Дополнительные материалы

к Отчету РНФ по гранту № 24-24-00563

«Разработка цифровых образовательных программ в биомедицине»

за 2025 год

Материалы представлены на сайте

http://www.orlov-bio.ru/about/

научная группа:

рук. Гранта - Орлов Юрий Львович, д.б.н., проф. РАН

осн. исполнитель - Анашкина Анастасия Андреевна, к.ф.-м.н., доцент

исполнители - Савина Екатерина Алексеевна, ассистент, преподаватель

Сафрыгина Анастасия Александровна, студент

Исполнители, вспомогат. персонал: Волков Игорь Алексеевич, к.м.н.

Втюрина Дарья Николаевна, к.х.н.

Студенты - Туркина Василиса, Головач Алена, Терещенко Ярослав, Белоусова Анна, Каширин Андрей

Министерство здравоохранения Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего образования
ПЕРВЫЙ МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ имени И.М.СЕЧЕНОВА
Министерства здравоохранения Российской Федерации
(Сеченовский Университет)

УДК: 577.2
Рег. № НИОКТР 124040100050-7

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

«Компьютерный анализ структуры генных сетей и промоторных районов генов, ассоциированных с комплексными заболеваниями»

В рамках гранта РНФ № 24-24-00563
«Разработка цифровых образовательных программ в биомедицине»
Соглашения от 29.12.2023г.

(

Исполнитель НИР _____ Сафрыгина А.А.

Москва, 2025

РЕФЕРАТ

Отчет 42с., 12 рис., 43 источника, 2 таблицы.

Ключевые слова: промоторные районы генов, генные сети, муковисцидоз, CMA1, белок-белковые взаимодействия, трехмерная структура белка, биоинформатика, медицинская информатика, ДНК, РНК, компьютерное моделирование.

Объект исследования: открытые базы данных структур белков, генетической информации, сети белок-белковых взаимодействий, нуклеотидные и аминокислотные последовательности.

Цель работы: Компьютерный анализ генов комплексных заболеваний в рамках проекта «Разработка цифровых образовательных программ в биомедицине».

Работы включали следующие задачи: Подготовка списка генов, связанных с комплексными заболеваниями - глиомой, деменцией, муковисцидозом. Компьютерный анализ промоторных районов этих генов по банкам данных и предсказательным моделям, подготовка методического материала (текстового описания) по анализу сложности генетического текста промоторных и энхансерных районов генов, с подготовкой текста научной публикации (для представления на российской научной конференции).

Задачи: 1) Подготовка списков генов, связанных с комплексными заболеваниями, по базам данных GeneBank, OMIM, GeneCards (для исследуемых заболеваний - глиома, деменция, муковисцидоз).

2) Компьютерные расчеты и визуализация структуры прямых и инвертированных повторов в промоторных районов генов, расположение сайтов связывания транскрипционных факторов.

3) Подготовка текста тезисов научной конференции и научной публикации в журнальном формате.

4) Подготовка шаблона методических материалов по анализу промоторных районов генов используя интеграцию данных из баз данных (EPD, JASPAR, TRANSFAC, GeneCards).

СОДЕРЖАНИЕ:

РЕФЕРАТ.....3
СОДЕРЖАНИЕ:.....4
ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ 5
ВВЕДЕНИЕ 6
ОТРАБОТКА МЕТОДИКИ ПОСТРОЕНИЯ СПИСКА ГЕНОВ 7
ПРИМЕР АНАЛИЗА ГЕНОВ: КИСТОЗНЫЙ ФИБРОЗ (МУКОВИСЦИДОЗ) 8
РАСЧЁТ И АНАЛИЗ КАТЕГОРИЙ ГЕННЫХ ОНТОЛОГИЙ ГЕНОВ МУКОВИСЦИДОЗА.....9
РЕКОНСТРУКЦИЯ И АНАЛИЗ ГЕННОЙ СЕТИ МУКОВИСЦИДОЗА.....10
БЕЛОК CFTR, ЕГО ФУНКЦИИ И РОЛЬ В ПАТОГЕНЕЗЕ МУКОВИСЦИДОЗА..14
ОБСУЖДЕНИЕ ПЕРСПЕКТИВ ТЕРАПИИ МУКОВИСЦИДОЗА 17
СПИНАЛЬНАЯ МЫШЕЧНАЯ АТРОФИЯ I ТИПА 20
МЕТОДИКА. СОСТАВЛЕНИЕ СПИСКА ГЕНОВ 23
РАСЧЁТ И АНАЛИЗ КАТЕГОРИЙ ГЕННЫХ ОНТОЛОГИЙ.....24
ГРАФИК ТОЧЕЧНЫХ ЗНАЧЕНИЙ КАТЕГОРИЙ ГЕННЫХ ОНТОЛОГИЙ 27
РЕКОНСТРУКЦИЯ И АНАЛИЗ ГЕННОЙ СЕТИ 28
СТРУКТУРЫ БЕЛКОВ.....31
РЕЗУЛЬТАТЫ АНАЛИЗА ГЕНОВ АССОЦИИРОВАННЫХ СМА1 31
СЛОЖНОСТЬ ВИРУСНОГО ГЕНОМА И СТРУКТУРА ПОВТОРОВ 32
ЗАКЛЮЧЕНИЕ.....37
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 39

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

В настоящем отчете о НИР применяют следующие сокращения и обозначения:

CFTR (*Cystic Fibrosis Transmembrane Conductance Regulator)* - Трансмембранный Регулятор Проводимости Муковисцидоза).

SBML (Systems Biology Markup Language) – формат разметки данных системной биологии

FASTA – текстовый формат для нуклеотидных или полипептидных последовательностей, в котором нуклеотиды или аминокислоты обозначаются при помощи однобуквенных кодов.

РНК - рибонуклеиновая кислота.

ДНК - дезоксирибонуклеиновая кислота.

Виртуальный скрининг – процедура отбора из библиотеки химических соединений лигандов, которые наилучшим образом связываются с заданной биологической мишенью.

Докинг – вычислительная процедура поиска положения молекулы лиганда в активном центре белка-мишени, соответствующего положению, наблюдаемому в эксперименте.

ВВЕДЕНИЕ

Целью данной НИР являлся компьютерный анализ генов комплексных заболеваний в рамках проекта «Разработка цифровых образовательных программ в биомедицине»

В соответствии с планом работ по проекту, были составлены списки генов, и выполнен биоинформационный анализ молекулярных механизмов, определяющих аспекты экспрессии генов в комплексных заболеваниях на примере генов заболеваний глиомой, деменцией, болезнью Паркинсона, муковисцидоза (кистозного фиброза), CMA1, с реконструкцией генной сети и исследованием трехмерной структуры белков, кодируемых этими генами.

Для достижения поставленной цели необходимо решить ряд **научно-технических задач** по сбору и интеграции информации из онлайн баз данных структурной и функциональной аннотации генов.

В ходе работы решены следующие задачи:

1) Подготовлены списки генов, связанных с комплексными заболеваниями, по базам данных GeneBank, OMIM, GeneCards (для исследуемых заболеваний глиомой, деменцией, болезнью Паркинсона, муковисцидоз, CMA1).

2) Выполнены компьютерные расчеты и визуализация структуры прямых и инвертированных повторов в промоторных районов генов, расположение сайтов связывания транскрипционных факторов.

3) Подготовлены тексты тезисов научной конференции и научной публикации в журнальном формате (файл с иллюстрациями). Тезисы поданы на конференции «Сознание-2025» в Красновидово, II Международная конференция «Инновационные лекарственные средства от молекулы до пациента», XXXI Symposium on Bioinformatics and Computer-Aided Drug Discovery (BCADD-2025), III Всероссийская конференция с международным участием и школа молодых ученых «Физико-химические методы в междисциплинарных экологических исследованиях», Международный молодёжный научный форум «Медицинская весна – 2025», и приняты для публикации в сборники конференций.

Подготовлен текст научной публикации (для журнала «Вавиловский журнал генетики и селекции»)

4) Подготовка шаблона методических материалов по анализу промоторных районов генов используя интеграцию данных из баз данных (EPD, JASPAR, TRANSFAC, GeneCards) (файл с иллюстрациями).

Для решения образовательных и исследовательских задач большое значения имеет отработка использования компьютерного конвейера обработки данных (списков генов,

ассоциированных с заболеванием) на основе использования открытых (бесплатных) онлайн-программ и ресурсов.

В данной работе использовался компьютер стандартной конфигурации и биоинформатические онлайн-ресурсы, которые находятся в открытом доступе.

ОТРАБОТКА МЕТОДИКИ ПОСТРОЕНИЯ СПИСКА ГЕНОВ

База данных OMIM (Online Mendelian Inheritance in Man) использовалась для извлечения генов, связанных с наследственной предрасположенностью к заданному заболеванию. OMIM (https://omim.org/) - сборник генов человека и генетических фенотипов.

Использовались следующие базы данных:

PANTHER - Protein ANalysis THrough Evolutionary Relationships.

Ресурс PANTHER (http://pantherdb.org/) использовался для анализа генных онтологий. Инструмент классифицирует белки для анализа на основании эволюционных взаимосвязей.

DAVID - The Database for Annotation, Visualization and Integrated Discovery. Ресурс DAVID (https://david.ncifcrf.gov/summary.jsp) использовался для анализа категорий генных онтологий.

Программа g:GOSt. Ресурс g:GOSt (http://biit.cs.ut.ee/gprofiler/gost) использовался для визуализации генных онтологий. Ресурс позволяет построить график точечных значений категорий генных онтологий.

Ресурс STRING-DB (https://string-db.org/) использовался для реконструкции генной сети взаимодействий генов астроцитомы. Это база данных известных и прогнозируемых белок-белковых взаимодействий. Взаимодействия включают прямые (физические) и косвенные (функциональные) ассоциации.

Запись структуры генной сети проводилась в формате SBML (systems biology markup language) (https://sbml.org/).

MalaCards. База данных MalaCards (https://www.malacards.org/) использовалась для поиска заболеваний, ассоциированных с заданным, в том числе глиомой. База данных обладает возможностью поиска, которая предоставляет информацию обо всех аннотированных заболеваниях человека.

Биоинформационный анализ с использованием приведенных выше баз данных применялся для генов глиомы и других комплексных заболеваний, таких как болезнь Паркинсона [20,21].

Рассмотрим описание заболевания и известных ключевых белков муковисцидоза [1], рассмотрим их известную трехмерную структуру и белок-белковые взаимодействия.

ПРИМЕР АНАЛИЗА ГЕНОВ: КИСТОЗНЫЙ ФИБРОЗ (МУКОВИСЦИДОЗ)

Кистозный фиброз (муковисцидоз, MB, Cystic fibrosis, CF), код МКБ-10: E84 — это тяжелое системное моногенное заболевание с аутосомно-рецессивным типом наследования, вызванное мутациями в гене *CFTR* (*Cystic Fibrosis Transmembrane Conductance Regulator - Трансмембранный Регулятор Проводимости Муковисцидоза*). При этом заболевании происходит поражение всех экзокринных желёз, а также внутренних органов и систем, в первую очередь дыхательной, пищеварительной и репродуктивной [1].

Ген *CFTR* имеет длину около 215 kb, находится в локусе 7q31.2 и содержит 27 экзонов. Этот ген кодирует белок CFTR, представляющий собой селективный ионный канал и относящийся к подсемейству ABCC суперсемейства ABC-транспортеров (ATP-binding cassette transporter superfamily, ABC) [1, 2, 3, 4, 5, 11]. CFTR локализован на плазматической мембране в апикальной части эпителиальных клеток, образующих выстилку экзокринных желез, в частности желез бронхов, кишечника, урогенитального тракта, потовых и слюнных желез, поджелудочной железы, и играет важную роль в регуляции транспорта ионов Cl^- и HCO_3^-, а также воды между клетками и межклеточной жидкостью, обеспечивая поддержание водного баланса [4]. Мутации в гене *CFTR* вызывают образование нефункционального белка CFTR или нарушают его транспорт к апикальной мембране эпителиоцитов [8]. Вследствие этого нарушается их работа, в том числе мукоцилиарный транспорт, приводя к повышению концентрации хлоридов в поте и сбоям в работе жизненно важных внутренних органов, в особенности легких, поджелудочной железы и кишечника [5]. У пациентов с муковисцидозом в бронхах и протоках экзокринных желез накапливается вязкий секрет, нарушается микробиом дыхательных путей, развивается дыхательная и сердечно-сосудистая недостаточность, снижается толерантность к инфекциям, что существенно снижает качество и продолжительность жизни пациентов [8, 9].

Помимо прочего, муковисцидоз считается одним из наиболее часто встречающихся наследственных заболеваний, с частотой от 1:600 до 1:17000 среди новорожденных европеоидов [1]. Согласно данным «Регистра пациентов с муковисцидозом в Российской Федерации» частота заболевания в России составляет 1:9000 новорожденных, причем

наиболее частым является патогенный вариант *CFTR* F508del (dbSNP: rs113993960), с аллельной частотой 51,40% [1, 8].

С момента первого описания этого заболевания в 1938 году и идентификации гена *CFTR* в 1989 г. были достигнуты значительные успехи в понимании этиологии и патогенеза муковисцидоза [6]. Ведется активная разработка модуляторы *CFTR* — препаратов, которые непосредственно влияют на мутантный белок CFTR [7, 8, 9]. Также совершенствуется поддерживающая терапия. Благодаря этому для некоторых пациентов, получающих лечение, удается достичь продолжительности жизни в 72 года, что почти вдвое превышает такой показатель без применения модуляторов CFTR [8, 9]. Однако, проблема муковисцидоза остается актуальной, значительное количество пациентов по всему миру продолжают страдать от симптомов и осложнений этой болезни, а также сопутствующих заболеваний, сохраняется высокий риск преждевременной смерти.

Список генов, ассоциированных с муковисцидозом был составлен по записям из открытых баз данных, после удаления повторов осталось 4219 генов.

РАСЧЁТ И АНАЛИЗ КАТЕГОРИЙ ГЕННЫХ ОНТОЛОГИЙ ГЕНОВ МУКОВИСЦИДОЗА

Был произведен GO Enrichment Analysis и выполнен расчет категорий генных онтологий биологических процессов (GO biological process complete) и молекулярных функций (GO molecular function complete) для найденных генов муковисцидоза.

Наиболее значимыми категориями биологических процессов, в которых задействованы продукты ассоциированных с муковисцидозом генов, являются:

– реакции на различные воздействия (response to chemical (GO:0042221), response to stress (GO:0006950), response to stimulus (GO:0050896), cellular response to chemical stimulus (GO:0070887), response to oxygen-containing compound (GO:1901700), response to external stimulus (GO:0009605)),

– транспортные процессы (regulation of localization (GO:0032879), regulation of transport (GO:0051049), establishment of localization (GO:0051234), regulation of multicellular organismal process (GO:0051239), transport (GO:0006810)),

– процессы, связанные с иммунными реакциями (immune system process (GO:0002376), defense response (GO:0006952), immune response (GO:0006955), inflammatory response (GO:0006954), regulation of immune system process (GO:0002682), response to cytokine (GO:0034097), response to bacterium (GO:0009617))

– регуляторные процессы (regulation of multicellular organismal process (GO:0051239), positive regulation of biological process (GO:0048518), positive regulation of transport (GO:0051050), homeostatic process (GO:0042592))

Важно разрабатывать новые более эффективные методы патогенетического лечения. Реконструкция генной сети и определение кластеров может помочь установить новые биомаркеры муковисцидоза и потенциальные мишени для новых модуляторов CFTR.

РЕКОНСТРУКЦИЯ И АНАЛИЗ ГЕННОЙ СЕТИ МУКОВИСЦИДОЗА

Была реконструирована сеть, состоящая из 495 узлов и 657 ребер, имеющая среднюю степень связности узла - 2.65, коэффициент кластеризации - 0.376. Также были найдены естественные кластеры (рис. 1).

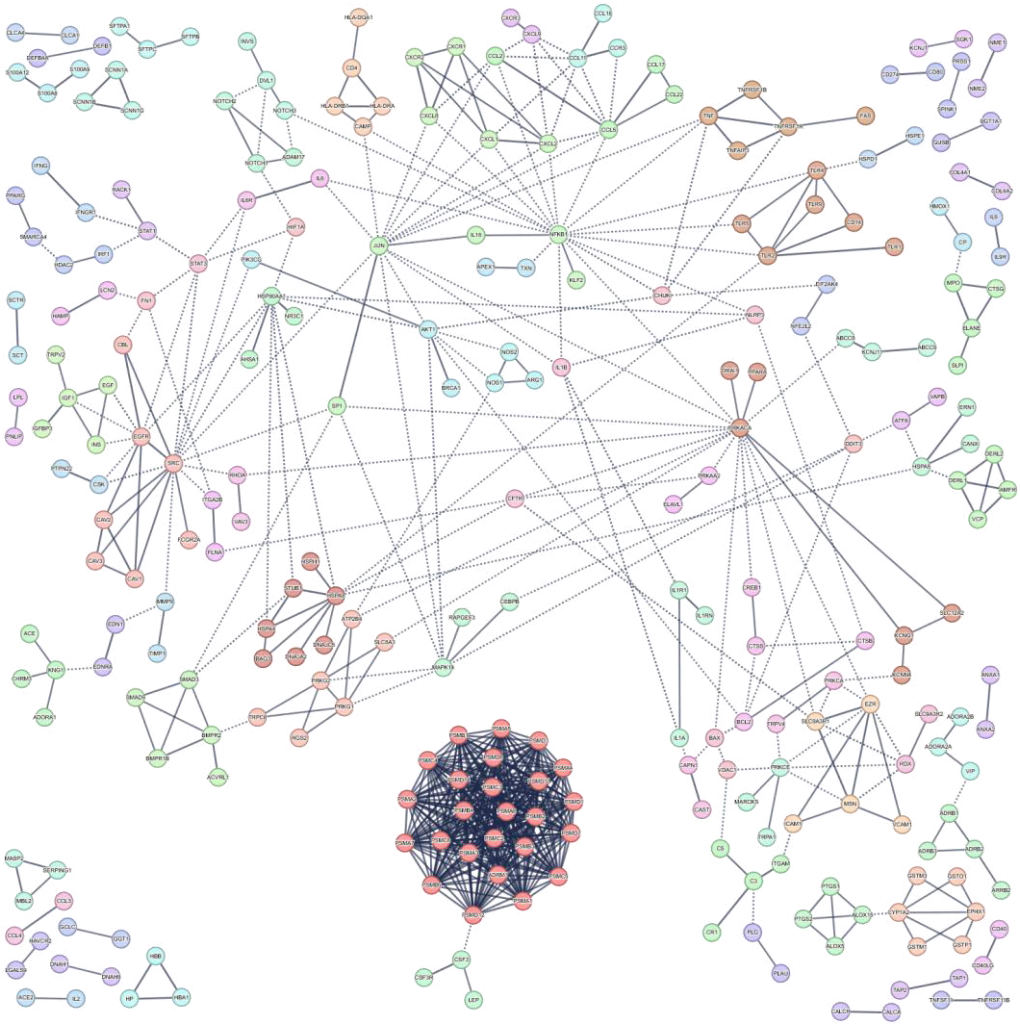


Рисунок 1. Сеть генов, ассоциированных с муковисцидозом, и естественные кластеры

Наибольшую степень узла имеют следующие гены: *PSMD12, ADRM1, PSMA1 - PSMA7, PSMB1 - PSMB5, PSMC2 - PSMC6, PSMD1, PSMD11, PSMD14, PSMD3, PSMD7, PSMD8, NFKB1, PRKACA, SRC, JUN, EGFR, HSP90AA1, HSPA8*. Ген CFTR связан с генами *PRKACA*, *PRKAA2*, *FLNA*, *HSPA8*, *SLC9A3R1* (синонимы *EBP50, NHERF, NHERF1*). Роль найденных генов в патогенезе муковисцидоза изучена плохо, опубликовано мало работ на эту тему.

PRKACA (Protein Kinase CAMP-Activated Catalytic Subunit Alpha) — одна из каталитических субъединиц протеинкиназы А, основного эффектора цАМФ-сигнализации во всех тканях, которая также обеспечивает открытие канала CFTR путем фосфорилирования R-домена [13, 14].

PRKAA2 (Protein Kinase AMP-Activated Catalytic Subunit Alpha 2) — каталитическая субъединица AMP-активируемой протеинкиназы (AMPK), имеются данные о её прямой связи с CFTR: AMPK способна in vitro фосфорилировать CFTR, и кроме того она, как и CFTR, локализована в апикальной части эпителиоцитов (PMID: 10862786) [16].

FLNA (Filamin A) — мультидоменный адаптер цитоскелета, который образует поперечные связи между актиновыми филаментами и соединяет их с мембранными гликопротеинами, связывается участком на N-конце CFTR (PMID: 23636454) [17]. FLNA также напрямую взаимодействует с RACK1 (Receptor For Activated C Kinase 1), что стабилизирует CFTR на поверхности клеток, нарушение взаимодействия между RACK1 и FLNA, а также CFTR и FLNA снижает количество CFTR на мембране [17].

HSPA8 (Heat Shock Protein Family A (Hsp70) Member 8) — относится к белкам теплового шока и выполняет функцию шаперона. Установлено, что подавление экспрессии HSPA8 с помощью алкалоида мартина (Matrine) повышает высвобождение CFTR с мутацией F508del из эндоплазматического ретикулума в цитозоль, и, как следствие, уровень белка CFTR на клеточной мембране (PMID: 22170045) [42].

SLC9A3R1 (Sodium-hydrogen antiporter 3 regulator 1, NHERF1) — каркасный белок (scaffolding protein), регулятор натрий-водородного антипортера 3 (Sodium–hydrogen antiporter 3, sodium–hydrogen exchanger 3 (NHE3), solute carrier family 9 member 3 (SLC9A3)), связывает различные белки, находящиеся на плазматической мембране, с представителями семейства эзрин/моэзин/радиксин (ezrin/moesin/radixin), что обеспечивает их взаимодействие с актиновым цитоскелетом и модулирует их экспрессию на мембрану. SLC9A3R1 стабилизирует CFTR на плазматической мембране, избыточная экспрессия SLC9A3R1 способствует переходу CFTR из цитоплазмы на плазматическую мембрану и восстанавливает активность CFTR с мутацией F508del за счёт реорганизации актинового цитоскелета путём образования мультибелкового комплекса NHERF1-RhoA-ROCK-Ezrin-

actin (PMID: 33530577) [43]. CFTR и SLC9A3R1, а также SLC26A3 и SLC26A6 взаимодействуют друг с другом и играют важную роль в регуляции капацитации (capacitation) сперматозоидов (PMID: 21976599) [44].

Было найдено 84 кластера, состоящих из более чем одного гена, примечательно, что ген *CFTR* не вошел в кластер, состоящий из нескольких генов. Наибольший кластер состоит из 25 генов (рис. 2). Он включает компоненты протеасом: *ADRM1* (Proteasomal ubiquitin receptor, component of the 26S proteasome), *PSMA1 - PSMA7* (альфа субъединицы протеасом 20S), *PSMB1 - PSMB5* (бета-субъединицы протеасом 20S), *PSMC2 - PSMC6* (регуляторные субъединиты протеасом 26S), *PSMD1, PSMD11, PSMD12, PSMD14, PSMD3, PSMD7, PSMD8* (субъединицы протеасом 26S, не обладающие АТФ-азной активностью (26S proteasome non-ATPase regulatory subunit)).

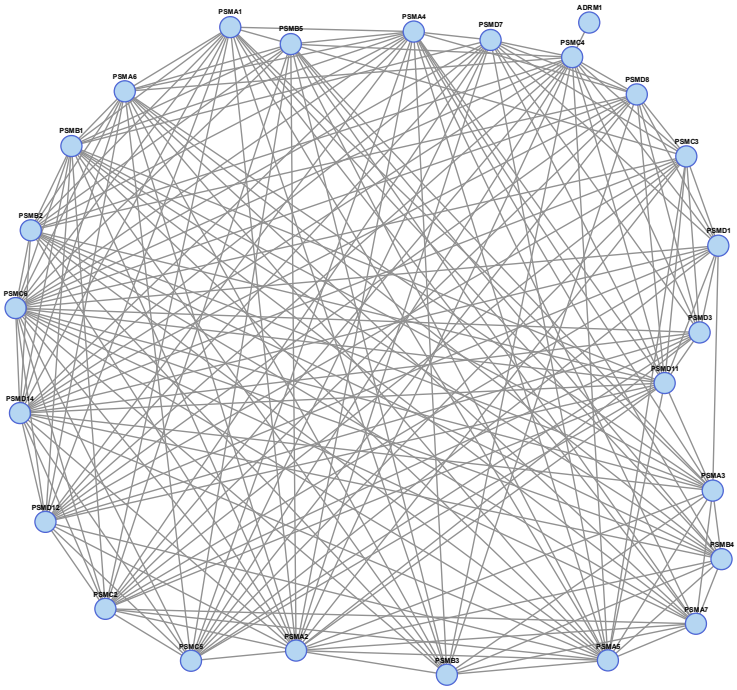


Рисунок 2. Кластер 1, наибольший кластер взаимодействующих генов в генной сети муковисцидоза.

Другие кластеры сети включают:

– гены кавеолинов (*CAV1, CAV2, CAV3*) и гены, вовлеченные в ответ на бактериальную инвазию в эпителиальные клетки. *CBL* (E3 ubiquitin-protein ligase CBL, адаптерный белок, негативный регулятор многих сигнальных путей, переносит убиквитин на субстраты, подлежащие деградации протеасомой, распознает активированные рецепторные тирозинкиназы (KIT, FLT1, FGFR1, FGFR2, PDGFRA, PDGFRB, EGFR, CSF1R, EPHA8 и KDR) и прекращает передачу сигналов), *EGFR* (рецептор эпидермального фактора роста), *FCGR2A* (низкоаффинный рецептор IgG),

SRC (протоонкогенная тирозин-протеинкиназа, участвует в сигнальных путях, контролирующих широкий спектр биологических процессов, включая транскрипцию, иммунный ответ, клеточную адгезию, прогрессирование клеточного цикла, апоптоз, миграцию и трансформацию)

Кластеры генной сети муковисцидоза включают

– гены, вовлечённые в шаперон-опосредованную аутофагию. *BAG3* (ко-шаперон шаперонов HSP70 и HSC70)*, DNAJA2* (один из шаперонов Hsc70)*, DNAJC5* (ко-шаперон SNAP-25, участвует в кальций-опосредованном контроле поздней стадии экзоцитоза)*, HSPA4, HSPA8, HSPH1, STUB1*.

Другие кластеры включают (рис. 3):

– гены кавеолинов (*CAV1, CAV2, CAV3*) и гены, вовлеченные в ответ на бактериальную инвазию в эпителиальные клетки. *CBL* (E3 ubiquitin-protein ligase CBL, адаптерный белок, негативный регулятор многих сигнальных путей, переносит убиквитин на субстраты, подлежащие деградации протеасомой, распознает активированные рецепторные тирозинкиназы (KIT, FLT1, FGFR1, FGFR2, PDGFRA, PDGFRB, EGFR, CSF1R, EPHA8 и KDR) и прекращает передачу сигналов), *EGFR* (рецептор эпидермального фактора роста), *FCGR2A* (низкоаффинный рецептор IgG), *SRC* (протоонкогенная тирозин-протеинкиназа, участвует в сигнальных путях, контролирующих широкий спектр биологических процессов, включая транскрипцию, иммунный ответ, клеточную адгезию, прогрессирование клеточного цикла, апоптоз, миграцию и трансформацию)

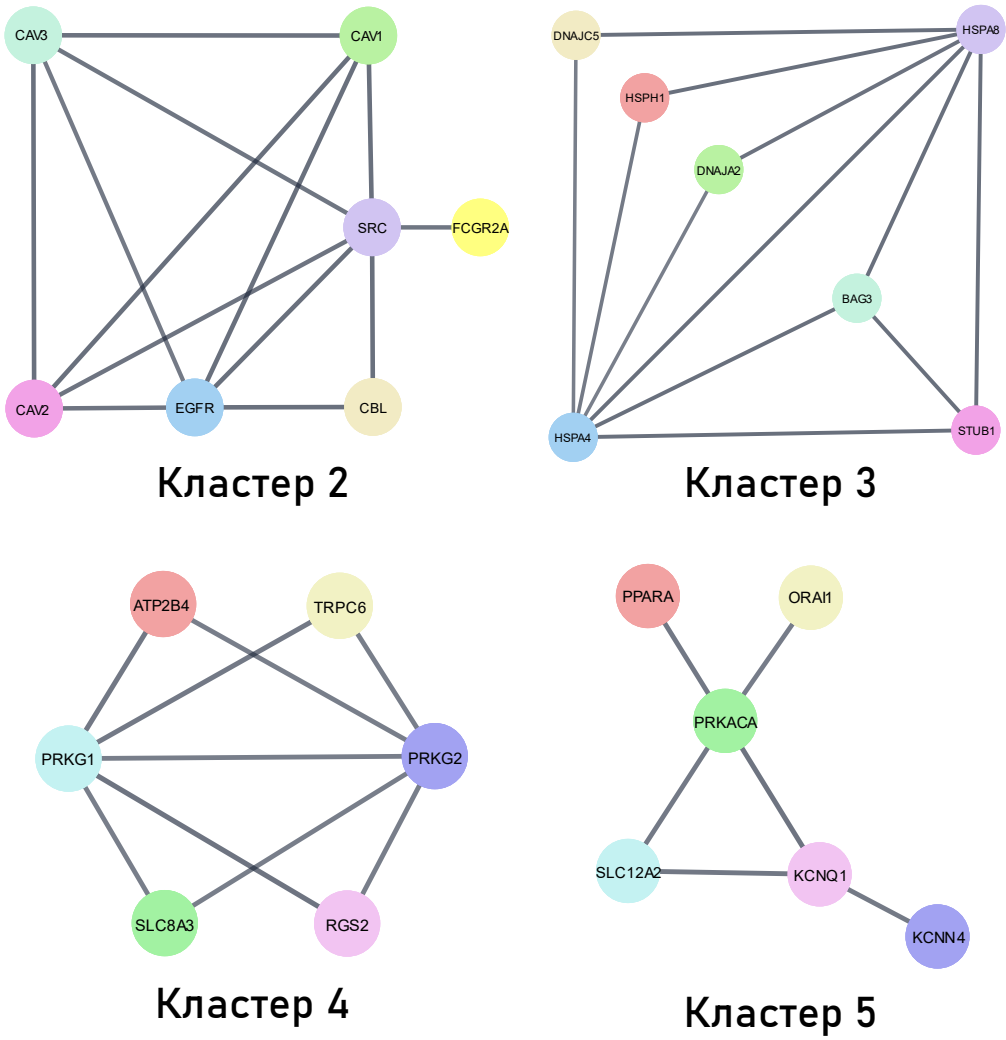


Рисунок 3. Найденные кластеры генной сети муковисцидоза. Ключевыми генами являются EGFR, HSPA4, PRKG1, PRKACA.

Кластеры генной сети муковисцидоза включают

– гены, вовлечённые в шаперон-опосредованную аутофагию. *BAG3* (ко-шаперон шаперонов HSP70 и HSC70), *DNAJA2* (один из шаперонов Hsc70), *DNAJC5* (ко-шаперон SNAP-25, участвует в кальций-опосредованном контроле поздней стадии экзоцитоза), *HSPA4, HSPA8, HSPH1, STUB1*.

БЕЛОК CFTR, ЕГО ФУНКЦИИ И РОЛЬ В ПАТОГЕНЕЗЕ МУКОВИСЦИДОЗА

Белок CFTR локализован на плазматической мембране в апикальной части эпителиальных клеток, образующих выстилку экзокринных желез, в частности потовых, слюнных, желез бронхов, поджелудочной железы, желез кишечника, урогенитального

тракта. Он выполняет функцию селективного ионного канала и участвует в регуляции транспорта ионов Cl^- и HCO_3^-, а также воды между клетками и межклеточной жидкостью и, как следствие в регуляции процессов поглощения и секреции [1, 4, 13]. Ион HCO_3^-, играет важную роль в поддержании гомеостаза; внутриклеточного и внеклеточного pH дыхательных путей [4, 13, 24].

В ответ на связывание АТФ, CFTR изменяет свою конформацию и открывается, что позволяет ионом Cl^- пассивно перемещаться через мембрану, далее после гидролиза АТФ канала закрывается [4, 13, 20, 21, 23]. CFTR является единственным каналом семейства ABC, обеспечивающим пассивный транспорт ионов [25].

CFTR имеет доменную структуру и содержит два нуклеотид-связывающих (NBD1 и NBD2) и два трансмембранных (TMD1 и TMD2) домена, NBD1 непосредственно связан с TMD1, который находится на N-конце и начинается с мотива «лассо» (lasso motif), содержащего две α-спирали со стороны цитоплазмы наполовину погруженные в плазматическую мембрану и α-спираль, обернутую вокруг TMD2, который связан с NBD2, находящемся на C-конце. В отличие от других белков семейства ABC, CFTR имеет цитоплазматический неструктурированный регуляторный домен (R-домен) между NBD1 и TMD2 [20]. Каждый из трансмембранных доменов, состоит из шести гидрофобных трансмембранных α-спиралей (TM, TM). Спирали TM1-TM6 находятся в трансмембранном домене 1 (TMD1), а TM7-TM12 — в TMD2. Трансмембранные домены образуют ионный канал белка CFTR, а NBD обеспечивают его открытие и закрытие в ответ на связывание и гидролиз АТФ, R-домен разделяет TMD-домены [4, 20, 21].

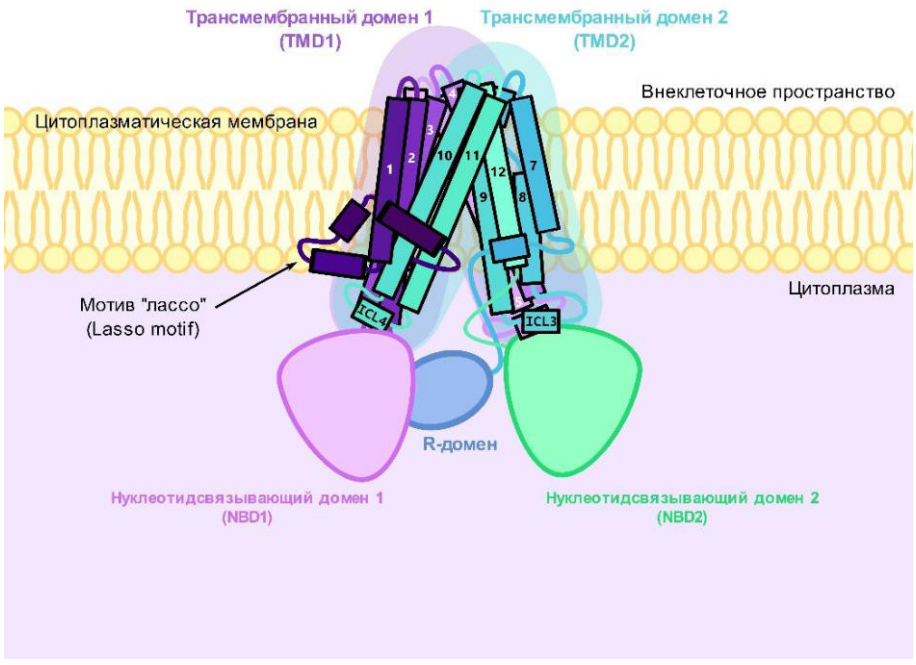


Рисунок 4. Структура дефосфорилированного, не связанного с АТФ CFTR. Адаптированно из [25].

В случае, если R-домен дефосфорилирован, и при отсутствии АТФ, CFTR образует закрытую конформацию, в которой NBD отдалены друг от друга примерно на 20 Å, а R-домен стерически препятствует их сближению [21]. Открытие канала CFTR осуществляется в ответ на фосфорилирование R-домена протеинкиназой А (ПКА, PKA) и связывание АТФ с NBD, что приводит к димеризации доменов NBD1 и NBD2, изменению конформации TMD доменов из обращенной внутрь, в конформацию, обращенную наружу, и выходу ионов Cl^- из эпителиальных клеток. После гидролиза АТФ приводит расхождение димера NBD, восстановление конформации TMD, и канал CFTR закрывается [20, 21].

Наличие ионов Cl^-, HCO_3^- и Na^+ в межклеточном пространстве возле апикальной поверхности эпителиальных клеток является необходимым условием поддержания нормального осмотического давления и необходимо для циркуляции жидкости около плазматических мембран клеток [13, 23]. Кроме того, для нормального функционирования клеток необходим постоянный и контролируемый ток этих ионов через мембрану, который обеспечивают мембранные транспортеры, к которым относиться CFTR. Этот канал регулирует транспорт ионов в мерцательном эпителии дыхательных путей и подслизистых железах, как следствие количество жидкости на поверхности дыхательных путей и степень вязкости слизи [24].

Таблица 1. Топ 15 наиболее распространенных вариантов гена CFTR по данным «Регистра пациентов с муковисцидозом в Российской Федерации. 2023 год» [8]

№	Традиционное название генетического варианта	Название по кодирующей ДНК	Номер в dbSNP	Аллельная частота, %	Класс
1	F508del	c.1521_1523delCTT	rs113993960	51,4	II
2	CFTRdele2,3	c.54-5940_273+10250del21kb	Не найден	6,1	I
3	E92K	c.274G>A	rs121908751	3,71	IV(?)
4	1677delTA	c.1545_1546delTA	rs121908776	2,51	I
5	3849+10kbC->T	c.3718-2477C>T	rs75039782	2,24	V
6	2143delT	c.2012delT	rs121908812	1,99	I
7	2184insA	c.2052_2053insA(c.2052dupA)	rs121908786	1,93	I
8	W1282X	c.3846G>A	rs77010898	1,73	I
9	L138ins	c.413_415dupTAC	rs397508679	1,71	IV
10	N1303K	c.3909C>G	rs80034486	1,62	II
11	G542X	c.1624G>T	rs113993959	1,47	I
12	394delTT	c.262_263delTT	rs121908769	0,92	I
13	L467F;F508del	c.[1399C>T;1521_1523delCTT]	Не найден	0,87	?

14	R334W	c.1000C>T	rs121909011	0,75	IV
15	W1282R	c.3844T>C	rs121908805	0,63	II

Снижение концентрации ионов Cl^- около поверхности эпителиоцитов, которое происходит при нарушении работы CFTR, активирует эпителиальный натриевый канал (epithelial sodium channel, ENaC), закачивающий ионы Na^+ внутрь клеток. Недостаток ионов Na^+ и Cl^- в межклеточный жидкости около эпителиоцитов приводит к снижению объема воды, поступающей клетке, что в свою очередь приводит к образованию более густого и вязкого секрета, утрате способности дыхательных путей к самоочищению, из-за нарушений работы ресничек клеток слизистой оболочки. Происходит застой слизи в просветах желез, что нарушает их работу и снижает сопротивляемость организма пациентов инфекциям [13].

ОБСУЖДЕНИЕ ПЕРСПЕКТИВ ТЕРАПИИ МУКОВИСЦИДОЗА

В настоящее время достигнут значительный прогресс в изучении механизмов развития и разработки терапии против муковисцидоза, что существенно повысило продолжительность и качество жизни пациентов. Благодаря правильно подобранной терапии средняя продолжительность жизни с данным заболеванием составляет 40, а в некоторых случаях более 50 лет [17]. Управление по санитарному надзору за качеством пищевых продуктов и медикаментов США (FDA) уже одобрило стратегию терапии муковисцидоза, основанную на применении малых молекул – модуляторов CFTR, способных в определенной степени восстанавливать его работу при наличие наиболее распространённых мутаций II-VI классов, встречающихся у 90% пациентов. Тем не менее 10% мутаций, вызывающих муковисцидоз, резистентны к модуляторам CFTR, поскольку такие варианты характеризуются полным отсутствием синтеза белка CFTR или настолько выраженными дефектами структуры, что мутантный белок не способен осуществлять транспорт ионов [2]. А также известно, что 10–20% пациентов имеют индивидуальную непереносимость этих препаратов [10, 13]. Модуляторы CFTR классифицируются по механизму действия следующим образом: потенциаторы (potentiators), корректоры (correctors), стабилизаторы (stabilizers), амплификаторы (amplifiers) и вещества, способствующие «прочтению» преждевременных стоп-кодонов (read-through agents, through agents) [5, 7, 10]. На сегодняшний день клинически одобрены четыре модулятора CFTR: потенциатор ивакафтор (VX-770) в качестве монотерапии или как часть комбинированной терапии с корректорами: лумакафтором (VX-809), тезакафтором (VX-661) и элексакафтором (VX-445) [18].

К ним относятся соединения, которые повышают вероятность открытия канала CFTR, увеличивая тем самым опосредованный CFTR ток ионов через плазматическую мембрану. Потенциаторы эффективны против мутаций III и IV класса (R117H, R334W, R347P и G551D), которые встречаются не часто и нарушают работу CFTR, но не снижают уровень его экспрессии. Эти соединения связываются с CFTR в определенных сайтах, способствуя правильному открытию канала, повышают активность CFTR, уже находящегося на мембране. Примерами потенциаторов являются ивакафтор и дирокафтор [5, 7, 10].

Ивакафтор (Ivacaftor, Kalydeco™, VX-770, PubChem CID: 16220172), выпускаемый немецкой компанией «Vertex Pharmaceuticals» — первый патогенетический препарат против муковисцидоза, был одобрен FDA и EMA в 2012 году [7]. Он обеспечивает правильное открытие канала CFTR, стабилизируя открытую конформацию, тем самым увеличивая секрецию ионов Cl^- и бикарбонатов. Данный препарат доказал свою безопасность и эффективность. Примечательно, что в 2010-х годах данный препарат применяется для лечения пациентов старше 2 лет с мутацией p.G551D, тогда, как с 2017 года FDA одобрило его применения при более широком спектре вариантов, включающем 38 мутаций.

Точный механизм действия ивакафтора долгое время оставался до конца не известным. В 2019 году с помощью криоэлектронной микроскопии была определена структура комплекса CFTR с ивакафтором (PDB: 6O2P), что позволило установить два сайта связывания данного препарата (рис. 5). Ивакафтор связывается с CFTR внутри липидного бислоя между TMD1 и TMD2, ключевую роль в этом играют аминокислотные остатки Phe931 и Arg933, а также в область внутриклеточной петли ICL4 (intercellular loop 4) на границе NBD1 и TMD2.

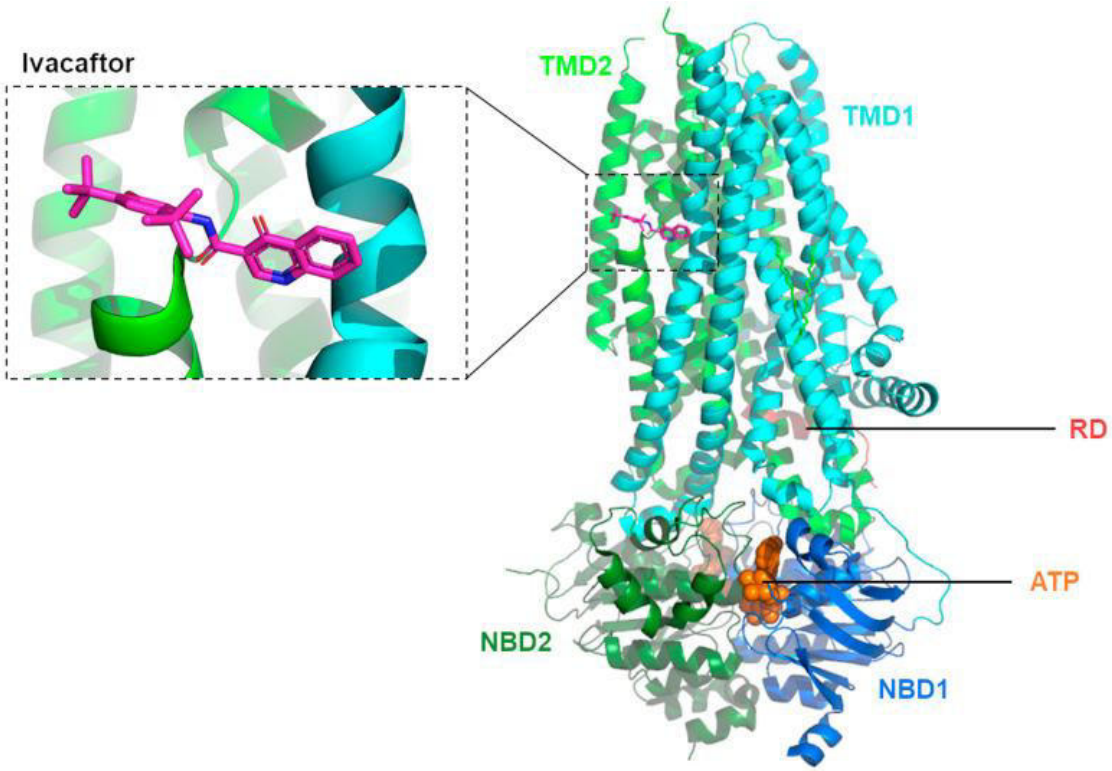


Рисунок 5. Структура фосфорилированного, связанного с АТФ белка CFTR человека в комплексе с Ivacaftor.

Однако применение только Ivacaftor для лечения пациентов, с мутацией F508del в гомозиготном состоянии оказалось недостаточно эффективным, из-за нехватки CFTR на мембране клеток. В таких случаях необходима комбинированная терапия, включающая корректор CFTR.

Соединения, которые улучшают фолдинг CFTR и, следовательно, облегчают процессинг, а также транспортировку на апикальную мембрану клеток, относятся к корректорам [5, 7, 10]. Нарушение этих процессов наблюдается при наличии наиболее распространенных мутаций, вызывающих муковисцидоз, которые относятся ко II классу (F508del, G85E, I507del, R560T и N1303K). Большинство корректоров, действуя по различным механизмам, повышают конформационную стабильность мутантного CFTR в процессе фолдинга в ЭПР, выступая в роли фармакологических шаперонов.

Перспективами терапии являются генноинженерные подходы, такие как генная заместительная терапия и редактирование генома. Они нацелены на коррекцию любых мутаций и в перспективе способны обеспечить полное выздоровление. Несмотря на значительный прогресс в области генетических технологий, пока не удается создать генную терапию муковисцидоза, на сегодняшний день нет ни одного одобренного генно-терапевтического препарата. Однако, активно ведутся исследования в этом направлении, и в будущем ожидается появление таких лекарств. Современные подходы к лечению не

эффективны против всех мутаций, некоторые пациенты плохо переносят доступные лекарства. Модуляторы CFTR не обеспечивают стойкого эффекта и не устраняют первопричину — генные мутации, вследствие чего больные вынуждены пожизненно принимать дорогие медикаменты. Муковисцидоз продолжает оставаться тяжелым и неизлечимым недугом, поражающим тысячи людей ежегодно, высокая стоимость препаратов затрудняет доступ к адекватному лечению и оказывает значительное давление на системы здравоохранения различных стран.

При этом активно совершенствуется стратегия терапии муковисцидоза с помощью модуляторов CFTR, ведутся поиски новых мишеней, разрабатываются новые действующие вещества. Биоинформатические подходы могут ускорить процесс поиска новых мишеней и прогностических биомаркеров, генные сети позволяют оценить функциональные связи между различными белками, ассоциированными с данным заболеванием.

Рассмотрим описание заболевания и известных ключевых белков CMA1, проведем анализ категорий генных онтологий, реконструировать и проанализировать генную сеть.

СПИНАЛЬНАЯ МЫШЕЧНАЯ АТРОФИЯ I ТИПА

Спинальная мышечная атрофия I типа (CMA1, болезнь Верднига-Гоффмана, Spinal muscular atrophy Type I, SMA1), код МКБ-10 G12.0, OMIM #253300 – тяжелое прогрессирующее моногенное нервно-мышечное заболевание с аутосомно-рецессивным типом наследования (AR), возникающее в результате потери или дисфункции Гена выживаемости мотонейронов 1 (Survival motor neuron 1, *SMN1)*, и приводящее к постепенной деградации и гибели мотонейронов в передних рогах спинного мозга.

Данная патология является наиболее тяжелой формой спинальной мышечной атрофии. Манифестирует в возрасте до 6 месяцев и характеризуется выраженной атрофией мышц, гипотонией, отсутствием спонтанных движений, а также снижением рефлексов. Дети с таким диагнозом не могут самостоятельно сидеть и держать голову, имеют трудности с глотанием и дыханием. Атрофия межреберных мышц и диафрагмы приводит к колоколообразной деформации грудной клетки и дыхательной недостаточности, пациентам часто требуется искусственная вентиляция легких. Без поддерживающей терапии продолжительность жизни с детской спинальной мышечной атрофией I типа в среднем не превышает двух лет, самой распространенной причиной смерти является дыхательная недостаточность, также высок риск аспирационной пневмонии.

Считается, что спинальная мышечная атрофия является ведущей генетической причиной младенческой смертности – это одно из самых распространенных орфанных

заболеваний. Частота встречаемости СМА1 составляет примерно 1 на 10 000 живорожденных, с частотой носительства от 1:38 до 1:70.

Клинические проявления

Клиническая картина СМА1 обусловлена проявлениями избирательной дегенерации двигательных нейронов передних рогов спинного мозга в сочетании с рядом сопутствующих патологий, что свидетельствует о системной значимости белка SMN.

Для детской спинальной мышечной атрофии I типа характерно резкое несоответствие моторной функции возрастной норме, включающее:

- Нервно-мышечные нарушения: выраженная атрофия мышц; генерализованная мышечная гипотония; снижение или отсутствие спонтанных движений; симптомокомплекс «вялого» ребенка; гипорефлексия, вплоть до арефлексии; нарушения бульбарной функции, слабый крик и кашель; фасцикуляции языка и фасцикуллярные подергивания в мышцах спины, туловища, проксимальных, реже в дистальных, отделов верхних и нижних конечностей; замещение мышечной ткани фиброзно-жировой.
- Изменения костно-суставной системы: колоколообразная деформация грудной клетки, зонтикообразная деформация ребер, деформация позвоночника (кифосколиоз), контрактуры суставов, патологии стоп.
- Нарушения дыхательной функции: дыхательная недостаточность; нарушения откашливания; гиповентиляции во время ночного сна; высокий риск аспирационной пневмонии и инфекционных заболеваний.
- Дисфункция ЖКТ: нарушения глотания, в том числе из-за бульбарного синдрома; нарушения моторики ЖКТ, среди которых запоры, задержка эвакуации содержимого желудка и гастроэзофагальный рефлюкс.
- Болевой синдром, как следствие патологий опорно-двигательного аппарата, остеопении и переломов.
- Нарушения сердечно-сосудистой системы, врожденные пороки сердца;

При этом для СМА1 не характерны нарушение чувствительности и задержка психического развития.

Этиология и патогенез

В 1990-х годах при помощи различных методов генетического и физического картирования было показано, что ген, ответственный за развитие спинальной мышечной атрофии, находится в локусе 5q13.2, то есть на длинном плече 5-ой хромосомы.

В этом локусе (5q12.2-q13.3) находится инвертированная дупликация длиной около 500 kbp, в которой закодированы 4 гена:

- SMN1 (survival motor neuron 1),
- NAIP (neuronal apoptosis inhibitor protein),
- GTF2H2A (general transcription factor IIH),
- SERF1A (small EDRK-rich factor 1A),

а также их копии, причем:

- SERF1B идентичен по нуклеотидной последовательности гену SERF1A,
- SMN2, также как SMN1, состоит из девяти экзонов (1, 2a, 2b, 3-8), но отличается от него пятью нуклеотидами и представлен несколькими копиями,
  чаще всего от 1 до 5, количество которых определяет тяжесть и тип СМА,
- копии генов GTF2H2A и NAIP являются псевдогенами ΨGTF2H2B и ΨNAIPΔ5, соответственно.

Причиной детской спинальной мышечной атрофии I типа являются мутации в гене SMN1, который в норме кодирует белок выживаемости мотонейронов (SMN). Данный белок массой 38 кДа локализован как в ядре, так и в цитоплазме клеток различных тканей в составе мультибелковых SMN комплексов (SMN complex), в которых несколько SMN белков связаны между собой и с другими белками. Нахождение в составе комплексов позволяет SMN выполнять множество функций. При этом в двигательных нейронах спинного мозга особенно много белка SMN. Показано, что SMN участвует в транспортировке зрелых мРНК, что имеет критическое значение для мотонейронов, имеющих очень длинные аксоны, поскольку нарушение транспорта мРНК в конец аксона приводит к гибели нейронов. В настоящее время одна из наиболее изученных функций белка SMN – это участие в сборке сплайсосом: в ядре SMN участвует в биогенезе малых ядерных рибонуклеопротеинов (мяРНП, small nuclear ribonucleic particles, snRNP), типа U1-5, U11 и U12, в цитоплазме он обеспечивает их включение в состав сплайсосом. Короткая аксональная изоформа a-SMN, отличающаяся от SMN вхождением только первых четырех экзонов, важна для аксоногенеза. Кроме того, имеются данные, что SMN участвует в процессах транскрипции и трансляции, а также необходим для нормального функционирования митохондриальной дыхательной цепи в двигательных нейронах и мышцах.

Наиболее значимое отличие паралогичных копий SMN1 и SMN2, лежащих в теломерной и центромерной области длинного плеча 5-ой хромосомы, соответственно, заключается в замене одного нуклеотида в экзоне 7 гена SMN2 (840C>T), которая приводит к преимущественному удалению 7 экзона в процессе сплайсинга и синтезу нестабильного белка SMNΔ7, который не способен к образованию функциональных SMN комплексов. Интересно, что замета 840C>T, единственная из пяти, отличающих ген SMN2 от SMN1, находится в кодирующей области гена, но при этом является синонимичной и не приводит к изменению в аминокислотной последовательности белка SMN, который ген SMN2 способен кодировать также как SMN1. Поэтому примерно из 10–20% продуктов гена SMN2, когда 7 экзон сохраняется при сплайсинге, получается функциональный белок SMN,

это объясняет модифицирующую роль гена SMN2, когда количество его копий определяет тяжесть СМА.

У большинства пациентов с СМА1 имеется одна или две копии SMN2, с СМА2 и СМА3 обычно – не менее трех копий SMN2, с СМА4 их обычно не менее четырех. Известны случаи бессимптомного носительства гомозиготных мутаций в гене SMN1 при наличии нескольких копий SMN2.

У 95% пациентов с СМА1 в гене SMN1 присутствуют делеции в 7 или в 7 и 8 экзоне, и примерно у 2% больных такие мутации возникли *de novo*, из-за нестабильности хроматина в этом районе. Это в большинстве случаев приводит к нарушению сплайсинга и синтезу укороченного нефункционального транскрипта.

Методы: Для составления списка генов использовались интернет-ресурс NCBI, базы данных GeneCards и MalaCards. Также в список добавлены гены, модифицирующие эффекты которых описаны в литературных источниках: PLS3, PFN2, ZPR1, CORO1C, GTF2H2, NRN1, SERF1A, NCALD, NAIP, TIA1. Для анализа категорий генных онтологий использовались ресурсы DAVID и PANTHER, для визуализации – инструмент GOST. Реконструкция генной сети выполнена с помощью интернет-ресурса STRING-DB.

МЕТОДИКА. СОСТАВЛЕНИЕ СПИСКА ГЕНОВ

1. Интернет-ресурс NCBI (The National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/gene/).
   Поиск по ключевому слову «Spinal muscular atrophy Type I» для организма «Homo sapiens (human)» дал 5 результатов (SMN1, GDNF, SMN2, HSPG2, SCO2).
2. Интернет-ресурс GeneCards (https://www.genecards.org/).
   Поиск по ключевому слову «Spinal muscular atrophy Type I» дал 315 результатов.
3. Интернет-ресурс MalaCards (https://www.malacards.org/).
   Поиск по ключевому слову «Spinal Muscular Atrophy, Type I» дал 26 результатов, из которых 5 с высокой степенью доказательности.
4. Также в список добавлены гены, модифицирующие эффекты которых описаны в литературных источниках (PLS3, PFN2, ZPR1, CORO1C, GTF2H2, NRN1, SERF1A, NCALD, NAIP, TIA1), причем для генов SERF1A, NAIP, NRN1 и PLS3 подтверждено влияние на фенотип при СМА.

Результаты поиска по интернет-ресурсам и литературным источникам были объединены в текстовый файл, после чего с помощью нескольких строчек кода на Python из файла были исключены повторы.

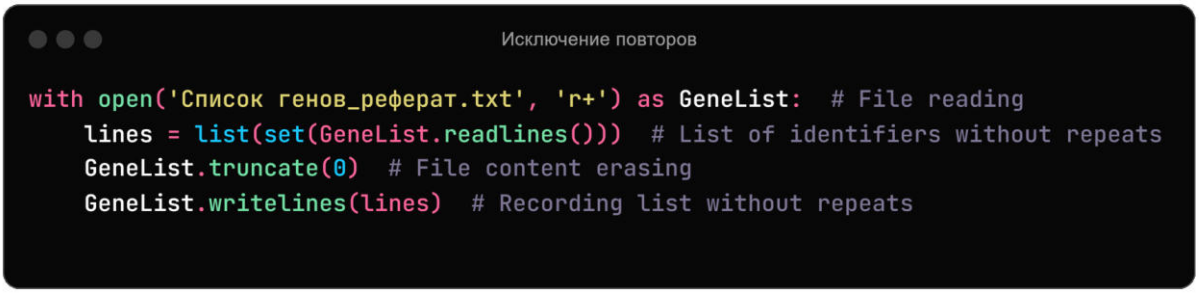


Рисунок 6. Исключение повторов из текстового файла с помощью Python
В результате осталось 319 генов.

РАСЧЁТ И АНАЛИЗ КАТЕГОРИЙ ГЕННЫХ ОНТОЛОГИЙ

Расчёт категорий генных онтологий для составленного списка генов был выполнен с использованием интернет-ресурсов DAVID (Database for Annotation, Visualization and Integrated Discovery, https://david.ncifcrf.gov/) и PANTHER (Protein ANalysis THrough Evolutionary Relationships).

DAVID

Список генов был загружен через интерфейс программы, в качестве организма сравнения выбран человек (Homo sapiens), в качестве идентификаторов были выбраны официальные символы генов «OFFICIAL_GENE_SYMBOL».

Полученная таблица, включает 1178 записей генных отнологий, из которых были отобраны 100 наиболее значимых по P-Value.

По данным таблицы, наиболее значимыми категориями для генов, ассоциированных со спинальной мышечной атрофией, I типа являются:

1. Neurodegeneration (Нейродегенерация). Дефектные продукты данных генов способны вызывать нейродегенерацию.
2. LSM_dom_sf (Суперсемейство доменов LSM). Суперсемейство доменов, встречающееся в качестве основной структуры в белках Lsm (like-Sm) и бактериальных белках Hfq, связанных с Lsm, а также в качестве среднего домена белка механочувствительных каналов MSC, содержит основную структуру, состоящую из открытого β-цилиндра с SH3-подобной топологией. Белки Lsm выполняют разнообразные функции, в том числе являются важными модуляторами биогенеза и функционирования РНК.
3. mRNA splicing (сплайсинг мРНК). Продукты данных генов участвуют в сплайсинге.
4. Sm_dom_euk/arc (Домен Sm, тип эукариот/архей).
5. Sm (Домен Sm). Продукты генов содержат домен, обнаруженный в белках Sm, Lsm (like-Sm) и Hfq (прокариоты), подобные белки участвуют в процессах, связанных с процессингом РНК и регуляцией экспрессии генов. У эукариот

белки Sm и Lsm образуют комплексы и играют важную роль в сплайсинге и декапировании мРНК.

6. mRNA processing (Процессинг мРНК). Продукты генов участвуют в процессах 5'-каппирования и 3'-полиаденилирования.
7. Fibroblast growth factor receptor signaling pathway (Сигнальный путь рецептора фактора роста фибробластов).
8. Fibroblast_GF_fam (Семейство факторов роста фибробластов). Факторы роста фибробластов (FGF) семейство многофункциональных белков, представляющих собой митогены, стимулирующие рост или дифференцировку клеток мезодермального или нейроэктодермального происхождения. FGF участвуют в процессах развития: индукция мезодермы, формирование передне-заднего паттерна, развитие конечностей, а также индукция и развитие нервной системы.
9. SMN-Sm protein complex (SMN-Sm комплексы). Белковые комплексы, образованные несколькими метилированными Sm-белками и SMN комплексом, а также дополнительными белками, среди которых белки Gemin2-8 и unrip, галектин-1 и -3. SMN-Sm комплексы участвуют в сборке snRNP в цитоплазме.
10. spliceosomal snRNP assembly (Сплайсосомальная сборка snRNP). Продукты генов участвуют в образовании сплайсосомы.
11. Cytoplasm (Цитоплазма). Клеточный компартмент
12. FGF (Семейство кислых и основных факторов роста фибробластов). FGF3, FGF4, FGF5, FGF17, FGF6, FGF16, FGF7, FGF8, FGF14, FGF9, FGF18, FGF20, FGF22, FGF10
13. Spinocerebellar ataxia (Спиноцеребеллярная атаксия). Дефектные продукты генов способны вызывать аутосомно-доминантную спиноцеребеллярную атаксию, заболевание клинически и генетически гетерогенной группы аутосомно-доминантных мозжечковых атаксий (ADCA).
14. RNA splicing (Сплайсинг РНК). Продукты генов участвуют в сплайсинге мРНК.
15. Positive regulation of gene expression (Положительная регуляция экспрессии генов).

PANTHER

Список генов был загружен через Gene Ontology Resource (https://www.geneontology.org/), где на основе ресурса PANTHER был произведен GO Enrichment Analysis, был выполнен расчет категорий генных онтологий биологических процессов (GO biological process complete) и молекулярных функций (GO molecular function complete) найденных генов. В качестве референса был выбран геном человека (Homo sapiens (all genes in database)), статистическая значимость оценивалась по точному критерию Фишера (Fisher's Exact), корректировка проводилась по Бонферрони. (Use the Bonferroni correction for multiple testing). Полученные таблицы, включают 421 запись генных онтологий биологических процессов, из которых были отобраны 100 наиболее значимых по P-Value, а также 50 записей генных онтологий молекулярных функций.

По данным таблицы, наиболее значимыми категориями биологических процессов, в которых задействованы гены, ассоциированные со спинальной мышечной атрофией, I типа являются:

1. RNA splicing (Сплайсинг РНК, GO:0008380)
2. Biosynthetic process (Процесс биосинтеза, GO:0009058)
3. mRNA metabolic process (Процесс метаболизма мРНК, GO:0016071)
4. Cellular process (Клеточный процесс, GO:0009987)
5. mRNA processing (Процессинг мРНК, GO:0006397)
6. RNA splicing, via transesterification reactions (Сплайсинг РНК посредством реакций переэтерификации, GO:0000375) Процесс сплайсинга РНК посредством серии из двух реакций переэтерификации с выпуклым остатком аденозина в точке ветвления интрона, который является инициирующим нуклеофилом.
7. Cellular metabolic process (Процесс клеточного метаболизма, GO:0044237)
8. Nucleobase-containing compound biosynthetic process (Процесс биосинтеза соединения, содержащего нуклеиновую основу, GO:0034654) Продукты генов участвуют в реакциях и путях, приводящих к образованию нуклеозидов, нуклеотидов и нуклеиновых кислот.
9. Nucleic acid biosynthetic process (Процесс биосинтеза нуклеиновых кислот, GO:0141187)
10. Cellular biosynthetic process (Процесс клеточного биосинтеза, GO:0044249) Продукты генов участвуют в осуществляемых отдельными клетками реакциях и путях, приводящих к образованию веществ.
11. Metabolic process (Процесс метаболизма, GO:0008152)
12. Gene expression (Экспрессия генов, GO:0010467)
13. mRNA splicing, via spliceosome (Сплайсинг мРНК посредством сплайсосом, GO:0000398)
14. RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (Сплайсинг РНК посредством реакций переэтерификации с выступающим аденозином в качестве нуклеофила, GO:0000377)
15. Cellular response to endogenous stimulus (Клеточный ответ на эндогенный стимул, GO:0071495)

Наиболее значимыми категориями молекулярных функций генов, ассоциированных со спинальной мышечной атрофией, I типа являются:

1. Protein binding (Связывание с белками, GO:0005515)
2. Fibroblast growth factor receptor binding (Связывание с рецептором фактора роста фибробластов, GO:0005104)
3. Growth factor activity (Активность фактора роста, GO:0008083) Продукты генов способны стимулировать рост и пролиферацию клеток.
4. RNA binding (Связывание с РНК, GO:0003723) Продукты генов способны связываться с РНК.
5. Enzyme binding (Связывание с ферментами, GO:0019899)
6. Binding (Связывающая активность, GO:0005488) Продукты генов способны связываться и взаимодействовать с другими агентами (белками, ферментами, нуклеиновыми кислотами и т. д.).
7. Protein-folding chaperone binding (Фолдинг белков, опосредов, GO:0051087) Продукты генов способны связываться с белками-шаперонами.
8. Identical protein binding (Идентичное связывание с белками, GO:0042802) Продукты генов способны связываться с идентичными белками или протеинами, способны к гомополимеризации.
9. Protein-containing complex binding (Связывание белоксодержащего комплекса, GO:0044877)

10. Organic cyclic compound binding (Связывание с органическими циклическими соединениями, GO:0097159)
11. Growth factor receptor binding (Связывание с рецептором фактора роста, GO:0070851)
12. Transcription factor binding (Связывание с факторами транскрипции, GO:0008134)
13. Signaling receptor binding (Связывание сигнального рецептора, GO:0005102) Продукты генов способны связываться с рецепторами внутриклеточных мессенджеров для инициирования изменения функции клетки.
14. Nucleic acid binding (Связывание с нуклеиновыми кислотами, GO:0003676)
15. Receptor ligand activity (Активность лиганда рецептора, GO:0048018). Продукты генов способны взаимодействовать с рецепторами, вызывая изменения клеточных процессов.
16.

ГРАФИК ТОЧЕЧНЫХ ЗНАЧЕНИЙ КАТЕГОРИЙ ГЕННЫХ ОНТОЛОГИЙ

Визуализация генных онтологий осуществлена с помощью инструмента GOST (https://biit.cs.ut.ee/gprofiler/gost) веб-сервера для анализа функционального обогащения и преобразования списков генов g:Profiler.

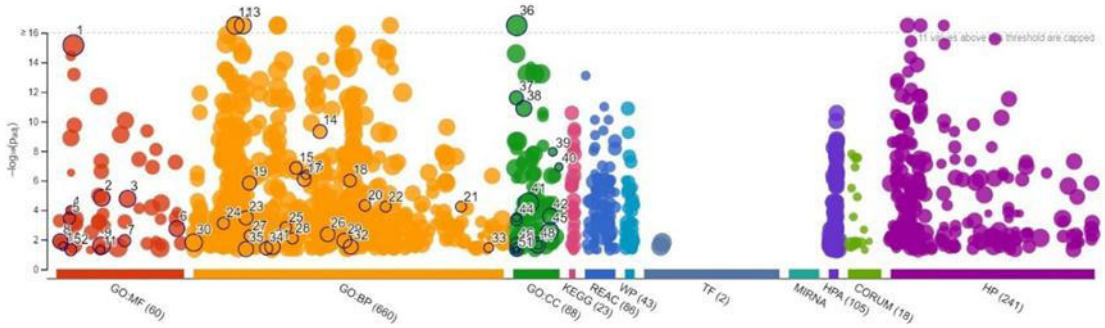


Рисунок 7. График точечных значений категорий генных онтологий генов, ассоциированных со спинальной мышечной атрофией I типа, рассчитанный с помощью программы g:Profiler.
(По оси Y —значения p-value в логарифмической шкале, по оси X — группы терминов генных онтологий).

По данным рисунка видно, что наиболее значимыми категориями молекулярных функций являются:

1. Protein binding (Связывание с белками, GO:0005515)
2. Protein-macromolecule adaptor activity (Белок-макромолекула адаптерная активность, GO:0030674). Продукты генов способны выступать в роле адаптеров, объединяя две или более макромолекул в контакте, что позволяет им функционировать скоординированно.
3. Transition metal ion binding (Перенос ионов переходных металлов, GO:0046914)
4. Prostaglandin E receptor activity (Активность рецептора простагландина-Е, GO:0004957)
5. Nuclear receptor activity (Активность ядерных рецепторов, GO:0004879)

А для биологических процессов –

1. Response to endogenous stimulus (Реакция на эндогенный стимул, GO:0009719)

2. Positive regulation of gene expression (Положительная регуляция экспрессии генов, GO:0010628)
3. Regulation of RNA splicing (Регуляция сплайсинга РНК, GO:0043484)
4. Cellular response to heat (Реакция клеток на тепло, GO:0034605)
5. 7-methylguanosine cap hypermethylation (Гиперметилирование 7-метилгуанозинового кэпа, GO:0036261) Продукты генов участвуют в процессе гиперметилирования 7-(моно)метилгуанозина (m(7)G) кэпа РНК во 2' положении остатка гуанозина для преобразования монометилированного кэпа в 2,2,7-триметилгуанозиновый кэп. Этот тип кэп-модификации происходит с малыми ядерными РНК (snRNA) и малыми ядрышковыми РНК (snoRNAs) и зависит от предшествующего метилирования гуанина-N7.
6. Response to topologically incorrect protein (Реакция на топологически неправильный белок, GO:0035966)
7. Regulation of steroid biosynthetic process (Регуляция процесса биосинтеза стероидов, GO:0050810)
8. Monocarboxylic acid transport (Транспорт монокарбоновых кислот, GO:0015718)
9. Cartilage development involved in endochondral bone morphogenesis (Развитие хряща, участвующего в морфогенезе эндохондральной кости, GO:0060351)
10. U2-type prespliceosome assembly (Сборка преспайсосомы типа U2, GO:1903241)

На основании анализа полученных категорий генных онтологий, можно сделать вывод, что многие продукты найденных генов с высокой достоверностью участвуют в процессах созревания РНК, инициации роста и дифференцировки клеток, развития нервной системы, а также регулируют экспрессию генов. Поэтому можно предположить, что дефекты этих генов способны оказать влияние на фенотип при спинальной мышечной атрофии.

РЕКОНСТРУКЦИЯ И АНАЛИЗ ГЕННОЙ СЕТИ

Для реконструкции генной сети данного заболевания использовался ресурс STRING-DB (https://string-db.org/). При минимально необходимом балле взаимодействия 0.99 (minimum required interaction score-custom value) была реконструирована сеть, состоящая из 303 узлов и 268 ребер, имеющая среднюю степень связности узла - 1.77, коэффициент кластеризации - 0.292. Как видно из рисунка, имеется 3 отдельные крупные сети, с разной степенью связанности.

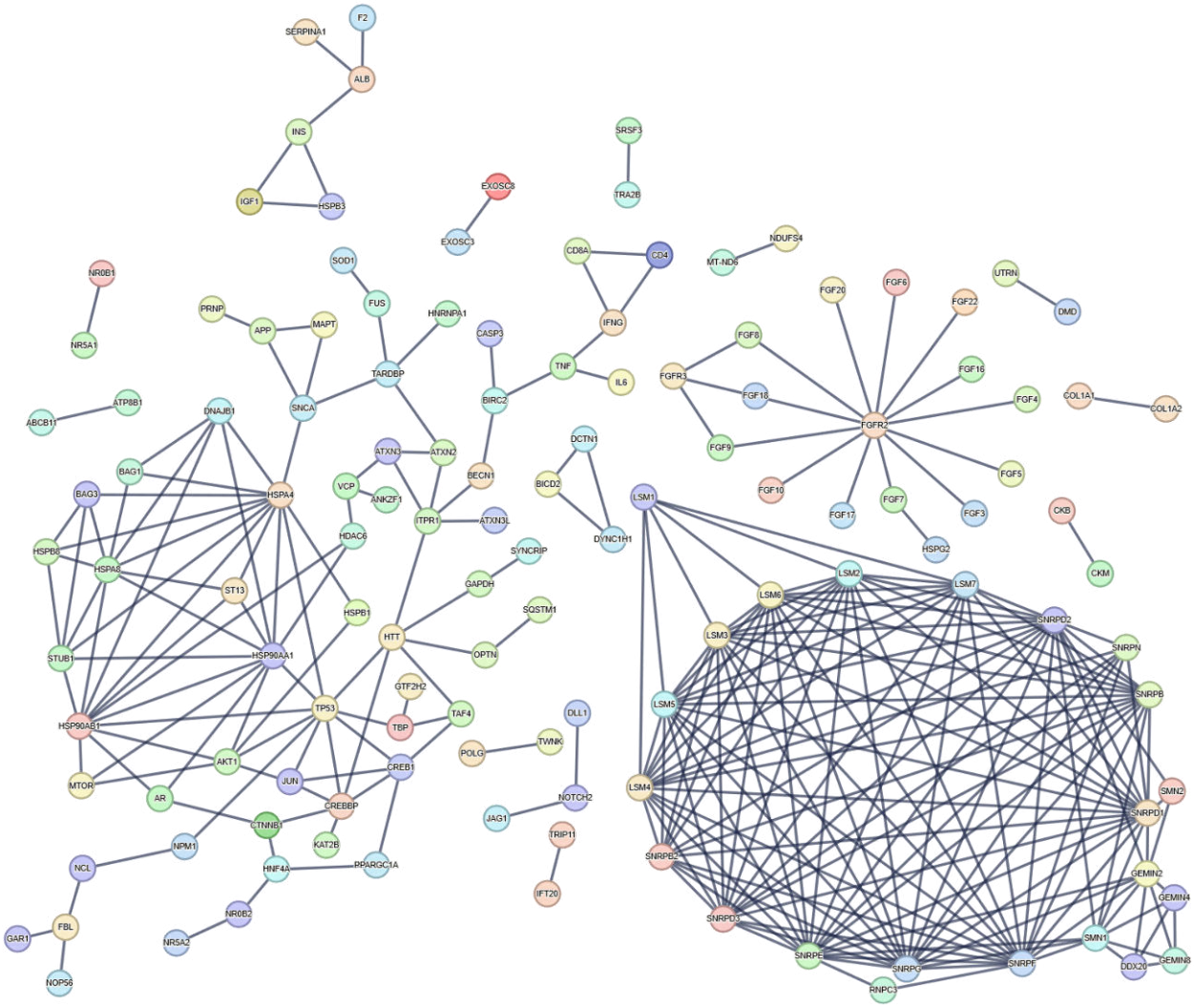


Рисунок 8. Генная сеть генов, ассоциированных со спинальной мышечной атрофией I типа

Также был проведен поиск кластеров, посредством опции «MCL clustering» были найдены естественные кластеры.

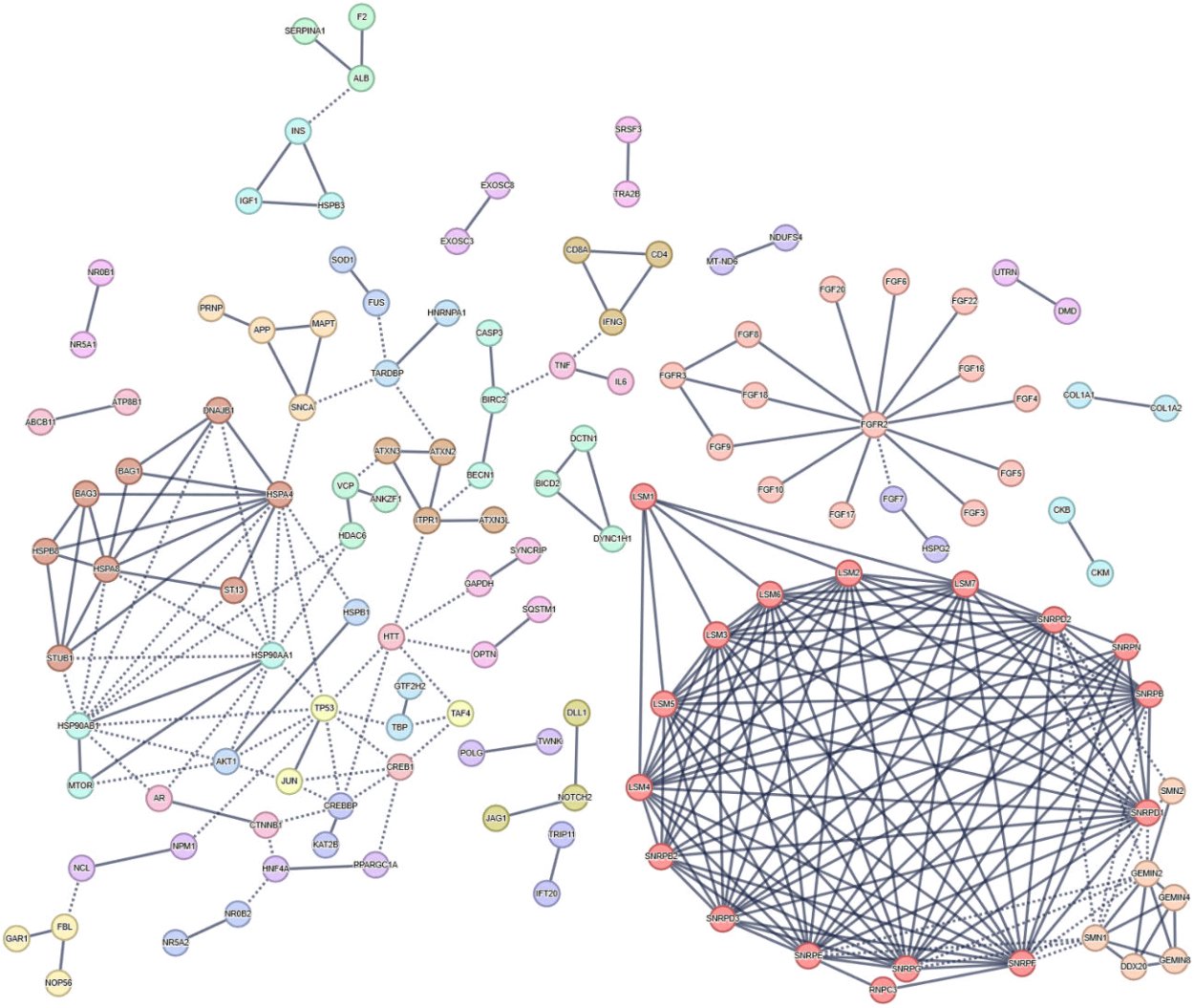


Рисунок 9. Естественные кластеры генной сети.

Ожидаемо, наибольшие кластеры включают гены, ответственные за биогенез snRNP и взаимодействие с рецептором фактором роста фибробластов.

Наибольшую степень узла имеют следующие гены: SNRPD1, SNRPD2, SNRPE, SNRPF, SNRPG, LSM4, LSM2, LSM3, LSM5, LSM6, SNRPB, SNRPD3, FGFR2, LSM7, HSPA4, SNRPB2, HSP90AA1, HSP90AB1, GEMIN2, SMN1, TP53. Таким образом, можно предположить, что именно они являются ключевыми генами данной сети и способны оказывать влияние на фенотип при спинальной мышечной атрофии I типа.

Полученная генная сеть коррелирует с данными пути сборки snRNP с интернет-ресурса Pathway Commons (https://www.pathwaycommons.org/).

СТРУКТУРЫ БЕЛКОВ

Структура некоторых наиболее значимых белков была определена с помощью инструмента AlphaFold (https://alphafold.ebi.ac.uk/) и PDBe-KB (https://www.ebi.ac.uk/pdbe/pdbe-kb/):

1. **SNRPD1 (Small nuclear ribonucleoprotein Sm D1)**

Играет роль в сплайсинге пре-мРНК в качестве основного компонента малых ядерных рибонуклеопротеинов U1, U2, U4 и U5 (snRNP). Входит в состав прекаталитического комплекса сплайсосомы B и активированного комплекса сплайсосомы C. В качестве компонента малой сплайсосомы участвует в сплайсинге интронов U12-типа в пре-мРНК. Может выступать в качестве заряженного белкового каркаса, способствующего сборке малых ядерных рибонуклеопротеинов или усиливающего взаимодействия малых ядерных рибонуклеопротеинов посредством неспецифических электростатических контактов с РНК

2 SNRPE (Small nuclear ribonucleoprotein E)

Также играет роль в сплайсинге пре-мРНК в качестве основного компонента snRNP U1, U2, U4 и U5.

3 SNRPG (Small nuclear ribonucleoprotein G)

Функции аналогичны представленным выше белкам: является основным компонентом сплайсосомальных малых ядерных рибонуклеопротеинов U1, U2, U4 и U5, в состав как прекаталитического комплекса сплайсосомы B, так и активированного комплекса сплайсосомы C

4 LSM4 (U6 snRNA-associated Sm-like protein LSm4)

Играет роль в сплайсинге пре-мРНК как компонент tri-snRNP комплекса U4/U6-U5, который участвует в сборке сплайсосомы, а также как компонент прекаталитической сплайсосомы B.

5 FGFR2 (Fibroblast growth factor receptor 2)

Рецептор на поверхности клеток для факторов роста фибробластов (FGF) представляет собой тирозинкиназу и играет важную роль в регуляции клеточной пролиферации, дифференцировки, миграции и апоптоза, а также участвует в регуляции эмбриогенеза. Необходим для нормального эмбрионального развития, функции трофобласта, развития зачатков конечностей, морфогенеза лёгких, развития кожи и остеогенеза (играет важную роль в регуляции дифференцировки, пролиферации и апоптоза остеобластов, поэтому необходим для нормального развития скелета). Фосфорилирует PLCG1, FRS2 и PAK4. Активация PLCG1 приводит к выработке диацилглицерина и инозитол-1,4,5-трифосфата, фосфорилирование FRS2 запускает рекрутирование GRB2, GAB1, PIK3R1 и SOS1, опосредует активацию RAS, MAPK1/ERK2, MAPK3/ERK1 и сигнального пути MAP-киназы, а также сигнального пути AKT1. Избыточная экспрессия FGFR2 способствует активации STAT1.

РЕЗУЛЬТАТЫ АНАЛИЗА ГЕНОВ АССОЦИИРОВАННЫХ СМА1

Был составлен список из 319 генов, для которого наиболее значимыми категориями генных онтологий по DAVID являются: Neurodegeneration, домены LSM и SM (LSM_dom_sf, Sm_dom_euk/arc, Sm; суперсемейство доменов LSM и домены Sm встречается в белках Lsm (like-Sm proteins) и Sm (Sm proteins), играющих важную роль в сплайсинге и декапировании мРНК), mRNA processing (процессинг мРНК), Fibroblast

growth factor receptor signaling pathway, Fibroblast_GF_fam (семейство факторов роста фибробластов), SMN-Sm protein complex (белковые комплексы, включающие Sm-белки и SMN), Spliceosomal snRNP assembly. По PANTHER наиболее значимыми категориями биологических процессов являются: RNA splicing (GO:0008380), процессы биосинтеза (Biosynthetic process GO:0009058, Nucleobase-containing compound biosynthetic process GO:0034654, Nucleic acid biosynthetic process GO:0141187, Cellular biosynthetic process GO:0044249), mRNA metabolic process (GO:0016071), Cellular process (GO:0009987), mRNA processing (GO:0006397), RNA splicing, via transesterification reactions (GO:0000375), Cellular metabolic process (GO:0044237). Наиболее значимые категории молекулярных функций генов: связывание различных агентов (Protein binding GO:0005515, Fibroblast growth factor receptor binding GO:0005104, RNA binding GO:0003723, Enzyme binding GO:0019899, Binding GO:0005488, Protein-folding chaperone binding GO:0051087, Identical protein binding GO:0042802, Protein-containing complex binding GO:0044877, Organic cyclic compound binding GO:0097159) и Growth factor activity GO:0008083. По данным GOST наиболее значимыми категориями молекулярных функций являются: Protein binding GO:0005515, Protein-macromolecule adaptor activity GO:0030674 (Продукты генов способны объединять макромолекулы обеспечивая их скоординированное функционирование), Transition metal ion binding GO:0046914, Prostaglandin E receptor activity GO:0004957, Nuclear receptor activity GO:0004879, а биологических процессов – Response to endogenous stimulus GO:0009719, Positive regulation of gene expression, GO:0010628, Regulation of RNA splicing GO:0043484, Cellular response to heat GO:0034605, 7-methylguanosine cap hypermethylation GO:0036261 (Этот тип кэп-модификации происходит с snRNA и snoRNA). Посредством STRING-DB была реконструирована генная сеть, содержащая 303 узлов и 268 ребер, имеющая среднюю степень связности узла - 1.77, коэффициент кластеризации - 0.292, и состоящая из трёх крупных сети, с разной степенью связанности. Также был проведен поиск естественных кластеров. Наибольшие кластеры включали гены, ответственны за биогенез snRNP и взаимодействие с рецептором фактором роста фибробластов.

СЛОЖНОСТЬ ВИРУСНОГО ГЕНОМА И СТРУКТУРА ПОВТОРОВ

Было изучено строение последовательности генома коронавируса, вызывающего COVID-19, с точки зрения структуры повторов с помощью инструментов биоинформатики, по базам данных[38]. Выравнивание последовательностей, поиск горячих точек мутации

(позиций последовательности, в которой мутации возникают с повышенной частотой относительно других позиций).

Новые расчеты на геномах вирусов показывают преобладание участков повышенной сложности, кодирующих полипротеин, участки низкой сложности связаны с повышенной изменчивостью. Максимальные длины повторов в геномах вирусов представляют собой прямые повторы, что может свидетельствовать об эволюционно недавних дупликациях.

Рассмотрено применение программы для анализа последовательности генома коронавируса SARS-CoV-2 – возбудителя опасного респираторного заболевания COVID-19.

SARS-CoV-2 относится к группе одноцепочечных РНК-вирусов и имеет один из самых крупных вирусных геномов, длина которого составляет 29,9 кб. Наиболее важными генами вируса SARS-CoV-2, кодирующими белки, являются (рис. 7):

1. Ген, кодирующий S-белок. S-белок представляет собой гликопротеин оболочки, который играет наиболее важную роль в прикреплении вируса, слиянии и проникновении в клетки-хозяева и служит основной мишенью для разработки нейтрализующих антител, ингибиторов проникновения вируса и вакцин.
2. Ген, кодирующий M-белок. Этот белок играет ключевую роль в формировании вирусной оболочки.
3. Ген, кодирующий E-белок. Данный белок участвует в сборке вирусной частицы.
4. Ген, кодирующий N-белок. Белок принимает участие в защите вирусной РНК от воздействия организма хозяина.

Кроме того, существуют коронавирусы, которые не заражают человека. Они отличаются от COVID-19 не более чем на 1%. Продолжаются дискуссии о возникновении патогенных участков вируса и его последующей эволюции.

Важно отметить, что длина РНК-последовательности генома коронавируса составляет около 30 000 нуклеотидов. РНК варианта Wuhan-Hu-1 (номер GenBank MN908947, RefCeq NC_045512) SARS-CoV-2 содержит 29 903 нуклеотида с нетранслируемыми участками длиной 281 и 325 нуклеотидов. Предполагаемые кодирующие области распределены по 10 белкам. Однако у вирусных частиц SARS-CoV-2 отсутствует белок Hemagglutinin esterase. Генетически вирус на 80 % идентичен SARS-CoV.

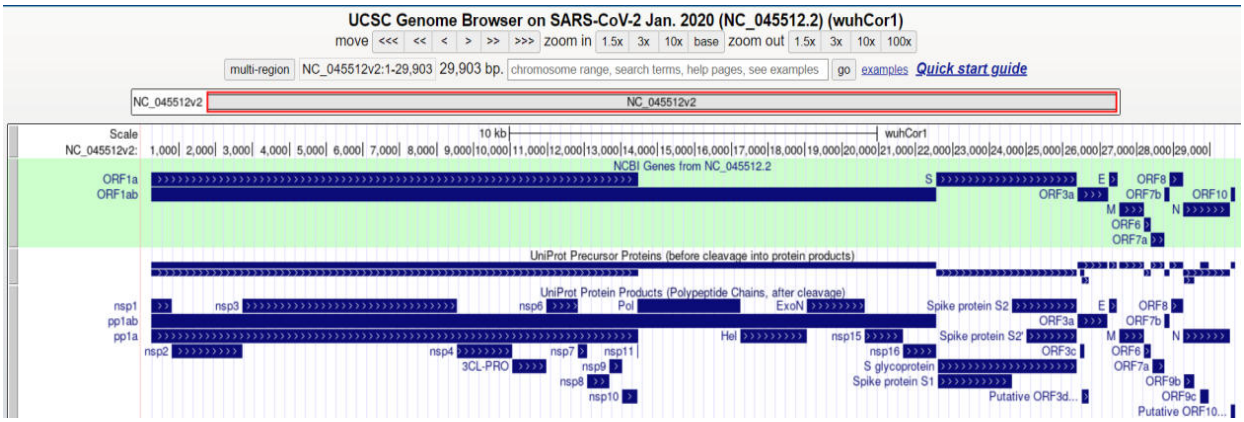


Рисунок 10. Структурная организация генома SARS-CoV-2. Скриншот из геномного браузера The UCSC Genome Browser (https://genome.ucsc.edu/).

С помощью специально созданной программы, анализирующей лингвистическую сложность генома, в нуклеотидной последовательности генома SARS-CoV-2 были обнаружены фрагменты с низкой сложностью, совпадающие с мононуклеотидными повторами. Для анализа использовалась полная геномная последовательность вируса, загруженная из базы данных NCBI (https://www.ncbi.nlm.nih.gov/).

При расчете сложности по Лемпелю-Зиву с помощью программного комплекса, разработанного в Институте цитологии и генетики СО РАН (http://wwwmgs.bionet.nsc.ru/mgs/programs/lzcomposer/), был получен следующий график (рис.8):

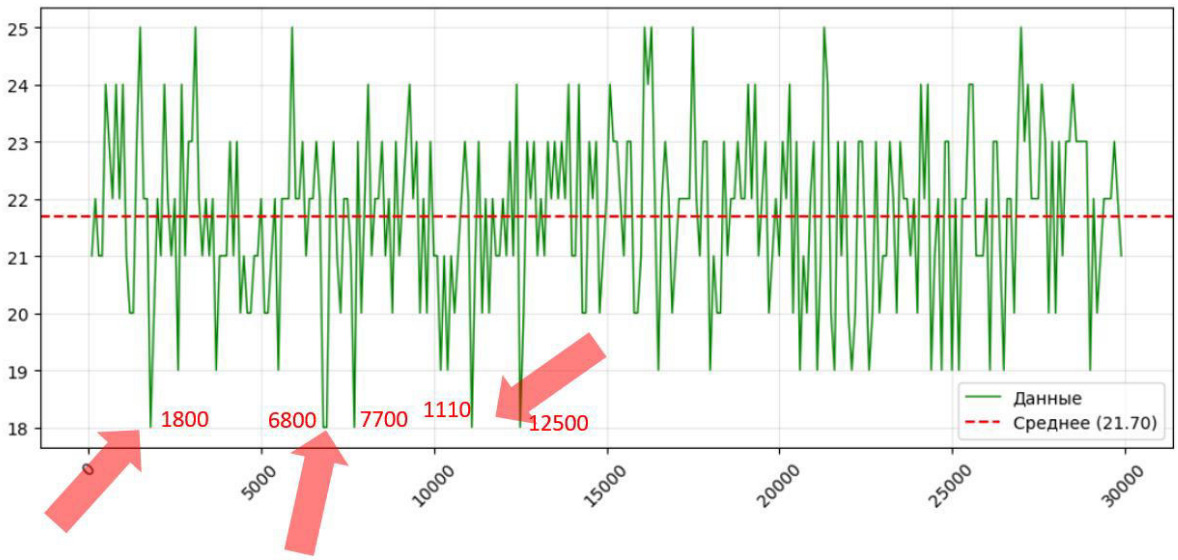


Рисунок 11. Расчет сложности по Лемпелю-Зиву для SARS-CoV-2 при шаге, равном 100, и длине окна, равной 50. Участки с наименьшими значениями сложности указаны стрелками.

В таблице 2 представлены гены коронавируса, входящие в участки низкой сложности:

Название гена	Участок генома, входящие домены	Координаты гена	NCBI Gene ID	Protein Product
ORF1a	NC_045512v2: 1261-2310 (CoV Nsp2 middle, CoV Nsp2 N-terminal) NC_045512v2:6301-7350 (Nucleic acid-binding, G2M, 3Ecto) NC_045512v2:7351-8260 (Region 2395-2485 Y1) NC_045512v2:11971-13020 (RdRp Nsp7 cofactor, RdRp Nsp8 cofactor, Nsp9 ssRNA-binding)	NC_045512v2:266-13483	43740578	YP_009725295.1
ORF1ab		NC_045512v2:266-21555	43740578	YP_009724389.1
Replicase polyprotein 1ab, pp1ab		NC_045512v2:266-21552		P0DTD1
Replicase polyprotein 1a, pp1a		NC_045512v2:266-13480		P0DTC1
Non-structural protein 2, nsp2	NC_045512v2: 1261-2310 (CoV Nsp2 middle, CoV Nsp2 N-terminal)	NC_045512v2:806-2719		P0DTC1
Papain-like protease nsp3	NC_045512v2:6301-7350 (Nucleic acid-binding, G2M, 3Ecto) NC_045512v2:7351-8260 (Region 2395-2485 Y1)	NC_045512v2:2720-8554		P0DTC1
Non-structural protein 7	NC_045512v2:11971-13020 (RdRp Nsp7 cofactor, RdRp Nsp8 cofactor, Nsp9 ssRNA-binding)	NC_045512v2:11843-12091		P0DTC1

Таблица 2. Гены коронавируса SARS-CoV-2, часть которых лежат в участках с низкой сложностью.

Участок низкой сложности вокруг 7700-ого нуклеотида (7351-8260 n.)

TTGGCTTATGTGGTTAATAATTAATCTTGTACAAATGGCCCCGATTTCAGCTATGGTTAGAATGTACATC
TTCTTTGCATCATTTTATTATGTATGGAAAAGTTATGTGCATGTTGTAGACGGTTGTAATTCATCAACTT
GTATGATGTGTTACAAACGTAATAGAGCAACAAGAGTCGAATGTACAACTATTGTTAATGGTGTTAGAAG
GTCCTTTTATGTCTATGCTAATGGAGGTAAAGGCTTTTGCAAACTACACAATTGGAATTGTGTTAATTGT
GATACATTCTGTGCTGGTAGTACATTTATTAGTGATGAAGTTGCGAGAGACTTGTCACTACAGTTTAAAT
GACCAATAAATCCTACTGACCAGTCTTCTTACATCGTTGATAGTGTTACAGTGAAGAATGGTTCCATCCA
TCTTTACTTTGATAAAGCTGGTCAAAAGACTTATGAAAGACATTCTCTCTCTCATTTTGTTAACTTAGAC
AACCTGAGAGCTAATAACACTAAAGGTTCATTGCCTATTAATGTTATAGTTTTTGATGGTAAATCAAAAT
GTGAAGAATCATCTGCAAAATCAGCGTCTGTTTACTACAGTCAGCTTATGTGTCAACCTATACTGTTACT
AGATCAGGCATTAGTGTCTGATGTTGGTGATAGTGCGGAAGTTGCAGTTAAAATGTTTGATGCTTACGTT
AATACGTTTTCATCAACTTTTAACGTACCAATGGAAAAACTCAAAACACTAGTTGCAACTGCAGAAGCTG
AACTTGCAAAGAATGTGTCCTTAGACAATGTCTTATCTACTTTTATTTCAGCAGCTCGGCAAGGGTTTGT
TGATTCAGATGTAGAAACTAAAGATGTTGTTGAATGTCTTAAATTGTCACATCAATCTGACATAGAAGTT

Участок низкой сложности вокруг 12500-ого нуклеотида (11971-13020 n.)

AGATACTACTGAAGCCTTTGAAAAAATGGTTTCACTACTTTCTGTTTTGCTTTCCATGCAGGGTGCTGTA
GACATAAACAAGCTTTGTGAAGAAATGCTGGACAACAGGGCAACCTTACAAGCTATAGCCTCAGAGTTTA
GTTCCCTTCCATCATATGCAGCTTTTGCTACTGCTCAAGAAGCTTATGAGCAGGCTGTTGCTAATGGTGA
TTCTGAAGTTGTTCTTAAAAAGTTGAAGAAGTCTTTGAATGTGGCTAAATCTGAATTTGACCGTGATGCA

GCCATGCAACGTAAGTTGGAAAAGATGGCTGATCAAGCTATGACCCAAATGTATAAACAGGCTAGATCTG
AGGACAAGAGGGCAAAAGTTACTAGTGCTATGCAGACAATGCTTTTCACTATGCTTAGAAAGTTGGATAA
TGATGCACTCAACAACATTATCAACAATGCAAGAGATGGTTGTGTTCCCTTGAACATAATACCTCTTACA
ACAGCAGCCAAACTAATGGTTGTCATACCAGACTATAACACATATAAAAATACGTGTGATGGTACAACAT
TTACTTATGCATCAGCATTGTGGGAAATCCAACAGGTTGTAGATGCAGATAGTAAAATTGTTCAACTTAG
TGAAATTAGTATGGACAATTCACCTAATTTAGCATGGCCTCTTATTGTAACAGCTTTAAGGGCCAATTCT
GCTGTCAAATTACAGAATAATGAGCTTAGTCCTGTTGCACTACGACAGATGTCTTGTGCTGCCGGTACTA
CACAAACTGCTTGCACTGATGACAATGCGTTAGCTTACTACAACACAACAAAGGGAGGTAGGTTTGTACT
TGCACTGTTATCCGATTTACAGGATTTGAAATGGGCTAGATTCCCTAAGAGTGATGGAACTGGTACTATC
TATACAGAACTGGAACCACCTTGTAGGTTTGTTACAGACACACCTAAAGGTCCTAAAGTGAAGTATTTAT
ACTTTATTAAAGGATTAAACAACCTAAATAGAGGTATGGTACTTGGTAGTTTAGCTGCCACAGTACGTCT

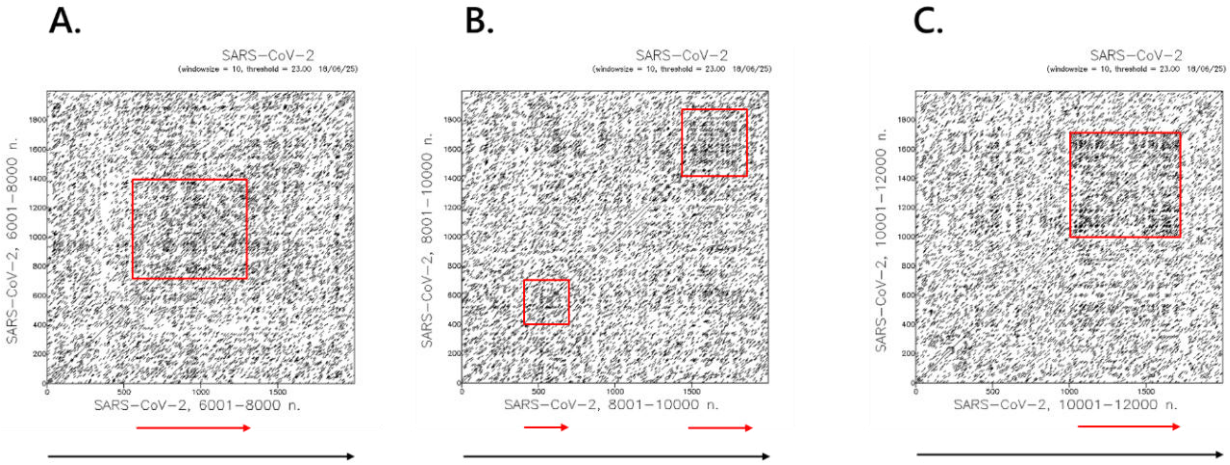


Рисунок 12. Визуализация участков низкой сложности текста в геноме коронавируса SARS-CoV-2 показывает участки, насыщенные повторами в районах 6547-7290 н. (A), 8404-8695 н. (B) и 11000-11715 н. (C), выделены красным цветом. Участок повторов виден в окне в дот-матрице.

Анализ лингвистической сложности генома выявил особенно значимый результат: участок с наименьшей сложностью кодирует S-белок. Это представляет особый научный интерес, поскольку подобные низкосложные последовательности, согласно современным представлениям, обладают повышенной склонностью к мутациям. Данное наблюдение хорошо согласуется с известной биологической ролью S-белка - именно этот белок является ключевой мишенью при разработке противовирусных средств. На его структуру направлено действие нейтрализующих антител, именно он служит основной мишенью для ингибиторов вирусного проникновения, и именно на его основе конструируются большинство современных вакцин. Таким образом, выявленная особенность геномной организации может иметь важное значение для понимания механизмов изменчивости вируса и разработки эффективных средств противодействия.

ЗАКЛЮЧЕНИЕ

В ходе работы «Компьютерный анализ структуры промоторных районов генов, ассоциированных с комплексными заболеваниями» решены следующие задачи –

Подготовлены списки генов (и кодируемых ими белков, связанных с комплексными заболеваниями - глиомой, деменцией, болезнью Паркинсона, муковисцидозом, CMA1). Использовались базы данных UniProt, GeneCards, согласно методике, представленной в работах [26-29].

Выполнены компьютерные расчеты и визуализация структуры генной сети муковисцидоза (STRING-DB, Cytoscape), выделение кластеров сети по полученному списку.

Подготовлен текст тезисов научной конференции и научной публикации в журнале («Вавиловский журнал генетики и селекции», «Журнал физической химии»). Тезисы поданы на конференции «Сознание-2025» в Красновидово, II Международная конференция «Инновационные лекарственные средства от молекулы до пациента», XXXI Symposium on Bioinformatics and Computer-Aided Drug Discovery (BCADD-2025), III Всероссийская конференция с международным участием и школа молодых ученых «Физико-химические методы в междисциплинарных экологических исследованиях», Международный молодёжный научный форум «Медицинская весна – 2025» и приняты для публикации в сборники конференций.

Научная публикация, подготовленная на английском языке, опубликована в журнале Biophysical Reviews в 2025 году.

Подготовлен шаблон методических материалов по реконструкции генной сети для списка генов используя интеграцию данных из баз данных (GeneCards, STRING-DB, Cytoscape) (с иллюстрациями).

Проведено биоинформационное исследование генов, ассоциированных с муковисцидозом, CMA1. Для найденных в открытых базах данных генов были рассчитаны категории генных онтологий биологических процессов и молекулярных функций.

По результатам исследования установлено, что наиболее значимые категории биологических процессов, в которых задействованы продукты ассоциированных с муковисцидозом генов, можно объединить в следующие группы: реакции на различные воздействия, транспортные процессы, процессы, связанные с иммунными реакциями и регуляторные процессы. И наиболее значимыми группами категорий молекулярных функций найденных генов является связывание белковых агентов и нуклеотидов.

Была реконструированная генная сеть муковисцидоза, состоящая из 495 узлов и 657 ребер, имеющая среднюю степень связности узла - 2.65, коэффициент кластеризации - 0.376. Наибольшую степень узла имеют следующие гены: *PSMD12, ADRM1, PSMA1 - PSMA7, PSMB1 - PSMB5, PSMC2 - PSMC6, PSMD1, PSMD11, PSMD14, PSMD3, PSMD7, PSMD8, NFKB1, PRKACA, SRC, JUN, EGFR, HSP90AA1, HSPA8*.

Было найдено 84 кластера, состоящих из более чем одного гена. Ген *CFTR* не вошел в малые кластеры сети, при этом он связан с генами *PRKACA*, *PRKAA2*, *FLNA*, *HSPA8*, *SLC9A3R1*. Установлена связь генов PRKACA, PRKAA2, FLNA, HSPA8 и SLC9A3R1 с муковисцидозом, возможность их использования в качестве прогностических маркёров и мишеней для терапии требует дальнейшего изучения и экспериментального подтверждения.

Был проанализирован список из 319 генов, ассоциированных со спинальной мышечной атрофией I типа.. Были определены категории генных онтологий и рассмотрены основные процессы, связанные с нарушениями работы генов при этом заболевании. Реконструирована и проанализирована генная сеть, определены естественные кластеры генов и ключевые гены. Многие продукты найденных генов участвуют в созревания РНК, инициации роста и дифференцировки клеток, развития нервной системы и регулируют экспрессию генов. Гены SNRPD1, SNRPD2, SNRPE, SNRPF, SNRPG, LSM4, LSM2, LSM3, LSM5, LSM6, SNRPB, SNRPD3, FGFR2, LSM7, HSPA4, SNRPB2, HSP90AA1, HSP90AB1, GEMIN2, SMN1 и TP53 являются ключевыми генами реконструированной сети и вероятно способны оказывать влияние на фенотип при СМА1.

Разработаны и оптимизированы программы оценки сложности генетических текстов, с использованием набора методов (сложность по Лемпелю-Зиву, лингвистическая сложность и энтропия). С использованием баз данных собраны выборки нуклеотидных последовательностей, содержащие участки нуклеотидных полиморфизмов. Показано, что сложность текста ДНК ниже на фланкирующих участках, что может быть связано с физико-химическими свойствами двойной спирали ДНК, приводящими к разрывам и мутациям. Выполнены оценки профилей сложности текста в скользящем окне для полных геномов микроорганизмов.

Показано, что участки пониженной сложности текста соответствуют границам структурных доменов в геноме коронавируса

В целом на примере такого комплексного заболевания как муковисцидоз отработана методика получения списка генов, их функциональной аннотации и реконструкции генной сети.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Союз педиатров России, Ассоциация медицинских генетиков, Российское респираторное общество, Российское трансплантологическое общество, Ассоциация детских врачей Московской области. Кистозный фиброз (муковисцидоз). Клинические рекомендации Минздрава РФ. 2021; https://cr.minzdrav.gov.ru/preview-cr/372_2
2. Plasschaert LW, MacDonald KD, Moffit JS. Current landscape of cystic fibrosis gene therapy. Front Pharmacol. 2024 Oct 8;15:1476331. doi: 10.3389/fphar.2024.1476331.
3. Dean M, Moitra K, Allikmets R. The human ATP-binding cassette (ABC) transporter superfamily. Hum Mutat. 2022 Sep;43(9):1162-1182. doi: 10.1002/humu.24418.
4. Hanssens LS, Duchateau J, Casimir GJ. CFTR Protein: Not Just a Chloride Channel? Cells. 2021 Oct 22;10(11):2844. doi: 10.3390/cells10112844.
5. Lee, JA., Cho, A., Huang, E.N. et al. Gene therapy for cystic fibrosis: new tools for precision medicine. J Transl Med 19, 452 (2021). doi: 10.1186/s12967-021-03099-4
6. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science. 1989 Sep 8;245(4922):1066-73. doi: 10.1126/science.2475911.
7. Terlizzi V, Lopes-Pacheco M. Cystic fibrosis: new challenges and perspectives beyond elexacaftor/tezacaftor/ivacaftor. Ther Adv Respir Dis. 2025 Jan-Dec;19:17534666251323194. doi: 10.1177/17534666251323194.
8. Регистр пациентов с муковисцидозом в Российской Федерации. 2023 год. /Под редакцией Е.Л. Амелиной, Н.Ю. Каширской, Е.И. Кондратьевой, С.А. Красовского, М.А. Стариновой, А.Ю. Воронковой, Е.К. Гинтера. – М.: ИД «МЕДПРАКТИКА-М», 2025, 70 с. DOI: 10.61726/1981.2025.26.16.001 ISBN 978-5-98803-469-8
9. Frost FJ, Peckham DG, Felton IC, Snowball JE, Gray RD, Jones AM, Simmonds NJ, Lord RW, Lip GYH, Chandler H, Murphy K, Downey DG, Sheppard DN, Davies JC, Bull J, Sommer P, Cupid B, Allen L, Duckers J. Managing an ageing cystic fibrosis population: challenges and priorities. Eur Respir Rev. 2025 May 14;34(176):240261. doi: 10.1183/16000617.0261-2024.
10. Maule G, Arosio D, Cereseto A. Gene Therapy for Cystic Fibrosis: Progress and Challenges of Genome Editing. Int J Mol Sci. 2020 May 30;21(11):3903. doi: 10.3390/ijms21113903.
11. Bareil C, Bergougnoux A. CFTR gene variants, epidemiology and molecular pathology. Arch Pediatr. 2020 Feb;27 Suppl 1:eS8-eS12. doi: 10.1016/S0929-693X(20)30044-0. PMID: 32172939.
12. Lomunova M.A., Gershovich P.M. Gene therapy for Cystic Fibrosis: recent advances and future prospects. Acta Naturae. - 2023. - Vol. 15. - N. 2. - P. 20-31. doi: 10.32607/actanaturae.11708
13. Simon MA, Csanády L. Molecular pathology of the R117H cystic fibrosis mutation is explained by loss of a hydrogen bond. Elife. 2021 Dec 6;10:e74693. doi: 10.7554/eLife.74693.
14. Levring J, Terry DS, Kilic Z, Fitzgerald G, Blanchard SC, Chen J. CFTR function, pathology and pharmacology at single-molecule resolution. Nature. 2023 Apr;616(7957):606-614. doi: 10.1038/s41586-023-05854-7.

15. Vilà-González, M., Pinte, L., Fradique, R. et al. In vitro platform to model the function of ionocytes in the human airway epithelium. Respir Res 25, 180 (2024). https://doi.org/10.1186/s12931-024-02800-7
16. Lei L, Traore S, Romano Ibarra GS, Karp PH, Rehman T, Meyerholz DK, Zabner J, Stoltz DA, Sinn PL, Welsh MJ, McCray PB Jr, Thornell IM. CFTR-rich ionocytes mediate chloride absorption across airway epithelia. J Clin Invest. 2023 Oct 16;133(20):e171268. doi: 10.1172/JCI171268.
17. Tümmler B. Puzzle resolved: CFTR mediates chloride homeostasis by segregating absorption and secretion to different cell types. J Clin Invest. 2023 Oct 16;133(20):e174667. doi: 10.1172/JCI174667.
18. McDonald EF, Meiler J, Plate L. CFTR Folding: From Structure and Proteostasis to Cystic Fibrosis Personalized Medicine. ACS Chem Biol. 2023 Oct 20;18(10):2128-2143. doi: 10.1021/acschembio.3c00310.
19. Orlov Y.L., Galieva A.G., Orlova N.G., Ivanova E.N., Mozyleva Y.A., Anashkina A.A. Reconstruction of gene network associated with Parkinson disease for gene targets search. Biomed Khim. 2021;67(3):222-230. Russian.
20. Туркина В.А., Ярема П.О., Майорова А.А., Орлова Н.Г., Савина Е.А., Орлов Ю.Л. Реконструкция генных и ассоциативных сетей заболеваний для поиска генов-мишеней. Актуальные вопросы биологической физики и химии, 2023, том 8, № 4, с. 458-467. (РИНЦ) Turkina V.A., Iarema P.O., Mayorova A.A., Orlova N.G., Savina E.A., Orlov Y.L. Reconstruction of gene and associative networks of diseases to search for target genes . Russian Journal of Biological Physics and Chemistry, 2023; 8(4): 458-467 DOI: 10.29039/rusjbpc.2023.0647 (In Russian)
21. Anashkina A.A., Orlova N.G., Kolchanov N.A., Orlov Y.L. New Sights into Bioinformatics of Gene Regulations and Structure. International Journal of Molecular Sciences. 2025; 26(13):6442. doi: 10.3390/ijms26136442
22. Hallows KR, Raghuram V, Kemp BE, Witters LA, Foskett JK. Inhibition of cystic fibrosis transmembrane conductance regulator by novel interaction with the metabolic sensor AMP-activated protein kinase. J Clin Invest. 2000 Jun;105(12):1711-21. doi: 10.1172/JCI9622.
23. Smith L, Litman P, Kohli E, Amick J, Page RC, Misra S, Liedtke CM. RACK1 interacts with filamin-A to regulate plasma membrane levels of the cystic fibrosis transmembrane conductance regulator. Am J Physiol Cell Physiol. 2013 Jul 1;305(1):C111-20. doi: 10.1152/ajpcell.00026.2013.
24. Basile A, Pascale M, Franceschelli S, Nieddu E, Mazzei MT, Fossa P, Turco MC, Mazzei M. Matrine modulates HSC70 levels and rescues ΔF508-CFTR. J Cell Physiol. 2012 Sep;227(9):3317-23. doi: 10.1002/jcp.24028.
25. Tamanini A, Fabbri E, Jakova T, Gasparello J, Manicardi A, Corradini R, Finotti A, Borgatti M, Lampronti I, Munari S, Dechecchi MC, Cabrini G, Gambari R. A Peptide-Nucleic Acid Targeting miR-335-5p Enhances Expression of Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Gene with the Possible Involvement of the CFTR Scaffolding Protein NHERF1. Biomedicines. 2021 Jan 26;9(2):117. doi: 10.3390/biomedicines9020117.
26. Chávez JC, Hernández-González EO, Wertheimer E, Visconti PE, Darszon A, Treviño CL. Participation of the Cl-/HCO(3)- exchangers SLC26A3 and SLC26A6, the Cl- channel

CFTR, and the regulatory factor SLC9A3R1 in mouse sperm capacitation. Biol Reprod. 2012 Jan 19;86(1):1-14. doi: 10.1095/biolreprod.111.094037.
27. Zhuri D, Gurkan H, Eker D, Karal Y, Yalcintepe S, Atli E, Demir S, Atli EI. Investigation on the Effects of Modifying Genes on the Spinal Muscular Atrophy Phenotype. Glob Med Genet. 2022 Sep 5;9(3):226-236. doi: 10.1055/s-0042-1751302.
28. Pascual-Morena C, Martínez-Vizcaíno V, Cavero-Redondo I, Martínez-García I, Moreno-Herráiz N, Álvarez-Bueno C, Saz-Lara A. Efficacy of risdiplam in spinal muscular atrophy: A systematic review and meta-analysis. Pharmacotherapy. 2024 Jan;44(1):97-105. doi: 10.1002/phar.2866.
29. Emmady PD, Bodle J. Werdnig-Hoffmann Disease. 2023 May 29. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–. PMID: 32644359.
30. Butchbach MER. Genomic Variability in the Survival Motor Neuron Genes (SMN1 and SMN2): Implications for Spinal Muscular Atrophy Phenotype and Therapeutics Development. Int J Mol Sci. 2021 Jul 23;22(15):7896. doi: 10.3390/ijms22157896. PMID: 34360669; PMCID: PMC8348669.
31. Mercuri E, Sumner CJ, Muntoni F, Darras BT, Finkel RS. Spinal muscular atrophy. Nat Rev Dis Primers. 2022 Aug 4;8(1):52. doi: 10.1038/s41572-022-00380-8. PMID: 35927425.
32. Nishio H, Niba ETE, Saito T, Okamoto K, Takeshima Y, Awano H. Spinal Muscular Atrophy: The Past, Present, and Future of Diagnosis and Treatment. Int J Mol Sci. 2023 Jul 26;24(15):11939. doi: 10.3390/ijms241511939. PMID: 37569314; PMCID: PMC10418635.
33. Gubitz AK, Feng W, Dreyfuss G. The SMN complex. Exp Cell Res. 2004 May 15;296(1):51-6. doi: 10.1016/j.yexcr.2004.03.022. PMID: 15120993.
34. Kolb SJ, Battle DJ, Dreyfuss G. Molecular functions of the SMN complex. J Child Neurol. 2007 Aug;22(8):990-4. doi: 10.1177/0883073807305666. PMID: 17761654.
35. Pellizzoni L, Yong J, Dreyfuss G. Essential role for the SMN complex in the specificity of snRNP assembly. Science. 2002 Nov 29;298(5599):1775-9. doi: 10.1126/science.1074962. PMID: 12459587.
36. Winkler C, Eggert C, Gradl D, Meister G, Giegerich M, Wedlich D, Laggerbauer B, Fischer U. Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy. Genes Dev. 2005 Oct 1;19(19):2320-30. doi: 10.1101/gad.342005. PMID: 16204184; PMCID: PMC1240041
37. Akbari Rokn Abadi S, Mohammadi A, Koohi S (2023) A new profiling approach for DNA sequences based on the nucleotides' physicochemical features for accurate analysis of SARS-CoV-2 genomes. BMC Genomics. 24(1): 266. https://doi.org/10.1186/s12864-023-09373-7
38. Orlov Y.L., Orlova N.G. (2023) Challenges in bioinformatics education courses organization. Biophysical Reviews, 15, 1367–1378.
39. Dergilev A.I., Orlova N.G., Dobrovolskaya O.B., Orlov Y.L. (2022) Statistical estimates of multiple transcription factors binding in the model plant genomes based on ChIP-seq data. Journal of Integrative Bioinformatics, 19(1), 20200036.
40. Orlov Y.L., Potapov V.N. (2004) Complexity: Internet-resource for analysis of DNA sequence complexity. Nucleic Acids Res. Web-issue 2004, V. 32: 628-633.
41. Montoro DT, et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature. 2018;560(7718):319–324. doi: 10.1038/s41586-018-0393-7.

42. Plasschaert LW, et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. Nature. 2018;560(7718):377–381. doi: 10.1038/s41586-018-0394-6.
43. Thornell IM, Lei L, McCray PB Jr, Welsh MJ. Do pulmonary ionocytes absorb chloride or secrete chloride? Am J Physiol Cell Physiol. 2025 Feb 1;328(2):C400-C403. doi: 10.1152/ajpcell.00672.2024.

Отчет составил
Исполнитель по гранту РНФ
№ 24-24-00563 от 29.12.2023г Сафрыгина А.А.

Отчет принял
Руководитель гранта РНФ
№ 24-24-00563 от 29.12.2023г Орлов Ю.Л.

2025 года

Москва