

*На правах рукописи*

**Протасова Мария Сергеевна**

**ГЕНЕТИЧЕСКИЕ ФАКТОРЫ РЕДКИХ ФОРМ ВРОЖДЁННЫХ  
МОЗЖЕЧКОВЫХ АТАКСИЙ**

1.5.7 - генетика

**АВТОРЕФЕРАТ**  
диссертации на соискание ученой степени  
кандидата биологических наук

Москва 2022

Работа выполнена в лаборатории эволюционной геномики Федерального государственного бюджетного учреждения науки Института общей генетики им. Н.И. Вавилова Российской академии наук.

**Научный руководитель:**

доктор биологических наук, профессор, академик РАН

**РОГАЕВ Евгений Иванович**

заведующий кафедрой генетики Биологического факультета Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В. Ломоносова», г. Москва

**Официальные оппоненты:**

доктор биологических наук, профессор

**ТАРАБЫКИН Виктор Степанович**

директор научно-исследовательского института нейронаук Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского», г. Нижний Новгород

доктор биологических наук

**ПЧЕЛИНА Софья Николаевна**

заведующий лабораторией молекулярной генетики человека Федерального государственного бюджетного учреждения «Петербургский институт ядерной физики им. Б.П. Константинова Национального исследовательского центра «Курчатовский институт», г. Санкт-Петербург

**Ведущая организация:**

Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет», г. Санкт-Петербург

Защита состоится «\_\_» \_\_\_\_\_ 2023 г. в \_\_\_\_ часов на заседании диссертационного совета 24.1.088.01 в Федеральном государственном бюджетном учреждении науки Институте общей генетики им. Н.И. Вавилова Российской академии наук по адресу: 119991, Москва, ул. Губкина, 3.

С диссертацией и авторефератом можно ознакомиться в библиотеке и на сайте Института [www.vigg.ru](http://www.vigg.ru), тел. 8-499-135-14-31.

**Автореферат разослан «\_\_» \_\_\_\_\_ 2023 года.**

Ученый секретарь диссертационного совета,  
доктор биологических наук

Горячева И. И.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### Актуальность темы исследования

Врождённые мозжечковые атаксии являются орфанными заболеваниями, которые приводят к нарушению моторных функций, задержке развития и инвалидности [Иллариошкин и др., 2006, Ashizawa, Xia, 2016]. По данным базы, объединяющей гены и клинические фенотипы, “Online Mendelian Inheritance in Man”, OMIM, известно более сотни различных мозжечковых атаксий. По типу наследования различают аутосомно-доминантные, аутосомно-рецессивные, X-сцепленные доминантные, X-сцепленные рецессивные и митохондриальные атаксии. Наиболее распространены атаксии, вызванные экспансией повторов и являющиеся нейродегенеративными заболеваниями с поздним началом и прогрессирующим течением [Клюшников и др., 2022]. К ним относятся аутосомно-доминантные спиноцеребеллярные атаксии (SCA1, 2, 3, 6, 7, 12 и 17), аутосомно-рецессивная атаксия Фридрейха, синдром мозжечковой атаксии, нейропатии и вестибулярной арефлексии (CANVAS) и синдром ломкой X-хромосомы с тремором/атаксией (FXTAS). Большинство аутосомно- и X-сцепленных рецессивных синдромов имеют раннее начало (до 25 лет). Частота встречаемости атаксий среди детей в европейской популяции составляет 26 на 100 000 человек [Musselman et al., 2014]. Атаксии с ранним началом могут быть вызваны как нейродегенерацией, например, атаксия-телеангиэктазия или атаксии с глазодвигательной апраксией 1-го и 2-го типа, так и связаны с нарушением развития мозжечка — врождённые заболевания. Врождённые мозжечковые атаксии представляют обширную группу редких клинических фенотипов, имеющих гетерогенную природу, большинство из которых отягощены умственной отсталостью и различными нарушениями со стороны других органов и тканей.

Геномные технологии массового параллельного секвенирования открывают большие возможности для исследования редких заболеваний при постановке диагноза и поиска причин заболеваний неизвестной этиологии, однако ввиду высокой стоимости и сложностей в биоинформатической интерпретации данных остаются недоступными для большинства пациентов [Claussnitzer et al., 2020]. Взамен используют панели кандидатных генов для конкретных заболеваний, однако в более чем половине случаев мозжечковых атаксий генетические причины остаются не идентифицированы [Ruano et al., 2014; Klockgether et al., 2019; Jiao et al., 2020; Galatolo et al., 2021; Sakamoto et al., 2022]. Одним из основных ограничений генетической диагностики с применением технологии массового параллельного секвенирования является отсутствие достаточных фундаментальных знаний об организации и регуляции генома, молекулярно-биологических знаний о строении и функциях белков и эволюционных процессах, лежащих в их основе. По последним данным геномного браузера Ensembl на 2022 год, в мировой человеческой популяции известно более 715 млн однонуклеотидных и 6 млн структурных вариантов, число которых растёт с увеличением геномных и экзомных данных [Cunningham et al., 2022]. Большая часть вариантов в геноме не критична для нормального

развития, однако малая часть может быть связана с орфанными заболеваниями [Xiao et al., 2021]. На сегодняшний день поиск патогенных вариантов наиболее успешен в белок-кодирующих последовательностях генов, имеющих связь с конкретными заболеваниями, но лишь в 26-33% случаев мозжечковых атаксий с неуточнённым диагнозом удаётся выявить патогенные варианты, ещё примерно в 15-20% случаев выявляемые варианты имеют неизвестное клиническое значение [Kim et al., 2020; Galatolo et al., 2021]. Среди кандидатных вариантов чаще всего (в 60-80% случаев) встречаются миссенс замены, что затрудняет предсказание их влияния на белок и требует дополнительных исследований [Galatolo et al., 2021, Xiao et al., 2021]. Мозжечковые атаксии характеризуются высокой гетерогенностью, поэтому стратегия анализа должна быть направлена на поиск патогенных вариантов не только в уже известных генах, но и в генах, ранее не связанных с мозжечковыми патологиями.

Для большинства мозжечковых атаксий отсутствует эффективное лечение, тем не менее, множество исследований направлены на разработку новых подходов в терапии данной группы заболеваний. Большинство исследований связано с разработкой подходов к лечению наиболее распространённых спиноцеребеллярных атаксий, вызванных экспансией повторов [Ashizawa et al., 2018; Vuijsen et al., 2019, VORBOLLA-JIMÉNEZ et al., 2021]. Тогда как для редких врождённых заболеваний более перспективным, вероятно, является развитие ранней диагностики и технологии геномного редактирования [Ilyas et al., 2020; Falabella et al., 2022; Reshetnikov et al., 2022].

Идентификация новых генетических дефектов является первоочередным шагом для понимания молекулярно-генетических механизмов патогенеза заболеваний как при определённой форме, так и для целой группы патологий. Расширение фундаментальных знаний, лежащих в основе развития мозжечка, моторных и когнитивных функций, играет важную роль для дальнейшей разработки лечения и профилактики заболеваний. Идентификация дефектных генов при врождённых мозжечковых атаксиях также полезна для избежания неэффективного и возможно даже приносящего вред симптоматического лечения. В современной мировой популяции наблюдается тенденция повышения генетического груза, поэтому выявление редких патогенных вариантов имеет важное значение при планировании семьи и в пренатальной диагностике.

Генетические причины наиболее распространённых форм мозжечковых атаксий известны и включены в диагностические панели, тогда как в настоящее время врождённые мозжечковые атаксии представляют наименее изученную группу наследственных атаксий, а спектр патогенных вариантов и кандидатных генов, связанных с данной группой заболеваний, требует изучения и расширения. В связи с этим является актуальным исследование генетических причин наиболее редких случаев врождённых патологий мозжечка. Одной из таких форм являются исследуемые в данной работе случаи редкой формы врождённой мозжечковой атаксии без клинически значимых когнитивных нарушений, выявленные в нескольких семьях из разных этнических групп [Шариошкин et al., 1996; Иллариошкин, 1997; Protasova et al., 2022]. Другим

случаем является синдром мозжечковой атаксии, умственной отсталости и квадропедии (CAMRQ), выявленный в близкородственной семье из Бразилии [Garcias, Roth, 2007].

### **Степень разработанности темы исследования**

Большинство врождённых мозжечковых атаксий отягощено умственной отсталостью и другими осложнениями [Wassmer et al., 2003; Poretti et al., 2014]. Отличительной особенностью исследуемых случаев является выраженное нарушение моторных функций, связанное с гипоплазией мозжечка, при котором сохранены когнитивные функции. На данный момент известны несколько заболеваний (SCAR6, SCAR19 и SCAR24), при которых у пациентов не было отмечено когнитивных нарушений, однако от исследуемых в данной работе врождённых мозжечковых патологий их отличает ряд симптомов. Наиболее близкая клиническая картина описана в семьях с патогенным вариантом в гене *ATP2B3*, приводящим к аминокислотной замене p.Gly1107Asp [Zanni et al., 2012].

Случаи синдрома CAMRQ встречаются крайне редко и ранее описаны в близкородственных семьях из Турции и Ирака [Türkmen et al., 2006; Ozelik et al., 2008; Tan, 2006; Tan, 2007; Valence et al., 2016]. Уникальным отличием от многих других врождённых патологий мозжечка, является тяжёлая форма умственной отсталости с полным отсутствием речи и локомоция исключительно на четырёх конечностях – квадропедия. Несмотря на близкую клиническую картину, наблюдается генетическая гетерогенность для данного синдрома. Наиболее распространены мутации в гене *VLDLR*, реже встречаются в генах *WDR81*, *ATP8A2*, *RELN*, *CA8* и *TUBB2B* [Moheb et al., 2008; Ozelik et al., 2008; Türkmen et al., 2009; Gulsuner et al., 2011; Onat et al., 2013; Breuss et al., 2017; Wali et al., 2021].

**Цель исследования** – выявить генетические причины редких случаев в семьях с врождённой мозжечковой атаксией и синдромом мозжечковой атаксии, умственной отсталости и квадропедии с использованием технологии массового параллельного секвенирования.

### **Задачи исследования:**

1. Провести массовое параллельное секвенирование образцов ДНК пациентов из разных семей с редкими формами мозжечковой атаксии.
2. Провести анализ данных массового параллельного секвенирования и выявить генетический дефект, являющийся причиной врождённой X-сцепленной рецессивной гипоплазии мозжечка у пациентов из большой бурятской семьи.
3. Провести генетический анализ кандидатных генов у пациентов с диагнозом врождённой мозжечковой атаксии из малых (ядерных) семей.
4. Разработать подход биоинформатического анализа данных массового параллельного секвенирования для идентификации новых патогенных генетических факторов в малых семьях или спорадических случаях с неопределённой этиологией.

5. С помощью разработанного подхода выявить генетические дефекты, приведшие к редкой спинномозжечковой атаксии в малых семьях.

6. Провести анализ данных массового параллельного секвенирования и выявить генетический дефект, являющийся причиной синдрома мозжечковой атаксии, умственной отсталости и квадропедии в близкородственной бразильской семье.

7. Установить эволюционный вклад генетического локуса, дефект которого связан с уникальным фенотипом в бразильской семье, и ответственного за формирование бипедальной локомоции, когнитивных функций и речи у современных людей.

### **Научная новизна**

В результате исследования редких форм врождённой мозжечковой атаксии были выявлены новые генетические варианты в генах, ответственных за развитие мозжечка. Более того, были выявлены высоко консервативные гены *LRCH2* и *CSMD1*, экспрессирующиеся в мозжечке, роль которых в развитии мозга мало изучена, и было впервые показано, что дефекты в этих генах могут приводить к мозжечковой атаксии без когнитивных нарушений. На примере генов, связанных с мозжечковой атаксией, был разработан подход биоинформатического поиска кандидатных генов генетических заболеваний на основе интеграции данных геномного секвенирования и “пространственно-временной” экспрессии паралогичных генов. Также были обнаружены особенности эволюции гена *GRID2*, вовлечённого в развитие мозжечка и отвечающего за формирование бипедальной локомоции, когнитивных функций и речи; обнаружена новая делеция в этом гене, приводящая к уникальному редкому синдрому мозжечковой атаксии, умственной отсталости и квадропедии.

### **Научно-практическая значимость**

Данное исследование представляет важное прикладное значение для молекулярно-диагностических исследований нарушений развития мозжечка, планирования семьи и пренатальной диагностики, расширяя список кандидатных генов, связанных с аутосомно- и X-сцепленными рецессивными атаксиям. Все новые мутации, выявленные у пациентов с гипоплазией и атрофией мозжечка, размещены в базе данных ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) и находятся в свободном международном доступе.

Более того, данное исследование полезно для дальнейших фундаментальных исследований молекулярных процессов развития и контроля моторных функций, так как оба гена *LRCH2* и *CSMD1*, ассоциированных по результатам исследования с мозжечковыми атаксиями, экспрессируются в интернейронах мозжечка и, вероятно, участвуют в одних и тех же биологических процессах, но на разных этапах онтогенеза.

### **Основные положения, выносимые на защиту:**

1. Редкая форма спиноцеребеллярной атаксии с ранним началом без когнитивных нарушений генетически гетерогенна, фенотипически близкие синдромы могут быть вызваны генетическими дефектами разных генов, вовлечённых в разные биологические пути в клетках мозжечка.

2. Мутации в лейцин-богатом повторах домене гена *LRCH2*, экспрессирующегося в мозжечке в эмбриональный период, могут приводить к врождённой X-сцепленной рецессивной спиноцеребеллярной атаксии с генерализованной полинейропатией без когнитивных нарушений.

3. Миссенс мутации в комплемент-связывающем домене гена *CSMD1*, экспрессирующегося в развивающемся мозжечке и участвующем в развитии нейрональных отростков и образовании синапсов, могут приводить к аутосомно-рецессивной спиноцеребеллярной атаксии без когнитивных нарушений с ранним началом. Выявленный патогенный вариант является редким и имеет эффект основателя.

4. Миссенс мутация в гене *ABCB7*, в области, кодирующей внутримитохондриальный домен, приводит к X-сцепленной рецессивной гипоплазии мозжечка и наследуется в общей группе сцепления с делецией в гене *ATP7A*, которая, возможно, является модифицирующим фактором фенотипа.

5. Миссенс мутация, расположенная в седьмом трансмембранном домене гена *GRM1*, приводит к врождённой аутосомно-рецессивной спиноцеребеллярной атаксии.

6. Делеция 5-7 экзонов в гене *GRID2*, кодирующей область в аминотерминальном домене, приводит к синдрому мозжечковой атаксии, умственной отсталости и квадропедии с полным отсутствием речи.

### **Личный вклад автора**

Все основные результаты диссертационной работы получены лично автором, либо при его непосредственном участии. Данные массива SNP6.0 на платформе Affymetrix были получены в сотрудничестве с Базельским Университетом в Швейцарии. Генетический анализ в семьях АХ и Q выполнен при непосредственном участии совместно с Григоренко А. П. и сотрудниками лаборатории эволюционной геномики. Автором лично проведено написание первичного текста статей с результатами исследований и непосредственное совместное участие в подготовке их для публикации в научных рецензируемых журналах, принято участие в написании тезисов и личное представление их в конференциях.

### **Апробация работы**

Апробация диссертационной работы проведена на межлабораторном семинаре ИОГен РАН (протокол № 8 от 28 ноября 2022 г.).

Результаты диссертационной работы были представлены на российских и международных конференциях: Конференция “Гены, Мозг, Поведение” Москва, 9 декабря 2019; 2nd International scientific conference “Science of the Future” Kazan,

September 20–23, 2016; VII съезд Российского общества медицинских генетиков 19-23 мая 2015 г. Санкт-Петербург; “Aging and Memory – 2015”, November 14, 2015 Moscow, Russia; Science of The Future, 17-20 September 2014, Saint-Petersburg.

### **Публикации**

Результаты исследования представлены в 7 научных публикациях, в том числе в 3 статьях в ведущих научных журналах, индексируемых в базах данных Scopus и Web of Science и рекомендованных ВАК для защиты диссертаций.

### **Структура и объем работы**

Диссертационная работа изложена на 206 страницах машинописного текста и включает следующие разделы: введение, обзор литературы, материалы и методы, результаты, обсуждение результатов, заключение, выводы, благодарности, список литературы и приложения. Иллюстративный материал диссертации представлен в виде 9 таблиц и 22 рисунков, приложение содержит 20 таблиц и 22 рисунка. Список литературы включает 490 литературных источников, из которых 476 источников зарубежной литературы.

### **Благодарности**

Автор выражает благодарность академику РАН, проф., д.б.н. Рогаеву Е. И., под руководством которого выполнена данная диссертационная работа. Автор благодарен к.б.н. Григоренко А. П., Гусеву Ф. Е., к.б.н. Андреевой Т. В., к.б.н. Кузнецовой И. Л., к.б.н. Кунижевой С. С. и всему коллективу лаборатории эволюционной геномики ИОГен РАН. Автор выражает благодарность академику РАН, проф., д.м.н. Иллариошкину С. Н., к.м.н. Ключникову С. А., док., проф., Гарсиасу Г. Л. и док., проф., Мартино-Роф М. Д. Г. за предоставленные ценные биологические образцы пациентов с редкими формами заболеваний.

### **СОДЕРЖАНИЕ РАБОТЫ**

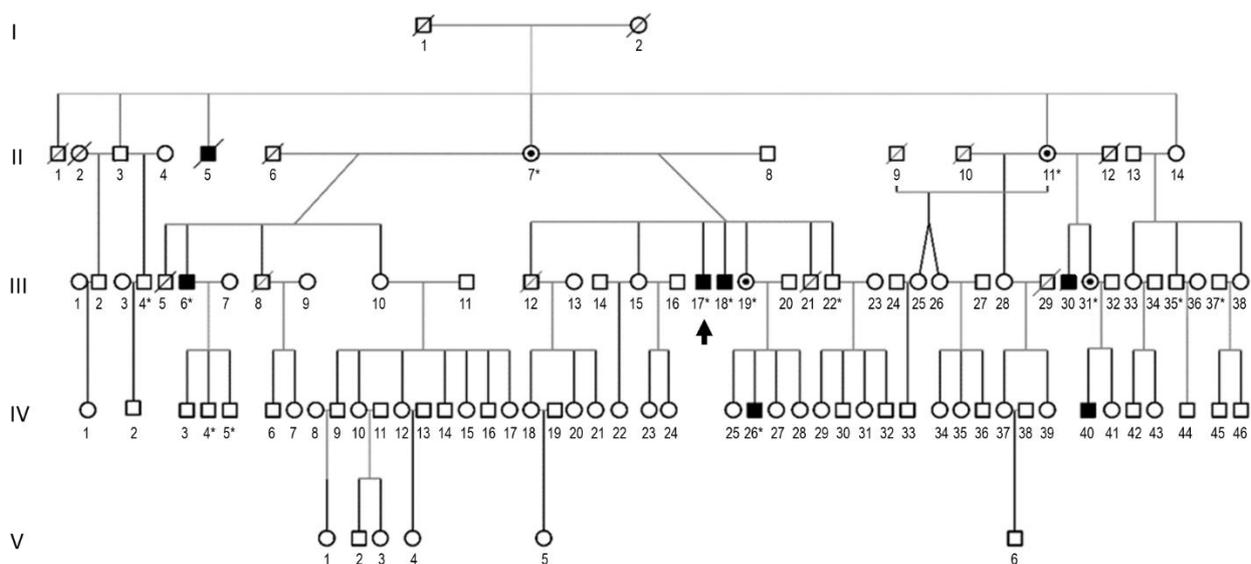
В диссертационной работе глава «**Обзор литературы**» состоит из пяти разделов. Первый раздел посвящён распространённости мозжечковых атаксий и генетике наиболее часто встречающихся форм атаксий. Во втором разделе рассматриваются общие молекулярно-генетические механизмы, нарушение которых лежит в основе мозжечковых патологий. В третьем и четвёртом разделах уделено внимание редким формам врождённых мозжечковых атаксий и синдромам мозжечковой атаксии, умственной отсталости и квадропедии, клинические формы которых близки к исследуемым в данной работе заболеваниям. В пятом разделе описана история развития и рассмотрены современные методы генетических исследований.

### **МАТЕРИАЛЫ ИССЛЕДОВАНИЯ**

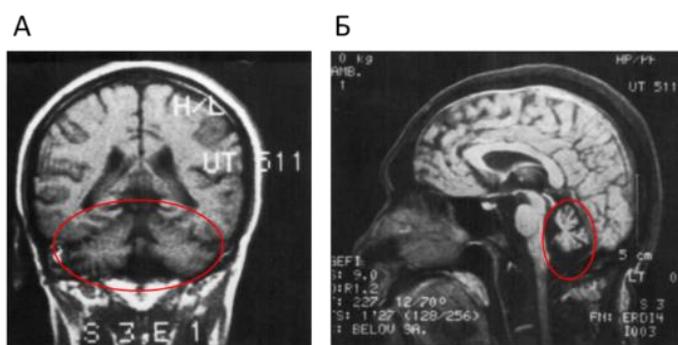
Образцы ДНК и клиническое описание пациентов с мозжечковой атаксией были получены из Научного Центра Неврологии. Образцы крови от пациентов с синдромом мозжечковой атаксии, умственной отсталости и квадропедии из близкородственной семьи были получены ранее в рамках экспедиции в регион Бразилии. Информированные согласия были получены от всех индивидов или их опекунов, участвующих в генетическом исследовании. Генетические

исследования были одобрены Этическим Комитетом Научного Центра Неврологии и Этическим Комитетом Института Общей Генетики им. Н.И. Вавилова Российской Академии Наук.

**Большая бурятская семья.** Большая бурятская семья (АХ) представлена родословной в 5 поколений, в которой у некоторых индивидов мужского пола диагностирована мозжечковая атаксия (Рисунок 1). У всех пациентов наблюдалась задержка моторного развития, атаксия конечностей и туловища, дизартрия, нистагм, офтальмоплегия и повышение сухожильных рефлексов. Течение заболевания было непрогрессирующим. Нарушения памяти или когнитивных функций не было. При МРТ выявлена гипоплазия полушарий и червя мозжечка (Рисунок 2). Признаков сидеробластной анемии или нарушений меди выявлено не было. Показатели общего анализа крови и мочи были в норме.



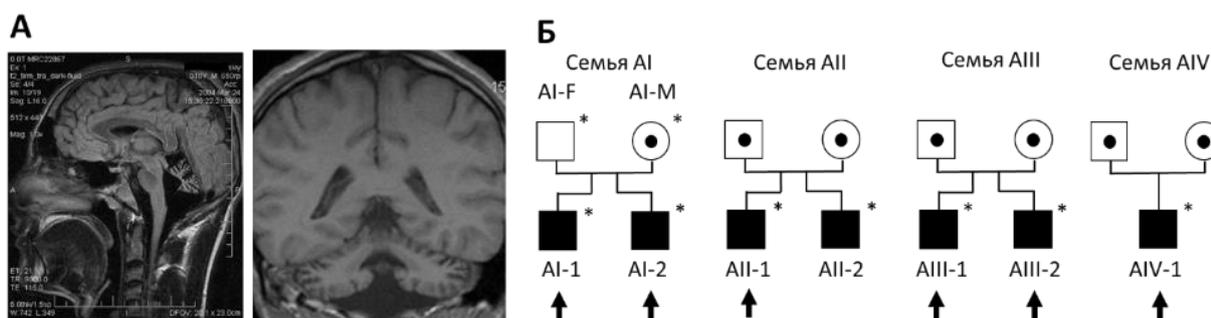
**Рисунок 1.** Родословная семьи АХ [Илариошкин et al., 1996]. Квадратные символы обозначают индивидов мужского пола, круглые символы — индивиды женского пола. Сплошными символами обозначены пациенты с диагнозом мозжечковой атаксии; символы с точками — облигатные гетерозиготные носители; прозрачные символы — здоровые индивиды; перечёркнутые символы — индивиды, умершие до момента исследования. Полногеномное секвенирование выполнено для пациента, обозначенного стрелкой; звёздочкой обозначены индивиды, для которых проведено генотипирование кандидатных локусов.



**Рисунок 2.** Снимки МРТ головного мозга пациента АХ-III-30 [Илариошкин et al., 1996]. А — Корональное T1-взвешенное изображение, выявлена гипоплазия полушарий и червя мозжечка Б — Срединно-сагитальное T1-взвешенное изображение, выявлена гипоплазия каудального

червя и парамедианных зон полушарий мозжечка, расширение борозды в переднем верхнем черве, увеличение четвёртого желудочка и большой цистерны, и небольшое истончение верхней части покрывки ствола мозга, также наблюдается умеренное расширение борозд в латеральных отделах полушарий головного мозга.

**Московская семья.** В Московской семье (AI) оба брата AI-1 и AI-2 с младенчества имели моторную задержку развития, нарушение координации и плавности движений, замедленную и скандирующую речь (Рисунок 3) [Protasova et al., 2022]. Течение заболевания было непрогрессирующим. При обследовании пациентов в возрасте 33 (AI-1) и 23 (AI-2) лет были выявлены: дизартрия со скандирующей речью, мышечная гипотония, дискоординация конечностей, интенционный тремор, положительная проба Ромберга, самостоятельная атаксическая походка на широком основании. У пациента AI-2 также отмечался выраженный осцилляторный тремор рук, туловища и головы. При МРТ выявлена выраженная гипоплазия полушарий и червя мозжечка (Рисунок 3А). Исследования нервной проводимости с помощью стимуляционной ЭНМГ показали генерализованное сенсомоторное поражение демиелинизирующего типа. Нарушений когнитивных функций не было отмечено.

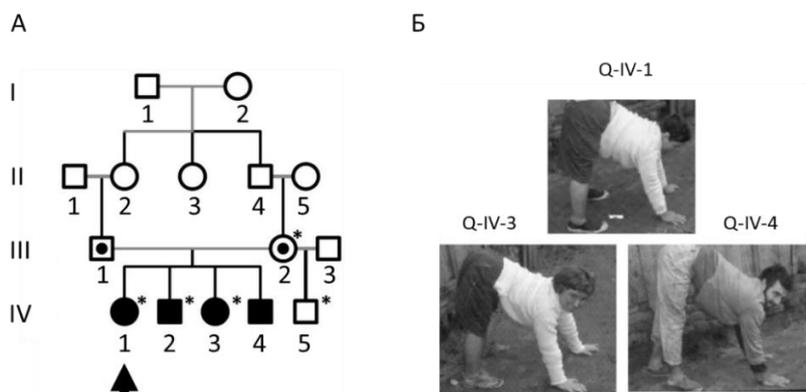


**Рисунок 3.** Малые семьи с мозжечковой атаксией. А – Снимки МРТ головного мозга пациента AI-1, показана выраженная гипоплазия червя (слева) и полушарий (справа) мозжечка. Б – Родословные семей AI-IV. Квадратные символы обозначают индивидов мужского пола, круглые символы – индивиды женского пола. Звёздочкой обозначены индивиды, для которых проведён генетический анализ; стрелками – индивиды, для которых проведено полноэкзомное секвенирование.

**Северо-Кавказские семьи.** В семьях AII и AIII раннее развитие было нормальным (Рисунок 3) [Protasova et al., 2022]. Дебют заболевания варьировал от 8 до 17 лет. В начале наблюдалась шаткость при ходьбе, нарушение координации и смазанность речи. У пациента AII-1 в возрасте 14 лет выявлен горизонтальный нистагм, нарушение координации и дисдиадохкинез, интенционный тремор правой руки, неустойчивость в пробе Ромберга и атаксическая походка без опоры. У пациентов AIII-1 и AIII-2 в возрасте 22 и 20 лет, соответственно, выявлены ротаторный нистагм, нарушение координации и дисдиадохкинез, дизартрия, положительная проба Ромберга и неустойчивая походка без поддержки. При МРТ была выявлена гипоплазия полушарий и червя мозжечка. У пациента AIV-1 с рождения наблюдалась задержка моторного развития. При обследовании в третьей декаде жизни были выявлены: дизартрия,

атаксия туловища, негрубая дискоординация при выполнении динамических проб в руках, врождённая гипоплазия мозжечка. У пациентов из семей АII-IV клинически значимых нарушений интеллекта не выявлено.

**Бразильская семья.** У всех пациентов из близкородственной бразильской семьи (Q) с рождения была задержка психомоторного развития, неспособность поддержания позы стоя без опоры, квадропедия (передвижение на всех четырёх конечностях), расходящееся косоглазие, у некоторых судороги, из внешних признаков: гирсутизм, грубые черты лица, короткие и маленькие кисти и стопы, небольшой рост. Все пациенты имели тяжёлую умственную отсталость с отсутствием речи, дружелюбное поведение без признаков аутизма (Рисунок 4) [Garcias, Roth, 2007].



**Рисунок 4.** Семья Q. А – родословная в четыре поколения (I - IV). Полногеномное секвенирование выполнено для пациента Q-IV-1, обозначенного стрелкой. Квадратные символы обозначают индивидов мужского пола, круглые символы – индивиды женского пола. Сплошные символы обозначают больных индивидов; символы с точками – облигатные гетерозиготные носители; прозрачные символы – здоровые индивиды; звёздочки – индивиды, для которых было проведено генетическое исследование. Б – фотографии пациентов [Garcias, Roth, 2007].

## МЕТОДЫ ИССЛЕДОВАНИЯ

**Приготовление библиотек ДНК и их массовое параллельное секвенирование.** Выделение ДНК проводилось из периферической крови с помощью набора QIAamp DNA Blood Mini Kit (QIAGEN, Хильден, Германия). Приготовление геномных библиотек проводилось из 2 мкг ДНК с помощью набора Paired-End DNA Sample Prep Kit (Illumina, Калифорния, США). Приготовление экзомных библиотек из 200 нг ДНК с помощью наборов TruSeq Exome kit (Illumina, Калифорния, США) в коммерческой лаборатории Генетико. Приготовление экзомных библиотек из 400 нг ДНК с помощью набора SureSelect Human All Exon V6 (Agilent Technologies, Калифорния, США) в лаборатории эволюционной геномики. Валидация размера и молярности библиотеки проводилась с помощью набора High Sensitivity DNA kit на электрофоретическом биоанализаторе Bioanalyzer 2100 (Agilent Technologies, Калифорния, США). Определение точной молярной концентрации проводилось с помощью количественной ПЦР в реальном времени с использованием набора

KAPA Library Quantification Kit (KAPA Biosystems, Массачусетс, США) на амплификаторе 7500 Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific, Калифорния, США).

Полноэкзомное и полногеномное массовое параллельное секвенирование проводилось на платформах HiSeq 2000/2500 System и NovaSeq 6000 System (Illumina, Калифорния, США) (Таблица 1). Секвенирование геномных библиотек (пациентов АХ-III-17 и Q-IV-1) и экзомных библиотек (пациентов АIII-1 и АIII-2) было проведено на платформе HiSeq 2000/2500 System в лаборатории Эволюционной геномики ИОГен РАН. Секвенирование экзомных библиотек пациентов AI-1, AI-2, АII-1 и AIV-1 было проведено на платформе NovaSeq 6000 System в коммерческой фирме Генетико.

**Таблица 1.** Анализ исследуемых семей с мозжечковыми атаксиями с помощью массового параллельного секвенирования

Диагноз	Семья, №	Пациенты	Метод	Прибор	Покрытие генома GRCh37
Врожд. рец. X-сцепл. мозж. атаксия	АХ	АХ-III-17	Полногеномный анализ	HiSeq 2000/2500	28X
Врожд. мозж. атаксия	AI	AI-1, AI-2	Полноэкзомный анализ (TruSeq Exome kit)	NovaSeq 6000	112-116X; 99,9% локусов прочтены
Мозж. атаксия	AII	AII-1	Полноэкзомный анализ (TruSeq Exome kit)	NovaSeq 6000	135X; 98,9% локусов прочтены
Мозж. атаксия	AIII	AIII-1, AIII-2	Полноэкзомный анализ (SureSelect Human All Exon V)	HiSeq 2000/2500	17X; 99,1% локусов прочтены
Врожд. мозж. атаксия	AIV	AIV-1	Полноэкзомный анализ (TruSeq Exome kit)	NovaSeq 6000	136X; 94,6% локусов прочтены
Синдром SAMRQ	Q	Q-IV-1	Полногеномный анализ	HiSeq 2000/2500	90X

**Методы биоинформатического анализа данных массового параллельного секвенирования.** Качество полученных с прибора необработанных чтений в формате fastq оценивалось с помощью программы FastQC. Чтения были выравнены на референсный геном человека GRCh37/hg19 с помощью программы BWA MEM [Li et al., 2013]. Дублицированные чтения были определены с помощью программы MarkDuplicate пакета The Genome Analysis Toolkit (GATK) [McKenna et al., 2010]. Качество выравнивания определялось с помощью программы SAMtools flagstat. Средняя глубина покрытия и глубина покрытия отдельных локусов определялась с помощью программы DepthOfCoverage из программного пакета GATK. ОНП и инделы были определены с помощью пакета программ GATK [McKenna et al., 2010; DePristo et al., 2011; Van der Auwera et al., 2013]. Варианты были аннотированы с помощью программы VEP [McLaren et al., 2016]. Структурные варианты были

определены с помощью программ Pindel и BreakDancer [Ye et al., 2009; Chen et al., 2009]. Были отобраны все редкие варианты с MAF < 5% по данным проекта 1000 Геномов. Затем были отобраны варианты, которые соответствуют следующим критериям: расположение в эволюционно консервативных сайтах, изменение белок-кодирующей последовательности, топологии аминокислотных замен в белке, предсказание патогенности на основе биоинформатических программ, экспрессии генов в мозжечке и наличия данных о связи с мозжечковой патологией. Предсказание пространственной структуры белков проводилось с помощью Phyre2 и iTasser [Kelley et al., 2015; Yang, Zhang, 2015]. Анализ экспрессии генов и их паралогов был проведён с помощью специально разработанной онлайн программы “Essential Parologue by Expression” (EPbE) [Protasova et al., 2022].

**Проверка кандидатных вариантов.** Проверка вариантов, а также анализ экзонов кандидатных генов проводился с помощью ПЦР и секвенирования по Сэнгеру на приборе 3730xl DNA Analyzer (Applied Biosystems, Thermo Fisher Scientific, Калифорния, США).

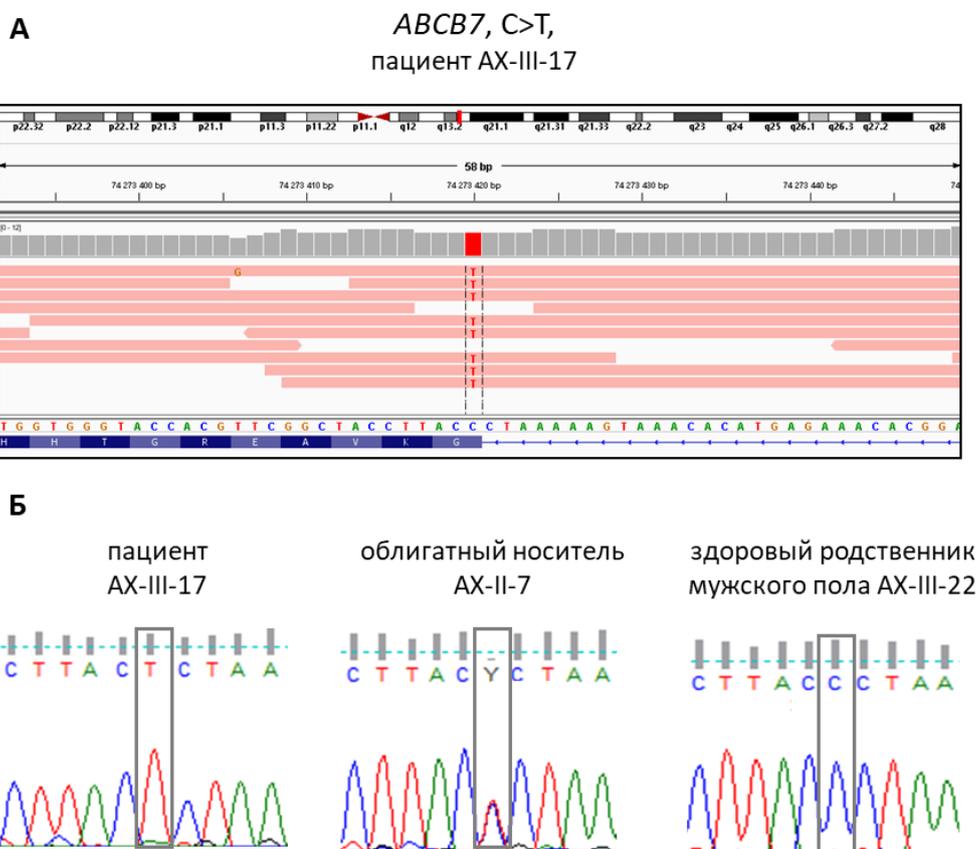
**Методы эволюционного анализа.** Белок-кодирующие последовательности гена *GRID2* были получены из базы данных NCBI [Pruitt et al., 2014], Ensembl [Aken et al., 2016] и геномного браузера UCSC [Karolchik et al., 2004]. Филогенетическое дерево включало 23 вида приматов и было реконструировано в соответствии с опубликованной филогенией [Perelman et al., 2011]. Эволюционный анализ проводился с помощью программы CodeML в пакете PAML v4.8 [Yang, 1997; Yang, 2007]. Для синонимичных замен был проведён анализ изменения свободной энергии Гиббса, рассчитанный для фрагментов длиной 25, 51, 75, 151 и 301 нуклеотидов и полноразмерных кодирующих последовательностей с помощью программы mfold [Zuker, 2003; Zuker, 1998; Waugh, 2002]. Уровень статистической значимости оценивался с помощью Критерия Знаков.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ

**Анализ генетических причин X-сцепленной рецессивной мозжечковой атаксии в семье АХ.** Было проведено массовое параллельное секвенирование генома пациента АХ-III-17 (Таблица 1, Рисунок 1). В общей сложности предсказано 4,2 млн однонуклеотидных вариантов, инделов и структурных вариантов. Поиск кандидатных вариантов был проведён в хромосомном локусе Xp11.21-q24 между STR маркерами DXS991 и DXS1001, соответствующим координатам X:55519052-119836873 генома человека версии GRCh37, ранее определённого с помощью анализа парного сцепления и реконструкции гаплотипов [Illarioshkin et al., 1996]. Заданным критериям соответствовало два варианта: миссенс вариант в гене *ABCB7* и большая делеция в гене *ATP7A*.

В гене *ABCB7* была найдена миссенс замена С>Т в позиции X:74273420 (NM\_001271696.3:c.2044G>A), отсутствующая в базах данных 1000 Геномов и gnomAD (позднее был присвоен rs797044558 в базе данных dbSNP) (Рисунок 5). Данный ген кодирует АТФ-связывающий транспортёр субсемейства В член 7,

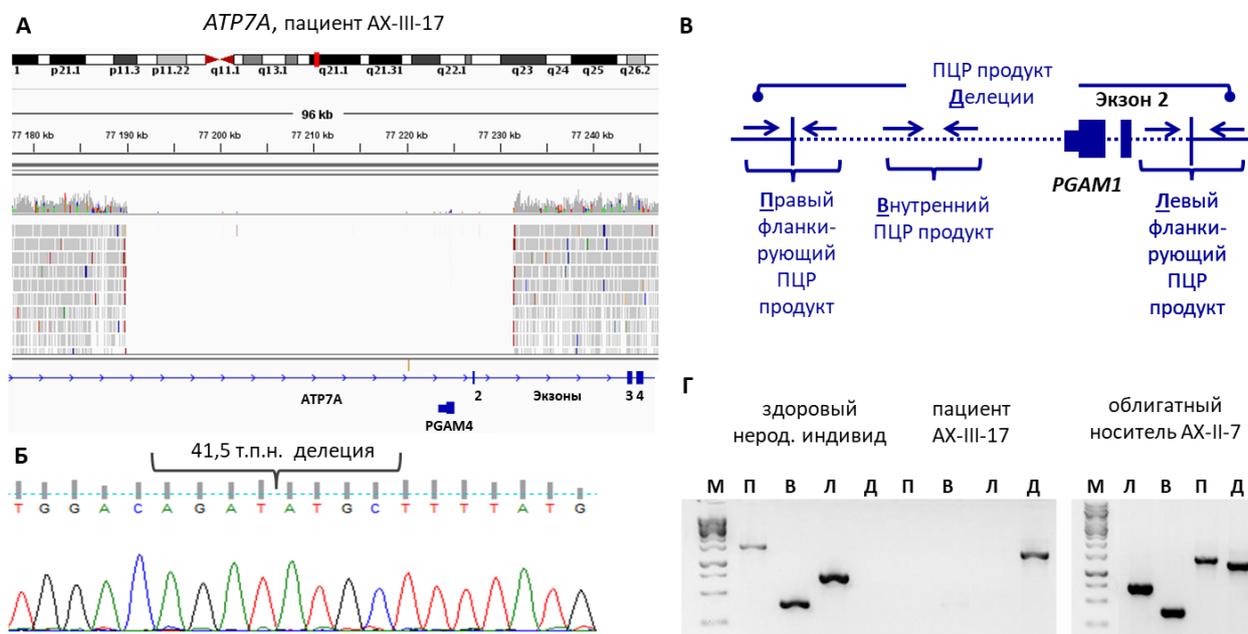
расположенный на внутренней митохондриальной мембране, играющий важную роль в транспорте ионов железа и синтезе гемопротейнов [Hollenstein et al., 2007]. В период пренатального развития *ABCB7* экспрессируется в мозжечке, а дефекты гена связаны с сидеробластной анемией и атаксией [Nikpour et al., 2013; D'Hooghe et al., 2012; Sato K., 2011; Boulwood et al., 2008; Pondarre et al., 2007].



**Рисунок 5.** Идентификация нового кандидатного варианта в гене *ABCB7* в семье АХ. А – Визуализация области генома, содержащей миссенс замену chrX:g.74273420C>T с помощью программы IGV; Б – Валидация гемизиготного варианта у пациента АХ-III-17, гетерозиготы у облигатного носителя, его матери АХ-II-7, и гомозиготы дикого типа у здорового родственника мужского пола с помощью секвенирования по Сэнгеру.

Также была выявлена крупная делеция X:77190006-77231471 в гене *ATP7A* (NC\_000023.10:g.77190006\_77231471del), приводящая к полному удалению второго экзона и значительной интронной области (Рисунок 6). Белок, кодируемый данным геном, является АТФазой транспортёром меди, располагающимся на клеточных мембранах [Uhlemann et al., 2021]. *ATP7A* экспрессируется во многих тканях, в значительной степени в печени, гематоэнцефалическом барьере, в том числе в мозге [Telianidis et al., 2013; Kapushesky et al., 2009; Fagerberg et al., 2014; Miller et al., 2014; Xu et al., 2019]. Мутации в *ATP7A* приводят к нарушению транспорта меди и могут приводить к нескольким заболеваниям: болезни Менкеса [Møller et al., 2013; Bull et al., 1993; Chelly et al., 1993], X-сцепленной дистально-спинально-мышечной атрофии 3-го типа [Kennerson et al., 2010], синдрому затылочного рога [Byers et al., 1980; Das et al., 1995; Ronce et al., 1997]. В результате делеции в гене *ATP7A* также

происходит полное удаление ретрогена *PGAM4*, представляющего функциональную ретрокопию гена фосфоглицератмутазы *PGAM1*. Так как *PGAM4* не экспрессируется в мозге, данный ретроген не является кандидатным фактором гипоплазии мозжечка.

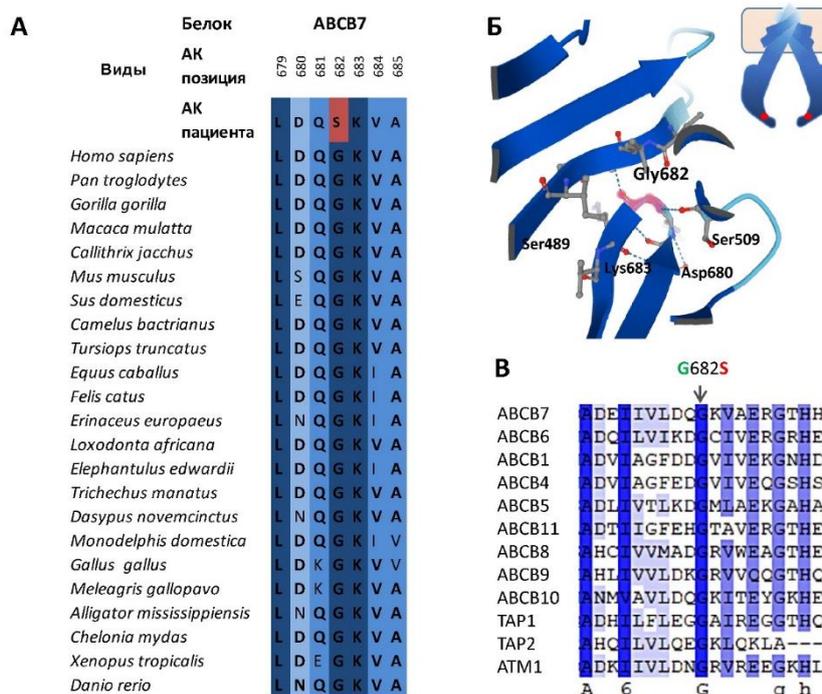


**Рисунок 6.** Идентификация делеции в гене *ATP7A* в семье АХ. А – Визуализация области делеции chrX:g.77190006\_77231471del с помощью программы IGV. Б – Валидация гемизиготной делеции у пациента АХ-III-17 с помощью секвенирования по Сэнгеру. В – Схема системы праймеров для валидации делеции в гене *ATP7A*: пара праймеров, комплементарных обеим геномным границам делеции, подтверждает наличие делеции (Д); три пары праймеров подтверждают отсутствие делеции, из них две комплементарны границам делеции (П и Л) и одна – области внутри делеции (В). Г – Электрофореграмма ПЦР продуктов: наличие трёх продуктов (П, В и Л) и отсутствие продукта Д подтверждает гемизиготу дикого типа у здорового неродственного индивида; наличие только продукта Д подтверждает гемизиготную делецию у пациента АХ-III-17; наличие всех четырех продуктов подтверждает гетерозиготную делецию у облигатного носителя АХ-II-7, матери исследуемого пациента.

Миссенс замена в гене *ABCB7* и делеция в гене *ATP7A*, были подтверждены с помощью ПЦР и секвенирования по Сэнгеру в гемизиготном состоянии у всех исследуемых пациентов (4 инд.), в гетерозиготном состоянии у всех исследуемых клинически здоровых женщин-носителей (4 инд.) и отсутствовали у клинически здоровых мужчин (4 инд.) из семьи АХ.

**Анализ пространственной структуры белков.** Так как замена в *ABCB7* происходит в первом кодоне 16 экзона рядом с сайтом акцептора сплайсинга был проведён биоинформатический тест, в основе которого лежит анализ по частотам нуклеотидов в каждом положении сайта (от -20 до +3), и было предсказано отсутствие изменений сайта сплайсинга [Stamm et al., 2000; Nurtdinov et al., 2006]. Дополнительный анализ с помощью онлайн программы SpliceAI [Jaganathan et al., 2019], также показал отсутствие изменений для всех

транскриптов. Затем было проведено исследование влияния аминокислотной замены p.Gly682Ser, расположенной во внутримитохондриальном нуклеотид-связывающем домене (NBD), (Рисунок 7Б). Сравнительный анализ ортологов и паралога ABC транспортёров показал высокую консервативность аминокислоты Gly682 (Рисунок 7А, В). С помощью дополнительного анализа в программе Phyre2 предсказано, что замена на серин в данном положении может приводить к появлению вращения в неподвижном сайте. В дополнение к этому замена p.Gly682Ser предсказана “патогенной” и “наиболее вероятно патогенной” с помощью программ SIFT и PolyPhen2.



**Рисунок 7.** Анализ области, содержащей замену p.Gly682Ser (G682S) в транспортёре ABCB7. А – Эволюционный анализ аминокислотных последовательностей ортологов; Б – пространственная структура области ABCB7 транспортёра, содержащей Gly682, представлена бета-складкой (создана AlfaFold). В верхнем правом углу схематичное изображение ABCB7 транспортёра, выявленная мутация выделена красным. В – Эволюционный анализ паралога человека и гомолога ATM1 (*Saccharomyces cerevisiae*). АК – аминокислота.

Отличительной особенностью случая в семье АХ является отсутствие классических симптомов сидеробластной анемии. Возможно, это связано с уникальным расположением мутации во внутримитохондриальном домене ABCB7, тогда как во всех ранее описанных случаях мутации были локализованы вблизи или внутри трансмембранных доменов [Nikpour et al., 2012; D'Hooghe et al., 2012; Sato, 2011; Boultwood et al., 2008, Pondarre et al., 2007]. Недавно в базе данных ClinVar опубликован независимый генетический анализ, в котором у гватемальского пациента с гипоплазией мозжечка выявлен тот же вариант в гене ABCB7, что и найденный в семье АХ [Landrum et al., 2018]. Описанная клиническая картина у гватемальского пациента имеет высокое сходство с

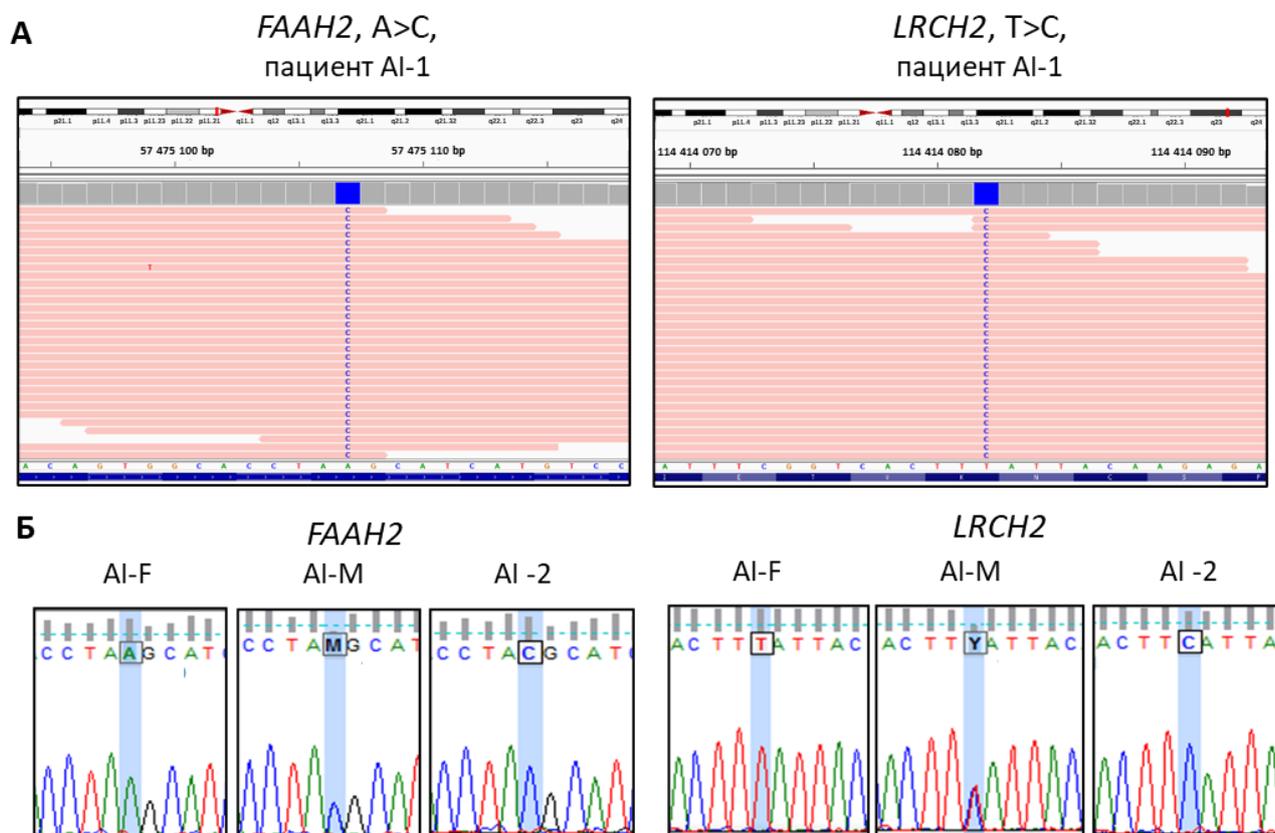
описанной в бурятской семье, и также не отмечено симптомов сидеробластной анемии.

Делеция в гене *ATP7A* приводит к удалению сайта инициации трансляции в мажорном транскрипте, в результате происходит усечение белка на 68 аминокислот. Наличие альтернативного сайта инициации трансляции позволяет сохранить остальную часть белка. Делеция приводит к удалению первого из шести металл-связывающих доменов (МД) во всех длинных изоформах (Рисунок 9). В малых изоформах, функциональное значение которых неизвестно, происходит нарушение рамки считывания и их полная утрата. Первый МД консервативен у большинства млекопитающих, а также у некоторых рептилий. Тогда как у низших эукариот и прокариот присутствуют только от одного до трёх металл-связывающих доменов в ортологе *ATP7A* транспортёра. В связи с чем можно предположить, что потеря одного из МД не разрушает функцию белка, тем не менее может влиять на изменение эффективности транспорта ионов меди. В связи с чем данная делеция не является основной причиной заболевания, однако нельзя исключать минорный эффект на фенотип.

**Анализ родства в семье АХ.** Поскольку делеция в гене *ATP7A* приводит к полному удалению ретрогена *PGAM4*, экспрессируемого исключительно в семенниках, и в литературе предполагается его роль в созревании мужских половых клеток (Illumina Human Body Map 2.0, номер доступа ArrayExpress E-MTAB-513, Okuda et al., 2012), нами был проведен анализ его возможной роли в фертильности. Несмотря на наличие спорных данных о влиянии мутаций в *PGAM4* на мужское бесплодие [Okuda et al., 2012; Jin et al., 2013], один из пациентов АШ-6 имеет трёх здоровых сыновей. У его исследуемых сыновей (2 инд.) было подтверждено отсутствие делеции, что согласуется с наследованием X хромосомы по материнской линии. Затем был проведён анализ STR генетических маркеров на родство данного пациента и его сыновей, и было подтверждено подлинное биологическое отцовство. Таким образом, отсутствие данного гена не ведёт к полному нарушению фертильности. Однако, учитывая, что все потомки пациента были мужского пола, нельзя исключать патогенный эффект потери гена на сперматозоиды, несущие X-хромосому.

**Анализ генетических причин мозжечковой атаксии в семьях АI-IV.** Клиническая картина в семьях АI-IV отличается от наиболее распространённых форм мозжечковых атаксий, однако имеет большое сходство со случаями в семье АХ и с описанным ранее [Zanni et al., 2012]. В семьях АI и АIV заболевание было выявлено с младенчества. В семьях АII и АIII первые симптомы появились в позднем детстве или в пубертатный период. В связи с этим, в первую очередь, был проведён анализ всех экзонов гена *ABCB7*, области делеции в гене *ATP7A* с помощью секвенирования по Сэнгеру, но не было выявлено мутаций в белок-кодирующей последовательности. Также не было выявлено мутаций в 20 экзоне гена *ATP2B3*, включая описанный ранее вариант p.Gly1107Asp [Zanni et al., 2012]. Далее было проведено полноэкзомное секвенирование у шести пациентов из семей АI-IV. В среднем было выявлено 13,8 тыс. гомозиготных и 26,1 тыс. гетерозиготных вариантов на образец. Затем были отобраны все редкие

варианты, которые присутствовали у обоих пациентов из одной семьи (MAF <0,01) и отсутствовали у здоровых индивидов. Было выявлено два варианта с неизвестной клинической значимостью, соответствующие заданным критериям патогенности (Рисунок 8).



**Рисунок 8.** Идентификация редких вариантов в семье AI. А – Визуализация областей генома, содержащих кандидатные миссенс варианты в генах *FAAH2* (hg19 chrX:g. 57475107A>C) и *LRCH2* (hg19 chrX:g.114414082T>C) с помощью программы IGV. Б – Валидация кандидатных гемизиготных вариантов у пациентов I-1 и I-2, гетерозиготных вариантов у их матери (I-M) и гемизиготных вариантов дикого типа у их отца (I-F).

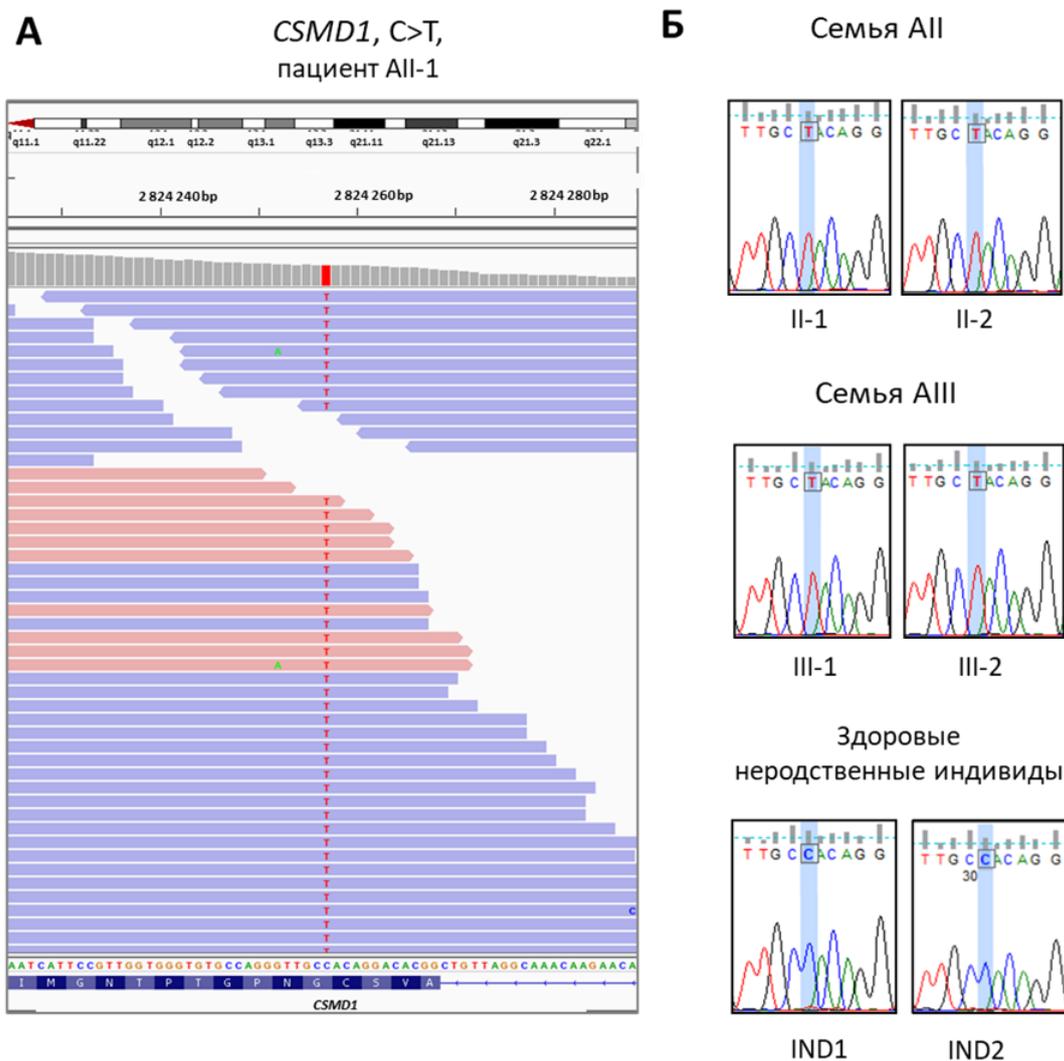
Первый вариант rs760074041, hg19 chrX:g.57475107A>C (NM\_174912.4:c.1379A>C), был обнаружен в гене *FAAH2*, приводящий к аминокислотной замене p.Lys460Thr. Вариант предсказан как «допустимый» с помощью программы SIFT и «возможно патогенный» с помощью PolyPhen2. *FAAH2* отсутствует в геноме некоторых животных, например, у мышей. Однако лизин консервативен во всех ортологах. *FAAH2* экспрессируется в различных тканях, включая мозжечок [Lindsay et al., 2016; GTEx Consortium, 2013; Miller et al., 2014]. Кроме того, есть публикация, в которой другая замена в данном гене найдена у пациента с атаксией и аутизмом [Sirrs et al., 2015]. Несмотря на некоторые различия в фенотипах, в обоих случаях наблюдалась врождённая атаксия, поэтому этот вариант был выбран для дальнейшего анализа.

Второй вариант, hg19 chrX:g.114414082T>C (NM\_020871.4:c.772A>G), был идентифицирован в высоко консервативной области гена *LRCH2* (Рисунок 8). Найденный вариант отсутствует в базах данных Проекта 1000 Геномов,

gnomAD и dbSNP. Вариант приводит к аминокислотной замене p.Lys258Glu, и с помощью программ SIFT и PolyPhen2 был предсказан как «патогенный» и «возможно патогенный». Ген *LRCH2* эволюционно высоко консервативен и экспрессируется в различных областях мозга, включая мозжечок [Lindsay et al., 2016; GTEx Consortium, 2013; Miller et al., 2014]. Однако функции гена практически не изучены, и нет данных о врождённых заболеваниях, вызванных дефектами в данном гене. С помощью секвенирования по Сэнгеру было подтверждено, что выбранные варианты в генах *FAAH2* и *LRCH2* являются гемизиготными у обоих пациентов, гетерозиготными у их матери и отсутствуют у их отца (Рисунок 8Б).

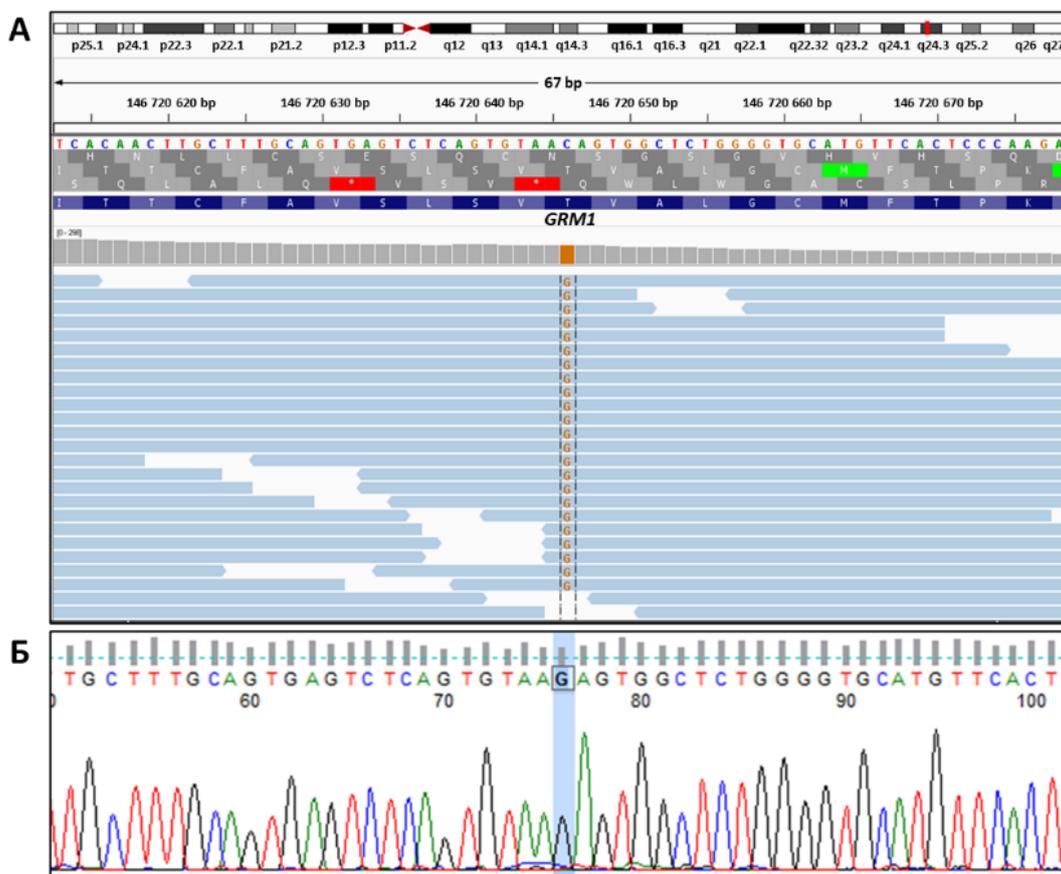
У всех четырех пациентов из семей АII и АIII был идентифицирован один редкий гомозиготный миссенс вариант rs117633452, hg19 chr8:g.2824257 C>T (NM\_033225.6:c.8935G>A), в гене *CSMD1* (Рисунок 9). Выявленный вариант приводит к аминокислотной замене p.Gly2979Ser в высоко консервативной области. Вариант предсказан «допустимым» в самом длинном транскрипте и «патогенным» в одном альтернативном транскрипте с помощью программы SIFT, тогда как с помощью программы PolyPhen2 данный вариант предсказан как «наиболее вероятно патогенный» во всех затронутых транскриптах. Ген *CSMD1* экспрессируется в различных областях мозга, включая мозжечок [Lindsay et al., 2016; GTEx Consortium, 2013; Miller et al., 2014]. *CSMD1* ассоциирован с шизофренией, аутизмом, когнитивными расстройствами и различными видами рака [Woo et al., 2020; Gialeli et al., 2018]. Данных о врождённых заболеваниях, вызванных мутациями в данном гене, не было на момент исследования.

Обе неродственные семьи АII и АIII произошли из Северо-Кавказского региона. При этом у пациентов из семей АII и АIII был обнаружен гомозиготный участок размером 911,088 т.п.н., в связи с чем можно предположить наличие эффекта основателя для выявленного варианта. Так как частота минорного аллеля может быть повышена в популяциях Северного Кавказа, было проведено сравнение частоты в популяциях из конкретного региона и в мировой популяции. Общий MAF варианта составляет 0,0042 для 2,5 тыс. человек базы данных Проекта 1000 Геномов и 0,005932 для 60 703 человек из ExAC v.1.0. Аналогично MAF данного варианта составляет 0,0056 у 533 человек с Кавказа по данным браузера RuSeq [Barbitoff YA et al., 2021]. Для оценки MAF варианта rs117633452 среди населения Северного Кавказа из географического региона этих семей было проведено генотипирование данного варианта у 42 неродственных индивидов без неврологических нарушений с помощью секвенирования по Сэнгеру. В результате не было выявлено носителей минорного аллеля, в связи с чем MAF в кавказских популяциях не превышает 0,0119. Таким образом, найденный вариант может быть фактором риска редкого заболевания.



**Рисунок 9.** Идентификация редкого варианта в семьях AII и AIII. А – Визуализация области генома, содержащей кандидатный миссенс вариант в гене *CSMD1* (hg19 chr8:g.2824257 C>T, rs117633452) с помощью программы IGV. Б – Валидация кандидатного гомозиготного варианта у пациентов (II-1, II-2, III-1 и III-2) и гомозиготного варианта дикого типа у здоровых неродственных индивидов (IND1-2).

У пациента AIV-1 был найден новый гомозиготный миссенс вариант hg19 chr6:g.146720646 C>G (NM\_000838.2:c.2471C>G) в гене *GRM1*. Новый вариант отсутствует в базах данных 1000 Геномов, gnomAD и dbSNP (Рисунок 10). Мутации в гене *GRM1* вызывают аутосомно доминантную спиноцеребеллярную атаксию 44, SCA44 (OMIM 617691), и аутосомно-рецессивную спиноцеребеллярную атаксию 13, SCAR13 (OMIM 614831). Найденный вариант расположен в высоко консервативной области (PhyloP счёт 7,7842). С помощью программ SIFT и PolyPhen2 новый вариант предсказан как “патогенный” и “наиболее вероятно патогенный”. Вариант подтвержден с помощью секвенирования по Сэнгеру.

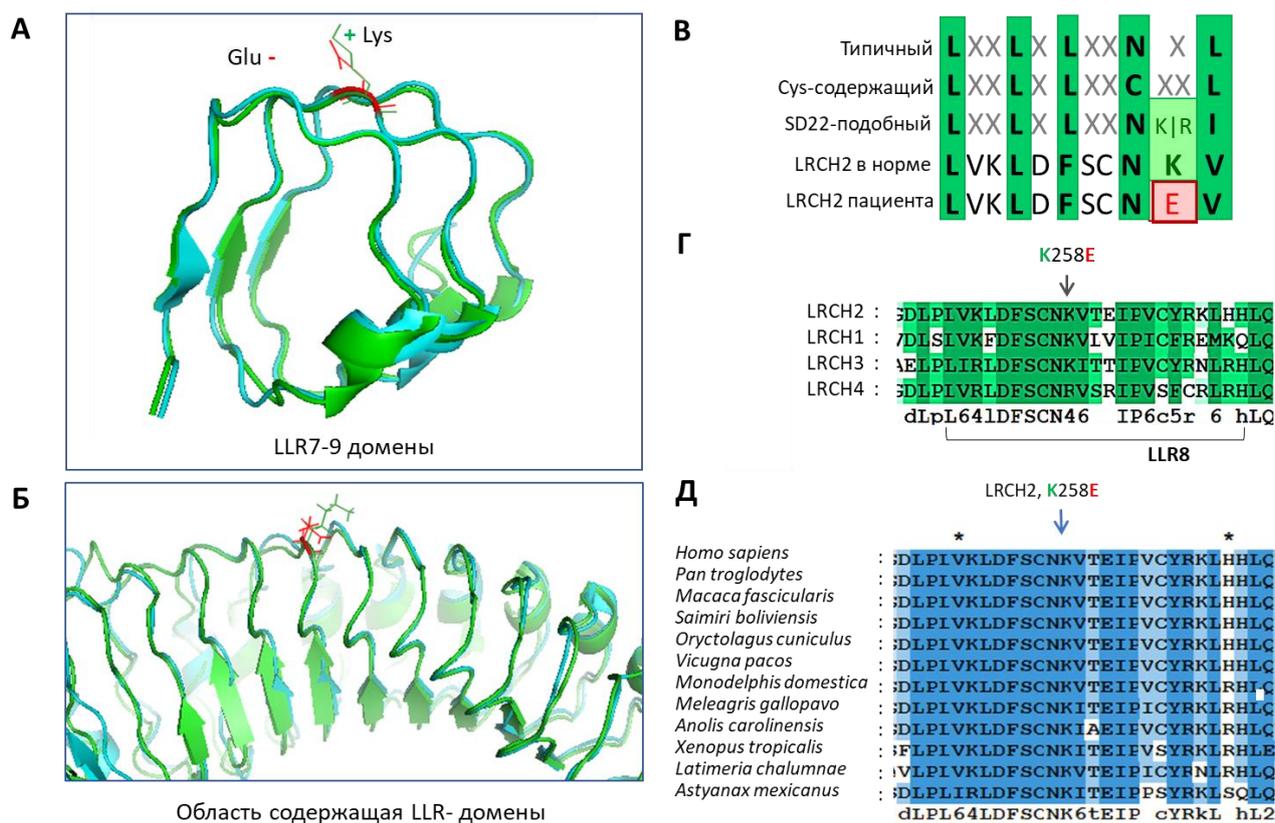


**Рисунок 10.** Идентификация нового варианта у пациента AIV-1. А – Визуализация области генома, содержащей кандидатный миссенс вариант в гене *GRM1* (hg19 chr6:g.146720646 C>G) с помощью программы IGV. Б – Валидация кандидатного гомозиготного варианта у пациента AIV-1.

**Анализ пространственной структуры белков, имеющих мутации в семьях AI-IV.** Было проведено предсказание пространственных структур белка FAАН2 с минорными вариантами: обнаруженным в семье AI p.Lys460Thr и опубликованным ранее p.Ala458Ser [Sirrs et al., 2015] и вариантом дикого типа. Оба минорных варианта расположены в консервативной области амидазного домена. Анализ пространственной структуры с использованием Phyre2 предсказал, что обе замены p.Ala458Ser и p.Lys460Thr уменьшают близлежащую альфа-спираль. Однако замена Ala458 на любую другую, кроме пролина, неблагоприятна, тогда как замена Lys460 менее критична. Сравнение белков, содержащих амидазный домен, показало, что сайт Ala458 более консервативен, чем сайт Lys460. Стоит отметить, что значение *FAАН2* для развития спорно. В одном исследовании нонсенс-вариант p.W392\* был обнаружен у пациента с летальным фенотипом, синдромом Зеллвегера [Shamseldin et al., 2021]. Но в другом исследовании не было выявлено каких-либо клинических проявлений при усечении белка FAАН2 с С-концевой области в результате нонсенс-варианта p.E432\* [Lim et al., 2013].

Замена p.Lys258Glu в белке LRCH2 происходит в высоко консервативной области лейцин-богатого повтора (LRR8). Анализ пространственной структуры

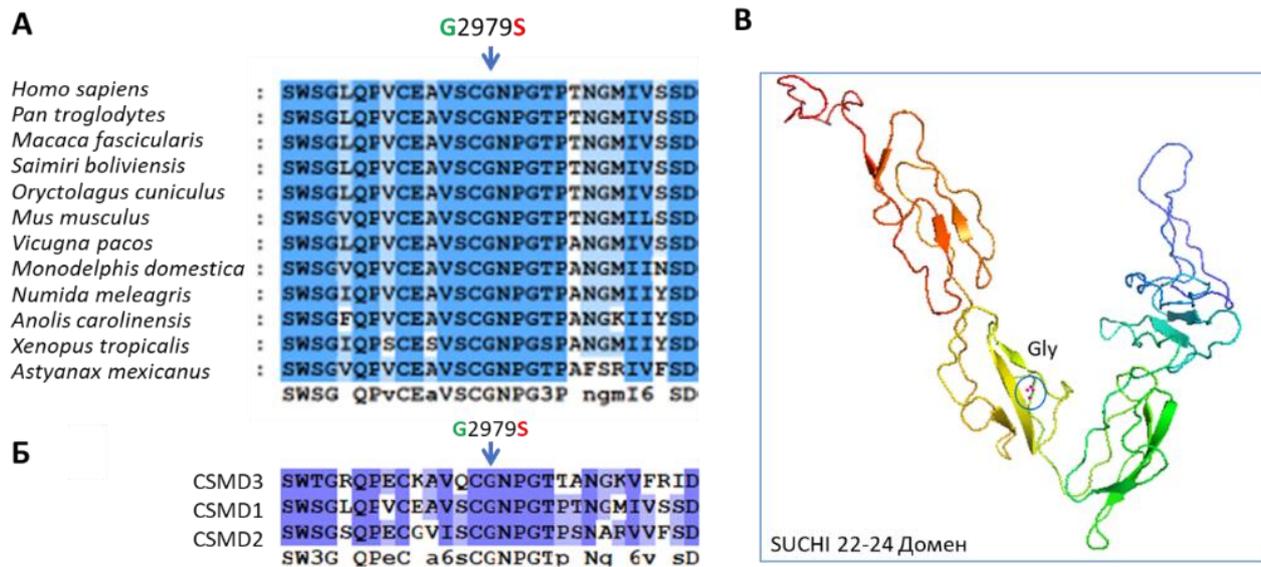
не выявил изменений в белке с мутантным вариантом (Рисунок 11А, Б). Однако эволюционный анализ семейства LRCH показал, что домен LRR8 является консервативным и содержит только аминокислоты с положительно заряженной боковой цепью (лизин или аргинин) в этом положении (Рисунок 11Г). Более того, повтор LRR8 в белке LRCH2 классифицируется как SD22-подобный; при этом типе LRR в исследуемом положении остаются только аминокислоты с положительной боковой цепью (Рисунок 11В) [Kajava, 1998; Rivière et al., 2020; Kobe, Kajava, 2001; Ng et al., 2011]. Найденный вариант приводит к замене на отрицательно заряженную глутаминовую кислоту. Таким образом, эта замена может влиять на аффинность связывания LRCH2 с взаимодействующими с ним белками.



**Рисунок 11.** Анализ влияния замены p.Lys258Glu на структуру белка LRCH2. А, Б – Предсказание пространственной структуры области LRR7-9 (А) и всей области LRR целиком (Б) с помощью программы Phyre2. Белок с вариантом дикого типа выделен зелёным, а с мутантным – голубым. В – Анализ положения замены в LRR домене (где L – Leu, Ile, Val или Phe; N – Asn, Thr, Ser или Cys; C – Cys, Ser или Asn) [Kajava, 1998; Rivière et al., 2020; Kobe, Kajava, 2001; Ng et al., 2011]. Эволюционный анализ консервативности области замены p.Lys258Glu (K258E) среди паралогов (Г), среди ортологов (Д).

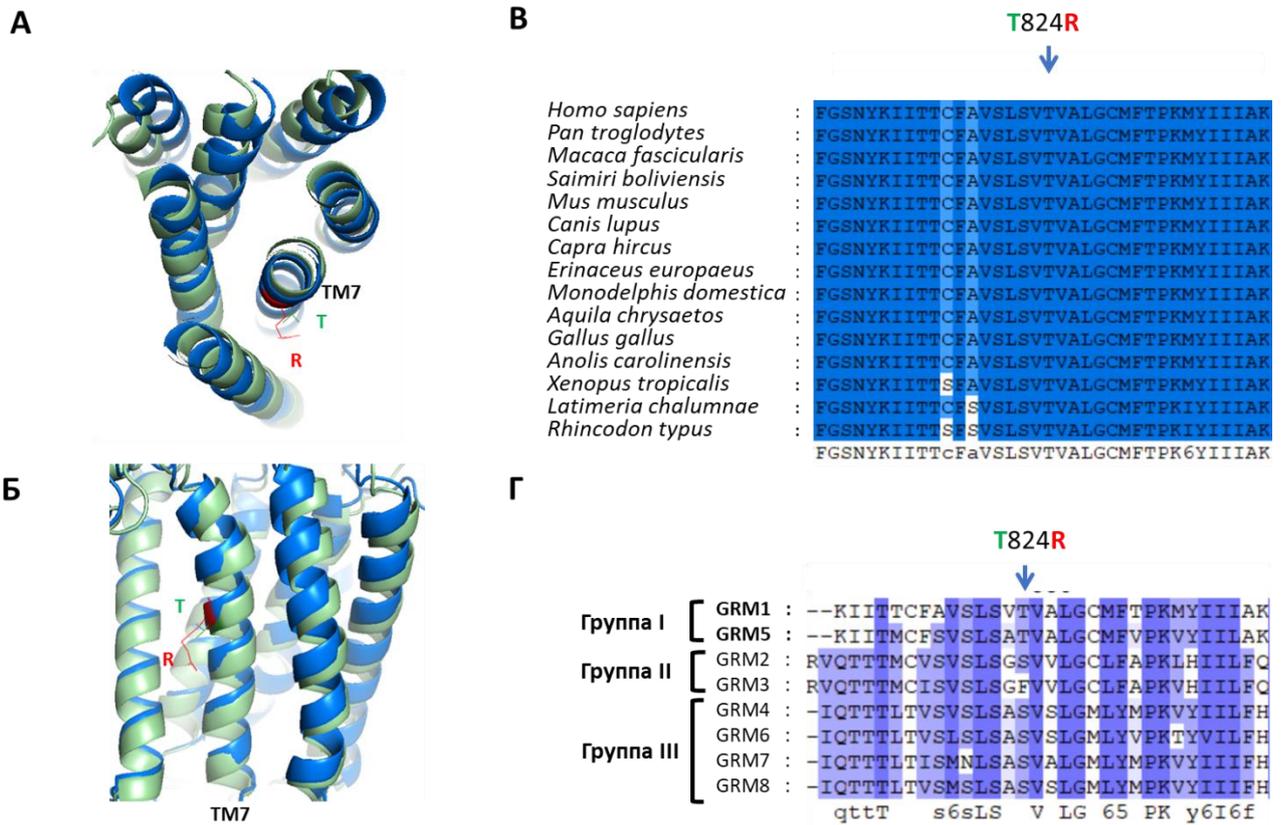
Миссенс вариант rs117633452 приводит к замене p.Gly2979Ser во внеклеточном повторе Such1 23 (комплемент-взаимодействующий домен) в белке CSMD1. Глицин консервативен у всех ортологов и паралогов (Рисунок 12). Анализ пространственной структуры белка не выявил изменений в случае замены. В одном исследовании семейных случаев с поздними формами болезни Паркинсона были выявлены гетерозиготные варианты в домене модуле белка

контроля комплемента (CCP) CSMD1, а у некоторых пациентов заболевание осложнялось дизартрией, атаксией и нарушениями ходьбы [Ruiz-Martínez et al., 2017]. CSMD1 с помощью своего CCP-модуля может ингибировать классический путь комплемента, ингибируя накопление основного компонента комплемента C3 на клеточной поверхности [Kraus et al., 2006; Kraus et al., 2007; Escudero-Esparza et al., 2013]. Более того, удаление CCP в CSMD3 приводит к уменьшению плотности шипиков и разветвления дендритов [Mizukami et al., 2016]. Поскольку вариант обнаружен в высоко консервативной области домена, предназначенного для взаимодействия с другими белками, найденная замена p.Gly2979Ser, скорее всего, влияет на эффективность этого взаимодействия.



**Рисунок 12.** Анализ влияния замены p.Gly2979Ser (G2979S) на структуру белка CSMD1. А, Б – Эволюционный анализ консервативности области белка, содержащей найденную замену среди ортологов (А) и паралогов (Б). В – Предсказание пространственной структуры области Suchi домена дикого типа, в котором найдена аминокислотная замена, с помощью Phyre2.

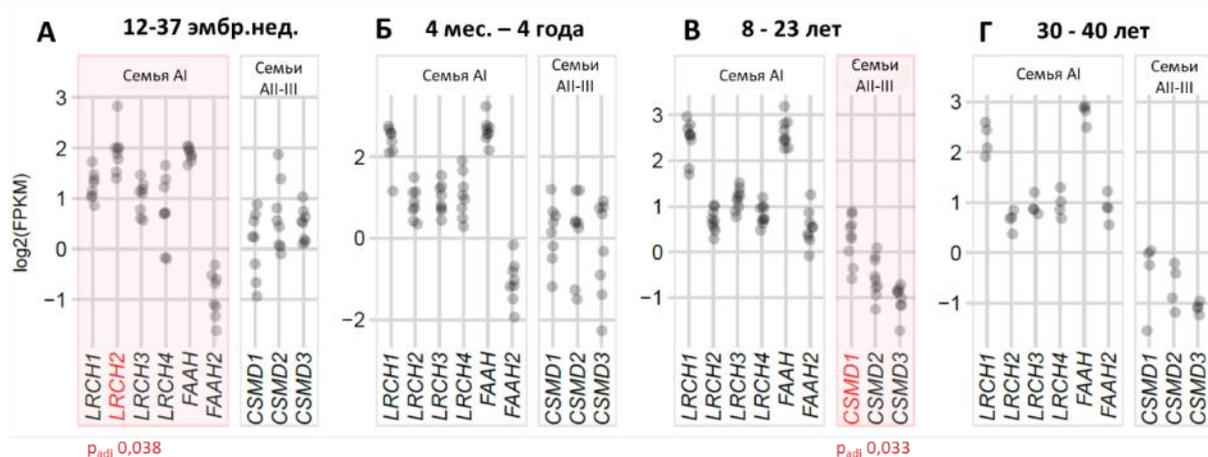
Найденная замена в гене *GRM1* приводит к замене p.Thr824Arg в длинной (альфа) и короткой (бета) изоформах глутаматного метаботропного рецептора 1 (mGluR1). Треонин расположен в большой высоко консервативной области, в трансмембранном домене 7 (TM7), и присутствует у всех ортологов от круглоротых до приматов (Рисунок 13). Треонин также консервативен в обеих изоформах mGluR5, который вместе с mGluR1 составляют группу I глутаматных метаботропных рецепторов [Mao et al., 2022]. Полное удаление TM7 приводит к нарушению структуры mGluR1 и аутосомно-рецессивной форме SCAR13 [Guerguelcheva et al., 2012]. TM7 необходим для присоединения лигандов, которые модулируют передачу сигналов, действуя как аллостерические модуляторы [Wu et al., 2014; Pin et al., 2016; Koehl et al., 2019; Zhang et al., 2021]. Аргинин по сравнению с треонином имеет длинную электрически положительно заряженную цепь, которая может образовывать избыточные водородные или ионные связи и, предположительно, мешать функциональному взаимодействию с лигандом.



**Рисунок 13.** Анализ влияния замены p.Thr824Arg на структуру mGluR1. А, Б – Пространственная структура трансмембранной области, содержащей замену p.Thr824Arg, структура белка предсказана с помощью Phyre2 (зелёным цветом выделен белок с вариантом дикого типа; синим – с мутантным вариантом). В, Г – Эволюционный анализ области mGluR1, содержащей замену p.Thr824Arg (T824R) среди ортологов (В) и паралогов (Г).

**Сравнительный анализ данных экспрессии генов и их паралогов.** Для анализа генной избыточности и определения генов-кандидатов при редких моногенных заболеваниях была разработана онлайн-программа «Основной Паралог по Экспрессии» (EPbE). Принцип EPbE заключается в следующем: программа получает предсказание паралогов из «Ортологической матрицы» (ОМА) [Altenhoff et al., 2021] и данные экспрессии из проекта «Нормальное развитие мозга» (BrainSpan) [Miller et al., 2014] и сравнивает экспрессию с помощью статистического критерия Данна «многие к одному» с корректировкой р значения с помощью одноэтапного метода. Для проверки программы EPbE было выбрано несколько известных генов, связанных с аутосомно-рецессивной формой мозжечковой атаксии, для которых есть данные о избыточности паралогов *SYNE1* и *SNX14* [Zhang et al., 2007; Lauzier et al., 2022], и для которых есть данные о незаменимости их функций *VPS13D* и *SPTBN2* [Baldwin et al., 2021; Fujishima et al., 2020]. Только для генов, функция которых может быть частично компенсирована их паралогами, программа EPbE предсказала статистически значимый высокий уровень экспрессии в коре мозжечка по сравнению с их паралогами. Для паралогов с дивергированными функциями (отрицательный контроль) не было обнаружено статистически значимых различий.

*FAAH2* имеет паралог *FAAH* (или *FAAH1*). Исследования ферментативной активности белков показали, что *FAAH* может гидролизовать тот же субстрат (олеамид или анандамид), в отношении которого активен *FAAH2* [Wei et al., 2015; Sirrs et al. 2015; Cravatt et al., 1996; Saghatelian et al., 2004]. Таким образом, отсутствие *FAAH2* может быть компенсировано его паралогом. Гипоплазия мозжечка связана с повреждением генов, действующих в эмбриональный период, однако в этот период экспрессия *FAAH2* в мозжечке чрезвычайно низка, в отличие от экспрессии *FAAH* (Рисунок 14А). Таким образом, маловероятно, что идентифицированный вариант в гене *FAAH2* вносит существенный вклад в гипоплазию мозжечка, однако, может иметь место минорный эффект.



**Рисунок 14.** Экспрессия генов-кандидатов и их паралогов в коре мозжечка. Значительно повышенная экспрессия *LRCH2* и *CSMD1* по сравнению с их паралогами совпадает с началом заболевания в исследованных семьях (А и В, выделены красным). А – пренатальное развитие, Б – постнатальное развитие в раннем детстве, В – период начиная со школьного возраста до ранней зрелости, Г – зрелый возраст. Для сравнительного анализа были получены данные проекта «Нормальное развитие мозга» (BrainSpan) [Miller et al., 2014]. Padj – максимальное скорректированное значение.

Ген *LRCH2* имеет три паралога, *LRCH1*, *LRCH3* и *LRCH4*, которые также имеют высокий уровень экспрессии в головном мозге. Структуры белков семейства LRCH очень похожи: все белки имеют 9-10 LRR в N-концевой области и СН-домен в С-концевой области [Kajava et al., 1998]. Более того, недавние исследования показывают, что эти гены могут взаимодействовать с белками DOCK6-8, которые играют роль в реорганизации цитоскелета [Müller et al., 2020]. В связи с этим компенсация гена *LRCH2* вполне возможна за счёт его паралогов. В коре мозжечка в пренатальный период наблюдается самый высокий уровень экспрессии гена *LRCH2* по сравнению с его паралогами (значение  $p = 0,038$ , Рисунок 14А). После рождения экспрессия *LRCH2* снижается, а экспрессия его паралогов, особенно *LRCH1*, увеличивается (Рисунок 14Б-Г). Таким образом, в семье АI развитие гипоплазии мозжечка в пренатальный период совпадает с наиболее высокой экспрессией гена *LRCH2*. В коре головного мозга в пренатальном периоде, кроме гена *LRCH2*, высокий уровень экспрессии

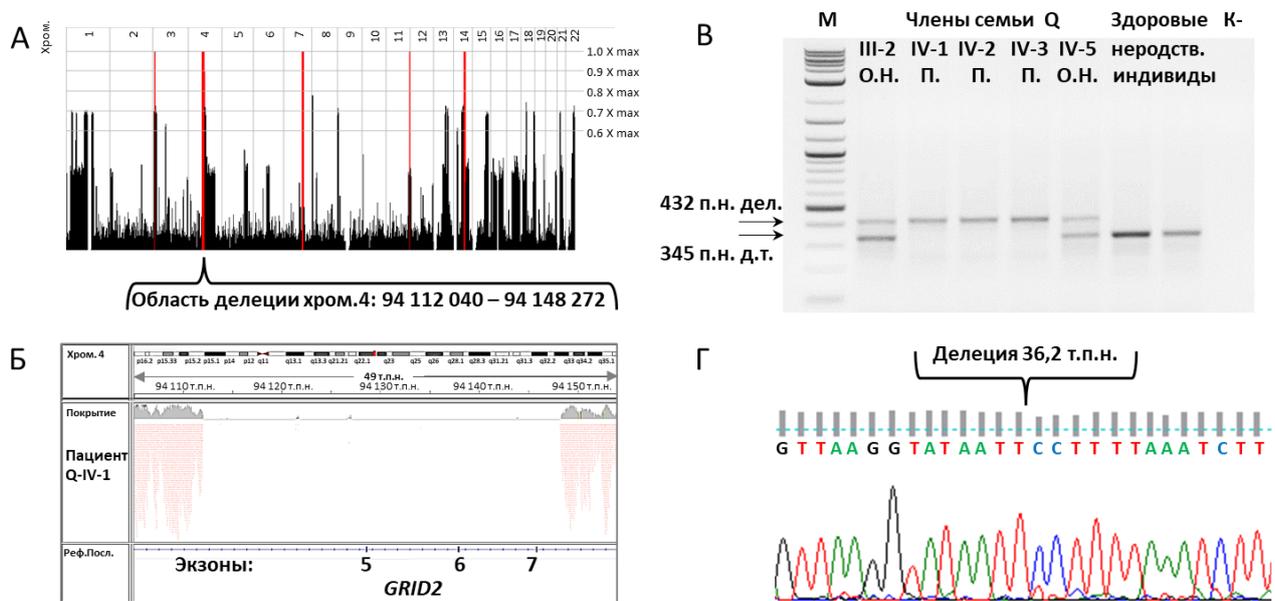
наблюдается и для паралога *LRCH1*, за исключением моторно-сенсорной коры и орбитофронтальной коры. При исследовании нервной проводимости у пациентов из семьи AI были обнаружены признаки демиелинизирующей моторной и сенсорной невропатии. Дифференциальная экспрессия *LRCH2* наблюдается в ганглиях задних корешков по сравнению с седалищным нервом [Sapio et al., 2016]. Кроме того, экспрессия гена *LRCH2* была обнаружена в шванновских клетках, формирующих миелиновую оболочку периферических нервов [Kim et al., 2017]. Полученные данные согласуются с генерализованным демиелинизирующим поражением сенсомоторных волокон у пациентов из семьи AI.

Ген *CSMD1* имеет два паралога, *CSMD2* и *CSMD3*, с высокой степенью сходства белков 61-65%. Все три белка имеют большой внеклеточный домен, содержащий повторы доменов CUB и Sushi, участвующий в связывании комплемента, и один трансмембранный домен с небольшой внутриклеточной C-концевой областью. Точных данных о генной избыточности данного семейства нет. Недавние исследования показали, что, подобно *CSMD1*, его паралоги участвуют в росте проекций нейронов [Kraus et al., 2007; Mizukami et al., 2016; Gutierrez et al., 2019]. С помощью EPbE было установлено, что разница в экспрессии генов заметна в период полового созревания и раннего взросления (11-23 года); в этот период самый высокий уровень экспрессии в коре мозжечка присутствует у гена *CSMD1*, а экспрессия его паралогов снижена (значение  $p = 0,033$ , Рисунок 14B). Этот период совпадает с появлением клинических симптомов в семьях AII-III. Эта разница снова исчезает с возрастом, и в более позднем возрасте все гены экспрессируются примерно на одном уровне. Экспрессия гена *CSMD1* в пренатальный период в коре головного мозга также не превышает экспрессию его паралогов. Затем экспрессия гена *CSMD1* увеличивается с возрастом (от 8 до 40 лет), в то время как экспрессия паралогов снижается или остаётся на том же уровне.

Интересно, по данным DropViz [Saunders et al., 2018], оба гена *CSMD1* и *LRCH2* экспрессируются в интернейронах мозжечка у взрослых мышей, что также согласуется с выявленной патологией развития мозжечка у пациентов. Более того, семейства данных генов участвуют в одних и тех же биологических процессах: врождённый иммунитет, формирование цитоскелета и миграцию клеток [Gialeli et al., 2018; Kajava, 1998]. Таким образом, оба гена-кандидата вовлечены в формирование моторных функций в интернейронах мозжечка, и, вероятно, вовлечены в общие биологические пути на разных этапах онтогенеза.

**Генетический анализ в семье Q с синдромом CAMRQ.** Чтобы определить локус с генетическим дефектом в семье Q, был проведён анализ областей гомозиготности с помощью Affymetrix Genome-Wide Human SNP Array 6.0. В результате было выявлено пять регионов с максимальной оценкой гомозиготности 3000 (Рисунок 15A). Затем для одного пациента Q-IV-1 был проведён анализ полногеномного секвенирования. Всего было выявлено 5,2 млн вариантов, из которых в кодирующей последовательности генов 3,78 тыс. были

редкими (MAF < 0,05 [1000 Genomes Project Consortium et al., 2012]). В области с максимальной гомозиготностью были отобраны три варианта, из которых два были исключены из дальнейшего анализа на основе критериев отбора. Затем было найдено ещё 3 варианта в малых гомозиготных регионах. Кроме того, вне областей гомозиготности, был обнаружен гомозиготный вариант rs140526335 (MAF 0,0005) в гене *VLDLR*, связанный с синдромом CAMRQ1. Генетический анализ вариантов с помощью ПЦР и секвенирования по Сэнгеру у членов семьи показал, что только для вариантов, расположенных в области максимальной гомозиготности, наследование соответствует статусу заболевания и аутосомно-рецессивному типу.



**Рисунок 15.** Идентификация делеции в гене *GRID2* в семье Q. А – локусы гомозиготности у пациентов Q-IV-1, Q-IV-2 и Q-IV-3 (отмечены красным). Б – область делеции в гене *GRID2*, в данных полногеномного секвенирования пациента Q-IV-1 визуализированная с помощью программы IGV. В – электрофореграмма продуктов ПЦР, подтверждающих делецию в гене *GRID2*: М – ДНК маркер; продукт размером 432 п.н., подтверждает наличие делеции в гене *GRID2*, выявлен в гомозиготном состоянии только у пациентов; продукт размером 345 п.н., подтверждающий отсутствие делеции, в гомозиготном состоянии, выявлен у здоровых неродственных индивидов. Оба продукта ПЦР выявлены у облигатных носителей: у матери пациентов – Q-III-2, сводного брата пациентов – Q-IV-5. Г – электрофореграмма секвенирования по Сэнгеру продукта, подтверждающего делецию. П. – пациент; О.Н. – облигатный носитель; дел. – делеция; д.т. – дикий тип.

В гене *GRID2* в области 4q22 хромосомы с максимальной гомозиготностью была отобрана крупная делеция hg19 chr4:g.94112040-94148272del (NM\_001510.4:c.736-16515\_1125+2346del) (Рисунок 15). Ген *GRID2*, кодирующий глутаматный рецептор дельта 2, имеет высокую экспрессию в клетках Пуркинью мозжечка в пре- и постнатальный период. Нарушения в *GRID2* связаны с аутосомно-рецессивной мозжечковой атаксией SCAR18, клиническая картина которой имеет сходство с исследуемым фенотипом, однако отличиям

является квадроцепь и полное отсутствие речи у пациентов в семье Q [Maier et al., 2014; Utine et al., 2013; Hills et al., 2013; Van Schil et al., 2015; Coutelier et al., 2015; Veerapandiyam et al., 2017; Taghdiri et al., 2019; Ceylan et al., 2020; Hetzelt et al., 2020]. Делеция охватывает с 5 по 7 экзоны самой длинной изоформы и влияет на все остальные изоформы, в результате которой происходит удаление области размером 130 аминокислот (V246–K375) во внеклеточном аминоконцевом домене (ATD, ранее LIVBP) рецептора GRID2. Данная область высоко консервативна у позвоночных. Домен ATD в рецепторе GRID2 необходим для образования комплекса с CBLN1 и  $\beta$ -NRXN1 [Elegheert et al., 2016], играющего важную роль в образовании синапсов между параллельными волокнами и клетками Пуркинье и передачи сигналов при долговременной депрессии. Потеря ATD, возможно, нарушила образование комплекса. Также есть данные, что крупные делеции в гене *GRID2* нарушают формирование третичной структуры и транспорт белка на поверхность клетки [Matsuda et al., 2002].

**Эволюционный анализ.** Кодирующая последовательность гена *GRID2* высоко консервативна среди всех позвоночных и достигает наибольшей консервативности (93–99%) у приматов, анализ с помощью ветвь и ветвь-сайт специфичных моделей, реализованных в программе CodeML пакета PAML не выявил статистически значимых изменений под действием направленной селекции. Значение  $\omega$  в ветвь-специфичной модели M0 было равным 0,027, указывая на сильную очищающую тенденцию. Несколько сайтов были определены с низкой байесовской вероятностью ( $0,5 < \text{BEV} < 0,8$ ), из них один N153H был определён в ветви человекообразных обезьян (Hominoidea). Так как белок-кодирующая последовательность гена *GRID2* современного человека и его ближайших родственных видов (*Pan troglodytes* и *Pan paniscus*) различается в основном синонимичными заменами, было проведено исследование влияния этих замен на стабильность мРНК с помощью программы mfold. В белок-кодирующей последовательности гена *GRID2* современного человека были выявлены два варианта G<sup>94436426</sup> и G<sup>94436547</sup> со значительно более низким уровнем свободной энергии Гиббса и три варианта A<sup>94159638</sup>, T<sup>94316763</sup> и G<sup>94376878</sup> с несколько более высоким уровнем свободной энергии Гиббса по сравнению с шимпанзе обыкновенным. Анализ свободной энергии Гиббса в более крупных таксонах приматов выявил три замены со статистически значимым снижением свободной энергии Гиббса. Две из них являются синонимичными заменами у Homini и одна несинонимичная замена N153H у Hominoidea. Таким образом, на основе проведённого биоинформатического анализа можно предположить, что эволюционный отбор вариантов в ветвях таксона Hominoidea для гена *GRID2* направлен на формирование оптимальной пространственной структуры РНК.

## ВЫВОДЫ

1. Была выявлена мутация в гене *ABCB7*, которая привела к X-сцепленной рецессивной гипоплазии мозжечка у пациентов из большой бурятской семьи. Данная мутация наследуется в группе сцепления с крупной делецией в гене *ATP7A*, приводящей к удалению второго экзона и ретрогена *PGAM4*. В результате делеции транспортёр ATP7A утрачивает первый из шести медь-связывающих доменов, сохраняя основную функцию. Отсутствие ретрогена *PGAM4*, экспрессируемого исключительно в тестикулах, не нарушает мужскую фертильность.

2. Анализ в четырёх малых семьях выявил генетическую гетерогенность клинически близких редких случаев мозжечковой атаксии. В одной из семей выявлена новая мутация в гене *GRM1*, ставшая причиной врождённой мозжечковой атаксии.

3. Разработан подход анализа данных массового параллельного секвенирования для выявления новых генетических факторов редких заболеваний в спорадических случаях и случаях в малых семьях с неопределённой этиологией, в основе которого лежит сравнительный анализ функций и экспрессии *in silico* кандидатных генов и их паралога на разных этапах онтогенеза.

4. С помощью разработанного подхода выявлены новые гены *LRCH2* и *CSMD1*, в которых миссенс варианты привели к гипоплазии мозжечка в случаях с редкой аутосомно- и X-сцепленной рецессивной формой спинномозжечковой атаксии.

5. Обнаружена делеция в гене *GRID2*, являющаяся причиной редкого синдрома мозжечковой атаксии, умственной отсталости, нарушения речи и квадропедии в бразильской семье. Эволюционный анализ гена *GRID2* у приматов выявил нуклеотидные замены, специфичные для линии *Homo*, которые влияют на стабильность мРНК.

## СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

### Публикации в журналах, соответствующих Перечню ВАК

1. **Protasova M.S.**, Gusev F.E., Andreeva T.V., Klyushnikov S.A., Illarioshkin S.N., Rogaev E.I. Novel genes bearing mutations in rare cases of early-onset ataxia with cerebellar hypoplasia. // *Eur J Hum Genet*. 2022. № 30. С. 703–711.
2. Grigorenko A.P., **Protasova M.S.**, Lisenkova A.A., Reshetov D.A., Andreeva T.V., Garcias G.L., Martino Roth M.D.G., Papassotiropoulos A., Rogaev E.I. Neurodevelopmental Syndrome with Intellectual Disability, Speech Impairment, and Quadrapedia Is Associated with Glutamate Receptor Delta 2 Gene Defect. // *Cells*. 2022. Т. 11. № 3. С. 400.
3. **Protasova M.**, Grigorenko A., Tyazhelova T., Andreeva T., Reshetov D., Gusev F., Laptenko A., Kuznetsova I., Goltsov A., Klyushnikov S., Illarioshkin S., Rogaev E. Whole Genome Sequencing Identifies a novel ABCB7 gene mutation for X linked congenital cerebellar ataxia in a large family of Mongolian ancestry. // *Eur J Hum Genet*. 2016. Т. 4. С.550-555.

### Тезисы докладов в материалах конференций

1. Grigorenko A.P., **Protasova M.S.**, Andreeva T.V., Reshetov D.A., Lisenkova A.A., Manakhov A.D., Rogaev E.I. Identification of the genetic cause for quadrapedal locomotion in human subjects by high-throughput sequencing. // 2nd International scientific conference “Science of the Future”. 2016, С. 279-280, *Устный доклад*
2. **Protasova M.S.**, Grigorenko A.P., Tyazhelova T.V., Andreeva T.V., Reshetov D.A., Gusev F.E., Kuznetsova I.L., Goltsov A.Y., Klyushnikov S.A., Illarioshkin S.N., Rogaev E.I. The identification of the genetic cause of the cerebellar disorder. // *Frontiers in Biomedicine*. 2015. №1. С.19. *Устный доклад*
3. **Протасова М.С.**, Григоренко А.П., Тяжелова Т.В., Гусев Ф.Е., Андреева Т.В., Решетов Д.А., Кузнецова И.Л., Ключников С.А., Иллариошкин С.Н., Робаев Е.И. Идентификация генетической причины X-сцепленной не прогрессирующей наследственной мозжечковой атаксии с использованием полногеномного секвенирования. // VII съезд Российского общества медицинских генетиков. Журнал "Медицинская Генетика" 2015. №3. С. 84. *Стендовый доклад*.
4. **Protasova M.S.**, Grigorenko A.P., Tyazhelova T.V., Gusev F.E., Andreeva T.V., Reshetov D.A., Kuznetsova I.L., Klyushnikov S.N., Illarioshkin S.N., Rogaev E.I. Complete Genome Re-Sequencing Identifies the Genetic Cause For X-linked Congenital Cerebral Ataxia. // *Science of the Future*, 2014. *Стендовый доклад*