экспрессии генов при ВН [16, 50], а также использованием современных вычислительных методов [97], появилась возможность определения генов-кандидатов врожденной нейтропении. Использование методов системной биологии и биоинформатики

ускорит и повысит точность выявления новых генов СН, что позволит глубже понять патогенетические механизмы этой патологии. Кроме того, это недорогой и быстрый метод, который при внедрении в клиническую практику поможет в диагностике случаев врожденной нейтропении с неизвестной генетической причиной.

### **1.5 - Синдром Хеннекама, фенотип и генотип**

Синдром Хеннекама - аутосомно-рецессивное заболевание, и это одна из самых редких форм первичного иммунодефицита, который в своём фенотипе имеет пороки развития лимфатической системы [87].

Первичная лимфедема-лимфангиэктазия, которая является основой синдрома Хеннекама, обусловлена дефектами развития и/или функционирования лимфатической системы. Она может поражать любую часть тела с преобладанием нижних конечностей, кишечника, брюшной и плевральной полостей. Кроме того, у таких пациентов часто уплощены черты лица, имеется широкая переносица, гипертелоризм, эпикантус, могут наблюдаться и другие аномалии [25]. В настоящий момент с первичной лимфедемой (либо изолированной, либо как часть синдрома) связали 27 различных генов. Считалось, что общим сигнальным путем в патогенезе лимфедемы является сигналинг рецептора VEGFR3. Однако на его долю приходится только треть всех случаев первичной лимфедемы, что подчеркивает существование дополнительных генетических факторов. Синдром лимфангиэктазии-лимфедемы Хеннекама может быть вызван мутациями в гене *CCBE1* (25% случаев), а также в гене *FAT4* и *ADAMTS3*, и при этом каждый из них так или иначе влияет на на сигнальные пути VEGF-C / VEGFR-3 [37, 67, 136, 226].

Синдром Хеннекама 1 типа, также известный как *CCBE1*-ассоциированный синдром Хеннекама, был первоначально описан голландским врачом Раулем Хеннекамом в 1989 году [25]. Основным

молекулярным механизмом лимфедемы при синдроме Хеннекама 1 типа является снижение способности мутировавшего ССВЕ1 (коллаген и кальций-связывающий белок 1, содержащий домен эпидермального фактора роста) ускорять и концентрировать активацию первичного лимфангиогенного фактора роста VEGF-C [226].

Для синдрома Хеннекама 2 типа причиной является гомозиготная или сложная гетерозиготная мутация в гене *FAT4* на хромосоме 4q28. Интересно, что мутация в гене *FAT4* может вызывать синдром Ван Мальдергема (VMLDS2), другое заболевание, при котором некоторые симптомы встречаются и при фенотипе синдрома Хеннекама [83].

В исследовании 2017 года группа авторов под руководством P. Brouillard выявили синдром Хеннекама 3-го типа, при котором обнаружилась гетерозиготная мутация в гене *ADAMTS3* на хромосоме 4q13. Что еще более важно, исследователи подчеркнули тесную функциональную связь между белками *ADAMTS3* и ССВЕ1 при активации молекулы VEGFR3, которая является краеугольным камнем для дифференциации и функционирования лимфоидных эндотелиальных клеток [124]. Однако мутации в этих генах обнаруживаются только у некоторых пациентов, и генетическая этиология большинства пациентов с синдромом Хеннекама остается неуточненной, в основном, потому, что синдром генетически гетерогенен.

Знания о генетической причине синдрома позволили выявить вовлеченность сигнального пути mTOR (mammalian target of rapamycin) и найти возможную цель для терапии, в частности, ингибиторы mTOR (рапамицин и его аналоги). mTOR - это белок, который играет роль в регуляции клеточного роста и метаболизма, и его дисрегуляция была вовлечена в патогенез нескольких генетических заболеваний, включая синдром Хеннекама. Ингибиторы mTOR останавливают прогресс лимфедемы и лимфэктазии, а также оказывают противовоспалительное и антифибротическое действие, однако не избавляют пациентов от уже

имеющихся нарушений лимфатической системы и их эффективность в целом не высокая. Поэтому изучение генетической природы и патогенеза синдрома Хеннекама поможет выявить более эффективные цели для терапии.

### **1.6 - Проблемы исследования первичных иммунодефицитов**

Существует множество фактов, которые делают первичный иммунодефицит сложной группой болезней как для практикующего врача, так и для научного исследователя [3, 5, 8].

В ближайшие годы будет крайне важным обеспечение всеобщего доступа к многочисленным научным достижениям и создание устойчивого механизма, позволяющего своевременно их учитывать в будущих разработках [118]. Хотя секвенирование следующего поколения является революционным способом диагностики ПИД, оно недоступно во многих странах, особенно в странах с низким уровнем дохода. Поэтому существует реальная задача достижения общедоступности этого метода диагностики и снижения стоимости генетических исследований. Необходимо также сделать легкодоступными другие экспресс-тесты для скрининга на синдромы дефицита антител, которые потенциально могут облегчить доступ к тестированию в отдаленных районах стран с ограниченными ресурсами. Кроме того, скрининг новорожденных на ТКИН и другие лимфопении представляет собой надежду на раннюю диагностику и лечение ПИД, но его необходимо внедрять более широко в государственных и частных медицинских учреждениях, поскольку он позволяет выявить ПИД на ранней стадии болезни [9]. Вслед за США, несколько европейских стран начали пилотные исследования по внедрению неонатального скрининга, или уже ввели его в повседневную практику как государственный проект, как это сделано в Российской Федерации [7].

Неравенство в доступе к лечению и уходу за пациентами с ПИД, включая вопросы возмещения расходов, доступность, создание организационной структуры доступа к медицинской помощи и т.д. Кроме

того, необходим количественный анализ потребностей в уходе в различных регионах мира - особенно в Азиатско-Тихоокеанском регионе - для поддержки усилий по увеличению государственных инвестиций в лечение и исследования ПИД [56].

Поскольку о ПИД еще слишком мало известно широкой общественности, они часто воспринимаются как «экзотические» болезни. Улучшение осведомленности, понимания и своевременное распознавание новых форм позволит изменить жизнь многих пациентов в будущем. Необходимо продолжать совместную работу по поддержанию поставок лекарственных препаратов из плазмы по всему миру, в том числе в периоды, когда системы здравоохранения испытывают трудности с поставками крови и плазмы.

Открытие новых, передовых методов лечения является постоянной необходимостью, поскольку открываются новые типы заболеваний и появляется лучшее понимание патогенеза болезни.

Вторым камнем преткновения является то, что фенотипические изменения, связанные с ПИД, как правило, очень разнообразны. Например, в случае пациентов с синдромом Вискотта-Олдрича точный характер дефекта гена, например, миссенс- или нонсенс-мутация, точное расположение аномалий сайта сплайсинга, может значительно изменить фенотип синдрома. Проявления мутаций этого гена могут быть от очень тяжелых до легких, например, X-сцепленная тромбоцитопения, В-клеточная лимфома, частые бактериальные и грибковые инфекции, экзема, низкий уровень тромбоцитов или нейтропения. Диагностика ПИД обычно требует углубленного анализа клинических проявлений в сочетании с оценкой анамнеза и генеалогии.

Важнейшей проблемой исследователей и клиницистов является сложность поиска информации. Базы данных в открытом доступе содержат мало образцов, разнообразие которых, по сравнению с данными по онкологии, также ограничено. Существует очень мало информационных ресурсов,

которые связывают клинические описания и функциональные геномные данные, белок-белковые взаимодействия и сигнальные пути. Примерами специализированных баз данных являются UniProt, IntAct, STRING и KEGG.

Несколько баз данных, реестров, баз знаний, инструментов прогнозирования и экспертных систем быстро развиваются под влиянием требований диагностики. Согласно статье, опубликованной Richardson A.M. и соавт., 2018, спектр заболеваний еще более уточняется, благодаря расширению иммунологических, генетических и эпигенетических знаний, тщательное применение этих диагностических инструментов и биоинформатических методов поможет не только понять эти сложные расстройства, но и позволит реализовать персонализированные терапевтические подходы для лечения заболеваний [61]. Крина Самаргитян и др. в 2009 году объяснили, что эффективные инструменты для совершенствования диагностики ПИД, можно разделить на следующие семь категорий (*рисунок 3*) [183].

Во-первых, основные ресурсы по ПИД предоставляют большое количество информации, ранжированной по уровням от генов до белковых структур, от моделей заболеваний до определенных групп диагнозов и так далее.

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

В-третьих, имеются лабораторные критерии и соответствующие инструменты.

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

В-шестых, доступны инструменты биоинформатики для прогнозирования или определения приоритетности новых генов-кандидатов

ПИД, которые также используются для принятия решений в диагностике ПИД – и это седьмая категория.



Рисунок 3 – Схематическая группировка биоинформационных ресурсов и инструментов, предоставляющих информацию о первичных иммунодефицитах [183]

Рациональный подход к выбору и интерпретации геномного анализа при первичных иммунодефицитах облегчает интеграцию клинических данных с иммунологическими и генетическими данными для установления диагноза [61].

В настоящее время проводятся геномные ассоциативные исследования распространенных заболеваний на предмет выявления распространенных причинных вариантов с низкой пенетрантностью. Часть из них будет изменять белковые последовательности, наиболее распространенными из которых

является несинонимичный однонуклеотидный полиморфизм (nsSNP). Преимуществом будет возможность прогнозирования функционального влияния nsSNP на структуру и функцию белка как для окончательной идентификации причинного варианта в ассоциированном с заболеванием хромосомном регионе, так и для дальнейшего функционального анализа nsSNP и ассоциированного с заболеванием белка [72].

Известно, что несинонимичные SNP (nsSNP) изменяют функцию белка и имеют больше шансов вызвать заболевание у человека. Последние исследования nsSNP с использованием вычислительных подходов показывают потенциальное влияние мутаций на понимание молекулярных механизмов различных заболеваний [17, 41].

Исследования продемонстрировали, что биоинформатический анализ профилей экспрессии генов имеет значительные перспективы в раскрытии потенциальных ключевых генов и путей развития сложных заболеваний [30, 180, 218].

Большое количество различных ПИД создает трудности для диагностики, в том числе на клиническом уровне. Кроме того, многие болезни настолько редки, что невозможно найти достаточное количество семей для анализа. Скрининг и раннее выявление ПИД является серьезной проблемой для врачей. За последние несколько лет с помощью высокопроизводительного секвенирования было получено большее количество известных в настоящее время генетических дефектов. Выявление новых генов-кандидатов для ПИД позволит в дальнейшем определить приоритетность генов при их подтверждении у пациентов с ПИД, у которых точный причинный ген еще не определен.

В 2009 году S. Keerthikumar и др. использовали машинный метод опорных векторов для классификации всех генов человека на гены ПИД и гены, не относящиеся к ПИД. Принцип этой классификации был основан на расчете показателя уверенности для каждого гена-кандидата на ПИД на

основе 69 признаков, наблюдаемых для 148 известных на тот момент генов ПИД [163]. На основании поиска литературы мы обнаружили, что внимание научных исследователей, занимающихся идентификацией ПИД-генов, также было сосредоточено на интеграции функциональных аннотаций из онтологии генов (GO) и составлении наборов данных сети белок-белковых взаимодействий. В 2018 году Guojun Liu и его коллеги идентифицировали 172 кандидатских гена для общей варибельной иммунной недостаточности (ОВИН), которые имеют сходные биологические функции с известными генами ОВИН, а восемь генов недавно были объявлены генами, связанными с ОВИН [38, 94], C. Ortutay et al., 2008, предсказали 26 генов-кандидатов ПИД путем анализа свойств сети белок-белковых взаимодействий (PPI) всех известных генов иммунной системы человека и их онтологии (GO). Кроме того, на основе расчета так называемого биологического расстояния (означающего функциональную взаимозависимость) было предсказано 3110 генов-кандидатов болезней, относящихся к ПИД [133, 159].

Исследователи, желающие принять участие в изучении проблемы ПИД, каждый раз сталкиваются с вопросом: являются ли ПИД мультигенными или моногенными болезнями? Первоначально ПИД считали врожденными и моногенными болезнями, которые подчиняются принципам менделевского наследования [174, 178]. Моногенные болезни являются результатом изменений в одном гене, которые происходят во всех клетках организма. Однако прогресс в секвенировании ДНК способствовал выявлению мультигенных и соматических причин ПИД, наблюдается широкая фенотипическая варибельность этих заболеваний [80, 211]. Понимание того, что большинство ПИД имеют мультигенную природу, является первым шагом в понимании патогенеза всех заболеваний. Согласно мультигенной концепции, ПИД являются результатом сложных взаимодействий между генами. Исходя из этого, ученые попытались найти «биологическое расстояние» между генами ПИД и другими протеинкодирующими генами

человека, было обнаружено, что гены ПИД, по сравнению с другими генами человека, обычно расположены в центральном узле геномной сети человека и более тесно взаимодействуют друг с другом [107]. Кроме того, гены ПИД образуют несколько тесно связанных субкластеров, большинство из них имеют как минимум одного функционально близкого соседа среди широкого спектра биологических механизмов [38, 107, 201]. Раскрытие этих взаимосвязей может обеспечить более глубокое понимание разнообразия генетических путей, лежащих в основе ПИД, что, в свою очередь, поможет раскрыть новые возможности для разработки лекарств и терапевтических подходов.

Генетические изменения могут приводить как к полной или частичной утрате белка (нонсенс-варианты), снижению его функции (LOF – lost of function), так и усилению функции (GOF – gain of function). Это справедливо для любых белков, в том числе ключевых молекул, участвующих в иммунном ответе. На сегодняшний день известны патологические варианты более чем 485 генов, вошедших в классификацию врожденных ошибок иммунитета (ВОИ) [89], но подробного описания ожидает более значительное количество генов, исследование которых находится на более ранних этапах изучения. Важным фактором, усложняющим поиск причиннозначимых генетических изменений при большом количестве болезней, является огромная фенотипическая гетерогенность врожденных ошибок иммунитета, включая дефициты антител, популяций и субпопуляций лимфоцитов, дефицит системы комплемента, аутоиммунные и аутовоспалительные патологии, лимфопролиферативные синдромы, костномозговую недостаточность и болезни иммунной дисрегуляции. В большой когорте иммунодефицитов можно наблюдать сочетанные иммунозависимые процессы, например, аутоиммунные и/или иммунную дисрегуляцию, особенно в тех случаях, при которых генетические ошибки приводят к изменениям молекул,

регулирующих иммунный ответ или участвующих в обеспечении процессов иммунной толерантности [146].

Несмотря на то, что до недавнего времени ВОИ считали редкими болезнями, а отдельные генетические нарушения могут встречаться нечасто, в совокупности они могут поражать значительное число людей. Более того, в результате улучшения диагностики, благодаря развитию технологий секвенирования последующих поколений (NGS), за последние годы констатируемая распространенность первичных иммунодефицитов (ПИД) увеличилась примерно до 40 на 100 000 населения [89, 164].

При этом необходимо учитывать сложные взаимосвязи всех генов и белков в организме, поскольку простая генотип-фенотипическая корреляция очень часто остается не востребованной – пациенты с дефектом одного и того же гена могут иметь принципиально отличную фенотипическую презентацию [11].

При наличии предположения о существовании связи фенотипа ПИД с геном, который ранее не был описан с этой точки зрения, необходимо проведение тщательных функциональных исследований, подтверждающих либо опровергающих эту ассоциацию, что позволит провести констатирование новой болезни, либо новых ее фенотипов. Исследование значений генетических изменений для обеспечения функционирования иммунной системы имеет уникальное преимущество, поскольку иммунные клетки легкодоступны, обычно требуется простой забор крови для получения соответствующих клеток, в отличие от мутаций, затрагивающих другие труднодоступные ткани [54].

Для того, чтобы доказать причинно-следственную связь, исследования должны продемонстрировать значимость конкретного патологического варианта гена с нарушениями конкретного иммунного процесса, приводящего к соответствующему фенотипу болезни. То есть должна быть проведена функциональная валидация, которая включает оценку количества и функции

белков, анализ сигнальных путей и биологический механизм реализации патологии [89, 164].

Определить причинно-следственную связь новых мутаций проще, когда выявлено несколько неродственных семей с аналогичными генетическими вариантами и фенотипом. Однако новые заболевания могут иметь единичные описания. Некоторые ограничения исследований с участием одного пациента заключаются в недостаточной статистической мощности или наличии сбивающих с толку генетических модификаторов, что снижает возможность определения конкретного варианта как мутации, приводящей к болезни. Преодолевают эти ограничения экспериментальное моделирование генетических изменений в клеточных линиях или на животных моделях. Для подтверждения нового гена, патологические варианты которого могут приводить к развитию ПИД у одного человека, необходимо соблюдение следующих критериев: найденный генотип не может быть у лиц, не имеющих клинического фенотипа; экспериментальные исследования должны продемонстрировать, что этот вариант повреждает, разрушает или изменяет функцию или экспрессию генного продукта; причинно-следственная связь между генотипом и клиническим фенотипом должна быть подтверждена на соответствующей клеточной или животной модели [80].

Следует отметить, что в настоящее время биоинформатика становится все более заметной частью различных областей биологии, включая молекулярную биологию, статистику и генетику, которые играют решающую роль в анализе экспрессии и регуляции генов и белков [195]. Изучение влияния однонуклеотидных полиморфизмов - SNP - в кодирующей части генома, непосредственно влияющих на структуру белков, находится в центре внимания подавляющего большинства научного сообщества. По оценкам различных исследователей, около 90% генетических вариаций у человека обусловлены однонуклеотидными полиморфизмами. Они определяются с частотой от 1% до 5% в зависимости от исследуемой патологии. Значения

частот распространения аллелей важны для определения актуальности SNP в конкретной популяции и для понимания потенциального влияния этого SNP на восприимчивость к заболеваниям или другим интересующим характеристикам [179]. В рамках проекта НарМар, который является международным коллаборативным усилием, направленным на идентификацию общих генетических вариаций среди людей, было описано и генотипировано более 4 миллионов образцов ДНК. Это позволило не только валидировать SNP и оценить частоту их аллелей в общей популяции, но также оценить степень неравновесия связи между ними. Более того, технологии генотипирования SNP в последнее время продвинулись настолько, что сотни тысяч SNP могут быть типированы у тысяч людей, например, с использованием метода случай-контроль. Следовательно, открытие причинных вариантов для распространенных заболеваний будет ускоряться, и было бы полезно, если бы функциональные эффекты SNP можно было предсказать биоинформатически, чтобы направить функциональные исследования и сузить круг лучших кандидатных SNP в областях генома, которые демонстрируют высокую степень неравновесия [46]. Именно поэтому наука биоинформатика становится неотъемлемой частью современных исследований.

Наиболее идентифицируемой категорией SNP является небольшая доля мутаций (менее 1%), изменяющая последовательность белка, и, как правило, это несинонимичные замены (nsSNP). Инструменты предсказания nsSNP используются для прогнозирования потенциального структурного и функционального воздействия, вызванного этими вариантами. Для того, чтобы более точно оценить структурное воздействие, вызванное изменениями в аминокислотной последовательности, требуется проведение биоинформатического анализа и моделирования белковой структуры с учетом внесенных изменений в последовательность аминокислот. Знание трехмерной структуры генного продукта оказывает большую помощь в предсказании и

понимании его функции, его роли во внутриклеточных процессах и в формировании патологии, можно провести моделирование молекулярной динамики, позволяющее наблюдать изменения многих параметров, таких как стабильность и гибкость белка. Междисциплинарная модель (биоинформатика, патофизиология, генетика и иммунология) постепенно становится основной тенденцией в развитии технологий для клинических исследований [46].

При выявлении генов-кандидатов для различных видов патологии необходима их верификация, для чего требуется использование не только клинических данных, но и экспериментальных, а также анализ коэкспрессии генов, активации биологических сигнальных путей, белок-белкового взаимодействия и оценка функционирования измененного белка в моделировании.

Важной частью проводимых исследований являются методы интеграции профилей экспрессии и данных о белок-белковом взаимодействии (PPI). Методы биоинформатики используются для изучения дифференциальных механизмов взаимодействия белков во всех линиях иммунных клеток, транскрипционных активаторов и модулей, которые анализируются в контексте примеров, полученных путем кластеризации сети PPI. Результаты таких исследований демонстрируют, что для создания гипотез о механизмах, лежащих в основе дифференциации и дифференциальной функциональной активности по всем линиям иммунных клеток может быть использована интеграция сетей белковых взаимодействий с наиболее полной базой данных профилей экспрессии генов иммунных клеток. Сравнительный анализ выявляемых различий между больным и здоровым состоянием организма помогает получить патогенетическую характеристику иммунозависимых заболеваний и, в конечном итоге, приводят к разработке новых куративных методов коррекции патологии.

В настоящее время исследования дифференциально экспрессированных генов определяют одно из особых научных направлений, при котором предполагается выявление генов, дифференциально экспрессирующихся при заболеваниях. В фармацевтических и клинических исследованиях результаты оценки дифференциально экспрессированных генов могут быть ценными объектами для определения биомаркеров-кандидатов, терапевтических мишеней и генных сигнатур для диагностики. Хотя конкретные изменения экспрессии генов не всегда приводят к последующей биологической активности, такие данные, тем не менее, могут быть объединены с другими биологическими данными, и, располагая возможностью обеспечения высокой пропускной способности для создания комплексных анализов, таких как построение целевого ландшафта болезни [123, 213], может стать незаменимым инструментом исследований [63].

В нашей работе, направленной на поиск значимых патофизиологических механизмов формирования отдельных видов иммунозависимой патологии, с помощью исследовательских методов биоинформатического анализа при нарушениях, характерных для первичных иммунодефицитов, в качестве моделей иммунозависимой патологии были выбраны различные объекты с врожденными иммунными нарушениями, включая врожденную нейтропению, аутовоспалительный синдром дефицита *RBSK1* и синдром Хеннекама.

### **Список публикаций по материалам данной главы**

1. Ретроспективный анализ случаев первичных иммунодефицитов у детей с врожденными пороками сердца / С.С. Дерябина, Д.А. Черемохин, И.А. Тузанкина, Х. Шинвари, // Российский иммунологический журнал. – 2020. – Т. 23, № 4. – С. 505-514. (RSCI, ВАК К1).
2. Классификация врожденных ошибок иммунитета человека,

обновленная экспертами комитета международного союза иммунологических обществ в 2019 году / М.А. Болков, И.А. Тузанкина, Х. Шинвари, Д.А. Черемохин // Российский иммунологический журнал. – 2021. Т. – 24 (1). – С. 7-68. (RSCI, ВАК К1).

3. Роль врожденных ошибок иммунитета в группе детей с летальными исходами на первом году жизни / Д.А. Черемохин, И.А. Тузанкина, В.А. Черешнев, М.А. Болков, Х. Шинвари, // Российский иммунологический журнал. – 2022. – Т. 25 (4). – С. 555-560. (RSCI, ВАК К1).

4. Analysis of the TREC and KREC Levels in the Dried Blood Spots of Healthy Newborns with Different Gestational Ages and Weights / D.A. Cheremokhin, K. Shinwari, S.S. Deryabina, M.A. Bolkov, I.A. Tuzankina, D.A. Kudlay // Acta naturae. –2022. –V. 14 (1). – P. 101–108. (RSCI, ВАК К1).

## ГЛАВА 2 - МАТЕРИАЛЫ И МЕТОДЫ, ИСПОЛЬЗОВАННЫЕ В РАБОТЕ

### 2.1 – Материалы исследования

Сбор данных был осуществлён из открытых источников данных, различных баз данных вариантов генов, а также двух ослепленных результатов секвенирования пациентов Свердловской области, которые были предоставлены для изучения Институту иммунологии и физиологии УрО РАН, ранее одобрены этическим комитетом и опубликованы.

Для выполнения задачи исследования особенностей патогенеза дефицита *RVCK1* были использованы два датасета (набора данных) из базы данных NCBI GEO, а именно: GSE40561, включающий данные, полученные по результатам исследования цельной крови, собранной у 2 пациентов с болезнью CINCA/NOMID, 5 пациентов с синдромом Макл-Уэллса, двух пациентов с дефицитом мевалонаткиназы, 1 пациента с дефицитом *RVCK1* и 41 здорового ребенка (для проведения сравнительного анализа); GSE31064, включающий данные, полученные из клеток фибробластов кожи 2 пациентов с дефицитом *RVCK1*, одного пациента с дефектом *MYD88*, одного пациента с синдромом NEMO и 3 здоровых пациентов (из контрольной группы).

Для выполнения задачи по исследованию генов-кандидатов, ассоциированных с врожденной нейтропенией, использовались наборы данных из NCBI GEO: GSE142347 (пациенты с врожденной нейтропенией - 93 женщины и 95 мужчин, и 193 контрольных пациента); GSE6322 (семейный случай - 2 здоровых родителя и 2 ребенка с врожденной нейтропенией). Список из 442 известных на момент исследования генов ПИД (и микроделаций), включая 31 ген, ассоциированный с врожденной нейтропенией, были получены с сайта Европейского общества по иммунодефициту (<https://esid.org>): *CD3D*, *CD3E*, *CD3Z*, *CORO1A*, *IL2RG*, *IL7R*, *JAK3*, *LAT*, *PTPRC*, *ADA*, *AK2*, *DCLRE1C*, *LIG4*, *NHEJ1*, *PRKDC*, *RAC2*, *RAG1*, *RAG2*, *B2M*, *BCL10*, *CARD11*, *CD3G*, *CD40 (TNFRSF5)*, *CD40LG (TNFSF5)*, *CD8A*, *CIITA*,

*DOCK2, DOCK8, FCHO1, ICOS, ICOSLG, IKBKB, IKZF1, IL21, IL21R, ITK, LCK, MALT1, MAP3K14, MSN, POLD1, POLD2, REL, RELB, RFX5, RFXANK, RFXAP, RHOH, STK4, TAP1, TAP2, TAPBP, TFRC, TNFRSF4, TRAC, ZAP70, ZAP70, ARPC1B, WAS, WIPF1, ATM, BLM (RECQL3), CDCA7, DNMT3B, GINS1, HELLS, LIG1, MCM4, NBS1, NSMCE3, PMS2, POLE1, POLE2, RNF168, ZBTB24, 11q23del, 22q11.2, CHD7, Del10p13-p14, FOXN1, FOXN1, SEMA3E, TBX1, EXTL3, MYSM1, RMRP, RNU4ATAC, SMARCAL1, CARD11, ERBB2IP, IL6R, IL6ST, PGM3, SPINK5, STAT3, TGFBR1, TGFBR2, ZNF341, MTHFD1, SLC46A1, TBH2, IKBKB, IKBKG, NFKBIA, ORAI1, STIM1, BCL11B, CCBE1, EPG5, FAT4, KDM6A, KMT2A, KMT2D (MLL2), NFE2L2, PNP, RBCK1, RNF31, SKIV2L, SP110, STAT5B, STAT5B, TTC37, TTC7A, BLNK, BTK, CD79A, CD79B, IGHM, IGLL1, PIK3CD, PIK3R1, SLC39A7, TCF3, TCF3, TOP2B, ARHGEF1, ATP6AP1, CD19, CD20, CD21, CD81, IKZF1, IRF2BP2, MOGS (GCS1), NFKB1, NFKB2, PIK3CDGOF, PIK3R1, PTEN, RAC2, SEC61A1, SH3KBP1, TNFRSF13B, TNFRSF13C, TNFSF12, TRNT1, AICDA, AICDA, INO80, MSH6, UNG, CARD11, IGKC, Mutation or chromosomal deletion at 14q32, FAAP24, PRF1, SLC7A7, STX11, STXBP2, UNC13D, AP3B1, AP3D1, LYST, RAB27A, BACH2, CTLA4, DEF6, FERMT1, FOXP3, IL2RA, IL2RB, LRBA, STAT3, AIRE, AIRE, ITCH, JAK1, PEPD, TPP2, IL10, IL10RA, IL10RB, NFAT5, RIPK1, TGFB1, CASP10, CASP8, FADD, TNFRSF6, TNFSF6, CARMIL2, CD27, CD70, CTPS1, MAGT1, PRKCD, RASGRP1, SH2D1A, TNFRSF9, XIAP, CEBPE, CLPB, CSF3R, DNAJC21, EFL1, ELANE, G6PC3, G6PT1, GF11, HAX1, HYOU1, JAGN1, LAMTOR2, SBDS, SMARCD2, SRP54, TAZ, USB1, VPS13B, VPS45, WAS, ACTB, CFTR, CTSC, FERMT3, FPR1, ITGB2, MKL1, RAC2, SLC35C1, WDR1, CYBA, CYBB, NCF1, NCF2, NCF4, CYBC1, G6PD, GATA2, CSF2RA, CSF2RB, CYBB, IFNGR1, IFNGR1, IFNGR2, IL12B, IL12RB1, IL12RB2, IL23R, IRF8, IRF8, SG15, JAK1, RORC, SPPL2A, STAT1, TYK2, TYK2, CIB1, CXCR4, TMC6, TMC8, FCGR3A, IFIH1, IFNAR1, IFNAR2, IRF7, IRF9, POLR3A, POLR3C, POLR3F, STAT1, STAT2, DBR1, IRF3, TBK1, TICAM1, TLR3, TLR3, TRAF3, UNC93B1, CARD9,*

*IL17F, IL17RA, IL17RC, STAT1, TRAF3IP2, IRAK1, IRAK4, MYD88, TIRAP, APOL1, CLBH7, HMOX, NBAS, NCSTN, OSTM1, PLEKHM1, PSEN, PSENE1, RANBP2, RPSA, SNX10, TCIRG1, TNFRSF11A, TNFSF11, IL18BP, IRF4, ACP5, ADA2, ADAR1, DNASE1L3, DNASE2, IFIH1 (GOF), OAS1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, TMEM173, TREX1, USP18, POLA1, MEFV, MEFV, MVKNLRC4, NLRP1, NLRP1, NLRP12, NLRP3, NLRP3, NLRP3, PLCG2, ADAM17, ALPI, AP1S3, CARD14, COPA, HAVCR2, IL1RNIL36RN, LPIN2, NOD2, OTULIN, PSMB8\*, PSMB8\*, PSMG2, PSTPIP1, SH3BP2, SLC29A3, TNFAIP3, TNFRSF1A, TRIM22, C1QA, C1QB, C1QC, C1R, C1R, C1S, C1S, C2, C3C3, C4A, C4B, C5, C6, C7, C8A, C8B, C8G, C9, CD46, CD55, CD59, CFB, CFB, CFD, CFH, CFH, CFHR1, CFHR2, CFHR3, CFHR4CFHR5, CFHR1CFHR2, CFHR3CFHR4, CFHR5, CFI, CFP, FBH3, MASP2, SERPING1, THBD, ACD, ACD, BRCA1, BRCA2, BRIP1, CTC1, DKC1, DNAJC21, ERCC4, ERCC6L2, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCI, FANCL, FANCM, MAD2L2, NOLA2, NOLA3, PALB2, PARN, RAD51, RAD51C, RFD3, RTEL1, RTEL1, SAMD9, SAMD9L, SLX4, SRP72, STN1, TERC, TERT, TERT, TIN2, TIN2, TP53, UBE2T, WRAP53, XRCC2, XRCC9.*

Список из 31 гена, ассоциированного с врожденной нейтропенией, использованный в данном исследовании: *MTHFD1, LYST, CSF3R, ELANE, JAGN1, LAMTOR2, SMARCD2, VPS13B, WAS, WDR1, CXCR4, TCIRG1, HAX1, G6PC3, GFII1, GATA2, SLC37A4, SBDS, STK4, CLPB, AP3B1, USB1, VPS45, CXCR2, EIF2AK3, RAB27A, AK2, RMRP, TBH2, TAZ, CD40LG.*

Исследование вариантов генов у пациентов с врожденной нейтропенией и синдромом Хеннекама из Свердловской области выполнено на обезличенных данных, добровольно предоставленных родителями пациентов для биоинформационного анализа по согласованию с этическим комитетом. Были использованы только VCF-файлы с миссенс-мутациями гена *FAT4* (синдром Хеннекама) и *TCIRG1* (врожденная нейтропения), а также обезличенные клинические данные. Медицинское наблюдение пациентов и

клинические исследования у них проводились до начала нашего исследования в медицинских организациях Свердловской области.

Данные по различным миссенс-мутациям для генов всех исследуемых болезней были получены из открытых баз данных. В частности, *FAT4*, *ADAMTS3*, *CBEE1*, *ELANE* и *TCIRG1* были получены из общедоступной базы dbSNP с портала Национального центра биотехнологической информации (National Center for Biotechnology Information – NCBI, <https://www.ncbi.nlm.nih.gov/snp/>), а также базы Ensembl (<https://www.ensembl.org/index.html>), базы данных Swiss-Prot (<http://expasy.org/>), OMIM (<https://www.omim.org/>) и HGMD (<https://www.hgmd.cf.ac.uk/>).

База dbSNP - это онлайн-ресурс, созданный для помощи исследователям в области биологии. Его целью является создание единой базы данных, содержащей все идентифицированные генетические вариации (однонуклеотидные полиморфизмы), которые могут быть использованы для исследования широкого спектра генетически обусловленных природных явлений. В частности, доступ к молекулярным вариациям, каталогизированным в dbSNP, помогает проводить фундаментальные исследования, такие как физическое картирование, популяционная генетика, изучение эволюционных взаимоотношений, а также позволяет быстро и количественно оценить вариации на определенном участке генома (*рисунок 4*). Большинство этих вариаций нуклеотидных последовательностей были выявлены путем секвенирования и генотипирования образцов ДНК из общей популяции, помимо группы больных (*рисунок 5*).

База данных Ensembles (США) позволяет проанализировать транскрипцию конкретных генов, а также соответствующие последовательности белка и их различные варианты. В частности, для выполнения анализа мы загружали в базу CSV-файл вариантов для исследуемых нами генов (*рисунок 6*).

Center for Biotechnology Information

SNP  [Create alert](#) [Advanced](#)

Display Settings: Summary, 20 per page, Sorted by SNP\_ID Send to: ▼

**Search results**

Items: 1 to 20 of 108976 << First < Prev Page 1 of 5449 Next > Last >>

rs632899 [*Homo sapiens*]

1.

Variant type: SNV  
 Alleles: A>G,T [Show Flanks]  
 Chromosome: 18:59469637 (GRCh38)  
 18:57136869 (GRCh37)  
 Canonical SPDI: NC\_000018.10:59469636:A:G,NC\_000018.10:59469636:A:T  
 Gene: CCBE1 (Varview)  
 Functional Consequence: intron\_variant  
 Clinical significance: benign  
 Validated: by frequency,by alfa,by cluster  
 MAF: G=0.452517/23482 (ALFA)  
 A=0.315789/12 (Siberian)  
 G=0.340256/213 (Chileans)

...more

HGVS: NC\_000018.10:g.59469637A>G, NC\_000018.10:g.59469637A>T,  
 NC\_000018.9:g.57136869A>G, NC\_000018.9:g.57136869A>T,

Рисунок 4 – Представления данных о миссенс-замен на примере гена *CCBE1* в базе данных dbSNP на портале NCBI.

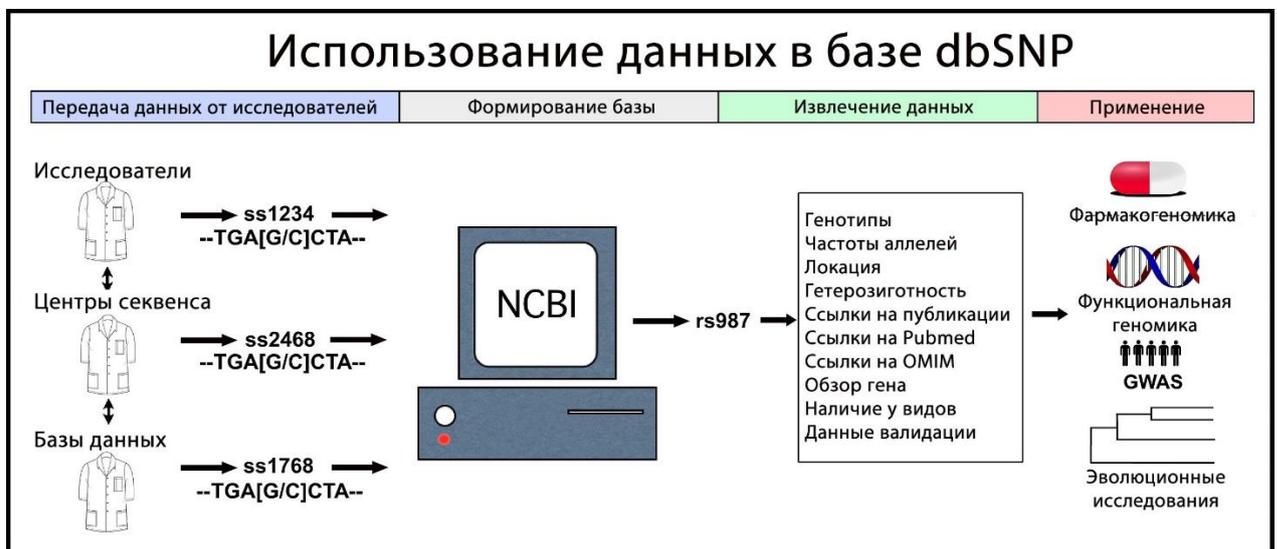


Рисунок 5 – Источники данных и пути использования данных в базе dbSNP (несинонимичных однонуклеотидных замен) Национального центра биотехнологической информации (NCBI, США).

Gene: **CCBE1** ENSG00000183287

Description: collagen and calcium binding EGF domains 1 [Source:HGNC Symbol;Acc:HGNC:29426]

Gene Synonyms: FLJ30681, KIAA1983

Location: [Chromosome 18: 59,430,939-59,697,662 reverse strand](#)  
GRCh38:CM000680.2

About this gene: This gene has 7 transcripts ([splice variants](#)), [185 orthologues](#) and is associated with [3 phenotypes](#).

Transcripts: [Hide transcript table](#)

Transcript ID	Name	bp	Protein	Biotype	CCDS	UniProt Match	RefSeq Match	Flags
<a href="#">ENST00000439986.9</a>	CCBE1-201	6271	<a href="#">406aa</a>	Protein coding	<a href="#">CCDS32838</a>	<a href="#">Q6UXH8-1</a>	<a href="#">NM_133459.4</a>	MANE Select v0.95 Ensembl Canonical GENCODE basic
<a href="#">ENST00000649564.1</a>	CCBE1-204	6403	<a href="#">406aa</a>	Protein coding	<a href="#">CCDS32838</a>	<a href="#">Q6UXH8-1</a>	-	GENCODE basic APPRIS
<a href="#">ENST000006650467.2</a>	CCBE1-205	6065	<a href="#">332aa</a>	Protein coding		<a href="#">A0A3B3IRL6</a>	-	GENCODE basic
<a href="#">ENST00000695904.1</a>	CCBE1-207	3290	<a href="#">435aa</a>	Protein coding		-	-	GENCODE basic
<a href="#">ENST00000695903.1</a>	CCBE1-206	3341	<a href="#">367aa</a>	Nonsense mediated decay		-	-	-
<a href="#">ENST00000589419.2</a>	CCBE1-203	1944	No protein	Processed transcript		-	-	TSL4
<a href="#">ENST00000589416.2</a>	CCBE1-202	1692	No protein	Retained intron		-	-	TSL4

Рисунок 6 – База данных Ensembl на примере поиска сведений о гене CCBE1.

Аналогичным образом для поиска вариантов генов иммунной системы использовались другие базы данных. Для перекрестной проверки данных несинонимичных однонуклеотидных замен (nsSNP) проводился поиск сведений в базах Консорциум по агрегации экзонов (Exome Aggregation Consortium – ExAC), на Сервере вариаций генома (Genome Variation Server), в Функциональной базе данных однонуклеотидных полиморфизмов (A Functional Single Nucleotide Polymorphism – F-SNP), базе данных генных мутаций человека (Human Gene Mutation Database - HGMD), которая объединяет известные генные мутации, ответственные за наследственные заболевания человека.

Аналогичным образом использовалась база данных генетических ассоциаций (Genetic Association Database - GAD), которая содержит архив более 3600 записей dbSNP, и База данных вариаций генома человека (Human /Genome Variation database – HGVBase), в которых сообщается об ассоциациях

с заболеваниями по результатам опубликованных клинических исследований, хотя из этих статистических ассоциаций до настоящего времени подтверждены очень немногие.

Используемая в работе онлайн база данных Менделевское наследование у человека (Online Mendelian Inheritance in Man - OMIM) представляет собой каталог генетических нарушений наследственных заболеваний, сопоставленных с генами человека не только высоко пенетрантными, но и редкими (MAF – минорная частота распространенности в популяции < 0,01).

Мы собрали данные о несинонимичных SNP с этих порталов, связанных с исследуемыми генами *FAT4*, *ADMATS3*, *CCBE1*, *ELANE* и *TCIRG1*; были исключены данные, относящиеся к другим факторам.

Количество SNP для перечисленных генов отображено в *таблице 1*.

Таблица 1 – Количество SNP, загруженных из баз данных dbSNP и Ensemble

Ген	SNP	nsSNP
<i>ELANE</i>	3646	301
<i>TCIRG1</i>	5627	811
<i>CCBE1</i>	73845	407
<i>FAT4</i>	68257	3434
<i>ADAMTS3</i>	70876	911

## 2.2 – Методы, использованные в работе

**2.2.1 - Анализ дифференциальной экспрессии генов по данным пациентов с дефицитом *HOIL-1/RBCK1* и пациентов с врожденной нейтропенией.**

Анализ дифференциальной экспрессии генов (ДЭГ) - это процесс, используемый для выявления генов, дифференциально экспрессированных

между двумя или более состояниями, такими как нормальное и болезнь или состояние при различных видах лечения. Этот анализ может быть выполнен с использованием инструментов и конвейеров биоинформатики [222].

Процедура анализа ДЭГ включает несколько шагов, включая контроль качества исходных данных секвенирования, отображение чтения на ссылочный геном или транскриптом, количественное определение уровней экспрессии генов, нормализацию данных, статистический анализ для выявления дифференциально экспрессированных генов и функциональный анализ выявленных генов.

Для определения различий в экспрессии генов в наборе данных при исследовании дефицита *RBCK1* был использован стандартный метод дифференциальной экспрессии генов. Анализ выполнен с использованием пакета Volstad R. Различия в экспрессии генов между пациентами с дефицитом *RBCK1* и нормальными образцами оценивались как значимые при значении  $P\text{-value} < 0,05$ ,  $|\log FC| > 1$ , а частота ложных обнаружений (FDR)  $p\text{-value}$  от 0,57 использовалась в качестве порогового значения [28].

Определение частоты ложных обнаружений (FDR) - это метод концептуализации уровня ошибок первого рода при проверке нулевых гипотез при множественных сравнениях.  $\text{Log}^2$ -изменение является отсекающим значением, важным для расчета разницы между уровнями экспрессии.

Метод определения частоты ложных обнаружений - один из главных статистических инструментов при аннотировании генов с использованием GO.

GO – это стандартизированный словарь терминов, которые используются для описания функций генов, клеточных компонентов и биологических процессов в различных организмах. Каждый ген может быть аннотирован к одному или нескольким терминам GO, что может быть использовано для вывода о функции гена и для сравнения функций различных генов [71].

Расчёт отличий в экспрессии генов проводился с помощью пакета R Limma. Анализ функционального обогащения генов, характерных для различных врожденных первичных иммунодефицитов и аутовоспалительных заболеваний – с помощью пакета R Bioconductor.

Анализ обогащения набора генов (GSEA) - это комплекс методов, позволяющих связать набор генов с изменением фенотипа [225]. Такие методы часто используют базы данных по ранее аннотированным наборам генов для формализации существующих данных о фенотипе (например, термины проекта Gene Ontology (GO): молекулярные функции, биологические процессы или клеточные компоненты [134]. Результатом применения метода (выпуска программы) в данном случае является набор предварительно проаннотированных наборов, которые помогают определить, зависит ли упорядоченный список генов от фенотипа или они просто случайны [225]. Такие предварительно проаннотированные наборы называются перепредставленными (если частота выше фоновой) или недопредставленными (если частота ниже фоновой).

Коэффициент обогащения (ES) - это статистический коэффициент, определяемый методом Колмогорова-Смирнова, отражающий степень перепредставленности генов в верхней или в нижней части ранжированного списка генов.

Анализ избыточного обогащения (ORA) и анализ онтологии генов (Gene Ontology) и вовлеченных сигнальных путей проводились с помощью анализа задействованных сигнальных путей в базах KEGG, WikiPathways, reactome и DAVID. Впоследствии DAVID был использован для проведения анализа в базе KEGG и аннотации генов (GO) [71, 109, 200]. Основные гены были выбраны по уровню связности и изображены с помощью программного обеспечения Metaphase [221].

В дальнейшем проводилось исследование межбелковых взаимодействий для выявления функциональных взаимоотношений между

различно экспрессированными генами. Сеть белок-белковых взаимодействий для генов с различиями в экспрессии была построена с помощью программы STRING (версия 10.0) (<http://string-db.org>) с комбинированной оценкой  $>0,4$  в качестве порога для статистически значимого взаимодействия [192]. Для дальнейшего исследования интерактивной сети использовали программу Cytoscape (версия 3.4.0) с плагином Molecular Complex Detection (MCODE) для классификации важных молекул в сети PPI, при этом критериями распознавания служили оценки MCODE  $>5$ , degree cut-off=2, node score cut-off=0.2, Max depth=100 и k-score=2. Биологические процессы узловых генов были визуализированы с помощью плагина Cytoscape (BiNGO) (версия 3.0.3) с порогом значимости 0,01 и Homo sapiens в качестве выбранного организма [58, 126].

### **2.2.2 - Предсказание генов-кандидатов врожденной нейтропении**

Для предсказания генов-кандидатов врожденной нейтропении нами были предприняты следующие действия.

Во-первых, использовалась база данных STRING для получения данных белок-белкового взаимодействия (PPI) для генов ПИД и врожденной нейтропении. Данные включают геномный контекст, коэкспрессию, а также известные и предсказанные взаимодействия из предыдущих данных. Минимальное необходимое значение взаимодействия было установлено на 0,4 [192].

При оценке плотности генной сети (Dnetwork) и биологического расстояния для генов врожденной нейтропении и генов других первичных иммунодефицитов использовалась программа Cytoscape (версия 3.5.1) [58]. Плотность сети (Dnetwork) является наиболее широко используемым понятием в регуляции генов и изучении сетей межбелковых взаимодействий (PPI) и может использоваться для определения того, является ли сеть плотной или нет. Плотность сети (Dnetwork) определяется по формуле [88].

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

где  $a_{ij}$  - попарная смежность,  $\sum$  представляет связность (связность сети равна невзвешенная связность сети равна количеству генов, которые непосредственно  $i$ -го гена), а  $n$  - количество генов в сети.

Следует обратить внимание на то, что  $a_{ij} = 1$ , если взаимодействие гена  $i$  и гена  $j$  имеет место в базе данных STRING, тогда как  $a_{ij} = 0$  в обратном случае.

Данные PPI группы врожденной нейтропении (основанные на опубликованных 32 генах врожденной нейтропении) и десяти случайных групп (каждая группа состоит 41 гена ПИД) были соответственно преобразованы в симметричную матрицу смежности ( $a_{ij}$ ,  $i, j = 1, \dots, n$ ) с помощью пакета «igraph» R [55].

Плотность сети использовалась для сравнения их функциональной сплоченности и близости. Чем выше плотность сети в группе, тем тем теснее взаимодействие генов в группе. Понятие биологического расстояния ( $B_{i,j}$ ) было впервые введено Итаном Й. и др, в 2013 году. С помощью биологического расстояния исследователи, изучающие функциональные взаимоотношения генов в сети геномных взаимодействий, подразумевают не фактическое расстояние между генами в молекуле ДНК или на хромосоме, а функциональную близость между парами генов или внутри группы генов [201].

Используя значение биологического расстояния, Itan Y. et al. показали, что гены первичных иммунодефицитов, как правило, расположены в центре геномной сети человека и образуют несколько тесно взаимосвязанных подгрупп в соответствии с различными биологическими механизмами [107 201]. Биологическое расстояние ( $B_{i,j}$ ) определяется по формуле:

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

где  $S_{i,j}$  - комбинированный показатель между геном  $i$  и геном  $j$ , полученный из базы данных STRING, и  $C$  - количество прямых связей между геном  $i$  и искомым. Чем меньше биологическое расстояние между группами, тем теснее биологическая взаимосвязь между генами в группе.

Был проведен расчет биологического расстояния группы известных генов врожденной нейтропении (32 генов) и двух случайных групп генов ПИД (каждая группа состояла из 41 гена ПИД) с помощью пакета Python для коннектома генов человека (HGC), предоставленного Y. Itan et al., 2015 [107].

Используя пакет "igraph" R, данные PPI группы врожденной нейтропении и 10 случайных групп (каждая с 41 геном PID) были преобразованы в симметричную матрицу смежности ( $a_{ij}$ ,  $1 \leq j = 1, n$ ) [55]. Сплоченность или плотность сети определялась с помощью анализа плотности сети (более высокая плотность сети представляет более тесное взаимодействие генов в группе).

В дальнейшем проведена оценка биологической дистанции между генами ( $B_{ij}$ ) которая может быть использована для расчета кратчайших функциональных расстояний между всеми возможными парами генов человека [201].

Расчет биологической дистанции между группой генов врожденной нейтропении (31 гена) и двумя случайными группами ПИД (41 ген ПИД в каждой) произведен с использованием инструмента Human Gene Connectome (HGC) в Python [201].

Непосредственный поиск генов-кандидатов после подготовительных этапов проводился тремя способами.

1) Проведен корреляционный анализ Пирсона (PCC) для оценки экспрессии 31 гена врожденной нейтропении и каждого белок-кодирующего гена (или гена-кандидата) на основе наборов данных GSE142347 и GSE6322 (Загруженные из NCBI с помощью GEO транскриптомные профили пациентов с врожденной нейтропенией). Использовался  $|r| > 0,9$  и  $p < 0,05$ .

2) Данные PPI для всех белковых кодирующих генов человека были получены из источника J. Cheng et al., 2006 [40], включая 217160 взаимодействий, предоставленных одиннадцатью базами данных, например, BioGRID [29], HI-II-14\_Net [19], HPRD [91], Instruct [101], InnateDB [100], IntAct [102], MINT [205], PINA [160], Signalink2.0 [191], KinomeNetworkX [172] и PhosphositePlus [118]. Ген-кандидат затем сохранялся, если взаимодействие между геном врожденной нейтропении и геном-кандидатом, полученным на предыдущем этапе, обнаруживалось в данных PPI.

3) Анализ Киотской геной и геномной энциклопедии (KEGG) был проведен с помощью пакета R «clusterProfiler» для оценки обогащения биологических функций генов врожденной нейтропении [217]. Затем KEGG-анализ был проведен для оставшихся генов-кандидатов врожденной нейтропении. Ген определялся как истинный ген-кандидат врожденной нейтропении, если он был обогащен в том же пути, что и ген врожденной нейтропении.

Для того, чтобы определить, подходит ли наш метод для предсказания генов-кандидатов врожденной нейтропении, мы рассчитали биологические расстояния ( $B_i, j$ ) предсказанных генов-кандидатов и сравнили их с 32 известными генами врожденной нейтропении. Затем были проведены «функциональное геномное выравнивание» (FGA) и филогенетический кластерный анализ. Эти этапы были выполнены с использованием пакета APE, доступного в R для оценки биологической корреляции между генами-

кандидатами и известными генами [152, 201]. В частности, сначала мы создали матрицу биологического расстояния между генами врожденной нейтропении и генами-кандидатами врожденной нейтропении, а затем применили алгоритм объединения соседей (функция  $pnj$ ) для создания филогенетического веерообразного дерева, показывающего иерархическую кластеризацию известных и генов-кандидатов врожденной нейтропении. Если гены-кандидаты были равномерно распределены по всему диапазону известных генов врожденной нейтропении, то это означало, что эти гены-кандидаты тесно связаны с известными генами. Если гены-кандидаты и известные гены были разделены на две или более групп, это означало обратное.

Применяя пакет R "limma", мы искали гены с различиями в экспрессии между пациентами с врожденной нейтропенией НС и здоровым контролем и показали  $|\log 2|$ -кратное изменение. Значения принимались отсекающими (Threshold)  $> 0,4$  и P-значение  $< 0,05$  [120]. Перекрытие данных между информацией о дифференциально экспрессированных генах, полученной из анализа датасетов GSE142347 и GSE6322, было определено с помощью диаграммы Венна в пакете R [39].

### **2.2.3 - Оценка последовательности несинонимичных однонуклеотидных замен (миссенс-SNP) генов *CCBE1*, *FAT4*, *ADAMTS3*, *TCIRG1*, *ELANE* и предсказание патогенности замен.**

Мы использовали различные инструменты *in silico* для проверки функциональной оценки перечисленных генов иммунной системы с nsSNP, имеющих патологический или доброкачественный характер. Мы использовали следующие инструменты: SIFT [145], POLYPHEN-2 [70] PROVEAN [42], FATHMM [69], LRT [215], M-CAP [125], VEST3 [59], CAAD [52], MetaLR [176], Mutation Assessor [105], MutationTaster [135, 141], и FATHMM-MKL [22], SNP&GO, PhD-SNP [77], PANTHER [150], SNAP2 [82]. Все эти инструменты были доступны через VarCard [212] и MutPred [99].

SIFT (Sorting Intolerant From Tolerant) - это алгоритм биоинформатики, используемый для прогнозирования возможного влияния аминокислотных замен на функцию белка. Алгоритм работает путем сравнения аминокислоты в данной позиции в белковой последовательности с набором родственных белковых последовательностей и оценки того, насколько аминокислота сохранилась у разных видов.

Алгоритм SIFT рассчитывает балл для каждой аминокислотной замены в диапазоне от 0 до 1. Оценка 0 означает, что замена с высокой вероятностью может быть вредной, в то время как оценка 1 означает, что замена, скорее всего, является доброкачественной. Для сортировки вариантов генов на патологические и доброкачественные пороговое значение в SIFT было установлено  $>0,5$  (рисунок 7).

### SIFT results (dbSNP)

Processing... If your browser times out before results are shown, html results can be seen at [https://sift.bii.a-star.edu.sg/www/sift/tmp/58b8f27d5f\\_dbSNP.html](https://sift.bii.a-star.edu.sg/www/sift/tmp/58b8f27d5f_dbSNP.html) and tsv results at [https://sift.bii.a-star.edu.sg/www/sift/tmp/58b8f27d5f\\_dbSNP.tsv](https://sift.bii.a-star.edu.sg/www/sift/tmp/58b8f27d5f_dbSNP.tsv) Both files are stored for 24 hours before being deleted.

Done.

SNP	ORGANISM/BUILD	CHR	COORDINATE	REF ALLELE	ALT ALLELE	AMINO ACID CHANGE	GENE NAME	GENE ID	TRANSCRIPT ID	PROTEIN ID	REGION	SIFT SCORE	SIFT MEDIAN	N SE POS
rs2288598	Homo_sapiens/GRCh37.74	18	57363917	G	A	I52I	CCBE1	ENSG00000183287	ENST00000439986	ENSP00000404464	CDS	1	3.71	13
rs61745250	Homo_sapiens/GRCh37.74	18	57106987	G	A	P279P	CCBE1	ENSG00000183287	ENST00000439986	ENSP00000404464	CDS	1	3.32	385
rs80008675	Homo_sapiens/GRCh37.74	18	57364452	G	T	D41E	CCBE1	ENSG00000183287	ENST00000439986	ENSP00000404464	CDS	0.016	4.25	11
rs116596858	Homo_sapiens/GRCh37.74	18	57133983	G	A	P181S	CCBE1	ENSG00000183287	ENST00000439986	ENSP00000404464	CDS	0.007	3.58	22
rs116596858	Homo_sapiens/GRCh37.74	18	57133983	G	A		CCBE1	ENSG00000183287	ENST00000398179	ENSP00000381241	UTR_5			
rs116596858	Homo_sapiens/GRCh37.74	18	57133983	G	A		CCBE1	ENSG00000183287	ENST00000589419	ENSP00000467710	UTR_5			
rs116675104	Homo_sapiens/GRCh37.74	18	57134025	G	A	R167W	CCBE1	ENSG00000183287	ENST00000439986	ENSP00000404464	CDS	0.017	3.52	25
rs116675104	Homo_sapiens/GRCh37.74	18	57134025	G	A		CCBE1	ENSG00000183287	ENST00000398179	ENSP00000381241	UTR_5			
rs116675104	Homo_sapiens/GRCh37.74	18	57134025	G	A		CCBE1	ENSG00000183287	ENST00000589419	ENSP00000467710	UTR_5			

Рисунок 7 – Пример представления результатов анализа патогенности SNP в SIFT

PolyPhen-2 (Polymorphism Phenotyping v2) - это инструмент биоинформатики, используемый для прогнозирования возможного функционального влияния аминокислотной замены в белке. Алгоритм анализирует аминокислотную последовательность белка, положение варианта и свойства аминокислот, участвующих в замене, чтобы предсказать, будет ли замена повреждающей или доброкачественной.

PolyPhen-2 использует комбинацию эволюционного сохранения и структурной информации для составления прогнозов. Сначала алгоритм выравнивает аминокислотную последовательность белка с последовательностями других родственных видов, чтобы определить, какие аминокислоты являются высококонсервативными и, следовательно, могут быть функционально важными. Затем алгоритм использует ряд структурных характеристик, включая доступность растворителя и наличие водородных связей, чтобы предсказать влияние аминокислотной замены на структуру и функцию белка.

Результатом работы PolyPhen-2 является предсказание функционального влияния аминокислотной замены, которое выражается в баллах от 0 до 1. Варианты с оценкой более 0,5 балла считаются вредными, а варианты с оценкой менее 0,5 балла - доброкачественными. PolyPhen-2 показал высокую точность в предсказании влияния аминокислотных замен, что делает его полезным инструментом для исследователей, изучающих влияние генетических вариаций на функции белков [70, 92] (рисунк 8).

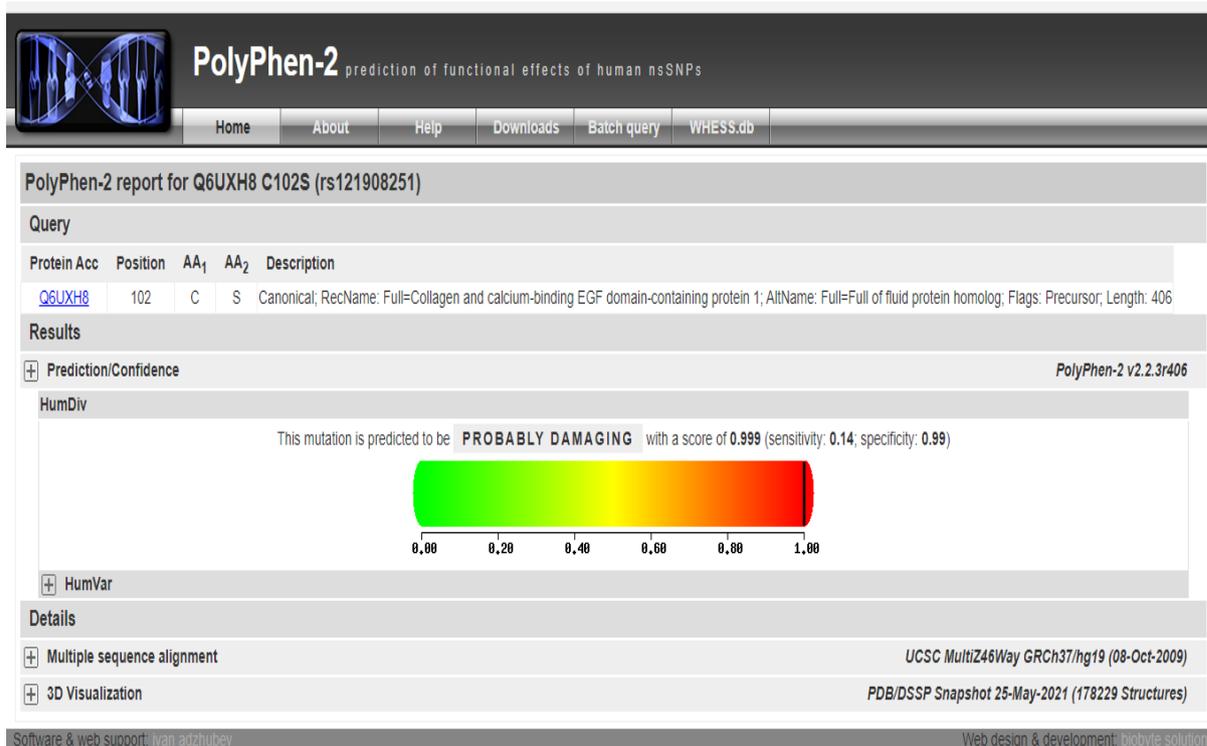


Рисунок 8 – Пример представления результатов анализа патогенности SNP в PolyPhen-2

VarCards - это инструмент биоинформатики, используемый для анализа генетических вариантов и прогнозирования их потенциального влияния на здоровье человека. Он объединяет данные из различных источников, включая публичные базы данных, литературу и экспериментальные данные, чтобы предоставить полную информацию о генетических вариантах и их потенциальном клиническом значении.

Алгоритм VarCards работает в несколько этапов:

1) Сбор данных о вариантах: Сначала алгоритм собирает информацию о генетических вариантах из различных источников, включая публичные базы данных, такие как dbSNP и ClinVar, а также литературные и экспериментальные данные.

2) Аннотирование вариантов: Алгоритм аннотирует варианты с информацией об их геномном расположении, функциональном влиянии и

частоте в популяции. Эта информация берется из различных источников, включая Ensembl, dbNSFP и ExAC.

3) Прогнозирование патогенности: Алгоритм использует различные вычислительные инструменты, такие как SIFT и PolyPhen-2, для прогнозирования потенциального влияния каждого варианта на функцию белка и оценки его влияния на изменение функции белка.

4) Ассоциация с заболеваниями: Алгоритм также объединяет информацию об ассоциации каждого варианта с заболеваниями человека из различных источников, включая базу данных мутаций генов человека (HGMD), ClinVar и PubMed.

5) Клиническая интерпретация: Наконец, алгоритм предоставляет клиническую интерпретацию каждого варианта, включая его потенциальную патогенность, ассоциацию с заболеваниями и соответствие конкретным клиническим условиям.

VarCards предоставляет удобный интерфейс для запроса и анализа генетических вариантов, а также настраиваемый конвейер для интеграции дополнительных источников данных и инструментов анализа. Он широко используется в клинических и исследовательских условиях для анализа генетических вариантов и выявления потенциальных мутаций, вызывающих заболевания.

VarCARD использовался нами для получения результатов таких инструментов, как: LRT, Mutation Taster, Mutation Accessor, PROVEAN, FATHMM, VEST3, MTA SVM, METALR, M-CAP, CADD, DANN, FATHMM-MKK, PhD-SNP, PANTHER, SNP-GO, P-MUT [212] (рисунк 9).

Отдельно использовался такой инструмент, как SNAP2. SNAP2 использует алгоритм на основе хэширования для быстрого сопоставления чтений секвенирования с эталонным геномом, что позволяет ему быть намного быстрее других инструментов выравнивания. В нем также

используется алгоритм динамического программирования для повышения точности и обработки пробелов и несоответствий в выравниваниях [72].

Пороговыми значениями для перечисленных выше инструментов были следующие значения: Mutation Taster:  $< 0.5$ ; CADD:  $> 15$ ; MetaLR:  $> 0.5$ ; M-Cap:  $> 0.025$ ; PANTHER: вероятно повреждающая (probably) замена при time  $> 450\text{my}$  возможно повреждающая (possibly, «при возможности», менее вероятная) при  $450\text{my} > \text{time} > 200\text{my}$ , вероятно доброкачественная при time  $< 200\text{my}$ ; VEST3:  $> 0.5$ ; LRT:  $> 0.001$ ; PROVEAN:  $> -2.667$ ; FATHMM-MKK:  $< 0.5$ ; PhDSNP:  $> 0.5$ ; SNP-GO:  $> 0.5$ ; SNAP2: шкала от  $-100$  (полностью нейтральная) до  $+100$  (оказывает сильный эффект); DANN:  $> 0.5$ ; Mutation Assessor:  $> 0.65$  (от  $-5.545$  до  $5.975$ , чем выше значение, тем более повреждающая); FATHMM:  $> 0.453$ ; PON-P2:  $> 0.5$ .

Details	Location	Ref	Alt	Gene	Effect	Amino acids change	D:A algorithms	damaging score
-	chr18:57103144-57103144	G	A	CCBE1	+ nonsynonymous nonsy...	+ CCBE1:NM_133459:ex...	16:23	0.70

Tool	Score	Impact
LRT	0.001	Neutral
MutationTaster	1.000	Disease_causing
MutationAssessor	2.08	Medium
FATHMM	-2.08	Damaging
PROVEAN	-3.04	Damaging
VEST3	0.352	Tolerable
MetaSVM	0.533	Damaging
MetaLR	0.729	Damaging
M-CAP	0.115	Damaging
CADD	24.5	Damaging
DANN	0.997	Damaging
FATHMM_MKL	0.838	Damaging
Eigen	0.027	Damaging

dataset	population	allele frequency
gnomAD_exome	ALL	-
gnomAD_exome	African American	-
gnomAD_exome	Latino	-
gnomAD_exome	Ashkenazi Jewish	-
gnomAD_exome	East Asian	-
gnomAD_exome	Finnish	-
gnomAD_exome	Non-Finnish European	-
gnomAD_exome	Other	-

database
denovo-db
InterVar
COSMIC
ICGC
GWAS Catalog
dbSNP
InterPro
ClinVar
Segmental_duplication

Рисунок 9 – Пример представления результатов анализа патогенности SNP в VarCards

Онлайн-сервис MutPred (<http://mutpred.mutdb.org/>) используется в качестве поискового инструмента для предсказания молекулярной основы заболевания, связанного с аминокислотной заменой в мутантном белке. Он использует несколько атрибутов, связанных со структурой, функцией и эволюцией белка. В MutPred используются три других сервиса - PSI-BLASAT, SIFT и Pfam, а также алгоритмы TMHMM, MARCOIL и DisProt. Это позволяет прогнозировать большинство структурных повреждений и достигать еще большей точности предсказания путем объединения оценок всех этих трех сервисов [99].

#### **2.2.4 - Оценка влияния nsSNP генов *CCBE1*, *FAT4*, *ADAMTS3*, *TCIRG1*, *ELANE* с помощью инструментов *in silico* на структуру и функцию белков.**

Метод Mupro использует машинное обучение вектора поддержки для прогнозирования изменений стабильности белка при одноаминокислотных мутациях, используя как информацию о последовательности, так и структурную информацию, как и метод iMutant 3.0.

iMutant 3.0 - это веб-сервер, который предсказывает влияние одноточечных мутаций на стабильность белка и выдает оценку, указывающую на вероятность дестабилизации или стабилизации мутации. Алгоритм основан на методе опорных векторов (SVM), обученную на большом наборе данных экспериментально охарактеризованных мутантов, для прогнозирования влияния мутации на стабильность белка. Модель SVM обучена различать стабилизирующие и дестабилизирующие мутации на основе извлеченных признаков.

iMutant 3.0 показал высокую точность в предсказании влияния одноточечных мутаций на стабильность белка, что делает его полезным инструментом для исследователей, изучающих влияние генетических вариаций на функцию белка. Алгоритм может быть использован для широкого

спектра приложений, включая конструирование стабильных и функциональных белков и выявление мутаций, вызывающих заболевания.

Некоторые методы используют сохранение последовательности определенных аминокислот в семействе последовательностей или ищут определенные особенности структуры белка, чтобы предсказать, влияет ли замена на функцию белка. Аминокислотные замены, вызванные nsSNP, могут изменить стабильность нативного белка, что может привести к влиянию на белок и, в конечном итоге, к заболеваниям [40].

С помощью классификатора met, iStable 2.0, мы предсказывали изменения, вызванные миссенс-заменой SNP, на стабильность белка. Классификатор met использует машинное обучение и исследует, увеличивается или уменьшается стабильность белка. Это происходит из-за аминокислотной замены, которая основана на предсказании 8 структурных (I-Mutant 3.0, CUPSAT, PoPMuSiC, AUTO-MUTE2.0, SDM, DUET, mCSM, MAESTRO и SDM2) и 3 последовательных (I-Mutant2.0, MUpro и iPTREESTAB) инструментов предсказания стабилизации белка. В качестве входных данных используется 4-буквенный PDB-код или последовательность белка в формате FASTA, но структурный предиктор достигает лучших показателей, чем последовательный. На веб-сервере <http://ncblab.nchu.edu.tw/iStable2> можно найти iStable 2.0. [106]. I-Mutant 3.0 <https://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi>.

Набор данных  $\Delta\Delta G$  Mut от Pro Therm был использован для предварительного обучения алгоритма. Значение  $\Delta\Delta G$  (ккал/моль) может быть использовано для определения односайтовой мутации, которая зависит от структуры или последовательности белка. Значение  $\Delta\Delta G$  меньше нуля указывает на то, что вариант изменяет структуру или последовательность белка [35].

Для оценки структурных последствий искомой замены при исследовании ELANE использовался проект HOPE. Project HOPE - это веб-

сервер (<https://www.projecthope.org/>), который предлагает оценить измененный белок в наблюдаемой 3D структуре при взаимодействии с UniProt и 3DAS алгоритмом предсказания. Последовательность белка используется в качестве входного источника в Project HOPE, а затем проводится структурное сравнение с диким типом.

Кроме того, оценка вторичной структуры белка ELANE проводилась при помощи программы SOPMA [74]. Это более сложная версия метода самооптимизированного предсказания (SOPM), который может предсказать вторичную структуру (спираль, поворот и виток) 69,5% аминокислот в базе данных 126 негомологичных (менее 25% гомологичных) белковых цепей. SOPMA и нейросетевой подход (PhD) правильно предсказывают 82,2% остатков и 74% предсказанных аминокислот при совместном использовании.

### **2.2.5 – Оценка влияния nsSNP на посттрансляционную модификацию белков иммунной системы.**

Оценка влияния аминокислотных замен в участках, влияющих на посттрансляционную модификацию белка, проводилось для предсказания изменений его структуры и функции [4, 143]. Программное обеспечение, доступное онлайн – GPSMSP v3.0 (<https://msp.biocuckoo.org/online.php>) использовалось для предсказания сайтов метилирования.

Мы использовали NetPhos 3.177 (<https://www.cbs.dtu.dk/services/NetPhos/>) [156] и GPS 5.078 (<https://gps.biocuckoo.cn/>) [76] для предсказания потенциальных сайтов фосфорилирования. Сервис NetPhos 3.1 предсказывает сайты фосфорилирования серина, треонина и тирозина в белках с помощью ансамблей нейронных сетей. Сайты фосфорилирования с баллом более 0,5 более вероятно будут фосфорилированы [31].

Для оценки потенциальных сайтов метилирования и убиквитилирования мы использовали GPSMSP 1.0 (<https://msp.biocuckoo.org/>), BDMPUB (<https://www.bdmpub.biocuckoo.org/>) и UbPred [93]

(<https://www.ubpred.org>). NetOglyc4.0 дополнительно предсказывает сайты гликозилирования, используя гликозилирование [162] (<https://www.cbs.dtu.dk/services/NetOGlyc/>). Сайты гликозилирования с баллом более 0,5 более вероятно будут гликозилированы.

### **2.2.6 – Оценка влияния nsSNP на функционально различные участки белков иммунной системы.**

Консервационный анализ - это метод биоинформатики, используемый для выявления функционально важных регионов в белковых структурах путем анализа эволюционного сохранения в родственных белковых последовательностях. Метод основан на принципе, что эволюционно сохраненные области в белковых структурах, скорее всего, являются функционально важными, в то время как изменчивые области, скорее всего, менее важны для функции белка. Для консервационного анализа использовался нейросетевой алгоритм и соответствующий веб-сервис ConSurf [53].

Алгоритм преобразует расчетную скорость эволюции в балл сохранности относительно других родственных белковых последовательностей, который обычно варьируется от 1 (высокая изменчивость) до 9 (высокая сохранность). Оценки сохранности затем наносятся на структуру белка для выявления сохраненных и изменчивых областей. Это можно визуализировать с помощью различных инструментов, таких как PyMOL или Chimera.

Анализ консервации может быть использован для выявления функционально важных регионов в белковых структурах, таких как активные сайты, сайты связывания и структурные домены. Он также может быть использован для изучения эволюции функции белка и для разработки экспериментов по проверке функциональной важности определенных регионов в структуре белка.

Исходя из расположения и функциональной важности различных участков белка, аминокислотные участки в белке можно разделить на несколько категорий, включая функциональные, открытые, заглубленные и структурные остатки.

Функциональные остатки - это аминокислоты, которые вносят непосредственный вклад в функцию белка, например, активные участки, участки связывания или каталитические остатки. Функциональные остатки, как правило, высоко консервативны в родственных белках и часто расположены на поверхности белка, где они могут взаимодействовать с другими молекулами.

Открытые остатки - это аминокислоты, которые расположены на поверхности белка и доступны для окружающей среды. Открытые остатки могут играть роль в белок-белковых взаимодействиях, связывании лигандов и других функциях, требующих взаимодействия с внешней средой.

Заглубленные остатки - это аминокислоты, которые находятся во внутренней части белка и недоступны для окружающей среды. Захороненные остатки играют важную роль в поддержании общей структуры и стабильности белка, поскольку они участвуют в формировании ядра белка.

Структурные остатки - это аминокислоты, которые не участвуют непосредственно в функционировании белка, но важны для поддержания его структуры и стабильности. К структурным остаткам относятся те, которые формируют вторичную структуру белка, например, альфа-спирали и бета-листы, а также те, которые способствуют общей стабильности белка, например, дисульфидные связи.

Классификация аминокислот по этим категориям может дать представление о структуре и функции белка, а также о его эволюционной истории и потенциале для инженерии или модификации.

### **2.2.7 – Построение 3D-модели структуры белков иммунной системы для выявления влияния аминокислотных замен**

**Источником данных**, чтобы получить последовательность дикого (оригинального) типа белка, использовалась база UniProt (Universal Protein Resource, <https://www.uniprot.org/>) - это онлайн база данных последовательности белков и функциональной информации о белках, находящаяся в свободном доступе. UniProt представляет собой централизованное хранилище белковых последовательностей, аннотаций и другой сопутствующей информации, которая поступает из различных баз данных [171].

**Предсказание трехмерных моделей белков** с целью дальнейшего сравнения трехмерных моделей диких (оригинальных) типов и мутантных (измененных) типов белков проводилось путем их 3D-моделирования (в программах Phyre2, I-Tasser, HHpred и AlphaFold2), наложения структур, сравнения и в дальнейшем с помощью симуляции молекулярной динамики (MDS). В результате работы этих программ получались файлы .pdb, содержащие координаты атомов в 3D-пространстве [154].

При этом, программы HHpred и AlphaFold2 позволяют проводить MDS без использования сторонних приложений от Schrodinger, о которых будет сказано ниже, и позволяют оценить среднеквадратическое отклонение (RMSD) расстояний между углеродными основами природных и мутантных моделей во времени.

Приложение HHpred (Обнаружение гомологии и предсказание структуры путем сравнения скрытых Марковских моделей) - это инструмент биоинформатики, который использует профильное сравнение скрытых Марковских моделей (HMM) для выявления гомологичных последовательностей и предсказания структуры белков. Алгоритм HHpred сравнивает целевую последовательность с базой данных HMM, полученных

из семейств белков в базе данных Pfam, для выявления гомологичных последовательностей и предсказания структуры белка.

HNPred является широко используемым инструментом для предсказания структуры белка, обладает высокой точностью и успешностью в определении гомологичных последовательностей и в предсказании структуры белка. Он особенно информативен для исследования белков, которые не имеют значительного сходства последовательностей с белками с известной структурой. HNPred находится в свободном доступе в виде веб-сервера и может быть использован для предсказания структуры и функции широкого спектра белков.

AlphaFold 2 - это программное обеспечение для предсказания структуры белков на основе глубокой нейронной сети, разработанное исследовательской группой искусственного интеллекта компании DeepMind. AlphaFold 2 использует методы глубокого обучения для предсказания 3D-структуры белков с высокой точностью, достигая в некоторых случаях точности, близкой к атомному уровню. Программное обеспечение использовалось Jumper J. et al., 2021, для предсказания структуры многих белков, включая те, которые вовлечены в такие заболевания, как COVID-19, и имеют потенциал для ускорения поиска лекарств и разработки белков. AlphaFold 2 был выпущен как инструмент с открытым исходным кодом, что делает его свободно доступным для исследователей во всем мире [85].

Phyre2 - это набор инструментов, доступных в Интернете для предсказания и анализа структуры, функции и мутаций белка. Основной задачей Phyre2 является предоставление биологам простого и интуитивно понятного интерфейса к самым современным инструментам белковой биоинформатики [207].

I-Tasser – Сервер итеративного уточнения сборки нитей – представляет собой интегрированную платформу для автоматизированного предсказания структуры и функции белка на основе парадигмы «последовательность-

структура-функция». Начиная с аминокислотной последовательности, I-TASSER сначала генерирует трехмерные (3D) атомные модели на основе многопоточного выравнивания и итеративного моделирования структурной сборки. Функция белка определяется путем структурного сопоставления 3D-моделей с другими известными белками. Результат работы типичного сервера содержит прогнозы полноразмерной вторичной и третичной структуры, а также функциональные аннотации по сайтам связывания лигандов, номерам ферментных комиссий и терминам Gene Ontology. Оценка точности предсказаний предоставляется на основе доверительного балла моделирования. Этот протокол дает новое понимание и рекомендации по проектированию серверных систем для современных прогнозов структуры и функции белков. Сервер доступен по адресу <http://zhanglab.ccmb.med.umich.edu/I-TASSER> [182].

**Полученные файлы .pdb были визуализированы** в программах PyMOL, Chimera и в онлайн-сервисе Discover Studio.

Chimera UCSF - программа для интерактивной визуализации и анализа молекулярных структур и связанных с ними данных, включая карты плотности, траектории и выравнивания последовательностей [210]. PyMOL это кросс-платформенный инструмент молекулярной графики и широко используется для 3D-визуализации макромолекул.

Возможности PyMOL были значительно расширены различными плагинами, включая макромолекулярный анализ, гомологическое моделирование, докинг белок-лиганд, фармакофорное моделирование, VS и MD моделирование. Для доступа к этим программам мы использовали языки программирования R и Python.

Discover Studio (<https://discover.3ds.com/>) – это программа для молекулярного моделирования и различных способов 3D-визуализации полученных моделей.

При исследовании белков ELANE расчет отличий моделей дикого и мутантного вариантов белков после создания моделей в Phyre2 и I-Tasser проводился с помощью онлайн-сервиса Zhanggroup (<https://zhanggroup.org/TM-score/>, Медицинский факультет Мичиганского университета, США).

**Валидация 3D-моделей** проводилась с помощью PROCHECK и сервиса Ramachandran plot.

PROCHECK - это программа, используемая для валидации трехмерных структур белков. Она была разработана Романом Ласковски в Европейском институте биоинформатики (EBI) и в настоящее время широко используется в области структурной биологии.

PROCHECK анализирует белковые структуры с точки зрения их геометрии, включая длины связей, углы связей и торсионные углы, и сравнивает их с идеальными значениями для хорошо функционирующих структур. Программа генерирует серию графических результатов, которые суммируют качество структуры и выделяют любые области, которые могут быть проблематичными [170].

Графики Рамачандрана служат косвенным инструментом проверки стереохимии и геометрии комплекса путем установления того, что ни одна из геометрий не находится в запрещенных электростатически неблагоприятных областях графика [170, 173]. В работе использовался данный онлайн-сервис, применяющий этот метод: <https://swift.cmbi.umcn.nl/servers/html/ramaplot.html> (Нидерланды).

Аналогичным методом, дополняющим результаты моделирования, является MolProbity. Это веб-приложение всеатомной проверки структуры для макромолекулярной кристаллографии, которое объединяет программы проверки из лаборатории Ричардсона в Университете Дьюка, предназначенные для оценки качества трехмерных белковых структур.

Одной из основных особенностей MolProbity является анализ графика Рамачандрана, который исследует углы кручения основной цепочки белковой структуры и сравнивает их с ожидаемыми значениями для правильно свернутого белка. Программа также оценивает качество геометрии белка, включая длины связей, углы и не связанные с связями взаимодействия, и выявляет потенциальные столкновения или стерические перекрытия. Кроме того, MolProbity включает инструменты для оценки соответствия структуры белка экспериментальным данным, таким как карты электронной плотности или данные ядерного магнитного резонанса. Программа также предоставляет рекомендации по оптимизации сети водородных связей в структуре белка и выявлению потенциальных ошибок при размещении лигандов или других небелковых молекул.

В смоделированной молекуле белка в MolProbity выделяют благоприятную область, разрешенную область и область отклонений, которые относятся к различным областям на графике Рамачандрана, который представляет собой графическое изображение углов кручения основной части белковой структуры.

Область благоприятствования соответствует области диаграммы, где расположено большинство высококачественных белковых структур. В этой области углы кручения основной части близки к идеальным значениям для хорошо свернутого белка, что указывает на хорошо работающую и стабильную структуру.

Допустимая область примыкает к благоприятной области и представляет собой область, в которой углы кручения основы немного менее идеальны, но все еще приемлемы. Белковые структуры с крутящимися углами в пределах этой области считаются достаточно качественными, хотя они могут иметь некоторые незначительные структурные проблемы.

Область отклонений - это область графика, где углы кручения основной части значительно отличаются от идеальных значений, что указывает на

потенциально нестабильную или плохо сложенную структуру белка. Белковые структуры с кручениями в этой области считаются низкокачественными и могут потребовать значительного структурного уточнения или коррекции.

**Сравнение 3D-моделей** дикого и мутантного вариантов белков проводились с учетом метрики сравнения моделей (TM-score).

TM-score (Template Modeling score) - это широко используемая метрика для сравнения структурного сходства между двумя белковыми структурами. Это мера структурного совпадения между двумя белковыми структурами, учитывающая как среднеквадратичное отклонение (RMSD) выровненных остатков, так и длину выровненной области.

TM-score варьируется от 0 до 1, причем более высокие значения указывают на большее структурное сходство между двумя белками. Оценка TM-score, равная 1, указывает на полное совпадение структур двух белков, в то время как оценка TM-score, равная 0, указывает на отсутствие структурного сходства.

### **2.2.8 – Методы докинга для изучения влияния замен на функцию анализируемых белков иммунной системы**

Анализ докинга белков представляет собой моделирование молекулярных взаимодействий между двумя белками с целью определения, какие конкретные атомы одного белка связываются с атомами другого белка в трехмерном пространстве. Данный анализ может помочь понять, каким образом два белка могут связываться и какие конкретно атомы участвуют в этом процессе.

Для этого использовались программы Discovery Studio и PyMol, описанные ранее. Взаимодействия атомов в определенных аминокислотных остатках были рассчитаны для выявления сил связывания, которые были решающими в стабилизации образования комплексов рецептор-лиганд.

Кроме того, при исследовании белка ELANE использовалась программа AutoDock (Исследовательского института Скриппса). AutoDock предназначен для проведения как жесткого, так и гибкого моделирования докинга. При жесткой стыковке белок остается неподвижным, и только лиганд может перемещаться во время моделирования. При гибкой стыковке и белок, и лиганд могут перемещаться во время моделирования. Такая гибкость позволяет AutoDock моделировать конформационные изменения белка, которые могут произойти при связывании лиганда [24].

### **2.2.9 – Молекулярно-динамическое моделирование для оценки патогенности новых идентифицированных nsSNP.**

Симуляция молекулярной динамики (MDS) - это метод компьютерного моделирования для анализа физического движения атомов и молекул. Атомам и молекулам разрешается взаимодействовать в течение фиксированного периода времени, что дает представление о динамической «эволюции» системы. В наиболее распространенном варианте траектории атомов и молекул определяются путем численного решения уравнений движения Ньютона для системы взаимодействующих частиц, при этом силы между частицами и их потенциальные энергии часто рассчитываются с помощью межатомных потенциалов или силовых полей молекулярной механики.

Особенно важным применением симуляции молекулярной динамики является определение того, как биомолекулярная система будет реагировать на некоторое возмущение. В каждом из этих случаев, как правило, необходимо провести несколько симуляций как возмущенной, так и невозмущенной системы, чтобы выявить устойчивые различия в результатах.

Симулирование молекулярной динамики проводилось с помощью пакетов Maestro и Gromacs 4.5.3 от компании Schrödinger LLC [78].

Maestro создаёт подготовку к симуляции, в частности, добавляет в виртуальное окружение атомы водорода, назначает водородные связи и

минимизирует энергии молекулы. Кроме того, проводится «растворение» молекулы.

Белки дикого типа и мутанты были предварительно обработаны с помощью Protein Preparation Wizard в Maestro, который включал оптимизацию и минимизацию комплекса. Инструмент, доступный в программном пакете Maestro, который предназначен для предварительной обработки белковых структур перед проведением молекулярно-динамических симуляций. Он автоматически оптимизирует геометрию белковой структуры, добавляет водородные атомы, исправляет пропущенные или неправильные атомы, удаляет воду и лиганды, создает расширенный вакуумный слой вокруг белка и многое другое. Все эти шаги помогают устранить возможные проблемы с белковой структурой и готовят ее к проведению молекулярно-динамических симуляций. Все системы были подготовлены с помощью инструмента System Builder. Была выбрана TIP3P, модель растворителя с орторомбической ячейкой. (Transferable Intermolecular Intermolecular Interaction Potential 3 Points). При моделировании использовалось силовое поле OPLS 2005 [196]. Чтобы сделать модели нейтральными, были введены противоионы. Для имитации физиологических условий был добавлен 0,15 М хлорид натрия (NaCl). Для всего моделирования был выбран ансамбль NPT с температурой 300 К и давлением 1 атм. Модели были «расслаблены» перед моделированием. Траектории сохранялись для изучения через каждые 100 пс, а стабильность моделирования проверялась путем сравнения среднеквадратичного отклонения (RMSD) белка и лиганда с течением времени.

Gromacs выдаёт следующие результаты симуляции:

1) Среднеквадратичное отклонение (RMSD) – мера отклонения структуры во времени по сравнению со структурой при  $T = 0$  нс. RMSD рассчитывается путем измерения среднего расстояния между атомами двух белковых структур после их выравнивания друг с другом. Выравнивание обычно выполняется путем сопоставления положений атомов основы двух

структур. Значение RMSD отражает степень отклонения между двумя структурами, при этом меньшие значения RMSD указывают на большую степень сходства или совпадения.

2) Среднеквадратичная флуктуация (RMSF) – это среднеквадратическое колебание и является мерой степени подвижности или гибкости каждого атома или остатка в структуре белка. RMSF рассчитывается путем взятия среднеквадратичного отклонения каждого атома или остатка в структуре белка от его среднего положения в течение заданной симуляции или траектории. Полученные значения RMSF являются мерой изменчивости или флуктуации положения каждого атома или остатка, что может свидетельствовать о степени его подвижности или гибкости.

3) Отличия во вторичной структуре белка

4) Радиус гирации ( $R_g$ ) – это мера "расширения" белка. Радиус гирации рассчитывается как среднеквадратичное расстояние всех атомов в белке от центра масс белка. Таким образом производится оценка общей формы и компактности белка. На величину  $R_g$  влияют различные факторы, такие как размер, форма и гибкость белка. Например, более компактный белок будет иметь меньшее значение  $R_g$ , в то время как более вытянутый или гибкий белок будет иметь большее значение  $R_g$ . Значение  $R_g$  можно использовать для мониторинга стабильности и сворачивания белка с течением времени при моделировании молекулярной динамики.

5) Количество водородных связей, образованных между различными группами атомов в ходе моделирования молекулярной динамики. Наиболее часто используемым инструментом для расчета водородных связей в GROMACS является команда "g\_hbond", которая идентифицирует водородные связи между донорными и акцепторными группами атомов на основе геометрических критериев. В частности, инструмент рассчитывает критерии расстояния и угла для каждой потенциальной водородной связи и

сообщает о количестве водородных связей, удовлетворяющих этим критериям.

6) Площадь поверхности, доступной для растворителя (SASA, Solvent Accessible Surface Area) - мера площади поверхности белка или другой биомолекулы, которая подвергается воздействию окружающего растворителя. Он обычно используется в моделировании молекулярной динамики для анализа конформационных свойств белков и их взаимодействий с растворителем. В GROMACS SASA рассчитывается как площадь поверхности белка или другой биомолекулы, которая доступна для сферы зонда с определенным радиусом, обычно молекулы воды. Расчет производится путем разделения поверхности биомолекулы на сетку маленьких треугольников или квадратов и вычисления площади каждого элемента сетки, которая подвергается воздействию растворителя.

7) Анализ главных компонент (РСА) - это статистический метод, используемый в моделировании молекулярной динамики для анализа движения и конформационных изменений белков и других биомолекул. РСА-анализ проводится на файлах траекторий, полученных в ходе моделирования молекулярной динамики. Первый шаг в РСА-анализе включает построение ковариационной матрицы из атомных координат белка или другой биомолекулы на каждом временном интервале моделирования. Затем ковариационная матрица диагонализируется для получения набора собственных векторов и собственных значений, которые описывают коллективные движения системы. Результаты анализа РСА могут быть использованы для определения наиболее важных коллективных движений белка или другой биомолекулы, таких как движения доменов, изгиб петель или колебания петель. Эти движения могут дать представление о функциональных свойствах белка, таких как катализ фермента, связывание лигандов или белок-белковые взаимодействия.

8) Ландшафт свободной энергии - это графическое представление свободной энергии системы как функции одной или нескольких коллективных переменных, которые обычно выбираются для описания важных степеней свободы системы. Это мощный инструмент, используемый в моделировании молекулярной динамики для изучения термодинамики и кинетики сложных систем, таких как сворачивание белков, связывание лигандов или конформационные изменения. В GROMACS ландшафты свободной энергии часто строятся с помощью метода зонтичной выборки, который предполагает применение внешнего смещающего потенциала для ограничения системы по выбранной коллективной переменной. Затем проводится несколько симуляций, каждая с различным значением смещающего потенциала, для выборки всего диапазона выбранной коллективной переменной.

Свободная энергия системы как функция выбранной коллективной переменной может быть получена из распределения вероятности коллективной переменной, которое оценивается по данным, полученным в результате моделирования зонтичной выборки. Это распределение вероятностей может быть далее проанализировано с помощью таких методов, как метод анализа взвешенных гистограмм (WHAM), для получения ландшафта свободной энергии. Ландшафт свободной энергии может дать ценные сведения о термодинамике и кинетике исследуемой системы. Например, он может показать стабильные и метастабильные состояния системы, барьеры для конформационных изменений или связывания лиганда, а также механизмы сворачивания или агрегации белка.

**В качестве отправной точки для симуляции молекулярной динамики** в Maestro была использована структура нативной и мутантной киназы Aurora-A. Параметры моделирования были установлены в соответствии с нашими предыдущими работами, проведенными для белка Aurora-A и других белков. Системы были сольватированы (растворены) в прямоугольной коробке с молекулами воды TIP3P с краевым радиусом 10 Å.

Системы были нейтрализованы путем добавления 3 ионов натрия ( $\text{Na}^+$ ) в поле моделирования с помощью инструмента «genion», который сопровождает пакет Gromacs. Минимизация энергии проводилась в течение 5000 итераций методом сопряженного градиента с использованием силового поля GROMOS96 43a1. Критерий сходимости Emtol, который служит мерой стабильности молекулярной динамики, был установлен на значение 1000 кДж/моль/нм. Для регулирования температуры внутри симуляционного бокса был применен метод температурной связи Берендсена. Этот метод обеспечивает постоянство температуры системы во время симуляции путем регулирования температуры бокса на основе мгновенной температуры системы.

Электростатические взаимодействия рассчитывались с помощью метода Эвальда с сеткой частиц. Системы подвергались моделированию с ограничением положения в течение 5 нс, а затем моделированию без ограничения в течение 200 нс. Затем проводили сравнительный анализ структурных отклонений в нативной и мутантной структуре. Анализ RMSD, RMSF, SAS и Rg проводился с помощью инструментов `g_rms`, `g_rmsf`, `g_sas` и `g_gyrate` соответственно. Количество отдельных водородных связей (NHbonds) было рассчитано с помощью `g_hbond`.

Кроме того, мы использовали `g_densmap` для получения распределения атомной плотности нативного и мутантного белка. Все графики были построены с использованием инструментария Grace GUI версии 5.1.22. Далее мы провели анализ главных компонент с использованием метода существенной динамики (ED) в соответствии с протоколом в программном пакете Gromacs. Данный раздел представляет собой сокращенную версию нашей ранее опубликованной работы.

Мы использовали 100 наносекунд, Desmond, программное обеспечение от Schrödinger LLC, было использовано для моделирования молекулярной динамики. Интегрируя классическое уравнение движения Ньютона, МД-

симуляции обычно рассчитывают движение атомов с течением времени. Моделирование использовалось для прогнозирования стабильности белка в физиологической среде [86, 132, 185].

### **2.2.10 - Поиск вероятных генетических причин болезни у пациентов с клиническими диагнозами «врожденная нейтропения» и «синдром Хеннекама»**

Результаты секвенирования были выровнены со стандартной последовательностью hg38 генома человека с помощью программы Burrows-Wheeler Aligner (BWA) [119]. Затем для изменения порядка, сортировки и преобразования SAM-файлов использовалась программа SAMtools [203]. Однонуклеотидные варианты (SNVs) и варианты вставки / делеции (инсерции) были определены с помощью программы The Genome Analysis Toolkit версии 4.1.2.0 (GATK4, <http://www.broadinstitute.org/gatk/>) [199]. Только варианты в экзонической области с охватом чтения (или глубиной)  $>10\times$  и минимальным баллом качества картирования 30 были сохранены с помощью программы VCFtools, чтобы уменьшить количество некачественных вызовов, возникающих из-за ошибок картирования [64, 204].

Затем были отброшены все синонимичные SNV, вставки и делеции без сдвига рамки координат (Indels), а также варианты, чья экзоническая функция обозначена как «не применимо» или «неизвестно». Кандидатные SNV и Indels, полученные на предыдущих этапах, были далее отфильтрованы на наличие SNV и инделов в генах, связанных с первичным иммунодефицитом (ПИД). Кандидат SNV (или Indel) затем сохранялся и оценивался, как менее редкий, редкий или нераспространенный, если частота минорного аллеля (MAF) SNV (или Indel) была менее 0,01 во всех данных Консорциума Exome Aggregation Consortium (ExAC), 1000 Genomes (1000g) и The Genome Aggregation Consortium (ExAC). (1000g), и The Genome Aggregation Database (gnomAD).

Все SNVs (или Indel) далее считались патогенными, если были номинированы как вредные хотя бы в одной модели из следующих.

Был выполнен функциональный анализ с помощью скрытых марковских моделей (FATHMM), анализатора эффектов вариаций белка (PROVEAN) и комбинированного аннотационного анализа зависимой деплеции (CADD). FATHMM и PROVEAN были выполнены с использованием программы ANNOVAR [206], а CADD был выполнен с помощью онлайн-сервера (<https://cadd.gs.washington.edu/snv>, версия: GRCh38-v1.5). Все потенциально патогенные SNV и инделы были вручную проверены с использованием программы Базы данных однонуклеотидных полиморфизмов (dbSNP) (<https://www.ncbi.nlm.nih.gov/snp/>) и программного обеспечения Integrative Genome Viewer (IGV) версии 2.4.5. Если два MAF SNV (или Indel), полученных соответственно из ANNOVAR и dbSNP, были неоднозначны, MAF, полученный из dbSNP, оценивался как истинный. Кандидатная мутация считалась истинной, если мутация, идентифицированная GATK4, была обнаружена с помощью приложения «Интегративный геномный просмотрщик» (IGV).

### ГЛАВА 3 – ОЦЕНКА РАЗЛИЧИЯ ЭКСПРЕССИИ ГЕНОВ И ИССЛЕДОВАНИЕ ОСНОВНЫХ СИГНАЛЬНЫХ ПУТЕЙ У ПАЦИЕНТОВ С ДЕФИЦИТОМ *RVCK1*

Для того, чтобы приблизиться к пониманию механизмов, лежащих в основе патогенеза этой патологии, был проведен сравнительный анализ экспрессии генов между данными транскрипции пациентов с дефицитом *RVCK1*, с синдромом CINCA/NOMID, с синдромом Макл-Уэллса, с дефицитом мевалонаткиназы (MVK) и данными транскрипции здоровых детей (рисунк 10). Полученные в результате анализа гены с разницей в экспрессии были аннотированы и функционально обогащены, то есть была получена информация об их роли в функционировании организма, в каких сигнальных путях задействованы данные гены и при каких состояниях они экспрессировались по сведениям, полученным ранее другими исследователями.

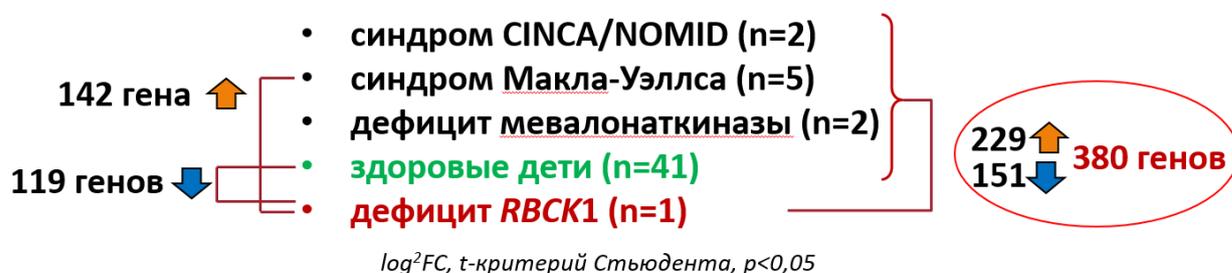


Рисунок 10 – Достоверные различия экспрессии генов при дефиците *RVCK1* в сравнении со здоровыми и пациентами с другими аутовоспалительными синдромами

Из набора данных GSE40561, включающего в сумме 48803 генов у разных лиц, было обнаружено 380 различных генов с дифференциальной экспрессией: 229 генов имели повышенную экспрессию, а 151 ген - сниженную. Сравнительный анализ транскрипции образцов здоровых лиц и

пациентов с дефицитом *RBCK1* показал наибольшее количество DEG – 119 генов с достоверно сниженной экспрессией (таблица 2). Кроме того, при сравнении образцов *RBCK1* и *MWS* было выявлено значительное отличие в сравнительно высокой экспрессии 142 генов при дефиците *RBCK1* (рисунок 10).

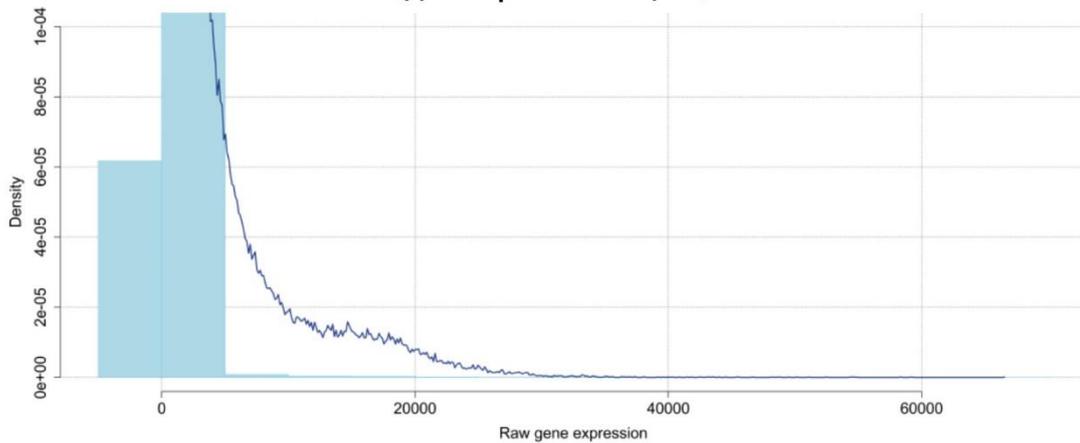
Таблица 2 – Список наиболее низко- и высоко-экспрессируемых генов при дефиците *RBCK1* в сравнении со здоровыми образцами

Гены со сниженной экспрессией				Гены с повышенной экспрессией			
№	Название	p	logFC	№	Название	p	logFC
1	<i>CISD2</i>	7.537936e-18	-0.5969954	1	<i>HS:551123</i>	6.458791e-13	3.1687011
2	<i>EPB41</i>	4.108343e-16	-0.6256048	2	<i>HS:552143</i>	1.583777e-07	1.2937618
3	<i>LOC253012</i>	5.739984e-16	-0.6810695	3	<i>F LJ00312</i>	3.640115e-07	1.6176159
4	<i>FAM83A</i>	1.703811e-13	-0.7337758	4	<i>HS:19339</i>	6.034732e-07	0.4643016
5	<i>NUP98</i>	5.965986e-13	-0.4364548	5	<i>ANKMY 2</i>	6.517125e-06	0.4189648
6	<i>CHD2</i>	6.279939e-12	-0.4504437	6	<i>RPS29</i>	1.978019e-05	0.4123195
7	<i>RAP1GAP</i>	1.338901e-11	-0.9886351	7	<i>HS:531457</i>	7.585137e-05	0.4194266
8	<i>HS:563750</i>	1.539916e-11	-0.5622475	8	<i>HS:542923</i>	0.000385	1.1559277
9	<i>ABCC13</i>	1.908913e-11	-0.5476611	9	<i>HIST1H2BI</i>	0.000278	0.9167271
10	<i>MAOA</i>	3.456357e-11	-0.7114969	10	<i>PLA2R1</i>	0.000193	0.8924394

Все неэкспрессированные гены были удалены, и была создана новая диаграмма главных компонент (PCA). Распределение данных до и после нормализации можно увидеть на гистограммах (рисунок 11) и боксплотах (рисунок 12). Получены данные об общей численности 532 DEG (генов с измененной экспрессией) в нашем втором наборе данных GSE31064 после стандартизации результатов микрочипов, среди которых было обнаружено 211 генов с пониженной экспрессией и 321 генов с повышенной экспрессией.

Следующим этапом стало определение вовлеченности дифференциально экспрессируемых генов при дефиците *RBCK1* в ключевые сигнальные пути и оценка их влияния на биологические функции.

**А** Гистограмма распределения необработанных данных  
(до нормализации)



**Б** Гистограмма распределения необработанных данных  
(после нормализации)

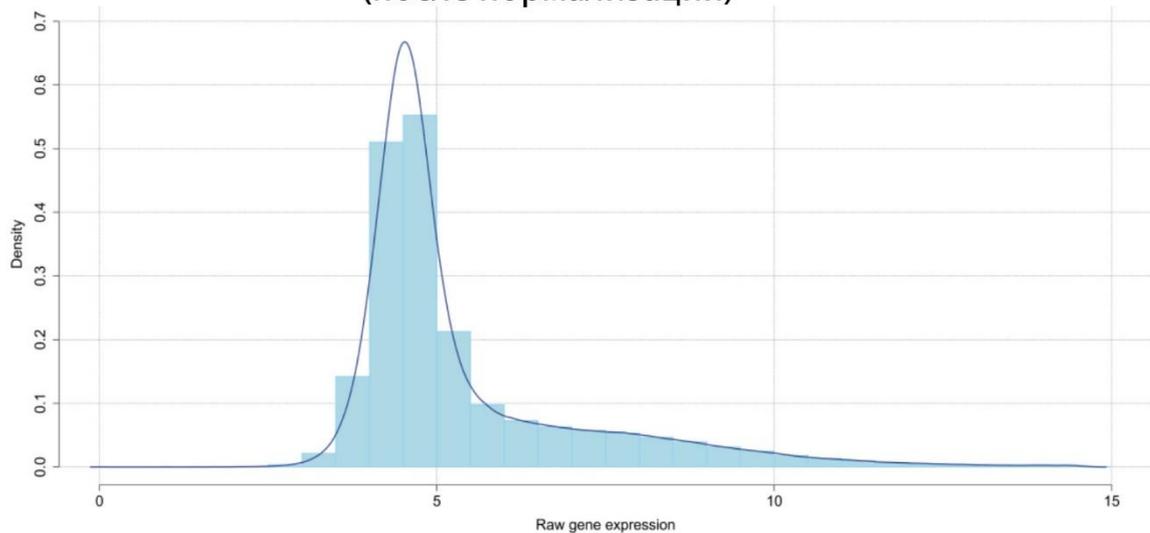


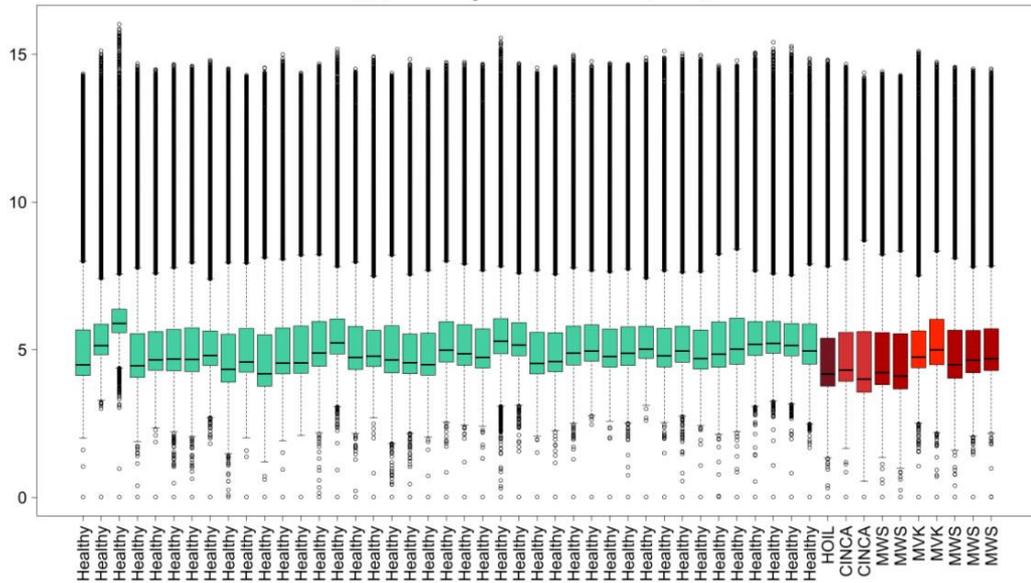
Рисунок 11 – Гистограмма распределения необработанных данных экспрессии до нормализации (А) и после (Б).

Примечание. По оси абсцисс - необработанные значения экспрессии в условных единицах, по оси ординат – количество генов с определенным уровнем экспрессии.

Пути иммунологического ответа, воспалительного ответа и фосфорилирования белков в категории биологических процессов были

перепредставлены в ГО и путях, полученных из кластеров коэкспрессированных генов (Таблица 3, Рисунок 13).

### А Боксплоты необработанных данных экспрессии (до нормализации)



### Б Боксплоты необработанных данных экспрессии (после нормализации)

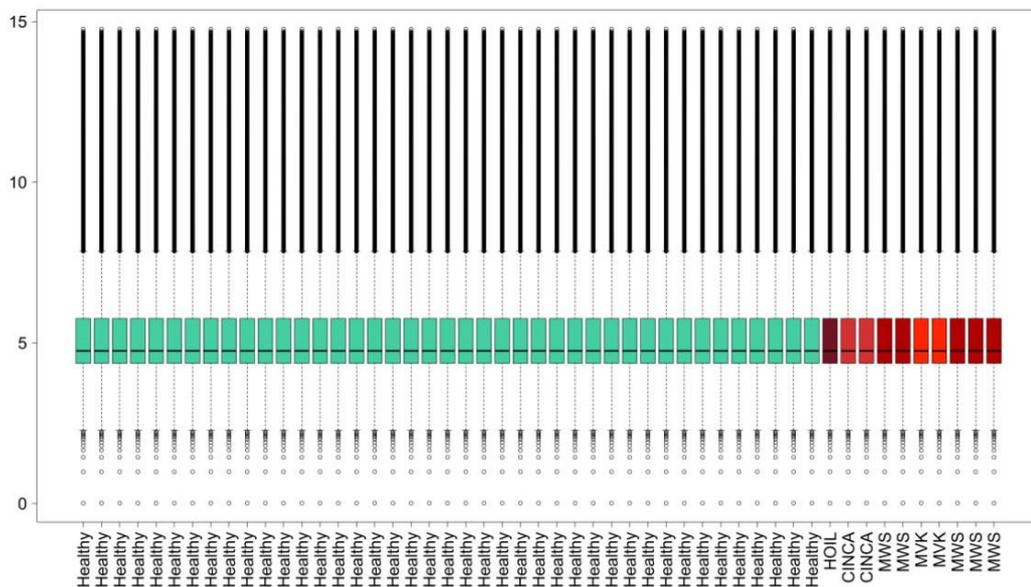


Рисунок 12 – Боксплоты необработанных данных экспрессии до нормализации (А) и после (Б).

Примечание. По оси абсцисс – образцы транскриптов, по оси ординат – уровень экспрессии.

Пути связывания белков и РНК-связывания были наиболее обогащенными в категории молекулярных функций. Наконец, цитоплазма, цитозоль и ядро были наиболее обогащенными путями в категории клеточных компонентов. После загрузки идентификаторов даунрегулярных генов (со сниженной экспрессией) из сравнения между здоровыми образцами и образцами с дефицитом *RBSK1* в базу данных WikiPathways, было отобрано для анализа 425 сигнальных путей. Среди обращающих на себя внимание были сигнальные пути, связанные с вирусом SARS-CoV-2: WP5115, WP5039, WP5098, среди которых было 4 гена из нашего набора: *FAM83A*, *IFI27*, *NUP98* и *TSC1*. Кроме того, было установлено, что ген *HP*, который находился в группе даун-регулируемых генов в сравнении между здоровыми и MWS образцами, был вовлечен в путь, связанный с COVID-19.

Таблица 3 – GO-аннотации с наименьшим коэффициентом ложных обнаружений.

Номер GO аннотации	Описание GO (функции генов, клеточных компонентов и биологических процессов)	Доля аннотированных генов	Коэффициент ложных обнаружений
GO:0071799	клеточный ответ на стимул простагландина D	2/5	0.0125
GO:0021796	регионализация коры головного мозга	2/7	0.0162
GO:0030656	регуляция процесса метаболизма витаминов	2/12	0.0349
GO:0051712	позитивная регуляция киллинга клеток другого организма	2/13	0.0376
GO:0001829	дифференцировка трофэктодермальных клеток	2/15	0.0456

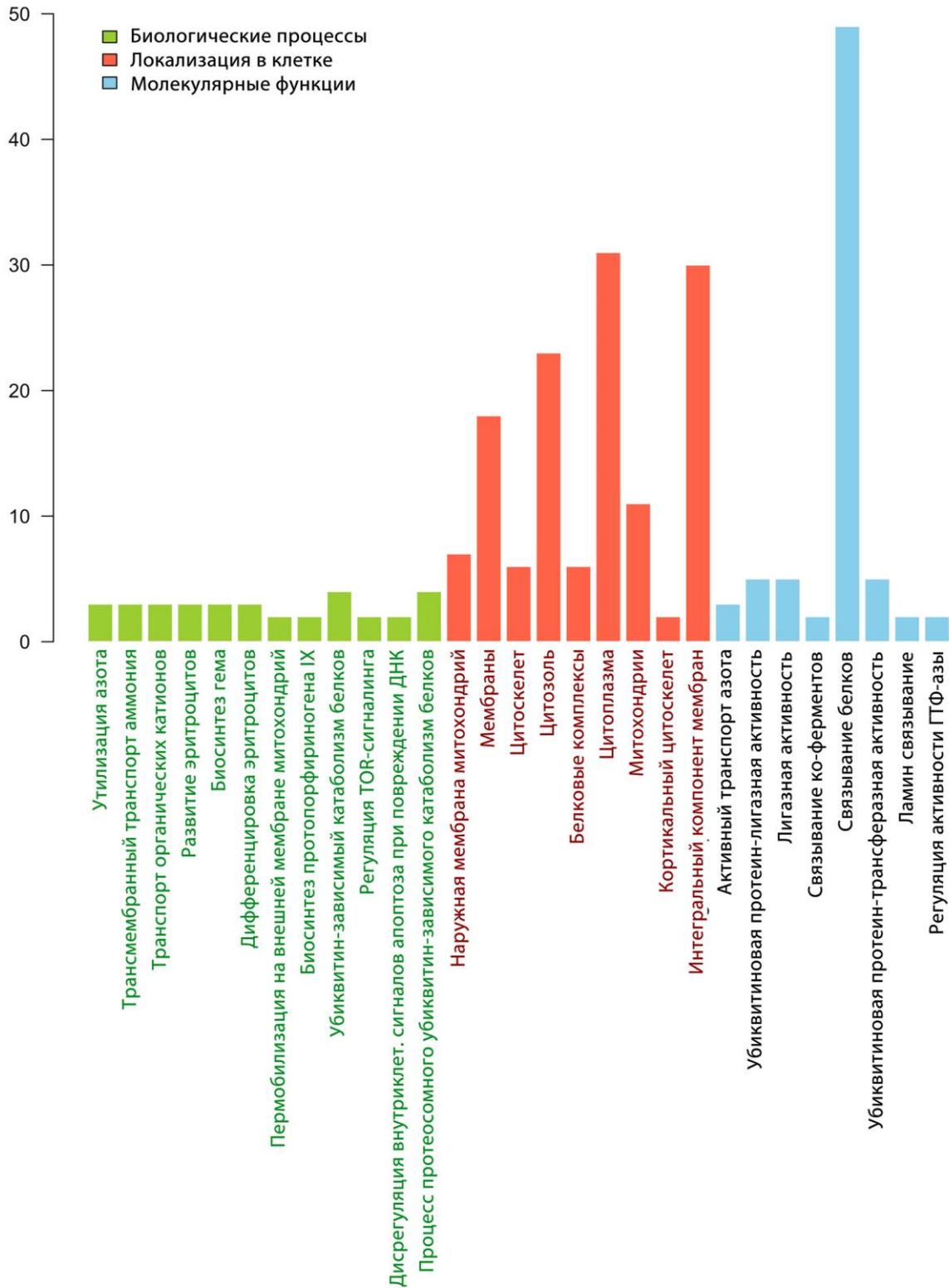


Рисунок 13 – Гистограммы генной онтологии (GO), переобогащенных путей среди генов с пониженной экспрессией, полученных при сравнении между здоровыми и пациентами с дефицитом RBSK1.

*Примечание:* Термины GO относятся к «молекулярной функции», «клеточному компоненту» и «биологическому процессу».

Исходя из кластеров, сформированных с помощью *semiTool*, и ранее определенных ДЭГ, было сформировано 30 отдельных графиков взаимодействия белков-протеинов. Только один график, который изображает взаимодействие между 54 белками одного из 14 кластеров *semiTool*, показал статистическую значимость (рисунк 14).

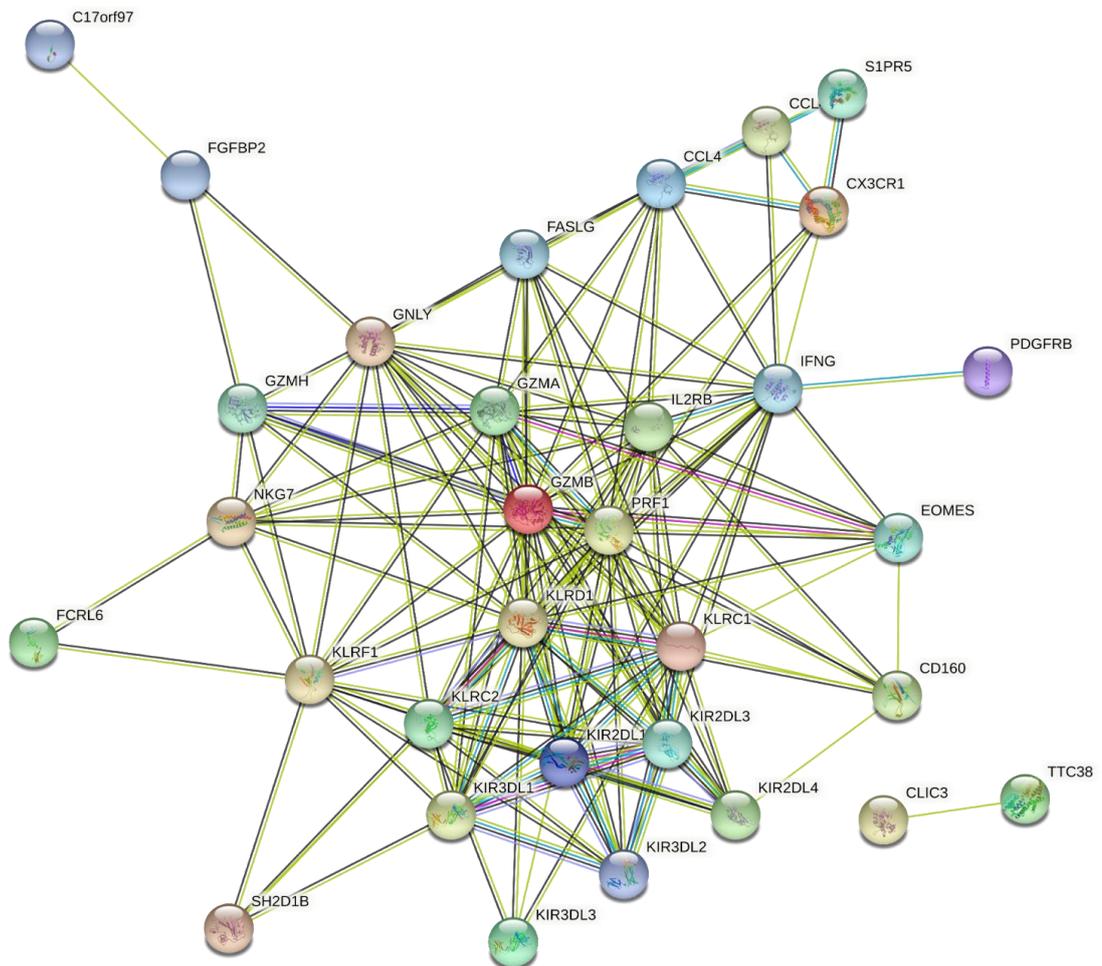


Рисунок 14 – Статистически значимые белок-белковые взаимодействия между 54 белками по результатам анализа сервиса STRING

Подсемейства лектиноидных рецепторов киллерных клеток, а именно *KLRD1*, *KLRC1*, *KIR2DL1*, *KIR2DL2*, *KIR2DL3*, *KIR2DL4*, *KIR3DL2* и *KIR3DL3* имели наиболее тесную взаимосвязь на схеме белковых взаимодействий. По сравнению со здоровыми людьми и людьми с синдромом CINCA, было

обнаружено понижение экспрессии нескольких генов. Кроме того, гены хемокинов (*CXCL8* и *CXCL10*) были особо выделены, потому что дефицит их активности может привести к серьезным ошибкам в функционировании клеток или их гибели из-за нарушения хемокиновой передачи. Р-значение для обогащения РРІ созданной сети составило  $1,0e-16$  (рисунк 14).

При функциональном обогащении дифференциально экспрессированных генов при дефиците *RBCK1* (относительно здоровых), были выявлены вовлеченности таких генов в несколько значимых сигнальных путей. В частности, задействованы сигнальные пути развития лейшманиоза, восприимчивости к стафилококковой инфекции, холеры, НК-клеточной цитотоксичности и различные другие пути, влияющие на иммунный ответ. Это не означает, что при дефиците *RBCK1* повышается вероятность соответствующей патологии, однако становится более понятным системное влияние дефицита одного белка на различные процессы, так или иначе затрагивающие иммунную систему и противoinфекционную защиту (таблицы 4 и 5).

Таблица 4 – Группы из дифференциально экспрессированных генов при дефиците *RBCK1* (относительно здоровых), вовлеченных в несколько значимых сигнальных путей, полученные по результатам анализа функционального обогащения с помощью базы KEGG.

Передача внутриклеточных сигналов	Линия гемопоэтических клеток	Дилатационная кардиомиопатия	Биосинтез пантотената и CoA	Инфекция <i>Vibrio cholerae</i>
<i>TIAM1</i>	<i>MME</i>	<i>TPM3</i>	<i>VNN1</i>	<i>TCIRG1</i>
<i>FOXO3</i>	<i>HLA-DRB5</i>	<i>ACTG1</i>	<i>UPB1</i>	<i>ACTG1</i>
<i>STAT1</i>	<i>CD8B</i>	<i>ITGB1</i>	<i>VNN2</i>	<i>PRKACB</i>
<i>GNB4</i>	<i>IL1R2</i>	<i>ITGA4</i>	<i>ZNF586</i>	
<i>GRB2</i>	<i>CSF3R</i>	<i>PRKACB</i>	<i>VNN3</i>	
<i>CXCL5</i>	<i>CD8A</i>			
<i>PIK3CG</i>	<i>IL1B</i>			
<i>ROCK1</i>				
<i>PIK3R1</i>				
<i>VAV3</i>				

Таблица 5 - Другие группы дифференциально экспрессированных генов при дефиците *RBCK1* (относительно здоровых), вовлеченных в несколько значимых сигнальных путей, полученные по результатам анализа функционального обогащения с помощью базы KEGG.

NK-опосредованная цитотоксичность	Процессинг и презентация антигена	Развитие лейшманиоза	Инфекция <i>Staphylococcus aureus</i>	Сигнальный путь RIG-I-подобных рецепторов
<i>KIR2DL3</i>	<i>KIR2DL3</i>	<i>FCGR2A</i>	<i>FCGR2A</i>	<i>CXCL8</i>
<i>KIR2DL1</i>	<i>KIR2DL1</i>	<i>PTGS2</i>	<i>KRT23</i>	<i>MAPK13</i>
<i>KIR2DS5</i>	<i>KIR2DS5</i>	<i>HLA-DRB5</i>	<i>FCGR1A</i>	<i>CXCL10</i>
<i>KIR3DL1</i>	<i>KIR3DL1</i>	<i>FCGR3B</i>	<i>FCAR</i>	<i>ISG15</i>
<i>KLRD1</i>	<i>KLRD1</i>	<i>NCF4</i>	<i>FPR2</i>	
<i>KLRC2</i>	<i>KLRC2</i>	<i>TLR2</i>		
<i>KIR3DL2</i>	<i>KIR3DL2</i>	<i>IL1B</i>		
<i>KIR2DS3</i>	<i>KIR2DS3</i>	<i>NCF2</i>		
<i>KIR3DL3</i>	<i>KIR3DL3</i>			
<i>KLRC1</i>	<i>KIR2DL4</i>			
<i>SH2D1B</i>	<i>KLRC1</i>			
<i>GZMB</i>	<i>IFNG</i>			
<i>FAS3LG</i>				
<i>PRF1</i>				

Таким образом, в данной главе представлены результаты биоинформационного анализа, проведенного для получения информации, помогающей раскрыть патогенез формирования патологии пациентов с дефицитом *RBCK1*.

Проведенные нами ранее исследования показали, что в повышенной восприимчивости к вирусным инфекциям (таким как грипп или герпес) могут иметь значение 14 новых генов, которые вовлечены в биогенез рибосомы, процессинг рРНК, экспрессию генов, процессинг мРНК, ядерный просвет, внутриклеточные немембранно-ограниченные органеллы, нуклеоплазму, процессомы малых субъединиц, путь обработки и презентации антигенов и биогенез рибосом эукариот. При данном синдроме данные сигнальные пути оказались не вовлеченными в патологический процесс, что ещё раз подчеркивает особенность патогенеза дефицита *RBCK1* и это совпадает с

наблюдениями о повышенной восприимчивости именно к бактериальным инфекциям.

Что касается противовирусной активности иммунного ответа у *RVCK1*-дефицитных пациентов, то, согласно клиническим данным, повышенная восприимчивость к ним - редкий случай. Наше исследование продемонстрировало участие сигнального пути, ответственного за ответ на коронавирусную инфекцию. Однако не было доказано, что *RVCK1*-дефицитные пациенты подвержены риску заболевания COVID-19. Это имеет объяснение, связанное с убиквитинированием интерферонового регуляторного фактора 3, важной сигнальной молекулы, связанной с Toll/IL-1R домен-содержащим адаптером, индуцирующим IFN и TLR 3.

В результате исследований выявлено высоко достоверное снижение экспрессии гена *CISD2* у данного пациента с дефицитом *RVCK1*. Известно, что дефект *CISD2*, ведет к ЭПР-стрессу (стрессу эндоплазматического ретикулума) и апоптозу [44], в том числе мононуклеаров периферической крови. Принимая во внимание тесную функциональную взаимосвязь данного белка с процессами апоптоза и клеточного стресса, можно предположить, что влияние низкой экспрессии данного гена на патогенез оказывает негативное влияние на устойчивость мононуклеаров периферических крови к апоптозу и клеточной гибели.

Вместе с тем, изменение активности генов сигнальных путей mTOR, PI3K/AKT, Rho и Nf-kB, прямо или косвенно влияет на экспрессию генов иммунной системы.

Все найденные отличия в экспрессии генов не объясняют неопосредственную причину повышенной восприимчивости к пиогенным инфекциям, однако раскрывают некоторые аспекты молекулярных взаимосвязей, позволяя приблизить понимание патогенеза дефицита *RVCK1*.

**Список публикаций по материалам данной главы**

1. Is Up-Regulation Gene Expression of the Certain Genes During the Viral Respiratory Tract Infection Would Have Any Influence in Pathogenesis of the SAR-CoV-2 Infection? / K. Shinwari, G. Liu, M. Bolkov, I. Ahmad, M. Daud, I. Tuzankina, V. Chereshevnev // *Acta Med. Iran.* – 2020. – V. 58 (5). – P. 246-248. (Scopus Q4).
2. Identifying Main Genes and Pathways by Using Gene Expression Profiling in Primary Immunodeficiency HOIL-1/RBCK1 Disorder Patients / K. Shinwari, G. Liu, M.A. Bolkov, M. Ullah, I. Tuzankina // *Acta Med. Iran.* – 2021. – V. 59 (5). – P.265-279 (Scopus, ~~Web of Science~~ Q4).
3. Gene expression and pathway analysis in patients with inborn error of TLRs and IL-IRs signaling using microarray data / K. Shinwari, G. Liu, M.A. Bolkov, I.A. Tuzankina, V.A. Chereshevnev // *AIP Conference Proceedings.* – 2022. – V. 2390. – 030088 (Scopus).
4. Additional Pathogenic Pathways in RBCK1 Deficiency / E.I. Demicheva, K. Shinwari, K.S. Ushenin, M.A. Bolkov // *Mathematical Biology and Bioinformatics.* –2022. –V. 17 (2). – P. 174-187. (RSCI).
5. Checking gene expression profile associated with IRF7 and UNC93B deficient patient peripheral blood mononuclear cells infected with pH1N1 influenza virus / K. Shinwari, G. Liu, M.A. Bolkov, I.A. Tuzankina, V.A. Chereshevnev // *AIP Conference Proceedings.* – 2022. – V. 2390. – 030089 (Scopus).

## ГЛАВА 4 – ИССЛЕДОВАНИЕ ВЛИЯНИЯ ВЫЯВЛЕННЫХ НЕСИНОНИМИЧНЫХ ОДНОНУКЛЕОТИДНЫХ ЗАМЕН В ГЕНАХ *ELANE* И *TCIRG1* НА СТРУКТУРУ И ФУНКЦИЮ БЕЛКОВ *ELANE* И *TCIRG1*

Синдромы врожденной нейтропении - это группа редких болезней, проявляющихся с рождения и характеризующихся низким уровнем нейтрофилов, необходимых для борьбы с инфекциями. Наиболее частым и серьезным иммунодефицитом с врожденной нейтропенией является тяжелая врожденная нейтропения – редкое заболевание кроветворной системы, встречающееся по данным Donadieu J. et al., 2013, примерно у 1 из 100 000 человек европейского происхождения, многие случаи которого наследуются по аутосомно-доминантному типу [65]. Несмотря на то, что было выявлено несколько причинных генов, генетическая основа >30% случаев остается неизвестной.

Около половины всех случаев тяжелой врожденной нейтропении вызваны вариантами в гене *ELANE*. На долю других связанных генов, включая *TCIRG1*, приходится лишь небольшой процент всех случаев этого заболевания.

В работе были данные о nsSNP в генах *TCIRG1* и *ELANE*, полученные из онлайн-базы данных NCBI dbSNP, а также данные о nsSNP для гена *TCIRG1* из NGS-данных одного пациента, проанализированных с помощью биоинформационных методов, включая *in silico*-моделирование и симуляцию молекулярной динамики, что позволило выявить их потенциал дестабилизации структуры и функции белков *TCIRG1* и *ELANE*.

#### 4.1 – Определение вредоносности несинонимичных однонуклеотидных замен в генах *TCIRG1* и *ELANE* с помощью инструментов SIFT и PolyPhen-2

База данных NCBI сообщает о 5627 SNP в гене *TCIRG1*. Первым этапом нужно было выбрать только те полиморфизмы, которые вызывают замену аминокислоты. Выяснилось, что менее 2% замен - 811 из 5627, являются кодирующими несинонимичными (миссенс) заменами (nsSNP). В гене *ELANE* было определено всего 301 nsSNP из 3646 SNP.

Программы SIFT и PolyPhen-2 рассчитывают влияние nsSNP на функцию белка и оценивают, приемлема ли индуцированная аминокислота в конкретном месте. SIFT классифицирует каждый nsSNP по очкам, и те, которые имеют очки ниже порога, признаются «допустимыми» или доброкачественными, а те, которые имеют очки выше порога, считаются «пагубными» или вредоносными. Для SIFT порог отнесения к пагубным nsSNP был определен количеством очков  $>0.5$ .

В гене *TCIRG1* программа SIFT предсказала 118 вероятно вредоносных nsSNP, PolyPhen-2 – 64, а их взаимно пересекающийся результат объединенного анализа позволил отобрать только 34 nsSNP, которые привели к замене аминокислот из общего числа 811 nsSNP. В *таблице 6* представлена часть полученных результатов анализа.

Для замен в гене *ELANE* программа SIFT определила 21 nsSNP как вредоносный полиморфизм, в то время как объединенный анализ SIFT и PolyPhen-2 указал только 8 nsSNP как вредоносные из общего числа 301 nsSNP (*таблица 7*).

Для подтверждения деструктивности полиморфизмов, отобранных через SIFT и PolyPhen-2, были использованы дополнительные инструменты *in silico*.

Таблица 6 – Патогенные несинонимичные однонуклеотидные замены (nsSNP) в гене *TCIRG1* по результатам анализа в SIFT и PolyPhen-2

ID nsSNP	A.К.	SIFT	Порог	PolyPhen-2	Порог
rs36027301	R56W	Del	0	Pd	0.999
rs368945298	M546V	Del	0	Pd	0.999
rs115854062	P572L	Del	0	Pd	1
rs150260808	I721N	Del	0	Pd	1
rs137853150	G405R	Del	0	Pd	1
rs137853151	R444L	Del	0	Pd	1
rs147580611	F610S	Del	0	Pd	1.00
rs148921764	E722K	Del	0	Pd	1.00
rs140963213	A417T	Del	0.002	Pd	1
rs144775787	A778V	Del	0.46	Pd	0.883
rs145080707	R213W	Low	0.012	Pd	1
rs150648332	R57H	Del	0.001	Pd	1.00
rs150260808	I721N	Del	0	Pd	1
rs201329219	R109W	Del	0.014	Pd	1.00
rs367703865	R191H	Del	0.32	Pd	0.999
rs371214361	S532C	Del	0.001	Pd	1.00
rs199914625	S474W	Del	0	Pd	1
rs200851583	G458S	Del	0	Pd	1
rs371658110	G192S	Del	0.003	Pd	1.00
rs370319355	R50C	Del	0	Pd	1
rs376351835	F529L	Del	0.013	Pd	1.00
rs371004297	G379S	Del	0.011	Pd	1.00
rs200209146	N730S	Del	0.022	Pd	1.00
rs200415611	V375M	Del	0.001	Pd	1.00

*Примечание:* ID nsSNP – идентификатор несинонимичной однонуклеотидной замены, A.К. – положение аминокислоты, Del – высокая вероятность патогенности мутации; Low – низкая вероятность патогенности мутации, Pd – предсказана вероятная патогенность мутации

Результаты предсказания вредоносности значимых nsSNP в гене *TCIRG1* с помощью 17 дополнительных инструментов анализа представлены на рисунке 15 и в таблице 8.

Таблица 7 – Патогенные несинонимичные однонуклеотидные замены (nsSNP) в гене *ELANE* по результатам анализа в SIFT и PolyPhen-2

ID nsSNP	А.К.	SIFT	Порог	PolyPhen-2	Порог	Частота аллели
rs201163886	R34W	Del	0.002	Pd	1	
rs28931611	C71R	Del	0	Pd	1	6.076e-06
rs137854449	V101M	Del	0.005	Pd	0.964	
rs137854448	P139L	Del	0	Pd	1	
rs199558534	R143C	Del	0.048	Pd	1	
rs57246956	C151Y	Del	0	Pd	1	
rs201788817	A166T	Del	0.33	Pd	0.976	
rs199891906	A166V	Del	0.23	Pd	0.582	
rs193141883	T175M	Del	0.008	Pd	1	gnomAD_exome0.0005
rs200449787	R182H	Del	0.015	Pd	1	
rs367663236	V190M	Del	0.047	Pd	1	
rs201723157	R193W	Del	0.006	Pd	1	
rs201139487	G203S	Del	0	Pd	1	4.094e-06
rs137854446	L206F	Del	0	Pd	1	
rs201664319	N209K	Del	0.03	Pd	0.983	
rs140880838	G210R	Del	0.019	Pd	1	
rs137854451	G214R	Del	0.002	Pd	1	
rs200384291	F218L	Del	0.011	Pd	0.998	

*Примечание:* ID nsSNP – идентификатор несинонимичной однонуклеотидной замены, А.К. – положение аминокислоты, Del – высокая вероятность патогенности мутации; Pd – предсказана вероятная патогенность мутации

Все перечисленные аминокислотные замены предсказывались как вредоносные большинством алгоритмов (FATHMM-MKL, SNP-GO, PHD-SNP, PANTHER, SNAP2, P-MUT PROVEAN, FATHMM, LRT, M-CAP, CAAD, META SVM, METALR, Mutation Assessor и Mutation Taster), использованных в исследовании. Каждый из использованных в данном исследовании алгоритмов имеет свой уникальный порог и критерий оценки, на основе которого определяются пагубность или допустимость замены.

Для гена *TCIRG1* комбинация алгоритмов SIFT и VEST 3 определили только 6 nsSNP (10% из отобранных ранее) как пагубные, а 51 - как допустимые. PolyPhen-2, FATHMM, M-CAP и PANTHER показали

наибольший процент пагубных предсказаний. При использовании метода SNAP2 - 41 замена была признана пагубной (71%), а 16 предсказаний не имели никакого эффекта (оценка SNAP2 равна 100). Для предсказания воздействия 54 (92%) nsSNP на белок TCIRG1 использовался алгоритм PANTHER, 48 nsSNP из которых, скорее всего, могут оказывать разрушающее воздействие, 6 nsSNP могут оказывать возможно разрушающее воздействие, а 3 nsSNP, вероятно, безвредны. В частности, те, у которых время больше 450 мс, были классифицированы как возможно разрушающие, те, у которых время между 450 мс и 200 мс, были классифицированы как вероятно безвредные, а те, у которых время менее 200 мс, не были классифицированы. Программа PROVEAN, предназначенная для предсказания воздействия SNPs на белковую функцию, определила 22 (38%) nsSNP в гене *TCIRG1* как значительно вредоносные (относительно их влияния на структуру и функцию белка), в то время как 35 nsSNP были определены как нейтральные в соответствии с критериями порогового значения PROVEAN ( $> -2.667$ ). Используя порог ( $> 0,65$ , от 5,545 до 5,975), оценщик мутаций классифицировал 24 nsSNP как вредные, из которых 12 были классифицированы как высокие, 17 как умеренные, 5 как низкие, а 19 не были обнаружены. FATHNMM и FATHMM-MKK ( $<0.5$ ), CADD ( $>15$ ), DANN ( $>0.5$ ), Mutation Taster ( $<0.5$ ), и их соответствующие показатели предсказывают более 75-90% nsSNP как вредные/разрушающие. P-Mut предсказывает 45 (75.21%) вредных, 7 нейтральных и для 5 nsSNP данные отсутствовали с порогом ( $<0.5$ ). LRT предсказывает 42 (77%) вредных nsSNP с результатом ( $>0.001$ ) и 13 нейтральных. PhD-SNP, SNP-GO и M-CAP определили 47 (82%), 35 (61%) и 54 (94.73%) nsSNP как вредные соответственно. Кроме того, MetalR и MTA-SVM определили 10 (17%) и 37 (64%) nsSNP как вредные, соответственно (рисунки 15).

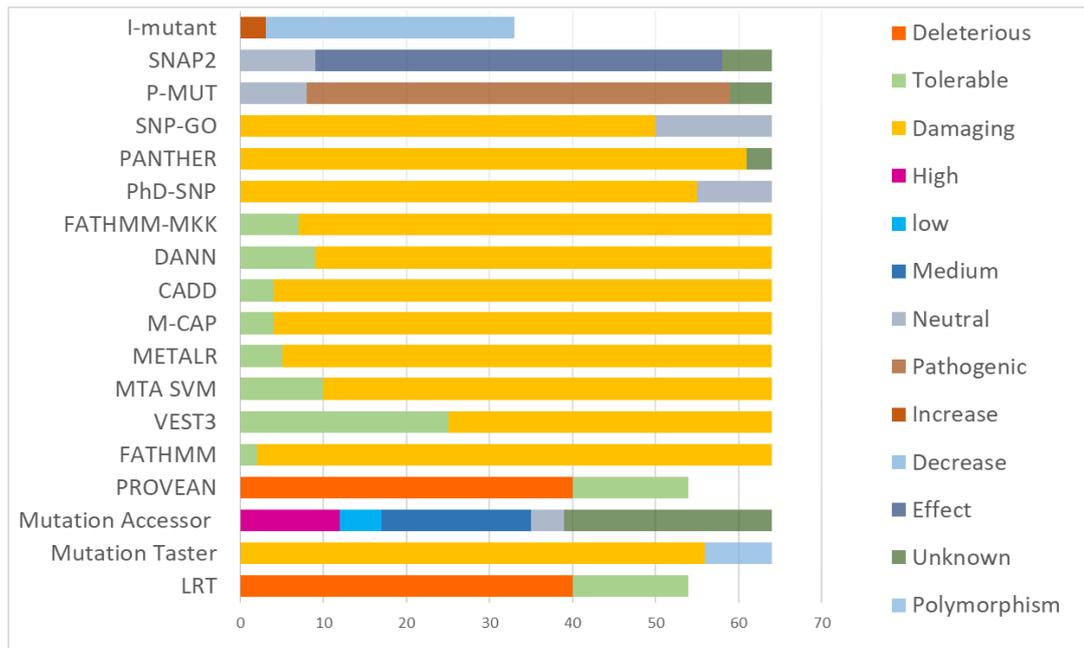


Рисунок 15 – Результаты предсказания влияния 64 nsSNP в гене *TCIRG1*, проанализированных восемнадцатью вычислительными инструментами

В заключение, основываясь на оценке позиций замещения с использованием PANTHER, PROVEAN score, SIFT score, SNP&GO, FATHMM, LRT, M-CAP, VEST3, CAAD, METALR, Mutation Assessor, Mutation Taster, FATHMM-MKL, PHD-SNP score и PolyPhen-2, группа из 15 nsSNP, включая P572L, M546V, I721N, F610S, A732T, F51S, A717D, E722K, R57H, R109W, R191H, S532C, G192S, F529L и H804Q, была признана значительно вредоносной всеми современными методами. Только LRT не подтвердил результаты влияния замены A717D, предсказанным другими инструментами. Результаты применения всех алгоритмов предсказания были статистически значимыми и сильно коррелировали друг с другом (р-значение для Т-теста Стьюдента между инструментами составило 0,001).



Продолжение таблицы 8.

Замена	LRT	Mutation Taster	Mutation Accessor	PROVEAN	FATHMM	VEST3	MTA SVM	METALR	M-CAP	CADD	DANN	FATHMM- MCK	PhD- SNP	PANTHER	SNP- GO	P-MUT	SNAP2
R444L	D	D	-	D	D	D	D	D	D	D	D	D	D	D	D	D	D
R56P	D	D	-	D	D	D	D	D	D	D	D	D	D	D	D	D	D
G379S	D	D	-	D	D	D	D	D	D	D	D	D	D	D	D	D	D
R757C	D	D	M	D	D	D	D	D	D	D	T	D	D	D	D	D	D
N730S	D	D	M	D	D	T	D	D	D	D	D	D	D	D	D	D	D
V375M	D	D	-	D	D	D	D	D	D	D	D	D	D	D	D	D	D
T314M	D	D	-	D	D	D	D	D	D	D	D	D	D	D	D	D	D
D517N	D	D	H	D	D	T	D	D	D	D	D	D	D	D	D	D	D
R92W	D	D	M	D	D	T	D	D	D	D	D	D	D	D	D	D	D
T368M	D	D	-	D	D	D	D	D	D	D	D	D	D	D	D	D	D
A417T	D	D	H	D	D	D	D	D	D	D	D	T	D	D	D	D	D
R363C	D	D	-	D	D	D	D	D	D	D	D	D	D	D	D	D	D
R56W	D	D	H	D	D	T	D	T	-	D	D	D	D	D	D	D	D
A778V	D	D	M	D	D	D	D	D	D	D	D	D	D	D	D	N	N
R50C	D	D	M	D	D	T	D	D	D	D	D	D	D	D	D	D	-
V52L	D	D	M	T	D	T	D	D	-	D	D	D	D	D	D	D	D

*Примечание:* Замена – аминокислотная замена в молекуле; D – вредоносная замена, T – толерантная, N – нейтральная, M – средняя вероятность, L – низкая, H – высокая, P – патогенная, - – нет данных.

Таблица 9 - Оценка патогенности выявленных замен в *ELANE* с помощью различных инструментов предсказания

Замена	LRT	Mutation n Taster	Mutation Accessor	PROVEAN	FATHM M	VEST3	MTA SVM	METALR	M-CAP	CADD	DANN	FATHMM- MKK	PhD-SNP	PANTHE R	SNP-GO	P- MUT	SNAP2
C71Y	0.001 -	1 D	4.21 H	-11.22 D	-4.7 D	0.944 D	1.09 7 D	0.968 D	0.973 D	23.9 D	0.987 D	0.817 D	0.933 D	0.907 D	0.982 D	0.71 D	77 D
P139L	0.002 -	1.00 D	2.72 M	-8.79 D	-3.56 D	0.912 D	0.95 6 D	0.886 D	0.942 D	26.6 D	0.999 D	0.747 D	0.733 D	0.888 D	0.855 D	0.85 D	59 D
C151Y	0.001 -	1.00 D	2.745 M	-10.41 D	-3.34 D	0.925 D	1.00 5 D	0.900 D	0.923 D	25.2 D	0.996 D	0.916 D	0.939 D	0.988 D	0.939 D	0.9 D	82 D
T175M	0.002 -	0.94 0 D	1.98 M	-4.76 D	-2.48 D	0.774 D	0.79 8 D	0.812 D	0.891 D	33 D	0.999 D	0.685 D	0.242 N	0.833 D	0.719 D	0.7 D	20 D
G203S	0.001 -	0.99 7 D	3.865 H	-5.4 D	-7.34 D	0.737 D	0.91 5 D	0.966 D	0.966 D	27 D	0.998 D	0.826 D	0.898 D	0.909 D	0.909 D	0.89 D	77 D
G214R	0.001 -	1.00 D	4.2 H	-6.2 D	-6.13 D	0.965 D	0.99 7 D	0.989 D	0.949 D	26.7 D	0.999 D	0.934 D	0.924 D	0.98 D	0.904 D	0.9 D	94 D
R193W	0.071 -	1.00 D	1.755 M	-5.69 D	-2.45 D	0.805 D	0.06 8 D	0.733 D	0.806 D	27 D	0.998 D	0.240 D	0.566 D	0.909 D	0.834 D	0.63 D	55 D
F218L	0.001 D	0.99 0 D	2.915 M	-4.82 D	-3.25 D	0.766 D	0.89 8 D	0.861 D	0.807 D	24.7 D	0.998 D	0.904 D	0.769 D	0.531 D	0.899 D	0.84 D	76 D

*Примечание:* Замена – аминокислотная замена в молекуле; D – вредоносная замена, T – толерантная, N – нейтральная, M – средняя вероятность, L – низкая, H – высокая, P – патогенная, - – нет данных.

Результаты нашего исследования nsSNP гена *ELANE* показали, что 21 nsSNP были определены как вредоносные с помощью алгоритма SIFT. Из этих 21 SIFT-предсказанных вредоносных nsSNP 18 также были предсказаны как вредоносные алгоритмами PolyPhen-2 и FATHMM-MKL. Однако другие алгоритмы, использованные в этом исследовании, не показали 100% согласия (рисунки 16).

Среди всех 21 SIFT-предсказанных вредоносных nsSNP алгоритмы LRT и FATMANH предсказали наименьшее количество соответствий. Оба алгоритма предсказали только 10 патогенных nsSNP для *ELANE* и 11 были определены как толерантные, нейтральные или с неизвестным значением.

Платформа PolyPhen-2 определила 18 патогенных nsSNP; платформы VEST, CADD и DANN предсказали по 19; M-Cap и Mutation Taster предсказали по 20 патогенных nsSNP. При использовании подхода SNAP2 было обнаружено 18 повреждающих мутаций, в то время как три не имели связи с патологией.

Для программы PANTHER 17 nsSNP считались отклоняющимися мутациями, среди которых 10 nsSNP были классифицированы как вероятно патогенные, 7 - как возможно патогенные, 2 - как вероятно доброкачественные, и 2 варианта с неизвестным значением. При анализе с помощью PROVEAN, 14 из 21 nsSNP в гене *ELANE* были предсказаны как сильно разрушительные, а 7 были признаны нейтральными.

Mutation Assessor посчитал 20 nsSNP разрушительными, включая 3 с высокой патогенностью, со средней 6, с низкой 12 и для одна с неизвестным значением. P-Mut предсказал 10 мутаций как патологические, 10 – с неизвестным значением и одну без результата. PhD-SNP предсказал 13 мутаций как патологические, SNP-GO - 10, MetalR - 17, а MTA-SVM - 15.

Все современные методы оценки патогенности nsSNP, использованные вместе, выявили 8 перекрывающихся общих мутаций в гене *ELANE*: C71R, P139L, C151Y, T175M, G203S, G214R, R193W и F218L (таблица 9).

Согласно используемому программному обеспечению, известно, что частота аллеля C71R у латиноамериканцев составляет  $3,655e-05$ , T175M у африканцев -  $0,0002$ , у латиноамериканцев -  $0,0023$ , у жителей Восточной Азии -  $5,832e-05$ , у

европейцев -  $3,632e-05$ , у латиноамериканцев -  $2,979e-05$ . Результаты всех алгоритмов прогнозирования были статистически значимы и тесно связаны между собой. Значение коэффициента Стьюдента между инструментами имеет  $p$ -значение  $=0,001$ .

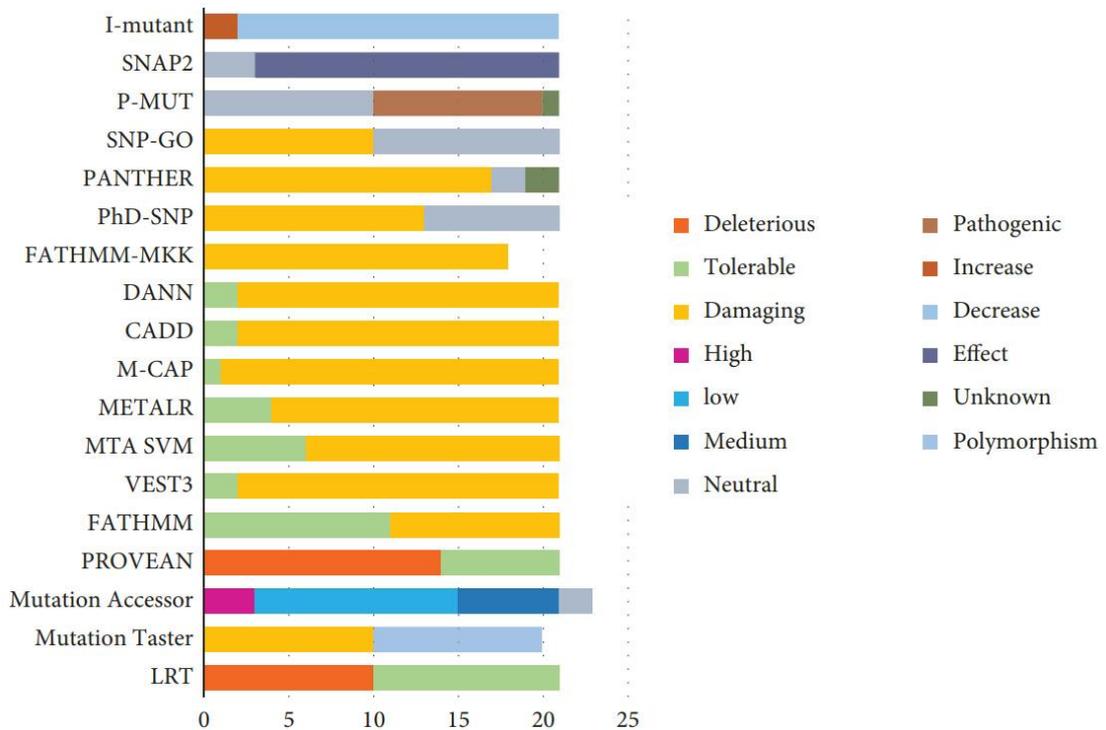


Рисунок 16 – Результаты предсказания влияния 21 nsSNP в гене *ELANE*, проанализированных восемнадцатью вычислительными инструментами

Таким образом, после данного этапа исследований было отобрано 15 наиболее вредоносных несинонимичных однонуклеотидных замен (и соответствующих аминокислотных замен), которые по предсказаниям использованных алгоритмов окажут наибольшее влияние на структуру и функцию белка TCIRG1: rs199902030, rs200149541, rs372499913, rs267605221, rs374941368, rs375717418, rs80008675, rs149792489, rs116675104, rs121908250, rs121908251, rs121908251, rs149792489, rs116675104, rs118141250. Одна из замен (rs118141250, Val52Leu) ранее была выявлена у пациента из Свердловской области по результатам полногеномного секвенирования.

Для гена *ELANE* в итоге было отобрано 8 nsSNP – rs28931611, rs57246956, rs137854448, rs193141883, rs201723157, rs201139487, rs137854451 и rs20038429.

Эти nsSNP (и соответственно аминокислотные замены) были в дальнейшем отправлены на анализ другими методами с целью установления их влияния на 3D-структуру белков и их функции.

#### **4.2 – Предсказание влияния nsSNP на устойчивость белка с помощью вычислительных инструментов I-Mutant, MU-pro**

Влияние nsSNP с высоким риском патогенности на стабильность и функцию белка TCIRG1 было оценено с помощью веб-сервиса I-Mutant 3.0. Результаты показали, что аминокислотные замены G405R, S474W и A778V повышают стабильность белка, в то время как P572L, M546V, I730N, F610S, A732T, F51S, A717D, E722K, R57H, R109W, R191W, S532C, G192S, F529L, H804Q, G458S, R444L, R56P, G379S, R757C, N730S, V375M, T314M, D517N, R92W, T368M, A417T, R363C, R56W и R50C снижают его расчётную стабильность (*таблица 10*).

Результаты влияния nsSNP с высоким риском патогенности на стабильность и функцию белка ELANE показали, что аминокислотные замены V101L и A166V повышают стабильность белка, тогда как R34W, C71R, V101M, P139L, R143C, C151Y, A166T, T175M, R182H, V190M, R193W, G203S, L206F, N209K, G210R, G214R, F218L, P262S и P262L R50C снижают его расчётную стабильность.

В то же время, алгоритм Mu-pro показал, что все выявленные по предыдущим тестам nsSNP с высокой оценкой патогенности снижают стабильность белков.

Таблица 10 – результаты анализа высоко вредоносных nsSNP в гене *TCIRG1* в программе-I-Mutant 3.0

Аминокислотная замена	Балл уверенности	Расчетная стабильность белка
P572L	-0.35	Снижение
M546V	-0.56	Снижение
I730N	-1.74	Снижение
F610S	-1.43	Снижение
A732T	-0.69	Снижение
F51S	-1.78	Снижение
A717D	-0.51	Снижение
E722K	-0.44	Снижение
R57H	-1.47	Снижение
R109W	-0.06	Снижение
R191W	-0.37	Снижение
S532C	-0.58	Снижение
G192S	-1.00	Снижение
F529L	-0.95	Снижение
H804Q	-0.10	Снижение
G405R	-0.28	Повышение
S474W	-0.10	Повышение
G458S	-1.26	Снижение
R444L	-0.23	Снижение
R56P	-0.85	Снижение
G379S	-1.41	Снижение
R757C	-1.00	Снижение
N730S	-0.34	Снижение
V375M	-1.06	Снижение
T314M	0.02	Снижение
D517N	-0.98	Снижение
R92W	-0.24	Снижение
T368M	-0.37	Снижение
A417T	-0.78	Снижение
R363C	-1.00	Снижение
R56W	-0.49	Снижение
A778V	-0.15	Повышение
R50C	-1.20	Снижение

### **4.3 – Анализ влияния nsSNP в генах *TCIRG1* и *ELANE* на консервативные участки белков.**

Согласно результатам исследования с использованием сервиса ConSurf, в высококонсервативных областях (7-9 баллов по шкале консервативности) белка *TCIRG1* обнаружены 22 вредоносных nsSNP. Ещё 16 замен – S7K, V52L, G379S, M403I, G405R, G458S, D517N, F529L, S532C, M546V, A640S, D683H, I732N, N730S, A732T, H804Q – были предсказаны как замены в функциональных и открытых аминокислотных остатках белка. Такие 10 замен, как A20V, R56P, R57H, R191H, G192C, E321K, R366H, T368M, R444L и E722K были предсказаны в регионе функциональных и открытых остатков, ещё 16 – S7K, V52L, G379S, M403I, G405R, G458S, D517N, F529L, S532C, M546V, A640S, D683H, I732N, N730S, A732T и H804Q – были предсказаны как заглубленные и структурные остатки. Следующие 18 замен - S3F, R28W, S45A, R50C, R92W, R109W, R166T, T314M, D328M, S340L, R363C, R382H, R467H, S474W, P572L, Y626S, R628W, и R757C были предсказаны как замены в открытых участках, а оставшиеся 9 - F51S, V348M, V375M, A417T, T570M, F610S, A717D, A778V и M783I – как замены в заглубленных аминокислотах. Результаты представлены на *рисунке 17*.

## ConSurf Results

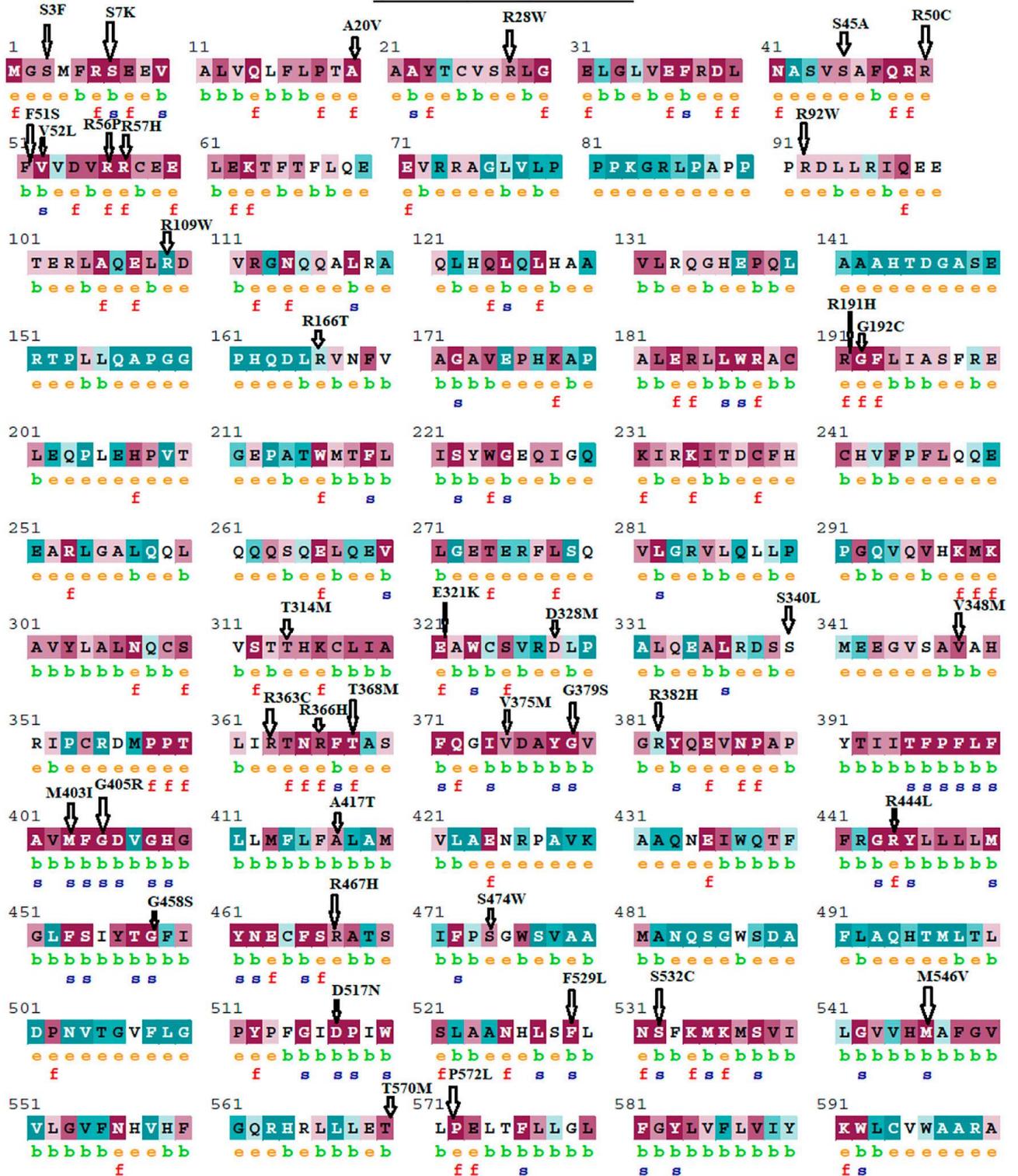


Рисунок 17. Расположение аминокислотных замен в белке TCIRG1 с учетом эволюционной сохранности и расположения различных участков белка по данным сервиса ConSurf

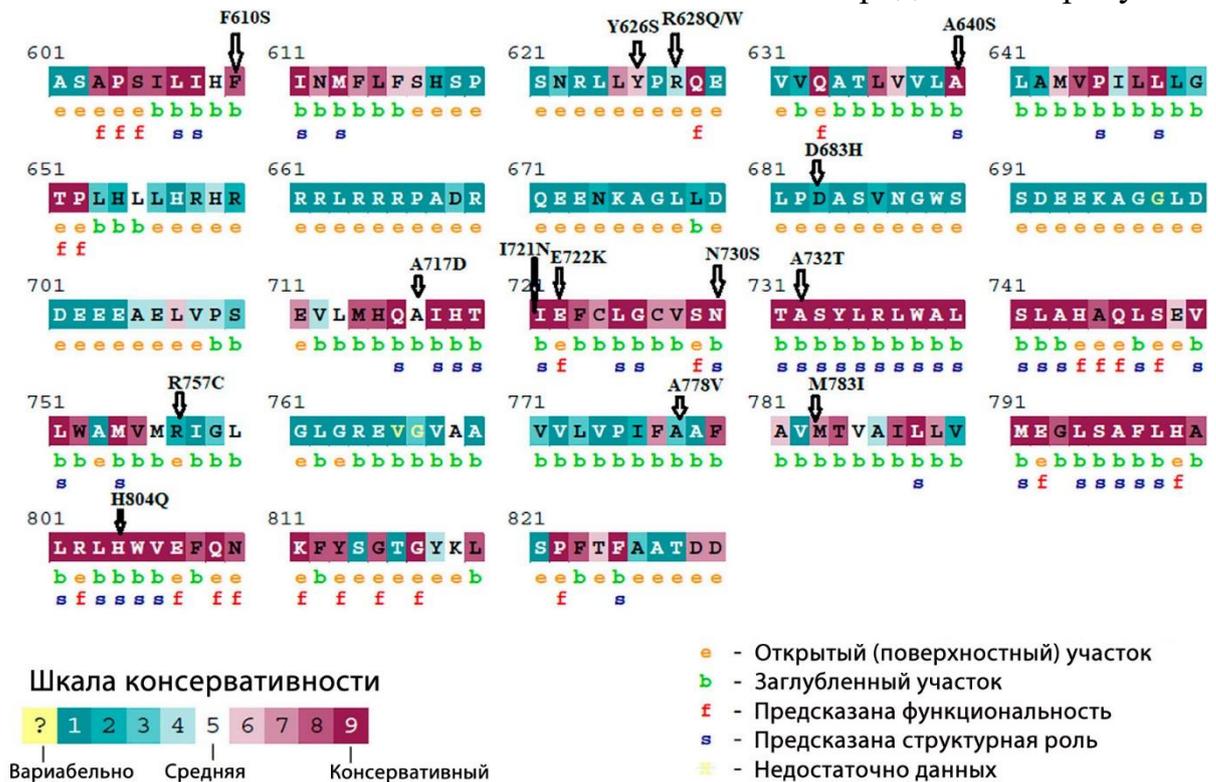


Рисунок 17 – Расположение аминокислотных замен в белке TCIRG1 с учетом эволюционной сохранности и расположения различных участков белка по данным сервиса ConSurf

*Примечание:* Значение 1 указывает на высокую изменчивость региона, 9 – на самый консервативный регион с точки зрения эволюции.

Для белка ELANE, согласно результатам анализа ConSurf, было определено 22 опасных nsSNP в высококонсервативных областях белка (7-9 баллов по шкале консервативности). Среди этих 22 миссенс-вариантов 8 были расположены в высококонсервативных позициях, 2 – P139L и C71R были предсказаны как функциональные и открытые остатки, а остальные 3 – G214R, C151Y и C71Y были предсказаны как заглубленные и структурные остатки. Следующие 12 замен – R34W, R143C, A166T, A166V, T175M, R182H, V190M, R193W, N209K, G210R, P262S и P262L были предсказаны как замены аминокислот на открытых участках белка, а F218L, V101L и V101M – как замены в заглубленных остатках.

#### **4.4 – 3D-моделирование структур белков TCIRG1 и ELANE с учетом выявленных потенциально вредоносных аминокислотных замен**

Предсказание структур белков с учетом отобранных после предыдущих этапов исследования аминокислотных замен, было проведено с помощью Phyre2, I-Tasser, HHPred и AlphaFold2.

Для белка TCIRG1 это было 15 аминокислотных замен, в том числе та, что выявлена у пациента в Свердловской области. Все эти мутации были включены в одну трехмерную структуру белка TCIRG1, поскольку все они располагались в разных участках. Таким образом при наложении и сравнении моделей получилось увидеть отличия в 3D-структуре для каждого из участков, где могла бы появиться замена.

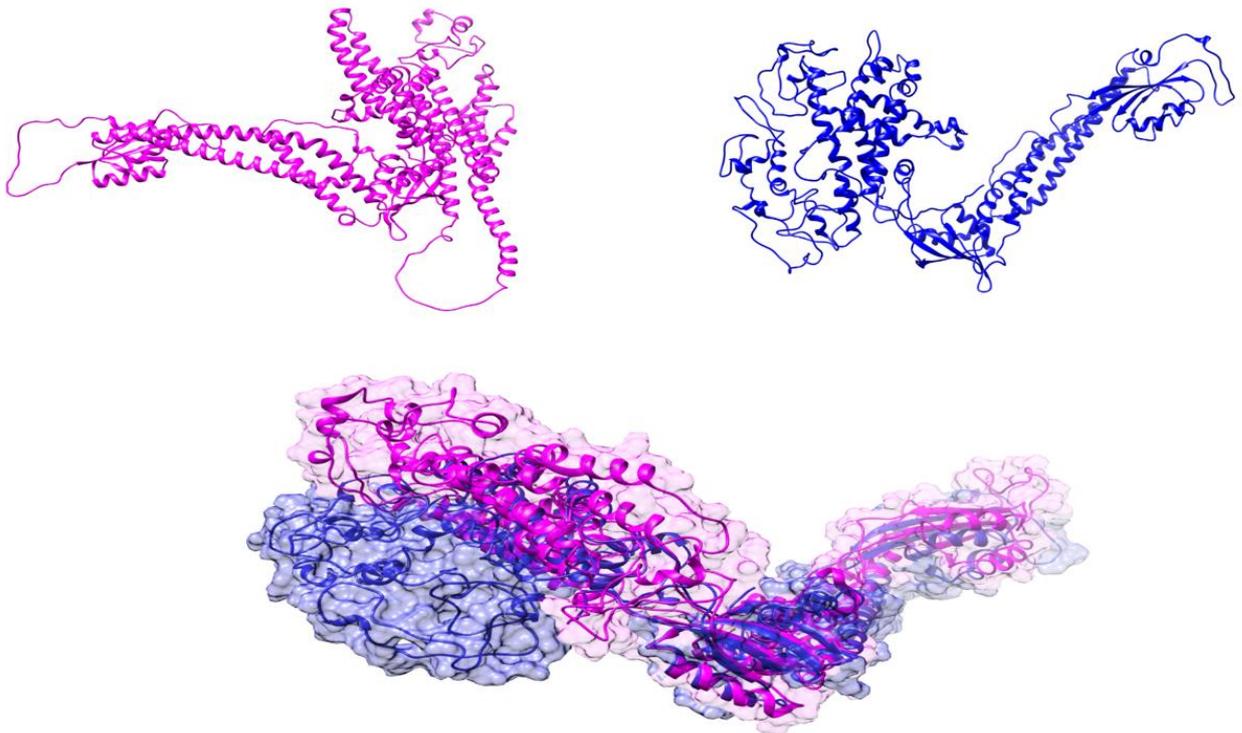
В качестве шаблона для предсказания влияния замен на TCIRG1 в Phyre2 был выбран шаблон модели C6VQ7A (шаблон с наибольшим сходством, согласно данным сервера Phyre2).

Phyre2 использовался для создания 3D-структур белка TCIRG1 с учётом его 56 мутаций. Замены nsSNP в последовательности белка TCIRG1 были смоделированы отдельно и затем переданы в Phyre2, который предсказал 3D-структуры мутантных белков. Однако, наши сравнительные исследования показали, что AlphaFold2 показал гораздо более качественные результаты для анализа TCIRG1, чем Phyre2, поэтому дальнейшие исследования MDS белка TCIRG1 мы проводили без использования Phyre2.

Пример 3D-структуры белка TCIRG1 в AlphaFold2 с включением отобранных в исследовании аминокислотных замен представлен на *рисунке 18*. Структура дикого типа была ранее предсказана AlphaFold2 и доступна в UniProt для загрузки (идентификатор Q13488).

При сравнении 3D-моделей дикого и мутантного типов белков были определены метрики сравнения моделей (TM-score) и среднеквадратическое отклонение (RMSD) расстояний между углеродными основаниями природных и мутантных моделей (при симуляции молекулярной динамики в течение 50 и 100 нс).

Низкие TM-score и высокие значения RMSD указывали на то, что структура мутанта отличается от структуры дикого типа. Соответствующий анализ 34 nsSNP, определенных как вредоносные для белка TCIRG1 при совместном анализе SIFT и PolyPhen2 представлен в *таблице 11*.



**Наложение 3D структуры дикого (сиреневый) и мутантного (синий) типов белка TCIRG1**

Рисунок 18 – 3D-структура дикого и мутантного типов белка TCIRG1, предсказанная AlphaFold2, и наложение трехмерных структур (внизу)

Мутант R92W (rs371907380) имеет самое высокое значение RMSD - 0,89Å, за ним следуют R444L (rs137853151), N730S (rs200209146) и S532C (rs371214361) с 0,84Å, 0,84Å и 0,81Å, соответственно. F610S, M546V и P572L имеют значения RMSD 0,81Å, 0,78Å и 0,78Å, соответственно, что указывает на отсутствие структурных отличий от дикого типа.

Таблица 11 – Значения TMscore и RMSD для 34 вредоносных nsSNP в TCIRG1

nsSNP	замена	TM-Score	RMSD
rs371907380	R92W		0,89
rs199902030	P572L	0.99626	0.78
rs200149541	M546V	0.99626	0.78
rs372499913	I721N	0.99760	0.53
rs267605221	F610S	0.99312	0.81
rs374941368	A732T	0.99621	0.78
rs375717418	F51S	0.99626	0.78
rs80008675	A717D	0.99661	0.73
rs149792489	E722K	0.99830	0.46
rs116675104	R57H	0.99790	0.48
rs121908250	R109W	0.99626	0.78
rs121908251	R191H	0.99785	0.49
rs121908251	S532C	0.99092	0.81
rs149792489	G192C	0.99626	0.78
rs116675104	F529L	0.99435	0.58
rs121908251	G405R	0.99674	0.62
rs116675104	G458S	0.99674	0.48
rs121908251	R56P	0.99657	0.48
rs121908252	R56W	0.99621	0.78
rs121908254	G379C	0.99435	0.58
rs147974432	R757C	0.99790	0.48
rs192224843	N730S	0.99275	0.84
rs115982879	V375M	0.99743	0.54
rs139059968	T314M	0.99626	0.78
rs141125426	D517N	0.99785	0.49
rs147208835	R92W	0.96213	0.89
rs147681552	T368M	0.99626	0.78
rs148498685	A417T	0.99790	0.48
rs149531418	R363C	0.99626	0.78
rs149531418	A778V	0.99661	0.76
rs147208835	R50C	0.99621	0.78
rs121908250	H804Q	0.99790	0.48
rs149792489	S474W	0.99760	0.53
rs121908250	R444L	0.99270	0.84

Другие nsSNP показали незначительные различия, которые включали I721N (0.53B RMSD), A732T (0.78B RMSD), R51C (0.78B RMSD), A717D (0.73B RMSD), E722K (0. 46B RMSD), R57H (0.48B RMSD), R109W (0.78B RMSD), R191H (0.49B RMSD), G192C (0.78B RMSD), F529L (0.58B RMSD), H804Q (0.48B RMSD), G405R

(0.48B RMSD) S474W (0.53B RMSD), G458S (0.48B RMSD), R56P (0.48B RMSD), R56W (0.78B RMSD), G379C (0.58B RMSD), R757C (0.48B RMSD), V375M (0.54B RMSD), T314M (0.78B RMSD), D517N (0.49B RMSD), T368M (0.78B RMSD), A417T (0.40B RMSD), R363C (0.78B RMSD), A778V (0.76B RMSD) и R50C (0.78B RMSD).

Четыре nsSNP с наибольшими значениями RMSD (R92W, R444L, N730S и S532C) были выбраны и переданы в I-Tasser для моделирования. Однако, сравнительный анализ результатов показал, что более качественные результаты моделирования белков позволили получить HHpred и AlphaFold2.

Поэтому ниже представлено сравнение 3D-моделей дикого и мутантного вариантов TCIRG1 в AlphaFold2 – до начала MDS (рисунок 19), при 50-наносекундной симуляции (рисунок 20) и при 100-наносекундной симуляции молекулярной динамики (рисунок 21).

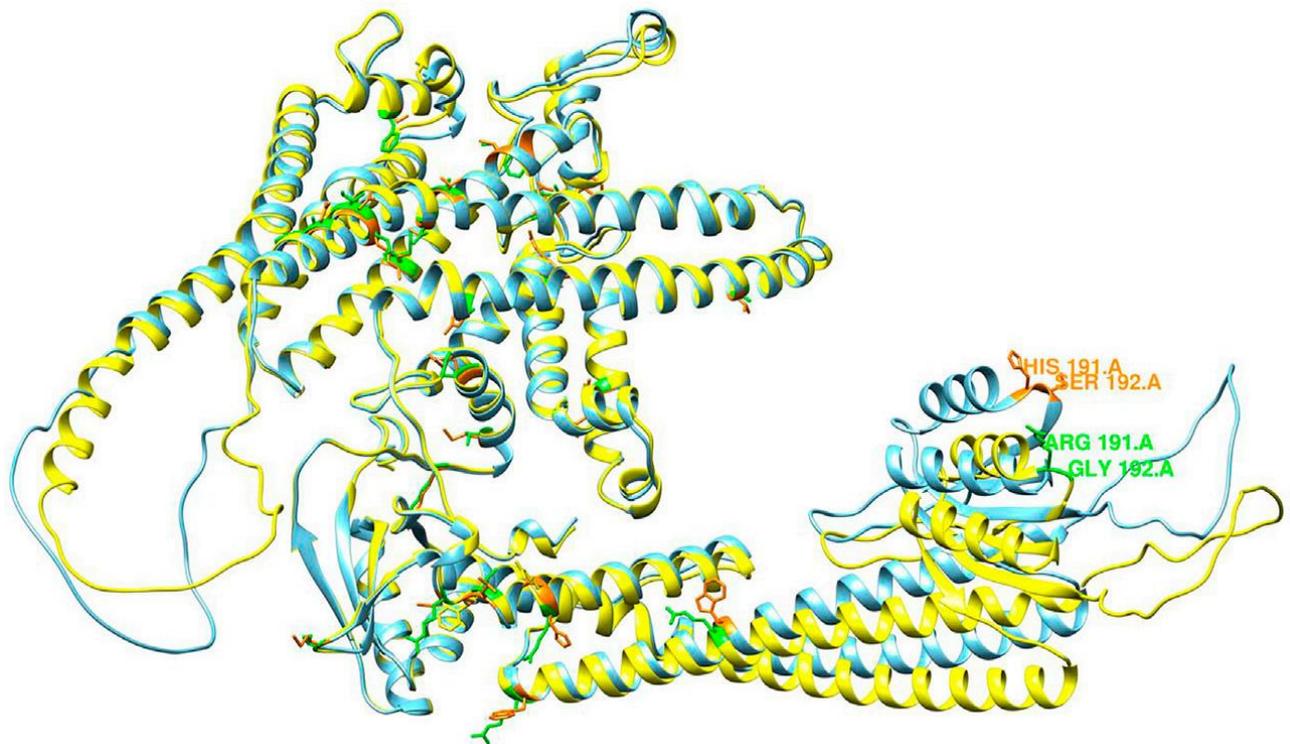


Рисунок 19 – Наложение 3D-структуры дикого (желтый) и мутантного типа (голубой) белка TCIRG1 до начала симуляции молекулярной динамики  
*Примечание:* Оранжевым обозначены участки внедренных в модель наиболее вредоносных замен, зеленым – соответствующие участки на модели дикого типа

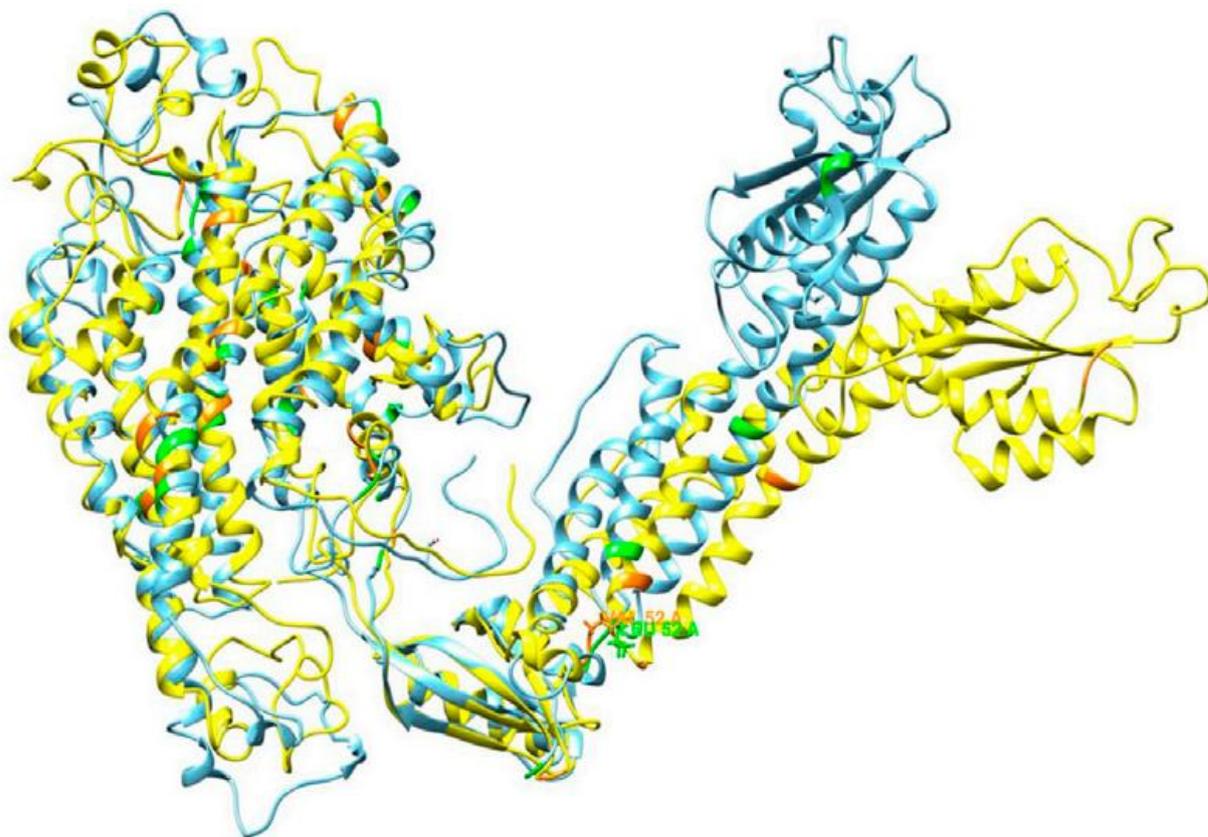


Рисунок 20 – Наложение 3D-структуры дикого (желтый) и мутантного типа (голубой) белка TCIRG1 при 50-наносекундной симуляции молекулярной динамики

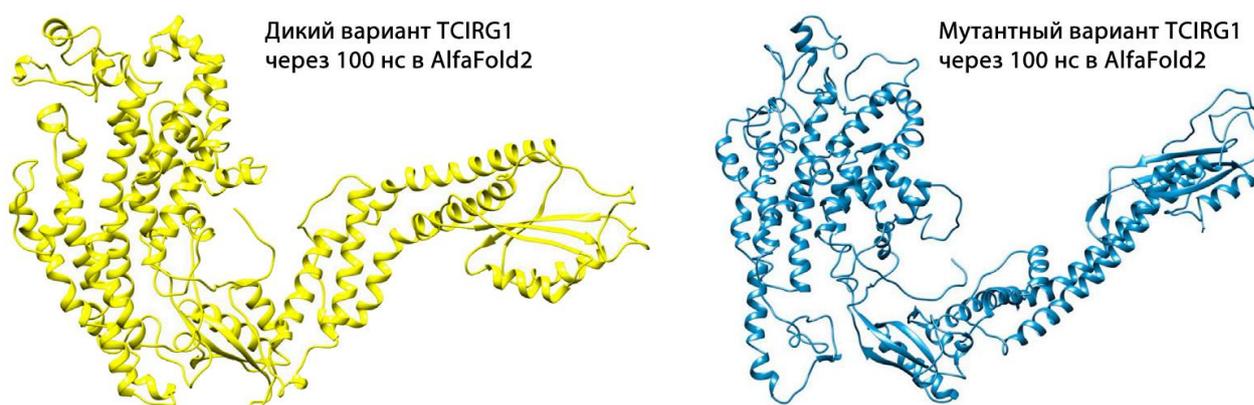


Рисунок 21 – 3D-структуры дикого (желтый) и мутантного типа (голубой) белка TCIRG1 при 100-наносекундной симуляции молекулярной динамики

В дальнейшем эти избранные мутантные типы TCIRG1 были проверены с помощью пакетов Schrodnger в симуляции молекулярной динамики.

Для моделирования 3D-структур белка ELANE дикого типа и мутантного типов был использован Phyre2. В качестве шаблона для предсказания 3D-модели ELANE в Phyre2 была выбрана модель sb01gA (*рисунок 22*). Предсказанные 3D-структуры мутантных белков представлены на *рисунке 23*.

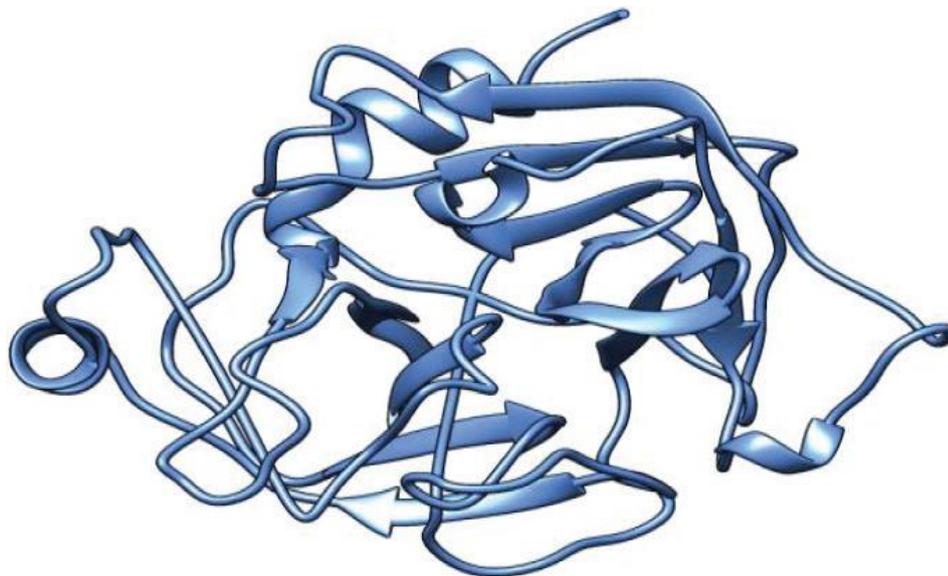


Рисунок 22 – Модель дикого типа белка ELANE по данным Phyre2

Для трехмерного моделирования белка ELANE была использована программа I-Tasser. Затем 3D-модели из I-Tasser были загружены в онлайн-сервис Zhanggroup, который выдал результаты в виде метрики сравнения моделей (TM-score) и среднеквадратичного отклонения (RMSD).

Мутантная модель C71Y (на основе nsSNP rs28931611) имела наибольшее отклонение от шаблона дикого типа ELANE со значением RMSD 2,05Å, за ней следуют R34W (rs201163886), F218L (rs200384291) и G214R (rs137854451) со значениями RMSD 1,98Å, 1,96Å и 1,12Å соответственно.

P139L, G203S и R193W имеют значения RMSD 0,04Å, 0,49Å и 0,96Å, соответственно, что указывает на отсутствие структурных отличий от дикого типа. В *таблице 12* приведены TM баллы и значения RMSD для мутантных типов ELANE.

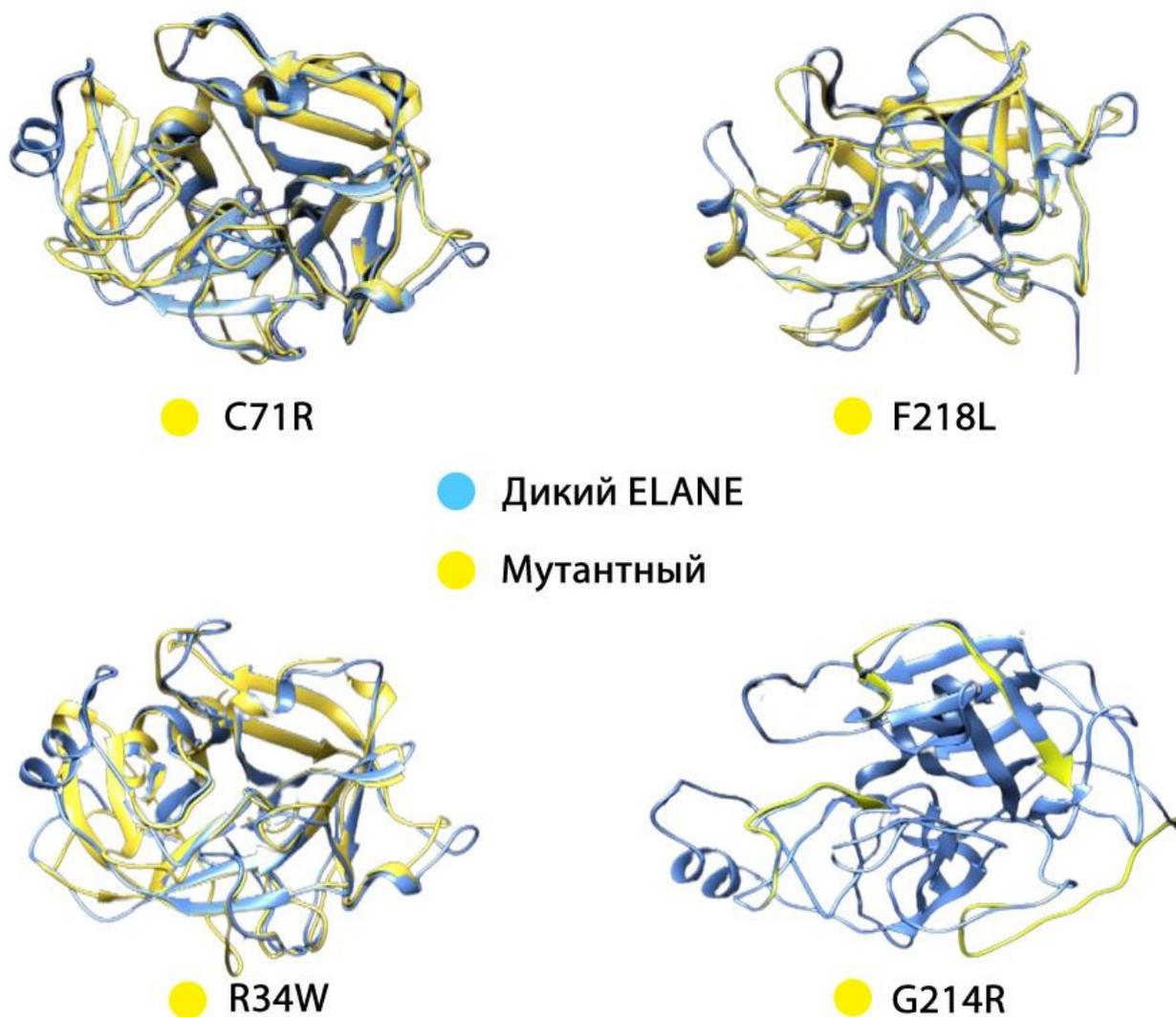


Рисунок 23 – Наложение дикого (голубой) и мутантного (желтый) типов белков ELANE с включением 4 наиболее значимых аминокислотных замен: C71R, F218L, R34W и G214R

Таблица 12 – TM-score и значения среднеквадратичного отклонения (RMSD) для выбранных nsSNP в белке ELANE

nsSNP	замена	TM-Score	RMSD
rs28931611	C71Y	0.85993	2.05
rs201163886	R34W	0.86482	1.98
rs200384291	F218L	0.87828	1.96
rs137854451	G214R	0.96114	1.12
rs201723157	R193W	0.95176	0.96
rs201139487	G203S	0.99524	0.49
rs137854448	P139L	0.99994	0.04

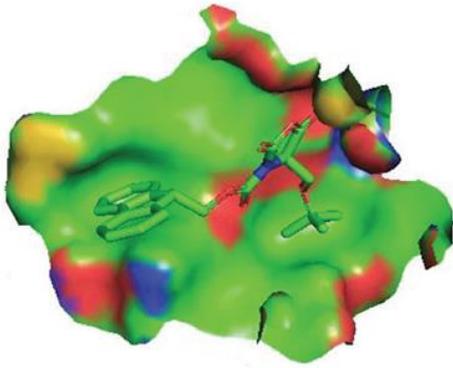
Четыре nsSNP с наибольшими значениями RMSD (C191Y, G214R, R34W и F218L) были выбраны и переданы в I-Tasser для ремоделирования и сравнения структур с диким типом (*рисунок 23*). Результаты проверки для дикого и мутантного вариантов 3D-моделей были удовлетворительными.

В дальнейшем эти избранные мутантные типы ELANE были использованы в эксперименте *in silico* при скрининге молекулярного докинга.

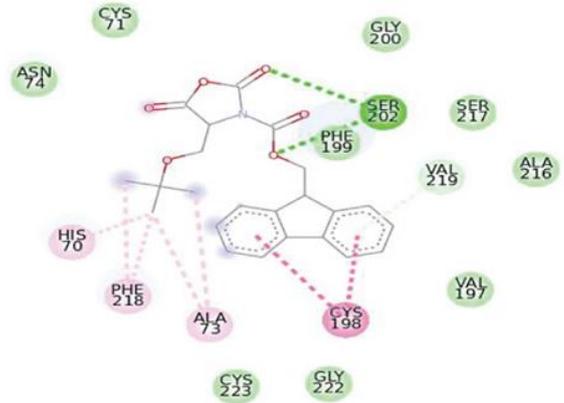
#### 4.5 - Оценка взаимодействия мутантных типов ELANE методом докинга

Докинг визуально оценивался с помощью Discovery Studio и PyMol. Взаимодействия участков докинга были рассчитаны для выявления сил связывания, которые были решающими в стабилизации образования комплексов рецептор-лиганд. Лиганд ANH был докирован в активный сайт белка дикого типа, а также четырех мутантных белков. Оценка докинга для дикого типа составила -8,4 КДж/моль, а 2D-взаимодействие показало, что дикий тип имеет две водородные связи с SER202, а также семь Ван-дер-Ваальсовых и семь гидрофобных контактов (*рисунок 24*). Оценка докинга для мутаций G214R, R34W, C71Y и F218L составила -9,2, -7,5, -7,1 и -6,8 КДж/моль, соответственно. 2D взаимодействие G214R показывало 2 водородные связи с SER202, 7 Ван-дер-Ваальсовых и 5 гидрофобных взаимодействий. Замена R34W показывала 2 водородные связи с SER202 и VAL219, 6 Ван-дер-Ваальсовых взаимодействий и 8 гидрофобных взаимодействий. Одна водородная связь присутствует в мутации C71Y с ARG81, 4 Ван-дер-Ваальсовых и 8 гидрофобных взаимодействий. На *рисунке 26* отражены связи для ELANE с заменой C71Y. Аналогично, F218L показывает одну водородную связь с ASN74, 6 Ван-дер-Ваальсовых и 4 гидрофобных взаимодействия. Две мутации, G214R и R34W, имеют взаимодействия, довольно близкие к дикому типу. Все они вовлечены в водородную связь с SER202. Две другие замены – C71Y и F218L – имели меньше водородных связей, что указывает на то, что эти две мутации могут влиять на стабильность и энергию белка (*рисунки 24-26*).

## 3D ELANE докинг



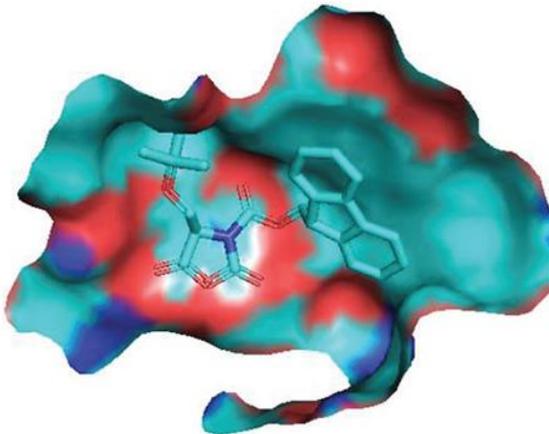
## 2D ELANE докинг



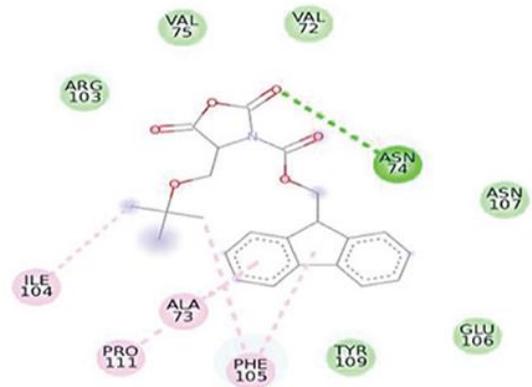
- |                             |                   |
|-----------------------------|-------------------|
| силы Ван-дер-Ваальса        | Амид-п-связь      |
| Обычная водородная связь    | Алкильная связь   |
| π-донорная водородная связь | π-алкильная связь |

Рисунок 24 – 2D- и 3D-графики поверхности ELANE дикого типа с лигандом внутри активного кармана

## 3D ELANE (C71Y) докинг



## 2D ELANE (C71Y) докинг



- |                          |
|--------------------------|
| силы Ван-дер-Ваальса     |
| Обычная водородная связь |
| Алкильная связь          |
| π-алкильная связь        |

Рисунок 25 – 2D- и 3D-графики поверхности ELANE с заменой C71Y с лигандом внутри активного кармана

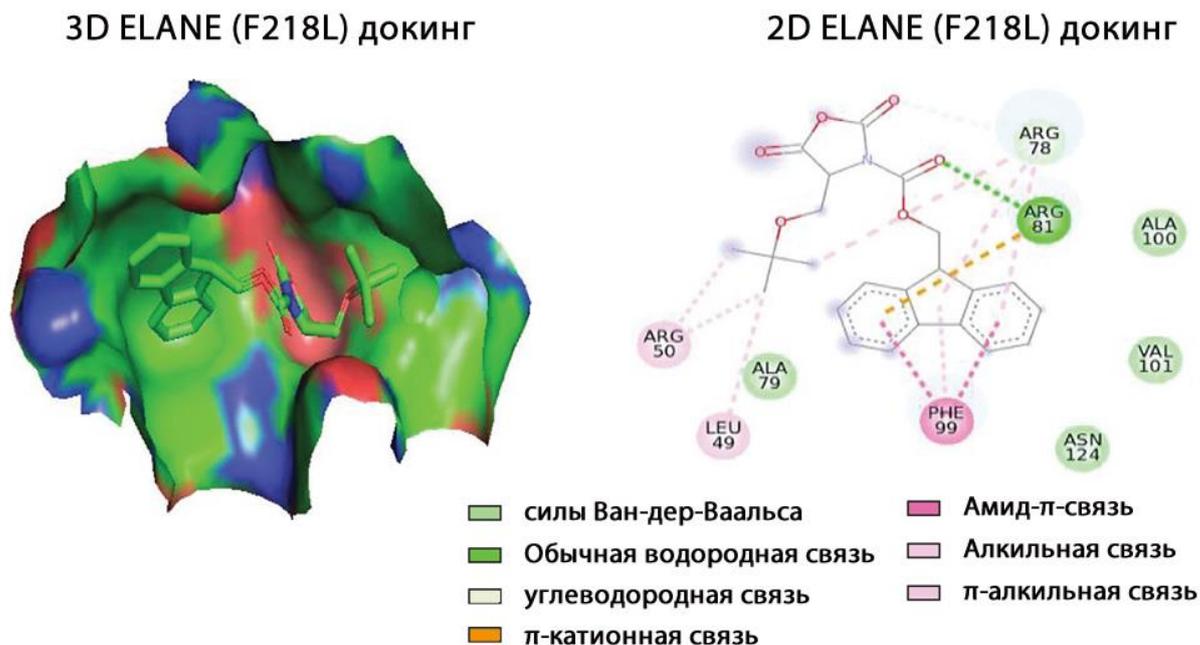


Рисунок 26 – 2D- и 3D-графики поверхности ELANE с заменой F218L с лигандом внутри активного кармана

#### 4.6 – Молекулярно-динамическое моделирование белка TCIRG1 дикого типа и его мутантов

При симуляции молекулярной динамики в программах HHPred (рисунок 27) и AlfaFold2 (рисунок 28) была сгенерирована эволюция значений среднеквадратичных отклонений (RMSD) альфа-атомов углерода (C $\alpha$ ) в молекуле белка с течением времени. График на *рисунке 28*, полученный в HHPred, показывал, что белок достигает стабильности при 20 000 пс. После этого, в течение всего времени моделирования, колебания значений RMSD для дикого типа оставались в пределах 2,0 Ангстрем, что является приемлемым [78]. Значения RMSD мутантного белка колебались в пределах 3,5 Ангстрем после того, как они были уравновешены. Эти результаты указывают на то, что мутантный белок имеет более высокий RMSD в течение всего периода моделирования. На графике RMSF пики представляют части белков, которые наиболее сильно колеблются во время моделирования (рисунок 29).

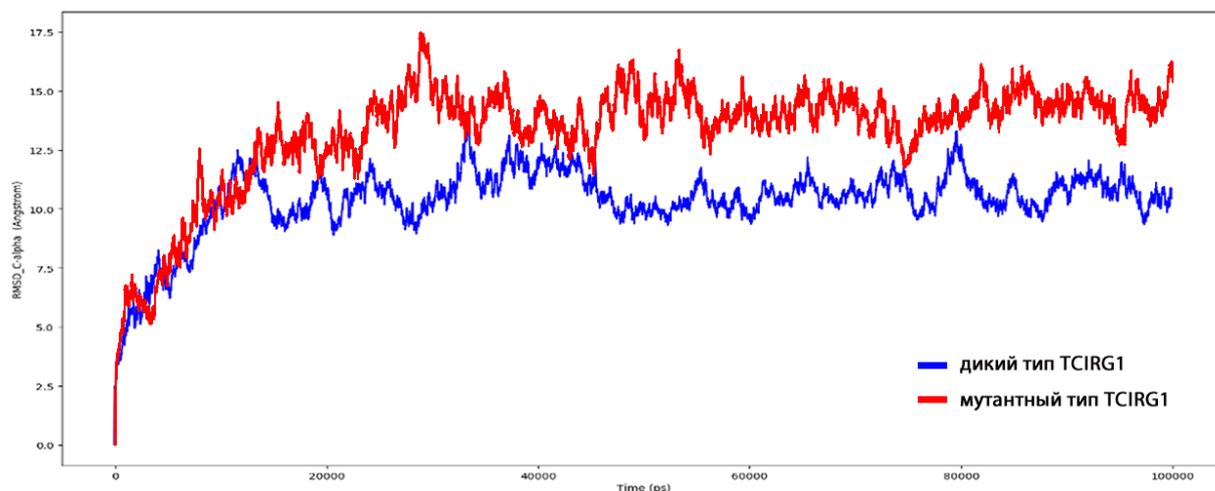


Рисунок 27 – Среднеквадратичное отклонение (RMSD)  $C\alpha$  атомов дикого типа и мутанта по данным HHpred с течением времени (100 нс)  
*Примечание:* По оси абсцисс – время в пс, по оси ординат – RMSD ( $\text{\AA}$ ).

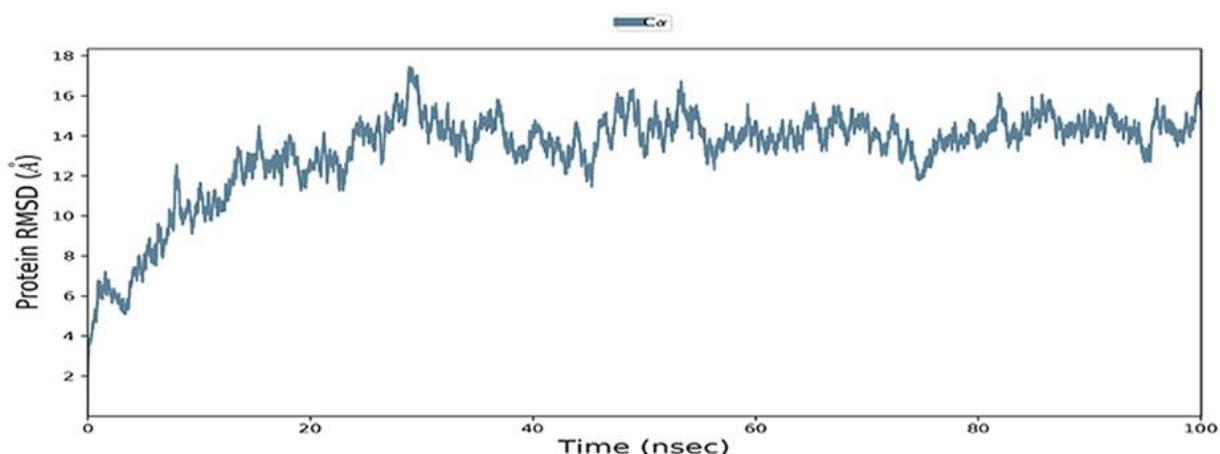


Рисунок 28 – Среднеквадратичное отклонение (RMSD)  $C\alpha$  атомов дикого мутантного типа TCIRG1 по данным AlfaFold2 с течением времени (100 нс)  
*Примечание:* По оси абсцисс – время в нс, по оси ординат – RMSD ( $\text{\AA}$ ).

На *рисунке 30* показаны общая энергия мутантного и дикого типов белка TCIRG1, на *рисунке 31* - энергия Ван-дер-Ваальсовых сил дикого типа и мутантного TCIRG1.

Хвосты белков (как N-, так и C-концевые) обычно изменяются сильнее, чем любая другая часть белка. Альфа-спирали и бета-слои, например, обычно жестче, чем неструктурированная часть белка, и колеблются меньше, чем части петель. Согласно рассчитанным MD-траекториям, остатки с большими пиками относятся к участкам петель или N- и C-концевым зонам. Альфа-спирали и бета-слои

отслеживаются как элементы вторичной структуры (SSE) во время моделирования. На *рисунке 32* показано распределение вторичных структур по индексу остатков для всех структур белка, а на *рисунке 33* - распределение элементов вторичной структуры в течение моделируемого времени 100нс. Все эти результаты свидетельствуют о том, что стабильность молекулы TCIRG1 мутантного типа снижена относительно дикого типа белка.

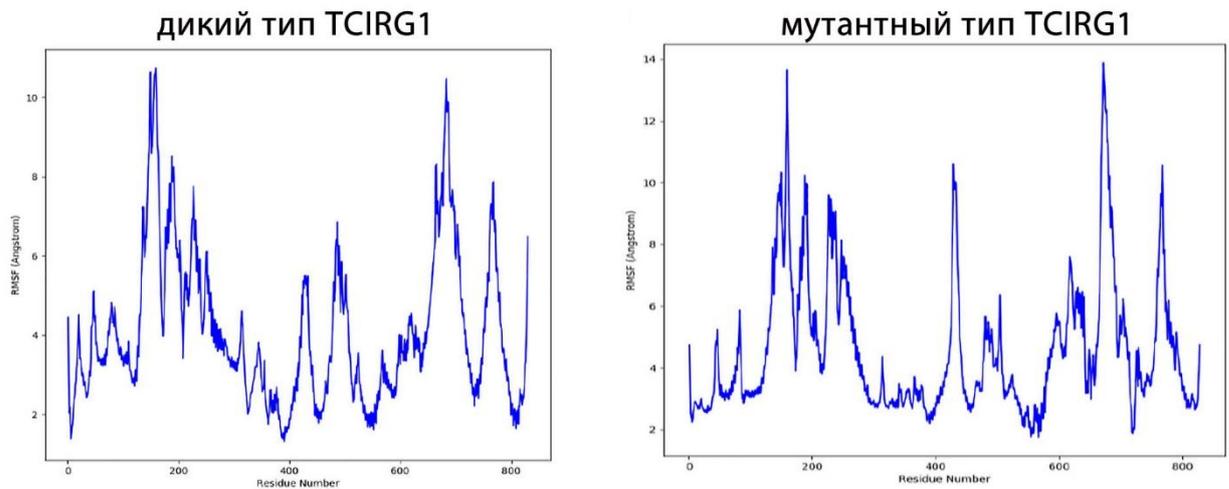


Рисунок 29 – Остаточное среднеквадратичное колебание (RMSF) белка TCIRG1 дикого типа (слева) и мутантного (справа)

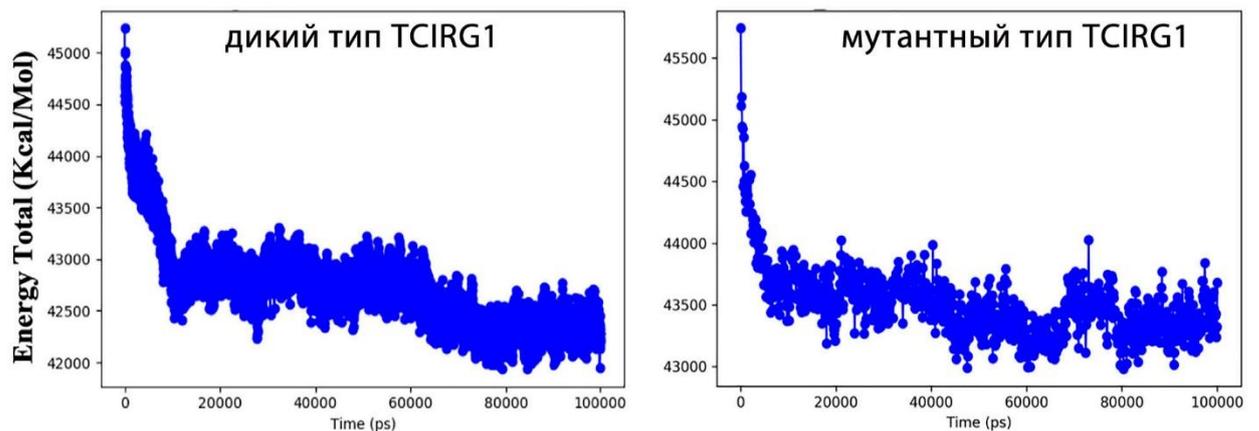


Рисунок 30 – Общая энергия белка дикого типа и мутантного TCIRG1

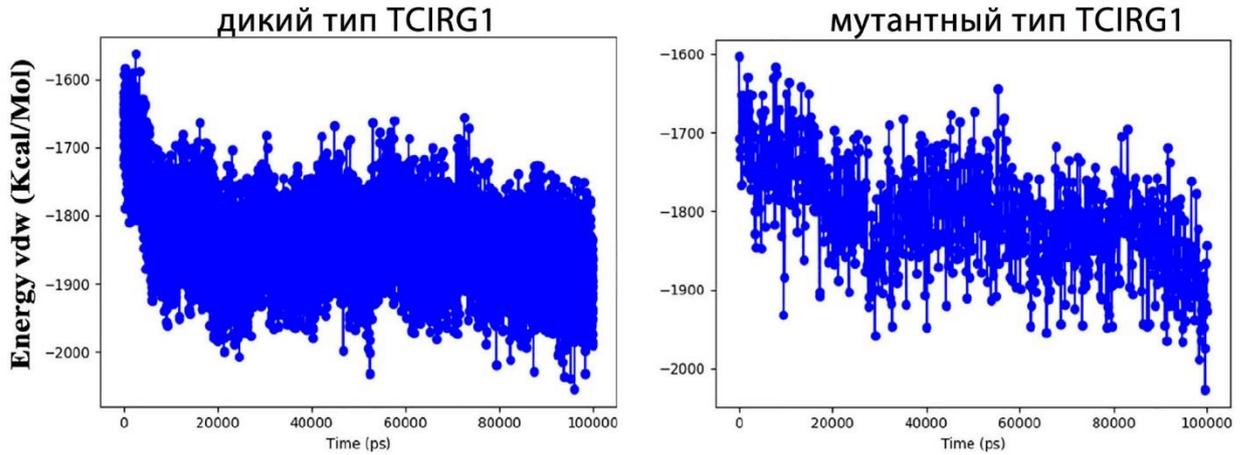


Рисунок 31 – Энергия Ван-дер-Ваальсовых сил дикого типа и мутантного TCIRG1

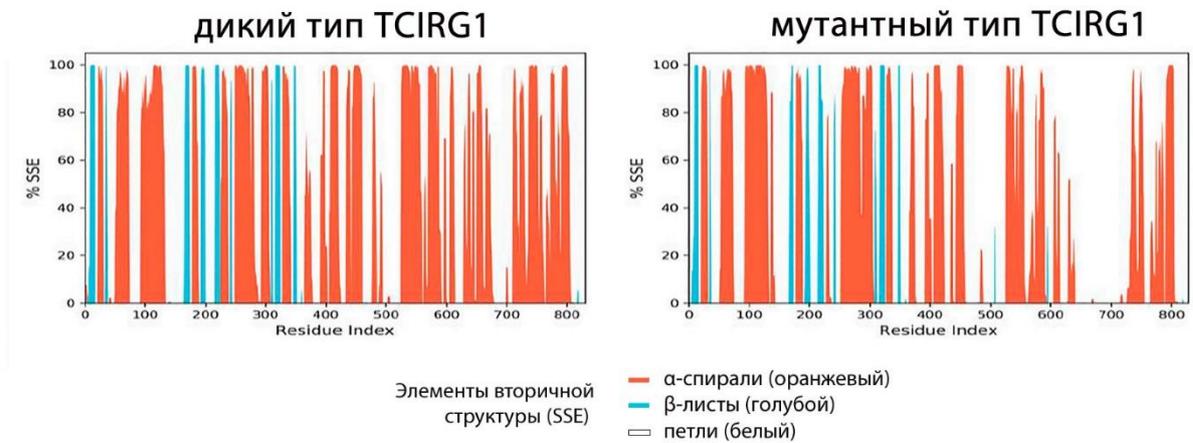


Рисунок 32 – Процент элементов вторичной структуры дикого типа и мутантного TCIRG1

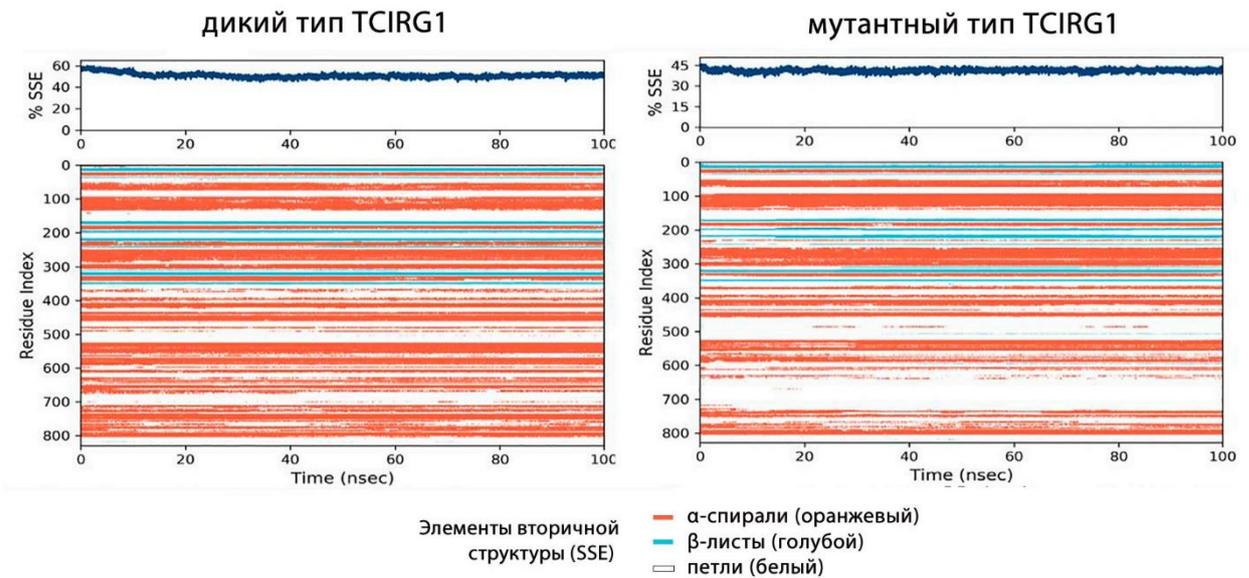


Рисунок 33 – Распределение элементов вторичной структуры в течение моделируемого времени 100нс

Большинство значимых внутримолекулярных взаимодействий, обнаруженных с использованием молекулярно-динамической симуляции, являются водородными связями (рисунк 34). Временная шкала изображает взаимодействия и контакты. Распределение атомов в белке вокруг его оси известно, как радиус гирации ( $R_g$ ). Скорость сворачивания белка напрямую связана с его компактностью, которую можно отследить с помощью передового компьютерного подхода для определения радиуса гирации (рисунк 35).

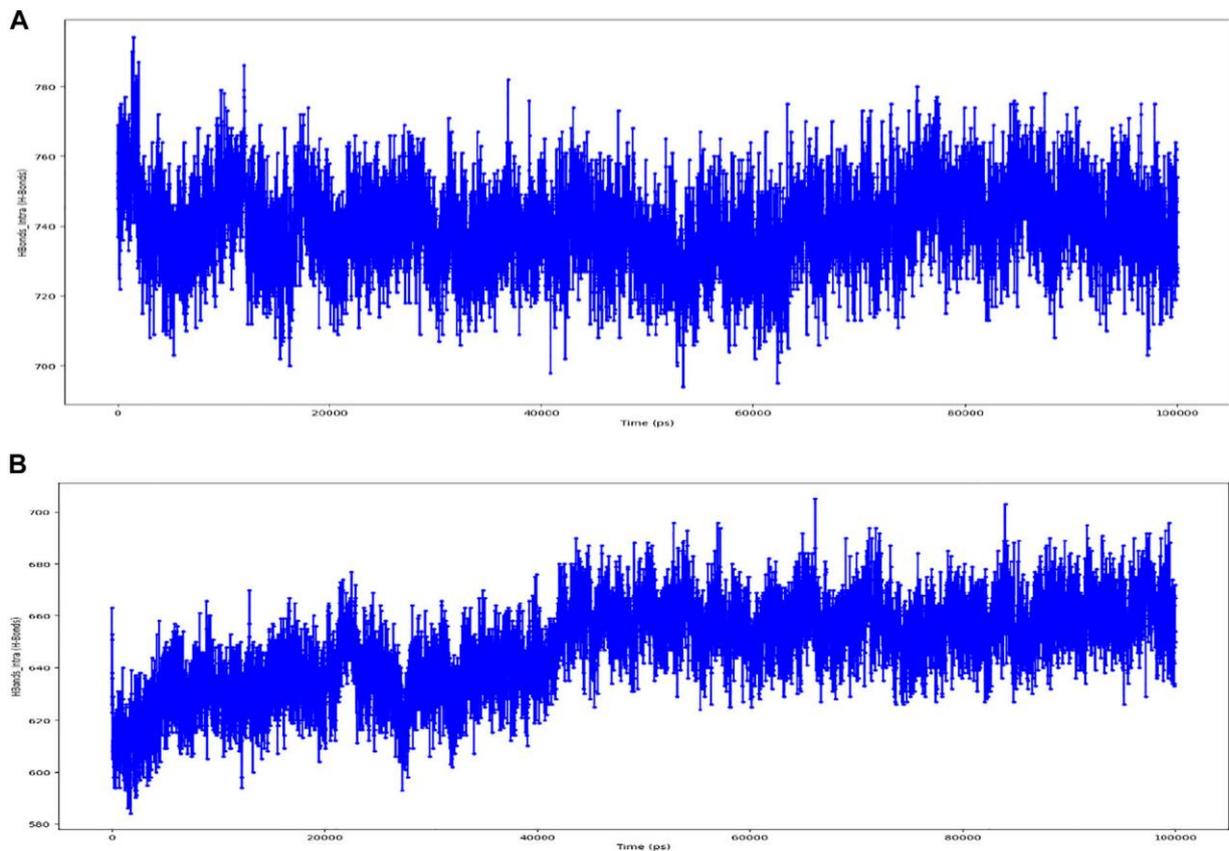


Рисунок 34 – Временная шкала представления взаимодействий и контактов водородных связей в белке TCIRG1 дикого (A) и мутантного (B) типов

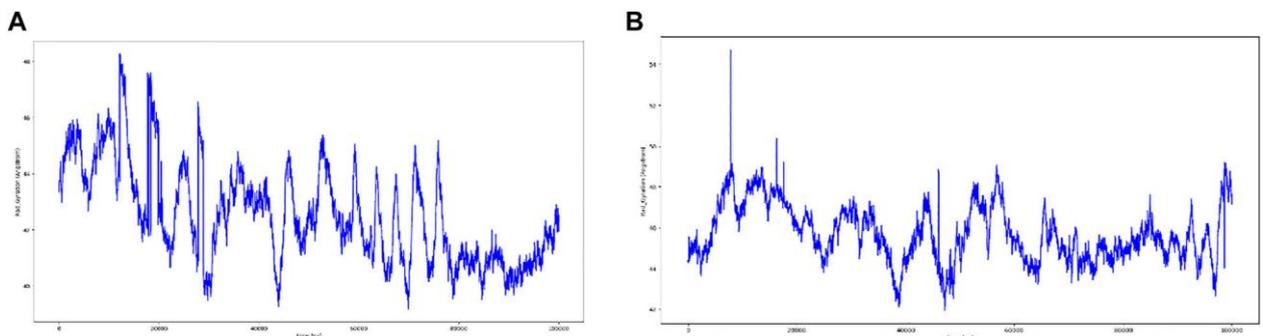


Рисунок 35 – Радиус гирации белка TCIRG1 дикого (A) и мутантного (B) типов. Примечание. По оси абсцис – время (нс), ординат – радиус гирации (Å)

Симуляция молекулярной динамики и соответствующая оценка влияния аминокислотных замен для белка ELANE нами не проводилась, был проведен только докинг. Фокус нашего внимания именно на TCIRG1 объясняется тем, что мы располагали результатами полногеномного секвенирования пациента с фенотипическими проявлениями врожденной нейтропении наряду с другими проявлениями, не характерными для классической врожденной нейтропении. Анализ данных полногеномного секвенирования позволил выявить несколько вероятно причинно-значимых мутаций, и только одна из них имела отношение к нейтропении. В частности, была выявлена несинонимичная однонуклеотидная замена g. 68041789G >C в гене TCIRG1 (аминокислотная замена V52L), однако её влияние на белок и фенотип было не известно. Поэтому было принято решение применить новые методы для оценки патогенности нового варианта гена *TCIRG1* и подтвердить нашу гипотезу.

Данная замена была включена в список замен, анализируемых выше, и это позволило более обоснованно предполагать, что выявленный у пациента вариант гена может иметь клиническое значение.

Таким образом, несинонимичные однонуклеотидные замены в генах *TCIRG1* (rs199902030, rs200149541, rs372499913, rs267605221, rs374941368, rs375717418, rs80008675, rs149792489, rs116675104, rs121908250, rs121908251, rs121908251, rs149792489 и rs116675104) и *ELANE* (rs200384291, rs201163886, rs193141883, rs201139487 и rs201723157) дестабилизируют белки TCIRG1 и ELANE.

#### **4.7 – Исследование кандидатных генов врожденной нейтропении**

Первым шагом для поиска или предсказания новых генов-кандидатов для врожденного заболевания является обзорный анализ опубликованной информации, анализ информации в базах данных генома и наследственных болезней, а также обзор генетических исследований, связанных с конкретным заболеванием.

Простой поиск информации в генетических базах данных, таких как OMIM (Online Mendelian Inheritance in Man) и HGMD (Human Gene Mutation Database),

помог выявить ранее зарегистрированные мутации, приводящие к болезням, а анализ публикаций в PubMed подсказал поиск направления дальнейших исследований.

Анализ белок-белковых взаимодействий (protein-protein interaction, PPI) известных генов врожденной нейтропении в геноме человека явился ключом к пониманию мультигенной природы врожденной нейтропении и дальнейшей идентификации генов-кандидатов.

При помощи базы данных STRING, была извлечена информация о белок-белковых взаимосвязях (PPI) для всех известных генов первичных иммунодефицитов (ПИД). Её визуальное представление с помощью программного обеспечения Cytoscape представлено на *рисунке 3б*, на котором гены врожденной нейтропении смещены к центру сети взаимосвязей. Это позволило предположить, что гены врожденной нейтропении взаимодействует чаще, чем случайные гены ПИД, что логично, поскольку, несмотря на различную генетическую причину врожденных нейтропений, фенотипы разных вариантов болезни схожи и при этом задействованы достаточно общие сигнальные пути, обеспечивая схожий патогенез. При врожденной нейтропении, в первую очередь, нарушены механизмы, связанные с регуляцией числа и функций нейтрофилов.

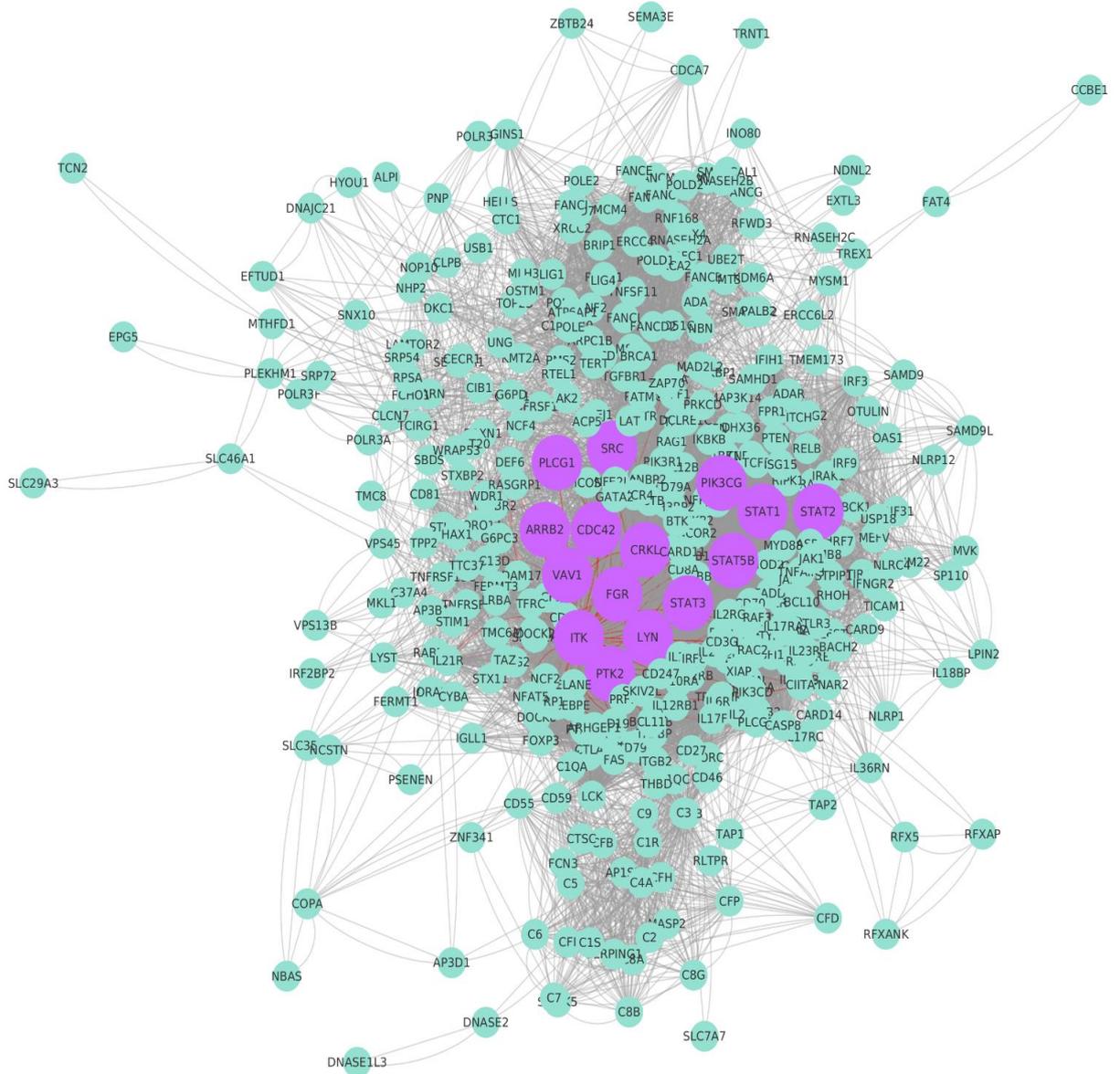


Рисунок 36 – Визуализированная сеть белок-белковых взаимодействий генов ПИД, извлеченных из базы данных STRING (результат анализа в Cytoscape)  
*Примечание:* Известные до момента исследования гены врожденной нейтропении представлены фиолетовыми узлами, а гены ПИД - зелеными.

На *рисунке 37* представлен график функциональных взаимоотношений между известными генами врожденной нейтропении.

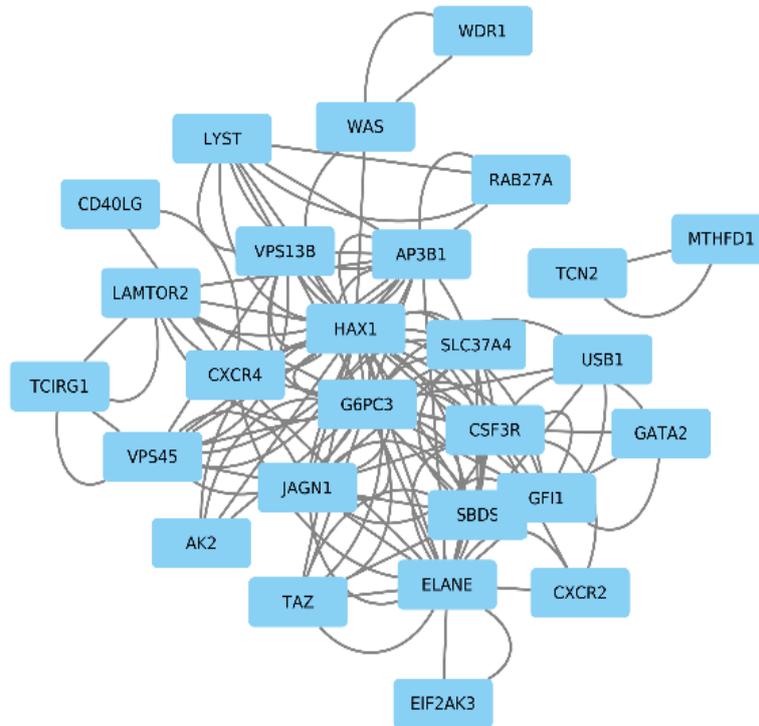


Рисунок 37 – Визуализированная сеть белок-белковых взаимодействий между известными генами врожденной нейтропении, извлеченных из базы данных STRING (результат анализа в Cytoscape)

Для дальнейшего исследования сложных взаимодействий генов врожденной нейтропении была оценена плотность сети группы из 31 гена врожденной нейтропении и десяти случайных групп ПИД, каждая из которых состояла из 41 гена. Сцепление и плотность сетей PPI в каждой группе были затем измерены и подвергнуты сравнительному анализу с помощью метода оценки плотности сети (D сети), и наши результаты показали более высокую плотность сети в группе врожденной нейтропении, чем в 10 случайных группах. Полученные результаты продемонстрировали сильное взаимодействие генов врожденной нейтропении друг с другом (рисунок 38).

Мы также проанализировали распределение биологической дистанции между группой известных генов врожденной нейтропении и двумя случайными группами, учитывая, что меньшее биологическое расстояние свидетельствует о более сильной ассоциации между генами в группе (рисунок 39).

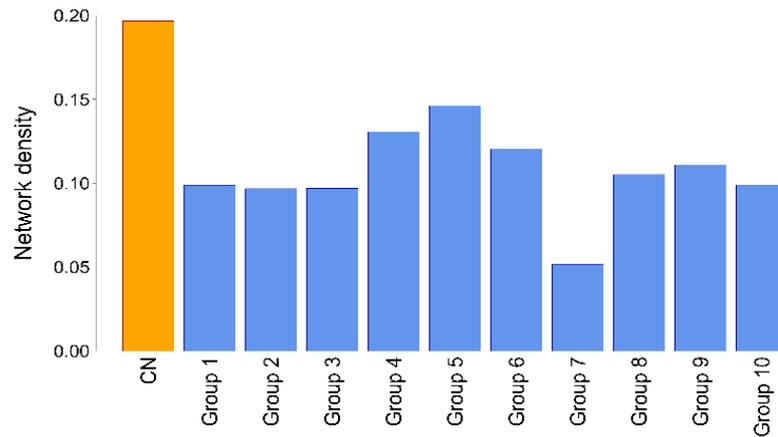


Рисунок 38 – Сравнение плотности сети группы известных генов врожденной нейтропении и десяти случайных групп ПИД

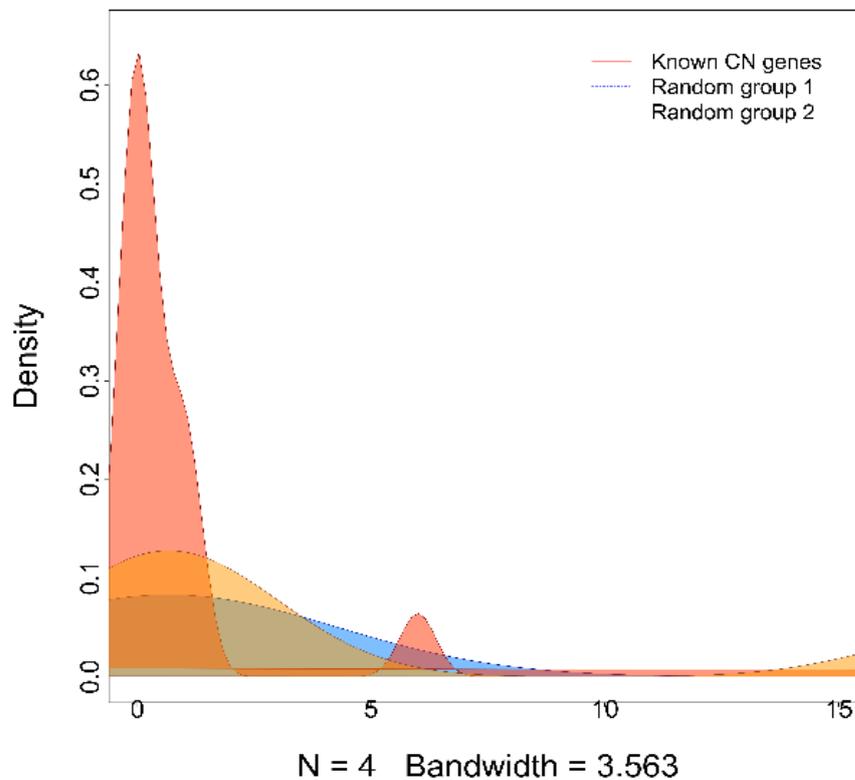


Рисунок 39 – Сравнение биологической дистанции в группе известных генов врожденной нейтропении и 2 случайных групп ПИД

Результаты показали, что медианная плотность распределения группы известных генов врожденной нейтропении составляет 1,067, в то время как медианная плотность распределения случайной 1-й и случайной 2-й групп составляет около 2,6, что указывает на более тесную функциональную взаимосвязь между известными генами врожденной нейтропении (рисунок 39).

Эти результаты позволяют сделать вывод о том, что гены CN взаимодействуют более тесно и имеют больше биологических взаимодействий друг с другом, чем гены ПИД.

На основе корреляционного анализа Пирсона (PCC) и белок-белковых взаимодействий, предоставленных Cheng F., et al. (2018) [203], было получено 4 613 специфических взаимодействий генов, функционально связанных с врожденной нейтропенией, и 177 генов-кандидатов.

Используя данные KEGG, мы провели анализ функционального обогащения известных генов врожденной нейтропении, фактически, связав гены из списка с их биологическими функциями. Наш анализ KEGG-путей показал 5 статистически значимых сигнальных путей ( $p < 0,05$ ), таких как взаимодействие цитокин-цитоклиновых рецепторов, хемокиновые сигнальные пути и прочие (рисунк 40).

Мы провели поиск конкретных генов-кандидатов, которые функционально схожи с известными генами врожденной нейтропении и обогащены по крайней мере одним из вышеупомянутых пяти путей KEGG.

Таким образом было получено 15 новых генов-кандидатов врожденной нейтропении: *STAT1*, *STAT2*, *STAT3*, *STAT5B*, *LYN*, *FGR*, *SRC*, *PIK3CG*, *ITK*, *VAV1*, *CDC42*, *PTK2*, *CRKL*, *PLCG1*, *ARRB2*.

На рисунке 41 показана сеть PPI известных генов и генов-кандидатов врожденной нейтропении.

Анализ функционального обогащения генов врожденной нейтропении, включая 15 генов-кандидатов, показал, что в общей сложности задействовано 15 статистически значимых сигнальных путей, описанных в базе данных KEGG (например, инфекция вируса Эпштейна-Барр, взаимодействие цитокин-цитоклиновый рецептор и сигнальный путь рецептора В-клеток).

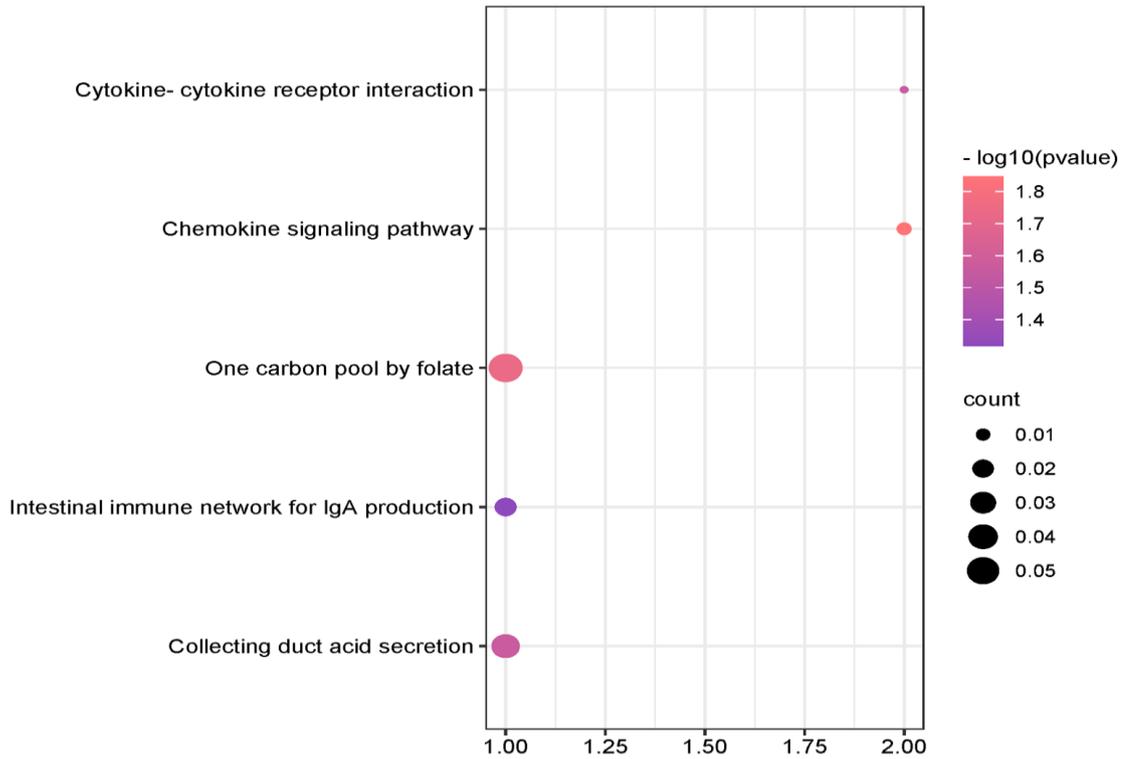


Рисунок 40 – Анализ функционального обогащения генов-кандидатов врожденной нейтропении на основе базы данных KEGG

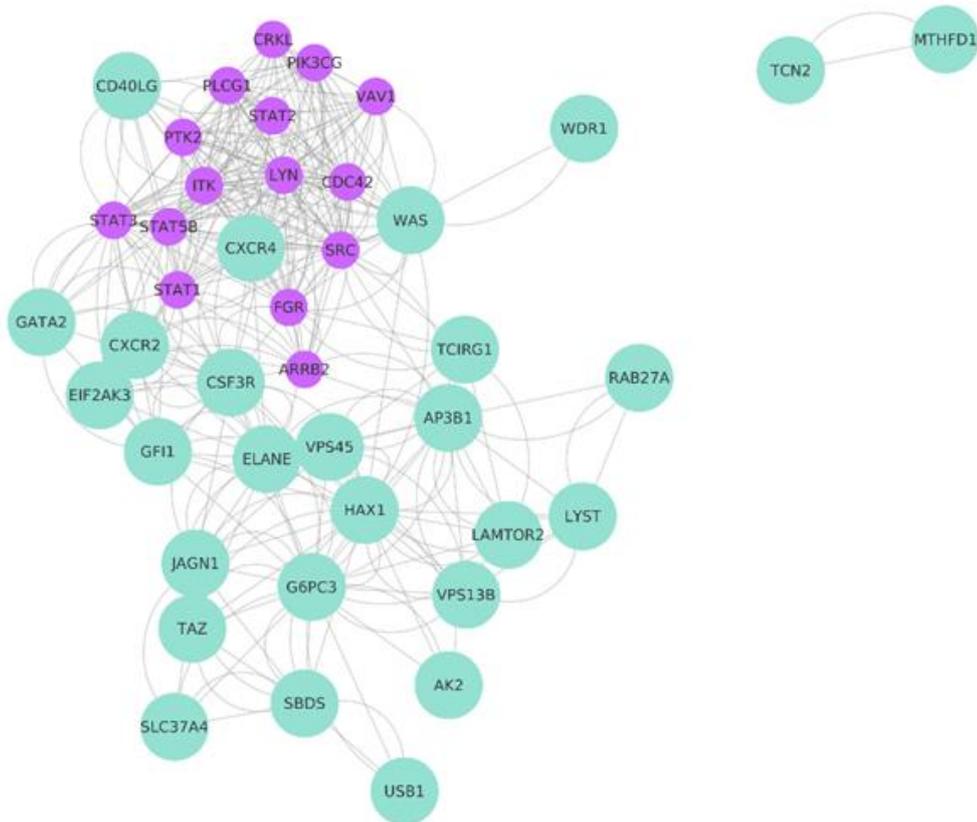


Рисунок 41 – Сеть белок-белковых взаимодействий известных генов и генов-кандидатов врожденной нейтропении (Cytoscape)

Биологические расстояния 15 генов-кандидатов врожденной нейтропении были сравнены с биологическими расстояниями 31 известного гена врожденной нейтропении. В результате, среднее биологическое расстояние генов-кандидатов составило 6,08, что было меньше (или аналогично), чем у известных генов врожденной нейтропении. Это свидетельствует о том, что гены-кандидаты врожденной нейтропении имеют сопоставимые сильные биологические связи (рисунок 42).

Затем ген-кандидат врожденной нейтропении был смешан с известными генами врожденной нейтропении, и биологическое расстояние смешанного гена было определено заново.

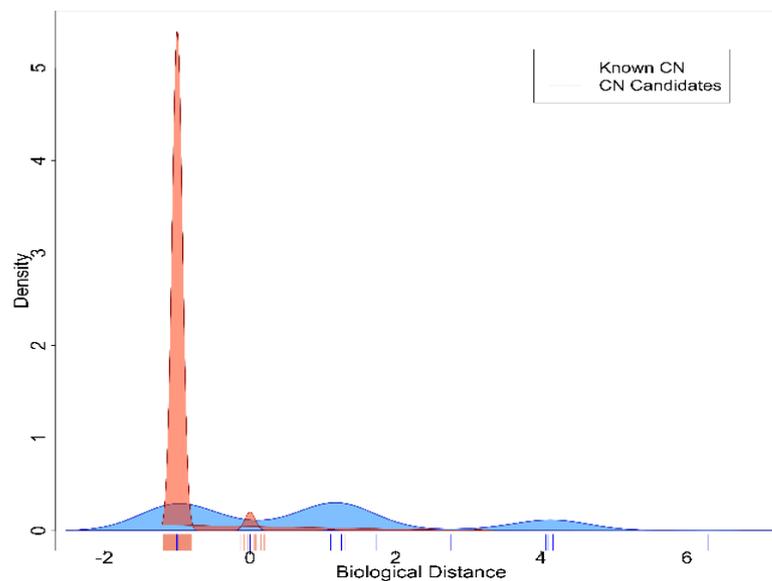


Рисунок 42 – График плотности биологических расстояний известных генов врожденной нейтропении и предсказанных генов-кандидатов

Затем смешанные гены были подвергнуты филогенетическому анализу FGA для определения биологического родства между геном врожденной нейтропении и геном-кандидатом врожденной нейтропении. Результаты показали, что гены-кандидаты врожденной нейтропении равномерно распределены по всему диапазону известных генов, что подразумевает их тесную связь с известными генами врожденной нейтропении (рисунок 43).

Также благодаря анализу функционального обогащения выявленных генов-кандидатов с помощью KEGG была сформирована диаграмма взаимосвязей генов-

кандидатов врожденной нейтропении и связанных с ними сигнальных путей (рисунк 44).

Оценка разницы экспрессии генов в нейтрофилах периферической крови пациентов позволила подойти к поиску новых кандидатных генов с другой стороны и предварительные результаты были подтверждены.

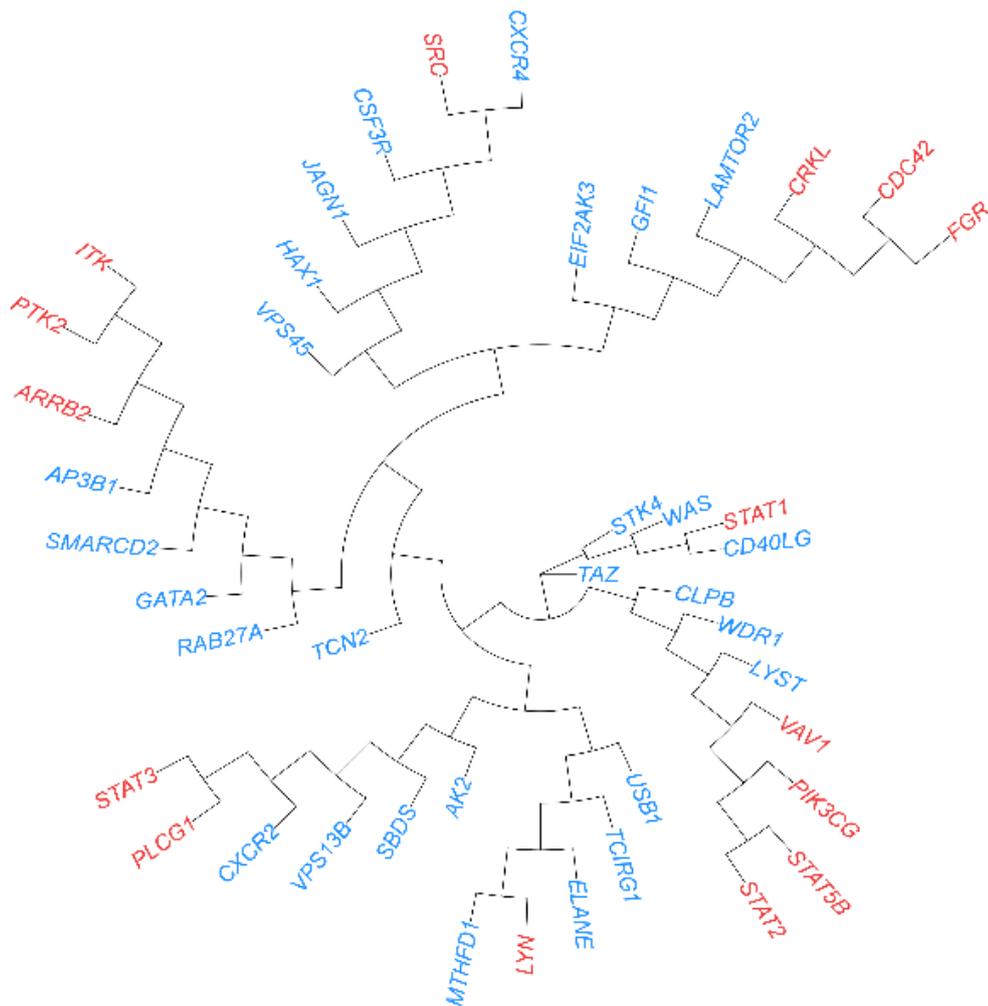


Рисунок 43 – Филогенетическое дерево биологических расстояний, сгенерированное FGA, показывающее иерархическую кластеризацию всех известных генов врожденной нейтропении (синий) и предсказанных генов врожденной нейтропении (красный).

*Примечание:* Длина ветви указывает на силу разделения между субъектами.

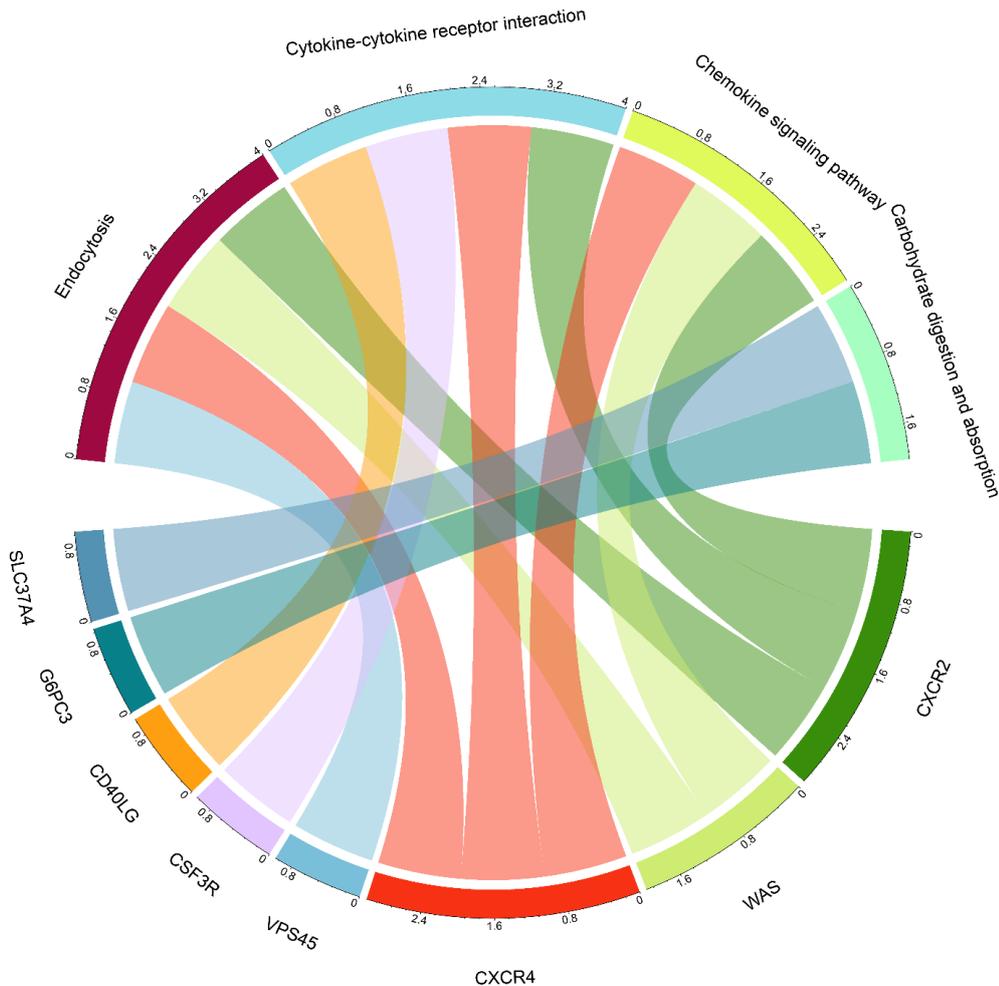


Рисунок 44 – Хордовая диаграмма генов-кандидатов врожденной нейтропении и связанных с ними сигнальных путей (по данным KEGG)

В наборе данных GSE142347 в нейтрофилах периферической крови пациентов с врожденной нейтропенией по сравнению с уровнем экспрессии у здоровых (контрольные образцы) экспрессия 1327 генов значительно отличалась – у 739 генов она была повышена, а у 558 генов – понижена. В наборе данных GSE6233 в В-клетках пациентов с врожденной нейтропенией по сравнению с уровнем в контрольных образцах было обнаружено 573 генов с достоверной разницей в экспрессии. Из этих 573 генов 274 гена были с повышенной, а 299 генов с пониженной экспрессией (рисунок 45).

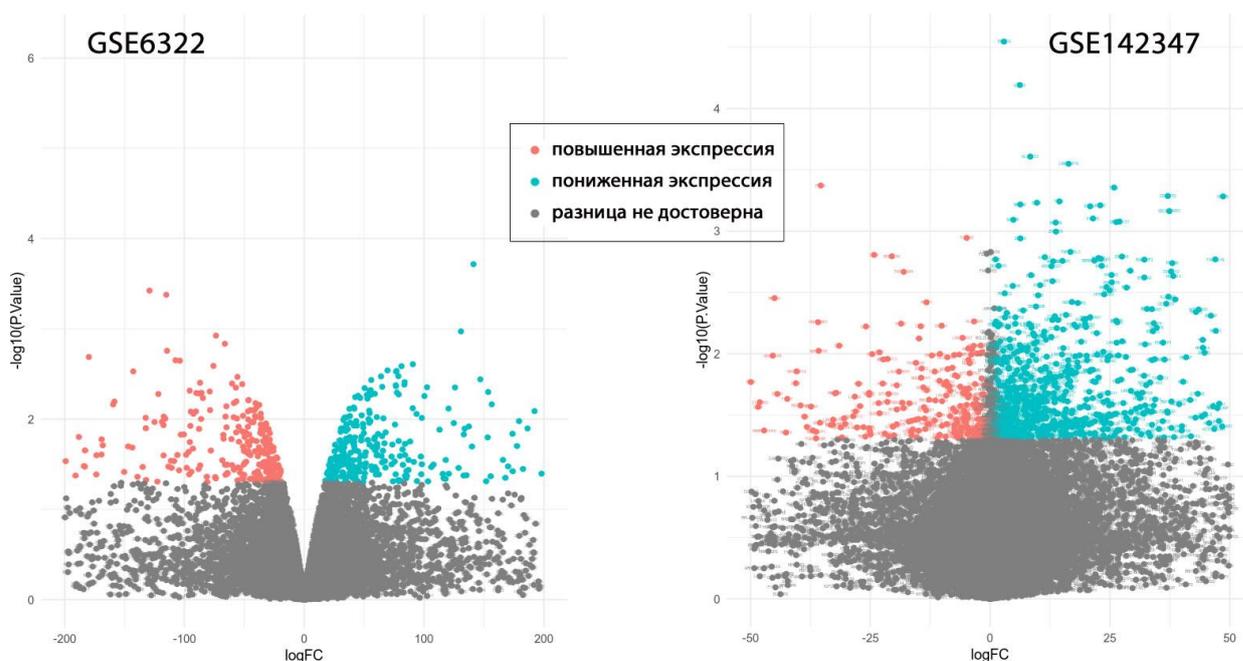


Рисунок 45 – Вулкано-плот дифференциально экспрессированных генов в наборах данных GSE6322 (слева) и GSE142347 (справа).

Кроме того, при сравнении генов с повышенной экспрессией в нейтрофилах периферической крови и В-клетках пациентов с врожденной нейтропенией был выявлен один общий ген, а при сравнении генов с пониженной экспрессией в нейтрофилах периферической крови и В-клетках пациентов с врожденной нейтропенией было выявлено 7 общих генов со сниженной экспрессией относительно контрольных образцов (рисунки 46). Фактически это означало выявление общих свойств транскриптома нейтрофилов и В-клеток пациентов с врожденной нейтропенией.



Рисунок 46 – Диаграммы Венна перекрывающихся высоко и низко экспрессируемых генов при врожденной нейтропении в разных датасетах

Некоторые известные гены ПИД также значительно отличались по экспрессии. В наборе данных GSE6233 10 генов имели повышенную и 7 - пониженную экспрессию. В наборе данных GSE142347 3 гена с повышенной, а 18 – с пониженной экспрессией. Генами с повышенной экспрессией в GSE6233 были следующие: *LAMTOR2*, *SmarCD2*, *CD81*, *ZBTB24*, *ACTB*, *CASP10*, *APOL1*, *PARN*, *ITGB2* и *IRF3*. Генами с повышенной экспрессией в датасете GSE142347 являлись *SEC61A1*, *MASP2* и *RAD51*. Среди генов с пониженной экспрессией в датасете GSE6233 - *SEC61A1*, *MTHFD1*, *STIM1*, *EXTL3*, *TGFBR1*, *CEBPE* и *HAX1*. Гены с пониженной экспрессией в датасете GSE142347: *PTPRC*, *RAC2*, *BRCA1*, *PRF1*, *FCGR3A*, *ACTB*, *COPA*, *IL2RG*, *MSN*, *IKZF1*, *KDM6A*, *CD55*, *AP1S3*, *NFKB1*, *WDR1*, *JAK1*, *IFIH1*, *RAD51C*. Все известные и кандидатные гены врожденной нейтропении, за исключением *CXCR4*, *LAMTOR2*, *STAT1* и *STAT2*, были более высоко экспрессируемы у пациентов с врожденной нейтропенией, чем в контрольных образцах (рисунок 47).

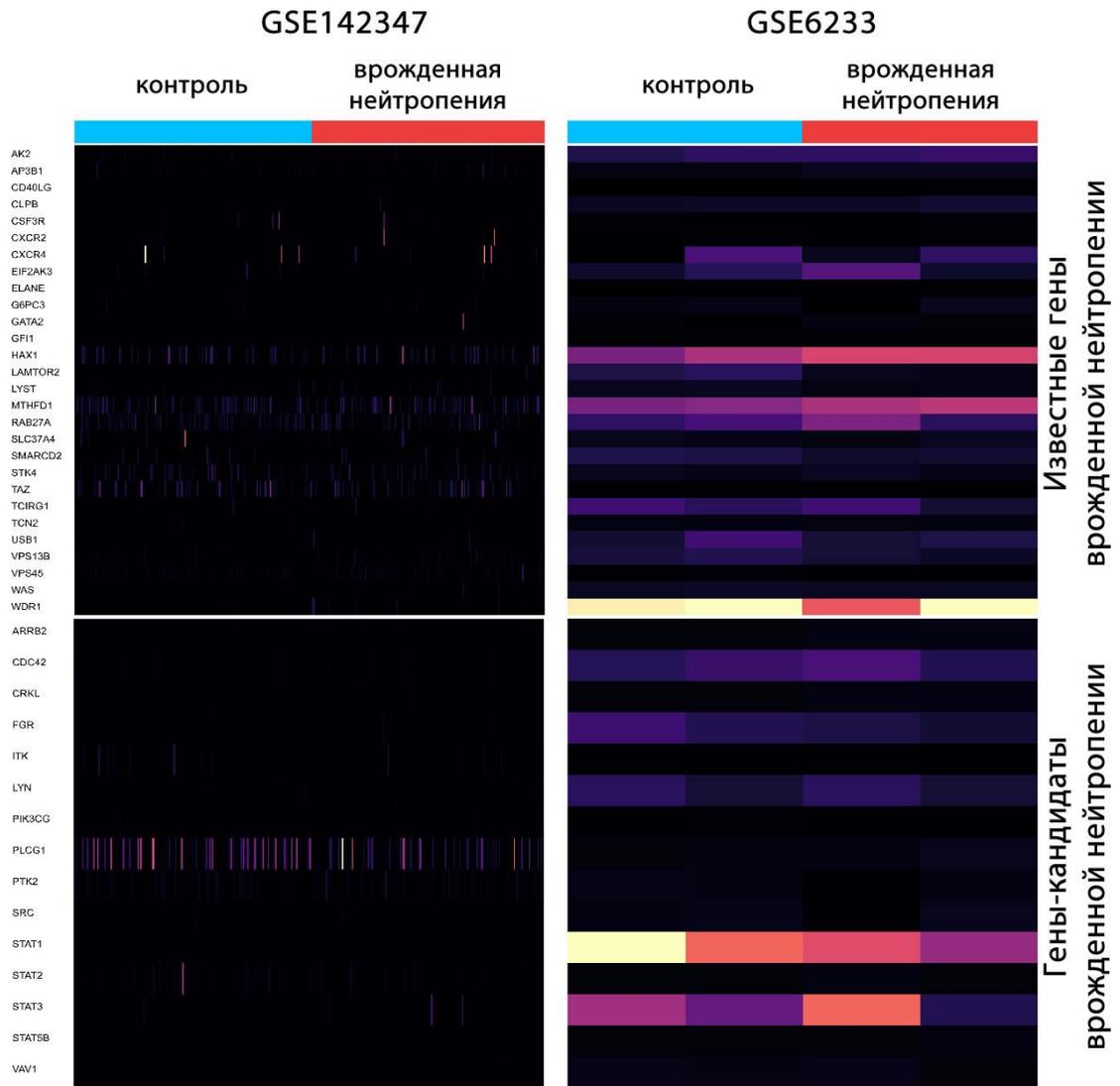


Рисунок 47 – Тепловая карта дифференциально экспрессированных известных и кандидатных генов врожденной нейтропении.

*Примечание:* Высокая экспрессия показана светло-желтым цветом, низкая экспрессия обозначена черно-фиолетовым цветом.

Определив новые гены-кандидаты, мы провели исследование качества прогноза, ознакомившись с некоторыми исследованиями. В частности, десять генов-кандидатов врожденной нейтропении (*STAT1*, *STAT2*, *STAT3*, *STAT5B*, *LYN*, *FGR*, *SRC*, *PIK3CG*, *ITK*, *VAV1*, *CDC42*), которые отсутствовали в перечне ESID, но были предсказанные нами, были выявлены другими исследователями в клинических случаях врожденной нейтропении. Это подтверждает значимость выявленных нами генов-кандидатов врожденной нейтропении (таблица 13).

Таблица 13 – Гены-кандидаты врожденной нейтропении с недавно полученными сообщениями об ассоциации с врожденной нейтропенией

Ген	Сигнальные пути и основная роль	Эффект на нейтрофилы	Ссылки на публикации в библиотеке PubMed
CDC42	АТФ-связывающий компонент комплекса Agr2/3 через WASp	Actin polymerisation and phagocytosis	PMID: 19082760 PMID: 21178275 PMID: 10360578 PMID: 34425130
CRKL	АТФ-связывающий компонент комплекса Agr2/3 через WAVE	Actin polymerisation and phagocytosis	PMID: 11313252 PMID: 22837718 PMID: 12504004 PMID: 23934128
FGR	Тирозинкиназы/Адапторы и регуляция динамики актина для фагоцитоза	G-CSF Neutrophil regulation	PMID: 1895577 PMID: 8634424
SRC	Цитокиновый сигналинг в иммунной системе и Пигментный фактор эпителиального происхождения	G-CSF	PMID: 16772601
LYN	Передача сигнала от антигена	Initiation of the B-cell response, B-cell differentiation	PMID: 10643150 PMID: 23001182 PMID: 19201855
PLCG1	Путь фактора роста нейронов Путь CCR5 в макрофагах	-	PMID: 29543328 (?)
ARRB2	Цитокиновый сигналинг в иммунной системе и тирозиновые киназы/адапторы.	IL8-опосредованный выброс гранул в нейтрофилах	• PMID: 24657625
PIK3CG	Путь NF-каппаВ, Иммунный ответ сигналинг CCR3 в эозинофилах.	Вместе с PIK3CD участвует в респираторном взрыве нейтрофилов, хемотаксисе и экстравазации нейтрофилов	PMID: 29233821 PMID: 29191916 PMID: 31964785
PTK2	Путь NF-каппаВ, Иммунный ответ Цитокиновый сигналинг	-	Вопрос о болезни накопления гликогена 1b
STAT1	Путь пегинтерферона альфа-2а/пегинтерферона альфа-2b (гепатоцит), Фармакодинамика, Цитокиновая сигнализация в иммунной системе	Роль в иммунном ответе	PMID: 27879260 PMID: 29202461 PMID: 33344614 PMID: 27222657
STAT2	Пегинтерферон альфа-2а/Пегинтерферон альфа-2b путь (гепатоцит), Фармакодинамика, Иммунный ответ ИФН-гамма сигнальный путь	транскрипционный фактор	PMID: 27881648 PMID: 27713294

Ген	Сигнальные пути и основная роль	Эффект на нейтрофилы	Ссылки на публикации в библиотеке PubMed
STAT3	Цитокиновая сигнализация в иммунной системе Сигнальные пути ИЛ-4.	Г-КСФ, Созревание клеток иммунной системы, особенно Т- и В-клеток	PMID: 29330115 PMID: 28253502
STAT5B	Цитокиновый сигналинг в иммунной системе и сигнальные пути ИЛ-4.	Дифференцировка гранулоцитов	PMID: 29160632 PMID: 33255665 PMID: 24512550 PMID: 31585621
VAV1	Цитокиновый сигналинг в иммунной системе и развитие Трансактивация EGFR рецептором дофамина D2	Дифференцировка клеток Развитие и активация Т- и В-клеток	PMID: 12874226 PMID: 31456807 PMID: 10879282
ITK	Тирозинкиназы / адапторы и Т-клеточный рецептор и ко-стимулирующий сигналинг.	Регулирует развитие, функцию и дифференцировку Т- и НКТ-клеток	PMID: 32306816 PMID: 34365077 PMID: 34368657 PMID: 33007409 PMID: 32049330

Таким образом, в нашей работе было выявлено 15 генов-кандидатов врожденной нейтропении, которые могут оказывать влияние на функции нейтрофилов: *STAT1*, *STAT2*, *STAT3*, *STAT5B*, *LYN*, *FGR*, *SRC*, *PIK3CG*, *ITK*, *VAV1*, *CDC42*, *PTK2*, *CRKL*, *PLCG1*, *ARRB2*.

#### Список публикаций по материалам данной главы

1. Novel Disease-Associated Missense Single-Nucleotide Polymorphisms Variants Predication by Algorithms Tools and Molecular Dynamics Simulation of Human TCIRG1 Gene Causing Congenital Neutropenia and Osteopetrosis / K. Shinwari, H.M. Rehman, G. Liu, M.A. Bolkov, I.A. Tuzankina, V.A. Chereshevnev // Front. Mol. Biosci. – 2022. – 9. – 879875. (WoS Q2, Scopus Q1).

2. In Silico Analysis Revealed Five Novel High-Risk Single-Nucleotide Polymorphisms (rs200384291, rs201163886, rs193141883, rs201139487, and rs201723157) in ELANE Gene Causing Autosomal Dominant Severe Congenital Neutropenia 1 and Cyclic Hematopoiesis / K. Shinwari, M.A. Bolkov, M. Yasir Akbar, L. Guojun, S.S. Deryabina, I.A. Tuzankina, V.A. Chereshevnev // Scientific World Journal. – 2022- V. 2022. – 3356835. (Scopus Q1)

## ГЛАВА 5 - ИДЕНТИФИКАЦИЯ НОВЫХ МИССЕНС-МУТАЦИЙ ГЕНОВ *CCBE1*, *FAT4*, *ADAMTS3*, ПРИВОДЯЩИХ К СИНДРОМУ ХЕННЕКАМА

Задачей исследования была проверка потенциальной патогенности новых миссенс-замен (генов *CCBE1*, *FAT4* и *ADAMTS3*), выявленных в базах данных NCBI dbSNP, а также однонуклеотидных несинонимичных замен в гене *FAT4*, выявленных у пациента с клиническим диагнозом синдром Хеннекама, на структуру и функцию белков. Были отобраны наиболее вероятные вредоносные замены в этих генах и проверено их влияние на структуру и функцию белков путем включения замен в структуру белка дикого типа с помощью молекулярно-динамического моделирования.

### 5.1 - Идентификация вредоносных nsSNP в генах *FAT4*, *ADAMTS3*, *CCBE1*, приводящих к развитию синдрома Хеннекама

В общей сложности было проверено 407 nsSNP в гене *CCBE1* на предмет их влияния на структуру и функцию белка *CCBE1*. Из 407 nsSNP 23 были признаны вредоносными в обеих программах – SIFT и PolyPhen-2. Для 11 nsSNP получена информация о частоте минорного аллеля (MAF). За исключением T153N, G107D, P249S, S19N, C75S, C102S, G327R, C174R, D397Y, R125W, P87W и G330E, расчетная встречаемость других nsSNP в популяции была ниже 1% (таблица 14). В дальнейшем все 23 отобранных nsSNP были проанализированы с помощью дополнительных 16 инструментов для биоинформационного предказания вредоносности замен на структуру и функцию белка (таблица 15, рисунок 48).

Таблица 14 – Несинонимичные однонуклеотидные замены в гене *CCBE1*, оцененные в SIFT и PolyPhen2 как вредоносные

nsSNP	замена	SIFT	балл	PolyPhen-2	балл	MAF
rs199902030	D336N	Del	0.003	Prob damage	1	< 0.001 (T)
rs200149541	T153N	Del	0.001	Prob damage	1	
rs372499913	G107D	Del	0	Prob damage	1	
rs267605221	P249S	Del	0.007	Prob damage	1	
rs374941368	S19N	Del	0.004	Prob damage	0.981	
rs375717418	R301W	Del	0.004	Prob damage	1	< 0.001 (T)
rs80008675	D41E	Del L	0.016	Prob damage	0.982	0.017 (T)
rs116596858	P181S	Del L	0.007	Prob damage	0.906	< 0.001 (A)
rs116675104	R167W	Del L	0.017	Prob damage	0.990	0.003 (A)
rs121908250	C75S	Del L	0.002	Prob damage	0.981	
rs121908251	C102S	Del L	0	Prob damage	0.999	
rs121908252	G327R	Del	0	Prob damage	1	
rs121908254	C174R	Del	0.001	Prob damage	0.984	
rs147974432	T144M	Del L	0.002	Prob damage	1	< 0.001 (A)
rs192224843	Q353R	Del	0.011	Prob damage	0.993	< 0.001 (C)
rs115982879	R118L	Del L	0.001	Prob damage	0.910	< 0.001 (T)
rs139059968	K355T	Del	0.002	Prob damage	0.883	< 0.001 (G)
rs141125426	D397Y	Del L	0.002	Prob damage	0.828	
rs147208835	R125W	Del L	0	Prob damage	0.995	
rs147681552	P290L	Del	0.005	Prob damage	1	< 0.001 (A)
rs148498685	P87S	Del L	0.002	Prob damage	1	
rs149531418	G330E	Del	0	Prob damage	0.999	
rs149792489	A96G	Del L	0.004	Prob damage	1	< 0.001 (C)

*Примечание:* Замена – аминокислотная замена в белке; Del – вредоносная замена, Del L – вероятно менее повреждающая замена, Prob damage – вероятно повреждающая замена. MAF - минорная частота аллелей.

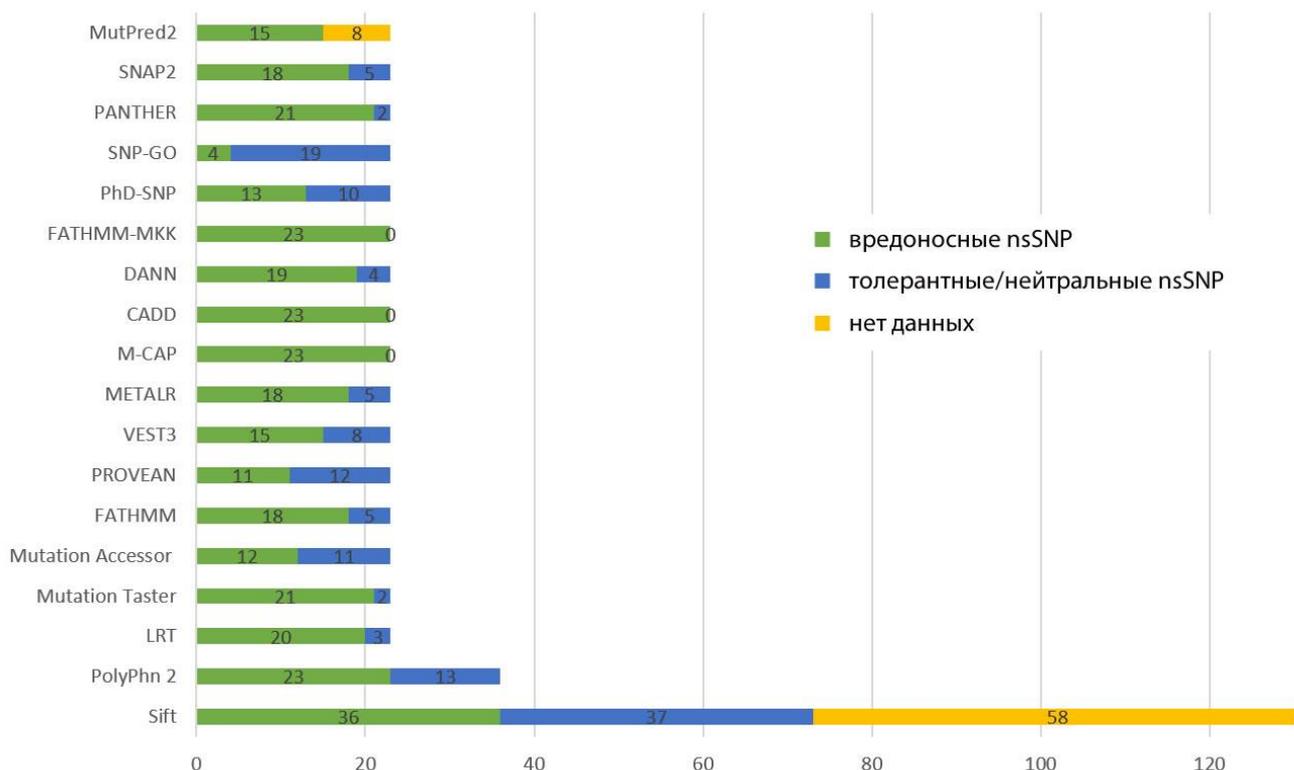


Рисунок 48 – Результаты предсказания влияния выявленных nsSNP в гене *CCBE1*, проанализированных восемнадцатью вычислительными инструментами

Для гена *ADAMTS3* было проверено 919 nsSNP. Только 50 из 919 nsSNP были оценены программой SIFT как вредоносные, и эти 50 были затем подвергнуты анализу нескольким инструментам (рисунок 49). Информация о частоте минорных аллелей (MAF) была доступна для 16 из них, MAF других nsSNP может составлять менее 1%.

Визуальное представление результатов фильтрации nsSNP гена *ADAMTS3* через 19 инструментов биоинформатики для предсказания патогенности замен (включая SIFT и Polyphen-2) представлено на рисунке 49. Все методы предсказания дали статистически значимые результаты. Для всех инструментов р-значение Т-теста Стьюдента составило 0,001.

Таблица 15 – Проверка вредоносности 23 выявленных nsSNP гене CCBE1 другими инструментами

Замена	LRT	Mutation Taster	Mutation Accessor	PROVEAN	FATHMM	VEST 3	MetaL R	M- CAP	CADD	DANN	FATHMM- МКК	PhD-SNP	PANTHER	SNP- GO	SNAP2
G330E	D	D	H	D	D	D	D	D	D	D	D	D	D	D	E
C102S	D	D	M	D	D	D	D	D	D	D	D	D	D	D	E
C174R	D	D	H	D	D	D	D	D	D	D	D	D	D	D	E
G107D	D	D	L	D	D	D	D	D	D	D	D	D	D	D	E
R125W	D	D	L	D	D	T	D	D	D	D	D	D	D	N	E
G327R	D	D	H	D	D	D	D	D	D	D	D	N	D	N	E
P290L	D	D	M	D	T	D	D	D	D	D	D	N	D	N	E
K355T	D	D	M	N	D	D	D	D	D	D	D	D	D	N	E
Q353R	D	D	M	N	D	D	D	D	D	D	D	D	D	N	E
D336N	D	D	M	N	D	T	D	D	D	D	D	D	D	N	E
T153N	D	D	M	N	D	T	D	D	D	D	D	D	D	N	E
C75S	D	D	L	D	D	D	D	D	D	T	D	N	D	N	E
P87S	D	D	L	N	D	D	D	D	D	D	D	D	D	N	E
T144M	D	D	L	N	D	D	D	D	D	D	D	N	D	N	E
R118L	D	D	L	D	D	D	T	D	D	D	D	D	D	N	E
D397Y	N	D	M	D	D	T	D	D	D	T	D	D	D	N	E
R301W	D	D	M	D	T	D	T	D	D	D	D	N	D	N	E
P249S	D	D	M	N	T	T	D	D	D	D	D	N	D	N	N
D41E	D	P	L	N	D	T	T	D	D	T	D	D	D	N	N
S19N	N	P	L	N	D	T	D	D	D	D	D	N	D	N	N
R167W	N	D	L	N	D	T	D	D	D	D	D	N	N	N	E
A96G	D	D	L	N	T	D	T	D	D	D	D	N	D	N	N
P181S	N	D	L	N	T	D	T	D	D	T	D	N	N	N	N

*Примечание:* Замена – аминокислотная замена в молекуле; следующие колонки – программы предсказания патогенности мутации. D – вредоносная замена, T – толерантная; N – нейтральная; L – низкая, M – средняя, H – высокая вероятность вредоносности; P – патогенная, E – оказывающая влияние

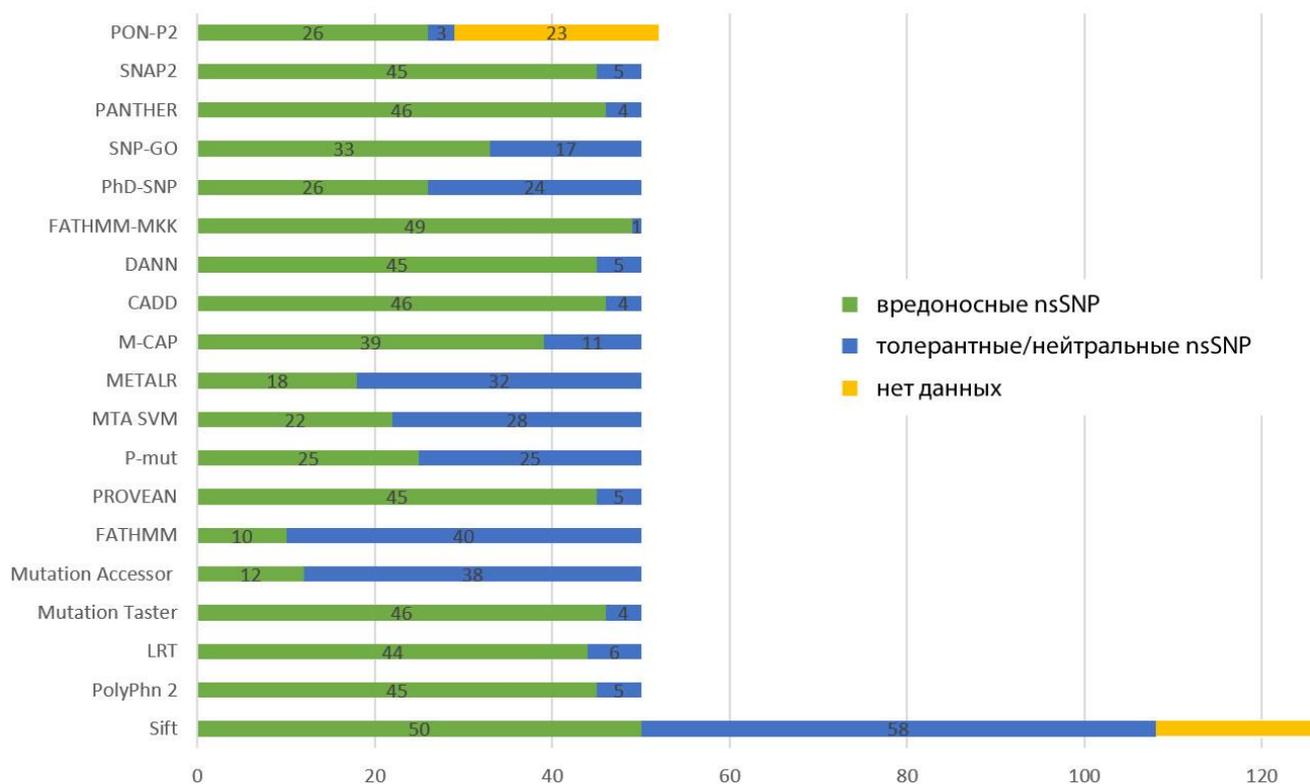


Рисунок 49 – Результаты предсказания влияния выявленных nsSNP в гене *ADAMTS3*, проанализированных 19 инструментами биоинформатики для предсказания патогенности замен (включая SIFT и Polyphen-2)

Для гена *FAT4* из 3434 nsSNP SIFT и PolyPhen-2 в совокупности предсказали 298 вредоносных или повреждающих nsSNP. Всего 70 nsSNP имели информацию о минорной частоте аллелей (MAF). За исключением замен G4361, S4710, A785, G1822, D1124N, R2285, R3128, R4726, H1513, S2098, A2959, T1914, V543I, R2285, K294R и G412, другие MAF для nsSNP гена *FAT4* имели значение менее 1%. После применения 18 различных инструментов биоинформатики для предсказания патогенности замен только 11 nsSNP – D2978G, V986D, Y1912C, R4799C, D1022G, G4786R, D2439E, E2426Q, R4643C, N1309I, Y2909H – считались полиморфизмами с высоким риском влияния на структуру и функцию *FAT4*, несмотря на то, что SIFT все 11 замен считал повреждающими с небольшой вероятностью (выше  $>0,5$ , но ниже  $<0,8$  по шкале SIFT) (таблица 16-17, рисунок 50).

Таблица 16 – Проверка вредоносности 11 выявленных nsSNP в гене *FAT4* другими инструментами in-silico

Замена	Mutation Taster	Mutation Accessor	FATH MM	PROVEAN	VEST3	MTA SVM	METALR	M-CAP	CADD	DANN	FATHMM-MKK	PhD-SNP	PANTHER	SNP-GO	SNAP2	P-Mut
D2978G	D	M	T	D	D	D	D	D	D	D	D	D	D	D	D	D
V986D	D	H	D	D	D	D	D	D	D	D	D	D	B	D	D	D
Y1912C	D	M	T	D	D	D	D	D	D	D	D	D	D	D	D	D
R4799C	D	L	D	D	D	D	D	D	D	D	D	D	D	N	D	D
D1022G	D	H	T	D	D	D	D	D	D	D	D	D	D	D	N	D
G4786R	D	L	D	D	D	D	D	D	D	D	D	D	D	N	D	D
D2439E	D	M	T	D	D	D	D	D	D	D	D	D	D	N	D	D
E2426Q	D	H	T	D	D	D	D	D	D	D	D	D	D	D	N	D
R4643C	D	L	D	D	D	D	D	D	D	D	D	D	D	D	D	N
N1309I	D	H	T	D	D	D	D	D	D	D	D	D	D	D	N	D
Y2909H	D	H	T	D	D	D	D	D	D	D	D	D	D	D	N	D

*Примечание:* Замена – аминокислотная замена в белке; D – вредоносная замена; N – нейтральная; L – низкая, M – средняя, H – высокая вероятность вредоносности; B – доброкачественная замена

Таблица 17 – Оценка отфильтрованных 11 nsSNP в гене *FAT4* и их минорная частота в популяции

nsSNP	Замена в белке	SIFT	Балл	PolyPhen-2	Балл	MAF
rs147663284	D2978G	менее вероятно, что повреждающая	0.005	вероятно повреждающая	0.99	
rs192514171	V986D	менее вероятно, что повреждающая	0	вероятно повреждающая	1.00	
rs138137489	Y1912C	менее вероятно, что повреждающая	0.001	вероятно повреждающая	1.00	
rs199895179	R4799C	менее вероятно, что повреждающая	0	вероятно повреждающая	1.00	<0.001 (T)
rs372060616	D1022G	менее вероятно, что повреждающая	0	вероятно повреждающая	1.00	
rs138173652	G4786R	менее вероятно, что повреждающая	0	вероятно повреждающая	1.00	<0.001 (A)
rs142184187	D2439E	менее вероятно, что повреждающая	0	вероятно повреждающая	0.99	
rs147633644	E2426Q	менее вероятно, что повреждающая	0	вероятно повреждающая	1.00	
rs181607904	R4643C	менее вероятно, что повреждающая	0	вероятно повреждающая	1.00	<0.001 (T)
rs184971791	N1309I	менее вероятно, что повреждающая	0	вероятно повреждающая	0.99	
rs148655455	Y2909H	менее вероятно, что повреждающая	0	вероятно повреждающая	1.00	

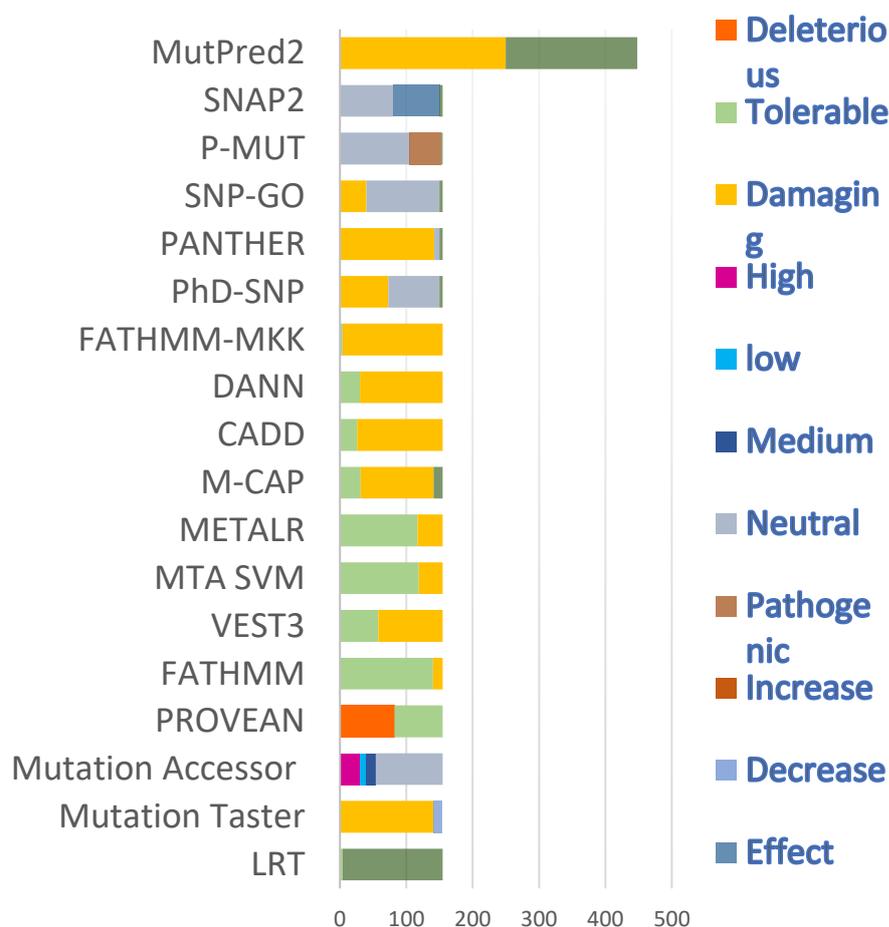


Рисунок 50 – Результаты предсказания влияния выявленных nsSNP в гене *FAT4*, проанализированных 18 инструментами биоинформатики для предсказания патогенности замен (после фильтрации в SIFT и Polyphen-2)

## 5.2 - Прогнозирование стабильности белков с включением несинонимичных замен в генах *FAT4*, *ADAMTS3* и *CCBE1*, выявленных на предыдущем этапе.

Для анализа предсказания стабильности *CCBE1* использовался веб-инструмент iStable 2.0. Этот веб-инструмент состоит из 11 инструментов предсказания на основе последовательности и структуры, и для всех результатов использовался подход машинного обучения. Результаты показали, что замены G330E, C174R, G327R, P290L, D41E, A96G, T114M, D397Y, S19N и Q359RT повышают стабильность, в то время как аминокислотные замены P249S, R167W, R301W, C75S, P87S, R118L, T153N, D336N, R125W, K355T, G107D и C102S снижают стабильность белка *CCBE1*.

Данных о замене P181S в программе iStable 2.0 получить не удалось. В *таблице 18* представлены прогнозы iStable 2.0.

Таблица 18 – прогнозы iStable 2.0 на стабильность белка ССВЕ1 с учетом выявленных аминокислотных замен.

Замена	Балл достоверности	Стабильность
G330E	-0.002680719	Повышает
C174R	0.021838337	Повышает
C102S	-1.2213084	Понижает
G107D	-0.86388123	Понижает
R125W	-0.85255766	Понижает
G327R	0.0042461157	Повышает
P290L	0.2298831	Повышает
K355T	-0.052274585	Понижает
Q353R	0.8725257	Повышает
D336N	-1.2082165	Понижает
T153N	-0.546193	Понижает
C75S	-1.0542232	Понижает
P87S	-1.9976869	Понижает
T144M	0.23297998	Повышает
R118L	-0.5704589	Понижает
D397Y	0.071232796	Повышает
R301W	-0.3441298	Понижает
P249S	-1.1325055	Понижает
D41E	0.4703572	Повышает
S19N	0.77003396	Повышает
R167W	-0.4350294	Понижает
A96G	-0.041893244	Повышает

Для оценки выявленных 50 nsSNP с высоким риском влияния на стабильность белка ADAMTS3, использовались программы I-Mutant 3.0 и MUpro. Прогноз нарушения стабильности белка ( $\Delta\Delta G$ ) в I-Mutant 3.0 показал,

что 47 nsSNP снижают стабильность ( $\Delta\Delta G < 0$ ), а 3 nsSNP повышают стабильность ( $\Delta\Delta G > 0$ ). MUpro идентифицировал 48 nsSNP, которые снижают стабильность белка по отдельности. Вариации с заменами S1038F, S58F и D791V (по данным I-Mutant), а также R576L, R954H и G412S (по данным MUpro) определены, как повышающие стабильность белка. Расчеты показывают, что структура и функция белка должны быть нарушены 19 вариациями, которые включали V395I, A336V, G298R, Q616H, Q927H, S1038F, G374S, D815Y, R94L, G983S, Q588H, G25H, R565W, R817C, R713L, R55L, N98S, Y636S, R576L, R1053C, D791V, G412S и L801F. Все эти варианты показали значения  $\Delta\Delta G$  менее - 1 ккал/моль, определенные этими двумя инструментами.

Используя те же инструменты I-Mutant 3.0 и MUpro (путем сравнения свободных энергий), оценивалось влияние выявленных на предыдущем этапе 11 nsSNP в гене *FAT4* на стабильность соответствующего белка (таблица 19).

Таблица 19 – Прогноз влияния выявленных аминокислотных замен на стабильность белка FAT4 (I-Mutant 3.0 и MUpro)

Замена	Стабильность	Замена	Стабильность
D2978G	Понижает	D2439E	Понижает
V986D	Понижает	E2426Q	Нет данных
Y1912C	Понижает	R4643C	Понижает
R4799C	Понижает	N1309I	Понижает
D1022G	Нет данных	Y2909H	Понижает
G4786R	Понижает		

### 5.3 - Анализ сохранения выявленных замен на консервативные участки белков CCBE1, ADAMTS3 и FAT4.

Исследование влияния 23 замен в гене *CCBE1* на белок *CCBE1* с использованием сервиса ConSurf показало, что 13 замен были расположены в высококонсервативных регионах белка. Одинадцать из них (C75S, P87S,

P290L, A96G, G107D, R118L, G330E, D336N, R125W, Q353R и T153N) были предсказаны как функциональные и открытые остатки, а два других - C102S и C174R – как заглубленные и структурные аминокислотные остатки. Замена S19N была предсказана как консервативный и заглубленный остаток, а оставшиеся 8 (T144M, R167W, P249S, R301W, G327R, K355T, D397Y и D41E) - как открытые аминокислотные остатки (на поверхности белка). Результаты показаны на *рисунке 51*.

Аналогичное исследование было проведено для ADAMTS3 и выявленных 50 nsSNP. Двадцать шесть из 50 миссенс-вариантов были определены, как располагающиеся в высококонсервативной области. И 19 из 26 (Q927R, G298R, C567Y, C567R, Q616H, R565W, R565Q, P371S, P513T, R248H, T668M, R435H, N98S, R883C, G412S, L801F, S1038F, G983S, R959W), как ожидается, являются функциональными и открытыми остатками, а остальные 7 (I291T, V395I, A336V, G374S, S58F, I287F и A370T), как ожидается, являются заглубленными и структурными остатками.

Также консервированными и заглубленными остатками определены замены F81L, Y148C, R435H, Y536C, M731T, F777L, R94L, R270H, P510A, R572C, R572H, Q588H, R713L, R817C, R943H и R954H. Кроме того, восемь замен расположены на поверхности (G25V, R55L, P77T, R137W, Y636C, D791V, D815Y и R1053C). Среди 11 nsSNP с высоким риском вредоносности 7 (D1022G, N1309I, D2439E, E2426Q, R4799C, G4786R и R4643C) предсказаны как функциональные и открытые остатки, в то время как остальные 3 (V986D, Y1912C, Y2909H) должны быть заглубленными (*рисунк 52*).

Для белка FAT4 ConSurf показал, что многие из аминокислотных замен, определенных ранее как имеющие высокий риск влияния на белок, оказались в высококонсервативных регионах. Семь из 11 nsSNPs (в позициях 1022, 1309, 2439, 2426, 4799, 4786 и 4643) должны были быть функциональными и открытыми остатками, в то время как остальные считались заглубленными. Структурно заглубленным аминокислотным остатком признан 4786 (замена

G4786R), в то время как замена G298 воздействовала на открытый участок белка. Для белка FAT4 в силу его размера изображение не приводится.

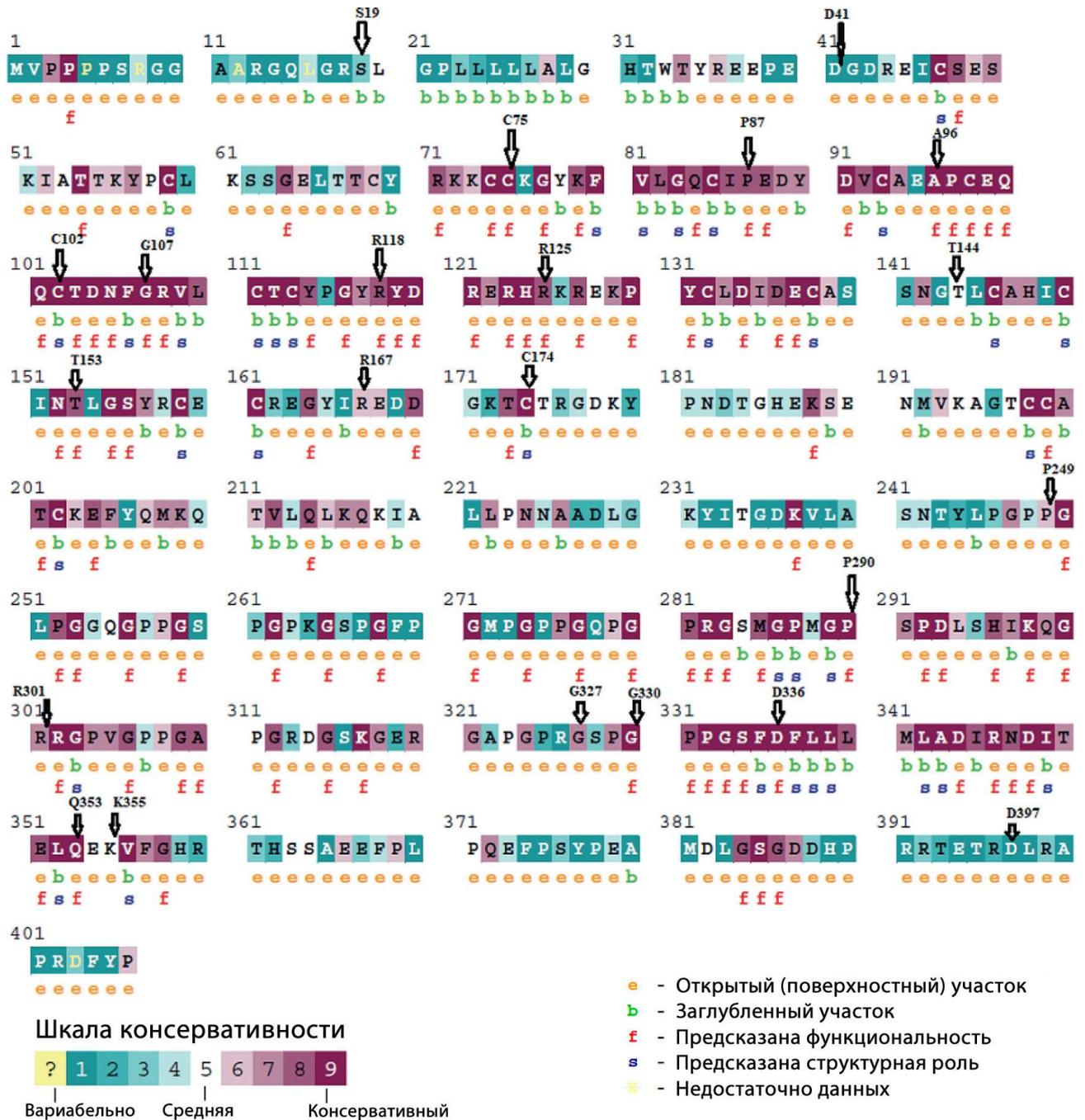
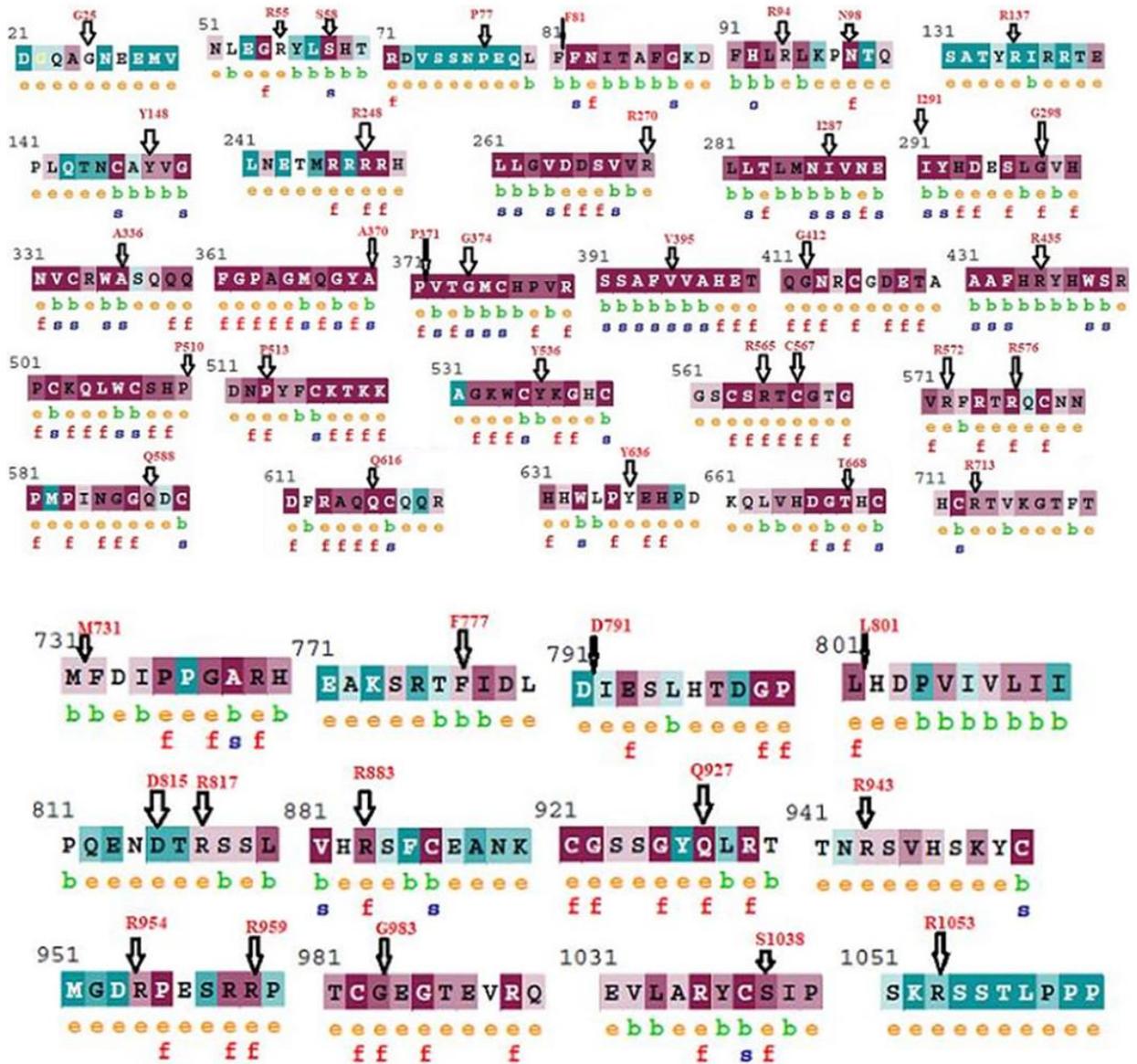


Рисунок 51 – Расположение аминокислотных замен в белке CCBE1 с учетом эволюционной сохранности и расположения различных участков белка по данным сервиса ConSurf

*Примечание* Значение 1 указывает на высокую изменчивость региона, 9 – на самый консервативный регион с точки зрения эволюции.



**Шкала консервативности**



- e - Открытый (поверхностный) участок
- b - Заглубленный участок
- f - Предсказана функциональность
- s - Предсказана структурная роль
- x - Недостаточно данных

Рисунок 52 – Расположение аминокислотных замен в белке ADAMTS3 с учетом эволюционной сохранности и расположения различных участков белка по данным сервиса ConSurf

*Примечание:* Значение 1 указывает на высокую изменчивость региона, 9 – на самый консервативный регион с точки зрения эволюции.

#### **5.4 - 3D-моделирование структур белков ССВЕ1 дикого и мутантного типов**

Для моделирования 3D-структур белка ССВЕ1 дикого типа и 22 мутантных типов был использован Phyre2, который предсказал 3D-структуры мутантных белков. В качестве шаблона для предсказания 3D-модели ССВЕ1 в Phyre2 была выбрана модель c5to3B.

Модель для мутанта R118L (rs115982879) показала наибольшее отклонение, имея значение RMSD 1,56Å, за ним следуют A96G (rs149792489), S19N (rs374941368) C174R (rs121908254) со значениями RMSD 1,50Å, 1,44Å и 1,46Å соответственно. R125W, C75S и T153N показали значения RMSD 0,89Å, 0,90Å и 0,85Å, то есть, не демонстрируя никаких изменений в структуре по сравнению с диким типом.

Другие замены аминокислот показали незначительное влияние на 3D-структуру ССВЕ1. Это были замены G327R (1.36Å RMSD), P290L (1.36Å RMSD), Q353T (1.32Å RMSD), P290L (1.25Å RMSD), D336N (1.25Å RMSD), C102R (1.22Å RMSD), R167W (1.16Å RMSD), P87L (1.14Å RMSD), G107D (1.13Å RMSD), T144M (1.13Å RMSD), G330R (1.12Å RMSD), D41E (1.12Å RMSD) D297Y (1.06Å RMSD), R301W (1.02Å RMSD) и K355T (1.01Å RMSD). Четыре nsSNP (R118L, A96G, S19N и C174R) с самыми высокими значениями RMSD были отобраны и переданы в I-TASSER для ремоделирования. Структура белка, полученная с помощью I-TASSER, является наиболее надежной, поскольку это самый современный инструмент моделирования. Каждый из этих трех мутантов был оценен и наложен с помощью Chimera 1.11 на белок ССВЕ1 дикого типа, как показано на *рисунке 53*.

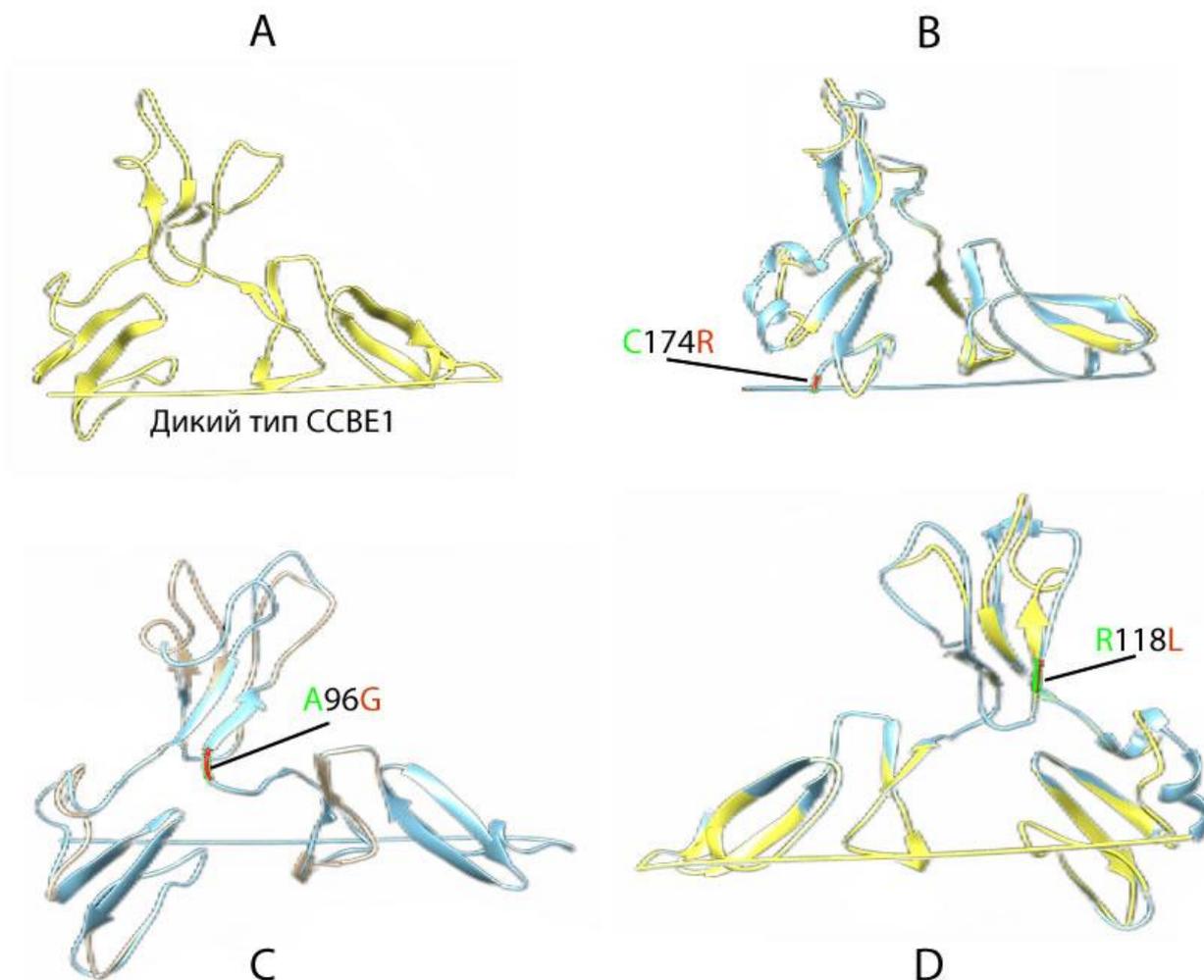


Рисунок 53 – (А) Структура белка ССВЕ1 дикого типа. (В) Наложенная структура ССВЕ1 и его мутанта С174R. (С) Наложенная структура ССВЕ1 и его мутанта А96G. (D) Наложенная структура ССВЕ1 и его мутанта R118L  
*Примечание:* Представлен результат визуализации модели Phyre2 в программе Chimera 1.11.

Нами проводилась оценка сайтов связывания лигандов с помощью FTSite, то есть анализ докинга белка ССВЕ1 и оценка влияния на докинг выявленных однонуклеотидных замен. Сайты связывания лигандов были предсказаны алгоритмами FTSite, визуализированы и далее проанализированы с помощью PyMOL. С помощью этого инструмента в белке ССВЕ1 человека было идентифицировано 3 сайта связывания лигандов. Сайт 1 состоял из 14 аминокислотных остатков; сайт 2 и сайт 3 состояли из 7 и 5 остатков соответственно. Некоторые из 22 аминокислотных замен,

предсказанных сервером SIFT как потенциально вредоносные, локализованы в предполагаемых сайтах связывания лигандов (T153N и R167W). Данные результаты не получили дальнейшего использования, но представлены в приложении к опубликованной статье по данному материалу, а в данную диссертационную работу не включены.

Для предсказания посттрансляционных модификаций был использован набор различного программного обеспечения. В частности, GPS-MSP 3.0 показала отсутствие сайтов метилирования в CCBE1. Программы GPS 3.0 и NetPhos 3.1 предсказали сайты фосфорилирования CCBE1, участки, имеющие потенциал фосфорилирования. Программы BDM-PUB и UbPred были использованы для предсказания убиквитилирования. В частности, BDM-PUB предсказал убиквитилирование 11 остатков лизина). Программа NetOGlyc4.0 была использована для предсказания потенциальных сайтов гликозилирования и потерю гликозилирования в определенных участках вследствие описанных замен. Данные результаты не получили дальнейшего использования, но представлены в приложении к опубликованной статье по данному материалу, а в данную диссертационную работу не включены.

### **5.5 - 3D-моделирование структур белков ADAMTS3 дикого и мутантного типов**

Структуры ADAMTS3 дикого и мутантных типов предсказаны с помощью AlphaFold 2. Визуализация осуществлена в Chimera 1.3. В моделирование структуры мутанта были включены 25 мутаций включены. 21 из них – S58F, I291T, G298R, A336V, A370T, P371S, G374S, G412S, R435H, Y536C, R565W, C567R, R572C, R576L, Q616H, Y636C, T668M, R883C, R954H, R959W и G983S – подтверждены несколькими программами как пагубные (C567Y не включен, поскольку занимает ту же позицию, что и C567R), а 4 из них (R138K, R574C, C578L и Q606H) были обнаружены клинически другими

исследователями. Модели дикого типа и мутанта подтверждены графиком Рамачандрана и анализом всеатомных контактов программой MolProbity. Модель дикого типа представляла 1032 остатка (85,8%) в благоприятной области, 77 (6,4%) в разрешенной области и 94 (7,8%) в области отклонений, а общее количество остатков в благоприятной и разрешенной областях составило 1109 (92,2%). Мутантная модель представляла 1008 остатков (83,8%) в благоприятной области, 110 (9,1%) в разрешенной области и 85 (7,1%) в области отклонений, а общее количество остатков в благоприятной и разрешенной областях составляла 1118 (92,9%).

Для всеатомного контактного анализа MolProbity белок дикого типа показал уровень очков 3,61, а мутантный - 1,88, что является приемлемым значением.

Структуру белка можно разделить на 3 сегмента (сегмент 1: Met1-Pro466; сегмент 2: Lys467-Val831; сегмент 3: Pro832-Arg1205), которые соединены петлями (*рисунок 54*). Сегмент 3 обоих белков состоит в основном из петель без множества вторичных структур, поэтому мы считаем его неточным предсказанием и игнорируем для дальнейшего анализа.

Наши последующие исследования были сосредоточены, в основном, на сегментах 1 и 2, которые содержат обширные вторичные структуры. Мы предполагаем, что между сегментами существуют незначительные взаимодействия, поэтому мутации в одном сегменте не окажут существенного влияния на другой. Наложение структур дикого типа и мутанта ADAMTS3 (*рисунок 55*) демонстрирует значение RMSD в 30,367 Å.

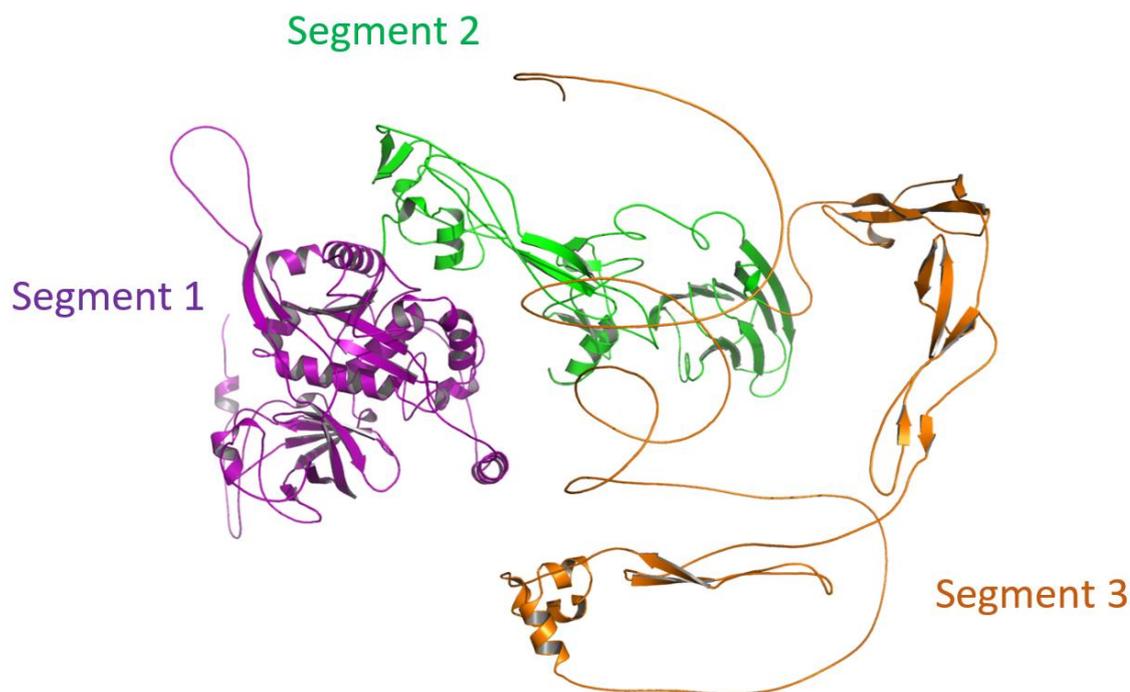


Рисунок 54 – Сегментированная структура белка ADAMTS3 на примере молекулы дикого типа

*Примечание:* Сегмент 1 (остатки 1-466), сегмент 2 (остатки 467-831) и сегмент 3 (остатки 832-1205). Три сегмента соединены петлями. Сегмент 3 состоит, в основном, из петель. На рисунке представлен результат визуализации модели из AlfaFold2 в программе Chimera 1.3.

Также для оценки последствия аминокислотных замен с высоким уровнем риска патогенности на сайты связывания лигандов был проведен анализ докинга мутантных типов ADAMTS3. Сайты связывания лигандов белка ADAMTS3 были предсказаны с помощью сервера RaptorX Binding (установлено значение «кратности кармана» более 40) и сервера предсказания сайтов связывания лигандов COACH. Анализ RaptorX Binding определил кратность кармана 151, что является самым высоким показателем, и связал его с остатками, которые подвержены заменам G365, M366, Q367, G368, Y369, V395, H398, E399, H402, H408, A426, P427, L428 и V429 с ожидаемым лигандом катиона  $Zn^{2+}$ . Сервер COACH предсказал сайт связывания катиона  $Zn^{2+}$  с C-score 0,15, расположенный на остатках H398, H402 и H408. Сайты второго ранга, идентифицированные COACH, были связаны с катионом  $Co^{2+}$  в остатках E259, L334, 351, 355 и 356.

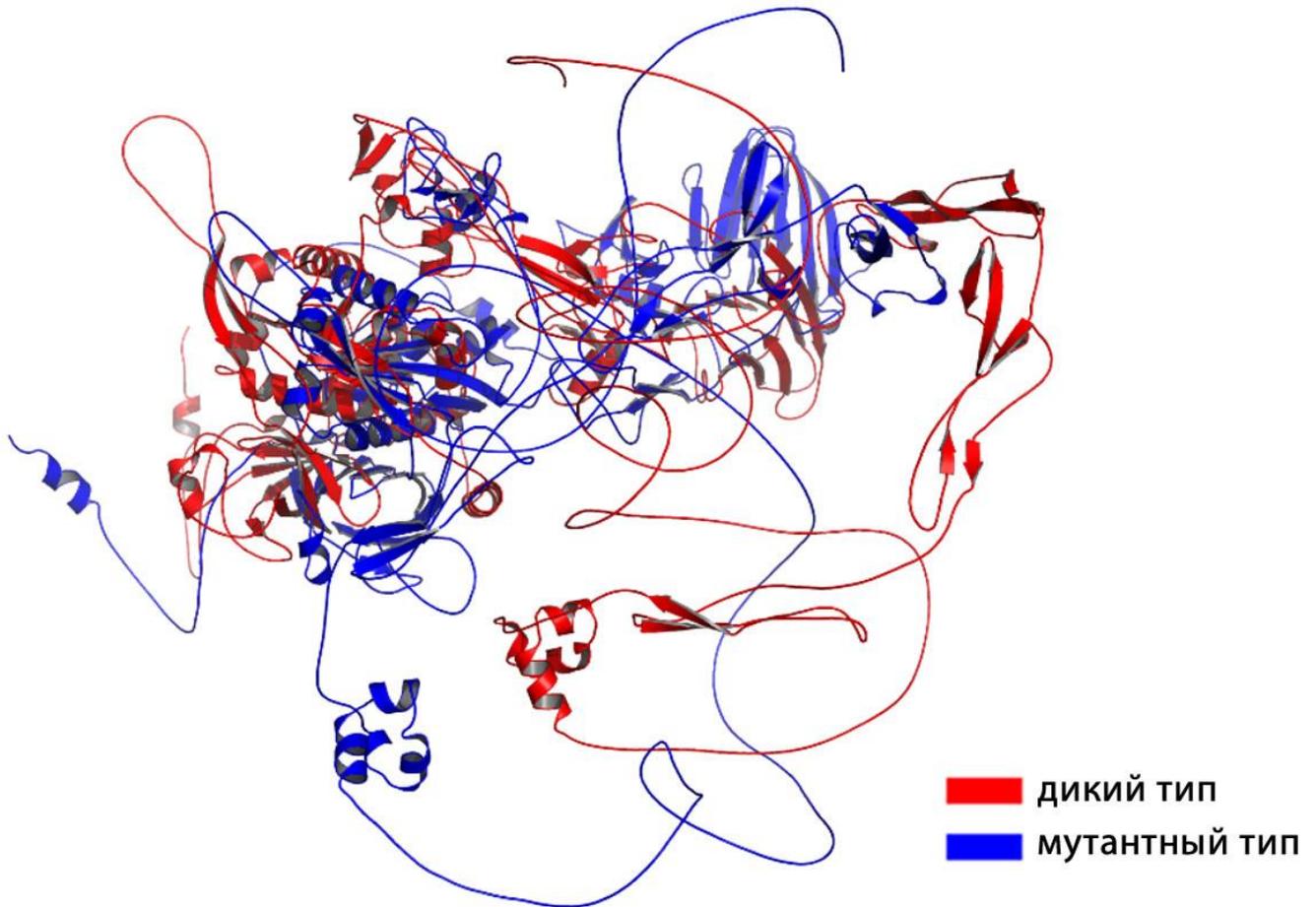


Рисунок 55 – Наложение структур дикого типа и мутанта ADAMTS3 (результат визуализации модели из AlfaFold2 в программе Chimera 1.3)

Кроме того, мы изучили эффекты каждой мутации и то, как они влияют на близлежащие структуры. В Project HOPE проанализировано влияние 50 отобранных несинонимичных однонуклеотидных полиморфизмов (nsSNP) в ADAMTS3 на размеры аминокислот, заряды и гидрофобность. Среди этих nsSNP 26 приводили к уменьшению размера аминокислот, в то время как 22 - к увеличению. В 23 участках был изменен заряд, причем в 20 он изменился с положительного на нейтральный, в одном - с нейтрального на положительный, а в двух - с отрицательного на нейтральный. Гидрофобность уменьшалась в семи мутациях, в то время как 22 другие приводили к ее увеличению. Эти результаты свидетельствуют о том, что изменения свойств аминокислот в этих позициях могут влиять на структуру белка и его взаимодействие с другими молекулами, что в конечном итоге сказывается на функции белка. Также были

исследованы локальные 3D-структуры вышеупомянутых 25 мутаций, включенных в модели белков AlphaFold. Результаты показывают, что большинство мутаций не оказывают большого влияния на структуру последовательности вблизи позиции аминокислотной замены в 25 мутациях. Только замена Y536C имеет существенное нарушение вторичной структуры вблизи по сравнению с другими мутациями (рисунк 56). Остальные изображения 3D-структур, измененных после аминокислотной замены представлены в приложении к опубликованной статье по данному материалу и в данную диссертационную работу не включены.

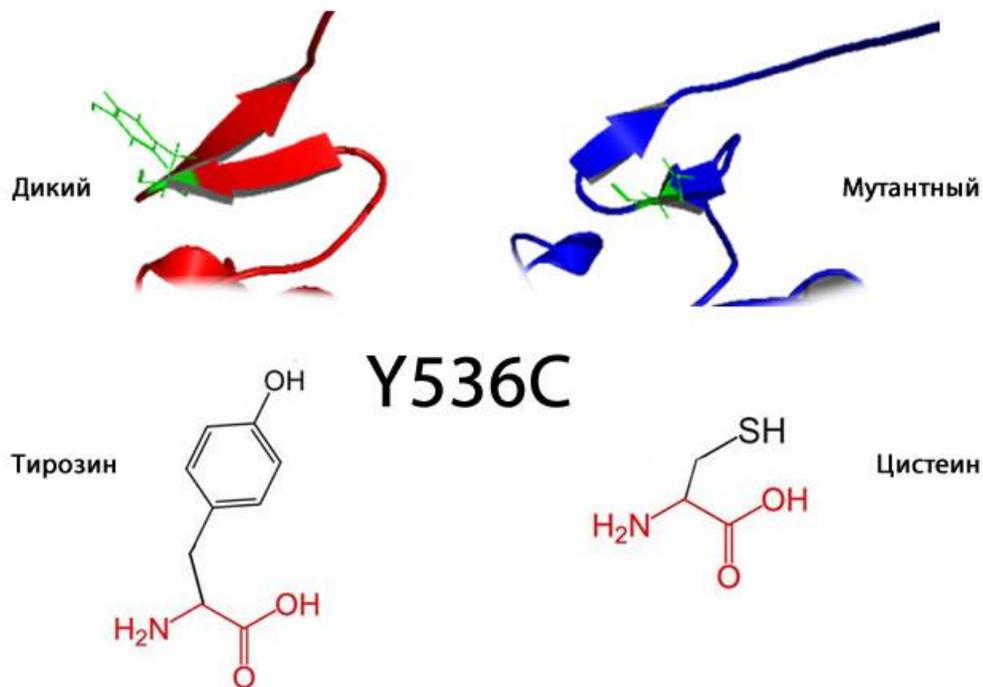


Рисунок 56 – Изменение в трехмерной структуре ADAMTS3 при аминокислотной замене Y536C. Сайты мутаций окрашены в зеленый цвет. Результат Project HOPE

Также был проведен расчет общих посттрансляционных модификаций, включая метилирование, фосфорилирование, убиквитинирование и гликозилирование для структуры дикого типа ADAMTS3 и мутанта. GPS-MSP предсказал отсутствие метилированных сайтов в ADAMTS3. Предсказанные сайты фосфорилирования серина, треонина и тирозина различными киназами

отличаются между NetPhos 3.1 и GPS 6.0, причем GPS 6.0 предсказывает больше сайтов фосфорилирования, чем NetPhos 3.1 для обеих структур. Примечательно, что некоторые сайты фосфорилирования появляются и исчезают после мутации. GPS 6.0 показывает исчезновение сайтов на Ser58, Tyr536, Tyr636 и Thr668 и появление новых сайтов на Ile291, Ala370, Pro371, Gly374, Gly412 и Gly983, в то время как GPS 6.0 показывает исчезновение сайтов фосфорилирования на Ser58, Tyr536, Tyr636 и Thr668 и появление новых сайтов на Ile291, Ala370, Pro371, Gly374, Gly412 и Gly983, а NetPhos 3.1 показывает исчезновение сайтов на Tyr56, Ser58 и Ser957 и появление новых сайтов на Ile291, Pro371, Gly374, Gly412 и Gly983. Большинство этих изменений приходится на сайты мутаций с участием серина, треонина и тирозина. Примечательно, что больше изменений в сайтах фосфорилирования находится в сегменте 1. Для убиквитинирования, UbPred обнаружил 9 лизиновых остатков сайтов убиквитинирования в структуре дикого типа и мутанта, в то время как BDM-PUB обнаружил 37 и 36 убиквитинированных лизиновых остатков в диком типе и мутантной белке, соответственно, и после мутации, есть несколько новых и исчезновение сайтов убиквитинирования, большая часть которых находятся в сегменте 3. Анализ с использованием NetOGlyc4.0 предсказал все возможные сайты O-гликозилирования в обеих белках, и некоторые мутанты потеряли или приобрели гликозилирование в определенных позициях, большинство из которых расположены в сегменте 3. Данные результаты не получили дальнейшего использования, но опубликованы и приведены в дополнительных файлах к публикации по этой теме.

## **5.6 - 3D-моделирование структур белков FAT4 дикого и мутантного типов**

Мы создали 5 моделей белка FAT4 с помощью I-Tasser и проверили влияние замен на качество и структуру мутантных типов белка. Поскольку FAT4 является большим белком, мы моделировали только те

последовательности белка, где были обнаружены мутации нашего клинического случая (уже опубликованные), а также наиболее пагубные мутации, полученные по результатам вышеописанного *in-silico* исследования.

Затем, используя оценку QMEAN, Prosa Z-score и анализ диаграммы Рамачандрана, каждая структура была проверена на надежность. На лучшую модель белка указывала более положительная оценка QMEAN, Prosa сканировала структуры моделей и сравнивала их с кристаллическими структурами PDB для определения качества моделей. На основе анализа QMEAN, Prosa Z-score и графика Рамачандрана выбранные модели были оптимизированы путем минимизации энергии с помощью UCSF Chimera (*рисунок 57*). Аналогичные 3D-модели последовательностей 2-5 участков белка FAT4 не получили дальнейшего использования, но опубликованы и приведены в дополнительных файлах к публикации по этой теме.

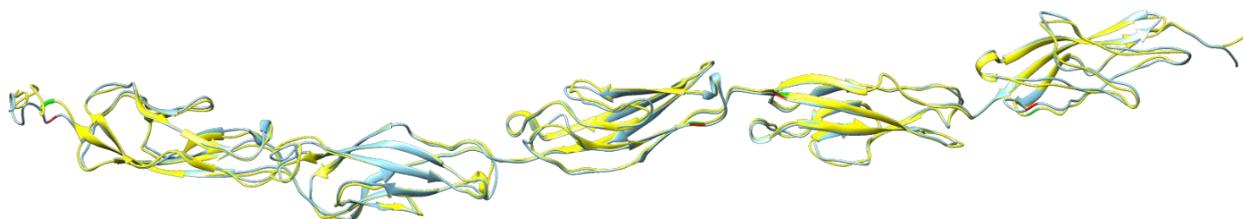


Рисунок 57 – Наложение 3D-моделей FAT4, последовательность 1, содержащая замены A807V, V986D, D1022G и N1309I. Желтый - дикий тип FAT4, голубой - мутантный вариант FAT4 с красным выделением мутаций

В Project NOPE проанализированы последствия 11 исследуемых аминокислотных замен в FAT4 на размеры аминокислот, заряды и гидрофобность. Четыре мутантные аминокислоты были крупнее своих аналогов дикого типа, тогда как шесть мутантных аминокислот имели меньший размер. Заряд был изменен на восьми различных участках: 2 с положительного на нейтральный, 1 с нейтрального на положительный, 3 с нейтрального на отрицательный и 2 с отрицательного на нейтральный. Анализ показал, что гидрофобность была уменьшена в пяти мутациях и увеличена в четырех других. Эти результаты позволили предположить, что аминокислотные мутации влияют на функцию белка через изменение

структуры белка и взаимодействие с другими компонентами. Эти nsSNP, которые оказались наиболее пагубными, приведены с возможным модельным шаблоном: Y1912C (5DZY), D2439E (1L3W), E2426Q (1L3W), D2978G (5W1D) и Y2909H (1L3W). Они обеспечивают уникальную конформацию центральной оси молекулы. Тем не менее, данные результаты в дальнейшем не использовались, но опубликованы и приведены в дополнительных файлах к публикации по этой теме.

Ещё до построения 3D-модели FAT4 мы провели предсказание вторичной структуры FAT4 с помощью программы SOPMA, которая помогла уточнить распределение альфа-спиралей, бета-слоев и случайных витков. Анализ вторичной структуры показал наличие 49,31% случайных витков (1148), затем 36,90% протяженных витков (859), 8,98% альфа-спиралей (209) и 4,81% бета-слоев (111). Распределение аминокислотных замен по вторичным структурам в дальнейшем не рассматривалось, данные результаты в дальнейшем не использовались, но опубликованы и приведены в дополнительных файлах к публикации по этой теме.

Анализируя возможное влияние аминокислотных замен на посттрансляционные модификации в белке FAT4, программа GPSMSP 3.0 не предоставила информации о метилировании в этом белке. NetPhos 3.1 предсказал сайт фосфорилирования для 579 остатков. Инструмент UbPred предсказал, что ни один из остатков лизина не может быть убиквитинирован. В отличие от этого, BDMPUB предсказал, что 101 остаток лизина может быть убиквитинирован, но ни один из них не попадает в список анализируемых аминокислотных замен. Также были оценены сайты гликозилирования с помощью SUMOylation. Данные результаты не получили дальнейшего использования, но опубликованы и приведены в дополнительных файлах к публикации по этой теме.

## 5.7 - Молекулярно-динамическое моделирование дикого и мутантного типа ADAMTS3

Изменение значений среднеквадратичного отклонения (RMSD)  $C\alpha$  атомов дикого и мутантного типов ADAMTS3 представлено на *рисунке 58*. Для сегмента 1 системы как дикого типа, так и мутантных структур достигают равновесия через 130 нс, после чего значения RMSD двух структур существенно не отличаются, что показывает, что мутации в сегменте 1 не сильно влияют на структуру (дикий тип: среднее: 10,830 Å, SD: 0,169 Å; мутант: среднее: 11,109 Å, SD: 0,157 Å). Однако для сегмента 2 белок дикого типа достигает стабильности чуть менее чем за 10 нс. После этого система уравнивается, и моделирование сходится в течение всего времени работы, но значения RMSD мутантного белка колеблются сильнее по сравнению со структурой дикого типа на протяжении всего моделирования. Мутантная структура имеет больший RMSD. Это указывает на то, что мутации в сегменте 2 сильнее дестабилизируют этот участок протеина (дикий тип: среднее: 5.31 Å, SD: 0.344 Å; мутант: среднее: 14.312 Å, SD: 0.584 Å).

Области белков, которые больше всего колеблются во время моделирования, показаны пиками на графиках RMSF (*рисунок 59*).  $\beta$ -слои и  $\alpha$ -спирали часто более жесткие и менее изменчивые, чем неструктурированный компонент белка. В сегменте 1, хотя пик на остатках Asn119-Pro129 больше для структуры дикого типа, RMSF дикого типа и мутантных структур в целом схожи. Это показывает, что мутации в сегменте 1 не стабилизируют и не дестабилизируют структуру в значительной степени. В сегменте 2 общий RMSF мутантной структуры выше, чем у дикого типа, что говорит о том, что мутации дестабилизировали структуру в этом сегменте. Существует большая разница в RMSF в остатках Met478-Pro523, что указывает на то, что эта область наиболее дестабилизирована.

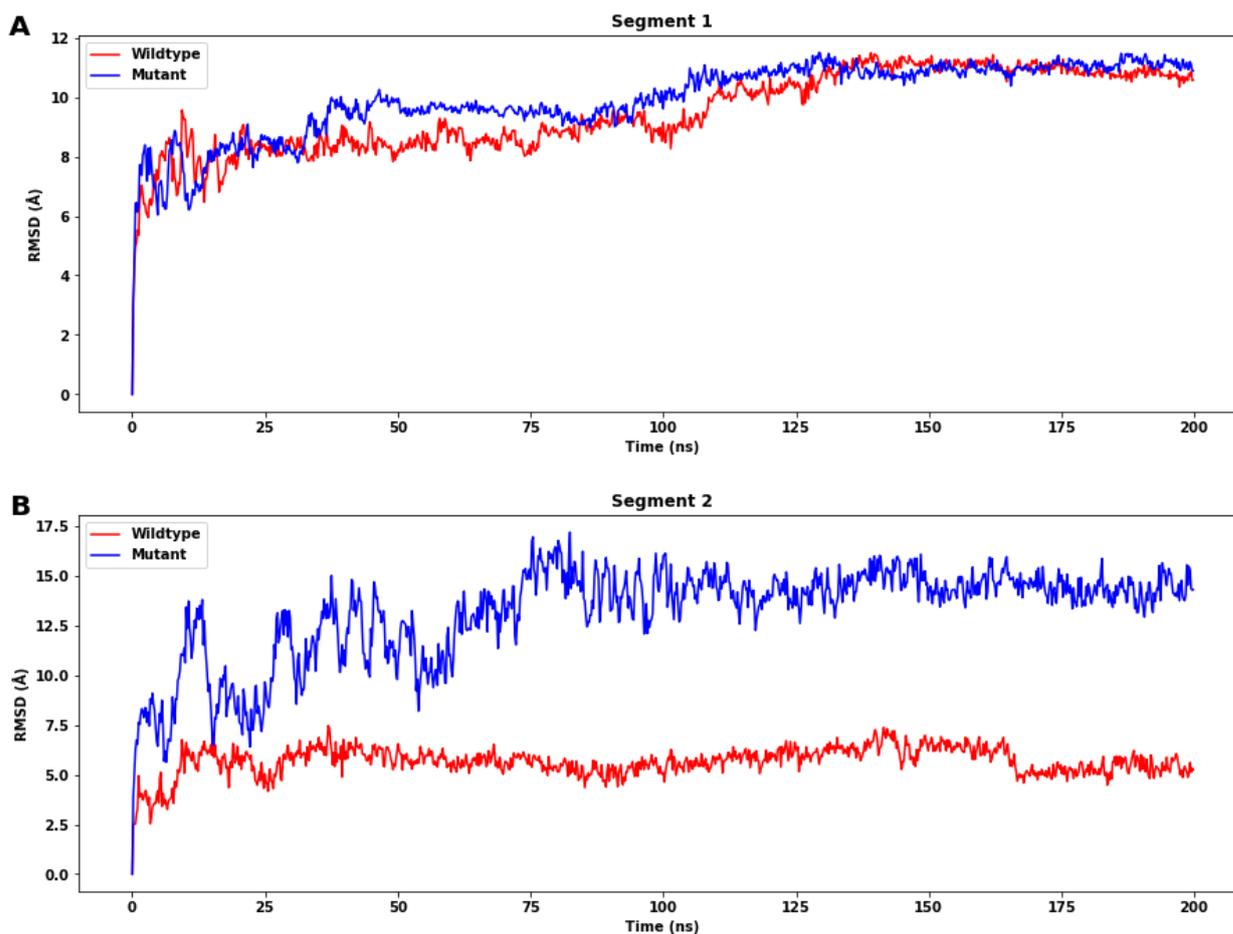


Рисунок 58 - Среднеквадратичное отклонение (RMSD) C $\alpha$  атомов белка ADAMTS3 дикого типа (красный) и мутанта (синий) сегментов 1 (A) и 2 (B) с течением времени

*Примечание:* Для сегмента 1 (A) разница в равновесных RMSD незначительна. Для сегмента 2 (B) существует значительная разница между RMSD. Все средние значения и SD рассчитаны со значениями после 170 нс.

Также рассчитываются средние распределения элементов вторичной структуры белка во время моделирования через 170 нс. Для сегмента 1 (рисунок 60) процент средней вторичной структуры  $\alpha$ -спиралей снизился на 5,15% в мутантной структуре по сравнению с диким типом, но наблюдается увеличение процент  $3_{10}$ -спиралей на 3,86%, что может стабилизировать мутантную структуру и противодействовать дестабилизирующему влиянию нарушения  $\alpha$ -спиралей мутациями, и увеличение процента витков на 1,50%.

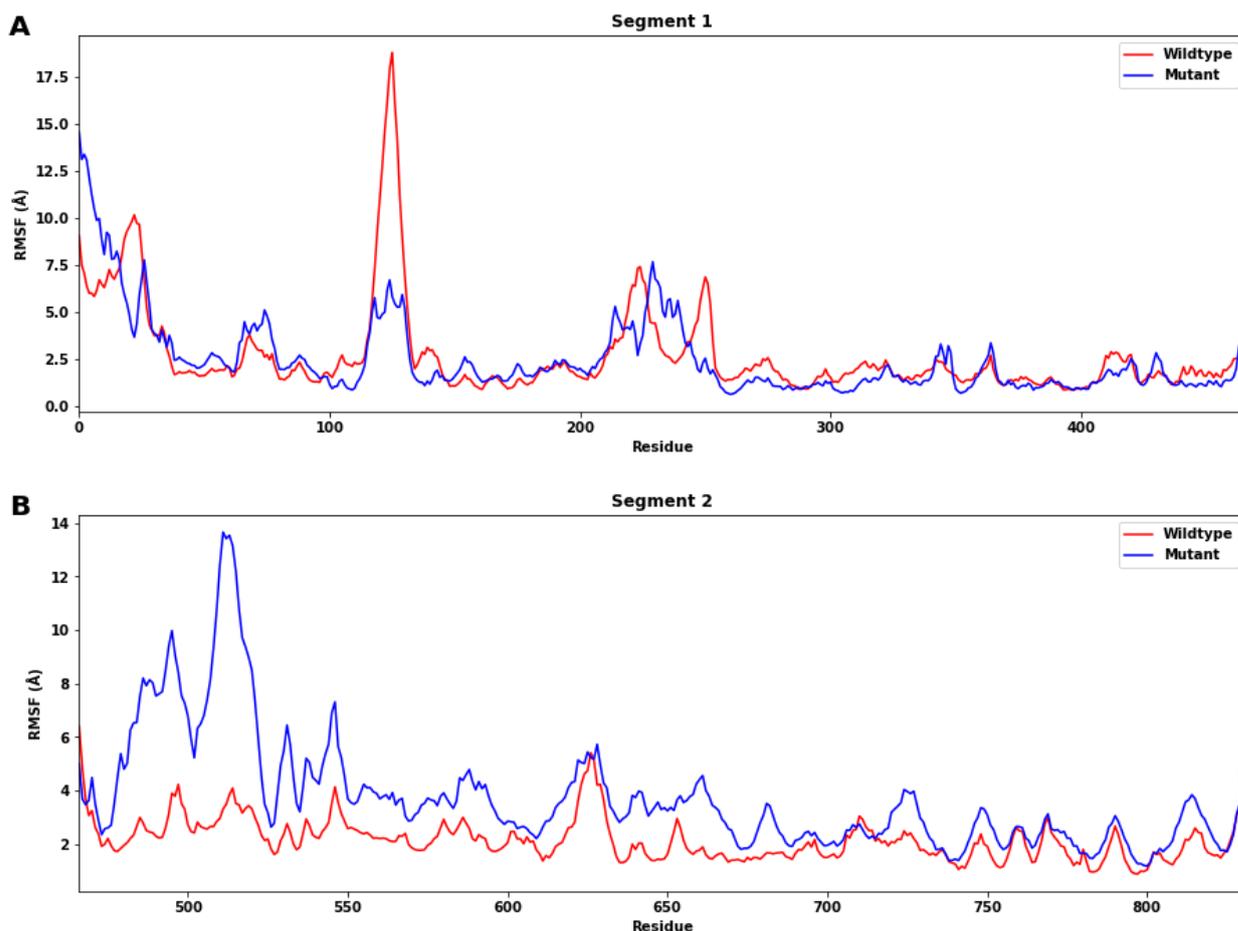


Рисунок 59 – Среднеквадратичная флуктуация (RMSF) для каждого участка белка дикого (красный) и мутантного типа (синий) сегментов 1 (А) и 2 (В). *Примечание:* Для сегмента 1 RMSF структур дикого типа и мутанта в целом похожи. Для 2 сегмента общий RMSF мутантной структуры выше, чем у структуры дикого типа.

Для сегмента 2 (рисунок 61) наблюдается снижение процента  $\beta$ -листов на 5,48% и увеличение процента витков на 4,11%, что может дестабилизировать общую структуру мутантного белка в этом сегменте. Кроме того, процент  $\alpha$ -спиралей также увеличился на 1,10%.

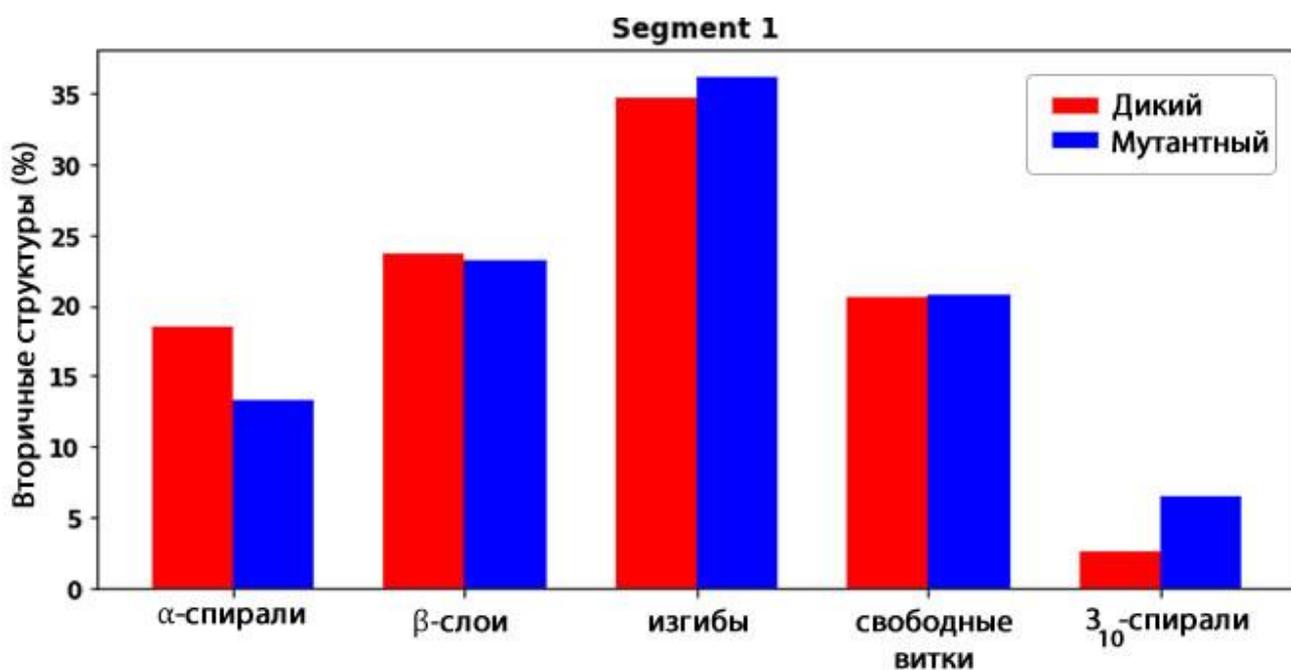


Рисунок 60 – Распределение вторичных структур в сегменте 1 после 170 нс симуляции молекулярной динамики

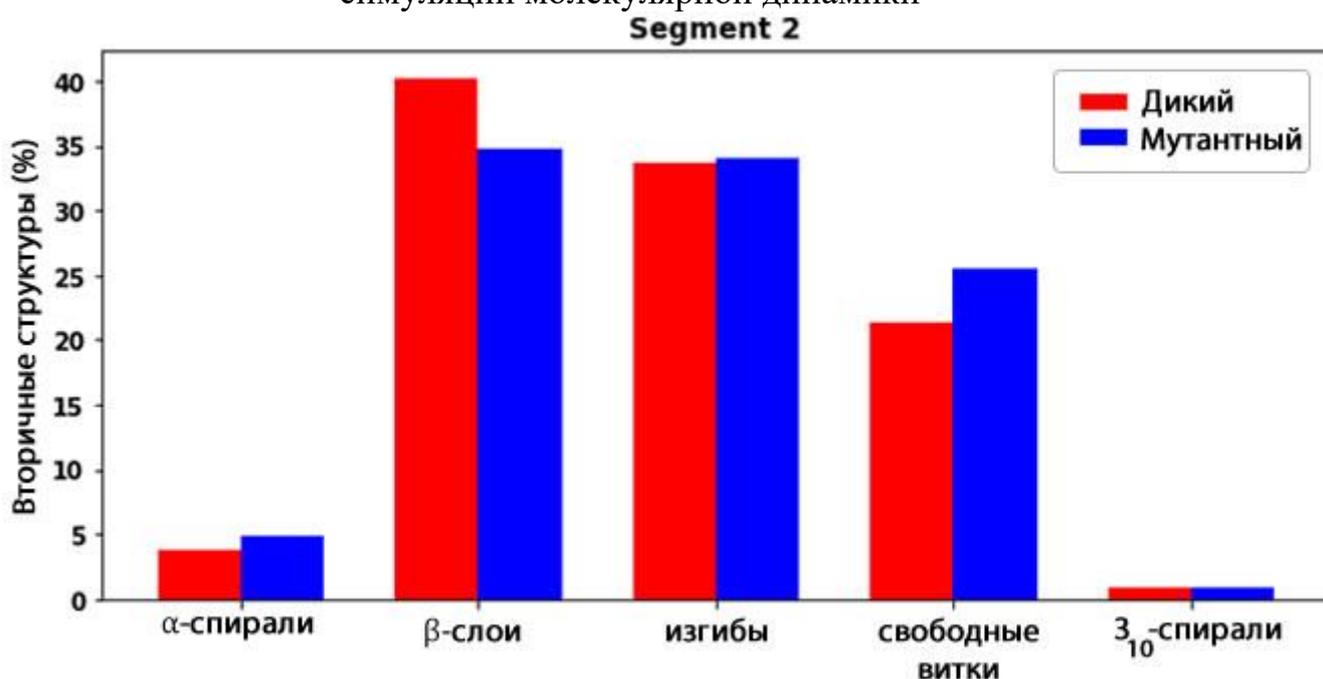


Рисунок 61 – Распределение вторичных структур в сегменте 2 после 170 нс симуляции молекулярной динамики

Элементы вторичных структур дикого и мутантного типов ADAMTS3 в сегменте 1 показано на *рисунке 62*.

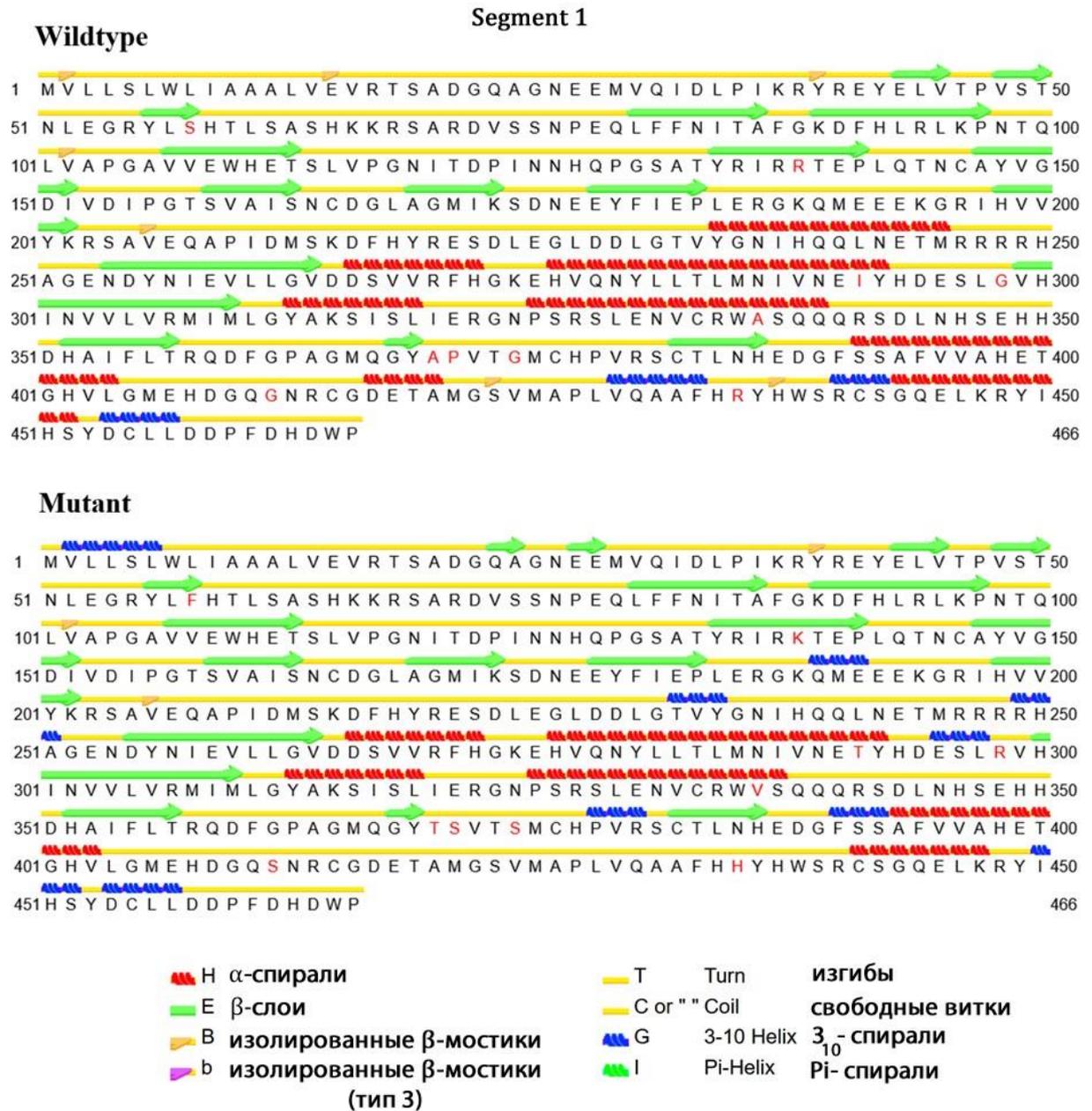
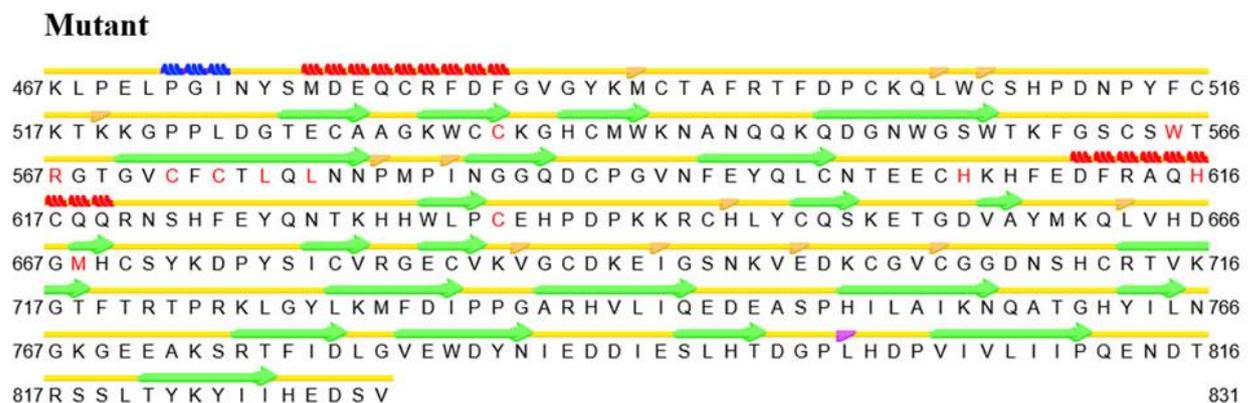
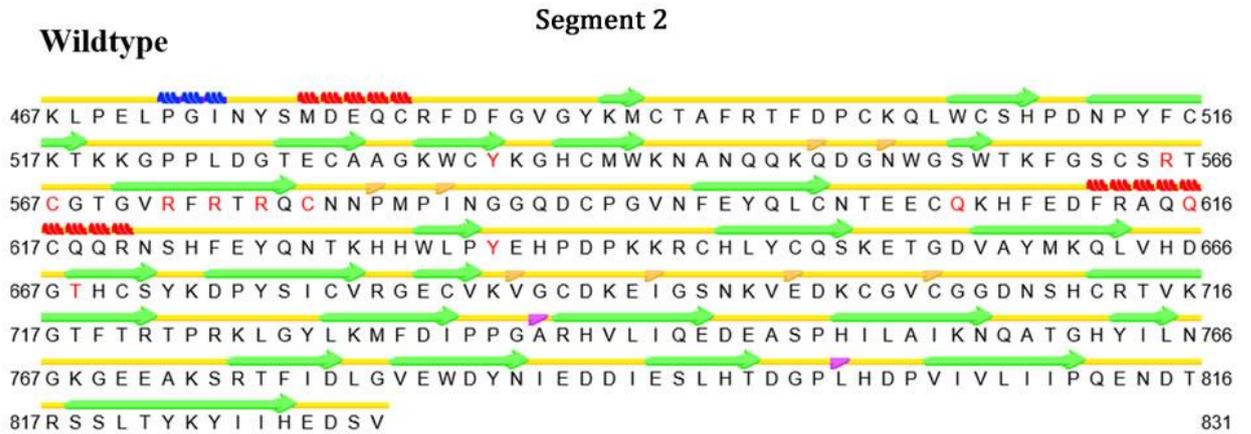


Рисунок 62 – Анализ вторичной структуры в сегменте 1 после 170 нс симуляции молекулярной динамики

*Примечание:* Представлены элементы вторичных структур дикого и мутантного типов ADAMTS3. Мутировавшие аминокислоты отмечены красным цветом.

Рассматривая распределение изменений вторичных структур в различных остатках в сегменте 2 (рисунок 63), мы видим, что в остатках Lys491-Met492, Trp506-His509 и Asn512-Thr518 нарушены β-листы, что может быть причиной увеличения RMSFs в остатках Met478-Pro523 в мутантной структуре.



H $\alpha$ -спирали	T Turn	изгибы
E $\beta$ -слои	C or " " Coil	свободные витки
B изолированные $\beta$ -мостики	G 3-10 Helix	$3_{10}$ -спирали
b изолированные $\beta$ -мостики (тип 3)	I Pi-Helix	Pi-спирали

Рисунок 63 – Анализ вторичной структуры в сегменте 1 после 170 нс симуляции молекулярной динамики

*Примечание:* Представлены элементы вторичных структур дикого и мутантного типов ADAMTS3. Мутировавшие аминокислоты отмечены красным цветом.

Мы также видим, что вблизи положения большинства исследуемых замен нет сильных изменений во вторичных структурах. Нет ни одной аминокислотной замены в остатках вышеупомянутых нарушений  $\beta$ -листов.

Также проводился анализ радиуса гирации ( $R_g$ ). Двумя наиболее важными показателями для определения структурной активности

макромолекулы являются определением  $R_g$  и расчет расстояния до центра массы молекулы. Скорость, с которой белок складывается, пропорциональна его компактности и может быть измерена с помощью сложного компьютерного метода расчета радиуса гирации (рисунок 64).

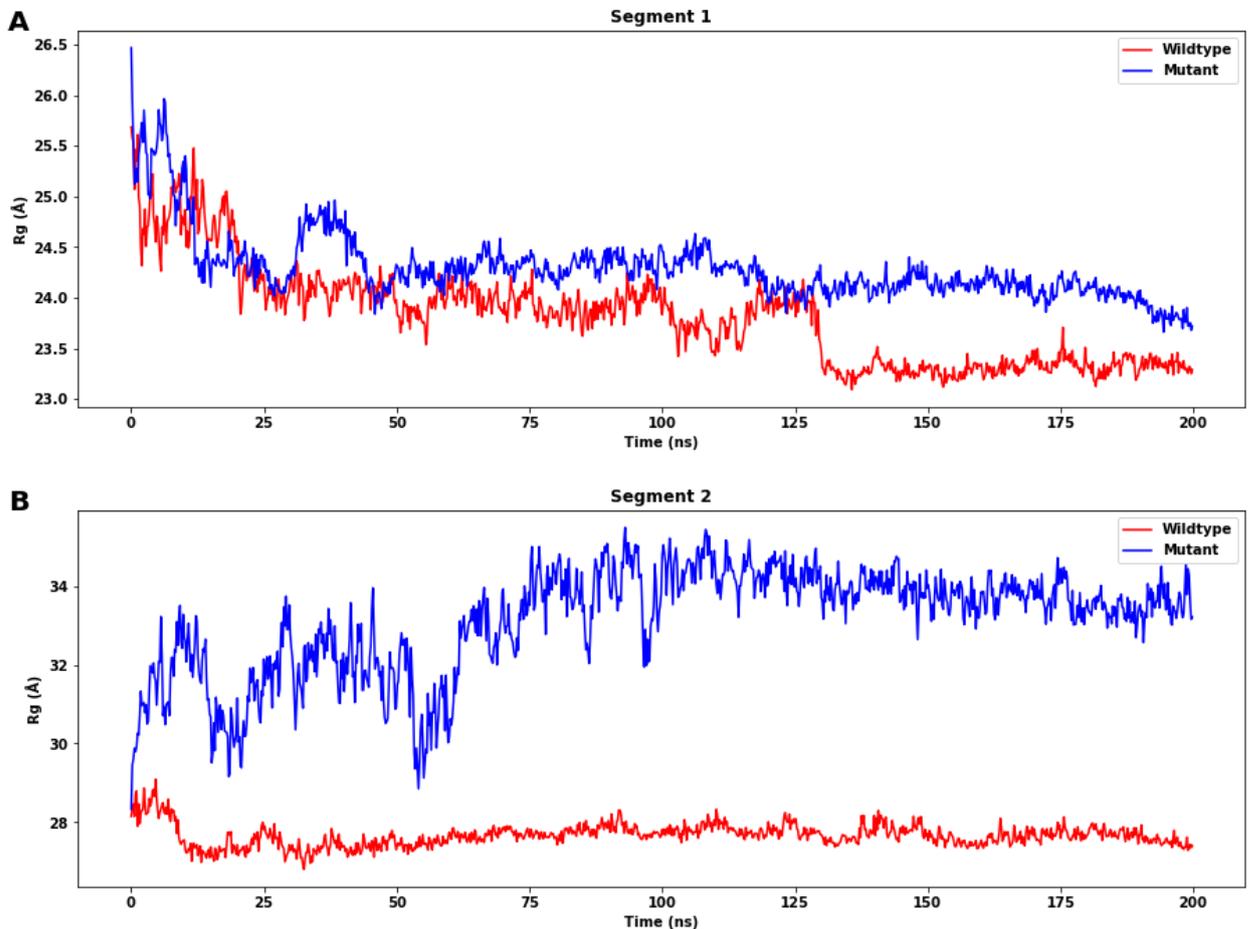


Рисунок 64 – Радиусы гирации белка дикого и мутантного типа сегментов 1 и 2  
*Примечание:* Все средние значения и SD рассчитаны со значениями после 170 нс.

Из анализа радиуса гирации структур дикого типа и мутанта ADAMTS3 можно заметить, что мутантный тип показал в целом более высокие значения  $R_g$  в течение шкалы времени моделирования по сравнению с диким типом в сегментах 1 и 2, но разница для сегмента 1 не столь значительна (дикий тип: среднее: 23.339 Å, SD: 0.082 Å; мутант: среднее: 23.984 Å, SD: 0.139 Å), как для сегмента 2 (дикий тип: среднее: 27.648 Å, SD: 0.163 Å; мутант: среднее: 33.564 Å, SD: 0.402 Å). В результате гибкость мутантного типа белка увеличивалась (рисунок 64).

Анализ площади поверхности, доступной для растворителя (SASA), показал, что мутантная структура имеет более высокое значение SASA, чем дикий тип, для сегментов 1 и 2 (рисунки 65). (Сегмент 1 дикого типа: среднее значение:  $21906.066 \text{ \AA}^2$ , SD:  $282.987 \text{ \AA}^2$ ; мутант: среднее:  $22675.036 \text{ \AA}^2$ , SD:  $453.033 \text{ \AA}^2$ ; сегмент 2 дикого типа: среднее:  $21565.973 \text{ \AA}^2$ , SD:  $245.14 \text{ \AA}^2$ ; мутант: среднее:  $22160.942 \text{ \AA}^2$ , SD:  $269.095 \text{ \AA}^2$ ).

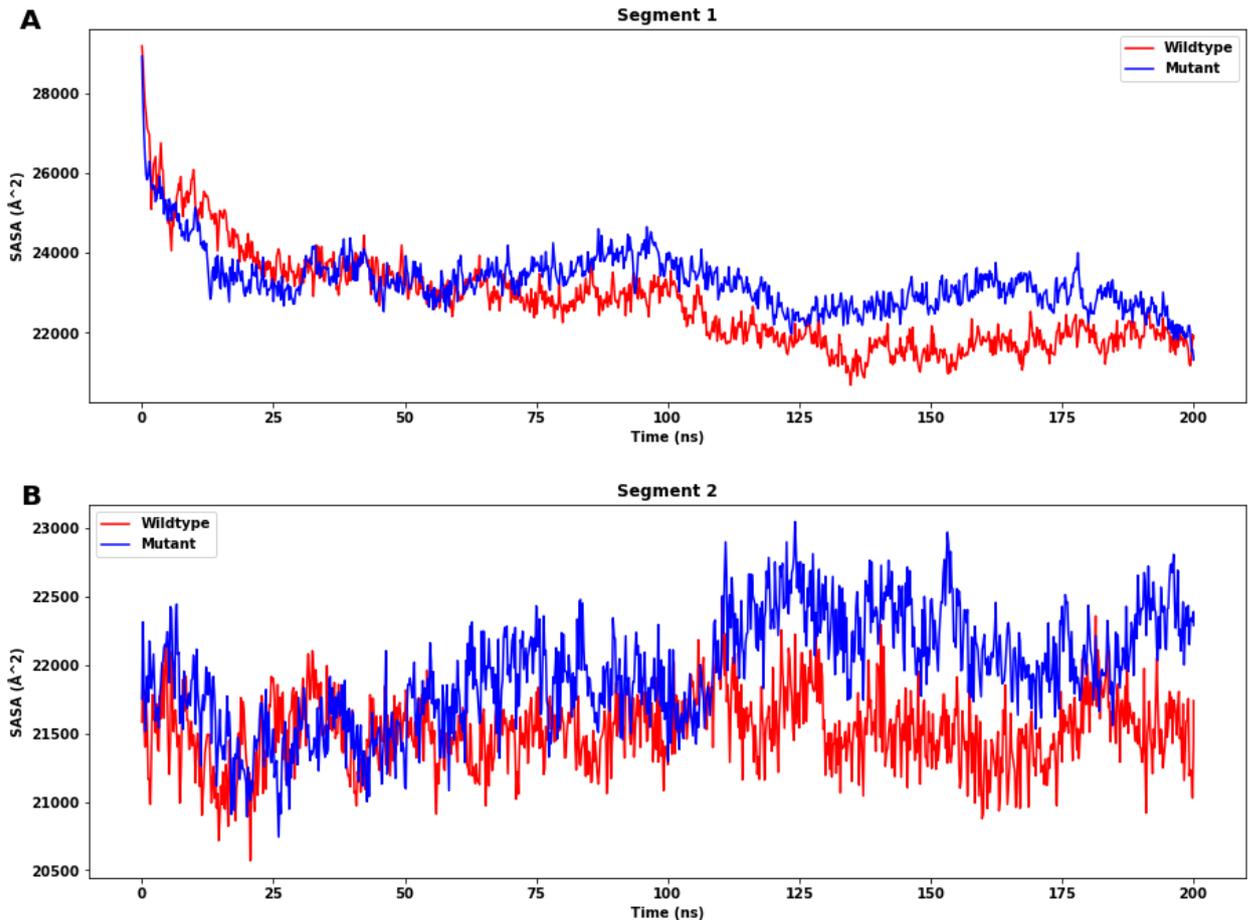


Рисунок 65 – Площадь поверхности, доступной для растворителя (SASA) (в  $\text{\AA}^2$ ), дикого и мутантного типа белка ADAMTS3 сегментов 1 (A) и 2 (B)  
*Примечание:* Для обоих сегментов SASA мутантной структуры выше, чем у дикого типа. Все средние значения и SD рассчитаны со значениями после 170 нс.

Поскольку более высокое значение SASA указывает на расширение белка, можно предположить, что дикий тип более стабилен, чем мутантный белок. Причиной более значительного изменения значения SASA может быть эффект аминокислотной замены, изменяющей размер поверхности белка, её гидрофильность и другие характеристики.

Также во время симуляции молекулярной динамики подсчитана разница между количествами водородных или Н-связей (*рисунок 66*). Для сегмента 1 она была незначительна, что еще раз указывает на то, что эффект дестабилизации для замен в этом сегменте мал (дикий тип: среднее: 396,854, SD: 7,427; мутант: среднее: 392,351, SD: 9,540). Для сегмента 2 можно отметить, что структура дикого типа образовывала большее количество Н-связей, в то время как мутантная структура демонстрировала меньшее количество Н-связей, что может повлиять на стабильность мутантного белка (дикий тип: среднее: 263,192, SD: 6,421; мутант: среднее: 255,709, SD: 9,473).

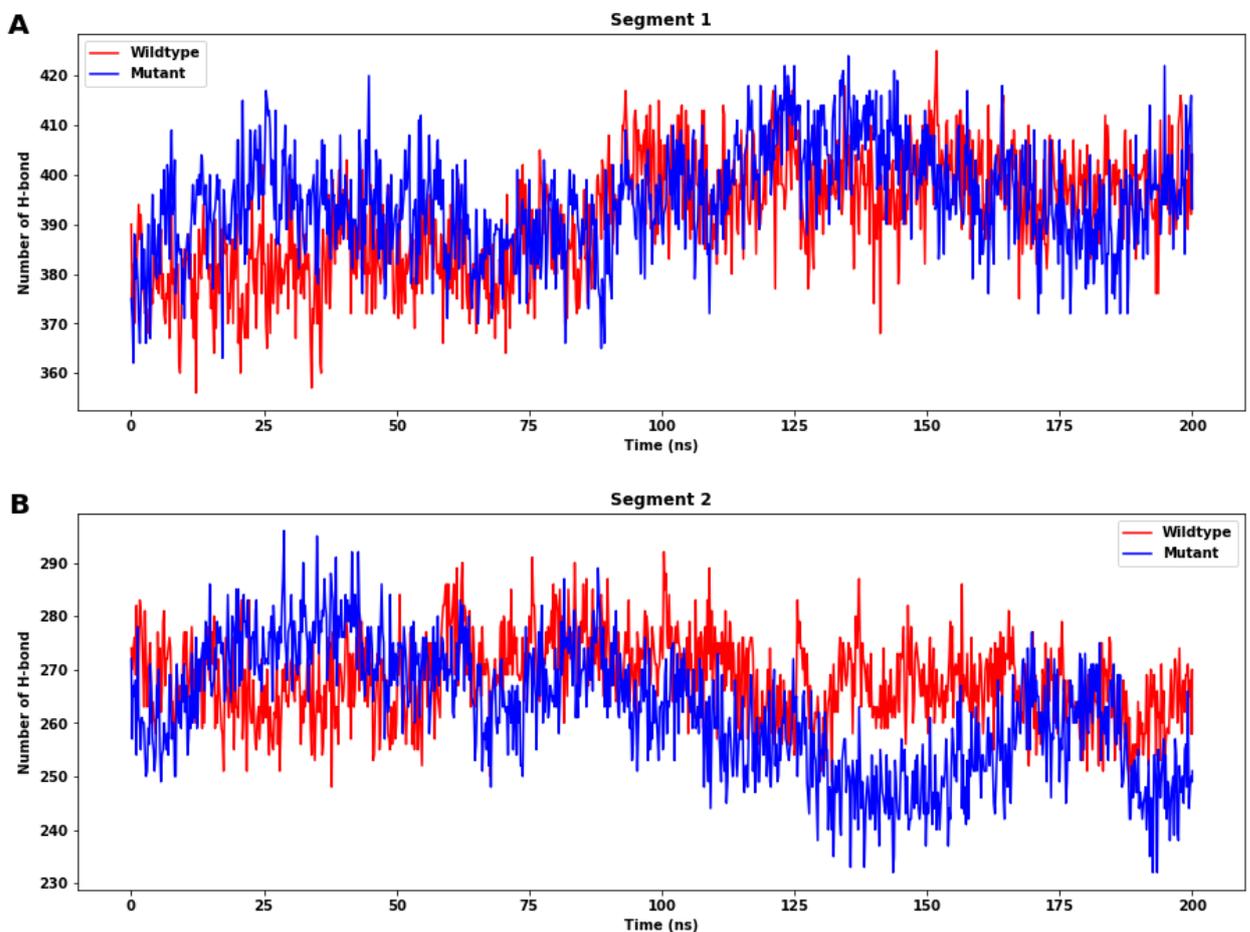


Рисунок 66 – Общее количество Н-связей на протяжении всего моделирования белков ADAMTS3 дикого и мутантного типов сегмента 1 и 2  
*Примечание:* Все средние значения и SD рассчитаны со значениями после 170 нс.

Анализ главных компонент (PCA) в данном исследовании использовался для анализа траекторий и структур белков дикого типа и

мутантных белков ADAMTS3 в сегментах 1 и 2. Графики, полученные в результате анализа PCA, показывают коллективные движения системы, спроецированные на первые две главные компоненты.

Графики сегментов 1 и 2 (рисунок 67) показывают, что существовала значительная разница в траекториях и движениях систем белков дикого типа и мутантных белков. В сегменте 1 графики дикого типа и мутантных структур в значительной степени перекрывались, что указывает на то, что мутации в этом сегменте оказывают незначительное влияние на коллективные движения белка. Однако в сегменте 2 наблюдалось меньшее перекрытие между графиками структур дикого типа и мутанта, что свидетельствует о том, что мутации в этом сегменте оказывают большее влияние на коллективные движения белка.

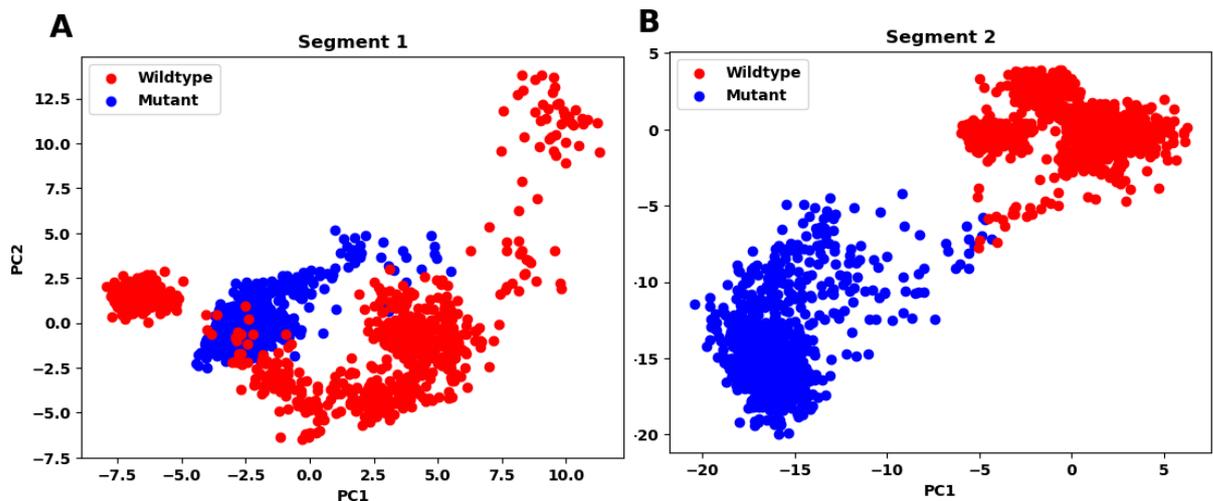


Рисунок 67 – PCA-плоты анализа траекторий сегментов 1 (А) и 2 (В)  
*Примечание:* Данные траекторий дикого типа и мутанта представлены синими и красными точками соответственно.

Графики ландшафта свободной энергии (FEL) строились после анализа PCA с использованием первых двух главных компонент. На графиках FEL (рисунок 68) конформация с наименьшей энергией обозначена темно-синим цветом. Для сегмента 1 самая низкая энергия для структуры дикого типа составляла 12,2 кДж/моль, а для мутантной структуры - 10,9 кДж/моль.

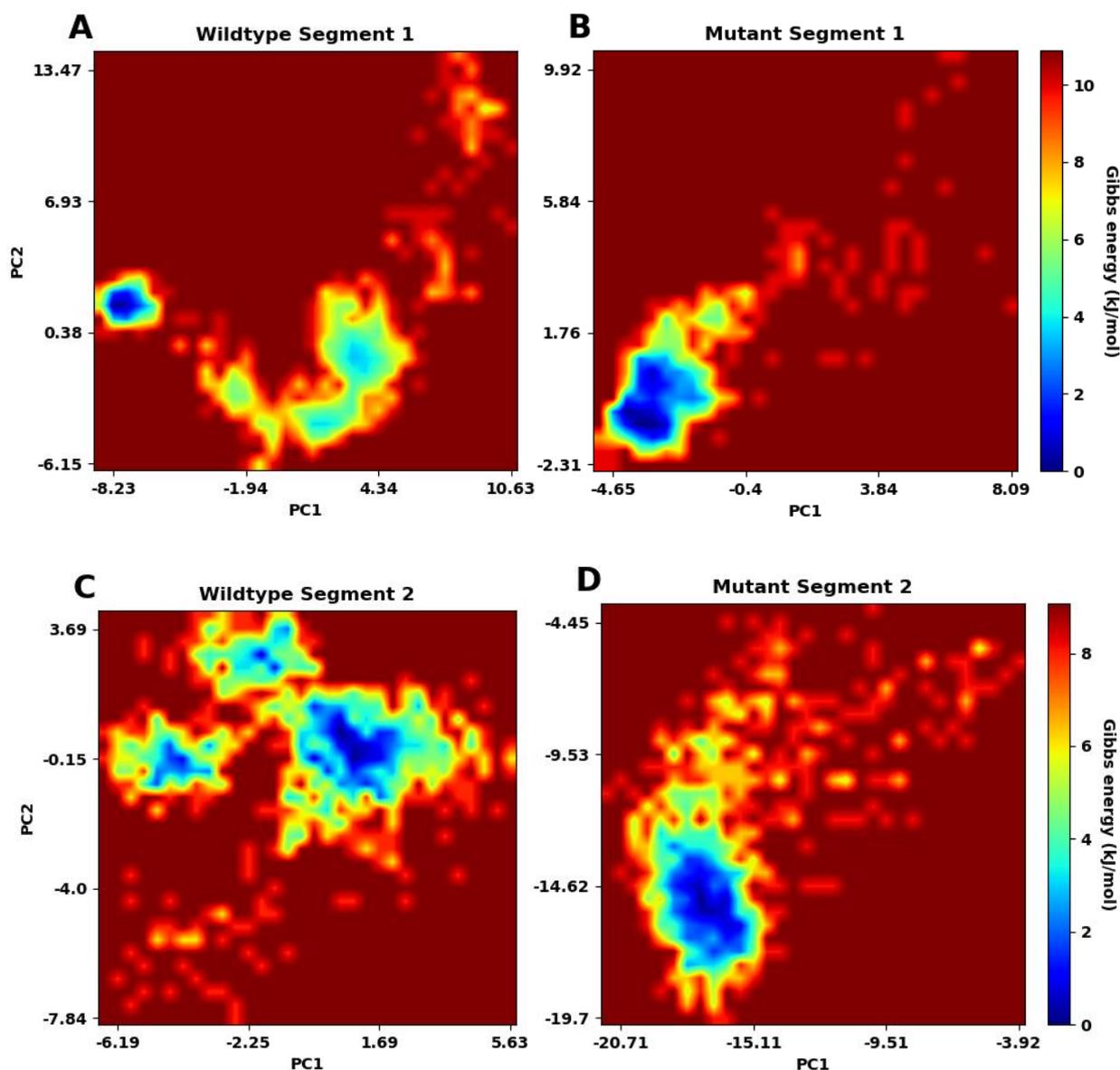


Рисунок 68 – Анализ ландшафта свободной энергии (FEL)

*Примечание:* Энергия Гиббса построена как функция первых двух главных компонент (PC1 и PC2) для сегментов 1 и 2 дикого и мутантного типов ADAMTS3. Конформация с наименьшей энергией обозначена темно-синим цветом.

Для сегмента 2 самая низкая энергия для структуры дикого типа составляла 7,80 кДж/моль, а для мутантной структуры - 9,08 кДж/моль. Для обоих сегментов структуры дикого типа и мутантные структуры в ADAMTS3 демонстрируют различия в количестве и положении стабильных конформаций, которые соответствуют локальным минимумам на графиках FEL. Это позволяет предположить, что мутации повлияли на общую конформационную стабильность белка.

## 5.8 - Молекулярно-динамическое моделирование дикого и мутантного типа FAT4

Белок FAT4 состоит из 4981 аминокислотного остатка. Это очень длинный протеин, который для моделирования необходимо разделять на несколько фрагментов. В данной работе мы создали 5 моделей: модель 1 = последовательность из 540 аминокислотных остатков (а.о.); модель 2 = 600 а.о.; модель 3 = 660 а.о.; модель 4 = 600 а.о.; и модель 5 = 420 а.о. Соответственно проводилось множество моделирований этих последовательностей.

Изменение значений среднеквадратичного отклонения (RMSD)  $C\alpha$  атомов дикого и мутантного типов FAT4 представлено на *рисунках 69-73*. График на *рисунке 69* показывает, что белок дикого типа достигал стабильности почти через 30 нс, а затем через 60 нс система сходилась и уравнивалась. Значения RMSD мутантного белка отклонялись, и после 60 нс RMSD постоянно увеличивалось до конца моделирования. Следовательно, белок модели 1 отклонялся больше по сравнению с диким на протяжении всей симуляции. Эти результаты показывают, что белок дикого типа более стабилен, чем мутантный белок для модели 1.

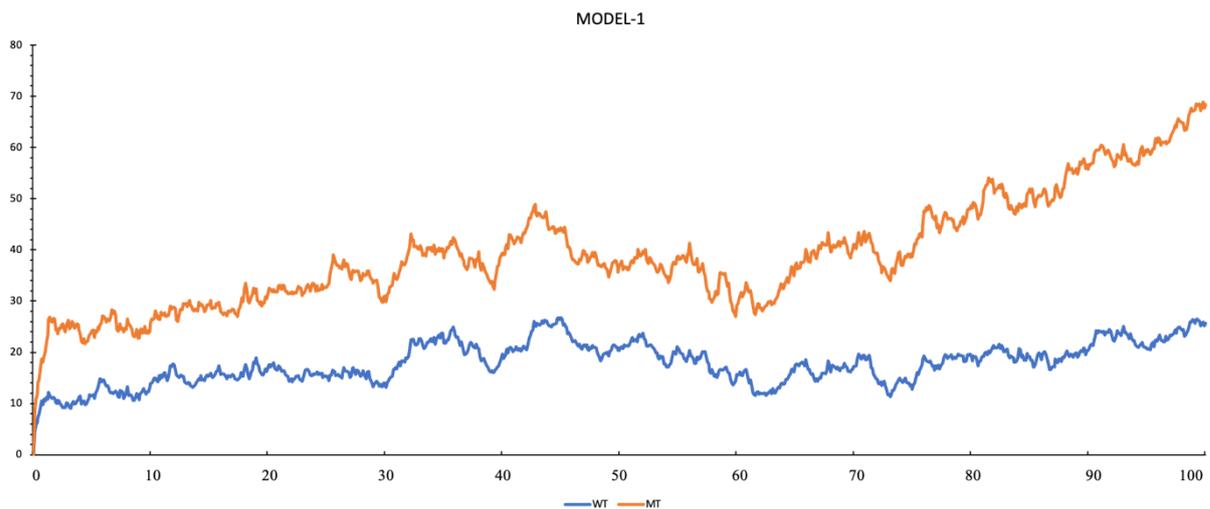


Рисунок 69 – Среднеквадратичное отклонение (RMSD)  $C\alpha$  атомов белка FAT4 дикого типа (синий) и мутантной модели 1 (красный) с течением времени

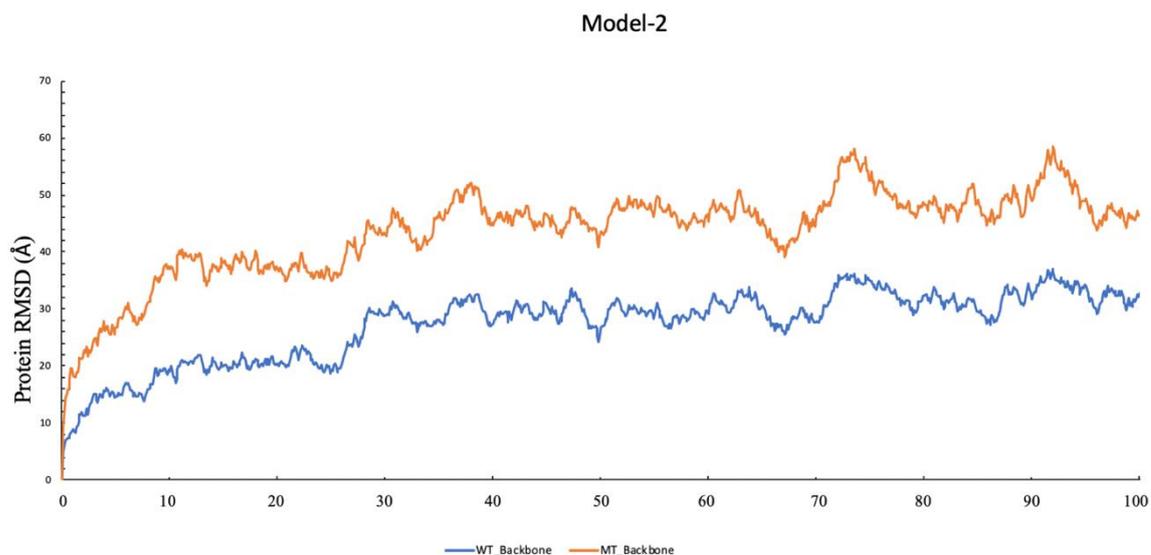


Рисунок 70 – Среднеквадратичное отклонение (RMSD)  $C\alpha$  атомов белка FAT4 дикого типа (синий) и мутантной модели 2 (красный) с течением времени

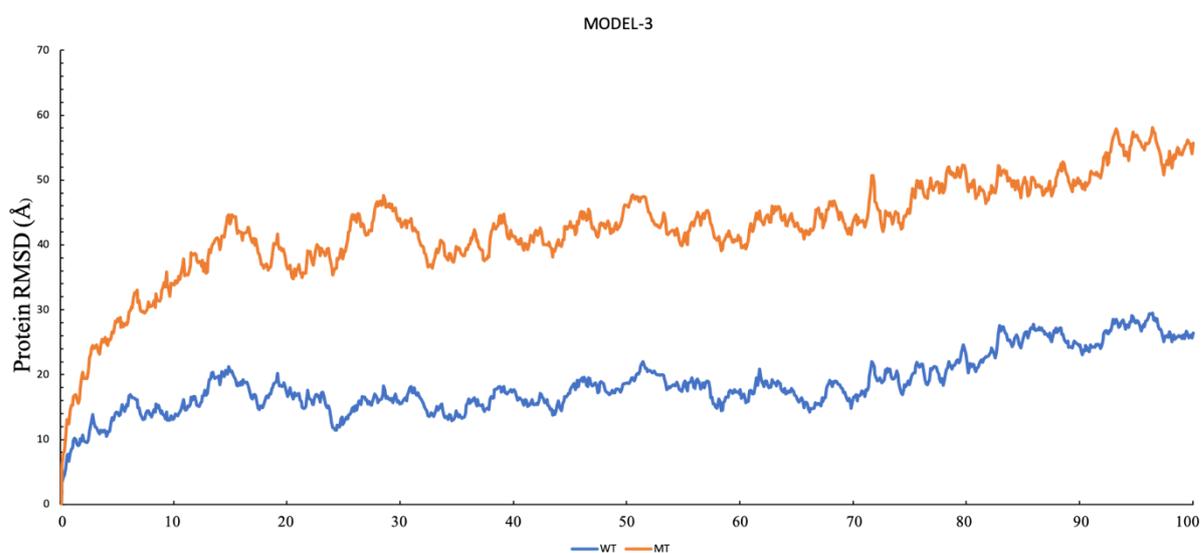


Рисунок 71 – Среднеквадратичное отклонение (RMSD)  $C\alpha$  атомов белка FAT4 дикого типа (синий) и мутантной модели 3 (красный) с течением времени

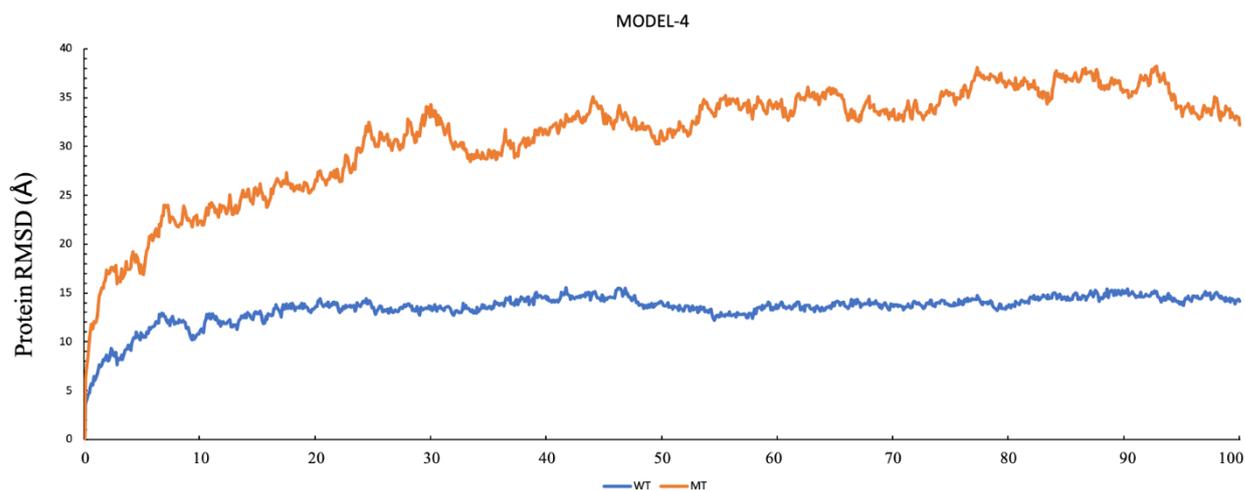


Рисунок 72 – Среднеквадратичное отклонение (RMSD)  $\text{C}\alpha$  атомов белка FAT4 дикого типа (синий) и мутантной модели 4 (красный) с течением времени

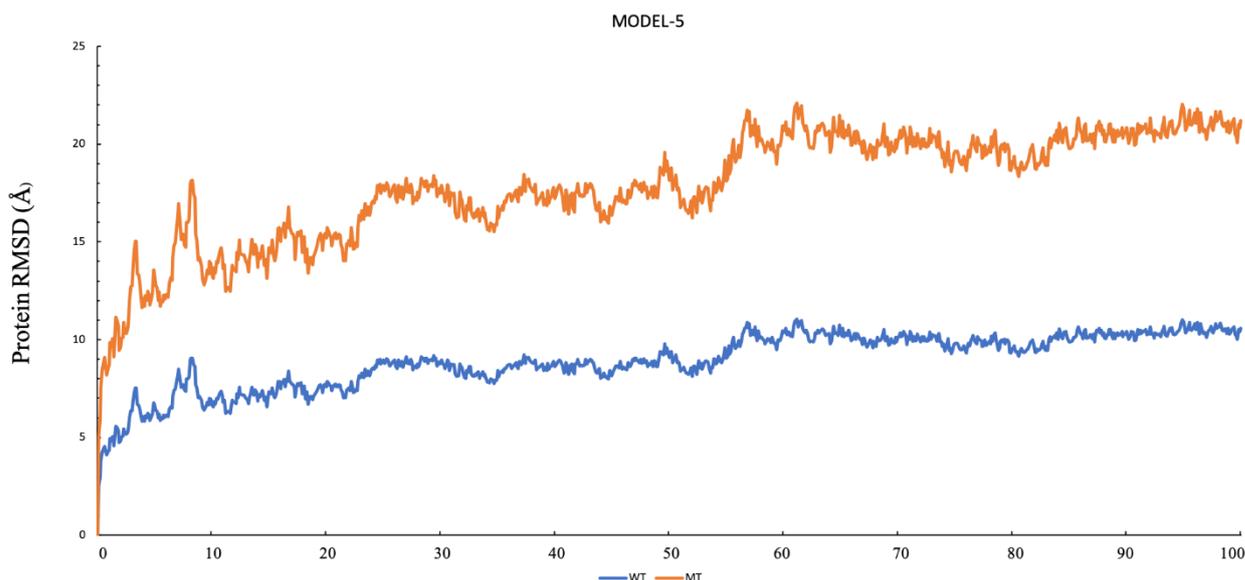


Рисунок 73 – Среднеквадратичное отклонение (RMSD)  $\text{C}\alpha$  атомов белка FAT4 дикого типа (синий) и мутантной модели 5 (красный) с течением времени

Аналогично для модели-2 после 28 нс в белке дикого типа наблюдалось меньше отклонений, и система была уравновешена на протяжении всего времени моделирования. Но для мутантного варианта почти через 30 нс наблюдалось увеличение RMSD и большее отклонение до 100 нс (рисунок 70).

На рисунке 71 представлены RMSD с учетом модели-3. Для белка дикого типа наблюдалось небольшое отклонение почти на 80 нс, после чего моделирование сходилось, в то время как для белка мутантного типа RMSD

постоянно увеличивался после 25 нс. Это указывает на большую стабильность белка дикого типа по сравнению с мутантной модели 3.

Аналогично, для модели-4 и модели 5, что указывает на стабильность дикого типа FAT4 относительно мутантных моделей.

Области белков, которые наиболее сильно колебались во время моделирования, показаны пиками на графике RMSF (рисунки 74-78). Хвосты белков (как N-, так и C-концевые) изменяются чаще, чем другие области белка. Альфа-спирали и бета-слои, например, обычно более жесткие и менее флуктуирующие, чем неструктурированная составляющая белка. Остатки с более высокими пиками, согласно траекториям, относятся к петлевым областям или N- и C-концевым зонам. RMSF показывает, что в мутантных моделях больше флуктуаций по сравнению с белком дикого типа FAT4.

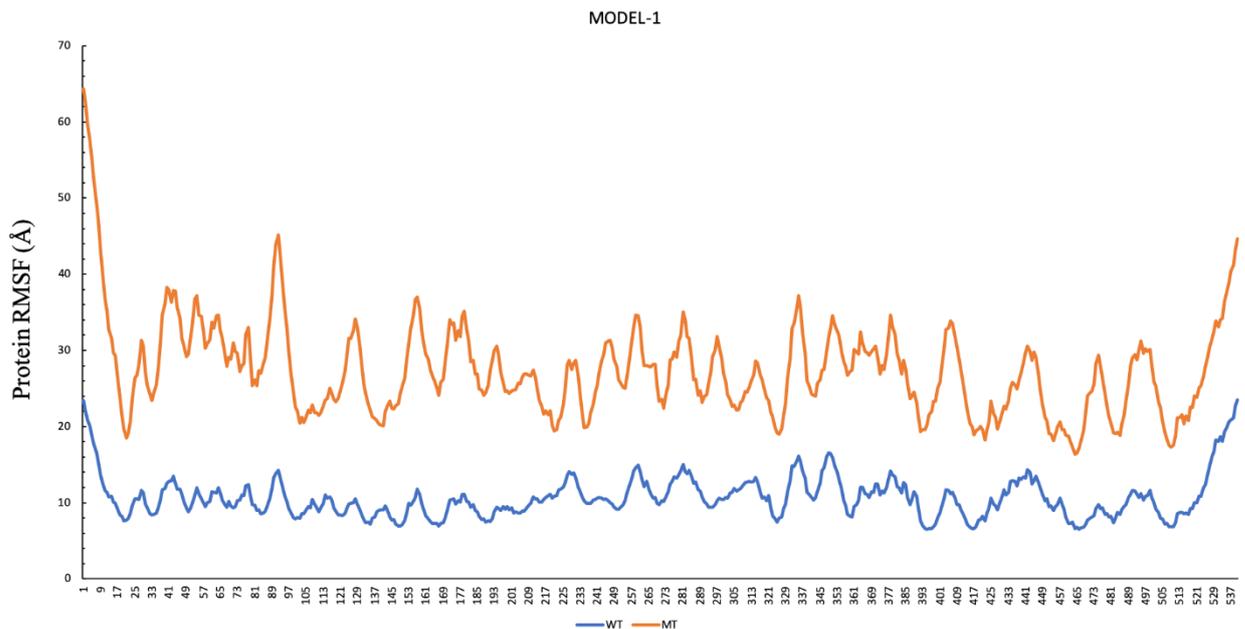


Рисунок 74 – Среднеквадратичная флуктуация (RMSF)  $\text{C}\alpha$  атомов белка FAT4 дикого типа (синий) и мутантной модели 1 (красный) с течением времени

Model-2

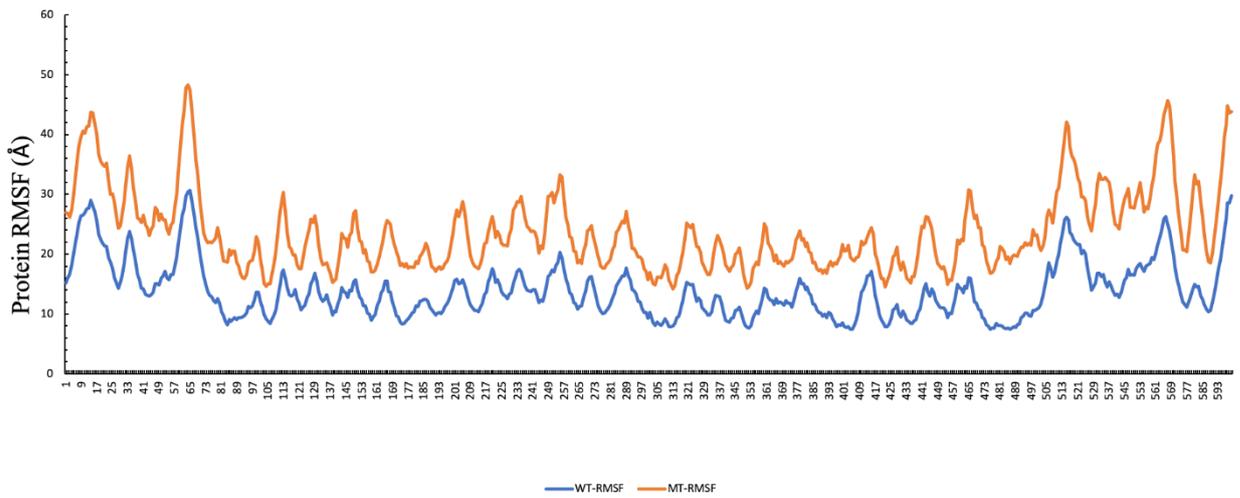


Рисунок 75 – Среднеквадратичная флуктуация (RMSF)  $\alpha$  атомов белка FAT4 дикого типа (синий) и мутантной модели 2 (красный) с течением времени

MODEL-3

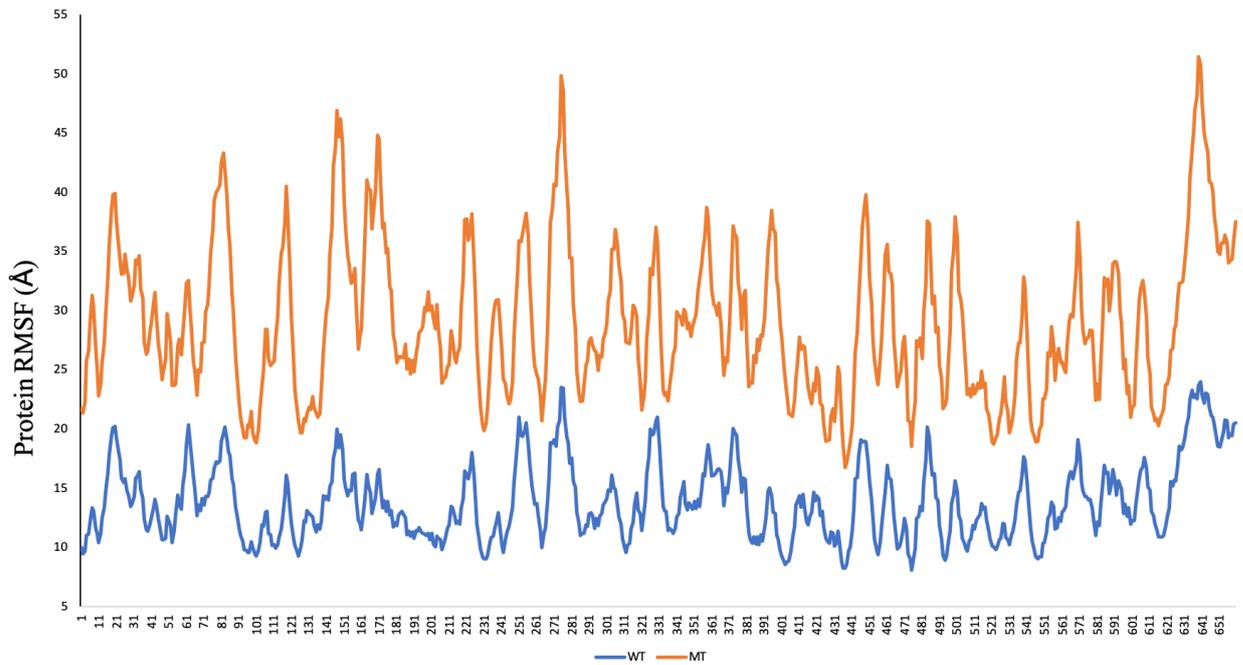


Рисунок 76 – Среднеквадратичная флуктуация (RMSF)  $\alpha$  атомов белка FAT4 дикого типа (синий) и мутантной модели 3 (красный) с течением времени

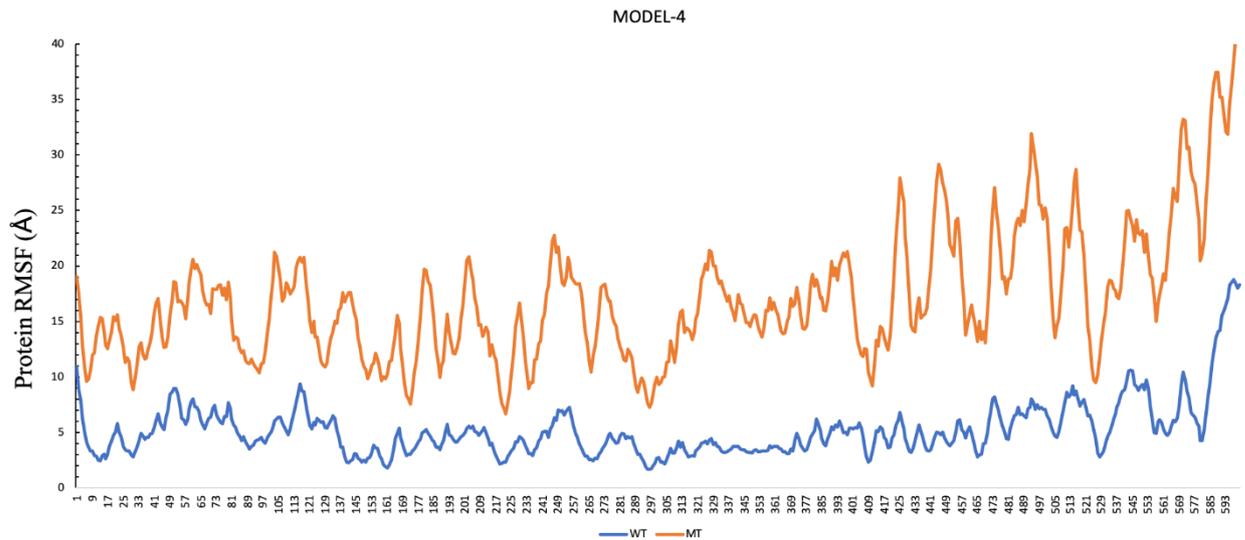


Рисунок 77 – Среднеквадратичная флуктуация (RMSF)  $\text{C}\alpha$  атомов белка FAT4 дикого типа (синий) и мутантной модели 4 (красный) с течением времени

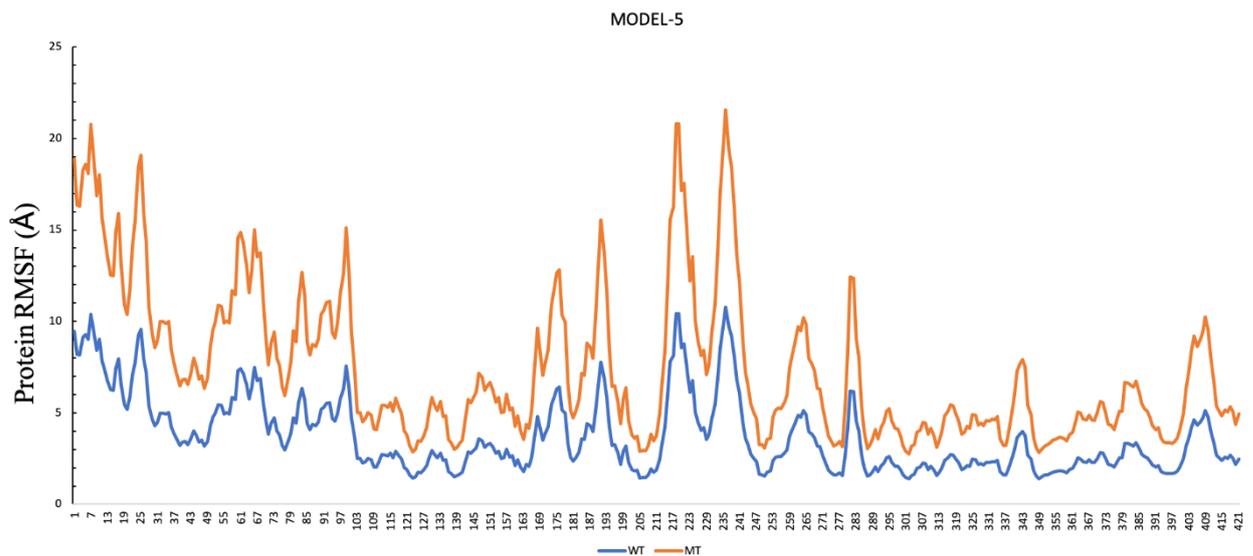


Рисунок 78 – Среднеквадратичная флуктуация (RMSF)  $\text{C}\alpha$  атомов белка FAT4 дикого типа (синий) и мутантной модели 5 (красный) с течением времени

Также рассчитывались средние распределения элементов вторичной структуры белка во время моделирования через 170 нс. Во время моделирования альфа-спирали (оранжевые) и бета-слои (голубые) являются отслеживаемыми элементами среди остальных элементов вторичной структуры (SSE). Белок дикого типа при моделировании с фрагментом 1 (модель-1) показал 0,83 % спиралей и 35,49 % бета-нитей наряду с 36,33 %

SSE, тогда как мутантный белок показал 1,33 % спиралей и 36,05 % бета-нитей наряду с 37,38 % SSE. Белок дикого типа для модели-2 показал 0,37 % спиралей и 35,25 % бета-нитей наряду с 35,62 % SSE, тогда как мутантный белок показал 1,63 % спиралей и 34,43 % бета-нитей наряду с 36,06 % SSE. Аналогичные исследования были проведены для других моделей и эти данные, как и графическое представление результатов, опубликованы в статье по данной теме. Для дальнейшего анализа принципиального значения данные результаты не имели, но учитывались.

Из анализа радиуса гирации ( $R_g$ ) структур дикого типа и мутантных моделей можно заметить, что мутантные варианты показали более высокое значение  $R_g$  на шкале времени моделирования по сравнению с диким типом. В результате гибкость мутанта относительно увеличивалась (рисунк 79). Аналогичные исследования были проведены для других моделей и эти данные, как и графическое представление результатов, опубликованы в статье по данной теме.

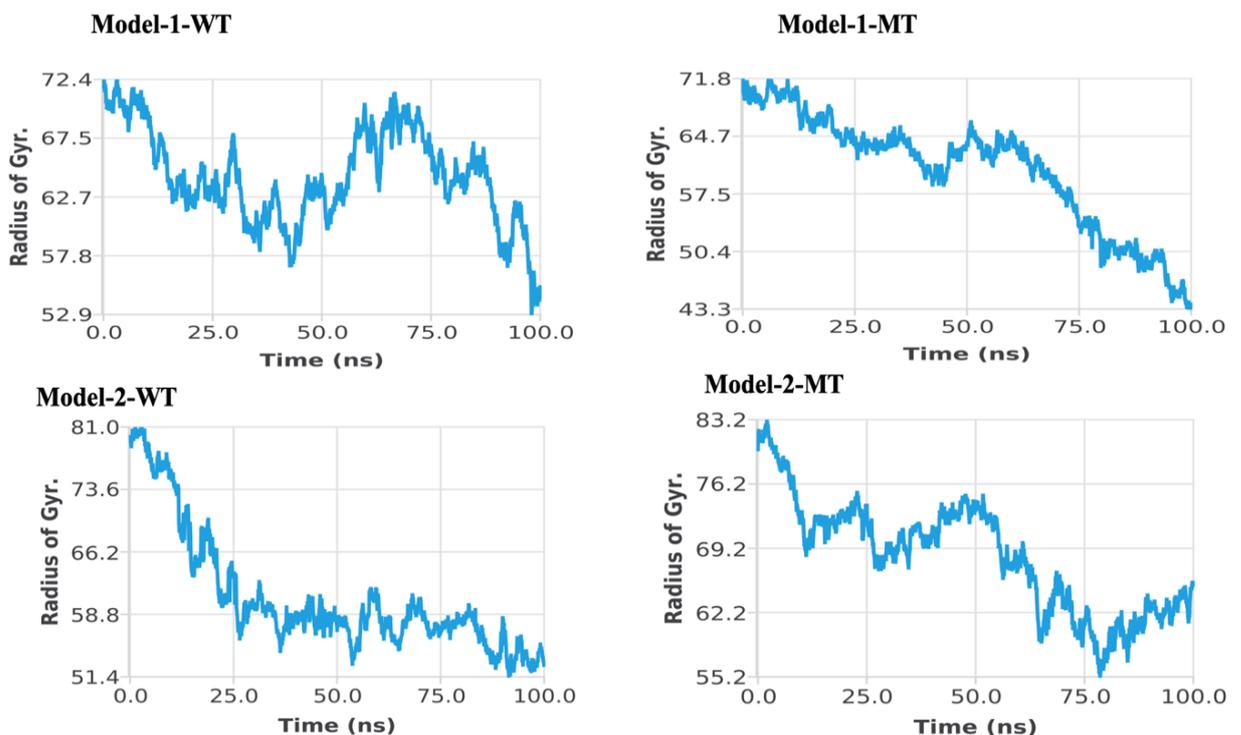


Рисунок 79 – Анализ радиуса гирации во времени при моделировании дикого (слева) и мутантного (справа) вариантов FAT4 моделей 1 (сверху) и 2 (снизу)

Анализ энергетических параметров для мутантных моделей и дикого типа показал, что общая энергия мутантной системы была увеличена по сравнению с диким типом, но не равнозначно для разных моделей. Так, наибольшие отличия наблюдались для модели 1, 2, 3, а наименьшие для 4 и 5.

### 5.9 - Оценка влияния на структуру и функцию белка FAT4 несинонимичных однонуклеотидных замен, выявленных у пациента с фенотипом синдрома Хеннекама

Мы провели идентификацию влияния миссенс-мутаций гена *FAT4*, а также миссенс-мутации p.A807V, p.G3526D и p.S3875N, которые получили через VCF из данных полного генома пациента с подозрением на синдром Хеннекама (таблица 20). С помощью многих алгоритмов и инструментов данные замены были оценены как вредоносные. Благодаря включению данных аминокислотных замен в модели при симуляции молекулярной динамики белка FAT4 была проверена гипотеза, что эти мутации нарушают структуру и функцию белка.

Таблица 20 – Обзор nsSNP в гене *FAT4*, полученных в результате анализа полногеномного секвенса пациента с фенотипом синдрома Хеннекама, включая сайт мутации, предсказания патогенности и информацию о распространенности в популяции

ген	nsSNP	генотип	Замена в белке	Protein mutation	Частота в популяции (1000g/gnomAD/ExAC)	Предсказание патогенности (FATHMM/PROVEAN/CADD)
FAT4	rs1039808	гетерози гота	c.C2420T	p.A807V	0.424/0.461/ 0.422	D/N/17.22
FAT4	rs1567047	гетерози гота	c.G10577A	p.G3526D	0.231/0.225/ 0.267	D/D/29.7
FAT4	rs12650153	гомозиг ота	c.G11624A	p.S3875N	0.009/0.007/ 0.002	T/N/20.1

Таким образом, в этой главе мы представили результаты исследований nsSNP в генах *CCBE1*, *ADAMTS3* и *FAT4*, часть которых уже была опубликована другими исследователями, а другие были вновь выявлены нами, а также обнаружены в данных полногеномного секвенса пациента с фенотипом синдрома Хеннекама. Мы провели проверку этих мутаций с помощью различных инструментов *in silico* для выявления среди них вредоносных nsSNP, подтвердили влияние отобранных потенциально вредоносных мутаций на структуру и функцию исследуемых белков, а затем провели моделирование белков, их анализ и симуляцию молекулярной динамики моделей белков дикого и мутантного типов, определив какие эффекты могут оказать выявленные аминокислотные замены на белки *CCBE1*, *ADAMTS3* и *FAT4*.

#### **Список публикаций по материалам данной главы**

1. Predicting the Most Deleterious Missense Nonsynonymous Single-Nucleotide Polymorphisms of Hennekam Syndrome-Causing *CCBE1* Gene, In Silico Analysis / K. Shinwari, L. Guojun, S.S. Deryabina, M.A. Bolkov, I.A. Tuzankina, V.A. Chereshevnev // *The Scientific World Journal*. – 2021. – V. 2021. – 6642626. (Scopus Q3).
2. Novel high-risk missense mutations identification in *FAT4* gene causing Hennekam syndrome and Van Maldergem syndrome 2 through molecular dynamics simulation / K. Shinwari, H.M. Rehman, N. Xiao, L. Guojun, M.A. Khan, M.A. Bolkov, I.A. Tuzankina, V.A. Chereshevnev // *Informatics in Medicine Unlocked*. – 2023. – V. 37. – 101160. (Scopus Q2).

## ЗАКЛЮЧЕНИЕ

С совершенствованием технологий и накоплением новых знаний происходит постоянное улучшение значимых аномалий развития, затрагивающих человеческую популяцию. Тем не менее, многое еще остается неизвестным и ждёт своего открытия. Прогноз любого отклонения в живой системе, которое может привести к болезни/синдромам/смерти, зависит от множества переменных. Исследование наследственных факторов открывает перед врачами и биологами значительные возможности для диагностики, прогнозирования и обеспечения персонифицированного лечения, однако, прежде всего, необходимо оценить влияние каждой отдельной замены в геноме на структуру и функцию белка, а также оценить её влияние на фенотип в целом.

Представленные в работе результаты были достигнуты благодаря развитию и подключению к процессу анализа современных технологий оценки влияния несинонимичных однонуклеотидных замен в гене на кодируемый белок. Наряду с другими методами оценки, интерпретация крупномасштабных и сложных многомерных данных системной биологии, это может стать очередным инструментом для регулярных исследований различной патологии человека, в том числе иммунопатологии.

Чтобы понять механизмы повышенной восприимчивости пациентов с дефицитом *RBCK1* к инфекциям, а также их взаимосвязь с амилопектинозом, присутствующими в фенотипе болезни, мы провели биоинформационное исследование ранее описанного случая дефицита *RBCK1*, сопровождающегося аутовоспалением и инфекционным синдромом, отличающим данную болезнь от других аутовоспалительных синдромов этой группы.

Выявленные в ходе нашего исследований гены с повышенной и пониженной экспрессией при аутовоспалительном синдроме дефицита *RBCK1* (также известный как дефицит *HOIL1*), позволили нам выявить ключевые сигнальные пути, задействованные при развитии данного заболевания (сигнальный путь, задействованный при инфекции *Staphylococcus aureus*, *Vibrio*

cholerae, лейшманиозе, в передаче внутриклеточных сигналов, процессинге и презентации антигена, NK-опосредованной цитотоксичности и другие).

Данные сигнальные пути и соответствующие белки прямо или косвенно отражают глубокую вовлеченность молекулярных процессов, связанных с иммунной системой, в патогенез заболевания. Кроме того, составлено общее представление о задействованных механизмах иммунитета.

В частности, было показано изменение активности генов сигнальных путей mTOR, PI3K/AKT, Rho и Nf-kB, влияющих на экспрессию генов иммунной системы, апоптоз клеток и чувствительность к ключевому цитокину иммунного ответа IL-1 $\beta$ .

Кроме того, существенное влияние на восприимчивость клеток к ЭПР-стрессу, апоптоз и клеточную гибель оказывает ген *CSID2*, экспрессия которого при дефиците *RBSK1* была значительно снижена в сравнении с экспрессией в мононуклеарах периферической крови здоровых детей ( $p=0.000000000000000007537936$ ).

Также нами было выявлено, что отличия в экспрессии генов, связанных с вирусными инфекциями, в том числе в сигнальном пути, вовлеченным в инфекцию SARS-CoV-2, являются незначительными. Это также подтверждается описанными клиническими наблюдениями в публикациях других исследователей, где особая восприимчивость у пациентов выявляется именно относительно бактериальных агентов [129].

В целом можно заключить, что повышенная восприимчивость к пиогенным инфекциям осложняется общими нарушениями убиквитирования белков, обширной глюкофажией с истощением гликогена при накоплении полисахаридов, а также выявленными отличиями в экспрессии генов, и, вероятнее всего, продукции различных белков иммунного ответа.

При проведенном нами исследовании плотности сети и так называемого биологического расстояния для генов, связанных с первичными иммунодефицитами (ПИД) и врожденной нейтропенией, в частности, было обнаружено, что гены врожденной нейтропении функционально схожи друг с

другом и тесно взаимодействуют по сравнению с другими генами ПИД. Используя эти данные, а также выявив гены, экспрессия которых значительно отличается от нормы при тяжелой врожденной нейтропении, и объединив полученную информацию с данными о белок-белковом взаимодействии, а также данные, характеризующих функции генов («биологическая дистанция» и «сетевая плотность»), нам удалось предсказать причиннозначимые гены для развития врожденной нейтропении, которые не описывались ранее в классификации первичных иммунодефицитов в этой роли.

В нашем исследовании было выявлено 15 новых генов-кандидатов для развития врожденной нейтропении, которые являются взаимозависимыми с уже известными генами врожденной нейтропении, и задействованные в тех же биологических путях, что демонстрирует высокую биологическую значимость корреляции с известными генами врожденной нейтропении. Подтверждение нескольких предсказанных генов и их влияния на функции нейтрофилов в недавно появившихся исследованиях у пациентов с дефектами нейтрофилов, убедительно демонстрируют значимость предсказанных генов-кандидатов для развития патологии.

Кроме того, при исследовании врожденной нейтропении был проведен анализ патогенности однонуклеотидных замен в генах *ELANE* и *TCIRG1*.

С помощью нескольких вычислительных инструментов мы определили в гене эластазы нейтрофилов (*ELANE*) 8 несинонимичных однонуклеотидных замен (rs28931611, rs57246956, rs137854448, rs193141883, rs201723157, rs201139487, rs137854451 и rs200384291), которые являются наиболее разрушительными для структуры и функции белка. Варианты с заменами F218L, R34W, G203S, R193W и T175M еще не были выявлены у пациентов, страдающих тяжелой врожденной нейтропенией, тогда как как C71Y, P139R, C151Y, G214R и G203C, о которых сообщалось в нашем исследовании, уже ассоциированы с обоими расстройствами. Показано, что эти мутации дестабилизируют структуру и нарушают активацию, сплайсинг и фолдинг

белка ELANE, а также могут снижать эффективность трипсиноподобной сериновой протеазы.

Дефект гена *TCIRG1* последнее время рассматривается различными учеными не только как причина наследственного остеопетроза (агрессивный остеопороз и повышенный риск переломов), но и как причина врожденной нейтропении. Имеющиеся в нашем распоряжении результаты полногеномного секвенирования пациента с врожденной нейтропенией позволили заподозрить вариант гена *TCIRG1*, как причину, при том что другие наследственные причины врожденной нейтропении специалистами ранее выявлены не были, а пациент получает специфическое лечение и наблюдается у иммунолога с этим диагнозом.

Для оценки патогенности найденной несинонимичной однонуклеотидной замены была проведена симуляция молекулярной динамики белка *TCIRG1* с учетом данной аминокислотной замены (V52L). И тесты показали, что итоговый измененный белок менее стабилен, то есть у него выше вероятность неадекватного функционирования в организме пациента.

Для убедительного доказательства необходимо провести проверку данного белка у пациента и его родителей, а также исключить другие причины, которые могут быть открыты в будущем. Тем не менее, метод исследования *in silico* позволил повысить значимость такой замены в белке и представить её, как замену, требующую особого внимания.

Кроме мутации *TCIRG1* V52L нами были протестированы другие одиночные аминокислотные изменения в высоко консервативных областях белка *TCIRG1* и всего было выявлено 15 nsSNP (rs199902030, rs200149541, rs372499913, rs267605221, rs374941368, rs375717418, rs80008675, rs149792489, rs116675104, rs121908250, rs121908251, rs121908251, rs149792489 и rs116675104), которые, вероятно, являются патогенными вариантами генов, поскольку, встраиваясь в белок «дикого» типа, они дестабилизируют его структуру и функцию. Некоторые из них расположены в консервативном

домене V-АТФазы I. Эти варианты еще не были выявлены у пациентов, страдающих врожденной нейтропенией и/или остеопетрозом, в то время как варианты G405R, R444L и D517N, о которых сообщалось в нашем исследовании, уже подтверждены другими исследователями, как варианты, ассоциированные с остеопетрозом [26, 34]. Результаты исследования могут помочь в дальнейшем понимании широкого спектра заболеваний, связанных с активацией каталитического домена киназы TCIRG1, и помочь в разработке эффективного лечения заболеваний, связанных с изменениями данного белка.

Аналогичные методы были применены при оценке влияния несинонимичных однонуклеотидных замен на структуру и функцию белков, ответственных за развитие синдрома Хеннекама – FAT4 и ADAMTS3. Кроме этих двух генов, данное аутомно-рецессивное заболевание, в патогенезе которого лимфангиэктазия и лимфедема играют ключевую роль, связано ещё с дефектами гена *CCBE1*. Три соответствующих белка влияют на активацию первичного лимфангиогенного фактора роста VEGF-C.

Используя современные инструменты *in silico*, в данном исследовании были исследованы наиболее патогенные несинонимичные однонуклеотидные полиморфизмы генов *CCBE1*, *FAT4* и *ADAMTS3*.

Наши результаты демонстрируют, что семь nsSNP в гене *CCBE1* (rs115982879, rs149792489, rs374941368, rs121908254, rs149531418, rs121908251 и rs372499913) вероятно оказывают патогенное влияние, а четыре из них (G330E, C102S, C174R и G107D) являются патогенные с очень высокой вероятностью, при том две из них (G330E и G107D) никогда не встречались в контексте синдрома Хеннекама.

Кроме того, были оценены две важные замены в гене *CCBE1* (rs374941368 и rs200149541), которые могут оказывать влияние на посттрансляционные изменения, поскольку они влияют на возможный участок фосфорилирования. Веб-сервис анализа связки лигандов FTSite участие этих замен в функции молекулы, то есть эти две замены потенциально

высокодеструктивны и их нужно учитывать при диагностике синдрома Хеннекама.

При анализе вариантов гена *ADAMTS3* из базы dbSNP предварительно было отсортировано 919 nsSNP, из которых 5 замен (G298R, C567Y, A370T, C567R и G374S) были предсказаны как наиболее опасные, которые могут быть связаны с заболеванием. Моделирование белка показало, что его можно разделить на сегменты 1, 2 и 3, которые соединены короткими петлями. С помощью инструментов симулирования молекулярной динамики было обнаружено, что некоторые замены значительно дестабилизируют структуру белка и нарушают вторичные структуры, особенно в сегменте 2. Патогенное воздействие мутаций в сегменте 1, возможно, связано не с дестабилизацией, а с другими факторами, такими как изменение фосфорилирования, как предполагают исследования посттрансляционной модификации.

Наша работа – первое исследование полиморфизмов гена *ADAMTS3* множеством инструментов, включая симуляцию молекулярной динамики. О некоторых из предсказанных замен в белке *ADAMTS3* сообщений в библиотеке PubMed еще нет, и мы надеемся, что полученные данные будут полезными для диагностических задач и поиска методов терапии.

При анализе различных вариантов гена *FAT4* из 3 343 nsSNP, имеющихся в библиотеке NCBI с помощью различных инструментов предсказания патогенности 11 замен в белке *FAT4* (D2978G, V986D, Y1912C, R4799C, D1022G, G4786R, D2439E, E2426Q, R4643C, N1309I и Y2909H) были предсказаны, как потенциально патогенные. Кроме того, три замены в гене *FAT4* (rs12650153, rs1567047 и rs1039808) у пациента с предполагаемым синдромом Хеннекама ранее были обнаружены с помощью фильтрации вариантов-кандидатов при полногеномном секвенировании, и *in silico* исследование этих мутаций показало, что они сильно дестабилизируют структуру и функцию белка.

В данном исследовании с помощью метода симуляции молекулярной динамики (MDS) мы сфокусировались на 19 мутациях в гене *FAT4* – из 11

предсказанных в нашем исследовании *in silico*, 3 nsSNP обнаружены у пациента и 5 nsSNP уже опубликованы как вероятные причины синдрома Хеннекама и синдрома Ван Мальдергема, фенотипически отличающегося от синдрома Хеннекама.

Результаты примененного метода симуляции молекулярной динамики подтвердили более низкую стабильность «мутантного» белка по сравнению с «диким» типом. Генетические варианты, обнаруженные в этой когорте исследований, ранее не были зарегистрированы как причины синдрома Хеннекама. Стоит отметить, что в силу ограниченности ресурсов суперкомпьютера и программного обеспечения, такой длинный белок, как FAT4, который состоит из 4981 аминокислот, мог быть симулирован только фрагментарно, участками, содержащими анализируемую замену длиной менее, чем 1000 аминокислот. Тем не менее, мы надеемся, что данные результаты могут способствовать лучшему пониманию предрасположенности к заболеваниям, связанным с активацией белка FAT4 и в дальнейшем могут помочь в разработке эффективных подходов для диагностики и лечения болезней, связанных с этим геном.

В целом, стоит заметить, что молекула FAT4 обладает огромным размером и сама по себе является гибкой структурой, обеспечивающей передачу не до конца выясненных межклеточных сигналов. Возможно эта молекула обеспечивает пространственную ориентацию, поляризацию клеток, передаёт сигнал о межклеточном контакте и так далее. А в силу её длины трудно предположить, на сколько серьёзное влияние оказывает единичная аминокислотная замена, скорее должно иметь значение накопление отличий и особенно поражение функциональных активных центров биндинга молекул, нежели одна аминокислотная замена.

### **Перспективы дальнейшего развития темы**

Идентификация специфических генетических изменений и определение молекулярной основы иммунопатологии позволит исследовать

патогенетические механизмы, дифференцировать нозологические формы из обширной гетерогенной группы врожденных ошибок иммунитета, а также приблизить создание специфической таргетной терапии, включая геноредактирование и антисмысловые олигонуклеотиды. Это позволит решать вопросы радикального излечения пациентов. Кроме того, даже простое ускорение диагностического процесса поможет своевременно поставить диагноз, назначить заместительную и патогенетическую терапию, улучшить прогноз и качество жизни пациентов.

Процесс верификации генов первичного иммунодефицита может быть улучшен за счет разработки программного обеспечения для предсказания генов-кандидатов различных иммунопатологий, а включение методов предсказания влияния генетических изменений на белок *in silico* обеспечивает возможность его эффективного применения в клинических исследованиях.

Кроме того, изучение редких случаев патологии человека позволяет решать общепатологические вопросы формирования болезней, обогащая науку знаниями о закономерностях функционирования иммунной системы и всего организма человека. Погружаясь на молекулярный уровень патологии, исследователям становятся доступны объективные обоснования для разработки и применения целенаправленной терапевтической тактики, что открывает перспективу создания новых таргетных лекарственных средств.

Таким образом, проведенные нами исследования позволили нам сделать следующие выводы.

## **ВЫВОДЫ**

1. Выявлены новые генетические нарушения при трех видах врожденных ошибок иммунитета (первичных иммунодефицитах): дефиците *RBCK1*, врожденной нейтропении и синдроме Хеннекама.

2. Выявлены достоверные отличия в экспрессии генов при дефиците RBSK1 относительно данных здоровых детей и пациентов с синдромом CINCA/NOMID, синдром Макл-Уэллса и дефицита мевалонаткиназы.

3. Выявленные несинонимичные однонуклеотидные замены в гене *TCIRG1* (rs199902030, rs200149541, rs372499913, rs267605221, rs374941368, rs375717418, rs80008675, rs149792489, rs116675104, rs121908250, rs121908251, rs121908251, rs149792489, rs116675104) и *ELANE* (rs200384291, rs201163886, rs193141883, rs201139487, rs201723157) дестабилизируют белки *TCIRG1* и *ELANE* в нейтрофилах.

4. Гены *CDC42*, *CRKL*, *FGR*, *CRC*, *NYK*, *PLCG1*, *ARRB2*, *PIK3CG*, *PTK2*, *STAT1*, *STAT2*, *STAT3*, *STAT5B*, *VAV1* и *ITK* являются новыми генами-кандидатами развития врожденной нейтропении.

5. Несинонимичные однонуклеотидные замены в генах *CCBE1* (rs115982879, rs149792489, rs374941368, rs121908254, rs149531418, rs121908251 и rs372499913), *FAT4* (rs147663284, rs192514171, rs138137489, rs199895179, rs372060616, rs138173652, rs142184187, rs147633644, rs181607904, rs184971791, rs148655455) и *ADAMTS3* (rs61757480, rs61741624, rs140806973, rs140595148, rs140914273, rs142268705, rs142781084, rs143059623, rs146979323, rs372067284, rs370857003, rs375983592, rs367831484, rs202031187 и rs150012152) приводят к дестабилизации структуры белков *CCBE1*, *FAT4* и *ADAMTS3*, и могут стать причиной развития синдрома Хеннекама.

6. Разработанная программа последовательного использования методов биоинформационного анализа эффективна при поиске генов, оказывающих влияние на патогенез болезней, ассоциированных с врожденными ошибками иммунитета.

## ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ

1. Пациентам при диагностике первичных иммунодефицитов (врожденных ошибок иммунитета) необходимо определять профиль

экспрессии генов путем анализа дифференциальной экспрессии генов, сигнальных путей и генетических онтологий, а также проводить определение биомаркеров патологии, что позволит снизить затраты на лечение и предотвратить развитие побочных эффектов.

2. При проведении исследований по прогнозированию новых генов-кандидатов при врожденной нейтропении необходимо включать в анализ факторы ко-экспрессии, белок-белковые взаимодействия и сигнальные пути.

3. Для дифференциальной диагностики врожденной нейтропении в дополнение к генам, перечисленным на сайте ESID и в классификации IUIS, следует включить дополнительные гены (*CDC42*, *CRKL*, *FGR*, *CRC*, *NYK*, *PLCG1*, *ARRB2*, *PIK3CG*, *PTK2*, *STAT1*, *STAT2*, *STAT3*, *STAT5B*, *VAV1* и *ITK*), определенных в нашем исследовании в качестве генов-кандидатов.

4. Для дифференциальной диагностики синдрома Хеннекама и врожденной нейтропении, в дополнение к перечисленным миссенс-мутациям генов *ADAMTS3*, *FAT4*, *CCBE1*, *ELANE* и *TCIRG1* необходимо оценить наличие миссенс-мутаций nsSNP, выявленных в нашем исследовании для следующих генов: *CCBE1* (rs115982879, rs149792489, rs374941368, rs121908254, rs149531418, rs121908251 и rs372499913), *ELANE* (rs200384291, rs201163886, rs193141883, rs201139487 и rs201723157), *TCIRG1* (rs199902030, rs200149541, rs372499913, rs267605221, rs374941368, rs375717418, rs80008675, rs149792489, rs116675104, rs121908250, rs121908251, rs121908251, rs149792489 и rs116675104), *FAT4* (rs147663284, rs192514171, rs138137489, rs199895179, rs372060616, rs138173652, rs142184187, rs147633644, rs181607904, rs184971791, rs148655455) и *ADAMTS3* (rs61757480, rs61741624, rs140806973, rs140595148, rs140914273, rs142268705, rs142781084, rs143059623, rs146979323, rs372067284, rs370857003, rs375983592, rs367831484, rs202031187, и rs150012152).

**СПИСОК ЛИТЕРАТУРЫ**

1. Анализ уровней TREC и KREC в высушенных пятнах крови здоровых новорожденных с различным сроком беременности и весом. / Черемохин, Д.А. и др. // *Acta Naturae*, 14(1), С.101-108.
2. Воронина Л.И. Факторы влияния на отношение медицинского сообщества к генетическим исследованиям и оказанию медицинских услуг детям с врожденными ошибками иммунитета. / Воронина Л.И., Зайцева Е.В., Тузанкина И.А. // *Социология медицины*. 2019, 18(2) С.92–97.
3. Изменчивость симптомокомплекса САТСН-22 в рамках синдрома делеции 22q11.2. / Черемохин Д.А. и др. // *Медицинская иммунология*. 2021, 23(6), С.1357-1366.
4. Классификация врожденных ошибок иммунитета человека, обновленная экспертами комитета Международного союза иммунологических обществ в 2019 году / Болков М.А. и др. // *Российский иммунологический журнал*. 2021. Т. 24. №1. С. 7-68.
5. Неонатальный скрининг на тяжелую комбинированную иммунную недостаточность в России: прекрасное далеко или завтрашняя реальность? / Дерябина С.С. и др. // *Вопросы современной педиатрии*, 2017, 16(1), С. 59-66.
6. Первичные иммунодефициты (врожденные ошибки иммунитета) в раннем возрасте / И.А.Тузанкина и др.; рецензенты: А.С.Симбирцев, З.Ж.Рахманкулова; – Ташкент: издательство «Adast poligraf», 2022. – 232 с.
7. Приказ Министерства здравоохранения Российской Федерации от 21.04.2022 274н.
8. Ретроспективный анализ случаев первичных иммунодефицитов у детей с врожденными пороками сердца / Дерябина С.С. и др. // *Российский иммунологический журнал*. 2020. Т. 23. №4. С. 505-514.
9. Ретроспективная диагностика первичных иммунодефицитных состояний у детей в Свердловской области. / Дерябина С.С. и др. // *Медицинская иммунология*. 2016, 18(6), С. 583-588.
10. Роль врожденных ошибок иммунитета в группе детей с летальными исходами на первом году жизни / Черемохин Д.А. и др. // *Российский иммунологический журнал*. 2022. Т.25. №4. С. 555-560.
11. Фенотипические варианты манифестации гомозиготной делеции сегмента хромосомы 1, захватывающей участки гена CFHR3. / Тузанкина И.А. и др. // *Медицинская иммунология*. 2020, 22(3), С. 569-576
12. Эпидемиология первичных иммунодефицитов в Российской Федерации / А.А. Мухина и др. // *Педиатрия им. Г.Н. Сперанского*. 2020; 99 (2): С. 16-32.

13. A catalytic-independent role for the LUBAC in NF- $\kappa$ B activation upon antigen receptor engagement and in lymphoma cells / S.M. Dubois, [et al.]. // *Blood*. 2014. № 14 (123). C. 2199–2203.
14. Aksentijevich, I. NF- $\kappa$ B Pathway in Autoinflammatory Diseases: Dysregulation of Protein Modifications by Ubiquitin Defines a New Category of Autoinflammatory Diseases / I. Aksentijevich, Q. Zhou // *Frontiers in Immunology*. 2017. (8). C. 399.
15. Al-Mousa, H. Primary Immunodeficiency Diseases in Highly Consanguineous Populations from Middle East and North Africa: Epidemiology, Diagnosis, and Care / H. Al-Mousa, B. Al-Saud // *Frontiers in Immunology*. 2017. (8). C. 678.
16. A Next-Generation Sequencing Test for Severe Congenital Neutropenia: Utility in a Broader Clinicopathologic Spectrum of Disease / S.N. McNulty, [et al.]. // *The Journal of molecular diagnostics: JMD*. 2021. № 2 (23). C. 200–211.
17. A novel computational and structural analysis of nsSNPs in CFTR gene / C. George Priya Doss, [et al.]. // *Genomic Medicine*. 2008. № 1–2 (2). C. 23–32.
18. A partial form of recessive STAT1 deficiency in humans / A. Chappier, [et al.]. // *The Journal of Clinical Investigation*. 2009. № 6 (119). C. 1502–1514.
19. A proteome-scale map of the human interactome network / T. Rolland, [et al.]. // *Cell*. 2014. № 5 (159). C. 1212–1226.
20. As Little as Needed: The Extraordinary Case of a Mild Recessive Osteopetrosis Owing to a Novel Splicing Hypomorphic Mutation in the TCIRG1 Gene / C. Sobacchi, [et al.]. // *Journal of Bone and Mineral Research*. 2014. № 7 (29). C. 1646–1650.
21. A Systematic Review on Predisposition to Lymphoid (B and T cell) Neoplasias in Patients With Primary Immunodeficiencies and Immune Dysregulatory Disorders (Inborn Errors of Immunity) / I.B. Riaz, [et al.]. // *Frontiers in Immunology*. 2019. Vol. 10. C. 777.
22. An integrative approach to predicting the functional effects of non-coding and coding sequence variation/ Shihab H. A. et al. // *Bioinformatics*. – 2015. – T. 31. – №. 10. – C. 1536-1543.
23. Atp6i-deficient mice exhibit severe osteopetrosis due to loss of osteoclast-mediated extracellular acidification / Y.P. Li, [et al.]. // *Nature Genetics*. 1999. № 4 (23). C. 447–451.
24. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. / J. Eberhardt, [et al.]. // *Journal of chemical information and modeling*. – 2021. – T. 61. – №. 8. – C. 3891-3898

25. Autosomal recessive intestinal lymphangiectasia and lymphedema, with facial anomalies and mental retardation / R.C. Hennekam, [et al.]. // *American Journal of Medical Genetics*. 1989. № 4 (34). C. 593–600.
26. Autosomal recessive osteopetrosis type I: description of pathogenic variant of TCIRG1 gene / L.E. Chávez-Güitrón, [et al.]. // *Boletín médico del Hospital Infantil de México*. 2018. № 4 (75). C. 255–259.
27. Bach, E.A. The IFN gamma receptor: a paradigm for cytokine receptor signaling / E.A. Bach, M. Aguet, R.D. Schreiber // *Annual Review of Immunology*. 1997. (15). C. 563–591.
28. Ballgown bridges the gap between transcriptome assembly and expression analysis / A.C. Frazee, [et al.]. // *Nature Biotechnology*. 2015. № 3 (33). C. 243–246.
29. BioGRID: a general repository for interaction datasets / C. Stark, [et al.]. // *Nucleic Acids Research*. 2006. № Database issue (34). C. D535-539.
30. Bioinformatics analysis of key genes and latent pathway interactions based on the anaplastic thyroid carcinoma gene expression profile / Y. Huang, [et al.]. // *Oncology Letters*. 2017. № 1 (13). C. 167–176.
31. Blom N. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. / N. Blom, S. Gammeltoft, S. Brunak // *Journal of Molecular Biology*, 1999, № 5 (294), C. 1351–1362,
32. Bondos, S.E. Physical and genetic interactions link hox function with diverse transcription factors and cell signaling proteins / S.E. Bondos, X.-X. Tan, K.S. Matthews // *Molecular & cellular proteomics: MCP*. 2006. № 5 (5). C. 824–834.
33. Borregaard, N., Granules of the Human Neutrophilic Polymorphonuclear Leukocyte / N. Borregaard, J.B. Cowland // *Blood*. 1997. № 10 (89). C. 3503–3521.
34. Buried in the Middle but Guilty: Intronic Mutations in the TCIRG1 Gene Cause Human Autosomal Recessive Osteopetrosis / E. Palagano, [et al.]. // *Journal of Bone and Mineral Research*. 2015. № 10 (30). C. 1814–1821.

35. Capriotti, E. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure / E. Capriotti, P. Fariselli, R. Casadio // *Nucleic Acids Research*. 2005. № Web Server issue (33). C. W306-310.
36. Casanova, J.-L. Genetic dissection of immunity to mycobacteria: the human model / J.-L. Casanova, L. Abel // *Annual Review of Immunology*. 2002. Vol. 20. C. 581–620.
37. CCBE1 is required for embryonic lymphangiogenesis and venous sprouting / B.M. Hogan, [et al.]. // *Nature Genetics*. 2009. № 4 (41). C. 396–398.
38. CDG: An Online Server for Detecting Biologically Closest Disease-Causing Genes and its Application to Primary Immunodeficiency / D. Requena, [et al.]. // *Frontiers in Immunology*. 2018. Vol. 9. C. 1340.
39. Chen H. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R / Chen H., Boutros P. C. // *BMC bioinformatics*. – 2011. – T. 12. – №. 1. – C. 1-7.
40. Cheng, J. Prediction of protein stability changes for single-site mutations using support vector machines / J. Cheng, A. Randall, P. Baldi // *Proteins*. 2006. № 4 (62). C. 1125–1132.
41. Chitrala, K.N. Computational screening and molecular dynamic simulation of breast cancer associated deleterious non-synonymous single nucleotide polymorphisms in TP53 gene / K.N. Chitrala, S. Yeguvapalli // *PloS One*. 2014. № 8 (9). C. e104242.
42. Choi Y. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels / Choi Y., Chan A. P. // *Bioinformatics*. – 2015. – T. 31. – №. 16. – C. 2745-2747.
43. Clinical-Epidemiological Pattern of Primary Immunodeficiencies in Malaysia 1987-2006: A 20 year experience in Four Malaysian Hospitals / L.M. Noh, [et al.]. // *The Medical Journal of Malaysia*. 2013. № 1 (68). C. 13–17.
44. CISD2 maintains cellular homeostasis / Shen Z. Q. et al. // *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. – 2021. – T. 1868. – №. 4. – C. 118954.

45. Clinical features of dominant and recessive interferon gamma receptor 1 deficiencies / S.E. Dorman, [et al.]. // *Lancet* (London, England). 2004. № 9451 (364). C. 2113–2121.
46. Comparison of predicted and actual consequences of missense mutations / L.A. Miosge, [et al.]. // *Proceedings of the National Academy of Sciences of the United States of America*. 2015. № 37 (112). C. E5189-5198.
47. Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins. / W. Deng et al. // *Briefings in Bioinformatics*, 2017 №4 (18), C. 647–658.
48. Congenital neutropenia: diagnosis, molecular bases and patient management / J. Donadieu, [et al.]. // *Orphanet Journal of Rare Diseases*. 2011. № 6. C. 26.
49. Congenital neutropenia in the era of genomics: classification, diagnosis, and natural history / J. Donadieu, [et al.]. // *British Journal of Haematology*. 2017. № 4 (179). C. 557–574.
50. Congenital neutropenia with variable clinical presentation in novel mutation of the SRP54 gene / L. Goldberg, [et al.]. // *Pediatric Blood & Cancer*. 2020. № 6 (67). C. e28237.
51. Condino-Neto, A. Changing the Lives of People With Primary Immunodeficiencies (PI) With Early Testing and Diagnosis / A. Condino-Neto, F.J. Espinosa-Rosales // *Frontiers in Immunology*. 2018. № 9. C. 1439.
52. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies / Dong C. et al. // *Human molecular genetics*. – 2015. – T. 24. – №. 8. – C. 2125-2137.
53. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. / H. Ashkenazy et al. // *Nucleic Acids Research*, 2016, 8;44(W1):W344-50
54. Contribution of Iran in Elucidating the Genetic Causes of Autosomal Recessive Intellectual Disability / R. Ataei, [et al.]. // *Archives of Iranian Medicine*. 2019. № 8 (22). C. 461–471.

55. Csardi, G. The Igraph Software Package for Complex Network Research / G. Csardi, T. Nepusz // *InterJournal*. 2005. (Complex Systems). C. 1695.
56. Current Perspectives and Unmet Needs of Primary Immunodeficiency Care in Asia Pacific / D. Leung, [et al.]. // *Frontiers in Immunology*. 2020. Vol. 11. C. 1605.
57. Current status and prospects of primary immunodeficiency diseases in Asia / R.K. Pilia, [et al.]. // *Genes & Diseases*. 2019. № 1 (7). C. 3–11.
58. Cytoscape: a software environment for integrated models of biomolecular interaction networks / P. Shannon, [et al.]. // *Genome Research*. 2003. № 11 (13). C. 2498–2504.
59. dbNSFP v3. 0: A one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs/ Liu X. et al. // *Human mutation*. – 2016. – T. 37. – №. 3. – C. 235-241.
60. Defects in TCIRG1 subunit of the vacuolar proton pump are responsible for a subset of human autosomal recessive osteopetrosis / A. Frattini, [et al.]. // *Nature Genetics*. 2000. № 3 (25). C. 343–346.
61. Diagnostic Tools for Inborn Errors of Human Immunity (Primary Immunodeficiencies and Immune Dysregulatory Diseases) / A.M. Richardson, [et al.]. // *Current Allergy and Asthma Reports*. 2018. № 3 (18). C. 19.
62. Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome / Y. Minegishi, [et al.]. // *Nature*. 2007. № 7157 (448). C. 1058–1062.
63. Editorial: Methods and Applications of Computational Immunology / B. Chain, [et al.]. // *Frontiers in Immunology*. 2019. № 10. C. 2818.
64. Ekblom, R. A field guide to whole-genome sequencing, assembly and annotation / R. Ekblom, J.B.W. Wolf // *Evolutionary Applications*. 2014. № 9 (7). C. 1026–1042.
65. Epidemiology of congenital neutropenia/ Donadieu J. et al. // *Hematology/Oncology Clinics*. – 2013. – T. 27. – №. 1. – C. 1-17.

66. Epidermodysplasia Verruciformis: Inborn Errors of Immunity to Human Beta-Papillomaviruses / S.J. de Jong, [et al.]. // *Frontiers in Microbiology*. 2018. (9). C. 1222.
67. Evaluation of Clinical Manifestations in Patients with Severe Lymphedema with and without CCBE1 Mutations / M. Alders, [et al.]. // *Molecular Syndromology*. 2013. № 3 (4). C. 107–113.
68. Expression of the neutrophil elastase gene during human bone marrow cell differentiation // *The Journal of Experimental Medicine*. 1989. № 3 (169). C. 833–845.
69. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features / M.F. Rogers, [et al.]. // *Bioinformatics (Oxford, England)*. 2018. № 3 (34). C. 511–513.
70. Flanagan S.E. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations/ . Flanagan S. E., Patch A. M., Ellard S. // *Genetic testing and molecular biomarkers*. – 2010. – T. 14. – №. 4. – C. 533-537.
71. Gene Ontology Consortium The Gene Ontology (GO) project in 2006 // *Nucleic Acids Research*. 2006. № Database issue (34). C. D322-326.
72. Genome bioinformatic analysis of nonsynonymous SNPs / D.F. Burke, [et al.]. // *BMC bioinformatics*. 2007. (8). C. 301.
73. Ghosh, S. NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses / S. Ghosh, M.J. May, E.B. Kopp // *Annual Review of Immunology*. 1998. (16). C. 225–260.
74. Global overview of primary immunodeficiencies: a report from Jeffrey Modell Centers worldwide focused on diagnosis, treatment, and discovery / V. Modell, [et al.]. // *Immunologic Research*. 2014. № 1 (60). C. 132–144.
75. Gomber, S. Vaccine Associated Paralytic Poliomyelitis Unmasking Common Variable Immunodeficiency / S. Gomber, V. Arora, P. Dewan // *Indian Pediatrics*. 2017. № 3 (54). C. 241–242.

76. GPS: A comprehensive www server for phosphorylation sites prediction. / Y. Xue et al. // *Nucleic acids research*. – 2005. – T. 33. – №. suppl\_2. – C. W184-W187.
77. Gassoum A., E Abdelraheem N., Elsadig N. Comprehensive analysis of rsSNPs associated with hypertension using in-silico bioinformatics tools. – 2016.
78. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers / Abraham M. J. et al. // *SoftwareX*. – 2015. – T. 1. – C. 19-25.
79. Growth in diagnosis and treatment of primary immunodeficiency within the global Jeffrey Modell Centers Network / J. Quinn, [et al.]. // *Allergy, Asthma, and Clinical Immunology: Official Journal of the Canadian Society of Allergy and Clinical Immunology*. 2022. № 1 (18). C. 19.
80. Guidelines for genetic studies in single patients: lessons from primary immunodeficiencies / J.-L. Casanova, [et al.]. // *The Journal of Experimental Medicine*. 2014. № 11 (211). C. 2137–2149.
81. Hauck, F. Pathogenic mechanisms and clinical implications of congenital neutropenia syndromes / F. Hauck, C. Klein // *Current Opinion in Allergy and Clinical Immunology*. 2013. № 6 (13). C. 596–606.
82. Hecht M. Better prediction of functional effects for sequence variants. / Hecht M, Bromberg Y, Rost B. // *BMC Genomics*. 2015;16 Приложение 8:S1.
83. Hennekam syndrome can be caused by FAT4 mutations and be allelic to Van Maldergem syndrome / M. Alders, [et al.]. // *Human Genetics*. 2014. № 9 (133). C. 1161–1167.
84. Hershfield, M.S. Genotype is an important determinant of phenotype in adenosine deaminase deficiency // *Current Opinion in Immunology*. 2003. № 5 (15). C. 571–577.
85. Highly accurate protein structure prediction with AlphaFold / J. Jumper, [et al.]. // *Nature*. 2021. № 7873 (596). C. 583–589.

86. Hildebrand, P.W. Bringing Molecular Dynamics Simulation Data into View / P.W. Hildebrand, A.S. Rose, J.K.S. Tiemann // Trends in Biochemical Sciences. 2019. № 11 (44). C. 902–913.
87. Hilliard, R.I. Congenital abnormalities of the lymphatic system: a new clinical classification / R.I. Hilliard, J.B. McKendry, M.J. Phillips // Pediatrics. 1990. № 6 (86). C. 988–994.
88. Horvath, S. Geometric interpretation of gene coexpression network analysis / S. Horvath, J. Dong // PLoS computational biology. 2008. № 8 (4). C. e1000117.
89. Human Inborn Errors of Immunity: 2022 Update on the Classification from the International Union of Immunological Societies Expert Committee / S.G. Tangye [et al.]. // Journal of Clinical Immunology. 2022. № 7 (42). C. 1473–1507.
90. Human Inborn Errors of Immunity: 2019 Update of the IUIS Phenotypical Classification / A. Bousfiha, [et al.]. // Journal of Clinical Immunology. 2020. № 1 (40). C. 66–81.
91. Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis/ Goel R. et al. //Molecular BioSystems. – 2012. – T. 8. – №. 2. – C. 453-463.
92. Hypermutable Non-Synonymous Sites Are under Stronger Negative Selection / S. Schmidt, [et al.]. // PLOS Genetics. 2008. № 11 (4). C. e1000281.
93. Identification, analysis, and prediction of protein ubiquitination sites./ P. Radivojac et al. // Proteins: Structure, Function, and Bioinformatics. – 2010. – T. 78. – №. 2. – C. 365-380.
94. Identification of candidate disease genes in patients with common variable immunodeficiency/ G. Liu, [et al.]. // Quantitative Biology. 2019. № 3 (7). C. 190–201
95. Immunodeficiency, autoinflammation and amylopectinosis in humans with inherited HOIL-1 and LUBAC deficiency / B. Boisson, [et al.]. // Nature Immunology. 2012. № 12 (13). C. 1178–1186.

96. Inducible expression of a disease-associated ELANE mutation impairs granulocytic differentiation, without eliciting an unfolded protein response/Garg B. et al. //Journal of Biological Chemistry. – 2020. – T. 295. – №. 21. – C. 7492-7500.
97. In Silico Analysis Revealed Five Novel High-Risk Single-Nucleotide Polymorphisms (rs200384291, rs201163886, rs193141883, rs201139487, and rs201723157) in ELANE Gene Causing Autosomal Dominant Severe Congenital Neutropenia 1 and Cyclic Hematopoiesis / K. Shinwari, [et al.]. // The Scientific World Journal. 2022. № 6. C. 3356835.
98. Increasing incidence of human melioidosis in Northeast Thailand / D. Limmathurotsakul, [et al.]. // The American Journal of Tropical Medicine and Hygiene. 2010. № 6 (82). C. 1113–1117.
99. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2 / V. Pejaver, [et al.]. // Nature Communications. 2020. № 1 (11). C. 5918.
100. InnateDB: systems biology of innate immunity and beyond-recent updates and continuing curation / K. Breuer, [et al.]. // Nucleic Acids Research. 2013. № Database issue (41). C. D1228-1233.
101. INstruct: a database of high-quality 3D structurally resolved protein interactome networks / M.J. Meyer, [et al.]. // Bioinformatics (Oxford, England). 2013. № 12 (29). C. 1577–1579.
102. IntAct: an open source molecular interaction database / H. Hermjakob, [et al.]. // Nucleic Acids Research. 2004. № Database issue (32). C. D452-455.
103. Intrinsic and extrinsic causes of malignancies in patients with primary immunodeficiency disorders / F. Hauck, [et al.]. // The Journal of Allergy and Clinical Immunology. 2018. № 1 (141). C. 59-68.e4.
104. Introduction on primary immunodeficiency diseases / N. Rezaei [et al.]; под ред. N. Rezaei, A. Aghamohammadi, L.D. Notarangelo. Springer Verlag, 2017. C. 1–81.

105. Investigation of deleterious effects of nsSNPs in the POT1 gene: a structural genomics-based approach to understand the mechanism of cancer development / M. Amir, [et al.]. // *Journal of Cellular Biochemistry*. 2019. № 6 (120). C. 10281–10294.
106. iStable 2.0: Predicting protein thermal stability changes by integrating various characteristic modules / C.-W. Chen. [et al.]. // *Computational and Structural Biotechnology Journal*. 2020. (18). C. 622–630.
107. Itan, Y. Novel Primary Immunodeficiency Candidate Genes Predicted by the Human Gene Connectome / Y. Itan, J.-L. Casanova // *Frontiers in Immunology*. 2015. (6). C. 142.
108. Iwasaki, A. Regulation of adaptive immunity by the innate immune system / A. Iwasaki, R. Medzhitov // *Science (New York, N.Y.)*. 2010. № 5963 (327). C. 291–295.
109. Kanehisa, M. KEGG: kyoto encyclopedia of genes and genomes / M. Kanehisa, S. Goto // *Nucleic Acids Research*. 2000. № 1 (28). C. 27–30.
110. Kohn L.A. Gene Therapies for Primary Immune Deficiencies / L.A. Kohn, D.B. Kohn. // *Frontiers in Immunology*. 2021. № 12 C.648951
111. Komander, D. The ubiquitin code / D. Komander, M. Rape // *Annual Review of Biochemistry*. 2012. (81). C. 203–229.
112. Ku, B.C. Neutropenia in the Febrile Child / B.C. Ku, C. Bailey, F. Balamuth // *Pediatric Emergency Care*. 2016. № 5 (32). C. 329–334.
113. Lee, P.P.W. Cellular and Molecular Defects Underlying Invasive Fungal Infections-Revelations from Endemic Mycoses / P.P. Lee, Y.-L. Lau // *Frontiers in Immunology*. 2017. (8). C. 735.
114. Lee, P. P.W. Endemic infections in Southeast Asia provide new insights to the phenotypic spectrum of primary immunodeficiency disorders / P.P.W. Lee, Y.-L. Lau // *Asian Pacific Journal of Allergy and Immunology*. 2013. № 3 (31). C. 217–226.

115. Lee, P.P.W. Improving care, education, and research: the Asian primary immunodeficiency network / P.P.W. Lee, Y.-L. Lau // *Annals of the New York Academy of Sciences*. 2011. Vol. 1238. C. 33–41.
116. Lee, P.P.W. Primary immunodeficiencies: «new» disease in an old country / P.P.W. Lee, Y.-L. Lau // *Cellular & Molecular Immunology*. 2009. № 6 (6). C. 397–406.
117. Leiding, J.W. Warts and All: HPV in Primary Immunodeficiencies / J.W. Leiding, S.M. Holland // *The Journal of allergy and clinical immunology*. 2012. № 5 (130). C. 1030–1048.
118. Lessons learned from the study of human inborn errors of innate immunity / G. Bucciol, [et al.]. // *The Journal of Allergy and Clinical Immunology*. 2019. № 2 (143). C. 507–527.
119. Li, H. Fast and accurate short read alignment with Burrows-Wheeler transform / Li H., Durbin R. // *Bioinformatics (Oxford, England)*. 2009. № 14 (25). C. 1754–1760.
120. Llimma powers differential expression analyses for RNA-sequencing and microarray studies/ Ritchie M. E. et al. // *Nucleic acids research*. – 2015. – T. 43. – №. 7. – C. e47-e47.
121. Linear ubiquitination prevents inflammation and regulates immune signalling / B. Gerlach, [et al.]. // *Nature*. 2011. № 7340 (471). C. 591–596.
122. Liu B. C. Host-intrinsic interferon status in infection and immunity / Liu B. C., Sarhan J., Poltorak A. // *Trends in molecular medicine*. – 2018. – T. 24. – №. 8. – C. 658-668.
123. Loging, W. High-throughput electronic biology: mining information for drug discovery / W. Loging, L. Harland, B. Williams-Jones // *Nature Reviews Drug Discovery*. 2007. № 3 (6). C. 220–230.
124. Loss of ADAMTS3 activity causes Hennekam lymphangiectasia-lymphedema syndrome 3 / P. Brouillard, [et al.]. // *Human Molecular Genetics*. 2017. № 21 (26). C. 4095–4104.

125. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity/ Jagadeesh K. A. et al. // *Nature genetics*. – 2016. – T. 48. – №. 12. – C. 1581-1586.
126. Maere, S. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks / S. Maere, K. Heymans, M. Kuiper // *Bioinformatics (Oxford, England)*. 2005. № 16 (21). C. 3448–3449.
127. Malaysia's First Transplanted Case of Chronic Granulomatous Disease: The Journey of Overcoming Obstacles / I.H. Ismail, [et al.]. // *Children (Basel, Switzerland)*. 2016. № 2 (3). C. 9.
128. Malaviya, A.N. Ataxia telangiectasia: immunological abnormalities in probands and first degree relatives in 5 families / A.N. Malaviya, K.K. Sachdeva, N. Singh // *The Journal of the Association of Physicians of India*. 1973. № 8 (21). C. 701–705.
129. McDermott, D.H. WHIM syndrome: Immunopathogenesis, treatment and cure strategies / D.H. McDermott, P.M. Murphy // *Immunological Reviews*. 2019. № 1 (287). C. 91–102.
130. Medzhitov, R. Decoding the patterns of self and nonself by the innate immune system / R. Medzhitov, C.A. Janeway // *Science (New York, N.Y.)*. 2002. № 5566 (296). C. 298–300.
131. Mehta, S. R. Agammaglobulinaemia / S.R. Mehta, L.R. Sarin, L.M. Sanghvi // *Journal of the Indian Medical Association*. 1964. Vol. 42. C. 539–541.
132. Molecular docking and structure-based drug design strategies / L.G. Ferreira, [et al.]. // *Molecules (Basel, Switzerland)*. 2015. № 7 (20). C. 13384–13421.
133. Molecular markers in bladder cancer / F. Soria, [et al.]. // *World Journal of Urology*. 2019. № 1 (37). C. 31–40.
134. Mooney, M.A. Gene set analysis: A step-by-step guide / M.A. Mooney, B. Wilmot // *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*. 2015. № 7 (168). C. 517–527.

135. MutationTaster evaluates disease-causing potential of sequence alterations / Schwarz J. M. et al. // Nature methods. – 2010. – T. 7. – №. 8. – C. 575-576.
136. Mutations in CCBE1 cause generalized lymph vessel dysplasia in humans / M. Alders, [et al.]. // Nature Genetics. 2009. № 12 (41). C. 1272–1274.
137. Mutations in ELA2, encoding neutrophil elastase, define a 21-day biological clock in cyclic haematopoiesis / M. Horwitz, [et al.]. // Nature Genetics. 1999. № 4 (23). C. 433–436.
138. Mutations in the  $\alpha 3$  subunit of the vacuolar H(+)-ATPase cause infantile malignant osteopetrosis / U. Kornak, [et al.]. // Human Molecular Genetics. 2000. № 13 (9). C. 2059–2063.
139. Mutations in the ELA2 gene correlate with more severe expression of neutropenia: a study of 81 patients from the French Neutropenia Register / C. Bellanné-Chantelot, [et al.]. // Blood. 2004. № 11 (103). C. 4119–4125.
140. Mutations in the gene encoding neutrophil elastase in congenital and cyclic neutropenia. / Dale D.C. et al. / Blood, The Journal of the American Society of Hematology. – 2000. – T. 96. – №. 7. – C. 2317-2322
141. MutationTaster evaluates disease-causing potential of sequence alterations / J.M. Schwarz, [et al.]. // Nature Methods. 2010. № 8 (7). C. 575–576.
142. Nationwide survey of patients with primary immunodeficiency diseases in Japan / M. Ishimura, [et al.]. // Journal of Clinical Immunology. 2011. № 6 (31). C. 968–976.
143. Network-based approach to prediction and population-based validation of in silico drug repurposing / F. Cheng, [et al.]. // Nature Communications. 2018. № 1 (9). C. 2691.
144. Neutropenia and primary immunodeficiency diseases / N. Rezaei, [et al.]. // International Reviews of Immunology. 2009. № 5 (28). C. 335–366.
145. Ng, P.C. SIFT: Predicting amino acid changes that affect protein function / P.C. Ng, S. Henikoff // Nucleic Acids Research. 2003. № 13 (31). C. 3812–3814.

146. Notarangelo, L.D. Primary immunodeficiencies: novel genes and unusual presentations / Notarangelo, L.D. G. Uzel, V.K. Rao // Hematology. American Society of Hematology. Education Program. 2019. № 1 (2019). C. 443–448.
147. Novel HAX1 mutations in patients with severe congenital neutropenia reveal isoform-dependent genotype-phenotype associations / M. Germeshausen, [et al.]. // Blood. 2008. № 10 (111). C. 4954–4957.
148. Osteopetrosis: genetics, treatment and new insights into osteoclast function / C. Sobacchi, [et al.]. // Nature Reviews Endocrinology. 2013. № 9 (9). C. 522–536.
149. Padron G.T. Autoimmunity in Primary Immunodeficiencies (PID) / G.T. Padron, V.P. Hernandez-Trujillo // Clinical Reviews in Allergy & Immunology. – 2022. – C. 1-18.
150. PANTHER: a library of protein families and subfamilies indexed by function/ Thomas P. D. et al. // Genome research. – 2003. – T. 13. – №. 9. – C. 2129-2141.
151. Papadopoulou-Alataki, E. Prevention of infection in children and adolescents with primary immunodeficiency disorders / E. Papadopoulou-Alataki, A. Hassan, E.G. Davies // Asian Pacific Journal of Allergy and Immunology. 2012. № 4 (30). C. 249–258.
152. Paradis, E. APE: Analyses of Phylogenetics and Evolution in R language / E. Paradis, J. Claude, K. Strimmer // Bioinformatics (Oxford, England). 2004. № 2 (20). C. 289–290.
153. Patients with Primary Immunodeficiencies Are a Reservoir of Poliovirus and a Risk to Polio Eradication / A. Aghamohammadi, [et al.]. // Frontiers in Immunology. 2017. Vol. 8. C. 685.
154. PDBe: towards reusable data delivery infrastructure at protein data bank in Europe / S. Mir, [et al.]. // Nucleic Acids Research. 2018. № D1 (46). C. D486–D492.
155. Penicillium marneffeii infection and impaired IFN- $\gamma$  immunity in humans with autosomal-dominant gain-of-phosphorylation STAT1 mutations / P.P.W. Lee,

- [et al.]. // *The Journal of Allergy and Clinical Immunology*. 2014. № 3 (133). C. 894-896.e5.
156. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse / P.V. Hornbeck, [et al.]. // *Nucleic Acids Research*. 2012. № Database issue (40). C. D261-270.
157. Phylogenetic perspectives in innate immunity / J.A. Hoffmann, [et al.]. // *Science* (New York, N.Y.). 1999. № 5418 (284). C. 1313–1318.
158. Picard, C. Infectious diseases in patients with IRAK-4, MyD88, NEMO, or I $\kappa$ B $\alpha$  deficiency / C. Picard, J.-L. Casanova, A. Puel // *Clinical Microbiology Reviews*. 2011. № 3 (24). C. 490–497.
159. PIDO: the primary immunodeficiency disease ontology / N. Adams, [et al.]. // *Bioinformatics* (Oxford, England). 2011. № 22 (27). C. 3193–3199.
160. PINA v2.0: mining interactome modules / M.J. Cowley, [et al.]. // *Nucleic Acids Research*. 2012. № Database issue (40). C. D862-865.
161. Polyglucosan body myopathy caused by defective ubiquitin ligase RBCK1 / J. Nilsson, [et al.]. // *Annals of Neurology*. 2013. № 6 (74). C. 914–919.
162. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. / C. Steentoft et al // *The EMBO journal*. – 2013. – T. 32. – №. 10. – C. 1478-1488.
163. Prediction of Candidate Primary Immunodeficiency Disease Genes Using a Support Vector Machine Learning Approach / S. Keerthikumar, [et al.]. // *DNA Research*. 2009. № 6 (16). C. 345–351.
164. Prevalence and Outcomes of Primary Immunodeficiency in Hospitalized Children in the United States / Z. Rubin, [et al.]. // *The Journal of Allergy and Clinical Immunology. In Practice*. 2018. № 5 (6). C. 1705-1710.e1.
165. Primary immunodeficiencies in India: a perspective / S. Gupta, [et al.]. // *Annals of the New York Academy of Sciences*. 2012. Vol. 1250. C. 73–79.
166. Primary Immunodeficiencies in Russia: Data From the National Registry / A.A. Mukhina, [et al.]. // *Frontiers in Immunology*. 2020. Vol. 11. C. 1491.

167. Primary Immunodeficiency Diseases; A 20 Years Experience in a Tertiary University Hospital at Ramathibodi / O. Luecha, [et al.]. // *Journal of Allergy and Clinical Immunology*. 2012. № 2 (129). C. AB158.
168. Primary immunodeficiency diseases in Singapore--the last 11 years / D.L. Lim, [et al.]. // *Singapore Medical Journal*. 2003. № 11 (44). C. 579–586.
169. Primary Immunodeficiency Disorders in India-A Situational Review / A.K. Jindal, [et al.]. // *Frontiers in Immunology*. 2017. Vol. 8. C. 714.
170. PROCHECK: a program to check the stereochemical quality of protein structures / Laskowski R. A. et al. // *Journal of applied crystallography*. – 1993. – T. 26. – №. 2. – C. 283-291.
171. Protein structure prediction / Agnihotry S. et al. // *Bioinformatics*. – Academic Press, 2022. – C. 177-188.
172. Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy / F. Cheng, [et al.]. // *Oncotarget*. 2014. № 11 (5). C. 3697–3710.
173. Ramachandran GN. Stereochemistry of polypeptide chain configurations. / Ramachandran GN, Ramakrishnan C, Sasisekharan V. // *J. mol. Biol.* – 1963. – T. 7. – C. 95-99.
174. RBCK1 negatively regulates tumor necrosis factor- and interleukin-1-triggered NF-kappaB activation by targeting TAB2/3 for degradation / Y. Tian, [et al.]. // *The Journal of Biological Chemistry*. 2007. № 23 (282). C. 16776–16782.
175. Recurrent staphylococcal cellulitis and subcutaneous abscesses in a child with autoantibodies against IL-6 / A. Puel, [et al.]. // *Journal of Immunology (Baltimore, Md.: 1950)*. 2008. № 1 (180). C. 647–654.
176. Reva B. Predicting the functional impact of protein mutations: application to cancer genomics/ Reva B., Antipin Y., Sander C. // *Nucleic acids research*. – 2011. – T. 39. – №. 17. – C. e118-e118.
177. Revisiting human IL-12R $\beta$ 1 deficiency: a survey of 141 patients from 30 countries / L. de Beaucoudrey, [et al.]. // *Medicine*. 2010. № 6 (89). C. 381–402.

178. Revisiting human primary immunodeficiencies / J.-L. Casanova, [et al.]. // *Journal of Internal Medicine*. 2008. № 2 (264). C. 115–127.
179. Robert, F. Exploring the Impact of Single-Nucleotide Polymorphisms on Translation / F. Robert, J. Pelletier // *Frontiers in Genetics*. 2018. Vol. 9. C. 507.
180. Rodriguez-Esteban, R. Differential gene expression in disease: a comparison between high-throughput studies and the literature / R. Rodriguez-Esteban, X. Jiang // *BMC medical genomics*. 2017. № 1 (10). C. 59.
181. Rosenzweig, S.D. Defects in the interferon-gamma and interleukin-12 pathways / S.D. Rosenzweig, S.M. Holland // *Immunological Reviews*. 2005. (203). C. 38–47.
182. Roy, A. I-TASSER: a unified platform for automated protein structure and function prediction / A. Roy, A. Kucukural, Y. Zhang // *Nature Protocols*. 2010. № 4 (5). C. 725–738.
183. Samarghitean, C. Bioinformatics services related to diagnosis of primary immunodeficiencies / C. Samarghitean, M. Vihinen // *Current Opinion in Allergy and Clinical Immunology*. 2009. № 6 (9). C. 531–536.
184. Sasaki, Y. Crucial Role of Linear Ubiquitin Chain Assembly Complex-Mediated Inhibition of Programmed Cell Death in TLR4-Mediated B Cell Responses and B1b Cell Development / Y. Sasaki, K. Iwai // *Journal of Immunology (Baltimore, Md.: 1950)*. 2018. № 10 (200). C. 3438–3449.
185. Scalable algorithms for molecular dynamics simulations on commodity clusters / K.J. Bowers, [et al.] // *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*. 2006. C. 84-es.
186. Second Case of HOIP Deficiency Expands Clinical Features and Defines Inflammatory Transcriptome Regulated by LUBAC / H. Oda, [et al.]. // *Frontiers in Immunology*. 2019. (10). C. 479.
187. Severe congenital neutropenia: inheritance and pathophysiology / J. Skokowa, [et al.]. // *Current Opinion in Hematology*. 2007. № 1 (14). C. 22–28.
188. Severe congenital neutropenias / J. Skokowa [et al.]. // *Nature Reviews. Disease Primers*. 2017. (3). C. 17032.

189. SHARPIN is a component of the NF- $\kappa$ B-activating linear ubiquitin chain assembly complex / Tokunaga F. [et al.]. // *Nature*. 2011. № 7340 (471). C. 633–636.
190. Shore, G.C. Signaling cell death from the endoplasmic reticulum stress response / G.C. Shore, F.R. Papa, S.A. Oakes // *Current Opinion in Cell Biology*. 2011. № 2 (23). C. 143–149.
191. SignaLink 2 - a signaling pathway resource with multi-layered regulatory networks / D. Fazekas, [et al.]. // *BMC systems biology*. 2013. № 7. C. 7.
192. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets / D. Szklarczyk, [et al.]. // *Nucleic Acids Research*. 2019. № D1 (47). C. D607–D613.
193. Survival of mature T cells depends on signaling through HOIP / K. Okamura, [et al.]. // *Scientific Reports*. 2016. Vol. 6. C. 36135.
194. Susceptibility to mycobacterial infections in children with X-linked chronic granulomatous disease: a review of 17 patients living in a region endemic for tuberculosis / P.P.W. Lee, [et al.]. // *The Pediatric Infectious Disease Journal*. 2008. № 3 (27). C. 224–230.
195. Systems Bioinformatics: increasing precision of computational diagnostics and therapeutics through network-based approaches / A. Oulas, [et al.]. // *Briefings in Bioinformatics*. 2017. № 3 (20). C. 806–824.
196. Systems Biology as a Comparative Approach to Understand Complex Gene Expression in Neurological Diseases / L. Diaz-Beltran, [et al.]. // *Behavioral Sciences*. 2013. № 2 (3). C. 253–272.
197. Takada, H. Primary immunodeficiency in Japan; epidemiology, diagnosis, and pathogenesis // *Pediatrics International: Official Journal of the Japan Pediatric Society*. 2013. № 6 (55). C. 671–674.
198. TCIRG1-dependent recessive osteopetrosis: mutation analysis, functional identification of the splicing defects, and in vitro rescue by U1 snRNA / L. Susani, [et al.]. // *Human Mutation*. 2004. № 3 (24). C. 225–235.

199. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data / A. McKenna, [et al.]. // *Genome Research*. 2010. № 9 (20). C. 1297–1303.
200. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists / D.W. Huang, [et al.]. // *Genome Biology*. 2007. № 9 (8). C. R183.
201. The human gene connectome as a map of short cuts for morbid allele discovery / Y. Itan, [et al.]. // *Proceedings of the National Academy of Sciences of the United States of America*. 2013. № 14 (110). C. 5558–5563.
202. The mutational spectrum of human malignant autosomal recessive osteopetrosis / C. Sobacchi, [et al.]. // *Human Molecular Genetics*. 2001. № 17 (10). C. 1767–143.
203. The Sequence Alignment/Map format and SAMtools / H. Li, [et al.]. // *Bioinformatics (Oxford, England)*. 2009. № 16 (25). C. 2078–2079.
204. The variant call format and VCFtools / P. Danecek, [et al.]. // *Bioinformatics (Oxford, England)*. 2011. № 15 (27). C. 2156–2158.
205. Using the MINT database to search protein interactions / Calderone A. et al. // *Current Protocols in Bioinformatics*. – 2020. – T. 69. – №. 1. – C. e93.
206. Wang, K. ANNOVAR functional annotation of genetic variants from high-throughput sequencing data / K. Wang, M. Li, H.K. Hakonarson // *ANNOVAR* // *Nucleic Acids Research*. 2010. № 16 (38). C. e164.
207. The Phyre2 web portal for protein modeling, prediction and analysis / L.A. Kelley, [et al.]. // *Nature Protocols*. 2015. № 6 (10). C. 845–858.
208. The prevalences [correction] and patient characteristics of primary immunodeficiency diseases in Turkey - two centers study / S.S. Kilic, [et al.]. // *Journal of Clinical Immunology*. 2013. № 1 (33). C. 74–83.
209. The Primary Immunodeficiency Database in Japan / K. Mitsui-Sekinaka, [et al.]. // *Frontiers in Immunology*. 2022. Vol. 12.

210. UCSF Chimera--a visualization system for exploratory research and analysis / E.F. Pettersen, [et al.]. // *Journal of computational chemistry*. 2004. № 13 (25). C. 1605–1612.
211. Uses of Next-Generation Sequencing Technologies for the Diagnosis of Primary Immunodeficiencies / M. Seleman, [et al.]. // *Frontiers in Immunology*. 2017. (8). C. 847.
212. VarCards: an integrated genetic and clinical database for coding variants in the human genome / J. Li, [et al.] // *Nucleic Acids Research*. 2018. № D1 (46). C. D1039–D1048.
213. Visualizing the drug target landscape / S.J. Campbell, [et al.]. // *Drug Discovery Today*. 2010. № 1–2 (15). C. 3–15.
214. Wiskott-Aldrich syndrome. A case report / M.C. Gupta, [et al.]. // *The Journal of the Association of Physicians of India*. 1964. Vol. 12. C. 531–533.
215. Wu J. Prediction of deleterious nonsynonymous single-nucleotide polymorphism for human diseases / Wu J., Jiang R. // *The Scientific World Journal*. – 2013. – T. 2013.
216. X-linked agammaglobulinemia: Twenty years of single-center experience from North West India / S. Singh, [et al.]. // *Annals of Allergy, Asthma & Immunology: Official Publication of the American College of Allergy, Asthma, & Immunology*. 2016. № 4 (117). C. 405–411.
217. clusterProfiler: an R package for comparing biological themes among gene clusters / G. Yu, [et al.]. // *Omics: A Journal of Integrative Biology*. 2012. № 5 (16). C. 284–287.
218. Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data / T. Wang, [et al.]. // *BMC bioinformatics*. 2019. № 1 (20). C. 40.
219. Distribution, clinical features and molecular analysis of primary immunodeficiency diseases in Chinese children: a single-center study from 2005 to 2011 / Z.-Y. Zhang, [et al.]. // *The Pediatric Infectious Disease Journal*. 2013. № 10 (32). C. 1127–1134.

220. Inborn errors of interferon (IFN)-mediated immunity in humans: insights into the respective roles of IFN-alpha/beta, IFN-gamma, and IFN-lambda in host defense / S.-Y. Zhang, [et al.]. // *Immunological Reviews*. 2008. (226). С. 29–40.
221. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets / Y., Zhou [et al.]. // *Nature Communications*. 2019. № 1 (10). С. 1523.
222. Wang, K. A brief procedure for big data analysis of gene expression / K. Wang, W. Wang, M. Li // *Animal Models and Experimental Medicine*. 2018. № 3 (1). С. 189–193.
223. Welte, K. Severe congenital neutropenia / K. Welte, C. Zeidler, D.C. Dale // *Seminars in Hematology*. 2006. № 3 (43). С. 189–195.
224. Yang, C.-H. *Chromobacterium violaceum* infection: a clinical review of an important but neglected infection / C.-H. Yang, Y.-H. Li // *Journal of the Chinese Medical Association: JCMA*. 2011. № 10 (74). С. 435–441.
225. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles | PNAS [Электр. ресурс]. URL: <https://www.pnas.org/doi/10.1073/pnas.0506580102> (дата обращения: 06.01.2023).
226. Lymphedema-lymphangiectasia-mental retardation (Hennekam) syndrome: a review - PubMed [Электронный ресурс]. URL: <https://pubmed.ncbi.nlm.nih.gov/12376947/> (дата обращения: 29.12.2022).
227. Primary Immunodeficiency Diseases: an Update on the Classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency / C. Picard [et al.]. // *Journal of Clinical Immunology*. 2015. № 8 (35). С. 696–726.
228. Primary Immune Deficiency Diseases in America: The Third National Survey of Patients (2007) | SCID Compass [Электронный ресурс]. URL: <https://scidcompass.org/publication/surveys/primary-immune-deficiency->

diseases-america-third-national-survey-patients-2007 (дата обращения: 19.12.2022).

### СПИСОК СОКРАЩЕНИЙ

<b>ВАК</b>	высшая аттестационная комиссия (Российской академии наук)
<b>ВВИГ</b>	внутривенные иммуноглобулины
<b>ВН</b>	врожденная нейтропения
<b>ВОИ</b>	врожденные ошибки иммунитета
<b>ДЭГ</b>	Дифференциально экспрессированные гены
<b>КДЦ</b>	Клинико-диагностический центр
<b>МГЦ</b>	медико-генетический центр
<b>НАЭПИД</b>	Национальная Ассоциация Экспертов по Первичным Иммунодефицитам
<b>ПИД</b>	первичные иммунодефициты
<b>РАН</b>	Российская академия наук
<b>США</b>	Соединенные Штаты Америки
<b>ТГСК</b>	трансплантация гемопоэтических стволовых клеток
<b>ТКИН</b>	Тяжелая комбинированная недостаточность
<b>УрО</b>	Уральское отделение (Российской академии наук)
<b>УрФУ</b>	Уральский федеральный университет
<b>ФГБУН</b>	федеральное государственное бюджетное учреждение науки
<b>ХГБ</b>	хроническая гранулематозная болезнь
<b>1000g</b>	1000 Genomes (база данных "1000 геномов")
<b>2D</b>	2 dimentiaonal (двумерное)
<b>3D</b>	3 dimentional (трехмерное)
<b>BWA</b>	Burrows-Wheeler Aligner (Выравниватель Берроуза-Уиллера)

<b>CADD</b>	Combined Annotation Dependent Depletion (Комбинированная аннотационная зависимая деплеция)
<b>CCBE1</b>	Collagen and calcium-binding protein 1 containing an epidermal growth factor-like domain (Коллаген и кальций-связывающий белок 1, содержащий эпидермальный фактор роста-подобный домен)
<b>CGD</b>	chronic granulomatous disease (хроническая гранулематозная болезнь)
<b>CINCA</b>	Chronic Infantile Neurological Cutaneous and Articular (Хроническая инфантильная неврологическая кожно-суставная болезнь)
<b>CISD2</b>	CDGSH iron sulfur domain 2 (CDGSH железо-серный домен 2)
<b>CXCR4</b>	Chemokine receptor 4 (Хемокиновый рецептор 4)
<b>DANN</b>	Annotating genetic variants, especially non-coding variants (Аннотирование генетических вариантов, особенно некодирующих вариантов)
<b>dbSNP</b>	Single Nucleotide Polymorphism Database (База данных однонуклеотидных полиморфизмов)
<b>ELANE</b>	Elastase, Neutrophil Expressed (Эластаза, экспрессируемая нейтрофилами)
<b>ExAC</b>	Exome Aggregation Consortium (Консорциум по агрегированию экзонов)
<b>FATHMM</b>	Functional Analysis through Hidden Markov (Функциональный анализ с помощью скрытых Марковских моделей)
<b>FATHMM-MKL</b>	Functional Analysis through Hidden Markov Models - Multiple Kernel Learning (Функциональный анализ с помощью скрытых марковских моделей - многоядерное обучение)
<b>FDR</b>	False Discovery Frequency (Частота ложных обнаружений)
<b>FEL</b>	Free energy landscape (Ландшафт свободной энергии)
<b>FGA</b>	Functional genomic alignment (функциональное геномное выравнивание)
<b>GAD</b>	Genetic Association Database (База данных генетических ассоциаций)

<b>GATK</b>	The Genome Analysis Toolkit (Набор инструментов для анализа генома)
<b>GEO</b>	Gene Expression Omnibus (Омнибус экспрессии генов)
<b>gnomAD</b>	The Genome Aggregation Database (База данных агрегации генома)
<b>GO</b>	Gene Ontology (Онтология генов)
<b>GOF</b>	Gain of Function (усиление функции)
<b>GSEA</b>	Gene set enrichment analysis (Анализ обогащения набора генов)
<b>НарMap</b>	Нарplotype Mapping Project (Проект картирования гаплотипов)
<b>HGC</b>	Human Genome Connectome (Коннектом генома человека)
<b>HGMP</b>	Human Genome Mapping Project (Проект картирования генома человека)
<b>HH-Pred</b>	Homology detection and structure prediction by HMM-HMM comparison (Обнаружение гомологии и предсказание структуры путем сравнения скрытых марковских моделей )
<b>HOL1</b>	Heme-oxidized IRP2 ubiquitin ligase-1 (Гем-окисленная убиквитиновая лигаза-1 IRP2)
<b>HOLP1</b>	HOIL-1-interacting protein (HOIL-1-взаимодействующий белок)
<b>IFN</b>	interferon (интерферон)
<b>IFNGIR1</b>	Interferon-gamma receptor 1 (Рецептор интерферона-гамма 1)
<b>IGV</b>	Integrative Genome Viewer (Интегративный просмотрщик геномов)
<b>IL-1<math>\beta</math></b>	Interleukin-1 beta (интерлейкин 1-бета)
<b>I-mutant</b>	In Silico MUTation Analysis (Мутационный анализ In Silico)
<b>I-Tasser</b>	Iterative Threading ASSEmbly Refinement (Итеративное уточнение нитей ASSEmbly)
<b>IUIS</b>	International Union of Immunological Societies (Международный союз иммунологических обществ)
<b>iVDPV</b>	immunodeficiency-associated polioviruses obtained through vaccination (полиовирусы, ассоциированные с иммунодефицитом, полученные в результате вакцинации)
<b>JMF</b>	Jeffrey Modell Foundation (Фонд Джеффри Моделла)

<b>KEGG</b>	Kyoto Encyclopedia of Genes and Genomes (Киотская энциклопедия генов и геномов)
<b>KREC</b>	Kappa-deleting recombination excision circle (каппа-делеционное рекомбинационное эксцизионное кольцо)
<b>LOF</b>	Lost of Function (потеря функции)
<b>LRT</b>	Likelihood Ratio Test (Тест отношения правдоподобия)
<b>LUBAC</b>	Linear Ubiquitin Chain Assembly Complex (Линейный комплекс сборки цепей убиквитина)
<b>MAF</b>	Minor allele frequency (Минорная частота аллелей)
<b>M-CAP</b>	Mendelian Clinically Applicable Pathogenicity (Менделевская клинически применимая патогенность)
<b>MDS</b>	Molecular Dynamics Simulation (симуляция молекулярной динамики)
<b>MSMD</b>	Mendelian susceptibility to mycobacterial diseases (Менделевская восприимчивость к микобактериальным болезням)
<b>MTOR</b>	Mechanistic Target of Rapamycin (Механистическая мишень рапамицина)
<b>MuPred</b>	Mutation Prediction (Предсказание мутаций)
<b>MU-Pro</b>	Mutation Protein (Мутационный белок)
<b>MYD88</b>	Myeloid Differentiation primary response gene 88 (Ген первичного ответа на миелоидную дифференцировку 88)
<b>NCBI</b>	National Center for Biotechnology Information (Национальный центр биотехнологической информации)
<b>NEMO</b>	NF-κB essential modulator (NF-κB незаменимый модулятор)
<b>NES</b>	Normalized enrichment score (Нормализованный показатель обогащения)
<b>NF-κB</b>	Nuclear factor-κB (ядерный фактор каппа-би)
<b>NGS</b>	next-generation sequencing (секвенирование следующего поколения)
<b>NK</b>	Natural killer (естественные киллеры)
<b>NKT</b>	Natural killer T-cells (естественные Т-киллеры)

<b>NOMID</b>	Neonatal-Onset Multisystem Inflammatory Disease (Мультисистемное воспалительное заболевание с неонатальным началом)
<b>nsSNPs</b>	non-synonymous single nucleotide polymorphisms (несинонимичные однонуклеотидные полиморфизмы)
<b>OMIM</b>	Online Mendelian Inheritance in Man (Онлайн база менделевского наследования у человека)
<b>PANTHER</b>	Protein ANalysis THrough Evolutionary Relationships (Анализ белков через эволюционные связи)
<b>PCA</b>	Principal Component Analysis (Анализ главных компонент)
<b>PCC</b>	Pearson's correlation analysis (корреляционный анализ Пирсона)
<b>PDB</b>	Protein data bank (Банк данных протеинов)
<b>PHD-SNP</b>	Predictor of Human Deleterious Single Nucleotide Polymorphisms (Предиктор вредных однонуклеотидных полиморфизмов человека)
<b>PIK3/АКТ</b>	Phosphoinositide 3-kinase/Protein Kinase B (Фосфоинозитид 3- киназа/протеинкиназа В)
<b>PolyPhen2</b>	Polymorphism Phenotyping v2 (Фенотипирование полиморфизмов версии 2)
<b>PPI</b>	Protein-protein interaction (Белок-белковое взаимодействие)
<b>PROVEAN</b>	Protein Variation Effect Analyzer (Анализатор влияния вариаций белка)
<b>RBCK1</b>	RanBP-type and C3HC4-type zinc finger-containing protein 1 (RanBP-тип и C3HC4-тип цинковый палец-содержащий белок 1)
<b>Rg</b>	Radius of gyration (радиус гирации)
<b>RMSD</b>	Root Mean Square Deviation (среднеквадратичное отклонение)
<b>RMSF</b>	Root Mean Square Fluctuation (среднеквадратичная флуктуация)
<b>SASA</b>	Solvent Accessible Surface Area (Площадь поверхности, доступная для растворителя)

<b>SHARPIN</b>	SHANK-associated RH domain interactor (SHANK-ассоциированный интерактор домена RH)
<b>SIFT</b>	Sorting Intolerant From Tolerant (Сортировка "нетерпимых" от "терпимых")
<b>SNAP2</b>	Scaled Neural Network Aligning Profile 2 (Профиль выравнивания масштабированной нейронной сети 2)
<b>SNP-GO</b>	Single Nucleotide Polymorphism Gene Ontology (Онтология генов и однонуклеотидных полиморфизмов)
<b>SNP</b>	Single nucleotide polymorphisms (Однонуклеотидные полиморфизмы)
<b>SNV</b>	Single nucleotide variants (Однонуклеотидные варианты)
<b>SOPMA</b>	Self-Optimized Prediction Method with Alignment (Метод самооптимизированного предсказания с выравниванием)
<b>SSE</b>	Secondary Structure Element (Элемент вторичной структуры)
<b>STAT1</b>	Signal Transducer and Activator of Transcription 1 (Трансдуктор сигнала и активатор транскрипции 1)
<b>STRINGS</b>	Search Tool for the Retrieval of Interacting Genes/Proteins (Инструмент для поиска взаимодействующих генов/белков)
<b>SVM</b>	Support vector machine (метод опорных векторов)
<b>TCIRG1</b>	T-cell immune regulator 1 (Т-клеточный иммунный регулятор 1)
<b>TLR</b>	Toll-like receptor (Толл-подобный рецептор)
<b>TMS</b>	Template Modeling score (Показатель моделирования шаблона)
<b>TNF</b>	Tumor necrosis factor (Фактор некроза опухоли)
<b>TREC</b>	T-cell receptor excision circles (эксцизионные кольца Т-клеточного рецептора)
<b>TYK2</b>	Tyrosine Kinase 2 (Тирозиновая киназа 2)
<b>Uniprot</b>	Universal Protein Resource (Универсальный белковый ресурс)
<b>VARCARDS</b>	Variant Interpretation Cards (Карточки интерпретации вариантов)
<b>VCF</b>	Variant Call Format (Формат вызова варианта)

<b>VEGF-C</b>	Vascular Endothelial Growth Factor C (Фактор роста эндотелия сосудов C)
<b>VEST</b>	Variant Effect Scoring Tool (Инструмент оценки влияния вариантов)
<b>WHIM</b>	Warts, Hypogammaglobulinemia, Infections, and Myelokathexis (Бородавки, гипогаммаглобулинемия, инфекции и миелокатетериоз)
<b>XLA</b>	X-linked agammaglobulinemia (X-сцепленная агаммаглобулинемия)