

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НОВОСИБИРСКИЙ НАЦИОНАЛЬНЫЙ  
ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ» (НОВОСИБИРСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ, НГУ)

Институт медицины и психологии В. Зельмана НГУ

Направление подготовки: 31.05.01 Лечебное дело

## НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА

Валеев Эмиль Салаватович

Направление подготовки: 31.05.01 Лечебное дело

Направленность (профиль): Лечебное дело

Форма обучения: очная

Группа 12451

Тема работы: «Разработка инструментов для поиска клинически значимых  
полиморфизмов в геноме человека на основе данных секвенирования  
ЗС-библиотек»

### Научный руководитель:

Фишман Вениамин Семенович,  
к.б.н., ведущий научный сотрудник,  
заведующий Сектором геномных  
механизмов онтогенеза, ИЦиГ СО РАН

ФИО: \_\_\_\_\_ / \_\_\_\_\_

«\_\_\_\_\_» \_\_\_\_\_ 20\_\_\_\_ г.

Оценка: \_\_\_\_\_

# Содержание

|  |           |
|--|-----------|
| <b>Список сокращений</b>   | <b>3</b>  |
| <b>Введение</b>  | <b>6</b>  |
| <b>1 Обзор литературы</b>  | <b>8</b>  |
| 1.1 Механизмы развития генетических патологий . . . . .  | 11        |
| 1.2 Типы генетических аномалий, лежащих в основе генетических патологий . . . . .  | 13        |
| 1.3 Функциональные классы генетических вариантов . . . . .   | 15        |
| 1.4 Методы детекции генетических вариантов . . . . .   | 16        |
| 1.5 Виды NGS . . . . .   | 22        |
| 1.6 Базовая схема обработки результатов высокопроизводительного секвенирования для поиска и клинической интерпретации однонуклеотидных полиморфизмов . . . . . | 25        |
| 1.7 Аннотация, фильтрация и интерпретация результатов . . . . .  | 32        |
| 1.8 Ехо-С: суть метода . . . . .   | 37        |
| <b>2 Материалы и методы</b>  | <b>38</b> |
| <b>3 Результаты</b>  | <b>46</b> |
| 3.1 Результаты секвенирования Ехо-С-библиотек . . . . .  | 47        |
| 3.2 Автоматизация обработки данных секвенирования . . . . .  | 48        |
| 3.3 Сравнение данных секвенирования клеточной линии K562 . .   | 50        |
| 3.4 Клинические случаи . . . . .   | 56        |
| <b>4 Обсуждение результатов</b>  | <b>56</b> |
| 4.1 Контрольные образцы . . . . .  | 58        |
| 4.2 Оценка результатов секвенирования Ехо-С-библиотек . . . . .  | 58        |
| 4.3 Клинические случаи . . . . .   | 60        |
| <b>Выводы</b>  | <b>60</b> |

|   |            |
|---|------------|
| <b>Список литературы</b>  | <b>62</b>  |
| <b>A Технические данные по секвенированию клеточной линии K562</b>                      | <b>79</b>  |
| <b>B Результаты фильтрации генетических вариантов по глубине альтернативного аллеля</b> | <b>79</b>  |
| <b>C Клиническое описание пациентов</b>   | <b>82</b>  |
| <b>D Медицинские учреждения, предоставившие пациентов</b>                               | <b>175</b> |

# Список сокращений

**3C** (англ. *Chromosome Conformation Capture*) — захват конформации хромосом

**BAM** (англ. *Binary sequence Alignment/Map*) — бинарный файловый формат, предназначенный для хранения информации о картированных прочтениях

**bp** (англ. *base pairs*) — пары оснований

**BQSR** (англ. *Base Quality Score Recalibration*) — рекалибровка качества прочтений

**cffDNA** (англ. *Cell-Free Fetal DNA*) — свободная ДНК плода

**CGH** (англ. *Comparative Genomic Hybridization*) — сравнительная геномная гибридизация

**CNV** (англ. *Copy Number Variation*) — вариация числа копий

**Exo-C** — метод приготовления NGS-библиотек, сочетающий таргетное обогащение экзона и технологии захвата конформации хромосом

**FISH** (англ. *Fluorescence In Situ Hybridization*) — флуоресцентная *in situ* гибридизация

**GATK** (англ. *Genome Analysis ToolKit*) — набор инструментов для биоинформационного анализа, созданный Broad Institute

**Hi-C** — метод захвата конформации хромосом «все против всех»

**kbp** (англ. *kilo base pairs*) — тысячи пар оснований

**LoF** (англ. *Loss of Function*) — потеря функции гена

**MAPQ** (англ. *MAPping Quality*) — качество картирования

**Mbp** (англ. *mega base pairs*) — миллионы пар оснований

**MIP** (англ. *Molecularly Imprinted Polymers*) — молекулярно импринтированные полимеры

**MLPA** (англ. *Multiplex Ligation-dependent Probe Amplification*) — мультиплексная лигаза-зависимая амплификация зонда

**NGS** (англ. *New Generation Sequencing*) — секвенирование нового поколения

**NIPT** (англ. *Non-Invasive Prenatal Testing*) — неинвазивное пренатальное тестирование

**NOR** (англ. *Nucleolus Organizer Region*) — ядрышковый организатор

**ПЕС** (англ. *Primer Extension Capture*) — захват с помощью расширения праймера

**RG** (англ. *Read Group*) — группа прочтения

**SKY** (англ. *Spectral Karyotyping*) — спектральное кариотипирование

**SMART** — анализ транскриптома одной клетки

**SNV** (англ. *Single Nucleotide Variant*) — однонуклеотидный генетический вариант

**UTR** (англ. *UnTranslated Regions*) — нетранслируемая область

**VCF** (англ. *Variant Call Format*) — формат записи генетических вариантов, найденных в результатах секвенирования

**WES** (англ. *Whole Exome Sequencing*) — полноэкзомное секвенирование

**WGS** (англ. *Whole Genome Sequencing*) — полногеномное секвенирование

**БД** — база данных

**ВИЧ** — вирус иммунодефицита человека

**ДНК** — дезоксирибонуклеиновая кислота

**мРНК** — матричная РНК

**ПЦР** — полимеразная цепная реакция

**РНК** — рибонуклеиновая кислота

**ТАД** — топологически ассоциированные домены

**ХМА** — хромосомный микроматричный анализ

# Введение

## Актуальность

Наследственные заболевания являются одной из основных причин младенческой и детской смертности в развитых странах. Взрослые люди с такими патологиями требуют огромных затрат средств на медикаменты, оперативные вмешательства, специальный уход и социальные льготы. Таким образом, доступные и точные методы диагностики наследственных заболеваний могут помочь в сокращении заболеваемости и смертности, а также повысить экономическое благополучие населения.

Несмотря на то, что в развитии наследственных заболеваний играют роль множество механизмов, в основе их всегда лежат изменения тех или иных участков ДНК. Эти генетические варианты существенно различаются по размеру, характеру изменения, а также функциональному значению. Существует множество методов выявления генетических вариантов, каждый метод имеет свои преимущества и границы применения.

Наиболее перспективными в диагностическом и исследовательском плане в настоящее время являются методы секвенирования — например, полногеномное и полноэкзомное секвенирование. В Секторе геномных механизмов онтогенеза ИЦиГ СО РАН был разработан новейший метод секвенирования — Ехо-С, сочетающий технологии экзомного обогащения с захватом конформации хромосом. Потенциальным преимуществом данного метода может быть возможность поиска как крупных перестроек, так и точечных полиморфизмов в экзоне при относительно небольшой глубине секвенирования, от которой напрямую зависит цена секвенирования. Широкий спектр применения метода и доступность в финансовом аспекте делают метод Ехо-С привлекательным как для медико-биологических научных исследований, так и для внедрения в клиническую практику.

## Цель

Целью нашей работы является сравнение эффективности методов Ехо-С, полногеномного секвенирования и экзомного секвенирования для поиска точечных полиморфизмов в геномах клеток человека.

## Задачи

Основные задачи, которые необходимо решить для достижения поставленной нами цели:

1. Разработать биоинформационный протокол анализа данных секвенирования Ехо-С-библиотек.
2. Проанализировать доступные данные полногеномного, полноэкзомного, Hi-C и Ехо-С-секвенирования для иммортализованной клеточной линии человека K562.
3. Сравнить точечные генетические варианты в геноме клеток K562, детектируемые при использовании полногеномного и экзомного секвенирования, с таковыми, найденными методом Ехо-С.
4. Применить разработанный нами метод для поиска клинически значимых полиморфизмов в геномах пациентов с наследственными заболеваниями.

## Авторский вклад

1. Фишман В. С.<sup>1</sup>: разработка метода Ехо-С; разработка архитектуры конвейера биоинформационных утилит; разработка идеи и дизайна текущего исследования; постановка задач текущего исследования; контроль результатов, полученных в текущем исследовании; установление связей с медицинскими учреждениями, предоставившими пациентов.

---

<sup>1</sup>Институт Цитологии и Генетики СО РАН, Новосибирский государственный университет. Email: [minja@bionet.nsc.ru](mailto:minja@bionet.nsc.ru).



2. Гридина М. М.<sup>2</sup>: разработка метода Echo-C; приготовление библиотек для секвенирования из клеток K562 и образцов пациентов.
3. Валеев Э. С.<sup>3</sup>: разработка архитектуры конвейера биоинформационных утилит; программная реализация, отладка и тестирование конвейера биоинформационных утилит; биоинформационная обработка данных секвенирования; статистическая обработка и визуализация результатов; интерпретация данных по пациентам в контексте медицинской генетики; поиск клинически значимых генетических вариантов, соответствующих искомому фенотипу; написание текста данной исследовательской работы, создание иллюстраций и таблиц.

Данные по пациентам (выписки из историй болезни, актуальная информация по лабораторным исследованиям) были предоставлены врачами-генетиками: Назаренко Л. П.<sup>4</sup>, Кузьменко Н. Б.<sup>5</sup>, Яблонская М. И.<sup>6</sup>, Курочкина Ю. Д.<sup>7</sup>, Шилова Н. В.<sup>8</sup>, Лукьянова Т. В.<sup>9</sup>.

---

<sup>2</sup>Институт Цитологии и Генетики СО РАН. Email: [gridina@bionet.nsc.ru](mailto:gridina@bionet.nsc.ru).

<sup>3</sup>Институт Цитологии и Генетики СО РАН, Новосибирский государственный университет. Email: [emil@bionet.nsc.ru](mailto:emil@bionet.nsc.ru).

<sup>4</sup>Томский НИМЦ. Email: [ludmila.nazarenko@medgenetics.ru](mailto:ludmila.nazarenko@medgenetics.ru).

<sup>5</sup>НМИЦ им. Рогачева. Email: [plunge@list.ru](mailto:plunge@list.ru).

<sup>6</sup>НИКИ Педиатрии.

<sup>7</sup>НИИКЭЛ. Email: [juli\\_k@bk.ru](mailto:juli_k@bk.ru).

<sup>8</sup>МГНЦ. Email: [nvsh05@mail.ru](mailto:nvsh05@mail.ru).

<sup>9</sup>КЦОЗСиР. Email: [luktat@mail.ru](mailto:luktat@mail.ru).

# Глава 1. Обзор литературы

Генетические варианты, их взаимодействие друг с другом и со средой определяет течение болезней. Существуют генетические варианты, которые определяют предрасположенность и проявляются только во взаимодействии со средой; примером могут служить варианты, определяющие предрасположенность к аддикциям (никотин, героин, алкоголь и пр.) [[Hiroi and Agatsuma, 2004](#)]. Бывают и такие генетические варианты, которые повышают восприимчивость к одному фактору среды и повышают устойчивость к другому, либо дают позитивный эффект в сочетании и негативный по отдельности. Примером может служить бета-талассемия [[Galanello and Origa, 2010](#)]. Особняком стоят те варианты, которые вне зависимости от средового компонента и генетического окружения приводят к развитию заболевания (например, нейрофиброматоз I типа, который наследуется по аутосомно-доминантному типу и имеет 100 % пенетрантность — [Jett and Friedman, 2009](#)).

Генетические заболевания остаются одной из основных причин младенческой и детской смертности в развитых странах. Врождённые аномалии являются причиной около 20 % смертности до 1 года, а также порядка 10 % в возрасте 1–4 года и 6 % в возрасте 5–9 лет. Злокачественные новообразования являются причиной смерти в 8 % случаев в возрасте 1–4 лет, и 15 % случаев в возрасте 5–9 лет. Порядка 3 % от смертности в возрасте 1–9 лет связаны с сердечными патологиями [[Field and Behrman, 2003](#)]. Взрослые люди с генетическими патологиями требуют огромных затрат средств — на радикальные и паллиативные операции, медикаментозную поддержку (иногда пожизненную), создание условий, учреждений и обучение персонала для обеспечения специализированного ухода.

Таким образом, доступные и точные методы диагностики генетических заболеваний могут помочь в сокращении заболеваемости и смертности, а также повысить экономическое благополучие населения.

**Частые и редкие (орфанные) патологии.** Генетические патологии делятся на группы по частоте встречаемости в популяции. Выделяют частые и

редкие (орфанные) заболевания. Определения орфанных заболеваний могут различаться — например, в США, согласно “Health Promotion and Disease Prevention Amendments of 1984”, редкими считаются патологии, поражающие менее 200 тыс. населения страны (примерно 1 : 1630 при текущей численности населения в 326 млн человек) [Herder, 2017]. Европейское Медицинское Агентство определяет границу как 1 : 2000. Систематический анализ показал, что существует более 290 определений, и среднее значение находится в интервале 40–50 на 100 тыс. населения [Richter et al., 2015].

Также сложность в определении орфанных заболеваний представляет неравномерность их распространённости в тех или иных регионах. Некоторые заболевания могут быть орфанными в одной популяции и частыми в другой (эффект основателя, а также сверхдоминирование). Частным случаем эффекта основателя является атаксия Каймановых островов, связанная с гипоплазией мозжечка и сопутствующими неврологическими проявлениями (задержка развития, дизартрия, нистагм, интенционное дрожание). Это аутосомно-рецессивное заболевание распространено исключительно в одном регионе — Большой Кайманов остров, гетерозиготные носители составляют около 18 % местного населения [Bomar et al., 2003]. Примером сверхдоминирования может служить бета-талассемия — заболевание, связанное с нарушением структуры гемоглобина. Несмотря на то, что у эритроцитов носителей в значительной степени снижена способность переносить кислород, дефектный гемоглобин представляет сложность для развития малярийного плазмодия и таким образом повышает устойчивость носителя бета-талассемии к малярии [Galanello and Origa, 2010]. Соответственно, бета-талассемия распространена в эпидемически опасных по малярии регионах — Средиземноморье и Юго-Восточная Азия, наибольшая частота встречаемости наблюдается на Кипре (14 %) и Сардинии (10,3 %) при средней частоте по земному шару в 1,5 %.

Несмотря на то, что каждое из орфанных заболеваний само по себе встречается редко, в сумме они поражают значительный процент населения (предположительно 5–8 % европейской популяции). Общее число орфанных болезней неизвестно по причине недостатков стандартизации, наиболее частая оценка — 5–8 тыс. [The Lancet Neurology, 2011]. Существуют различные базы

данных, собирающие информацию по орфанным заболеваниям, наиболее известными и часто используемыми из них являются:

1. Global Genes;
2. Online Mendelian Inheritance in Man (OMIM®) [[Amberger et al., 2014](#)];
3. Orphanet [[French National Institute of Health and Medical Research \(INSERM\), 1997](#)].

Около 80 % редких болезней имеют генетическую природу и начинаются в раннем детстве [[The Lancet Neurology, 2011](#)]. Таким образом, ключевым моментом для изучения данных заболеваний является понимание механизмов, лежащих в основе их развития. Количество орфанных заболеваний делает эту задачу крайне непростой. Тем не менее, многие механизмы на сегодняшний момент достаточно хорошо изучены. О них речь пойдёт далее.

## **1.1. Механизмы развития генетических патологий**

Механизмы развития генетических патологий делятся на две большие группы. В первую относят изменения белок-кодирующей последовательности гена, приводящие к прекращению синтеза белка либо к синтезу изменённого полипептида. Ко второй группе относятся эпигенетические механизмы, не затрагивающие непосредственно белок-кодирующие последовательности генов.

Изменения белок-кодирующей последовательности гена (экзонов и сплайс-сайтов) могут приводить к замене аминокислот, сдвигам рамки считывания, появлению преждевременных стоп-кодонов и нарушениям сплайсинга. Прекращение синтеза белка снижает дозу гена, а изменённый полипептид способен как потерять свою функцию, снизив таким образом дозу гена, так и приобрести новые свойства (токсичность). Классическим примером приобретения белком токсичности является известное наследственное

нейродегенеративное заболевание — аутосомно-доминантный вариант болезни Альцгеймера. Другое нейродегенеративное заболевание — аутосомно-рецессивная болезнь Паркинсона — может служить примером потери белком протективной функции [Winklhofer et al., 2008].

Также генетические патологии могут развиваться из-за эпигенетических механизмов, приводящих к изменению экспрессии генов. К таким механизмам можно отнести метилирование ДНК — изменение молекулы ДНК без изменения нуклеотидной последовательности, а также ацетилирование гистонов [Handy et al., 2011].

В частности, нарушение метилирования ДНК ответственно за развитие синдрома Беквита—Видемана. Экспрессия генов *CDKN1C* и *IGF2* регулируется в зависимости от того, на материнской или отцовской хромосоме они находятся (явление геномного импринтинга). Потеря импринтинга, вызванная изменениями регуляторного района, ведёт к изменению экспрессии этих генов и, как следствие, к тяжёлым порокам развития, включающим висцеромегалию, висцеральные грыжи, эмбриональные опухоли, пороки сердца и почек [Jin and Liu, 2018]. Изменение ацетилирования гистонов некоторых генов в клетках головного мозга связано с развитием такого заболевания, как шизофрения [Tang et al., 2011].

Кроме того, на экспрессию генов в значительной степени влияет трёхмерная структура хроматина. К примеру, энхансерный район не обязательно находится в непосредственной близости от гена, для его работы необходим физический контакт с промотором гена за счёт выпетливания ДНК. Белковый комплекс, связанный с энхансером, привлекает в эту область РНК-полимеразу и увеличивает вероятность её связывания с промотором. Известно, что большая часть промотор-энхансерных взаимодействий находится внутри топологически ассоциированных доменов (ТАДов) [Rao et al., 2014]. В результате разрушения старых или образования новых границ ТАДов формируются структурные варианты, характеризующиеся иными промотор-энхансерными взаимодействиями. Подобные изменения лежат в основе таких состояний, как FtM-инверсия пола (ген *SOX9*) и синдром Кукса (ген *KCNJ2*) [Spielmann et al., 2018].

Несмотря на то, что в развитии наследственных заболеваний эпигене-

тика безусловно играет важную роль, в основе их всегда лежат изменения тех или иных участков ДНК. Эти генетические варианты существенно различаются по размеру, характеру изменения, а также функциональному значению, которое напрямую зависит от затрагиваемых вариантов районов генома.

## **1.2. Типы генетических аномалий, лежащих в основе генетических патологий**

Генетические аномалии различаются по размеру. Размер непосредственно влияет на способность исследователя обнаружить эту аномалию. Самыми крупными являются хромосомные перестройки. Они делятся на две основных группы — геномные мутации и хромосомные перестройки, к которым в свою очередь относятся сбалансированные (без изменения количества генетической информации) и несбалансированные (с изменением количества генетической информации).

Геномные мутации у человека представлены анеуплоидиями — изменением числа хромосом. Примерами анеуплоидий могут служить синдром Дауна (трисомия 21 хромосомы), Эдвардса (трисомия 18 хромосомы), Патау (трисомия 13 хромосомы), а также вариации числа половых хромосом (синдромы Тёрнера, Клайнфельтера и другие). Пример частичной моносомии — синдром кошачьего крика (связан с утратой плеча 5 хромосомы). Прочие анеуплоидии ведут к несовместимым с жизнью нарушениям эмбрионального развития и, как следствие, спонтанным абортam; тем не менее, некоторые летальные анеуплоидии могут существовать в мозаичной форме.

Несбалансированные перестройки в большинстве своём приводят к летальному исходу (в эмбриональном или детском периодах) и грубым изменениям фенотипа. К несбалансированным относятся:

- Несбалансированные транслокации — перемещение фрагмента хромосомы с одного места на другое с изменением количества генетической информации. Несбалансированные транслокации могут приводить к значимым изменениям фенотипа (например, инверсия пола — [Rizvi, 2008](#)) и служить онкогенами [[O'Connor, 2008](#)].

- Вариации числа копий (англ. *Copy Number Variations, CNV*) — дупликации (мультипликации) и делеции хромосомных сегментов размером от тысячи до нескольких миллионов пар оснований. Могут возникнуть из несбалансированных транслокаций, амплификаций и собственно делеций. CNV способны увеличивать или уменьшать дозу гена, в значительной степени влияя на его экспрессию. Различия в количестве копий могут носить как положительный характер, так и отрицательный — в частности, дупликации в гене *CCL3L1* способны увеличить устойчивость к ВИЧ [Gonzalez, 2005], а крупные CNV в разных частях генома ассоциированы с расстройствами аутистического спектра [Sebat et al., 2007].

Сбалансированные перестройки чаще всего характеризуются более мягкими фенотипическими проявлениями, а иногда и их отсутствием. К сбалансированным перестройкам относятся:

- Инверсии — переворот фрагмента хромосомы. Крупные инверсии могут быть причиной изменения границы ТАД, а также загибания кроссинговера и образования гаплогрупп.
- Сбалансированные транслокации — перемещение фрагмента хромосомы с одного места на другое без изменения количества генетической информации. В свою очередь они делятся на реципрокные (взаимный обмен участками между негомологичными хромосомами) и Робертсоновские (слияние акроцентрических хромосом с образованием метацентрической или субметацентрической). Сбалансированные транслокации в основном не проявляются в фенотипе, сказываясь только на фертильности [Dong et al., 2012].

Самыми небольшими — но не менее важными — являются точечные полиморфизмы (англ. *Single Nucleotide Variants, SNV*) и короткие инсерции и делеции (англ. *indels*) размером 20–50 bp. Чаще всего эти генетические варианты нейтральные и не имеют фенотипических проявлений, но некоторые могут приводить как к генетическим, так и к эпигенетическим изменениям. Также варианты делятся на наследуемые, которые передаются от родителей



к детям, и варианты *de novo*. Согласно оценкам, предоставленным [Acuna-Hidalgo et al. \[2016\]](#), в среднем в каждом поколении у человека возникают 44–82 SNV *de novo*, из них 1–2 приходятся на белок-кодирующие регионы. Число небольших инсерций и делеций оценивается в 2.9–9 на геном, крупные перестройки встречаются значительно реже. Также известно, что количество генетических вариантов *de novo* непрерывно растёт в течение жизни человека.

### 1.3. Функциональные классы генетических вариантов

Как уже было упомянуто выше, значение генетических вариантов напрямую зависит от их положения относительно функциональных частей генома. Варианты могут находиться как внутри генов, так и вне их.

Области гена, в которые может попасть генетический вариант:

- Экзоны, непосредственно отвечающие за последовательность белка. Генетические варианты в экзонах могут быть синонимичными (без замены аминокислоты) и несинонимичными — миссенс (замена на другую аминокислоту), нонсенс (замена на стоп-кодон) либо сдвиг рамки считывания, приводящий к изменению значительной части белковой молекулы. Миссенс-варианты редко приводят к утрате функции белка, но они могут повлиять на экспрессию гена, если замена пришлась на регуляторный мотив [[Brea-Fernandez et al., 2011](#)].
- Интроны, которые содержат регуляторные области и сплайс-сайты, необходимые для процессинга транскрипта в готовую мРНК, а также 3'-нетранслируемая область (англ. *3'-untranslated region, 3'UTR*) и 5'-нетранслируемая область (англ. *5'-untranslated region, 5'UTR*), вовлечённые в регуляцию транскрипции, трансляции и деградации транскрипта. В частности, в 5'UTR находится так называемая консенсусная последовательность Козак, важная для инициации трансляции мРНК [[Kozak, 1987](#)]. Также известно, что в 5'UTR могут находиться открытые рамки считывания, которые влияют на поведение рибосомы — могут вызывать её



торможение, диссоциацию, либо перекрывать основной старт-кодон гена [Young and Wek, 2016]. Генетические варианты могут как разрушать канонические сплайс-сайты, так и способствовать образованию новых внутри интронных участков [Abramowicz and Gos, 2018]. Влияние генетических вариантов в этих областях недостаточно изучено, и их связь с конкретной патологией у пациента порой достаточно трудно доказать. Тем не менее, существуют специальные инструменты, позволяющие оценить патогенность таких вариантов. Интронные и UTR генетические варианты обычно рассматриваются в случае, если иного объяснения фенотипу пациента не было найдено.

Внегенные варианты могут приходиться на различные регуляторные последовательности, например, энхансеры, сайленсеры, а также сайты связывания белков, отвечающих за процессы метилирования или трёхмерную организацию хроматина.

Как мы видим, типов генетических вариантов существует огромное множество, они в значительной степени различаются между собой, и их определение может представлять трудность для исследователя. На сегодняшний день разработано множество методик, облегчающих эту задачу. О них речь пойдёт ниже.

## 1.4. Методы детекции генетических вариантов

**Кариотипирование.** Данный метод представляет собой микроскопическое исследование клеток, синхронизированных на стадии метафазы митоза. Однако простое микроскопическое исследование хромосом плохо подходит для обнаружения генетических вариантов, поэтому были разработаны различные методы окраски (бэндинга), позволяющие отдифференцировать отдельные хромосомы и хромосомные регионы [Schreck and Distèche, 2001]:

1. Q-окрашивание — позволяет отдифференцировать все хромосомы, применяется для исследования Y-хромосомы (быстрое определение генетического пола, выявление мозаицизма по Y-хромосоме, транслокаций между Y-хромосомой и другими хромосомами). Окрашивание легко

снимается, что позволяет использовать этот метод для последовательной окраски и изучения хромосом;

2. G-окрашивание — наиболее часто используемый метод. Позволяет дифференцировать все хромосомы, гарантирует стойкое окрашивание, легко поддаётся фотографированию.
3. R-окрашивание — визуализирует концы хромосом, а также специфические именно для этого окрашивания бэнды (так называемые R-позитивные бэнды).
4. C-окрашивание — применяется для анализа вариабельной дистальной части Y-хромосомы, а также центромерных регионов прочих хромосом, содержащих конститутивный гетерохроматин. Хорошо подходит для выявления перестроек, затрагивающих гетерохроматиновые регионы. Кроме того, C-окрашиванием хорошо определяются кольцевые и дицентрические хромосомы;
5. NOR-окрашивание — визуализирует ядрышковые организаторы (англ. *Nucleolus Organizer Region, NOR*), богатые рибосомальными генами;
6. DA–DAPI-окрашивание — применяется для идентификации центромерных гетерохроматизированных районов.

Окрашенные хромосомы далее изучаются на предмет формы, количества и наличия перестроек.

Кариотипирование — рутинная методика при диагностике врождённых патологий, аутопсии мертворожденных и злокачественных образований кроветворного ряда. Преимущества кариотипирования в том, что данным методом можно охватить весь геном, визуализации поддаются отдельные клетки и отдельные хромосомы. Ограничения — обязательно требуются живые клетки, также на эффективность влияет размер перестроек (не менее 1–5 Mbp) и процент поражённых клеток в образце (минимум 5–10 %) [[Sampson and McGuire, 2014](#)].

В целом классический метод кариотипирования, достаточно дешёвый и простой в исполнении, требует от исследователя значительного опыта при

интерпретации. Более поздние методы изучения хромосом, как будет показано далее, развивались не только в направлении увеличения разрешающей способности, но и облегчения интерпретации полученных данных.

**Флуоресцентная *in situ* гибридизация** (англ. *Fluorescence In Situ Hybridization, FISH*). Основой является гибридизация нуклеиновых кислот образца и комплементарных им проб, содержащих флуоресцентную метку. Гибридизация может производиться с ДНК (метафазные или интерфазные хромосомы) или с РНК. FISH позволяет определить число исследуемых локусов в геноме (при использовании метода 3D-FISH) или последовательность расположения на метафазной хромосоме. Метод является «золотым стандартом» в определении хромосомных патологий — как в клетках с врождёнными перестройками, так и в клетках опухолей.

Данные при помощи метода FISH можно получить, анализируя отсутствие или присутствие сигналов от использованных флюорофоров. Количество различимых цветовых меток равно  $(2^x - 1)$ , где  $x$  — количество флюорофоров. Это позволяет реализовать, например, спектральное кариотипирование (англ. *Spectral Karyotyping, SKY*), при котором каждая хромосома окрашивается в свой собственный цвет и межхромосомные перестройки видны даже начинающему специалисту [Guo et al., 2014]. Тем не менее, лимитирующими факторами остаются:

- потребность в хорошо обученном персонале. Относительная простота интерпретации результатов сочетается со сложностью протокола приготовления образца, который зависит от характера пробы и образца, и должен быть настроен эмпирически;
- цена реактивов;
- время гибридизации. Кинетика реакций гибридизации в ядре изучена недостаточно, и требуется достаточно долгое время, чтобы получить сигналы, которые можно измерить и сравнить между собой.
- разрешение. Детектировать сигнал от одной молекулы флюорофора очень сложно, такими молекулами должен быть покрыт протяжённый уча-

сток ДНК. Поэтому детектировать изменения участков размером менее 100 kbp достаточно затруднительно.

В настоящее время методика FISH значительно усложнилась. Биотехнологические компании предлагают панели олигонуклеотидов, определяющие специфические участки размером от десятков тысяч до миллиона пар оснований, а также олигонуклеотиды с высокой чувствительностью, позволяющие определить сплайс-варианты и даже SNV. Разрабатываются технологии *micro-FISH* ( $\mu$ FISH), сочетающие FISH с микрофлюидными технологиями (проведение реакций в микроскопических объёмах жидкости). При этом процесс удешевляется, автоматизируется, ускоряется (за счёт уменьшения объёмов, а соответственно, и времени гибридизации) и упрощается для использования в обширных исследованиях и для внедрения в клинику [Huber et al., 2018].

**Сравнительная геномная гибридизация** (англ. *Comparative Genomic Hybridization, CGH*). Как и в случае с методом FISH, основой данного метода является флуоресцентная гибридизация. Однако CGH использует два образца генома — тестовый и контрольный, каждый из которых метится флюорофором, а затем гибридизуется в соотношении 1 : 1. Таким образом в тестовом образце можно обнаружить CNV и перестройки.

В отличие от FISH, CGH проверяет весь геном на наличие несбалансированных перестроек и не требует знаний о целевом регионе. К ограничениям анализа относится невозможность выявления полиплоидии, мозаицизма и сбалансированных транслокаций.

В настоящее время CGH используется в виде *array-CGH* (aCGH), или хромосомного микроматричного анализа (ХМА), при котором CGH комбинируется с микрочиповой гибридизацией [Theisen, 2008]. ДНК-микрочипы, или микроматрицы, представляют собой сотни тысяч или миллионы однонитевых фрагментов ДНК (зондов), которые ковалентно пришиты к основанию (микрочипу). При ХМА на микрочип наносятся контрольные фрагменты генома либо контрольные последовательности генов, которые могут быть связаны с конкретной патологией. Порядок зондов на чипе строго определён, что упрощает локализацию и определение характера перестройки.

С помощью сравнительной гибридизации геномов могут быть обнаружены самые разные структурные вариации — CNV, инверсии, хромосомные транслокации и анеуплоидии. Для этого используются длинные зонды, которые позволяют проводить гибридизацию последовательностей, имеющих некоторые различия. Когда пробы ДНК короткие, эффективность гибридизации очень чувствительна к несовпадениям; такие зонды облегчают сравнение геномов на нуклеотидном уровне (поиск SNV).

Микроматрицы предлагают относительно недорогие и эффективные средства сравнения всех известных типов генетических вариаций. Однако для таких целей, как обнаружение сбалансированных перестроек, неизвестных или часто повторяющихся последовательностей, эти методы не подходят [Gresham et al., 2008].

**Мультиплексная лигаза-зависимая амплификация зонда** (англ. *Multiplex Ligation-dependent Probe Amplification, MLPA*). Основой MLPA является ПЦР-амплификация специальных проб, гибридизующихся с целевыми районами ДНК. Каждая проба представляет собой пару полу-проб; каждая полу-проба имеет комплементарную геному часть и технические последовательности — праймер для ПЦР и вставки, обеспечивающие большой размер продукта амплификации. Если полу-пробы гибридизуются с геномом без зазора, они лигируются и впоследствии амплифицируются; лигированные пробы отличаются от полу-проб с праймером по длине. Длину готового ПЦР-продукта определяют методом электрофореза.

Данная методика подходит для определения CNV, включающих целые гены, а также аномалий метилирования ДНК. Во втором случае используют метил-чувствительные рестриктазы — ферменты, которые по определённым сайтам гидролизуют исключительно метилированную ДНК. Для определения этих участков также применяют электрофорез, т.к. не подвергшаяся гидролизу ДНК по длине значительно превосходит фрагменты гидролизированных рестриктазой метилированных регионов.

Слабым местом MLPA остаётся интерпретация результатов. Определение гомозиготных CNV не представляет труда — их распознают по наличию/отсутствию пика в сравнении с контрольным образцом. Гетерозиготные

CNV видны как пики отличающейся высоты, и их поиск требует серьёзную биоинформационную обработку с учётом особенностей конкретной ПЦР-реакции и различий между образцами [[Stuppia et al., 2012](#)].

Как мы видим, перечисленные методы имеют один серьёзный недостаток — они могут определить наличие или отсутствие, совпадение или несовпадение, но не способны прочесть априори неизвестную последовательность ДНК. Специально для этого были разработаны методы секвенирования.

**Секвенирование по Сэнгеру.** Исторический метод, позволяющий с высокой точностью анализировать короткие (до 1 kbp) фрагменты ДНК [[Sanger et al., 1977](#)]. Суть его состоит в проведении обычной реакции амплификации ДНК, только в смесь дезоксирибонуклеотидов (dNTP) добавлены дидезоксирибонуклеотиды (ddNTP), которые при присоединении к ДНК обрывают синтез и имеют флуоресцентную или радиоактивную метку (соотношение примерно 100 : 1 соответственно). Таким образом, в процессе амплификации в пробирках образуется смесь из меченых цепей разной длины. При разделении этой смеси на электрофореze проявляется характерная «лестница», последовательность флуоресцентных сигналов в которой совпадает с последовательностью исследуемой ДНК.

Основным недостатком секвенирования по Сэнгеру является ограничение длины исследуемого фрагмента ДНК.

В настоящее время метод Сэнгера используется для подтверждения вариантов, найденных с помощью методов секвенирования нового поколения.

**Секвенирование нового поколения** (англ. *New Generation Sequencing, NGS*). Это комплекс технологий, позволяющих прочесть за сравнительно небольшое время миллионы последовательностей ДНК. Благодаря этому одновременно можно проанализировать несколько генов, либо весь геном.

В методах NGS наблюдается развитие двух основных парадигм, различающихся по длине прочтений. Секвенирование короткими прочтениями характеризуется меньшей ценой и более качественными данными, что позволяет применять данные методы в популяционных исследованиях и клиниче-

ской практике (поиск патогенных генетических вариантов). Секвенирование длинными прочтениями хорошо подходит для сборки новых геномов и изучения отдельных изоформ генов [Goodwin et al., 2016]. Количество различных методов в настоящее время значительно, но самым часто используемым является метод Illumina (короткие прочтения).

Основные проблемы данных NGS:

- Финансовые вложения и время, затраченные на секвенирование и анализ данных. По-прежнему остаются лимитирующим фактором применения NGS в клинической практике;
- Ошибки секвенирования и ПЦР. Их значимость уменьшается с увеличением покрытия, но не исчезает полностью;
- Неоднородность покрытия генома или таргетных регионов прочтениями. Это может быть связано как с недостатками приготовления библиотеки, так и с проблемами картирования.

## 1.5. Виды NGS

**Полногеномное секвенирование** (англ. *Whole Genome Sequencing, WGS*). Приготовление библиотек при полногеномном секвенировании производится из всего клеточного материала, либо только из ядер. ДНК фрагментируется таким образом, что достигается относительно ровное покрытие генома.

WGS при достаточной глубине покрытия вполне пригодно для поиска SNV, небольших делеций и инсерций. Полногеномное секвенирование со слабым покрытием может быть использовано для определения CNV — например, при неинвазивном пренатальном тестировании (англ. *Non-Invasive Prenatal Testing, NIPT*), когда используется свободная ДНК плода (англ. *Cell-Free Fetal DNA, cffDNA*), циркулирующая в крови матери [Yu et al., 2019].

**Таргетные панели.** Основой данных методов является обогащение целевых регионов генома. Методов обогащения существует достаточно много, но все они делятся на четыре основные категории [Teer and Mullikin, 2010]:

1. Твердофазная гибридизация. Для этого используют комплементарные целевым регионам короткие ДНК-пробы, зафиксированные на твёрдом основании (микрочипе). После гибридизации нецелевую ДНК вымывают, а целевые фрагменты остаются на чипе.
2. Жидкофазная гибридизация. Эти методы характеризуются тем, что ДНК-пробы находятся в растворе и помечены специальной молекулой (например, биотином). После гибридизации с целевой ДНК пробы вылавливают бусинами, поверхность которых способна связывать молекулы биотина.
3. Полимеразно-опосредованный захват. В этих методах ПЦР производят на стадии обогащения. Например, методы молекулярно импринтированных полимеров (англ. *Molecularly Imprinted Polymers, MIP*) и анализа транскрипта одной клетки (англ. *SMART*) используют длинные пробы, содержащие как праймер, так и регион для останковки элонгации и инициации лигирования. После элонгации и лигирования получают кольцевые молекулы, содержащие целевой регион; линейные молекулы в последующем удаляют из раствора. Метод захвата с помощью расширения праймера (англ. *Primer Extension Capture, PEC*) использует биотинилированные праймеры, которые гибридизуются с целевыми регионами и элонгируются; далее их вылавливают бусинами, как в методах жидкофазной гибридизации.
4. Захват регионов. Включает в себя сортировку и микродиссекцию хромосом, благодаря чему можно обогатить библиотеку фрагментов последовательностями отдельной хромосомы или даже её части. Это методы, требующие чрезвычайно сложных техник и хорошо обученный персонал, но очень полезные в отдельных ситуациях.

Данный вид тестов позволяет анализировать гены, ответственные за отдельные группы заболеваний — например, существуют таргетные панели для иммунодефицитов, почечных, неврологических болезней, болезней соединительной ткани, сетчатки, а также предрасположенности к отдельным видам онкологических заболеваний. Таргетные панели позволяют анализировать и



клетки опухолей — некоторые приспособлены к выявлению общих для многих раковых линий мутаций, другие же разработаны для специфического типа опухолей [Yohe and Thyagarajan, 2017].

**Полноэкзомное секвенирование** (англ. *Whole Exome Sequencing, WES*). Техника заключается в секвенировании обогащённого экзона — совокупности белок-кодирующих последовательностей клетки. Для этого используют специальные экзомные таргетные панели. Несмотря на то, что существует множество методов таргетного обогащения, конкретно для WES могут быть использованы лишь немногие из них, а именно — твердофазная и жидкофазная гибридизация [Teer and Mullikin, 2010].

У человека экзом составляет примерно 1 % от генома, или примерно 30 Mbp (суммарно). При этом более 80 % генетических вариантов, которые представлены в базе данных известных геномных вариантов CLINVAR [Landrum et al., 2017], и из них более 89 % вариантов, которые отмечены как «патогенные», относятся к белок-кодирующим областям генома; эта цифра приближается к 99 %, если учитывать ближайшие окрестности экзонов [Barbitoff et al., 2020]. Таким образом, полноэкзомное секвенирование намного лучше подходит для обычной клинической практики, нежели полногеномное. Кроме того, полноэкзомное секвенирование значительно дешевле, что увеличивает его доступность и позволяет, например, произвести тестирование ребёнка и родителей (так называемый трио-тест) и, как следствие, улучшить интерпретацию вариантов [Yohe and Thyagarajan, 2017].

**Технологии захвата конформации хромосом** (англ. *Chromosome Conformation Capture, 3C*). Данные методики позволяют определить расстояние в 3D-пространстве ядра между двумя точками генома. Принцип состоит в том, что интактное ядро фиксируют формальдегидом, ДНК гидролизуют, лигируют, затем продукты лигазной реакции секвенируют при помощи NGS. Во время лигирования ковалентно связанными могут оказаться только те участки, которые физически находятся близко друг от друга. Картирование химерных прочтений с помощью специальных инструментов позволяет узнать, какие именно участки генома были связаны, а значит, располагались близко друг к

другу в пространстве ядра [[Lieberman-Aiden et al., 2009](#)]. При обработке большого количества ЗС-данных геном разделяют на районы фиксированной длины, называемые бинами. Длина бинов называется разрешением; чем меньше длина, тем более высоким считается разрешение. Прочтение, части которого были картированы на два разных бина, называется контактом между этими районами. Практическое значение имеет информация об относительной частоте контактов между бинами.

В настоящее время существует множество вариантов протокола ЗС. Самым известным и широко применяемым является метод Hi-C, сочетающий ЗС с методами массового параллельного секвенирования. С его помощью можно подсчитать количество контактов во всём геноме — как внутри-, так и межхромосомные контакты [[Oluwadare et al., 2019](#)].

Результаты NGS представляют собой гигантские блоки данных, содержащие всевозможные ошибки. Обработка данных секвенирования — это высокотехнологичная отрасль, которая позволяет получить из этих данных практически значимую информацию и минимизировать влияние ошибок на эту информацию.

## **1.6. Базовая схема обработки результатов высокопроизводительного секвенирования для поиска и клинической интерпретации однонуклеотидных полиморфизмов**

**Демультиплексикация.** В процессе приготовления NGS-библиотеки к целевым фрагментам ДНК лигируют так называемые адаптерные последовательности, или адаптеры. Очень часто потенциальное количество прочтений, которое способен выдать секвенатор за один запуск, значительно превышает требуемое количество прочтений для отдельной библиотеки, поэтому из соображений экономии и повышения производительности на одном чипе секвенируют сразу несколько библиотек. Для этого в адаптеры вставляют баркоды — последовательности, с помощью которых можно отличить прочтения,

относящиеся к разным библиотекам или образцам. Процесс сортировки данных секвенирования по баркодам называется демультиплексикацией.



Рис. 1: Результаты секвенирования библиотек, содержащих короткие последовательности, могут быть контаминированы адаптерами. На рисунке показаны результаты секвенирования последовательностей, длина которых превышает количество циклов секвенирования (сверху), либо значительно меньше количества циклов (снизу).

**Удаление адаптерных последовательностей.** Если целевой фрагмент ДНК короче длины прочтения, то фрагменты адаптерной последовательности могут попасть в готовые данные (Рис. 1). Это замедляет работу алгоритма картирования, а порой в значительной степени ухудшает его результаты, поэтому встаёт вопрос об удалении адаптерных последовательностей. Также присутствие адаптера в прочтениях может быть признаком контаминации библиотеки, и такие прочтения следует исключить из дальнейшего анализа [Martin, 2011].

**Картирование прочтений.** Как уже упоминалось выше, результаты NGS — это прочтения, содержащие небольшие (в пределах 200 bp) фрагменты генома. Извлечение информации из необработанных результатов секвенирования затруднительно, так как эти фрагменты содержат много ошибок (как в результате ПЦР-реакции, так и допущенные в процессе секвенирования) и не имеют никакой информации о регионе, из которого они произошли. Поэтому прочтения необходимо картировать на некую референсную геномную последовательность. Алгоритм картирования представляет собой очень сложную систему, которая учитывает последовательность букв в прочтении и качество прочтения. Качество прочтения отражает вероятность того, что буква, прочитанная секвенатором, совпадает с реальным нуклеотидом в данной позиции. Обычно качество прочтения записывается в шкале Phred, к которой приводится формулой

$$Q = -10 \log_{10} P, \quad (1)$$

где  $P$  — вероятность того, что нуклеотид прочтен правильно. Было разработано множество алгоритмов картирования, но в настоящее время «золотым стандартом» являются утилиты, использующие алгоритм Берроуса—Уиллера [Burrows and Wheeler, 1994].

Обычно алгоритм картирования выставляет выравниванию коэффициент, называемый качеством выравнивания (англ. *MAPping Quality*, *MAPQ*). MAPQ отражает вероятность правильности картирования и также записывается в шкале Phred (Формула 1). В силу размеров референсной последовательности в ней существует огромное множество повторов и похожих регионов. Современные алгоритмы могут находить несколько потенциальных мест картирования для одного прочтения, и их количество влияет на качество выравнивания.

Также алгоритмы способны разделять прочтение на участки, которые могут быть картированы в разные места генома. По этому признаку прочтения делятся на линейные и химерные. В линейных прочтениях не может быть изменения направления картирования, т.е. картированная часть может иметь только прямое направление, либо только обратное направление относительно генома. Химерные прочтения имеют картированные части с разным направлением. Эти участки могут перекрываться, и количество перекрытий также

влияет на MAPQ.

Исходя из особенностей алгоритмов картирования, выравнивания делятся на следующие классы:

- Первичное выравнивание (англ. *primary*) — выравнивание наиболее крупного (и содержащего наименьшее количество перекрытий, в случае химерного прочтения) фрагмента прочтения с наиболее высоким MAPQ. Первичное выравнивание только одно. Первичное выравнивание химерного прочтения называется репрезентативным;
- Вторичное выравнивание (англ. *secondary*) — выравнивание наиболее крупного фрагмента прочтения с меньшим MAPQ. Вторичных выравниваний может быть несколько (в зависимости от выставленного нижнего порога MAPQ);
- Добавочное выравнивание (англ. *supplementary*) — выравнивание менее крупных (либо содержащих большее количество перекрытий) фрагментов прочтения. Добавочные выравнивания характерны только для химерных прочтений.

Картированный участок может содержать в себе несовпадения с референсной последовательностью, инсерции и делеции. Это могут быть как ошибки, так и генетические варианты, поэтому данная информация безусловно важна при анализе данных. Также в частично картированных прочтениях могут присутствовать некартируемые участки с 3'- или 5'-конца. В отличие от делеций внутри картированных участков, некартированные концы обычно подвергаются так называемому клипированию и в дальнейшем не учитываются при анализе. Клипирование бывает двух типов:

- Мягкое клипирование (англ. *soft-clip*) — отсечение невыровненного конца прочтения с сохранением полной последовательности прочтения. В отсечённых методом мягкого клипирования регионах могут быть адаптерные последовательности, а также часть химерного прочтения (в репрезентативном выравнивании).

- Жёсткое клипирование (англ. *hard-clip*) — отсечение невыравненного конца прочтения без сохранения его последовательности. В регионах, подвергшихся жёсткому клипированию, обычно находятся репрезентативные участки химерных прочтений (в добавочных выравниваниях).

Основные проблемы картирования:

- Высоковариативные регионы. Алгоритм картирования разработан для поиска наиболее полных соответствий, и при большом количестве несовпадений прочтение просто не сможет быть картировано на нужный регион генома;
- Вырожденные (неуникальные) регионы. Соответствие между регионами может привести к неправильному распределению прочтений между ними, а значит — и неправильному картированию генетических вариаций. Кроме того, генетические варианты в регионах с короткими повторами в принципе невозможно картировать точно, поэтому обычной практикой является левое смещение (англ. *left-align*) [Tan et al., 2015].
- Регионы с инсерциями и делециями. Помимо того, что сами по себе эти варианты сильно ухудшают картирование, содержащие их прочтения могут быть картированы неправильно (из-за того, что алгоритмы картирования используют случайно выбранные позиции в геноме для начала поиска соответствий). Из-за этого могут возникать ложные SNP, а пропорции аллелей могут быть посчитаны неправильно. Пример показан на Рис. 2.

**Удаление дубликатов.** Так как молекулы ДНК очень малы, вероятность их разрушения или возникновения в них ошибок велика, а полученные от них сигналы находятся за пределами чувствительности многих современных приборов. Решением этих проблем является амплификация молекул ДНК. Амплификация может быть как на стадии приготовления библиотеки (ПЦР), так и на стадии секвенирования. При секвенировании амплификация и последующее объединение ампликонов в кластер производятся для усиления сигнала

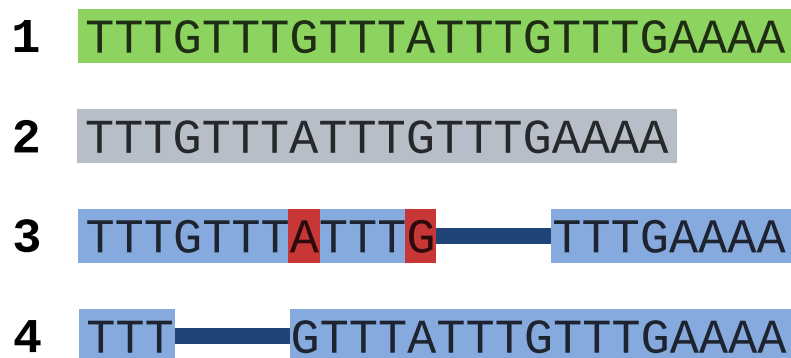


Рис. 2: Неоптимальное картирование прочтения, содержащего делецию. (1) — референсная последовательность, (2) — последовательность прочтения, (3) — картирование, произведённое алгоритмом, включающее две SNV и одну делецию, (4) — оптимальное местоположение делеции

и нивелирования ошибок, происходящих на каждом цикле секвенирования с отдельными молекулами. Соответственно, в процессе секвенирования возникают дубликатные прочтения, которые могут быть как ПЦР-дубликатами библиотеки, так и возникать из-за ошибок распознавания кластеров амплификации (оптические дубликаты). Согласно принятой практике, дубликаты должны быть удалены или помечены для улучшения поиска генетических вариантов [Auweru et al., 2013].

Однако было показано, что для WGS-данных удаление дубликатов имеет минимальный эффект на улучшение поиска полиморфизмов — приблизительно 92 % из более чем 17 млн вариантов были найдены вне зависимости от наличия этапа удаления дубликатов и использованных инструментов для поиска дубликатов [Ebbert et al., 2016]. Учитывая, что удаление дубликатов может занимать значительную часть потраченного на обработку данных времени, следует взвесить пользу и затраты данного этапа для конкретной прикладной задачи.

**Рекалибровка качества прочтений** (англ. *Base Quality Score Recalibration, BQSR*). В приборной оценке качества прочтений всегда имеют место систематические ошибки. Это связано как с особенностями физико-химических реакций в секвенаторе, так и с техническими недостатками оборудования.

Вычисление качества прочтения — сложный алгоритм, защищённый авторскими правами производителя секвенатора. Вместе с тем от качества прочтений напрямую зависит алгоритм поиска вариантов — он использует данный коэффициент как вес в пользу присутствия или отсутствия генетического варианта в конкретной точке генома.

Решением является рекалибровка качества прочтений, представляющая собой корректировку систематических ошибок, исходя из известных паттернов зависимости случайных величин. Следует заметить, что рекалибровка не помогает определить, какой нуклеотид в реальности находится в данной позиции — она лишь указывает алгоритму поиска генетических вариантов, выше или ниже вероятность правильного прочтения нуклеотида секвенатором.

Первоочередное влияние на ошибки оказывают:

1. Собственно прибор (секвенатор) и номер запуска. Большая часть секвенаторов выставляет прочтению более высокое качество прочтения по сравнению с ожидаемым, гораздо реже встречаются модели, занижающие качество прочтения [Auwera et al., 2013]. Каждый отдельный запуск может различаться по параметрам чипа и химических реагентов;
2. Цикл секвенирования. Качество прочтения уменьшается с каждым циклом за счёт накопления ошибок в кластере амплификации;
3. Нуклеотидный контекст. Систематические ошибки, связанные с физико-химическими процессами, влияют на качество прочтения нуклеотида в зависимости от предшествующего ему динуклеотида.

Кроме того, алгоритм рекалибровки учитывает изменчивость каждого отдельного сайта, используя базы данных известных генетических вариантов. Высокая изменчивость повышает вероятность правильного прочтения нуклеотида, не совпадающего с референсным в данной позиции генома.

Институт Броуд (англ. *Broad Institute of MIT and Harvard*) рекомендует BQSR к использованию для любых данных секвенирования [Auwera et al., 2013].



**Поиск генетических вариантов.** Невозможно точно сказать, какой нуклеотид находится в каждой позиции генома. Анализ производит специальный алгоритм, который оценивает качество прочтения, качество выравнивания и процент букв в данной позиции на картированных прочтениях. Отличие генома образца от референсного генома называется генетическим вариантом. Алгоритм выставляет каждому генетическому варианту коэффициент качества варианта (англ. *VCF QUAL*), записываемый в шкале Phred (Формула 1). Помимо определения генетического варианта, алгоритм может определять его зиготность.

Также важным этапом поиска вариантов является уже упомянутое выше левое выравнивание. Варианты в повторяющихся последовательностях с длиной менее длины одного прочтения невозможно точно локализовать, поэтому они всегда сдвигаются как можно левее относительно последовательности генома. Это чрезвычайно важно при аннотации генетических вариантов, так как все БД используют данные с левым выравниванием, и неправильная локализация может привести к отсеиванию потенциально патогенного варианта.

После того, как генетические варианты найдены, можно приступать к поиску тех, которые связаны с конкретной патологией у пациента. Однако только в кодирующих областях генома количество генетических вариантов достигает 100 тыс. (из них около 86 % SNV, 7 % инсерций и 7 % делеций) [Supernat et al., 2018], из них с патологиями связаны единицы. Даже после жёсткой фильтрации приходится работать минимум с сотней подходящих генетических вариантов. Это делает серьёзной проблемой поиск нужного варианта и интерпретацию полученных результатов.

## 1.7. Аннотация, фильтрация и интерпретация результатов

Первое, что следует сделать — это определить, насколько генетический вариант значим для нашего исследования, то есть аннотировать его. Существуют две основных парадигмы аннотации генетического варианта — это ан-

нотация по региону и аннотация по координате.

Основные методы аннотации по региону:

1. Функциональный класс. Для определения функционального класса генетического варианта существуют три основных базы данных: knownGene, refGene и ensGene. Они содержат информацию о генах, их частях и транскриптах — координаты, направление, а также номера экзонов и интронов. Координаты в этих базах данных могут различаться [McCarthy et al., 2014], поэтому, во избежание ошибок, рекомендуется использовать их все. Это особенно важно при дифференциации генетических вариантов с высокой вероятностью повреждающего эффекта (сдвиги рамок считывания, нонсенс-кодоны). Кроме того, различаются алгоритмы определения функционального класса в различных утилитах аннотации, что также создаёт определённые трудности [Jesaitis, 2014].
2. Клиническая значимость гена. Количество генетических вариантов для поиска можно сузить, зная, какие именно гены могут быть связаны с наблюдаемым у пациента фенотипом. Для поиска генов по клинической значимости существуют такие базы данных, как OMIM [Amberger et al., 2014] и OrphaData [French National Institute of Health and Medical Research (INSERM), 1997].
3. Потеря функции (англ. *Loss of Function, LoF*). Различные показатели, отражающие устойчивость функции гена, основанные на данных о стоп-кодонах, сдвигах рамки считывания и сплайс-вариантах. Одним из таких показателей является pLI. pLI отражает вероятность гаплонедостаточности гена, вычисляемую на основе количества белок-повреждающих генетических вариантов (сдвиги рамки считывания, нонсенс- и сплайс-варианты), взятых из контрольных баз данных, с поправкой на размер гена и глубину его покрытия при секвенировании [Ziegler et al., 2019].

Основные проблемы pLI [Ziegler et al., 2019]:

- Плохо приспособлен к распознаванию аутосомно-рецессивных вариантов (из-за того, что частота повреждающих вариантов в популяции может быть высокой) и X-сцепленных рецессивных вари-

антов (из-за наличия в популяции здоровых гетерозиготных носителей).

- Плохо приспособлен к распознаванию генетических вариантов в генах, ответственных за патологии, не влияющие на взросление и воспроизводство. Их частота в популяции также может быть высокой. К таким относятся варианты в генах *BRCA1* и *BRCA2*, ответственных за рак молочной железы.
- Сплайс-варианты априори рассматриваются как повреждающие, несмотря на то, что вариант в сайте сплайсинга может не иметь эффекта на сплайсинг, либо приводить к появлению изоформы белка без потери функции.
- Высокая частота распространения заболевания в контрольной группе. Пример — шизофрения.
- К миссенс-вариантам rLI применять следует с осторожностью, и без клинических данных следует исключить из анализа.
- Также следует отнестись с осторожностью к нонсенс-вариантам и сдвигам рамки считывания в последнем экзоне либо в С-терминальной части предпоследнего. Такие транскрипты избегают нонсенс-индуцированной деградации РНК и могут в результате как не привести к каким-либо функциональным изменениям, так и привести к образованию мутантного белка, обладающего меньшей активностью по сравнению с исходным, либо токсичного для клетки.
- В некоторых случаях соотношение rLI с гаплонедостаточностью конкретного гена в принципе сложно объяснить.

Таким образом, высокое значение rLI можно считать хорошим показателем LoF, низкое — с осторожностью.

Аннотация по координате обычно предназначена для миссенс-, интронных и сплайс-вариантов, связь которых с патологическим состоянием значительно сложнее выявить и доказать.

1. Частота аллеля в популяции. Многие тяжёлые генетические патологии испытывают на себе давление отбора, а значит, вызывающие их генетические варианты не могут иметь высокую частоту в популяции. Фильтрация по частоте является одним из базовых способов фильтрации генетических вариантов. Следует заметить, однако, что низкая частота генетического варианта далеко не всегда связана с его патогенностью, поэтому рассматривать низкую частоту как доказательство патогенности некорректно.

По мере развития методов NGS и увеличения их доступности, начали появляться базы данных, агрегирующие результаты секвенирования различных популяций, а значит — способные определить частоту генетических вариантов в популяции. В настоящее время наиболее крупной является gnomAD [[Karczewski et al., 2020](#)], поглотившая существовавший ранее ExAC, содержащий исключительно экзомные данные. Она содержит частоты генетических вариантов для всех основных рас, а также некоторых условно-здоровых групп.

Несмотря на то, что были созданы базы данных для всех рас, очень часто этого недостаточно и необходимо учитывать частоты в популяциях отдельных народов и стран. Такими базами данных являются GME [[Scott et al., 2016](#)], в которой отражены частоты по популяции Ближнего Востока, ABraOM [[Naslavsky et al., 2017](#)], предоставляющая частоты генетических вариантов среди практически здорового пожилого населения Бразилии. Также для анализа берутся популяции, в которых велика доля близкородственных связей, например, пакистанская [[Saleheen et al., 2017](#)].

2. Клинические данные из БД и статей. Наиболее достоверным источником данных о патогенности генетического варианта являются семейные и популяционные исследования конкретной патологии, а также базы данных, агрегирующие информацию из подобных статей. Наиболее используемыми в настоящее время являются HGMD [[Stenson et al., 2017](#)] и CLINVAR [[Landrum et al., 2017](#)]. Тем не менее, CLINVAR считается лишь дополнительным источником, так как часто содержит информа-

цию низкого качества [Ryzhkova et al., 2017].

3. Анализ и предсказание функционального эффекта *in silico*. *In silico* методы появились в ответ на необходимость как-то классифицировать генетические варианты, по которым недостаточно клинической информации. Существует множество способов проверить патогенность таких вариантов *in vitro*, но проверять таким образом все нецелесообразно, а иногда и невозможно. Даже в хорошо изученных генах варианты с неопределённой клинической значимостью могут занимать большую долю — например, в генах *BRCA* (*BRCA1* и *BRCA2*) количество таких вариантов даже в хорошо изученных популяциях (белые американцы) варьируется от 5 % до 15 %; этот показатель значительно выше в популяции чернокожих американцев и латиноамериканцев [Chern et al., 2019]. Менее изученные гены, а также пациенты, принадлежащие к популяциям с плохо изученным составом генетических вариантов, представляют ещё большую проблему.

Поэтому были разработаны инструменты на основе машинного обучения, предсказывающие консервативность районов и патогенность генетических вариантов на основе имеющихся данных — положения относительно гена и его функциональных элементов, характера замены, а также клинической информации об известных заменах [Brea-Fernandez et al., 2011]. Предсказательная способность отдельных инструментов оставляет желать лучшего, поэтому чаще всего в клинической практике используются агрегаторы, собирающие предсказания с большого числа известных *in silico* инструментов.

Значимость вклада каждого отдельного фактора достаточно сложно оценить. Эту проблему решают калькуляторы патогенности, которые по специальным критериям присваивают генетическому варианту ранг, отражающий вероятность повреждающего действия [Ryzhkova et al., 2017].

**Когортный и семейный анализ.** В случае, если исследователь имеет доступ к группе, представители которой связаны узами крови с пациентом, есть

возможность провести семейный анализ. Семейный анализ нужен для установления путей наследования тех или иных генетических вариантов в родословной. Это позволяет уточнить их связь с фенотипом. Также анализ нескольких родственных образцов помогает определить зиготность варианта, обнаружить генетические варианты *de novo*, либо импутировать район с недостаточным покрытием.

Если же в распоряжении исследователя находится группа, связанная одной патологией или вариантом фенотипа, можно провести когортный анализ. Когортный анализ позволяет, например, оценить частоты генетических вариантов в исследуемой и контрольной группе. Кроме того, когортный анализ образцов в конкретной лаборатории помогает детектировать систематические отклонения покрытия и артефакты выравнивания, связанные с конкретными районами генома и/или особенностями приготовления библиотек.

## 1.8. Ехо-С: суть метода

Как уже упоминалось выше, одним из основных ограничений NGS-технологий в настоящее время является их цена, напрямую зависящая от глубины секвенирования библиотеки. Есть ограничения и по возможностям поиска тех или иных генетических вариантов. Hi-C на сегодняшний момент является наиболее перспективным способом обнаружения хромосомных перестроек [Melo et al., 2020], но при небольшой глубине секвенирования в них обнаружение точечных полиморфизмов затруднительно [Sims et al., 2014]. WGS способно обнаруживать большую часть SNV, небольших инсерций и делеций, но требует большую глубину секвенирования [Sims et al., 2014]; WES, с другой стороны, позволяет выявить генетические варианты при небольшой глубине секвенирования, но только в экзоне. Возможности обнаружения хромосомных перестроек для последних двух методов ограничены.

Компромиссом между ценой и возможностями поиска генетических вариантов может служить новейший метод Ехо-С, сочетающий технологии таргетного обогащения с Hi-C. Суть его заключается в приготовлении Hi-C-библиотеки и последующем обогащении только тех последовательностей, которые связаны с экзоном. Таким образом, с его помощью можно как искать

точечные варианты в обогащённых регионах (за счёт большой глубины покрытия в них), так и хромосомные перестройки во всём геноме (за счёт Hi-C, дающей относительно небольшое, но доступное для анализа покрытие всего генома) [[Mozheiko and Fishman, 2019](#)].

Тем не менее, как выяснилось, уже существующие биоинформационные методы следует модифицировать для корректной обработки данных Ехо-С. Это связано в первую очередь с особенностями протокола Hi-C, к примеру, наличием технических последовательностей (бридж-адаптеров), которые приводят к появлению ложных SNV в экзомных регионах. Таргетное обогащение, со своей стороны, вносит определённые помехи в Hi-C-данные, так как изменяется представленность регионов генома в библиотеке, а значит, и пропорции контактов между регионами.

Данная работа посвящена разработке биоинформационных методов для поиска точковых генетических вариантов в Ехо-С-данных и последующего сравнения Ехо-С с методами полногеномного и полноэкзомного секвенирования. Поиск хромосомных перестроек в данных Ехо-С — отдельная большая задача, которой в данный момент занимается мой коллега.

## Глава 2. Материалы и методы

**Данные секвенирования.** Поиск данных полногеномного и полноэкзомного секвенирования производился в базах данных NCBI (GEO DataSets, SRA, PubMed) и ENCODE с использованием ключевых слов “K562”, “K562+WGS”, “K562+WES”, “K562+Hi-C”. Данные Ехо-С были получены путём секвенирования библиотек K562 и пациентов. Эти библиотеки были приготовлены в нашей лаборатории и далее секвенированы в компании BGI Genomics<sup>10</sup>.

**Контроль качества NGS-данных.** Для контроля качества прочтений мы использовали утилиту FastQC [Andrews, 2010], способную оценивать наличие адаптерных последовательностей, распределение прочтений по длине, GC-состав прочтений, а также производить анализ зависимости нуклеотидного состава от позиции в прочтении. Критерии качества были использованы согласно протоколу разработчика [Andrews, 2010].

**Удаление адаптерных последовательностей.** Удаление адаптерных последовательностей производилось с помощью утилиты cutadapt [Martin, 2011]. В Auwera et al. [2013] рекомендуется использовать в качестве входных данных некартированный BAM-файл (англ. *Unmapped Binary sequence Alignment/Map, uBAM*), а для удаления адаптеров использовать их собственный инструмент — MarkIlluminaAdapters, так как это позволяет сохранить важные метаданные. Тем не менее, был сделан акцент на том, что uBAM должен использоваться как выходной формат на уровне секвенатора, что не является общепринятой практикой.

Мы использовали данные секвенирования в формате FastQ. Преобразование FastQ-файлов в uBAM не предотвращает потерю метаданных, но значительно увеличивает время обработки данных. Сравнение эффективности cutadapt и MarkIlluminaAdapters в процессе удаления адаптеров не показало каких-либо значимых различий.

---

<sup>10</sup>BGI Genomics. Building No. 7, BGI Park, No. 21 Hongan 3<sup>rd</sup> Street, Yantian District, Shenzhen 518083, China. Tel: (+86) 400-706-6615.



**Картирование.** Картирование производилось с помощью инструментов Bowtie2 [Langmead and Salzberg, 2012] и BWA [Li and Durbin, 2009]. BWA показал лучшие результаты; кроме того, он значительно более эффективно работает с химерными ридами, что немаловажно для используемого нами метода Echo-S.

Для картирования был взят геном GRCh37/hg19, предоставленный NCBI. Из него были удалены так называемые неканоничные хромосомы (некартированные/вариативные референсные последовательности), что позволило улучшить качество выравнивания и значительно упростить работу с готовыми данными.

Кроме того, для правильного функционирования инструментов на дальнейших этапах был разработан скрипт, создающий метку группы прочтений (англ. *Read Group tag*, *RG*) для каждого файла. Конкретных рекомендаций по составлению RG не существует, поэтому мы разработали собственные, основанные на следующих требованиях [Auwera et al., 2013]:

- Поле SM является уникальным для каждого биологического образца и используется при поиске вариантов. Несколько SM в одном файле могут быть использованы при когортном анализе.
- Поле ID является уникальным для каждого RG в BAM-файле. BQSR использует ID как идентификатор самой базовой технической единицы секвенирования.
- Поле PU не является обязательным. Рекомендации GATK советуют помещать в него информацию о чипе секвенирования (баркод чипа), ячейке и баркоде (номере) образца. Во время BQSR поле PU является приоритетным по отношению к ID.
- Поле LB является уникальным для каждой библиотеки, приготовленной из биологического образца. Оно отражает различия в количестве ПЦР-дубликатов и потому используется инструментом MarkDuplicates.

Объединение BAM-файлов производилось инструментом MergeSamFiles. Сбор статистики по картированию мы осуществляли с помощью инструмента SAMTools flagstat [Li et al., 2009].

**Удаление ПЦР-дубликатов.** Для улучшения данных экзомного секвенирования в пайплайн был включён этап удаления ПЦР-дубликатов. Обычно этот процесс занимает много времени, но количество образцов у нас было относительно небольшим, и мы были заинтересованы в максимально качественной подготовке данных.

Удаление дубликатов производилось инструментом MarkDuplicates от Picard [[Broad Institute, 2019](#)], интегрированным в GATK. Оптимальные показатели скорости MarkDuplicates достигаются при запуске Java с параллелизацией сборщиков мусора и количеством сборщиков мусора равным двум [[Heldenbrand et al., 2019](#)]. Также, согласно рекомендациям разработчиков, прочтения были предварительно отсортированы по именам, чтобы удалению подверглись не только первичные, но и добавочные выравнивания [[Auwera et al., 2013](#)].

**Рекалибровка качества прочтений (BQSR).** Рекалибровка производилась с помощью инструментов GATK BaseRecalibrator и GATK ApplyBQSR. Для обучения машинной модели требуются генетические варианты в VCF-формате (согласно рекомендациям для *Homo sapiens* — dbSNP v132+).

К сожалению, предоставленная Broad Institute база данных оказалась сильно устаревшей и не вполне подходила для сделанной нами геномной сборки, поэтому было решено подвергнуть обработке dbSNP v150, предоставленную NCBI [[Sherry, 2001](#)]. База данных потребовала замену и сортировку контигов в соответствии с референсным геномом, а также удаление «пустых» вариантов, содержащих точки в полях REF и ALT. Далее база данных была архивирована с помощью bgzip, а затем проиндексирована GATK IndexFeatureFile (этот же инструмент одновременно проверяет БД на пригодность для BQSR).

В [Heldenbrand et al. \[2019\]](#) было показано, что оптимальные показатели скорости BaseRecalibrator достигаются, как и в случае с MarkDuplicates, запуском Java с двумя параллельными сборщиками мусора; кроме того, BaseRecalibrator поддается внешнему распараллеливанию путём разделения картированных прочтений на хромосомные группы. Хромосомные группы формировались вручную для используемой сборки генома, каждая запускалась с

помощью bash-скрипта. Нам удалось усовершенствовать данный этап — запуск BaseRecalibrator производился с помощью библиотеки Python subprocess, а параллелизация осуществлялась библиотекой multiprocessing, таким образом, можно было делить файл с картированными прочтениями по хромосомам и обрабатывать их отдельно, так как multiprocessing автоматически распределяет процессы по имеющимся потокам. Также для повышения отказоустойчивости скрипта у BaseRecalibrator и ApplyBQSR была устранена разница в фильтрации прочтений, из-за которой при малых размерах библиотек пайплайн экстренно завершал работу.

**Оценка покрытия и обогащения.** Покрытие и обогащение в экзOME оценивались с помощью скрипта на основе BEDTools [Quinlan and Hall, 2010].

**Поиск вариантов.** Поиск вариантов производился с помощью инструмента GATK HaplotypeCaller. Инструмент запускался с дополнительным параметром `--dont-use-soft-clipped-bases`, который не позволял использовать для поиска генетических вариантов клипированные химерные части и адаптеры.

Как и в случае с BaseRecalibrator, HaplotypeCaller поддаётся внешнему распараллеливанию [Heldenbrand et al., 2019]. Мы также осуществили параллелизацию с помощью сочетания subprocess и multiprocessing, достигнув 10–12-кратного ускорения по сравнению с запуском на одном потоке.

**Рекалибровка и ранжирование вариантов.** В GATK также присутствуют инструменты для рекалибровки и ранжирования вариантов, с использованием моделей машинного обучения и баз данных с частыми вариантами (CNNScoreVariants и FilterVariantTranches).

Анализ показал, что при наличии этапа рекалибровки вариантов время обработки результатов секвенирования увеличивается почти вдвое. Между тем, рекалибровка и ранжирование с помощью инструментов GATK не исключают необходимость фильтрации генетических вариантов. Таким образом, от этого этапа решено было отказаться.

**Аннотация вариантов.** Аннотация вариантов производилась с помощью инструмента ANNOVAR [Wang et al., 2010].

Используемые базы данных:

1. Human Gene Mutation Database (HGMD®) [Stenson et al., 2017]
2. Online Mendelian Inheritance in Man (OMIM®) [Amberger et al., 2014]
3. GeneCards®: The Human Gene Database [Stelzer et al., 2016]
4. CLINVAR [Landrum et al., 2017]
5. dbSNP [Sherry, 2001]
6. Genome Aggregation Database (gnomAD) [Karczewski et al., 2020]
7. 1000Genomes Project [The 1000 Genomes Project Consortium, 2015]
8. Great Middle East allele frequencies (GME) [Scott et al., 2016]
9. dbNSFP: Exome Predictions [Liu et al., 2016]
10. dbSCSNV: Splice site prediction [Jian et al., 2013]
11. RegSNPintron: intronic SNVs prediction [Lin et al., 2019]

Проблемные регионы:

1. ENCODE Blacklist [Amemiya et al., 2019] — содержит регионы, значительно чаще картируемые на несколько районов генома, а также картируемые на специфические участки — сателлитные, центромерные и теломерные районы.
2. Genome-In-A-Bottle [Zook et al., 2014] — содержит регионы, на которых сложно обнаружить генетические варианты (из-за низкого покрытия прочтениями, систематических ошибок секвенирования и локальных проблем картирования).
3. NCBI GeT-RM [Mandelker et al., 2016] — содержит набор регионов с высокой гомологией, которые сложно или невозможно анализировать с помощью секвенирования по Сэнгеру или NGS короткими прочтениями.

**Фильтрация генетических вариантов.** Аннотации были агрегированы для удобства использования. Так, агрегации подверглись:

- Имена генов по разным БД — для облегчения поиска;
- Описания функциональных классов из разных БД — для устранения несоответствий между ними;
- Ранги инструментов, предсказывающих патогенность генетического варианта. Трёхранговые системы (патогенный, вероятно патогенный и безвредный) были сведены к двухранговой (патогенный и безвредный). Отдельно были агрегированы предсказательные инструменты для экзонов, инструменты для интронов и сплайс-вариантов также учитывались отдельно;
- Ранги инструментов, предсказывающих консервативность нуклеотида. Эмпирическим путём было подобрано пороговое значение 0,7 — нуклеотид считался консервативным, если его предсказанная консервативность была выше, чем у 70 % всех нуклеотидов. Это максимальное пороговое значение, которое обеспечивает распределение балла агрегатора от минимального до максимального (от 0 до 7 баз данных, считающих данный нуклеотид консервативным);
- Популяционные частоты — из всех имеющихся в базах данных по конкретному генетическому варианту была выбрана максимальная частота.

Фильтрация происходила в две стадии:

1. Фильтрация отдельных генетических вариантов на основе имеющихся аннотаций. Самая жёсткая фильтрация, которой подвергались все варианты:
  - По глубине покрытия. Генетический вариант считался существующим, если он присутствовал в двух перекрывающихся парных прочтениях, либо в четырёх независимых прочтениях;

- Частота генетического варианта в популяции не более 3 % [[Ryzhkova et al., 2017](#)].
- По гомологии регионов — из анализа полностью исключались варианты, попавшие в «мёртвую зону» NGS и Сэнгер-секвенирования (геномные повторы длиной свыше 250 bp и 1 kbp соответственно).

Прочие фильтры были мягкими — генетический вариант отсеивался только в случае несоответствия всем указанным критериям:

- Присутствие описания связанной с геном патологии в базе данных OMIM;
- Присутствие генетического варианта в базе данных HGMD;
- Балл агрегатора патогенности экзомных вариантов не менее 3 [[Ryzhkova et al., 2017](#)];
- Ранг «патогенный» у агрегаторов интронных или сплайс-вариантов;
- Ранги «патогенный» и «возможно патогенный» по базе данных CLINVAR;
- По функциональному классу: сдвиги рамки считывания, потери стоп- и старт-кодонов, нонсенс- и сплайс-варианты.

2. Фильтрация значимых вариантов на основе аннотаций гена. Все эти фильтры были мягкими — ген мог соответствовать одному любому из перечисленных критериев:

- Значение rLI более 0,9, согласно рекомендациям в оригинальной статье [[Lek et al., 2016](#)];
- Наследование в гене значится как «доминантное» по базе данных OMIM, либо информации о доминантности нет;
- Любой значимый вариант в гене является гомозиготным;
- В гене более одного значимого варианта (вероятность цис-транс-положения).

**Интерпретация.** Интерпретация данных и составление отчёта производилось в соответствии с рекомендациями Американского колледжа медицинской генетики и геномики (англ. *American College of Medical Genetics, Bethesda, MD, USA*) и Ассоциации молекулярной патологии [[Richards et al., 2015](#)].

# Глава 3. Результаты

## 3.1. Результаты секвенирования

### Ехо-С-библиотек

Несмотря на то, что составляющие протокола Ехо-С — таргетное обогащение и Hi-C — в настоящее время достаточно отработаны, сочетание этих методик имеет свои подводные камни. Было разработано две вариации протокола Ехо-С (ЕхоС-19 и ЕхоС-20), обе этих вариации были использованы для приготовления библиотек клеточной линии K562 [Ramani et al., 2016; Ma et al., 2018; Gridina et al., 2021]. Критическим различием протоколов является использование дополнительных адаптеров в протоколе ЕхоС-19. Результаты секвенирования этих библиотек проверялись биоинформационными методами.

Базовыми параметрами качества библиотек были приняты:

- Доля дубликатов, отражающая качество стадии ПЦР;
- Доля участков, в которых покрытие прочтениями отсутствует, а также тех, в которых оно превышает минимальный порог для анализа (10 прочтений);
- Отношение среднего покрытия вне и внутри экзона, которое можно считать показателем качества таргетного обогащения.

Данные по качеству Ехо-С-библиотек представлены в Табл. 1.

Таблица 1: Данные по обогащению Ехо-С-библиотек

| Название | Глубина, прочтений | Дубликатов, % | Доля экзона с глубиной покрытия более 10, % | Среднее покрытие в экзоне | Среднее покрытие вне экзона | Обогащение экзона, раз | Доля непокрытых регионов в экзоне, % | Доля непокрытых регионов вне экзона, % |
|----------|--------------------|---------------|---|---------------------------|-----------------------------|------------------------|--------------------------------------|--|
| ЕхоС-19  | 136 609 179        | 18,86         | 91,68                                       | 60,51                     | 5,56                        | 10,89                  | 1,75                                 | 28,12                                  |
| ЕхоС-20  | 109 486 529        | 15,00         | 72,58                                       | 14,88                     | 7,74                        | 1,92                   | 1,66                                 | 11,62                                  |



## 3.2. Автоматизация обработки данных секвенирования

При обработке данных секвенирования приходится сталкиваться с проблемами различного характера. Одними из ключевых являются проблемы использования ресурсов компьютера. Результаты секвенирования даже в сжатом виде занимают десятки и сотни гигабайт дискового пространства, и многие инструменты создают файлы с промежуточными результатами, которые занимают дисковое пространство, не неся никакой практической пользы для исследования. Кроме того, из-за вычислительной сложности обработка таких больших блоков данных может занимать дни, недели и даже месяцы работы вычислительного кластера.

Вторая, не менее важная группа проблем, связана с используемыми для обработки инструментами. Как было показано выше, стадий у обработки значительное количество, и не все стадии нужны при обработке конкретного блока данных секвенирования. Ручная настройка и контроль процесса отнимают значительное количество времени исследователя; таким образом, встаёт вопрос стандартизации и автоматизации процесса обработки данных секвенирования.

Существующие инструменты для обработки данных секвенирования были разработаны независимыми группами людей. Эти инструменты различаются по многим аспектам. Так как разработка каждого отдельного инструмента является сложным и трудоёмким процессом, целесообразно использовать их как есть, а несоответствия устранять с помощью специально разработанной надстройки. Таким образом, для нами был создан пайплайн, интегрирующий все стадии обработки данных секвенирования. Блок-схема пайплайна представлена на Рис. 3.

Решённые задачи:

- Отказоустойчивость: максимально устранены несоответствия форматов входных и выходных данных; процесс разделён на стадии, и в случае экстренного прерывания вычислений (программного или аппаратного) предусмотрен автоматический откат.

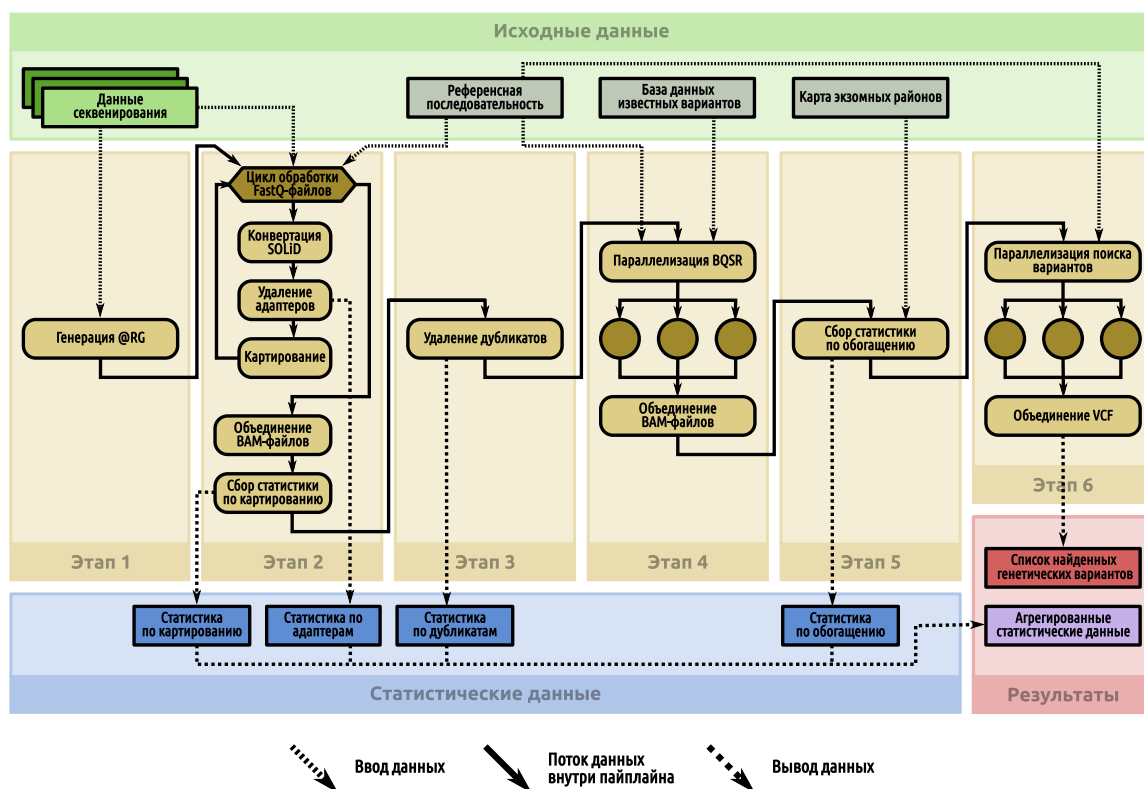


Рис. 3: Принципиальная схема пайплайна для обработки Eho-C-данных

- Оптимизация, параллелизация и масштабируемость: все процессы, которые способны использовать стандартные потоки ввода/вывода, объединены вместе, поддающиеся внешнему распараллеливанию были распараллелены, также были подобраны оптимальные параметры запуска приложений, использующих машину Java. Пайплайн может быть использован как на кластерах с большим количеством ядер и оперативной памяти, так и на относительно небольших мощностях офисных компьютеров;
- Значительно упрощены процессы развёртки и использования пайплайна: автоматизировано индексирование референсной последовательности, настройки вынесены в специальный конфигурационный файл, есть возможность обработки пула данных, используя один короткий сценарий;

Код пайплайна доступен на GitHub [[Valeev, 2020](#)].

### 3.3. Сравнение данных секвенирования клеточной линии K562

Следующим важным этапом работы было сравнение эффективности детекции генетических вариантов в данных, полученных на основе протокола Ехо-С и других протоколов (полногеномного и полноэкзомного секвенирования). Было решено использовать для этого распространённую иммортализованную клеточную линию K562, полученную от пациентки с хроническим миелолейкозом [Lozzio and Lozzio, 1975]. Данная клеточная линия была многократно секвенирована различными лабораториями с использованием различных методик приготовления библиотек. Таким образом, несмотря на то, что в этой клеточной линии наблюдается некоторая гетерогенность между лабораториями из-за большого количества пассажей, несмотря на наличие систематических ошибок при использовании разных методов секвенирования и приготовления библиотек, по K562 существует достаточное количество данных, чтобы использовать эту клеточную линию как стандарт для поиска генетических вариантов.

Результаты секвенирования клеточной линии K562 были взяты из публичных источников [Rao et al., 2014; Banaszak et al., 2018; Belaghzal et al., 2017; Ray et al., 2019; Zhou et al., 2019; Moquin et al., 2017; Dixon et al., 2018; Wang et al., 2020]. Использованные в этих статьях методики включают WGS, WES, Hi-C и Repli-seq. Из данных полноэкзомного секвенирования в дальнейшем были исключены все генетические варианты в интервале chr2:25,455,845-25,565,459 с фланкированием 1 kbp (ген *DNMT3A*), так как в одной из работ использовали генетически модифицированную линию с вариантами в данном гене [Banaszak et al., 2018]. В качестве тестовых Ехо-С-образцов мы использовали данные, полученные на основе клеточной линии K562, имеющейся в Институте Цитологии и Генетики СО РАН. Технические данные контроля качества по тестовым и контрольным образцам представлены в Приложении А.

В общей сложности, объединив варианты из всех контрольных образцов, мы получили 5 496 486 различных генетических вариантов. Также в библиотеках было найдено некоторое количество уникальных генетических ва-

риантов, встречающихся в одной библиотеке и не встречающихся в остальных (Табл. 2). Наибольший процент уникальных вариантов найден в данных Banaszak et al.

Таблица 2: Уникальные генетические варианты в данных секвенирования контрольных образцов клеточной линии K562

| Название         | Протокол  | Глубина секвенирования, прочтений | Общее число вариантов | Уникальные варианты | Доля уникальных вариантов, % |
|------------------|-----------|-----------------------------------|-----------------------|---------------------|------------------------------|
| Banaszak et al.  | WES       | 254 983 225                       | 408 008               | 41 830              | 10,25                        |
| Belaghzal et al. | Hi-C      | 72 914 268                        | 1 399 457             | 27 365              | 1,95                         |
| Dixon et al.     | WGS       | 366 291 496                       | 4 649 012             | 327 184             | 7,03                         |
| Moquin et al.    | Hi-C      | 256 500 659                       | 2 365 361             | 67 678              | 2,86                         |
| Rao et al.       | Hi-C      | 1 366 228 845                     | 4 218 233             | 320 508             | 7,59                         |
| Ray et al.       | Hi-C      | 428 306 794                       | 1 789 324             | 89 624              | 5,00                         |
| Wang et al.      | Repli-seq | 301 663 640                       | 2 207 451             | 37 578              | 1,70                         |
| Zhou et al.      | WGS       | 2 621 311 293                     | 4 412 455             | 166 451             | 3,77                         |

75 328 генетических вариантов были найдены в данных из всех восьми статей — их было решено использовать как «золотой стандарт». Сразу можно внимание на то, что это составляет лишь 1,37 % геномных SNV клеток K562. Такая ситуация может возникнуть в следующих случаях:

1. В одной или нескольких работах обнаружено очень много уникальных вариантов, которые дают существенный вклад в общее число вариантов, но не пересекаются с результатами других исследований;
2. В одной или нескольких работах не найдено подавляющее большинство вариантов, найденных во всех остальных работах;
3. Распределение уникальных вариантов и число общих вариантов между парами работ относительно равномерно, и низкое число общих для всех восьми работ вариантов не может объясняться особенностями какого-то одного или нескольких исследований.

Чтобы проверить, не связана ли низкая доля общих генетических вариантов с особенностями какого-то одного из использованных наборов данных, мы протестировали все комбинации из семи и шести работ. Результаты представлены на Рис. 4.

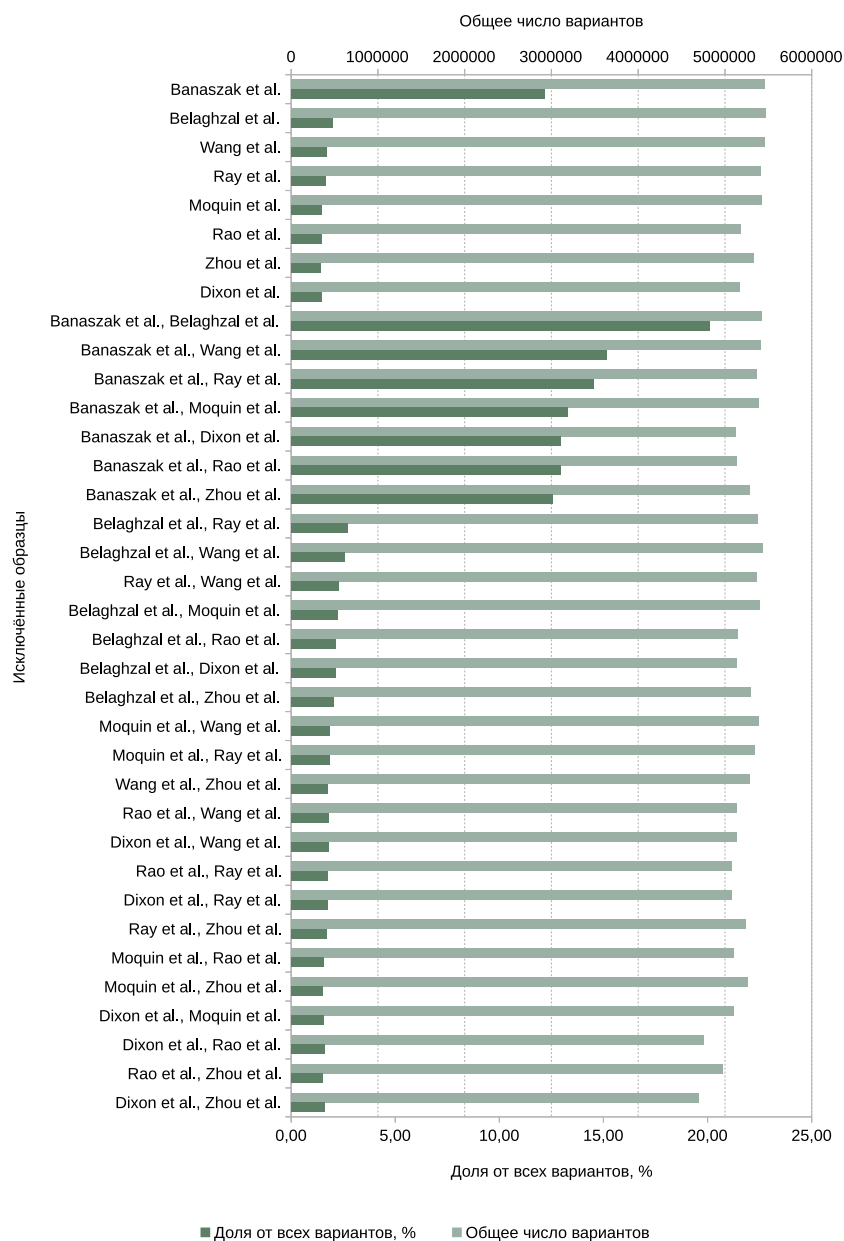


Рис. 4: Исключение отдельных образцов из контрольной выборки позволило увеличить количество генетических вариантов, которые можно использовать как стандарт. На рисунке показано суммарное количество вариантов в выборке (светло-зелёный) и процент общих для этой выборки вариантов (тёмно-зелёный) для выборок размером в 6 и 7 образцов. Слева указаны названия исключённых из выборки образцов.

Наибольшее число общих вариантов наблюдается при исключении из выборки данных [Banaszak et al.](#) и [Belaghzal et al.](#) — общими являются 1 091 331 (19,85 %) вариантов. Их решено было использовать как добавочный («серебряный») стандарт.

Далее генетические варианты контрольных образцов были проанализированы — выявлено соотношение инделов и SNP, количество альтернативных аллелей, а также зиготность. Результаты представлены в Табл. 3 и Табл. 4.

Таблица 3: Генетические варианты контрольных образцов по типу и количеству альтернативных аллелей

| Параметр                                  | Тип варианта |           | Количество альтернативных аллелей |       |      |      |
|---|--------------|-----------|-----------------------------------|-------|------|------|
|   | Indel        | SNP       | 1                                 | 2     | 3    | 4    |
| Вариантов «золотого стандарта»            | 3 546        | 71 782    | 75 112                            | 213   | 3    | 0    |
| Вариантов «серебряного стандарта»         | 51 656       | 1 039 675 | 1 088 792                         | 2 494 | 42   | 3    |
| Доля вариантов «золотого стандарта», %    | 4,71         | 95,29     | 99,71                             | 0,28  | 0,00 | 0,00 |
| Доля вариантов «серебряного стандарта», % | 4,73         | 95,27     | 99,77                             | 0,23  | 0,00 | 0,00 |

Таблица 4: Зиготность вариантов контрольных образцов с одним альтернативным аллелем. Отобраны только варианты, гомо- или гетерозиготность которых поддерживаются всеми библиотеками в выборке.

| Контрольная выборка  | GoldIndels | GoldSNP | SilverIndels | SilverSNP |
|--|------------|---------|--------------|-----------|
| Гетерозиготных вариантов   | 309        | 8 609   | 7 711        | 206 031   |
| Доля гетерозиготных вариантов от всех вариантов контрольной выборки, % | 0,41       | 11,43   | 0,71         | 18,88     |
| Гомозиготных вариантов   | 2 483      | 51 117  | 32 906       | 691 149   |
| Доля гомозиготных вариантов от всех вариантов контрольной выборки, %   | 3,3        | 67,86   | 3,02         | 63,33     |
| Общее число отобранных вариантов                                       | 2 792      | 59 726  | 40 617       | 897 180   |
| Доля отобранных вариантов от контрольной выборки, %                    | 3,71       | 79,29   | 3,72         | 82,21     |

Исходя из полученных данных, было решено использовать только генетические варианты с одним альтернативным аллелем, т.к. они составляют абсолютное большинство (99,7 %) вариантов, а из этих вариантов были отобраны только те, гетеро- или гомозиготность которых поддерживается 100 % биб-

лиотек в выборке. Таких вариантов в «золотом стандарте» — 62 518 (82,99 %), в «серебряном стандарте» — 937 797 (85,93 %).

Далее мы использовали указанную выше выборку вариантов «серебряного» и «золотого» стандартов для того, чтобы определить точность поиска генетических вариантов в наших Ехо-С-библиотеках. Для этого мы оценили количество генетических вариантов, являющихся общими для стандартных выборок и Ехо-С-библиотек, их долю от общего числа вариантов в Ехо-С-библиотеках. Подробные результаты показаны в Табл. 5.

Таблица 5: Валидация контрольных вариантов в Ехо-С-библиотеках. Проанализировано наличие позиции, наличие искомого варианта в позиции, а также совпадение зиготности.

| Параметр  | ExoC-19    |         |              |           | ExoC-20    |         |              |           |
|---|------------|---------|--------------|-----------|------------|---------|--------------|-----------|
|   | GoldIndels | GoldSNP | SilverIndels | SilverSNP | GoldIndels | GoldSNP | SilverIndels | SilverSNP |
| Общее число позиций в стандарте   | 2 792      | 59 726  | 40 617       | 897 180   | 2 792      | 59 726  | 40 617       | 897 180   |
| Найденных в образце позиций   | 2 701      | 57 948  | 36 778       | 813 125   | 2 163      | 50 219  | 20 606       | 518 831   |
| Найденных в образце позиций, % от общего числа позиций в стандарте            | 96,74      | 97,02   | 90,55        | 90,63     | 77,47      | 84,08   | 50,73        | 57,83     |
| Полное совпадение   | 2 600      | 56 656  | 35 014       | 786 430   | 2 067      | 48 615  | 18 950       | 480 497   |
| Полное совпадение, % от найденных позиций                                     | 96,26      | 97,77   | 95,20        | 96,72     | 95,56      | 96,81   | 91,96        | 92,61     |
| Несовпадение зиготности   | 93         | 1 257   | 1 607        | 26 357    | 85         | 1 552   | 1 547        | 37 719    |
| Несовпадение зиготности, % от найденных позиций                               | 3,44       | 2,17    | 4,37         | 3,24      | 3,93       | 3,09    | 7,51         | 7,27      |
| Несколько альтернативных аллелей при наличии искомого                         | 7          | 34      | 141          | 306       | 9          | 43      | 66           | 487       |
| Несколько альтернативных аллелей при наличии искомого, % от найденных позиций | 0,26       | 0,06    | 0,38         | 0,04      | 0,42       | 0,09    | 0,32         | 0,09      |
| Отсутствие искомого аллеля при наличии позиции                                | 1          | 1       | 16           | 32        | 2          | 9       | 43           | 128       |
| Отсутствие искомого аллеля при наличии позиции, % от найденных позиций        | 0,04       | 0,00    | 0,04         | 0,00      | 0,09       | 0,02    | 0,21         | 0,02      |

Также мы проанализировали количество и долю ложноположительных (отсутствующих в контрольных образцах) генетических вариантов. Стандартный метод определения чувствительности и специфичности (ROC-кривая) оказался неприменим к нашему случаю: доля ложноположительных вариантов вычисляется относительно суммы ложноположительных и истинно отрицательных, в контексте использованного нами метода количество истинно отрицательных вариантов с трудом поддаётся оценке. Поэтому доля ложноположительных результатов вычислялась относительно числа всех вариантов, найденных в Ехо-С-библиотеке, а доля истинно положительных — относительно размера контрольной выборки генетических вариантов. Варианты, которые присутствуют в нескольких контрольных образцах, но не являются

частью «серебряного стандарта» или «золотого стандарта», не рассматривались.

Также было решено проверить эффективность использованного нами базового фильтра — по глубине альтернативного аллеля. Мы проанализировали число истинно положительных и ложноположительных результатов при фильтрации вариантов с разной глубиной альтернативного аллеля. Результаты анализа показаны на Рис. 5.

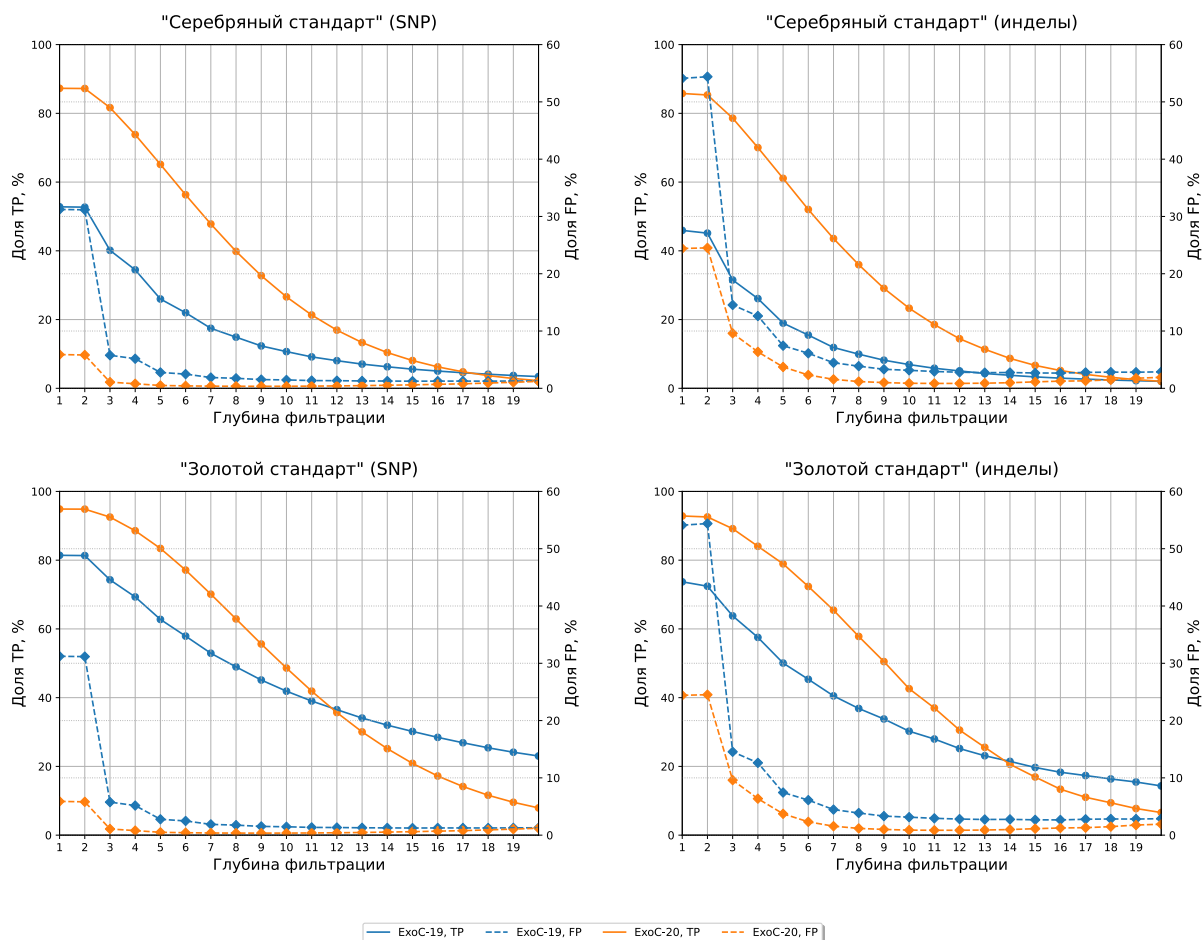


Рис. 5: Чувствительность и специфичность при фильтрации вариантов с разной глубиной альтернативного аллеля. Доля ложноположительных результатов вычислялась относительно числа всех вариантов, найденных в ExoC-библиотеке, доля истинно положительных — относительно размера контрольной выборки вариантов. Варианты, которые присутствуют в нескольких контрольных образцах, но не являются частью «серебряного стандарта» или «золотого стандарта», не рассматривались.



Как видно из полученных данных, фильтрация по глубине аллеля в 3 прочтения/позицию отсекает 80–90 % ложноположительных результатов. Потери истинно положительных вариантов при такой фильтрации невелики — 1 % и 7 % для EhoC-20 и EhoC-19 соответственно.

Поиск вариантов «серебряного» и «золотого» стандартов в наших библиотеках был произведён до и после фильтрации. Результаты фильтрации показаны в Приложении В.

### 3.4. Клинические случаи

Убедившись, что метод Eho-C подходит для использования в клинической практике, мы решили применить его для поиска генетических вариантов у пациентов с врождёнными патологиями. Для анализа мы взяли образцы пациентов, предоставленные врачами-генетиками крупных медицинских центров России (Приложение D). Из образцов 10 пациентов были приготовлены библиотеки Eho-C. Также в качестве контроля для биоинформационного пайплайна ещё для 11 пациентов мы подготовили полноэкзомные библиотеки. Спектр предполагаемых патологий включал психоневрологические синдромальные нарушения, иммунодефициты, лихорадку неясного генеза, ихтиоз и нейрофиброматоз I типа.

При анализе пациентов Eho-C клинически значимые варианты, соответствующие критериям поиска, были выявлены в 4 случаях из 10 (из них 2 патогенных и 2 варианта неопределённой клинической значимости), в одном случае мой коллега идентифицировал по данным Eho-C хромосомную перестройку. При анализе экзомных данных клинически значимые варианты, соответствующие критериям поиска, были выявлены в 5 случаях из 11, из них 1 патогенный и 4 неопределённой клинической значимости. Результаты представлены в Табл. 6. Подробное клиническое описание пациентов представлено в Приложении С.

Таблица 6: Список пациентов

| Код | Медицинское учреждение         | Пол | Дата рождения | Предварительный диагноз   | Родственные связи     | Тип библиотекы | Найденные варианты                  | Ген     | Нуклеотидная замена | Замена АК    | Этиотипность | Клиническая значимость | Заключение   | Подтверждение по Сигсеру |
|-----|--------------------------------|-----|---------------|---|-----------------------|----------------|-------------------------------------|---------|---------------------|--------------|--------------|------------------------|--|--------------------------|
| P1  | Томский НИМЦ <sup>1</sup>      | Ж   | 13.07.2018    | [E84.9] Муковисцидоз неутонченный   | —                     | Exo-C          | Не обнаружены                       | —       | —                   | —            | —            | —                      | —  | —                        |
| P2  | Томский НИМЦ <sup>1</sup>      | М   | 07.12.2012    | [F83] Задерка психо-речевого развития   | сбс: P4               | Exo-C          | Миссенс                             | ZMYND11 | c.570G>T            | p.Arg190Ser  | het          | Неопределенная         | [616083] Умственная отсталость, АД, 3                              | Подтверждено             |
| P3  | Томский НИМЦ <sup>1</sup>      | Ж   | 27.04.2013    | [F83] Задерка психо-речевого развития   | —                     | Exo-C          | Не обнаружены                       | —       | —                   | —            | —            | —                      | —  | —                        |
| P4  | Томский НИМЦ <sup>1</sup>      | М   | 13.10.2014    | [F83] Задерка психо-речевого развития   | сбс: P2               | Exo-C          | Миссенс                             | ZMYND11 | c.570G>T            | p.Arg190Ser  | het          | Неопределенная         | [616083] Умственная отсталость, АД, 3                              | Подтверждено             |
| P5  | Томский НИМЦ <sup>1</sup>      | М   | 25.07.2014    | [F83] Задерка психо-речевого развития   | —                     | Exo-C          | Сдвиг рамки                         | CJB2    | c.354del            | p.Gly12fs    | hom?         | Патогенный             | [220290] Глухота, AP, тип 1A                                       | —                        |
| P6  | НИКИ Педиатрия <sup>2</sup>    | М   | 16.02.2007    | [Q85.0] Нейрофиброматоз I типа  | ребенок P7            | Exo-C          | Не обнаружены                       | —       | —                   | —            | —            | —                      | —  | —                        |
| P7  | НИКИ Педиатрия <sup>2</sup>    | Ж   | —             | [Q85.0] Нейрофиброматоз I типа  | мать P6               | Exo-C          | Не обнаружены                       | —       | —                   | —            | —            | —                      | —  | —                        |
| P8  | НИМЦ им. Богачева <sup>3</sup> | Ж   | 01.04.2004    | [D84.9] Первичный иммунодефицит неутонченный, осложненный иммуной нейтропенией и тромбоцитопенией | —                     | Exo-C          | Не обнаружены                       | —       | —                   | —            | —            | —                      | —  | —                        |
| P9  | НИМЦ им. Богачева <sup>3</sup> | Ж   | 19.02.2013    | [Q87.8] Синдром повреждения Неймегена   | —                     | Exo-C          | Сдвиг рамки                         | NBN     | c.657_661del        | p.Lys219fs   | hom          | Патогенный             | [251260] Синдром Неймегена   | —                        |
| P10 | МГНЦ <sup>4</sup>              | Ж   | —             | [Q88.9] Комплекс врожденных аномалий  | —                     | Exo-C          | Транслокация по Exo-C               | —       | —                   | —            | —            | —                      | —  | —                        |
| P13 | НИИЭСЗ <sup>5</sup>            | Ж   | 10.03.2002    | [D61.3] Идиопатическая апластическая анемия   | —                     | WES            | Рекомендована консультация онколога | —       | —                   | —            | —            | —                      | —  | —                        |
| P14 | НИИЭСЗ <sup>5</sup>            | М   | 1993          | [R50] Лиоградия неясного генеза, сезонного характера  | —                     | WES            | Слайсинг                            | NLRP12  | g.27998G>A          | —            | het          | Неопределенная         | [611762] Семейный холодный аутоагглютинирующий синдром 2, АД       | —                        |
| P15 | KIIO3Cnр <sup>6</sup>          | Ж   | 24.09.2016    | [R27.0] Атактический синдром  | —                     | WES            | Миссенс                             | SCN4A   | c.4532C>T           | p.Ser1511Leu | het          | Неопределенная         | [608390] Калий-зависимая миотония, [168300] Миотония фон Эппенбург | —                        |
| P16 | KIIO3Cnр <sup>6</sup>          | Ж   | 30.09.2018    | [C93.0] Органическое поражение ЦНС генетической этиологии   | —                     | WES            | Миссенс                             | TUBB4A  | c.1178C>T           | p.Ala393Asp  | het          | Неопределенная         | [612438] Гипомиелинизирующая лейкодистрофия 6, АД                  | —                        |
| P17 | KIIO3Cnр <sup>6</sup>          | М   | 28.05.2015    | [Q80.9] Ихтиоз неутонченный   | —                     | WES            | Нопенс                              | ALOXE3  | c.10916C>T          | p.Arg366Ter  | hom          | Патогенный             | [606545] Врожденный ихтиоз 3, AP                                   | —                        |
| P18 | KIIO3Cnр <sup>6</sup>          | Ж   | 15.10.1988    | [Q80.9] Ихтиоз неутонченный   | —                     | WES            | Не обнаружены                       | —       | —                   | —            | —            | —                      | —  | —                        |
| P19 | KIIO3Cnр <sup>6</sup>          | М   | 11.09.2019    | [Q04.9] Врожденная аномалия головного мозга, предположительно X-сцепленная                        | —                     | WES            | Слайсинг                            | PLP1    | g.17892A>G          | —            | hem          | Неопределенная         | [312080] Болезнь Полиуса—Мерибомера, XLR                           | —                        |
| P20 | KIIO3Cnр <sup>6</sup>          | Ж   | 02.01.1977    | [C98] Нейтронные нарушения нервной системы  | мать P21 и P22        | WES            | Не обнаружены                       | —       | —                   | —            | —            | —                      | —  | —                        |
| P21 | KIIO3Cnр <sup>6</sup>          | Ж   | —             | [C98] Нейтронные нарушения нервной системы  | ребенок P20, сбс: P22 | WES            | Не обнаружены                       | —       | —                   | —            | —            | —                      | —  | —                        |
| P22 | KIIO3Cnр <sup>6</sup>          | М   | —             | [C98] Нейтронные нарушения нервной системы  | ребенок P20, сбс: P21 | WES            | Не обнаружены                       | —       | —                   | —            | —            | —                      | —  | —                        |
| P23 | KIIO3Cnр <sup>6</sup>          | М   | 17.04.2017    | [Q04.8] Синдром Дежне—Уоера   | —                     | WES            | Не обнаружены                       | —       | —                   | —            | —            | —                      | —  | —                        |

a1–6 См. Приложение D.

# Глава 4. Обсуждение результатов

## 4.1. Контрольные образцы

«Золотой стандарт» является набором генетических вариантов, относящихся к экзомным регионам, так как одна из библиотек представляла собой результаты WES. Это подтверждается пересечением с таргетной панелью, использованной при приготовлении данной библиотеки. Вариантов в «золотом стандарте» было обнаружено 75 тыс., что соответствует оценкам среднего количества генетических вариантов в кодирующих регионах у человека — 100 тыс. [Supernat et al., 2018]. Общее число несоответствий с референсным геномом у среднего человека составляет 4.1–5 млн [The 1000 Genomes Project Consortium, 2015], что с учётом гетерогенности клеточной линии K562 перекликается с общим количеством найденных нами генетических вариантов (5,5 млн).

Как видно из представленных выше данных, образец Banaszak et al. содержит наибольшее число уникальных вариантов (10,25 %). Это может быть связано с тем, что это данные полноэкзомного секвенирования, с высоким покрытием в экзонах, где и были найдены уникальные варианты. В качестве дополнительной гипотезы можно предположить, что в этой работе использовались линии клеток, в значительной степени отличающиеся от классической линии K562.

Прослеживается ожидаемая положительная связь между глубиной секвенирования Hi-C-библиотек и количеством уникальных вариантов в них. В двух WGS-библиотеках подобной связи не наблюдается. Вероятнее всего, это также связано с отличиями использованных линий K562.

## 4.2. Оценка результатов секвенирования

### Ехо-С-библиотек

В Ехо-С-библиотеках глубина секвенирования составляет 136,6 млн прочтений ( $2,05 \cdot 10^{10}$  bp) и 109,4 млн прочтений ( $1,64 \cdot 10^{10}$  bp), а среднее покры-

тие в экзOME — 60,51 и 14,88 прочтений для EhoC-19 и EhoC-20 соответственно. Глубину покрытия более 10 прочтений имеют 91,68 % и 72,58 % экзОМА для EhoC-19 и EhoC-20 соответственно. Согласно [Sims et al. \[2014\]](#), для репрезентативных результатов экзОМного секвенирования необходима глубина секвенирования не менее чем в  $10^{10}$  бр, а для Hi-C — не менее чем 100 млн прочтений. Минимальным порогом глубины для возможности поиска генетических вариантов считается 10 прочтений, практически все гомозиготные SNV могут быть найдены при глубине в 15 прочтений, а гетерозиготные требуют глубину прочтений не менее 33. Приемлемая доля экзОМА с репрезентативным покрытием (более 10 прочтений) составляет 90 %. Недостаточным остаётся обогащение, оно составляет 10,89 и 1,92 при норме в 75 для 40 Mbp таргетной панели [[Roc, 2018](#)]. Тем не менее, можно утверждать, что EhoC-19 отвечает требованиям для поиска SNV, а EhoC-20, во-первых, пригодна к поиску только гомозиготных генетических вариантов, а во-вторых, имеет недостаточно хорошее покрытие в экзОме.

«Золотой стандарт» покрыт нашими библиотеками на 82,75 % и 96,52 %, «серебряный стандарт» — на 56,48 % и 90,06 % (библиотеки EhoC-19 и EhoC-20 соответственно). Различия объясняются протоколами приготовления: у библиотеки EhoC-20 выше глубина покрытия в экзОме, но в 6 раз ниже обогащение в экзОмных районах. Кроме того, в библиотеке EhoC-19 были использованы адаптерные последовательности, дающие большое количество шума.

Одним из базовых методов фильтрации генетических вариантов в нашем методе является фильтрация по глубине альтернативного аллеля. Сразу можно обратить внимание на следующее:

- В библиотеке EhoC-19 потеряна большая доля вариантов, чем в EhoC-20 — как относительно общего числа, так и относительно вариантов «золотого» и «серебряного» стандартов.
- Доля ложноположительных (отсутствующих в контрольных образцах) генетических вариантов снижается в 5 раз уже при глубине альтернативного аллеля в 3 прочтения/позицию. Это можно объяснить наличием слабо покрытых регионов, в которых даже одно перекрывающееся парное прочтение способно привести к появлению ложного SNV.

Таким образом, фильтрация по глубине является эффективным способом улучшения данных низкого качества.

### 4.3. Клинические случаи

В данных Ехо-С клинически значимые варианты были найдены примерно в половине случаев. В данных экзоного секвенирования показан сравнимый уровень выявляемости. К сожалению, однозначных выводов относительно эффективности клинического применения метода сделать пока невозможно по следующим причинам:

- маленькая выборка (20 пациентов);
- широкий диапазон выявляемости клинически значимых вариантов в зависимости от нозологии. К примеру, выявляемость вариантов для первичных иммунодефицитов колеблется от 15 % до 70 % (среднее 38 % — [Vorsteveld et al., 2021](#)). Расстройства аутистического спектра выявляются лишь в 9 % случаев [[Du et al., 2018](#)];
- Общее описание фенотипа без предоставления подробных данных осмотра и лабораторных исследований — из-за политики конфиденциальности медучреждения либо по причине потери контакта с пациентом.

# Выводы

Таким образом, из приведённых нами данных можно сделать следующие выводы:

1. Конвейер биоинформационных утилит, созданный нами с учётом актуальных рекомендаций для биоинформационной обработки, позволяет обрабатывать данные Ехо-С-секвенирования, а также находить в этих данных SNV.
2. Сравнение генетических вариантов, полученных из контрольных образцов и Ехо-С-библиотек, позволяет утверждать, что метод Ехо-С способен детектировать около 75–90 % SNV, обнаруживаемых другими методами.
3. Применение метода Ехо-С для диагностики к когорты из 10 пациентов с различными наследственными заболеваниями позволило обнаружить 2 патогенных варианта и 2 варианта неопределенной клинической значимости. Таким образом, метод Ехо-С является эффективным инструментом геномной диагностики экзонных мутаций.

# Список литературы

1. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping / Rao Suhas S.P., Huntley Miriam H., Durand Neva C., Stamenova Elena K., Bochkov Ivan D., Robinson James T., Sanborn Adrian L., Machol Ido, Omer Arina D., Lander Eric S., and Aiden Erez Lieberman // *Cell*. — 2014. — Dec. — Vol. 159, no. 7. — P. 1665–1680. — Access mode: <https://doi.org/10.1016/j.cell.2014.11.021>.
2. Abnormal RNA splicing and genomic instability after induction of DNMT3A mutations by CRISPR/Cas9 gene editing / Banaszak Lauren G., Giudice Valentina, Zhao Xin, Wu Zhijie, Gao Shouguo, Hosokawa Kohei, Keyvanfar Keyvan, Townsley Danielle M., Gutierrez-Rodrigues Fernanda, del Pilar Fernandez Ibanez Maria, Kajigaya Sachiko, and Young Neal S. // *Blood Cells, Molecules, and Diseases*. — 2018. — Mar. — Vol. 69. — P. 10–22. — Access mode: <https://doi.org/10.1016/j.bcmd.2017.12.002>.
3. Abramowicz Anna, Gos Monika. Splicing mutations in human genetic disorders: examples, detection, and confirmation // *J Appl Genetics*. — 2018. — Apr. — Vol. 59, no. 3. — P. 253–268. — Access mode: <https://doi.org/10.1007/s13353-018-0444-7>.
4. Acuna-Hidalgo Rocio, Veltman Joris A., Hoischen Alexander. New insights into the generation and role of *de novo* mutations in health and disease // *Genome Biol*. — 2016. — Nov. — Vol. 17, no. 1. — Access mode: <https://doi.org/10.1186/s13059-016-1110-1>.
5. Amemiya Haley M., Kundaje Anshul, Boyle Alan P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome // *Sci Rep*. — 2019. — June. — Vol. 9, no. 1. — Access mode: <https://doi.org/10.1038/s41598-019-45839-z>.
6. Analysis of protein-coding genetic variation in 60,706 humans / Lek Monkol, , Karczewski Konrad J., Minikel Eric V., Samocha Kaitlin E., Banks Eric, Fennell Timothy, O'Donnell-Luria Anne H., Ware James S., Hill Andrew J.,

Cummings Beryl B., Tukiainen Taru, Birnbaum Daniel P., Kosmicki Jack A., Duncan Laramie E., Estrada Karol, Zhao Fengmei, Zou James, Pierce-Hoffman Emma, Berghout Joanne, Cooper David N., Deflaux Nicole, DePristo Mark, Do Ron, Flannick Jason, Fromer Menachem, Gauthier Laura, Goldstein Jackie, Gupta Namrata, Howrigan Daniel, Kiezun Adam, Kurki Mitja I., Moonshine Ami Levy, Natarajan Pradeep, Orozco Lorena, Peloso Gina M., Poplin Ryan, Rivas Manuel A., Ruano-Rubio Valentin, Rose Samuel A., Ruderfer Douglas M., Shakir Khalid, Stenson Peter D., Stevens Christine, Thomas Brett P., Tiao Grace, Tusie-Luna Maria T., Weisburd Ben, Won Hong-Hee, Yu Dongmei, Altshuler David M., Ardissino Diego, Boehnke Michael, Danesh John, Donnelly Stacey, Elosua Roberto, Florez Jose C., Gabriel Stacey B., Getz Gad, Glatt Stephen J., Hultman Christina M., Kathiresan Sekar, Laakso Markku, McCarroll Steven, McCarthy Mark I., McGovern Dermot, McPherson Ruth, Neale Benjamin M., Palotie Aarno, Purcell Shaun M., Saleheen Danish, Scharf Jeremiah M., Sklar Pamela, Sullivan Patrick F., Tuomilehto Jaakko, Tsuang Ming T., Watkins Hugh C., Wilson James G., Daly Mark J., and MacArthur Daniel G. // [Nature](#). — 2016. — Aug. — Vol. 536, no. 7616. — P. 285–291. — Access mode: <https://doi.org/10.1038/nature19057>.

7. Andrews Simon. FastQC: A Quality Control Tool for High Throughput Sequence Data. — Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. — 2010. — Accessed 2020/12/08.
8. Belaghzal Houda, Dekker Job, Gibcus Johan H. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation // [Methods](#). — 2017. — July. — Vol. 123. — P. 56–65. — Access mode: <https://doi.org/10.1016/j.ymeth.2017.04.004>.
9. Broad Institute. Picard Toolkit. — GitHub repository: <http://broadinstitute.github.io/picard/>. — 2019. — Accessed 2021/01/10.
10. A block-sorting lossless data compression algorithm : Rep. / Palo Alto, CA: Digital Equipment Corporation ; executor: Burrows M., Wheeler D.J. : 1994.



11. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery / Scott Eric M, Halees Anason, Itan Yuval, Spencer Emily G, He Yupeng, Azab Mostafa Abdellateef, Gabriel Stacey B, Belkadi Aziz, Boisson Bertrand, Abel Laurent, Clark Andrew G, Alkuraya Fowzan S, Casanova Jean-Laurent, and Gleeson Joseph G // [Nat Genet.](#) — 2016. — July. — Vol. 48, no. 9. — P. 1071–1076. — Access mode: <https://doi.org/10.1038/ng.3592>.
12. Choice of transcripts and software has a large effect on variant annotation / McCarthy Davis J, Humburg Peter, Kanapin Alexander, Rivas Manuel A, Gaulton Kyle, Cazier Jean-Baptiste, and and Peter Donnelly // [Genome Medicine.](#) — 2014. — Vol. 6, no. 3. — P. 26. — Access mode: <https://doi.org/10.1186/gm543>.
13. Chromatin conformation remains stable upon extensive transcriptional changes driven by heat shock / Ray Judhajeet, Munn Paul R., Vihervaara Anniina, Lewis James J., Ozer Abdullah, Danko Charles G., and Lis John T. // [Proc Natl Acad Sci USA.](#) — 2019. — Sep. — Vol. 116, no. 39. — P. 19431–19439. — Access mode: <https://doi.org/10.1073/pnas.1901244116>.
14. ClinVar: improving access to variant interpretations and supporting evidence / Landrum Melissa J, Lee Jennifer M, Benson Mark, Brown Garth R, Chao Chen, Chitipiralla Shanmuga, Gu Baoshan, Hart Jennifer, Hoffman Douglas, Jang Wonhee, Karapetyan Karen, Katz Kenneth, Liu Chunlei, Maddipatla Zenith, Malheiro Adriana, McDaniel Kurt, Ovetsky Michael, Riley George, Zhou George, Holmes J Bradley, Kattman Brandi L, and Maglott Donna R // [Nucleic Acids Research.](#) — 2017. — Nov. — Vol. 46, no. D1. — P. D1062–D1067. — Access mode: <https://doi.org/10.1093/nar/gkx1153>.
15. Comparison of three variant callers for human whole genome sequencing / Supernat Anna, Vidarsson Oskar Valdimar, Steen Vidar M., and Stokowy Tomasz // [Sci Rep.](#) — 2018. — Dec. — Vol. 8, no. 1. — Access mode: <https://doi.org/10.1038/s41598-018-36177-7>.
16. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome / Lieberman-Aiden E., van Berkum N. L.,

- Williams L., Imakaev M., Ragoczy T., Telling A., Amit I., Lajoie B. R., Sabo P. J., Dorschner M. O., Sandstrom R., Bernstein B., Bender M. A., Groudine M., Gnirke A., Stamatoyannopoulos J., Mirny L. A., Lander E. S., and Dekker J. // *Science*. — 2009. — ОКТ. — Vol. 326, no. 5950. — P. 289–293. — Access mode: <https://doi.org/10.1126/science.1181369>.
17. Comprehensive, integrated, and phased whole-genome analysis of the primary ENCODE cell line K562 / Zhou Bo, Ho Steve S., Greer Stephanie U., Zhu Xiaowei, Bell John M., Arthur Joseph G., Spies Noah, Zhang Xianglong, Byeon Seunggyu, Pattni Reenal, Ben-Efraim Noa, Haney Michael S., Haraksingh Rajini R., Song Giltae, Ji Hanlee P., Perrin Dimitri, Wong Wing H., Abyzov Alexej, and Urban Alexander E. // *Genome Res.* — 2019. — Feb. — Vol. 29, no. 3. — P. 472–484. — Access mode: <https://doi.org/10.1101/gr.234948.118>.
18. The Epstein–Barr Virus Episome Maneuvers between Nuclear Chromatin Compartments during Reactivation / Moquin Stephanie A., Thomas Sean, Whalen Sean, Warburton Alix, Fernandez Samantha G., McBride Alison A., Pollard Katherine S., and Miranda JJ L. // *J Virol.* — 2017. — Nov. — Vol. 92, no. 3. — Access mode: <https://doi.org/10.1128/jvi.01413-17>.
19. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches / Ebbert Mark T. W., Wadsworth Mark E., Staley Lyndsay A., Hoyt Kaitlyn L., Pickett Brandon, Miller Justin, Duce John, Kauwe John S. K., and Ridge Perry G. // *BMC Bioinformatics*. — 2016. — July. — Vol. 17, no. S7. — Access mode: <https://doi.org/10.1186/s12859-016-1097-3>.
20. Exomic variants of an elderly cohort of Brazilians in the ABraOM database / Naslavsky Michel Satya, Yamamoto Guilherme Lopes, de Almeida Tatiana Ferreira, Ezquina Suzana A. M., Sunaga Daniele Yumi, Pho Nam, Bozoklian Daniel, Sandberg Tatiana Orli Milkewitz, Brito Luciano Abreu, Lazar Monize, Bernardo Danilo Vicensotto, Amaro Edson, Duarte Yeda A. O., Lebrão Maria Lúcia, Passos-Bueno Maria Rita, and

- Zatz Mayana // *Human Mutation*. — 2017. — May. — Vol. 38, no. 7. — P. 751–763. — Access mode: <https://doi.org/10.1002/humu.23220>.
21. French National Institute of Health and Medical Research (INSERM). Orphanet: an online database of rare diseases and orphan drugs. — Available online at: <http://www.orpha.net>. — 1997. — Accessed 2020/12/08.
  22. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline / Auwera Geraldine A., Carneiro Mauricio O., Hartl Christopher, Poplin Ryan, del Angel Guillermo, Levy-Moonshine Ami, Jordan Tadeusz, Shakir Khalid, Roazen David, Thibault Joel, Banks Eric, Garimella Kiran V., Altshuler David, Gabriel Stacey, and DePristo Mark A. // *Current Protocols in Bioinformatics*. — 2013. — Okt. — Vol. 43, no. 1. — Access mode: <https://doi.org/10.1002/0471250953.bi1110s43>.
  23. Galanello Renzo, Origa Raffaella. Beta-thalassemia // *Orphanet J Rare Dis*. — 2010. — May. — Vol. 5, no. 1. — Access mode: <https://doi.org/10.1186/1750-1172-5-11>.
  24. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses / Stelzer Gil, Rosen Naomi, Plaschkes Inbar, Zimmerman Shahar, Twik Michal, Fishilevich Simon, Stein Tsippi Iny, Nudel Ron, Lieder Iris, Mazor Yaron, Kaplan Sergey, Dahary Dvir, Warshawsky David, Guan-Golan Yaron, Kohn Asher, Rappaport Noa, Safran Marilyn, and Lancet Doron // *Current Protocols in Bioinformatics*. — 2016. — June. — Vol. 54, no. 1. — Access mode: <https://doi.org/10.1002/cpbi.5>.
  25. Genetic Diagnostic Evaluation of Trio-Based Whole Exome Sequencing Among Children With Diagnosed or Suspected Autism Spectrum Disorder / Du Xiujuan, Gao Xueren, Liu Xin, Shen Lixiao, Wang Kai, Fan Yanjie, Sun Yu, Luo Xiaomei, Liu Huili, Wang Lili, Wang Yu, Gong Zhuwen, Wang Jianguo, Yu Yongguo, and Li Fei // *Front. Genet*. — 2018. — Nov. — Vol. 9. — Access mode: <https://doi.org/10.3389/fgene.2018.00594>.
  26. Gonzalez E. The Influence of CCL3L1 Gene-Containing Segmental Duplications on HIV-1/AIDS Susceptibility // *Science*. — 2005. —

- Mar. — Vol. 307, no. 5714. — P. 1434–1440. — Access mode: <https://doi.org/10.1126/science.1101160>.
27. Goodwin Sara, McPherson John D., McCombie W. Richard. Coming of age: ten years of next-generation sequencing technologies // *Nat Rev Genet*. — 2016. — May. — Vol. 17, no. 6. — P. 333–351. — Access mode: <https://doi.org/10.1038/nrg.2016.49>.
28. Gresham David, Dunham Maitreya J., Botstein David. Comparing whole genomes using DNA microarrays // *Nat Rev Genet*. — 2008. — Apr. — Vol. 9, no. 4. — P. 291–302. — Access mode: <https://doi.org/10.1038/nrg2335>.
29. Guidelines for the interpretation of massive parallel sequencing variants / Ryzhkova O.P., Kardymon O.L., Prohorchuk E.B., Konovalov F.A., Maslennikov A.B., Stepanov V.A., Afanasyev A.A., Zaklyazminskaya E.V., Kostareva A.A., Pavlov A.E., Golubenko M.V., Polyakov A.V., and Kutsev S.I. // *Medical Genetics*. — 2017. — Vol. 16, no. 7. — P. 4–17. — Access mode: <https://www.medgen-journal.ru/jour/article/view/308/224>.
30. Handy Diane E., Castro Rita, Loscalzo Joseph. Epigenetic Modifications // *Circulation*. — 2011. — May. — Vol. 123, no. 19. — P. 2145–2156. — Access mode: <https://doi.org/10.1161/circulationaha.110.956839>.
31. Herder Matthew. What Is the Purpose of the Orphan Drug Act? // *PLoS Med*. — 2017. — Jan. — Vol. 14, no. 1. — P. e1002191. — Access mode: <https://doi.org/10.1371/journal.pmed.1002191>.
32. Hi-C Identifies Complex Genomic Rearrangements and TAD-Shuffling in Developmental Diseases / Melo Uirá Souto, Schöpflin Robert, Acuna-Hidalgo Rocio, Mensah Martin Atta, Fischer-Zirnsak Björn, Holtgrewe Manuel, Klever Marius-Konstantin, Türkmen Seval, Heinrich Verena, Pluym Ilina Datkhaeva, Matoso Eunice, de Sousa Sérgio Bernardo, Louro Pedro, Hülsemann Wiebke, Cohen Monika, Dufke Andreas, Latos-Bieleńska Anna, Vingron Martin, Kalscheuer Vera, Quintero-Rivera Fabiola, Spielmann Malte, and Mundlos Stefan // *The American Journal of Human*

- Genetics*. — 2020. — June. — Vol. 106, no. 6. — P. 872–884. — Access mode: <https://doi.org/10.1016/j.ajhg.2020.04.016>.
33. Hiroi N, Agatsuma S. Genetic susceptibility to substance dependence // *Mol Psychiatry*. — 2004. — Dec. — Vol. 10, no. 4. — P. 336–344. — Access mode: <https://doi.org/10.1038/sj.mp.4001622>.
34. Huber D., von Voithenberg L. Voith, Kaigala G.V. Fluorescence *in situ* hybridization (FISH): History, limitations and what to expect from micro-scale FISH? // *Micro and Nano Engineering*. — 2018. — Nov. — Vol. 1. — P. 15–24. — Access mode: <https://doi.org/10.1016/j.mne.2018.10.006>.
35. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity / Saleheen Danish, Natarajan Pradeep, Armean Irina M., Zhao Wei, Rasheed Asif, Khetarpal Sumeet A., Won Hong-Hee, Karczewski Konrad J., O'Donnell-Luria Anne H., Samocha Kaitlin E., Weisburd Benjamin, Gupta Namrata, Zaidi Mozzam, Samuel Maria, Imran Atif, Abbas Shahid, Majeed Faisal, Ishaq Madiha, Akhtar Saba, Trindade Kevin, Mucksavage Megan, Qamar Nadeem, Zaman Khan Shah, Yaqoob Zia, Saghir Tahir, Rizvi Syed Nadeem Hasan, Memon Anis, Mallick Nadeem Hayyat, Ishaq Mohammad, Rasheed Syed Zahed, ur Rehman Memon Fazal, Mahmood Khalid, Ahmed Naveeduddin, Do Ron, Krauss Ronald M., MacArthur Daniel G., Gabriel Stacey, Lander Eric S., Daly Mark J., Frossard Philippe, Danesh John, Rader Daniel J., and Kathiresan Sekar // *Nature*. — 2017. — Apr. — Vol. 544, no. 7649. — P. 235–239. — Access mode: <https://doi.org/10.1038/nature22034>.
36. Impact of Chromosomal Translocations on Male Infertility, Semen Quality, Testicular Volume and Reproductive Hormone Levels / Dong Y, Du R-C, Jiang Y-T, Wu J, Li L-L, and Liu R-Z // *J Int Med Res*. — 2012. — Dec. — Vol. 40, no. 6. — P. 2274–2283. — Access mode: <https://doi.org/10.1177/030006051204000625>.
37. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls / Zook Justin M, Chapman Brad, Wang Jason, Mittelman David, Hofmann Oliver, Hide Winston, and Salit Marc // *Nat*

Biotechnol. — 2014. — Feb. — Vol. 32, no. 3. — P. 246–251. — Access mode: <https://doi.org/10.1038/nbt.2835>.

38. Integrative detection and analysis of structural variation in cancer genomes / Dixon Jesse R., Xu Jie, Dileep Vishnu, Zhan Ye, Song Fan, Le Victoria T., Yardımcı Galip Gürkan, Chakraborty Abhijit, Bann Darrin V., Wang Yanli, Clark Royden, Zhang Lijun, Yang Hongbo, Liu Tingting, Iyyanki Sriranga, An Lin, Pool Christopher, Sasaki Takayo, Rivera-Mulia Juan Carlos, Ozadam Hakan, Lajoie Bryan R., Kaul Rajinder, Buckley Michael, Lee Kristen, Diegel Morgan, Pezic Dubravka, Ernst Christina, Hadjur Suzana, Odom Duncan T., Stamatoyannopoulos John A., Broach James R., Hardison Ross C., Ay Ferhat, Noble William Stafford, Dekker Job, Gilbert David M., and Yue Feng // *Nat Genet.* — 2018. — Sep. — Vol. 50, no.10. — P. 1388–1398. — Access mode: <https://doi.org/10.1038/s41588-018-0195-8>.
39. The State of Variant Annotation: A Comparison of AnnoVar, snpEff and VEP : Rep. / The Golden Helix Blog (GHB) ; executor: Jesaitis Andrew. — <https://blog.goldenhelix.com/the-sate-of-variant-annotation-a-comparison-of-annovar-snpeff-and-vep/> : 2014. — June.
40. Jett Kimberly, Friedman Jan M. Clinical and genetic aspects of neurofibromatosis 1 // *Genet Med.* — 2009. — Dec. — Vol. 12, no. 1. — P. 1–11. — Access mode: <https://doi.org/10.1097/gim.0b013e3181bf15e3>.
41. Jian Xueqiu, Boerwinkle Eric, Liu Xiaoming. *In silico* tools for splicing defect prediction: a survey from the viewpoint of end users // *Genet Med.* — 2013. — Nov. — Vol. 16, no. 7. — P. 497–503. — Access mode: <https://doi.org/10.1038/gim.2013.176>.
42. Jin Zelin, Liu Yun. DNA methylation in human diseases // *Genes & Diseases.* — 2018. — Mar. — Vol. 5, no. 1. — P. 1–8. — Access mode: <https://doi.org/10.1016/j.gendis.2018.01.002>.
43. Kozak Marilyn. An analysis of 5'-noncoding sequences from 699 vertebrate

- messenger RNAs // *Nucl Acids Res.* — 1987. — Vol. 15, no. 20. — P. 8125–8148. — Access mode: <https://doi.org/10.1093/nar/15.20.8125>.
44. Langmead Ben, Salzberg Steven L. Fast gapped-read alignment with Bowtie2 // *Nat Methods.* — 2012. — Mar. — Vol. 9, no. 4. — P. 357–359. — Access mode: <https://doi.org/10.1038/nmeth.1923>.
45. Li H., Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform // *Bioinformatics.* — 2009. — May. — Vol. 25, no. 14. — P. 1754–1760. — Access mode: <https://doi.org/10.1093/bioinformatics/btp324>.
46. Lozzio C. B., Lozzio B. B. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome // *Blood.* — 1975. — Mar. — Vol. 45, no. 3. — P. 321–334. — Access mode: <https://pubmed.ncbi.nlm.nih.gov/163658>.
47. Mapping 3D genome architecture through in situ DNase Hi-C / Ramani Vijay, Cusanovich Darren A, Hause Ronald J, Ma Wenxiu, Qiu Ruolan, Deng Xinxian, Blau C Anthony, Disteche Christine M, Noble William S, Shendure Jay, and Duan Zhijun // *Nat Protoc.* — 2016. — Sep. — Vol. 11, no. 11. — P. 2104–2121. — Access mode: <https://doi.org/10.1038/nprot.2016.126>.
48. Martin Marcel. Cutadapt removes adapter sequences from high-throughput sequencing reads // *EMBnet j.* — 2011. — May. — Vol. 17, no. 1. — P. 10. — Access mode: <https://doi.org/10.14806/ej.17.1.200>.
49. Mozheiko E. A., Fishman V. S. Detection of Point Mutations and Chromosomal Translocations Based on Massive Parallel Sequencing of Enriched 3C Libraries // *Russ J Genet.* — 2019. — Окт. — Vol. 55, no. 10. — P. 1273–1281. — Access mode: <https://doi.org/10.1134/s1022795419100089>.
50. Mutations in a novel gene encoding a CRAL-TRIO domain cause human Cayman ataxia and ataxia/dystonia in the jittery mouse / Bomar Jamee M, Benke Paul J, Slattery Eric L, Puttagunta Radhika, Taylor Larry P, Seong Eunju, Nystuen Arne, Chen Weidong, Albin Roger L, Patel Paresh D, Kittles Rick A, Sheffield Val C, and Burmeister Margit // *Nat Genet.* — 2003. — Окт. — Vol. 35, no. 3. — P. 264–269. — Access mode: <https://doi.org/10.1038/ng1255>.



51. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing / Mandelker Diana, Schmidt Ryan J., Ankala Arunkanth, Gibson Kristin McDonald, Bowser Mark, Sharma Himanshu, Duffy Elizabeth, Hegde Madhuri, Santani Avni, Lebo Matthew, and Funke Birgit // *Genet Med*. — 2016. — May. — Vol. 18, no. 12. — P. 1282–1289. — Access mode: <https://doi.org/10.1038/gim.2016.58>.
52. Noninvasive prenatal testing for fetal subchromosomal copy number variations and chromosomal aneuploidy by low-pass whole-genome sequencing / Yu Dongyi, Zhang Kai, Han Meiyan, Pan Wei, Chen Ying, Wang Yunfeng, Jiao Hongyan, Duan Ling, Zhu Qiying, Song Xiaojie, Hong Yan, Chen Chen, Wang Juan, Hui Feng, Huang Linzhou, Chen Chongjian, and Du Yang // *Mol Genet Genomic Med*. — 2019. — Apr. — Vol. 7, no. 6. — Access mode: <https://doi.org/10.1002/mgg3.674>.
53. O'Connor Clare. Human chromosome translocations and cancer // Nature Education. — 2008. — Vol. 1, no. 1. — P. 56. — Access mode: <https://www.nature.com/scitable/topicpage/human-chromosome-translocations-and-cancer-23487/>.
54. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders / Amberger Joanna S., Bocchini Carol A., Schiettecatte François, Scott Alan F., and Hamosh Ada // *Nucleic Acids Research*. — 2014. — Nov. — Vol. 43, no. D1. — P. D789–D798. — Access mode: <https://doi.org/10.1093/nar/gku1205>.
55. Oluwadare Oluwatosin, Highsmith Max, Cheng Jianlin. An Overview of Methods for Reconstructing 3D Chromosome and Genome Structures from Hi-C Data // *Biol Proced Online*. — 2019. — Apr. — Vol. 21, no. 1. — Access mode: <https://doi.org/10.1186/s12575-019-0094-0>.
56. Quinlan Aaron R., Hall Ira M. BEDTools: a flexible suite of utilities for comparing genomic features // *Bioinformatics*. — 2010. — Jan. — Vol. 26, no. 6. — P. 841–842. — Access mode: <https://doi.org/10.1093/bioinformatics/btq033>.



57. Rare Disease Terminology and Definitions—A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group / Richter Trevor, Nestler-Parr Sandra, Babela Robert, Khan Zeba M., Tesoro Theresa, Molsen Elizabeth, and Hughes Dyfrig A. // *Value in Health*. — 2015. — Sep. — Vol. 18, no. 6. — P. 906–914. — Access mode: <https://doi.org/10.1016/j.jval.2015.05.008>.
58. Recommendations for performance optimizations when using GATK3.8 and GATK4 / Heldenbrand Jacob R, Baheti Saurabh, Bockol Matthew A, Drucker Travis M, Hart Steven N, Hudson Matthew E, Iyer Ravishankar K, Kalmbach Michael T, Kendig Katherine I, Klee Eric W, Mattson Nathan R, Wieben Eric D, Wierpert Mathieu, Wildman Derek E, and Mainzer Liudmila S // *BMC Bioinformatics*. — 2019. — Nov. — Vol. 20, no. 1. — Access mode: <https://doi.org/10.1186/s12859-019-3169-7>.
59. RegSNPs-intron: a computational framework for predicting pathogenic impact of intronic single nucleotide variants / Lin Hai, Hargreaves Katherine A., Li Rudong, Reiter Jill L., Wang Yue, Mort Matthew, Cooper David N., Zhou Yaoqi, Zhang Chi, Eadon Michael T., Dolan M. Eileen, Ipe Joseph, Skaar Todd C., and Liu Yunlong // *Genome Biol*. — 2019. — Nov. — Vol. 20, no. 1. — Access mode: <https://doi.org/10.1186/s13059-019-1847-4>.
60. Rizvi Ali A. 46, XX Man with SRY Gene Translocation: Cytogenetic Characteristics, Clinical Features and Management // *The American Journal of the Medical Sciences*. — 2008. — Apr. — Vol. 335, no. 4. — P. 307–309. — Access mode: <https://doi.org/10.1097/maj.0b013e31811ec1b4>.
61. SPIN reveals genome-wide landscape of nuclear compartmentalization / Wang Yuchuan, Zhang Yang, Zhang Ruochi, van Schaik Tom, Zhang Liguang, Sasaki Takayo, Hupkes Daniel Peric, Chen Yu, Gilbert David M., van Steensel Bas, Belmont Andrew S., and Ma Jian. — 2020. — Mar. — Access mode: <https://doi.org/10.1101/2020.03.09.982967>.
62. Sampson B., McGuire A. *Genetics and the Molecular Autopsy* // Pathobiology of Human Disease. — Elsevier, 2014. — P. 3459–3467. — Access mode: <https://doi.org/10.1016/b978-0-12-386456-7.06707-1>.

63. Sanger F., Nicklen S., Coulson A. R. DNA sequencing with chain-terminating inhibitors // [Proceedings of the National Academy of Sciences](#). — 1977. — Dec. — Vol. 74, no. 12. — P. 5463–5467. — Access mode: <https://doi.org/10.1073/pnas.74.12.5463>.
64. Schreck R. R., Distèche C. M. Chromosome banding techniques // [Current protocols in human genetics](#). — 2001. — May. — Vol. Chapter 4. — P. Unit4.2–Unit4.2. — 18428280[pmid]. Access mode: <https://pubmed.ncbi.nlm.nih.gov/18428280>.
65. Roche. — SeqCap EZ HyperCap Workflow User's Guide v2.3 : 2018. — June. — Available online at: <https://www.n-genetics.com/products/1295/1023/13869.pdf>.
66. Sequencing depth and coverage: key considerations in genomic analyses / Sims David, Sudbery Ian, Illott Nicholas E., Heger Andreas, and Ponting Chris P. // [Nat Rev Genet](#). — 2014. — Jan. — Vol. 15, no. 2. — P. 121–132. — Access mode: <https://doi.org/10.1038/nrg3642>.
67. Sherry S. T. dbSNP: the NCBI database of genetic variation // [Nucleic Acids Research](#). — 2001. — Jan. — Vol. 29, no. 1. — P. 308–311. — Access mode: <https://doi.org/10.1093/nar/29.1.308>.
68. Spectral karyotyping: an unique technique for the detection of complex genomic rearrangements in leukemia / Guo Bo, Han Xiaoping, Wu Zhanhe, Da Wanming, and Zhu Hongli // [Translational pediatrics](#). — 2014. — Apr. — Vol. 3, no. 2. — P. 135–139. — 26835331[pmid]. Access mode: <https://pubmed.ncbi.nlm.nih.gov/26835331>.
69. Spielmann Malte, Lupiáñez Darío G., Mundlos Stefan. Structural variation in the 3D genome // [Nat Rev Genet](#). — 2018. — Apr. — Vol. 19, no. 7. — P. 453–467. — Access mode: <https://doi.org/10.1038/s41576-018-0007-0>.
70. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology /

- Richards Sue, , Aziz Nazneen, Bale Sherri, Bick David, Das Soma, Gastier-Foster Julie, Grody Wayne W., Hegde Madhuri, Lyon Elaine, Spector Elaine, Voelkerding Karl, and Rehm Heidi L. // *Genet Med*. — 2015. — Mar. — Vol. 17, no. 5. — P. 405–423. — Access mode: <https://doi.org/10.1038/gim.2015.30>.
71. Strong Association of *de novo* Copy Number Mutations with Autism / Sebat J., Lakshmi B., Malhotra D., Troge J., Lese-Martin C., Walsh T., Yamrom B., Yoon S., Krasnitz A., Kendall J., Leotta A., Pai D., Zhang R., Lee Y.-H., Hicks J., Spence S. J., Lee A. T., Puura K., Lehtimäki T., Ledbetter D., Gregersen P. K., Agman, Sutcliffe J. S., Jobanputra V., Chung W., Warburton D., King M.-C., Skuse D., Geschwind D. H., Gilliam T. C., Ye K., and Wigler M. // *Science*. — 2007. — Apr. — Vol. 316, no. 5823. — P. 445–449. — Access mode: <https://doi.org/10.1126/science.1138659>.
72. Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage / Barbitoff Yury A., Polev Dmitrii E., Glotov Andrey S., Serebryakova Elena A., Shcherbakova Irina V., Kiselev Artem M., Kostareva Anna A., Glotov Oleg S., and Predeus Alexander V. // *Sci Rep*. — 2020. — Feb. — Vol. 10, no. 1. — Access mode: <https://doi.org/10.1038/s41598-020-59026-y>.
73. Tan Adrian, Abecasis Gonçalo R., Kang Hyun Min. Unified representation of genetic variants // *Bioinformatics*. — 2015. — Feb. — Vol. 31, no. 13. — P. 2202–2204. — Access mode: <https://doi.org/10.1093/bioinformatics/btv112>.
74. Tang B, Dean B, Thomas E A. Disease- and age-related changes in histone acetylation at gene promoters in psychiatric disorders // *Transl Psychiatry*. — 2011. — Dec. — Vol. 1, no. 12. — P. e64–e64. — Access mode: <https://doi.org/10.1038/tp.2011.61>.
75. Teer J. K., Mullikin J. C. Exome sequencing: the sweet spot before whole genomes // *Human Molecular Genetics*. — 2010. — Aug. — Vol. 19, no. R2. — P. R145–R151. — Access mode: <https://doi.org/10.1093/hmg/ddq333>.
76. The 1000 Genomes Project Consortium. A global reference for human genetic

- variation // [Nature](#). — 2015. — Sep. — Vol. 526, no. 7571. — P. 68–74. — Access mode:
77. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies / Stenson Peter D., Mort Matthew, Ball Edward V., Evans Katy, Hayden Matthew, Heywood Sally, Hussain Michelle, Phillips Andrew D., and Cooper David N. // [Hum Genet](#). — 2017. — Mar. — Vol. 136, no. 6. — P. 665–677. — Access mode: <https://doi.org/10.1007/s00439-017-1779-6>.
78. The Lancet Neurology. Rare neurological diseases: a united approach is needed // [The Lancet Neurology](#). — 2011. — Feb. — Vol. 10, no. 2. — P. 109. — Access mode: [https://doi.org/10.1016/S1474-4422\(11\)70001-1](https://doi.org/10.1016/S1474-4422(11)70001-1).
79. The Sequence Alignment/Map format and SAMtools / Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., and and R. Durbin // [Bioinformatics](#). — 2009. — June. — Vol. 25, no. 16. — P. 2078–2079. — Access mode: <https://doi.org/10.1093/bioinformatics/btp352>.
80. Theisen Aaron. Microarray-based Comparative Genomic Hybridization (aCGH) // [Nature Education](#). — 2008. — Vol. 1, no. 1. — P. 45. — Access mode: <https://www.nature.com/scitable/topicpage/microarray-based-comparative-genomic-hybridization-acgh-45432/>.
81. An Update of *in silico* Tools for the Prediction of Pathogenesis in Missense Variants / Brea-Fernandez Alejandro, Ferro Marta, Fernandez-Rozadilla Ceres, Blanco Ana, Fachal Laura, Santamarina Marta, Vega Ana, Pazos Alejandro, Carracedo Angel, and Ruiz-Ponte Clara // [CBIO](#). — 2011. — June. — Vol. 6, no. 2. — P. 185–198. — Access mode: <https://doi.org/10.2174/1574893611106020185>.
82. Use of the MLPA Assay in the Molecular Diagnosis of Gene Copy Number Alterations in Human Genetic Diseases / Stuppia Liborio, Antonucci Ivana, Palka Giandomenico, and Gatta Valentina // [IJMS](#). — 2012. — Mar. — Vol. 13, no. 3. — P. 3245–3276. — Access mode: <https://doi.org/10.3390/ijms13033245>.

83. Using DNase Hi-C techniques to map global and local three-dimensional genome architecture at high resolution / Ma Wenxiu, Ay Ferhat, Lee Choli, Gulsoy Gunhan, Deng Xinxian, Cook Savannah, Hesson Jennifer, Cavanaugh Christopher, Ware Carol B., Krumm Anton, Shendure Jay, Blau C. Anthony, Disteche Christine M., Noble William S., and Duan ZhiJun // [Methods](#). — 2018. — June. — Vol. 142. — P. 59–73. — Access mode: <https://doi.org/10.1016/j.ymeth.2018.01.014>.
84. Valeev Emil. Scissors: Exo-C Variants Search Pipeline. — GitHub repository: <https://github.com/regnveig/Scissors>. — 2020.
85. Vorsteveld Emil E., Hoischen Alexander, van der Made Caspar I. Next-Generation Sequencing in the Field of Primary Immunodeficiencies: Current Yield, Challenges, and Future Perspectives // [Clinic Rev Allerg Immunol](#). — 2021. — Mar. — Access mode: <https://doi.org/10.1007/s12016-021-08838-5>.
86. Wang K., Li M., Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data // [Nucleic Acids Research](#). — 2010. — July. — Vol. 38, no. 16. — P. e164–e164. — Access mode: <https://doi.org/10.1093/nar/gkq603>.
87. [When Children Die: Improving Palliative and End-of-Life Care for Children and Their Families](#) / ed. by Field Marilyn J., Behrman Richard E. — Washington (DC) : National Academies Press (US), 2003. — ISBN: 0-309-08437-7. — NBK220818[bookaccession]. Access mode: <https://pubmed.ncbi.nlm.nih.gov/25057608>.
88. Winklhofer Konstanze F, Tatzelt Jörg, Haass Christian. The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases // [EMBO J](#). — 2008. — Jan. — Vol. 27, no. 2. — P. 336–349. — Access mode: <https://doi.org/10.1038/sj.emboj.7601930>.
89. Yohe Sophia, Thyagarajan Bharat. Review of Clinical Next-Generation Sequencing // [Archives of Pathology & Laboratory Medicine](#). — 2017. — Aug. — Vol. 141, no. 11. — P. 1544–1557. — Access mode: <https://doi.org/10.5858/arpa.2016-0501-ra>.

90. Young Sara K., Wek Ronald C. Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response // *Journal of Biological Chemistry*. — 2016. — Aug. — Vol. 291, no. 33. — P. 16927–16935. — Access mode: <https://doi.org/10.1074/jbc.r116.733899>.
91. A cookbook for DNase Hi-C / Gridina Maria, Mozheiko Evgeniy, Valeev Emil, Nazarenko Ludmila P., Lopatkina Maria E., Markova Zhanna G., Yablonskaya Maria I., Voinova Viktoria Yu, Shilova Nadezhda V., Lebedev Igor N., and Fishman Veniamin // *Epigenetics & Chromatin*. — 2021. — Mar. — Vol. 14, no. 1. — Access mode: <https://doi.org/10.1186/s13072-021-00389-5>.
92. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs / Liu Xiaoming, Wu Chunlei, Li Chang, and Boerwinkle Eric // *Human Mutation*. — 2016. — Jan. — Vol. 37, no. 3. — P. 235–241. — Access mode: <https://doi.org/10.1002/humu.22932>.
93. The influence of BRCA variants of unknown significance on cancer risk management decision-making / Chern Jing-Yi, Lee Sarah S., Frey Melissa K., Lee Jessica, and Blank Stephanie V. // *J Gynecol Oncol*. — 2019. — Vol. 30, no. 4. — Access mode: <https://doi.org/10.3802/jgo.2019.30.e60>.
94. The mutational constraint spectrum quantified from variation in 141,456 humans / Karczewski Konrad J., , Francioli Laurent C., Tiao Grace, Cummings Beryl B., Alföldi Jessica, Wang Qingbo, Collins Ryan L., Laricchia Kristen M., Ganna Andrea, Birnbaum Daniel P., Gauthier Laura D., Brand Harrison, Solomonson Matthew, Watts Nicholas A., Rhodes Daniel, Singer-Berk Moriel, England Eleina M., Seaby Eleanor G., Kosmicki Jack A., Walters Raymond K., Tashman Katherine, Farjoun Yossi, Banks Eric, Poterba Timothy, Wang Arcturus, Seed Cotton, Whiffin Nicola, Chong Jessica X., Samocha Kaitlin E., Pierce-Hoffman Emma, Zappala Zachary, O'Donnell-Luria Anne H., Minikel Eric Vallabh, Weisburd Ben, Lek Monkol, Ware James S., Vittal Christopher, Armean Irina M., Bergelson Louis, Cibulskis Kristian, Connolly Kristen M., Covarrubias Miguel, Donnelly Stacey,

Ferriera Steven, Gabriel Stacey, Gentry Jeff, Gupta Namrata, Jeandet Thibault, Kaplan Diane, Llanwarne Christopher, Munshi Ruchi, Novod Sam, Petrillo Nikelle, Roazen David, Ruano-Rubio Valentin, Saltzman Andrea, Schleicher Molly, Soto Jose, Tibbetts Kathleen, Tolonen Charlotte, Wade Gordon, Talkowski Michael E., Neale Benjamin M., Daly Mark J., and MacArthur Daniel G. // *Nature*. — 2020. — May. — Vol. 581, no. 7809. — P. 434–443. — Access mode: <https://doi.org/10.1038/s41586-020-2308-7>.

95. A snapshot of some pLI score pitfalls / Ziegler Alban, Colin Estelle, Goudenège David, and Bonneau Dominique // *Human Mutation*. — 2019. — Apr. — Access mode: <https://doi.org/10.1002/humu.23763>.

# **Приложение А. Технические данные по секвенированию клеточной линии K562**

Полные технические данные по секвенированию клеточной линии K562 представлены в Табл. 7 (по библиотекам) и Табл. 8 (по образцам).



Таблица 7: Библиотеки данных секвенирования клеточной линии K562

| Библиотека         | Статьи           | Репозиторий | Коды доступа  | Тип данных       | Тип протекций | Глубина, протекций | Общее число протекций | Доля картированных, % от общего числа | Доля добавленных, % от общего числа | Картированные протекция | Картированные синглтоны | Дубликаты PE протекций | Дубликаты синглтонов | Доля дубликатов, % | Оценки размера библиотеки |
|--------------------|------------------|-------------|---|------------------|---------------|--------------------|-----------------------|---------------------------------------|-------------------------------------|-------------------------|-------------------------|------------------------|----------------------|--------------------|---------------------------|
| Контрольные данные |                  |             |   |                  |               |                    |                       |                                       |                                     |                         |                         |                        |                      |                    |                           |
| GSM1551618_HIC069  | Rao et al.       | GEO         | SRRI1658693   | Hi-C             | PE            | 4567577799         | 1 001 169 248         | 96,57                                 | 8,755                               | 424 945 100             | 29 290 805              | 17 648 021             | 13 182 626           | 5,56               | 4916 114 832              |
| GSM1551619_HIC070  | Rao et al.       | GEO         | SRRI1658694   | Hi-C             | PE            | 591 854 453        | 1 314 487 595         | 98,7                                  | 9,949                               | 575 565 379             | 15 452 072              | 98 778 796             | 8 811 532            | 17,69              | 1 478 944 337             |
| GSM1551620_HIC071  | Rao et al.       | GEO         | SRRI1658695   | Hi-C             | PE            | 79 905 895         | 173 931 529           | 98,81                                 | 8,118                               | 77 880 938              | 1 975 600               | 486 893                | 269 138              | 0,79               | 6 202 732 721             |
| GSM1551621_HIC072  | Rao et al.       | GEO         | SRRI1658697   | Hi-C             | PE            | 79 578 049         | 159 160 116           | 98,38                                 | 0,003                               | 77 155 821              | 2 265 995               | 366 805                | 285 395              | 0,65               | 8 088 955 029             |
| GSM1551622_HIC073  | Rao et al.       | GEO         | SRRI1658698   | Hi-C             | PE            | 77 353 816         | 154 710 364           | 98,33                                 | 0,002                               | 74 866 287              | 2 383 970               | 240 304                | 293 115              | 0,51               | 11 637 260 975            |
| GSM1551623_HIC074  | Rao et al.       | GEO         | SRRI1658702   | Hi-C             | PE            | 80 778 733         | 175 291 763           | 98,65                                 | 7,835                               | 78 467 294              | 2 254 814               | 644 986                | 321 965              | 1,01               | 4 746 870 162             |
| ENC50282GPG        | Zhou et al.      | ENCODE      | ENCFF52Y1LG<br>ENCFF921AXL<br>ENCFF598SXX                               | WGS              | SE            | 258 022 356        | 260 044 021           | 85,39                                 | 0,777                               | —                       | 220 029 156             | —                      | 50 680 083           | 23,04              | —                         |
| ENC50053AXS        | Zhou et al.      | ENCODE      | ENCFF004THU<br>ENCFF066GQD<br>ENCFF313MGL<br>ENCFF506TKC<br>ENCFF080MQF | WGS              | SE            | 1 472 492 722      | 1 592 540 515         | 91,19                                 | 7,538                               | —                       | 1 334 175 586           | —                      | 496 237 198          | 37,25              | —                         |
| ENC5071UUNY        | Zhou et al.      | ENCODE      | ENCFF471WSA<br>ENCFF068SYX<br>ENCFF598SXX                               | WGS              | SE            | 890 796 215        | 899 473 769           | 99,72                                 | 0,965                               | —                       | 889 239 055             | —                      | 203 498 352          | 22,91              | —                         |
| SRX3358201         | Dixon et al.     | GEO         | SRRG251264  | WGS              | PE            | 366 291 496        | 737 534 099           | 99,72                                 | 0,671                               | 364 794 328             | 923 254                 | 73 018 048             | 406 066              | 20,05              | 785 091 005               |
| GSE148362_G1       | Wang et al.      | GEO         | SRRI1518301   | Repli-seq        | SE            | 24 804 095         | 24 804 986            | 96,39                                 | 0,001                               | —                       | 23 908 072              | —                      | 921 353              | 3,85               | —                         |
| GSE148362_G2       | Wang et al.      | GEO         | SRRI1518308   | Repli-seq        | SE            | 32 032 314         | 32 033 010            | 97,61                                 | 0,002                               | —                       | 32 241 907              | —                      | 3 881 591            | 12,04              | —                         |
| GSE148362_S1       | Wang et al.      | GEO         | SRRI1518302   | Repli-seq        | SE            | 30 884 788         | 30 885 298            | 98,7                                  | 0,002                               | —                       | 30 481 936              | —                      | 2 156 480            | 7,07               | —                         |
| GSE148362_S2       | Wang et al.      | GEO         | SRRI1518303   | Repli-seq        | SE            | 45 359 773         | 45 360 305            | 98,39                                 | 0,002                               | —                       | 44 630 884              | —                      | 1 939 846            | 4,35               | —                         |
| GSE148362_S3       | Wang et al.      | GEO         | SRRI1518304   | Repli-seq        | SE            | 49 807 076         | 49 807 988            | 98,79                                 | 0,002                               | —                       | 49 205 535              | —                      | 2 889 464            | 5,87               | —                         |
| GSE148362_S4       | Wang et al.      | GEO         | SRRI1518305   | Repli-seq        | SE            | 44 149 029         | 44 149 770            | 98,46                                 | 0,002                               | —                       | 43 469 002              | —                      | 2 678 091            | 6,16               | —                         |
| GSE148362_S5       | Wang et al.      | GEO         | SRRI1518306   | Repli-seq        | SE            | 38 424 060         | 38 424 835            | 97,96                                 | 0,002                               | —                       | 37 640 056              | —                      | 3 600 260            | 9,57               | —                         |
| GSE148362_S6       | Wang et al.      | GEO         | SRRI1518307   | Repli-seq        | SE            | 35 203 065         | 35 203 676            | 97,51                                 | 0,002                               | —                       | 34 324 742              | —                      | 4 177 438            | 12,17              | —                         |
| INSITU_HIS1        | Ray et al.       | GEO         | SRR0919504  | Hi-C             | PE            | 86 294 895         | 172 589 790           | 93,3                                  | 0                                   | 75 521 119              | 9 982 274               | 1 841 061              | 1 615 286            | 3,29               | 1 523 677 153             |
| INSITU_HIS2        | Ray et al.       | GEO         | SRR0919505  | Hi-C             | PE            | 127 093 919        | 254 187 638           | 93,36                                 | 0                                   | 111 730 240             | 13 858 195              | 1 923 146              | 3 048 273            | 2,91               | 3 208 280 367             |
| INSITU_NHS1        | Ray et al.       | GEO         | SRR0919506  | Hi-C             | PE            | 86 445 594         | 172 891 188           | 93,43                                 | 0                                   | 75 893 138              | 9 737 847               | 1 903 981              | 1 649 376            | 3,38               | 1 487 154 386             |
| INSITU_NHS2        | Ray et al.       | GEO         | SRR0919507  | Hi-C             | PE            | 128 472 386        | 256 944 772           | 93,27                                 | 0                                   | 112 615 319             | 14 417 076              | 1 961 996              | 3 196 535            | 2,97               | 3 194 317 678             |
| PDDE_TRANSIENT     | Moquin et al.    | GEO         | SRRS470541<br>SRRS470540  | Hi-C             | PE            | 55 158 049         | 110 319 638           | 95,6                                  | 0,003                               | 51 158 920              | 3 140 556               | 3 917 308              | 721 938              | 8,11               | 316 780 447               |
| PD_STABLE_REP1     | Moquin et al.    | GEO         | SRRS470535<br>SRRS470534  | Hi-C             | PE            | 67 172 619         | 134 347 099           | 97,58                                 | 0,001                               | 64 767 511              | 1 565 427               | 5 573 966              | 376 260              | 8,79               | 354 373 851               |
| PD_STABLE_REP2     | Moquin et al.    | GEO         | SRRS470537  | Hi-C             | PE            | 52 872 167         | 105 745 908           | 98,23                                 | 0,001                               | 51 442 087              | 993 483                 | 2 058 449              | 217 598              | 4,17               | 625 522 723               |
| PD_TRANSIENT       | Moquin et al.    | GEO         | SRBS470639<br>SRBS470638  | Hi-C             | PE            | 81 297 824         | 162 600 928           | 95,28                                 | 0,003                               | 75 141 163              | 4 639 787               | 7 298 377              | 1 339 404            | 10,29              | 361 336 652               |
| GSM238881E_R1      | Bedigheal et al. | GEO         | SRBS479813  | Hi-C             | PE            | 72 914 268         | 172 533 452           | 99,39                                 | 15,478                              | 72 067 575              | 648 294                 | 9 684 590              | 210 273              | 13,54              | 243 264 112               |
| GSM236769_WT       | Banszak et al.   | GEO         | SRBS345331  | WES <sup>1</sup> | PE            | 39 211 003         | 78 464 649            | 99,46                                 | 0,054                               | 38 914 993              | 171 253                 | 7 621 960              | 91 145               | 20,17              | 83 342 746                |
| GSM236770_WT_IP    | Banszak et al.   | GEO         | SRBS345332  | WES <sup>1</sup> | PE            | 49 394 206         | 98 820 633            | 99,54                                 | 0,033                               | 49 068 605              | 193 565                 | 10 478 814             | 114 795              | 21,43              | 97 869 629                |
| GSM236771_MT2      | Banszak et al.   | GEO         | SRBS345333  | WES <sup>1</sup> | PE            | 42 020 936         | 84 093 776            | 99,63                                 | 0,062                               | 41 772 436              | 189 177                 | 8 755 216              | 104 927              | 21,04              | 85 177 236                |
| GSM236772_MT3      | Banszak et al.   | GEO         | SRBS345334  | WES <sup>1</sup> | PE            | 43 669 913         | 87 375 385            | 99,6                                  | 0,041                               | 43 414 109              | 164 448                 | 9 489 133              | 93 601               | 21,92              | 84 242 110                |
| GSM236773_MT4      | Banszak et al.   | GEO         | SRBS345335  | WES <sup>1</sup> | PE            | 39 879 463         | 79 788 947            | 99,53                                 | 0,038                               | 39 609 943              | 166 651                 | 7 989 943              | 90 809               | 21,76              | 77 577 055                |
| GSM236774_MT5      | Banszak et al.   | GEO         | SRBS345336  | WES <sup>1</sup> | PE            | 40 807 984         | 81 649 292            | 99,59                                 | 0,041                               | 40 559 969              | 163 957                 | 8 801 283              | 91 545               | 21,77              | 79 383 290                |

Тестовые данные

|                 |         |   |   |         |    |             |             |       |        |             |         |            |         |       |             |
|-----------------|---------|---|---|---------|----|-------------|-------------|-------|--------|-------------|---------|------------|---------|-------|-------------|
| FG_ExcCbe4001   | ExcC-19 | — | — | ExcC-19 | PE | 136 609 179 | 339 215 777 | 99,31 | 23,150 | 135 150 334 | 443 409 | 25 453 568 | 159 152 | 18,86 | 319 784 450 |
| FG_Quarantine-A | ExcC-20 | — | — | ExcC-20 | PE | 53 598 130  | 140 214 460 | 99,79 | 23,150 | 53 598 130  | 259 561 | 7 809 282  | 68 779  | 14,60 | 193 853 459 |
| FG_Quarantine-B | ExcC-20 | — | — | ExcC-20 | PE | 55 279 173  | 144 641 130 | 99,76 | 23,108 | 55 279 173  | 310 369 | 8 808 307  | 90 489  | 15,97 | 177 375 163 |

<sup>a</sup>Варианты в гене DNMT3A были исключены из выборки.

Таблица 8: Образцы данных секвенирования клеточной линии K562

| Образец            | Тип данных | Тип прочтений | Глубина, прочтений | Общее число прочтений | Доля картированных, % от общего числа | Доля добавочных, % от общего числа | FR PE прочтения, % от картированных | Картированные PE прочтения | Картированные синглеты | Картированные на разные хромосомы пары, % от картированных | Картированные на разные хромосомы пары (QMAP 4+), % от картированных на разные хромосомы |
|--------------------|------------|---------------|--------------------|-----------------------|---------------------------------------|------------------------------------|-------------------------------------|----------------------------|------------------------|--|--|
| Контрольные данные |            |               |                    |                       |                                       |                                    |                                     |                            |                        |  |  |
| Rao et al.         | Hi-C       | PE            | 1 366 228 845      | 2 978 750 615         | 97,95                                 | 8,268                              | 27,04                               | 2 617 761 638              | 53 623 256             | 21,03  | 84,23  |
| Zhou et al.        | WGS        | SE            | 2 621 311 293      | 2 752 058 305         | 93,43                                 | 4,751                              | —                                   | —                          | —                      | —  | —  |
| Dixon et al.       | WGS        | PE            | 366 291 496        | 737 534 099           | 99,72                                 | 0,671                              | 97,16                               | 729 588 656                | 923 254                | 1,25   | 51,22  |
| Wang et al.        | Repli-seq  | SE            | 301 663 640        | 301 669 278           | 98,09                                 | 0,002                              | —                                   | —                          | —                      | —  | —  |
| Ray et al.         | Hi-C       | PE            | 428 306 794        | 856 613 588           | 93,33                                 | 0                                  | 35,92                               | 751 519 632                | 47 995 392             | 22,77  | 76,00  |
| Moquin et al.      | Hi-C       | PE            | 256 500 659        | 513 013 573           | 96,56                                 | 0,002                              | 46,64                               | 485 019 362                | 10 339 253             | 17,76  | 75,56  |
| Belaghzal et al.   | Hi-C       | PE            | 72 914 268         | 172 533 452           | 96,39                                 | 15,478                             | 24,77                               | 144 135 150                | 648 294                | 34,02  | 88,02  |
| Banaszak et al.    | WES        | PE            | 254 983 225        | 510 192 582           | 99,56                                 | 0,044                              | 99,41                               | 506 680 110                | 1 049 051              | 0,11   | 81,38  |
| Тестовые данные    |            |               |                    |                       |                                       |                                    |                                     |                            |                        |  |  |
| ExoC-19            | Exo-C      | PE            | 136 609 179        | 359 215 777           | 99,31                                 | 23,94                              | 89,22                               | 270 300 668                | 443 409                | 5,01   | 66,93  |
| ExoC-20            | Exo-C      | PE            | 109 486 529        | 284 855 590           | 99,77                                 | 23,13                              | 70,02                               | 217 754 606                | 569 930                | 5,87   | 78,00  |

# Приложение В. Результаты фильтрации генетических вариантов по глубине альтернативного аллеля

Таблица 9: Параметры Ехо-С-библиотек. (F–) — до фильтрации по глубине альтернативного аллеля, (F+) — после фильтрации, ( $\Delta$ ) — изменение параметра после фильтрации в процентах

| Параметр   | ЕхоС-19   |           |              | ЕхоС-20   |           |              | В обеих |         |              | Ни в одной |         |              |
|--|-----------|-----------|--------------|-----------|-----------|--------------|---------|---------|--------------|------------|---------|--------------|
|  | F–        | F+        | $\Delta$ , % | F–        | F+        | $\Delta$ , % | F–      | F+      | $\Delta$ , % | F–         | F+      | $\Delta$ , % |
| Общее число вариантов в библиотеке   | 3 173 343 | 1 396 525 | –55,99       | 3 750 319 | 2 577 934 | –31,26       | —       | —       | —            | —          | —       | —            |
| Вариантов «золотого стандарта»   | 62 335    | 52 732    | –15,41       | 72 705    | 67 270    | –7,48        | 60 728  | 48 840  | –19,58       | 1 016      | 4 166   | +310,04      |
| Доля вариантов «золотого стандарта», %   | 82,75     | 70,00     | —            | 96,52     | 89,30     | —            | 80,62   | 64,84   | —            | 1,35       | 5,53    | —            |
| Вариантов «серебряного стандарта»  | 616 375   | 391 273   | –36,52       | 982 858   | 821 991   | –16,37       | 580 351 | 340 833 | –41,27       | 72 449     | 218 900 | +202,14      |
| Доля вариантов «серебряного стандарта», %  | 56,48     | 35,85     | —            | 90,06     | 75,32     | —            | 53,18   | 31,23   | —            | 6,64       | 20,06   | —            |
| Количество вариантов библиотеки, отсутствующих в контрольных образцах            | 1 130 049 | 84 770    | –92,50       | 354 044   | 41 719    | –88,22       | 14 455  | 2 981   | –79,38       | —          | —       | —            |
| Доля вариантов, отсутствующих в контрольных образцах, от вариантов библиотеки, % | 35,61     | 6,07      | –82,95       | 9,44      | 1,62      | –82,86       | —       | —       | —            | —          | —       | —            |

# Приложение С. Клиническое описание пациентов

## Пациент Р1

**Пол:** Женский

**Дата рождения:** 13.07.2018 (2 года)

**Дата обследования:** Неизвестна (2020)

### Жалобы

Сыпь на лице.

### Анамнез заболевания

С рождения желтуха, однократное срыгивание кофейной гущей, периодически срыгивание.

### Анамнез жизни

**Неонатальный период:** Беременность 4, роды 2. Во время беременности мама трижды болела ОРЗ. Роды в срок, самостоятельно. Закричала сразу. Вес 2 940 г, длина тела 49 см. Из роддома ребенок переведен в реанимацию, DS: [P39.8] Неонатальная инфекция с поражением органов дыхания.

**Генеалогический анализ и наследственность:** По данным клинико-генеалогического исследования наследственной отягощенности не выявлено.

**Прочее:** Данные по развитию, оперативным вмешательствам, сопутствующим заболеваниям, эпидемиологический, трансфузионный и аллергологический анамнез отсутствуют.

## **Клиническое обследование**

Данные по клиническому обследованию отсутствуют.

## **Лабораторные исследования**

**УЗИ плода от 2018:** Мультикистозная почка и подозрение на отсутствие желчного пузыря.

**УЗИ от 07.2018:** Тазовая дистопия правой почки, каликоектазия левой почки, эхопризнаки дисхолии (желчный пузырь сокращен. содержимое неоднородное).

**ОАК от 07.2018:** Полицитемия.

**Неонатальный скрининг от 16.07.2018:** Иммунореактивный трипсин 111 нг/мл.

**Неонатальный скрининг от 08.08.2018:** иммунореактивный трипсин 59 нг/мл.

**Анализ Гибсона—Кука от 2020:** Хлориды пота 42 ммоль/л.

## **Диагноз**

[E84.9] Муковисцидоз неуточненный.

## **Дифференциальный диагноз**

Консультация генетика.

## ЗАКЛЮЧЕНИЕ

по результатам секвенирования экзона от 06.03.2020 г.

|                       |                                |
|-----------------------|--------------------------------|
| Код пациента:         | P1                             |
| Вид материала:        | Кровь (венозная)               |
| Дата забора:          | 2019                           |
| Экзомная панель:      | SeqCap EZ MedExome Kit (Roche) |
| Количество прочтений: | 85,2 млн                       |
| Длина прочтений:      | 150 bp                         |
| Выравнивание:         | — %                            |
| Покрывание экзона:    |                                |
| – $\geq 10$ :         | — %                            |
| – среднее:            | — прочтений/позицию            |

**Заключение:** Значимых изменений, соответствующих критериям поиска, в данных экзомного секвенирования при индивидуальном анализе не обнаружено.

## Пациент Р2

**Пол:** Мужской

**Дата рождения:** 07.12.2012 (8 лет)

**Дата обследования:** Неизвестна (2020)

### Жалобы

Отсутствие речи, агрессивное поведение.

### Анамнез заболевания

До 1,5 лет лет разговаривал отдельными словами («мама», «папа»). После рождения младшего брата перестал разговаривать, изменилось поведение, стал агрессивным. В настоящее время не разговаривает, бывает неадекватное поведение, отмечается задержка психического развития.

### Анамнез жизни

**Неонатальный период:** Беременность 1, роды 1. Во время беременности мама лежала на сохранении. Роды в срок. Родился с обвитием пуповины. Вес 3 710 г, длина тела 54 см. Закричал сразу. Желтуха в течение недели. Кривошея с рождения.

**Формула развития:** Моторное развитие: сел в 9 мес., пополз в 8 мес., ходить начал в 10 мес. Речевое развитие: первые слова в 1 год и 3 мес., фразовой речи нет. После рождения младшего брата ребенок перестал разговаривать.

**Генеалогический анализ и наследственность:** У родного брата (пациент Р4) те же проявления.

**Эпидемиологический анамнез:** В 5 мес. переболел ротавирусной инфекцией.

**Прочее:** Данные по оперативным вмешательствам, сопутствующим заболеваниям, трансфузионный и аллергологический анамнез отсутствуют.

### **Клиническое обследование**

Данные по клиническому обследованию пациента отсутствуют.

### **Лабораторные исследования**

Данные по лабораторным исследованиям отсутствуют.

### **Диагноз**

[F83] Задержка психо-речевого развития.

### **Дифференциальный диагноз**

Необходимо исключить в плане дифференциальной диагностики наследственные заболевания, связанные с хромосомными нарушениями и наследственными нарушениями обмена веществ.



## ЗАКЛЮЧЕНИЕ

**по результатам секвенирования экзона от 24.08.2020 г.**

**Код пациента:** P2

**Вид материала:** Кровь (венозная)

**Дата забора:** 2019

**Экзомная панель:** SeqCap EZ MedExome Kit (Roche)

**Количество прочтений:** 113,7 млн

**Длина прочтений:** 150 bp

**Выравнивание:** 99,46 %

**Покрывание экзона:**

- $\geq 10$ : 92,55 %
- среднее: 68 прочтений/позицию

| Ген  | Положение (hg19) | Замена  | Генотип | Тип замены | Глубина | Ref/Alt |
|--|------------------|---|---------|------------|---------|---------|
| ZMYND11  | chr10:283578     | G>T   | het     | nonsyn SNV | 10      | 2/8     |
| <b>ОМIM фенотип:</b>   |                  | [616083] Умственная отсталость, АД, 3   |         |            |         |         |
| <b>HGMD:</b>   |                  | —   |         |            |         |         |
| <b>Экзон:</b>  |                  | 6/15  |         |            |         |         |
| <b>Частота аллеля:</b>   |                  | [rs757019491] 0,000 03  |         |            |         |         |
| <b>Критерии патогенности (согласно российским рекомендациям по интерпретации данных MPS [29]):</b> |                  | <p>[PM1] Вариант расположен в горячей точке и/или важных и хорошо исследованных функциональных доменах белка (например, активный сайт фермента), в которых не описаны доброкачественные изменения (Bromodomain;Bromodomain Zinc finger, FYVE/PHD-type Zinc finger, PHD-finger Zinc finger, PHD-type Zinc finger, PHD-type, conserved site Zinc finger, RING-type Zinc finger, RING/FYVE/PHD-type). [PM2] Вариант отсутствует в контрольных выборках (или встречается с крайне низкой частотой). [PP3] Результаты не менее трёх программ предсказания патогенности <i>in silico</i> подтверждают патогенное действие варианта на ген или продукт гена.</p> |         |            |         |         |
| <b>Комментарий:</b>  |                  | —   |         |            |         |         |

**Заключение:** Выявлен описанный ранее у больных вариант нуклеотидной последовательности в 6 экзоне гена *ZMYND11* (chr10:283578G>T, rs757019491) в гетерозиготном состоянии, приводящий к замене аргинина на серин в 190 кодоне (NP\_006615.2:p.Arg190Ser). Гетерозиготные варианты в гене *ZMYND11* описаны для пациентов с умственной отсталостью, АД, З (OMIM: 616083). Выявленный вариант встречается с крайне низкой частотой 0,003 % в контрольных базах данных. Он расположен в важном функциональном домене белка. Результаты 5 программ предсказания патогенности *in silico* подтверждают патогенное действие варианта на ген или продукт гена. По совокупности сведений, выявленный вариант нуклеотидной последовательности следует считать вариантом неопределенного клинического значения, который может иметь отношение к фенотипу пациента в случае появления новых подтверждающих данных. Других значимых изменений, соответствующих критериям поиска, не обнаружено. Вариант подтвержден по Сэнгеру.

## Пациент РЗ

**Пол:** Женский

**Дата рождения:** 27.04.2013 (7 лет)

**Дата обследования:** Неизвестна (2020)

### Жалобы

Отсутствие речи, неусидчивость, снижение концентрации внимания.

### Анамнез заболевания

07.2013: Мама заметила асимметрию подвижности рук. Консультация невролога, DS: [G80.9] Церебральный паралич неуточненный.

Открытое овальное окно, наблюдаются у кардиолога.

### Анамнез жизни

**Неонатальный период:** Беременность 1, роды 1. Беременность протекала без особенностей. Роды в сроке беременности 41 нед. Вес 3 800 г, длина тела 54 см. Закричала сразу, родилась с тройным обвитием пуповины, были направлены в перинатальный центр.

**Формула развития:** Моторное развитие: голову держать начала в 3,5 мес., сидеть в 10 мес., ходить в 3,5 года. Речевое развитие: Первое слово после 2 лет. Фразовой речи нет. На момент обследования словарный запас 2–3 слова.

**Генеалогический анализ и наследственность:** По данным клинико-генеалогического исследования, со слов матери, у двоюродной сестры пробанда ДЦП.

**Прочее:** Данные по оперативным вмешательствам, сопутствующим заболеваниям, эпидемиологический, трансфузионный и аллергологический анамнез отсутствуют.

## **Клиническое обследование**

Данные по клиническому обследованию пациента отсутствуют.

## **Лабораторные исследования**

Данные лабораторных исследований пациента отсутствуют.

## **Диагноз**

[F83] Задержка психо-речевого развития.

## **Дифференциальный диагноз**

Необходимо исключить в плане дифференциальной диагностики наследственные заболевания, связанные с наследственным нарушением обмена веществ, а также хромосомные нарушения, связанные со структурными или количественными перестройками хромосом.

## ЗАКЛЮЧЕНИЕ

по результатам секвенирования экзона от 20.08.2020 г.

|                       |                                |
|-----------------------|--------------------------------|
| Код пациента:         | P3                             |
| Вид материала:        | Кровь (венозная)               |
| Дата забора:          | 2019                           |
| Экзомная панель:      | SeqCap EZ MedExome Kit (Roche) |
| Количество прочтений: | — млн                          |
| Длина прочтений:      | 150 bp                         |
| Выравнивание:         | — %                            |
| Покрывание экзона:    |                                |
| – $\geq 10$ :         | — %                            |
| – среднее:            | — прочтений/позицию            |

**Заключение:** Значимых изменений, соответствующих критериям поиска, в данных экзомного секвенирования не обнаружено.

## Пациент Р4

**Пол:** Мужской

**Дата рождения:** 13.10.2014 (6 лет)

**Дата обследования:** Неизвестна (2020)

### Жалобы

Отсутствие речи, неадекватное поведение, агрессивность.

### Анамнез заболевания

В настоящее время ребенок не разговаривает, отмечается задержка психического развития.

### Анамнез жизни

**Неонатальный период:** Беременность 2, роды 2. Во время беременности мама лежала на сохранении. Роды в срок. Родился с обвитием пуповины. Вес 4 470 г, длина тела 57 см. Закричал сразу.

**Формула развития:** Моторное развитие: сел в 7 мес., ходить стал в 10 мес. Речевое развитие: в словарном запасе одно слово («уйди»), впервые сказал в 2 года, фразовой речи нет.

**Генеалогический анализ и наследственность:** У родного брата (пациент Р2) те же проявления.

**Прочее:** Данные по оперативным вмешательствам, сопутствующим заболеваниям, эпидемиологический, трансфузионный и аллергологический анамнез отсутствуют.

### Клиническое обследование

Данные по клиническому обследованию пациента отсутствуют.

## **Лабораторные исследования**

Данные лабораторных исследований пациента отсутствуют.

## **Диагноз**

[F83] Задержка психо-речевого развития.

## **Дифференциальный диагноз**

Необходимо исключить в плане дифференциальной диагностики наследственные заболевания, связанные с хромосомными нарушениями и наследственными нарушениями обмена веществ.

## ЗАКЛЮЧЕНИЕ

**по результатам секвенирования экзона от 23.08.2020 г.**

**Код пациента:** P4

**Вид материала:** Кровь (венозная)

**Дата забора:** 2019

**Экзомная панель:** SeqCap EZ MedExome Kit (Roche)

**Количество прочтений:** 242 млн

**Длина прочтений:** 150 bp

**Выравнивание:** 99,21 %

**Покрывание экзона:**

- $\geq 10$ : 97,99 %
- среднее: 132 прочтений/позицию

| Ген  | Положение (hg19) | Замена  | Генотип | Тип замены | Глубина | Ref/Alt |
|--|------------------|---|---------|------------|---------|---------|
| ZMYND11  | chr10:283578     | G>T   | het     | nonsyn SNV | 10      | 2/8     |
| <b>ОМIM фенотип:</b>   |                  | [616083] Умственная отсталость, АД, 3   |         |            |         |         |
| <b>HGMD:</b>   |                  | —   |         |            |         |         |
| <b>Экзон:</b>  |                  | 6/15  |         |            |         |         |
| <b>Частота аллеля:</b>   |                  | [rs757019491] 0,000 03  |         |            |         |         |
| <b>Критерии патогенности (согласно российским рекомендациям по интерпретации данных MPS [29]):</b> |                  | <p>[PM1] Вариант расположен в горячей точке и/или важных и хорошо исследованных функциональных доменах белка (например, активный сайт фермента), в которых не описаны доброкачественные изменения (Bromodomain;Bromodomain Zinc finger, FYVE/PHD-type Zinc finger, PHD-finger Zinc finger, PHD-type Zinc finger, PHD-type, conserved site Zinc finger, RING-type Zinc finger, RING/FYVE/PHD-type). [PM2] Вариант отсутствует в контрольных выборках (или встречается с крайне низкой частотой). [PP3] Результаты не менее трёх программ предсказания патогенности <i>in silico</i> подтверждают патогенное действие варианта на ген или продукт гена.</p> |         |            |         |         |
| <b>Комментарий:</b>  |                  | —   |         |            |         |         |



**Заключение:** Выявлен описанный ранее у больных вариант нуклеотидной последовательности в 6 экзоне гена *ZMYND11* (chr10:283578G>T, rs757019491) в гетерозиготном состоянии, приводящий к замене аргинина на серин в 190 кодоне (NP\_006615.2:p.Arg190Ser). Гетерозиготные варианты в гене *ZMYND11* описаны для пациентов с умственной отсталостью, АД, З (OMIM: 616083). Выявленный вариант встречается с крайне низкой частотой 0,003 % в контрольных базах данных. Он расположен в важном функциональном домене белка. Результаты 5 программ предсказания патогенности *in silico* подтверждают патогенное действие варианта на ген или продукт гена. По совокупности сведений, выявленный вариант нуклеотидной последовательности следует считать вариантом неопределенного клинического значения, который может иметь отношение к фенотипу пациента в случае появления новых подтверждающих данных. Других значимых изменений, соответствующих критериям поиска, не обнаружено. Вариант подтвержден по Сэнгеру.

## Пациент Р5

**Пол:** Мужской

**Дата рождения:** 25.07.2014 (6 лет)

**Дата обследования:** Неизвестна (2020)

### Жалобы

Отсутствие речи, не сформированы навыки опрятности, отсутствие контакта с окружающими.

### Анамнез заболевания

*01.2015:* Родители заметили, что ребенок не реагирует на своё имя, малоэмоциональный.

В настоящее время не разговаривает, с окружающими не контактирует, обращенную речь не понимает.

### Анамнез жизни

**Неонатальный период:** Беременность 4, роды 3. Роды в срок 40 нед. Вес 2 800 г, длина тела 50 см, окружность головы 34 см. Оценка по шкале Апгар 7 баллов. Грудное вскармливание до 1 года 4 мес., прикорм введен в 6 мес.

**Формула развития:** Моторное развитие: головку держит в 2 мес., сидит в 6 мес., ходит в 12 мес. Речевое развитие: первые слова в 12 мес., затем пропали, фразовая речь отсутствует.

**Генеалогический анализ и наследственность:** Со слов матери — у двоюродной сестры пробанда врожденная тугоухость.

**Прочее:** Данные по оперативным вмешательствам, сопутствующим заболеваниям, эпидемиологический, трансфузионный и аллергологический анамнез отсутствуют.

## **Клиническое обследование**

Общие данные объективного осмотра пациента отсутствуют.

**Нервная система:** Макроцефалия.

**Костно-мышечная система:** Мышечная гипотония. Увеличенные ушные раковины.

**Прочие:** Данные по кровеносной, дыхательной, пищеварительной, мочевыделительной, кроветворной и эндокринной системам, органам чувств отсутствуют.

## **Лабораторные исследования**

Данные лабораторных исследований пациента отсутствуют.

## **Диагноз**

[F83] Задержка психо-речевого развития.

## **Дифференциальный диагноз**

Необходимо исключить в плане дифференциальной диагностики наследственные заболевания, связанные с хромосомными нарушениями и наследственными нарушениями обмена веществ.

## ЗАКЛЮЧЕНИЕ

**по результатам секвенирования экзона от 24.08.2020 г.**

**Код пациента:** P5

**Вид материала:** Кровь (венозная)

**Дата забора:** 2019

**Экзомная панель:** SeqCap EZ MedExome Kit (Roche)

**Количество прочтений:** 80,7 млн

**Длина прочтений:** 150 bp

**Выравнивание:** — %

**Покрывание экзона:**

- $\geq 10$ : — %
- среднее: — прочтений/позицию

| Ген   | Положение (hg19) | Замена | Генотип    | Тип замены     | Глубина | Ref/Alt |
|---|------------------|--------|------------|----------------|---------|---------|
| GJB2  | chr13:20763686   | C>.    | het (hom?) | frameshift del | 10      | 2/8     |
| <p><b>OMIM фенотип:</b> [220290] Тугоухость, AP, вариант 1A</p> <p><b>HGMD:</b> [CD972240] Тугоухость, AP, вариант 1</p> <p><b>Экзон:</b> 1/1</p> <p><b>Частота аллеля:</b> [rs80338939] 0,023 2</p> <p><b>Критерии патогенности (согласно российским рекомендациям по интерпретации данных MPS [29]):</b> [BS1] Частота аллеля больше, чем ожидаемая для заболевания. [PVS1] LOF-варианты — варианты, приводящие к прекращению синтеза белка. [PP1] Вариант в гене, для которого точно установлена косегрегация с болезнью у нескольких поражённых членов семьи в двух поколениях. [PP5] Источники с хорошей репутацией указывают на патогенность варианта, но независимая оценка не проводилась.</p> <p><b>Комментарий:</b> —</p> |                  |        |            |                |         |         |

**Заключение:** Выявлен описанный ранее у больных вариант нуклеотидной последовательности в 1 экзоне гена *GJB2* (chr13:20763686C>., rs80338939) в гетерозиготном (вероятно, гомозиготном) состоянии, приводящий к сдви-

гу рамки считывания и терминации трансляции белка через 1 кодон (NM\_004004.6:c.35del(p.Gly12fs)). Гомозиготные LoF-варианты в гене *GJB2* описаны для пациентов с тугоухостью, AP, вариант 1A (OMIM: 220290). Выявленный вариант имеет частоту 2,3 % в финской популяции, в прочих же встречается с крайне низкой частотой. Согласно базе данных CLINVAR, более 60 публикаций оценивают этот вариант как патогенный. По совокупности сведений, выявленный вариант нуклеотидной последовательности следует считать патогенным, имеющим отношение к фенотипу пациента. Других значимых изменений, соответствующих критериям поиска, не обнаружено. Подтверждение по Сэнгеру не проводилось.

## Пациент Р6

**Пол:** Мужской

**Дата рождения:** 16.02.2007 (12 лет)

**Дата обследования:** Неизвестна (2020)

### Жалобы

Плохое зрение, периодические головные боли, пятна на коже цвета «кофе с молоком», подкожные образования.

### Анамнез заболевания

Врожденная глаукома, гнойный дакриоцистит.

05.2007: Однократный судорожный эпизод.

10.2014: OD — госпитализация по поводу оперированной врожденной глаукомы OD, субатрофия, Эпителиально-эндотелиальная дистрофия роговицы, отслойка сетчатки, хронический вялотекущий увеит. Прогноз по зрению бесперспективный, рекомендована госпитализация для энуклеации.

### Анамнез жизни

Обучается в общеобразовательной школе, программу усваивает.

**Неонатальный период:** Беременность 1, роды 1. Беременность протекала с токсикозом в I триместре на фоне уреоплазмоза. Роды оперативно путем кесарева сечения, в срок. Вес 4 040 г, длина тела 53 см. В период новорожденности гипоксически-ишемическое поражение ЦНС.

**Формула развития:** Физическое развитие выше среднего, дисгармоничное: длина тела 154 см (75–97 перцентиль), масса тела 54 кг (75 перцентиль по росту). Имеется макроцефалия — окружность головы 59 см (>97 перцентиль).

**Генеалогический анализ и наследственность:** у матери пробанда при осмотре отмечаются множественные пятна цвета «кофе с молоком», нейрофибромы. Со слов — у дедушки со стороны матери, сестры и брата дедушки, прабабушки — множественные пятна цвета «кофе с молоком», нейрофибромы.

**Оперативные вмешательства:** 12.2007: Реконструктивно-пластическая операция, OD трабекулотомия с эктомией, иридэктомией. Повторные операции в 2 и 4 года.

**Прочие:** Данные по сопутствующим заболеваниям, эпидемиологический, трансфузионный, аллергологический анамнез отсутствуют.

### **Клиническое обследование**

Сознание ясное. Состояние средней степени тяжести по совокупности клинических проявлений основного заболевания. ЧДД 26 /мин, ЧСС 82 уд/мин.

**Кровеносная система:** Область сердца визуально не изменена. Сердечные тоны громкие, ритмичные.

**Дыхательная система:** Носовое дыхание свободное. Грудная клетка цилиндрической формы. Вспомогательная мускулатура в акте дыхания участия не принимает. Перкуторный звук не изменен. Аускультативно выслушивается везикулярное дыхание по всем полям, хрипов и крепитации нет.

**Пищеварительная система:** Слизистые чистые, обычной окраски. Неправильный рост зубов. Язык чистый. Живот при пальпации мягкий, безболезненный во всех отделах. Печень по краю реберной дуги. Аппетит удовлетворительный. Стул регулярный, оформленный.

**Мочевыделительная система:** Мочеиспускание свободное, безболезненное. Половые органы сформированы по мужскому типу.

**Нервная система:** Ребенок уравновешенный. Сон спокойный. Общемозговой и менингеальной симптоматики нет. Черепно-мозговая иннервация: асимметрия лица в покое и при движениях (птоз правого верхнего века, гипертрофия правой половины лица, асимметрия оскала), девиация языка влево. Мышечный тонус физиологический. Активные и пассивные движения в суставах в полном объеме. Сухожильные рефлексy с рук и ног живые, D=S, патологических нет. Походка обычная. Координаторные пробы выполняет уверенно. В позе Ромберга устойчив.

**Кроветворная система:** Селезёнка не пальпируется.

**Эндокринная система:** Данные отсутствуют.

**Костно-мышечная система:** Толстая шея. Широкая грудная клетка, гипертелоризм сосков. Вальгусная установка коленных суставов.

**Органы чувств:** Субатрофия глазного яблока OD. Гиперметропия слабой степени OS.

**Местные изменения:** Множество пигментных пятен цвета «кофе с молоком» различного размера с округлыми краями, в складках кожи — мелкие пигментные пятна, напоминающие веснушки. Синофриз. Лимфатические узлы не пальпируются. Подкожная жировая клетчатка развита удовлетворительно. Отеков нет.

## Лабораторные исследования

**ОАК от 20.02.2018:** Лейкоциты  $5,9 \cdot 10^9$  кл/л, RBC  $5,16 \cdot 10^{12}$  кл/л, Hb 146 г/л, PLT  $216 \cdot 10^9$  кл/л, п/я нейтрофилы 1 %, с/я нейтрофилы 55 %, эозинофилы 6 %, моноциты 3 %, лимфоциты 35 %, СОЭ 3 мм/ч.

**Биохимический анализ крови от 20.02.2018:**  $\text{Na}^+$  141 ммоль/л,  $\text{K}^+$  4,4 ммоль/л, кальций общий 2,55 ммоль/л,  $\text{Ca}^{2+}$  1,22 ммоль/л, общий белок 77 г/л, глюкоза 6,0 ммоль/л, мочеви́на 4,8 ммоль/л, креатинин 62 мкмоль/л, мочева́я кислота 0,28 ммоль/л, билирубин общий 10,3 мкмоль/л, холестерин 3,1 ммоль/л,



АСТ 20 МЕ/л, АЛТ 11 МЕ/л, ЛДГ 346 Ед/л, КФК 87 Ед/л, ЩФ 533 Ед/л, Fe 13,3 мкмоль/л, фосфор неорганический 1,62 ммоль/л, осмолярность 282,1 мОсм/кг, СРБ 2,0 мг/л, альбумин 49,0 г/л, глобулины 28,0 г/л, альбумино-глобулиновый коэффициент 1,75, остеокальцин 87,4 нг/мл.

**Гормоны щитовидной железы от 20.02.2018:** ТТГ 1,09 мМЕ/л, Т4 свободный 10,43 пмоль/л, Т3 свободный 8,29 нмоль/л, АТ ТПО 1,5 мМЕ/л, АТ-ТГ 0,2 мМЕ/л, паратгормон 66 пг/мл, остаза 79,04 мкг/л

**Иммуноглобулины сыворотки от 20.02.2018:** IgG 11,5 г/л, IgA 1,0 г/л, IgM 0,9 г/л, IgE 249,6 МЕ/л.

**Исследование биоэнергетического обмена от 21.02.2018:** Глюкоза натощак 5,0 ммоль/л, молочная кислота натощак 1,0 ммоль/л, глюкоза через 1 ч. 5,2 ммоль/л, молочная кислота через 1 ч. 2,2 ммоль/л, глюкоза через 2 ч. 4,6 ммоль/л, молочная кислота через 2 ч. 2,0 ммоль/л.

**Исследование мочи по Зимницкому от 26.02.2018:** 6–9 ч. — количество 10 мл; 9–12 ч. — количество 90 мл, плотность 1 028 г/л; 12–15 ч. — количество 20 мл; 15–18 ч. — количество 80 мл, плотность 1 031 г/л; 19–21 ч. — количество 90 мл, плотность 1 030 г/л; 21–24 ч. — количество 20 мл; 0–3 ч. — количество 40 мл, плотность 1 030 г/л; 3–6 ч. — количество 50 мл, плотность 1 035 г/л. Дневной диурез 200 мл, суточный диурез 400 мл, ночной диурез 200 мл.

**Биохимическое исследование мочи от 21.02.2018:**  $\text{Ca}^{2+}$  0,19 ммоль/сут, Ca/Cr 0,03, P неорганический 10,8 ммоль/сут, P/Cr 1,89, оксалаты 0,192 мкмоль/сут, Ох/Cr 0,33, креатинин 5,71 ммоль/сут.

**ОАМ от 20.02.2018:** глюкоза — норма, белок отрицательно, билирубин отрицательно, уробилиноген 34 ммоль/л, рН 5,5, гемоглобин отрицательно, кетоны отрицательно, нитриты отрицательно, прозрачность полная, удельный вес 1 036 г/л, цвет желтый, эритроциты <1 в п/з, эпителий плоский <1 в п/з.

**ЭКГ от 27.02.2017:** Резко выраженная синусовая аритмия, периоды брадикардии, ЧСС 92–60 уд/мин. Отклонение ЭОС влево. Нарушение внутрижелудочкового проведения по правой ножке пучка Гиса. Ротационные изменения — поворот по часовой стрелке.

**ЭЭГ от 21.02.2018:** Корковая ритмика организована по возрасту. Ускорена с частотой  $\alpha$ -ритма 11,0 Гц, периодически экзальтирована. Очаговых изменений, эпилептиформной активности не отмечено.

**Эхо-КГ от 20.02.2018:** Пороков сердца не выявлено. Размеры полостей сердца в пределах нормы (размер левого желудочка на верхней границе нормы). Эхо-неоднородность миокарда межжелудочковой перегородки с участками гиперэхогенности. Систолическая и диастолическая функции не нарушены. Эктопические хорды и дисфункция хорд митрального клапана с регургитацией 1+.

**УЗИ органов БП и почек от 22.02.2018:** Реактивные изменения поджелудочной железы (неоднородная эхоструктура). Содержимое в желудке натошак. Увеличение количества и размеров лимфоузлов в воротах печени.

**Rg костей кистей с захватом предплечий от 20.02.2018:** Кости кистей сохранены. Структура костей не изменена. Эпифизы не изменены. Зоны роста четкие. Костный возраст 11 лет (соответствует паспортному), коэффициент оссификации в норме ( $1 \pm 0,2$ ). Соотношение костей в суставах стабильное.

**Rg костей голеней с захватом коленных суставов от 20.02.2018:** Кости голеней не изменены, структура без патологии. Эпифизы не изменены. Зоны роста четкие. Соотношение костей в суставах сохранное.

**МРТ головного мозга от 12.04.2012:** Желудочковая асимметрия, наружная гидроцефалия. Зона нарушения формирования борозд и извилин в области полюса правой височной доли.

**МРТ головного мозга от 02.02.2015:** Атрофия правого глазного яблока, правого зрительного нерва, гипоплазия хиазмы. Образование мягких тканей правой орбиты по задне-наружной стенке (возможно, фиброма). Единичный очаг у правой ножки мозга. Гипоплазия правой височной доли.

**МРТ головного мозга от 02.03.17:** МР картина нейрофиброматоза с поражением правой орбиты, костей черепа справа, формированием мягкотканного образования в правой височной области, единичным очагом в подкорковых ядрах справа. Атрофические изменения правого глазного яблока. Асимметрия, умеренное расширение боковых желудочков, наружных ликворных пространств в правой лобно-височной области. Асимметрия черепа.

## Диагноз

[Q85.0] Нейрофиброматоз I типа (возможно, микроделеционный синдром 17q11.2).

[H44.5] Дегенеративные состояния глазного яблока: вторичная субатрофия правого глазного яблока.

[H52.0] Гиперметропия, слабой степени, OS.

[I45.1] Другая и неуточненная блокада правой ножки пучка Гиса: идиопатическая блокада передней левой ветви пучка Гиса.

[M41.2] Другие идиопатические сколиозы: грудно-поясничный сколиоз I степени.

[M81.9] Остеопороз неуточненный.

Правостороннее сужение общего носового хода.

Синдром билиарной дисфункции.

### **Дифференциальный диагноз**

Необходима консультация генетика для уточнения диагноза.

## ЗАКЛЮЧЕНИЕ

по результатам секвенирования экзона от 24.08.2020 г.

|                       |                                |
|-----------------------|--------------------------------|
| Код пациента:         | P6                             |
| Вид материала:        | Кровь (венозная)               |
| Дата забора:          | 2019                           |
| Экзомная панель:      | SeqCap EZ MedExome Kit (Roche) |
| Количество прочтений: | 98,3 млн                       |
| Длина прочтений:      | 150 bp                         |
| Выравнивание:         | 99,76 %                        |
| Покрывание экзона:    |                                |
| – $\geq 10$ :         | 68,56 %                        |
| – среднее:            | 12 прочтений/позицию           |

**Заключение:** Значимых изменений, соответствующих критериям поиска, в данных экзомного секвенирования при индивидуальном и семейном анализе не обнаружено.

## **Пациент Р7**

**Пол:** Женский

**Дата рождения:** Неизвестна

**Дата обследования:** Неизвестна

### **Анамнез заболевания**

Со слов лечащего врача — симптомы сходны с таковыми у сына (пациент Р6).

## ЗАКЛЮЧЕНИЕ

по результатам секвенирования экзона от 09.09.2020 г.

|                       |                                |
|-----------------------|--------------------------------|
| Код пациента:         | P7                             |
| Вид материала:        | Кровь (венозная)               |
| Дата забора:          | 2019                           |
| Экзомная панель:      | SeqCap EZ MedExome Kit (Roche) |
| Количество прочтений: | 146,2 млн                      |
| Длина прочтений:      | 150 bp                         |
| Выравнивание:         | 99,82 %                        |
| Покрывание экзона:    |                                |
| – $\geq 10$ :         | 81,84 %                        |
| – среднее:            | 15 прочтений/позицию           |

**Заключение:** Значимых изменений, соответствующих критериям поиска, в данных экзомного секвенирования при индивидуальном и семейном анализе не обнаружено.

## Пациент Р8

**Пол:** Женский

**Дата рождения:** 01.04.2004 (15 лет)

**Дата обследования:** 16.01.2020

### Жалобы

Афты, гингивит.

### Анамнез заболевания

*02.2006:* Впервые возникла фебрильная лихорадка, стоматит. Лейкоцитарная формула в норме. В дальнейшем обострения стоматита отмечались ежемесячно.

*10.2006:* Впервые зафиксирован агранулоцитоз.

*01.2007:* Поставлен DS: [D70] Агранулоцитоз: аутоиммунная нейтропения. Исключены НЯК и болезнь Крона, назначен метипред 1 мг/кг по преднизолону.

*12.2008:* Пульс-терапия солумедролом. Назначен азатиоприн 2 мг/кг в качестве иммуносупрессивной терапии, но при попытке снижения дозы метипреда менее 3–4 мг/кг отмечались обострения агранулоцитоза, афтозного стоматита, явления колита.

*09.2009:* Пульс-терапия солумедролом. Начата терапия селлсептом 750 мг/сут. Отменена в связи с отсутствием эффекта.

*10.2012:* В связи с длительной ремиссией метипред полностью отменен.

*11.2012:* Назначен G-CSF.

*16.10.2013:* Госпитализирована с клиникой острого живота, проведена экстренная аппендэктомия лапароскопически. Послеоперационное состояние осложнено агранулоцитозом, резистентным к терапии G-CSF, сепсисом, ДВС-синдромом, парезом кишечника, однократным судорожным приступом.

*19.10.2013:* На фоне тяжелой полиорганной недостаточности нарушения микроциркуляции дистальных фаланг пальцев рук с последующим формированием некрозов ногтевых фаланг II, III, IV пальцев правой руки, III пальца

левой руки.

21.01.2014: Проведена некрэктомия пальцев.

24.01.2014: Начало терапии Селлсепт 1 000 мг/сут., преднизолон 7,5 мг/сут., стимуляция G-CSF пэгфилграстимом 1 раз раз в 2 нед., никотинамид 400 мг/сут. Далее селлсепт был заменен на рапамун, через 3 мес. рапамун отменен в связи с неэффективностью.

10.2014: Госпитализация в связи с лейкоцитарной гиперстимуляцией на фоне пролонгированного G-CSF. Подобрана доза G-CSF — 1 мкг/кг с хорошим гематологическим ответом.

07.07.2015: Травма III пальца правой руки, умеренный отек, болезненность.

13.07.2015: Хирургическое вскрытие, назначен клиндамицин, положительная динамика.

07.2015: Госпитализация. Проведена КПМ, данных за бластоз нет. Продолжена терапия лейкоцитимом 5 мкг/кг — 150 мкг через день. Постепенно отменяли метилпреднизолон, получала дифлюкан 100 мг/сут..

15.04.2016: Панариций I пальца правой руки, II пальца левой руки. Произведено вскрытие, дренирование, назначена антибиотикотерапия с положительной динамикой.

19.08.2016: Госпитализирована с DS: [N61] Острый гнойный нелактационный мастит слева. Терапия метронидазолом, левомеколь, цефазолин, флюконазол, филграстим 150 мкг/сут. Пункция и дренирование абсцесса.

14.09.2016: Госпитализирована с DS: [N61] Гнойный мастит. Проведено оперативное лечение, послеоперационный период прошел гладко.

10.2016: Тенденция к тромбоцитопении, снижение В-клеток, KREC. Терапия азатиоприном.

30.10.2016: [K81.0] Острый акалькулезный холецистит, воспалительные изменения в проекции малого сальника. Назначена комплексная противомикробная, симптоматическая терапия с положительной динамикой.

02.11.2016: Повышение СРР до 207 мг/л, панцитопения. В связи с тем, что нельзя было исключить развитие гематофагоцитарного синдрома, назначена пульс-терапия солумедролом (5 дней), далее поддерживающая терапия преднизолоном в дозе 0,5 мг/кг перорально.

Прогрессирует развитие медикаментозного гиперкортицизма, прибавка в ве-



се с 30 до 50 кг. Был перерыв в приеме азатиоприна (по причине отсутствия препарата в продаже), прочие рекомендованные препараты принимала в полном объёме.

07.2017: Госпитализация. Дозировка преднизолона снижена до  $1\frac{3}{4}$  табл.

18.07.2017: Снижение веса на 9 кг на фоне снижения дозы ГКС.

12.08.2017: Обильные длительные менструации, стабильное снижение HGB до 105 г/л.

04.10.2017: Панариций IV пальца, после спиртовых повязок с улучшением. Преднизолон снижен до  $\frac{1}{2}-\frac{1}{4}$  табл. через день, азатиоприн с коррекцией по весу  $1\frac{1}{4}$  табл. (62,5 мг), G-CSF 150 мкг через день.

02.2018: Афты. Заместительную терапию ВВИГ (интралект 20 мг) получает регулярно. Преднизолон регулярно в дозе  $\frac{1}{4}$  табл. Азатиоприн 75 мг, G-CSF через день, лейкостим 150 мкг/сут., профилактическая терапия азитромицином 250 мг/сут.

08.2018: Появление крупных афт на слизистой щёк, языка, повышение температуры до фебрильных цифр раз в 2 нед. Терапия азатиоприн 75 мг, G-CSF 150 мкг (3,8 мкг/кг), преднизолон  $\frac{1}{4}-\frac{1}{2}$  табл. через день. Заместительная терапия ВВИГ не проводилась по причине отсутствия препарата.

11.2018: Выявлены АТ к нейтрофилам 1:8–1:32.

12.2018: Заместительная терапия ВВИГ.

06.2019: G-CSF 300 мкг с умеренным гематологическим ответом, терапия азатиоприном и преднизолоном отменена в связи с неэффективностью. Специфическая терапия ритуксимабом: после второй инфузии нейтрофилы повысились до  $2,87 \cdot 10^9$  кл/л, с дальнейшим падением до  $0,5 \cdot 10^9$  кл/л. От проведения ТГСК как единственного куративного метода мама отказалась.

## **Анамнез жизни**

**Генеалогический анализ и наследственность:** Наследственность не отягощена.

**Травмы и оперативные вмешательства:** 15.11.2017: ЗЧМТ, сотрясение ГМ (упала на физкультуре).

**Трансфузионный анамнез:** ВВИГ регулярно.

**Прочие:** Данные по неонатальный периоду развития, сопутствующим заболеваниям, эпидемиологический и аллергологический анамнез отсутствуют.

## **Клиническое обследование**

Общее состояние стабильное, температура тела 36,7 °С, вес 48 кг. Сознание ясное.

**Кровеносная система:** Тоны сердца ясные, ритмичные.

**Дыхательная система:** Носовое дыхание свободное. Аускультативно: дыхание везикулярно, проводится по всем полям, хрипы не выслушиваются.

**Пищеварительная система:** Выраженный гингивит, язвы отсутствуют, язык с фестончатыми краями вследствие рубцовых изменений. Зев чистый, без гиперемии. Живот мягкий, безболезненный. Печень по краю реберной дуги. Стул в норме.

**Местные изменения:** Кожа и подкожная клетчатка без особенностей. ЛУ: подчелюстные плотные, до 1 см, шейные до 0,5 см, переднешейный болезненный, до 2 см.

**Прочие:** данные по мочевыделительной, нервной, кроветворной, эндокринной, костно-мышечной системам, а также органам чувств отсутствуют.

## **Лабораторные исследования**

**Таргетное секвенирование от 2013:** В генах *KRAS* (2, 3 экзоны), *NRAS* (2, 3 экзоны), *CBL*, *PTPN11* клинически значимых вариантов не обнаружено.

**Таргетное секвенирование от 04.10.2016:** В гене *ELANE* клинически значимых вариантов не обнаружено.

**КМП от 07.07.2017:** Данных за бластоз нет, цитогенетика в норме.

**Биохимический анализ крови от 07.09.2017:** Fe 20,1 мкмоль/л, ферритин 203 мкг/л.

Множественно — ОАК: нейтропения  $0,8 \cdot 10^9$  кл/л, за исключением анализа после введения ВВИГ — АКН  $2,2 \cdot 10^9$  кл/л.

**Биохимический анализ крови от 04.10.2017:** Hb 105 г/л, лейкоциты  $2,9 \cdot 10^9$  кл/л, АКН  $0,6 \cdot 10^9$  кл/л, тромбоциты  $147 \cdot 10^9$  кл/л.

**ОАК от 01.03.2018:** лейкоциты  $3,1 \cdot 10^9$  кл/л, АКН  $0,347 \cdot 10^9$  кл/л, тромбоциты  $160 \cdot 10^9$  кл/л, предтрансфузионный уровень IgG 1316 мг/дл, IgM 651 мг/дл, IgA 196 мг/дл, СРБ отрицательно.

**Биохимический анализ крови от 01.03.2018:** Без патологий, незначительно повышен АСЛО.

**ОАК от 08.2018:** Hb 104–110 г/л, тромбоциты  $209 \cdot 10^9$  кл/л, лейкоциты  $5,6 \cdot 10^9$  кл/л, АКН  $0,455 \cdot 10^9$  кл/л, относительный моноцитоз 20 % АКМ ( $0,838 \cdot 10^9$  кл/л).

**Цитогенетический анализ от 11.2018:** Моносомии 7 хромосомы, трисомии 8 хромосомы не обнаружено, митозов нет.

**Иммуноглобулины от 11.2018:** IgG 2100 мг/дл, IgM 800 мг/дл.

**Общий и биохимический анализы крови от 02.2019:** Тромбоциты  $90 \cdot 10^9$  кл/л, Hb 90 г/л.

**УЗИ БП от 02.2019:** Спленомегалия 150x46 мм.

**Иммунологическая панель от 05.2019:** Молекулярно-генетических дефектов не найдено.

## Диагноз

[D84.9] Первичный иммунодефицит неуточненный, осложненный иммунной нейтропенией и тромбоцитопенией.

## Дифференциальный диагноз

Необходим дальнейший поиск молекулярно-генетических дефектов.

## ЗАКЛЮЧЕНИЕ

по результатам секвенирования экзона от 24.08.2020 г.

|                       |                                |
|-----------------------|--------------------------------|
| Код пациента:         | P8                             |
| Вид материала:        | Кровь (венозная)               |
| Дата забора:          | 2019                           |
| Экзомная панель:      | SeqCap EZ MedExome Kit (Roche) |
| Количество прочтений: | 102 млн                        |
| Длина прочтений:      | 150 bp                         |
| Выравнивание:         | 99,7 %                         |
| Покрывание экзона:    |                                |
| – $\geq 10$ :         | 72,1 %                         |
| – среднее:            | 15 прочтений/позицию           |

**Заключение:** Значимых изменений, соответствующих критериям поиска, в данных экзомного секвенирования не обнаружено.

## Пациент Р9

**Пол:** Женский

**Дата рождения:** 19.02.2013 (7 лет)

**Дата обследования:** 10.04.2020

### Анамнез заболевания

*05.2013:* Госпитализирована в отделение неврологии. DS: [Q04.9] Врожденная аномалия мозга неуточненная. [Q02] Микроцефалия. [G80.0] Спастический церебральный парез, квадриплегия. [F82] Задержка психомоторного развития. ЦМВ инфекция, проведено лечение зовираксом.

*10.2013:* Госпитализация с DS: [G96.8] Органическое поражение ЦНС с формированием перивентрикулярных глиозно-атрофических изменений, двусторонний синдром пирамидной недостаточности. [Q02] Микроцефалия. [Q21.1] Дефект предсердной перегородки: открытое овальное окно. [B27.1] Цитомегаловирусный мононуклеоз в стадии активации. Далее отмечались рецидивирующие инфекционные эпизоды, требующие антибиотикотерапии.

*06.2018:* Консультирована иммунологом, DS: [Q87.8] Синдром повреждения Неймегена, кровь отправлена на исследование гена *NBN*, выявлена выраженная гипогаммаглобулинемия, инициирована регулярная терапия ВВИГ.

*08.2018:* Госпитализация. За время пребывания в отделении состояние оставалось стабильным, отмечалось течение гнойного риносинусита, с положительной динамикой на фоне проводимой антибиотико- и местной терапии. По данным КТ, очаговых изменений в легочной ткани и увеличенных лимфоузлов не выявлено. Случайная находка: в эпифизе правой плечевой кости два гиперденсивных очага размером 5 мм и 2 мм мм, подозрение на остеому, консультирована хирургом-ортопедом. Учитывая отсутствие рентгенологических признаков злокачественности, болевого синдрома, рекомендовано плановое дообследование (МРТ плечевого сустава) и динамическое наблюдение.

*15.05.2019:* Госпитализация с DS: [J41.0] Простой хронический бронхит в стадии обострения. [H72.H66.3] Хронический левосторонний гнойный сред-

ний отит в стадии обострения (перфорация барабанной перепонки пристеночная). Получала комплексную антибактериальную, противогрибковую и противовирусную терапию (Цефепим, Микамин, Бакперазон, Цимевен). Состояние расценено как EBV-ассоциированное поражение легких, инициирован курс терапии ритуксимабом (Мабтера). Терапию переносила удовлетворительно.

03.06.2019: Госпитализация в отделение иммунологии. В отделении была продолжена ранее инициированная специфическая терапия Ритуксимабом. КТ органов грудной клетки: картина интерстициальной лимфоцитарной болезни легких (ИЛБЛ). По данным исследования БАЛ, отмечено снижение вирусной нагрузки. В гемограмме выявлена нейтропения. Была инициирована терапия G-CSF (Зарсио) 3 мкг/кг ежедневно, уровень нейтрофилов поднялся до физиологической нормы.

11.09.2019: Госпитализация, проведена очередная инфузия ВВИГ Привиджен в дозе 10 г (претрансфузионный уровень IgG 5,8 г/л). Продолжена ранее инициированная терапия Ритуксимабом (Мабтера) 250 мг в/в.

21.10.2019: Поступила в отделение иммунологии для проведения контрольного обследования. На фоне проведенной ранее патогенетической терапии Ритуксимабом со стороны течения интерстициально-лимфоцитарной болезни легких отмечена положительная динамика (клинически и по данным МСКТ ОГК), отсутствует EBV-виремия и персистенция EBV в бронхоальвеолярном лаваже. Проведена инфузия ВВИГ Привиджен 7,5 г в/в, перенесла удовлетворительно. В стабильном состоянии выписана под наблюдение специалистов по месту жительства.

Постоянно получала профилактическую противомикробную терапию (азитромицин, бисептол, дифлюкан), проводилась регулярная иммунозаместительная терапия ВВИГ.

## **Анамнез жизни**

**Неонатальный период:** Беременность 1, роды 1. Родоразрешение оперативно на сроке 38 нед. Течение беременности на фоне фетоплацентарной недостаточности, задержки внутриутробного развития, угрозы прерывания, ма-

ловодия, анемии, тромбофилии, ОРВИ. Вес 2 600 г. Оценка по шкале Апгар 6/7 баллов. К груди приложена на 2 сутки. В возрасте 2 дней дней консультирована генетиком, DS: [Q87.8] Синдром Опица. Из роддома выписана на 5 день с DS: [Q04.9] ВПР мозга, [Q02] Микроцефалия. [P05.9] ЗВУР по гипопластическому типу. [Q70.3] Синдактилия 4–5 пальцев правой стопы, неполная кожная форма.

**Генеалогический анализ и наследственность:** Со слов, наследственность по онкогематологической патологии не отягощена.

**Оперативные вмешательства:** 20.04.2020: Проведена постановка центрального венозного катетера по типу Broviac, без осложнений.

**Сопутствующие заболевания:** Пансинусит.

**Эпидемиологический анамнез:** БЦЖ-М проведено в роддоме. 14.7.2017: V1 против кори, краснухи, паротита. 22.10.2019: Диаскинтест отрицательный.

**Трансфузионный анамнез:** Инфузия ВВИГ многократно.

**Аллергологический анамнез:** Аллергия на ванкорус — высыпания на коже. Поллиноз — полынь.

### **Клиническое обследование**

Вес 14,3 кг. Рост 110,5 см. АД 100/62 торр. ЧСС 95 уд/мин. Температура 36,6 °С. Состояние тяжелое по основному заболеванию. Самочувствие удовлетворительное.

**Кровеносная система:** Тоны сердца ясные, ритмичные.

**Дыхательная система:** Носовое дыхание свободное. Аускультативно в легких дыхание жесткое, равномерно проводится во все отделы, хрипов нет.

**Пищеварительная система:** Видимые слизистые чистые, розовые. Хронический гингивит в стадии ремиссии. Живот мягкий, безболезненный при пальпации во всех отделах. Печень пальпируется по краю реберной дуги. Стул в норме.

**Мочевыделительная система:** Мочеиспускание в норме.

**Нервная система:** Очаговой и менингеальной симптоматики нет. Микроцефалия.

**Кроветворная система:** Селезенка не пальпируется.

**Местные изменения:** Кожные покровы розовые, на коже лица, конечностей распространенные множественные элементы округлой эритематозной очаговой сыпи, местами с продуктивным компонентом и элементами с шелушением (гранулематозный дерматит).

**Прочее:** Данные по эндокринной и костно-мышечной системам, а также органам чувств отсутствуют.

**Органы чувств:** OU-по переднему отрезку и на глазном дне без грубой очаговой патологии.

### **Лабораторные исследования**

**Кариотипирование от 08.2013:** Хромосомная болезнь, мозаичный вариант, 47,XX,+mar(4)/46.XX(22).

**Бронхоальвеолярный лаваж от 16.05.2019:** EBV 1 839 500 копий/мл.

**Бронхоальвеолярный лаваж от 03.06.2019:** EBV 642 копий/мл.

**Гемограмма от 03.06.2019:** Нейтрофилы  $0,39 \cdot 10^9$  кл/л.

**ОАК от 13.04.2020:** Лейкоциты  $2,99 \cdot 10^9$  кл/л, RBC  $4,85 \cdot 10^{12}$  кл/л, Hb 121 г/л, PLT  $215 \cdot 10^9$  кл/л, нейтрофилы  $1,14 \cdot 10^9$  кл/л, базофилы  $0 \cdot 10^9$  кл/л, эозинофилы  $0,2 \cdot 10^9$  кл/л, моноциты  $0,47 \cdot 10^9$  кл/л, лимфоциты  $1,18 \cdot 10^9$  кл/л,



ретикулоциты 8 %, СОЭ 6 мм/ч, с/я нейтрофилы 44 %, эозинофилы 5 %, лимфоциты 39 %, моноциты 12 %.

**ОАК от 27.04.2020:** Лейкоциты  $1,85 \cdot 10^9$  кл/л, RBC  $4,24 \cdot 10^{12}$  кл/л, Hb 107 г/л, PLT  $223 \cdot 10^9$  кл/л, нейтрофилы  $0,77 \cdot 10^9$  кл/л, базофилы  $0 \cdot 10^9$  кл/л, эозинофилы  $0,18 \cdot 10^9$  кл/л, моноциты  $0,44 \cdot 10^9$  кл/л, лимфоциты  $0,46 \cdot 10^9$  кл/л.

**ОАК от 04.05.2020:** Лейкоциты  $3,77 \cdot 10^9$  кл/л, RBC  $4,21 \cdot 10^{12}$  кл/л, Hb 109 г/л, PLT  $247 \cdot 10^9$  кл/л, нейтрофилы  $2,21 \cdot 10^9$  кл/л.

**ОАК от 18.05.2020:** Лейкоциты  $1,01 \cdot 10^9$  кл/л, RBC  $3,98 \cdot 10^{12}$  кл/л, Hb 104 г/л, PLT  $222 \cdot 10^9$  кл/л, нейтрофилы  $0,4 \cdot 10^9$  кл/л, базофилы  $0,01 \cdot 10^9$  кл/л, эозинофилы  $0,2 \cdot 10^9$  кл/л, моноциты  $0,1 \cdot 10^9$  кл/л, лимфоциты  $0,3 \cdot 10^9$  кл/л.

**ОАК от 19.05.2020:** Лейкоциты  $4,33 \cdot 10^9$  кл/л, RBC  $3,95 \cdot 10^{12}$  кл/л, Hb 102 г/л, PLT  $222 \cdot 10^9$  кл/л, нейтрофилы  $3,41 \cdot 10^9$  кл/л, базофилы  $0,01 \cdot 10^9$  кл/л, эозинофилы  $0,24 \cdot 10^9$  кл/л, моноциты  $0,19 \cdot 10^9$  кл/л, лимфоциты  $0,48 \cdot 10^9$  кл/л.

**Биохимический анализ крови от 13.04.2020:** АЛТ 34 Ед/л, альбумин 43,4 г/л,  $\alpha$ -амилаза 69 Ед/л, П-амилаза 12,26 Ед/л, АСТ 39 Ед/л, общий белок 61 г/л, билирубин общий 5 мкмоль/л, билирубин прямой 2,6 мкмоль/л, ГГТ 0 Ед/л, глюкоза 4,48 ммоль/л, Fe 6,2 мкмоль/л, креатинин 43,1 мкмоль/л, ЛДГ 209 Ед/л, липаза 71 Ед/л, мочевины 5,8 ммоль/л, СРБ 0,5 мг/л,  $K^+$  3,8 ммоль/л,  $Mg^{2+}$  общий 0,83 ммоль/л,  $Na^+$  140 ммоль/л,  $Ca^{2+}$  1,25 ммоль/л, триглицериды 0,68 ммоль/л, ферритин 26 мкг/л, ЩФ 140 Ед/л, кортизол 10,5 мкг/дл.

**Биохимический анализ крови от 04.05.2020:** АЛТ 37 Ед/л, альбумин 41,4 г/л, АСТ 39 Ед/л, общий белок 63 г/л, билирубин общий 4,8 мкмоль/л, билирубин прямой 2,3 мкмоль/л, глюкоза 5,24 ммоль/л, креатинин 36,8 мкмоль/л, ЛДГ 320 Ед/л, мочевины 6,4 ммоль/л, СРБ 0,8 мг/л,  $K^+$  4,2 ммоль/л, мочевины кислота 0,118 ммоль/л,  $Na^+$  137 ммоль/л, триглицериды 0,48 ммоль/л, ферритин 29,1 мкг/л.

**Биохимический анализ крови от 08.05.2020:** Креатинин 35,8 мкмоль/л, мочевины 4,6 ммоль/л.

**Биохимический анализ крови от 18.05.2020:** АЛТ 51 Ед/л,  $\alpha$ -амилаза 48 Ед/л, П-амилаза 7,33 Ед/л, АСТ 43 Ед/л, билирубин общий 5,5 мкмоль/л, билирубин прямой 2,7 мкмоль/л, креатинин 38,5 мкмоль/л, ЛДГ 210 Ед/л, липаза 41 Ед/л, мочевины 4,8 ммоль/л.

**Цистатин С от 27.04.2020:** 0,85 мг/л.

**Цистатин С от 08.05.2020:** 1,04 мг/л.

**Иммунофенотипирование лейкоцитов периферической крови от 13.04.2020:**

Лейкоциты  $3,62 \cdot 10^9$  кл/л, гранулоциты 46 % ( $1,67 \cdot 10^9$  кл/л), моноциты 13 % ( $0,47 \cdot 10^9$  кл/л), лимфоциты 41 % ( $1,48 \cdot 10^9$  кл/л), CD3+ (Т-лимфоциты) 80,1 % ( $1,19 \cdot 10^9$  кл/л), CD3+4+ (Т-хелперы) 21 % ( $0,31 \cdot 10^9$  кл/л), CD3+8+ (Т-цитотокс.) 38 % ( $0,56 \cdot 10^9$  кл/л), CD3+4-8- 21 % ( $0,31 \cdot 10^9$  кл/л), CD19+ (В-лимфоциты) 0 % ( $0,0 \cdot 10^9$  кл/л), CD3-CD16+CD56+ (NK-cells) 19,8 % ( $0,29 \cdot 10^9$  кл/л), CD56 + high NK 11,6 % ( $0,03 \cdot 10^9$  кл/л), CD3+CD16+CD56+ (Т-NK-cells) 8,1 % ( $0,12 \cdot 10^9$  кл/л), CD14+high CD16- Mon 81,7 % ( $0,38 \cdot 10^9$  кл/л), CD14+high CD16+high Mon 14,6 % ( $0,07 \cdot 10^9$  кл/л), CD14+ CD16+high Mon 3,2 % ( $0,02 \cdot 10^9$  кл/л).

**Коагулограмма от 13.04.2020:** Фибриноген по Клауссу 1,97 г/л, протромбин по Квику 110 %, МНО 0,9, протромбиновое время 10,8 сек, АЧТВ 31,7 сек, тромбиновое время 29,1 сек.

**Иммуноглобулины сыворотки крови от 13.04.2020:** IgA <0.26 г/л, IgM 0,705 г/л, IgG 7,27 г/л.

**ОАМ от 13.04.2020:** Без патологий.

**Группа крови от 13.04.20:** В(III), Rh+. Фенотип C(+)c(-)E(-)e(+). Kell-отрицательный.

**Прямая проба Кумбса от 13.04.2020:** Отрицательно.

**Гормональный статус от 13.04.2020:** ТТГ 4,06 мМЕ/л, ИФР-1 70 нг/мл, Т4 свободный 1,35 нг/дл.

**Госпитальный скрининг от 13.04.2020:** HBsAg, HIV-1/2, анти-HCV, Сифилис суммарные АТ отрицательные.

**ПЦР-исследование крови от 14.04.20:** EBV, HHV VI типа отрицательно, CMV 43 копий/мл.

**ПЦР-исследование крови от 28.04:** CMV 67 копий/мл.

**ПЦР-исследование крови от 12.05.2020:** EBV, CMV, HHV VI типа отрицательно.

**Госпитальный скрининг на носительство MRSA (мазок из носа) от 11.04.2020:** Отрицательно.

**Госпитальный скрининг на носительство ESBL (ректальный мазок) от 11.04.2020:** Отрицательно.

**Микробиологическое исследование (посев) мазок из носа от 24.04.2020:**

*Streptococcus mitis* 10<sup>3</sup> КОЕ/мл, *Corynebacterium pseudodiphthericum* 10<sup>3</sup> КОЕ/мл.

**Микробиологическое исследование кала от 28.04.2020:** *Streptococcus gallolyticus* 10<sup>3</sup> КОЕ/мл.

**Микробиологическое исследование мазка из зева от 28.04.2020:** *Streptococcus mitis* 10<sup>4</sup> КОЕ/мл, *Haemophilus parainfluenzae* 10<sup>3</sup> КОЕ/мл.

**Микробиологическое исследование мазка из перианальной области от 23.04.2020:** *Corynebacterium pseudodiphthericum* 10<sup>2</sup> КОЕ/мл.

**Микробиологическое исследование мазка из носа от 23.04.2020:** *Streptococcus gallolyticus* 10<sup>3</sup> КОЕ/мл, *Corynebacterium tuberculostearicum* <10<sup>3</sup> КОЕ/мл, *Staphylococcus hominis* 10<sup>2</sup> КОЕ/мл.

**ПЦР-исследование крови от 16.04.2020:** HIV, HBV, HCV отрицательно.

**ПЦР исследование биоптата кожи от 20.04.2020:** EBV, CMV отрицательно, HHV VI типа 128 копий/мл, PCR-Rubella отрицательно.

**Бронхоальвеолярный лаваж от 17.04.2020:** HHV VI типа, CMV отрицательно, EBV 7 760 копий/мл, *Mycobacterium tuberculosis complex* (MTC) отрицательно, *Mycoplasma pneumoniae* отрицательно, *Chlamydia pneumoniae* отрицательно, ДНК-пневмоцисты отрицательно, *Adenovirus* положительный, *Coronavirus* отрицательный, *Bocavirus* положительный, *Rhinovirus* положительный, *Metapneumovirus* отрицательный, *Parainfluenzae virus* тип I, II, III, IV отрицательно, RSV отрицательный, галактоманнан отрицательный, при флуоресцентной микроскопии (калькофлюор белый, КФБ) структура мицелия не обнаружена, при посеве жидкости бронхоальвеолярного лаважа микроорганизмы не обнаружены.

**Гистологическое исследование биоптата кожи от 04.2020:** В пределах исследованного материала признаки васкулита мелких сосудов, поверхностного дерматита, признаков гранулематозного поражения не выявлены.

**ЭКГ от 17.04.2020:** Слабая синусовая аритмия. Нормальное положение ЭОС.

**УЗИ ОБП и мочевыделительной системы от 21.04.2020:** Данные за патологию отсутствуют.

**УЗИ вен шеи с доплерографией от 18.04.2020:** Данные за патологию отсутствуют.

**Эхо-КГ от 21.04.2020:** Камеры сердца не расширены. Сократительная способность миокарда ЛЖ не нарушена.

**МСКТ органов грудной клетки от 14.04.2020:** ОГК: Единичные мелкие очаги в обоих легких (возможно начальные воспалительные изменения, рекомендуется клинико-лабораторная корреляция). Очаги остеосклероза в головке правой плечевой кости (остеома?). ОБП: Диффузные слабо дифференцирующиеся изменения селезенки (рекомендуется контроль УЗИ).

**Бронхоскопия от 17.04.2020:** Двусторонний бронхит I степени интенсивности.

**MPT Total body от 22.04.2020:** Зоны изменений метафизов трубчатых костей нижних конечностей на фоне проводимой терапии или нарушения обменных процессов без структурной патологии, минимальные реактивные изменения в локтевых, коленных и голеностопных суставах в виде умеренного повышения количества жидкости.

**MPT головного мозга с контрастом от 26.04.2020:** МР-признаки перивентрикулярной узловой гетеротопии справа. Признаки полисинусита. В сравнении с данными от 20.06.2019 — МР-картина без динамики.

**ЭЭГ с видеомониторингом в состоянии бодрствования от 23.04.2020:** Легкая дезорганизация основного коркового ритма, развитого в соответствии с возрастом пациентки, медленноволновой активностью, полифазными комплексами. Асимметрия биопотенциалов зарегистрирована в лобно-центральных отделах со снижением амплитуды справа. Реакции на функциональные пробы адекватные. Типичной эпилептиформной активности за время исследования не выявлено. На синхронной видеозаписи патологических движений не зарегистрировано.

**Спирометрия от 18.05.2020:** Показатели спирометрии в границах нормы.

**Импульсная осциллометрия от 18.05.2020:** Показатели ИОС в границах нормы.

## **Диагноз**

[Q87.8] Синдром повреждения Неймегена

## **Дифференциальный диагноз**

Провести дифференциальную диагностику с прочими генетическими нарушениями, приводящими к иммунодефициту.

## ЗАКЛЮЧЕНИЕ

**по результатам секвенирования экзона от 24.08.2020 г.**

**Код пациента:** P9

**Вид материала:** Кровь (венозная)

**Дата забора:** 2019

**Экзомная панель:** SeqCap EZ MedExome Kit (Roche)

**Количество прочтений:** 81,2 млн

**Длина прочтений:** 150 bp

**Выравнивание:** — %

**Покрывание экзона:**

- $\geq 10$ : — %
- среднее: — прочтений/позицию

| Ген  | Положение (hg19) | Замена  | Генотип | Тип замены     | Глубина | Ref/Alt |
|--|------------------|---------|---------|----------------|---------|---------|
| NBN  | chr8:90983442    | TTTGT>. | hom     | frameshift del | 7       | 0/7     |
| <p><b>OMIM фенотип:</b> [251260] Синдром повреждения Неймегена, AP</p> <p><b>HGMD:</b> —</p> <p><b>Экзон:</b> 6/16</p> <p><b>Частота аллеля:</b> [rs587776650] 0,000 63</p> <p><b>Критерии патогенности (согласно российским рекомендациям по интерпретации данных MPS [29]):</b> [PM2] Вариант отсутствует в контрольных выборках (или встречается с крайне низкой частотой). [PVS1] LOF-варианты — варианты, приводящие к прекращению синтеза белка. [PP5] Источники с хорошей репутацией указывают на патогенность варианта, но независимая оценка не проводилась (CLINVAR).</p> <p><b>Комментарий:</b> —</p> |                  |         |         |                |         |         |

**Заключение:** Выявлен описанный ранее у больных вариант нуклеотидной последовательности в 6 экзоне гена *NBN* (chr8:90983442TTTGT>., rs587776650) в гомозиготном состоянии, приводящий к сдвигу рамки считывания и терминации трансляции через 16 кодонов (p.Lys219AsnfsTer16; K219Nfs\*16). Гомозиготные LoF-варианты в гене *NBN* описаны для пациентов с синдромом по-

вреждения Неймегена, AP (OMIM: 251260). Выявленный вариант не зафиксирован в контрольных выборках ESP, 1000 Genomes и ExAC. Согласно базе данных CLINVAR, 7 публикаций оценивают этот вариант как патогенный и как фактор риска. По совокупности сведений, выявленный вариант нуклеотидной последовательности следует считать патогенным, имеющим отношение к фенотипу пациента. Других значимых изменений, соответствующих критериям поиска, не обнаружено. Подтверждение по Сэнгеру не проводилось.

## Пациент Р10

**Пол:** Женский

**Дата рождения:** Неизвестна

**Дата обследования:** Неизвестна (2020)

### Жалобы

Мало прибавляет в весе.

### Анамнез заболевания

Рудименты 6 пальцев постаксиальные, удалены в 6 мес. Открытое овальное окно.

### Анамнез жизни

**Формула развития:** Задержка психомоторного развития: поворачивается на живот с 6 мес., не ползает, сама не садится.

**Прочее:** Данные по неонатальному периоду развития, наследственности, оперативным вмешательствам, сопутствующим заболеваниям, эпидемиологический, трансфузионный и аллергологический анамнез отсутствуют.

### Клиническое обследование

Вес 3 410 г, длина тела 53 см.

**Нервная система:** Микроцефалия, окружность головы 42 см см.

**Костно-мышечная система:** Гипотония. Большой родничок 1x1 см. Широкая переносица, Низкопосаженные диспластичные ушные раковины. Клинодактилия V пальцев кистей. Неправильное положение пальцев стоп, вальгусные стопы.



**Органы чувств:** Монголоидный разрез глаз.

**Прочее:** Данные по кровеносной, дыхательной, пищеварительной, мочевыделительной, кроветворной и эндокринной системам отсутствуют.

### **Лабораторные исследования**

Данные лабораторных исследований пациента отсутствуют.

### **Диагноз**

[Q89.9] Комплекс врожденных аномалий.

### **Дифференциальный диагноз**

Необходима консультация генетика.

## ЗАКЛЮЧЕНИЕ

по результатам секвенирования экзона от 03.09.2020 г.

|                       |                                |
|-----------------------|--------------------------------|
| Код пациента:         | P10                            |
| Вид материала:        | Кровь (венозная)               |
| Дата забора:          | 2019                           |
| Экзомная панель:      | SeqCap EZ MedExome Kit (Roche) |
| Количество прочтений: | 81,4 млн                       |
| Длина прочтений:      | 150 bp                         |
| Выравнивание:         | — %                            |
| Покрывание экзона:    |                                |
| – $\geq 10$ :         | — %                            |
| – среднее:            | — прочтений/позицию            |

**Заключение:** Значимых изменений, соответствующих критериям поиска, в данных экзомного секвенирования не обнаружено. По данным Ехo-С локализована хромосомная транслокация.

## Пациент Р13

**Пол:** Женский

**Дата рождения:** 10.03.2002 (18 лет)

**Дата обследования:** 15.04.2020

### Анамнез заболевания

*10.2019:* На фоне полного здоровья нарушение менструального цикла, слабость, дважды обморочное состояние. Осмотрена гинекологом, DS: [N91.1] Вторичная аменорея. Рекомендован приём препаратов железа. Осмотрена гематологом, DS: [D61.3] Идиопатическая апластическая анемия.

*10.12.2019:* Госпитализирована в гематологическое отделение, проведено лабораторное и инструментальное обследование, трепанобиопсия.

*13.01.2020:* Выписана с DS: [D61.3] Идиопатическая апластическая анемия. Рекомендован приём препаратов железа, наблюдение гинеколога, консультация генетика для исключения болезней накопления. Субъективно состояние не улучшилось.

*02.2020:* Повышение температуры тела до фебрильных цифр, прогрессирующая одышка.

*10.03.2020:* Госпитализирована в терапевтическое отделение.

*01.04.2020:* Выписана переводом в ревматологическое отделение с DS: [D61.3] Идиопатическая апластическая анемия, тяжёлой степени тяжести. [J18.1] Двусторонняя нижнедолевая пневмония на фоне иммунодефицита, ОДН 0. Подозрение на заболевание соединительной ткани. Проведён курс терапии системными ГКС в дозе 35 мг, незначительная положительная динамика в виде нормализации температуры тела.

*02.04.2020:* По эпидемиологическим показаниям переведена в инфекционное отделение (контакт по COVID, анализ отрицательный).

*08.04.2020:* Переведена в психоневрологическое отделение с DS: [J18.8] Двусторонняя пневмония, диссеминированный процесс в лёгких. Получала лечение: гидроксихлорохин, АЦЦ, сорбифер, амоксиклав, азитромицин, бакцефорт.

15.04.2020: Переведена в пульмонологию для лечения в условиях специализированного отделения.

### **Анамнез жизни**

Родилась в Алтайском крае. Образование неоконченное высшее. Материально-бытовые условия удовлетворительные, проживает с родителями. Перелом правого предплечья в анамнезе.

**Генеалогический анализ и наследственность:** Наследственность не отягощена.

**Вредные привычки:** Употребление алкоголя, наркотиков, курение отрицает.

**Оперативные вмешательства:** Отрицает.

**Сопутствующие заболевания:** Вторичная опсоменорея центрального генеза. Гиперпролактинемия. Анемия тяжёлой степени тяжести.

**Эпидемиологический анамнез:** Гепатиты, туберкулёз, ЗППП, паразитозы отрицает.

**Семейно-половой и гинекологический анамнез:** 09.2019: Аменорея. Половой жизнью не живёт.

**Трансфузионный анамнез:** 12.2019: Гемотрансфузия, без осложнений.

**Аллергологический анамнез:** Не отягощён, лекарственную непереносимость отрицает.

### **Лабораторные исследования**

**Биохимический анализ крови от 10.2019:** Нб 89 г/л.

## **Диагноз**

[R50] Лихорадка неясного генеза.

[J18.8] Интерстициальное поражение лёгких неясного генеза, двусторонняя полисегментарная пневмония.

[D61.3] Идиопатическая апластическая анемия, тяжёлой степени тяжести.

[N91.1] Вторичная опсоменорея центрального генеза.

[E22.1] Гиперпролактинемия.

[I82.8] Тромбоз поверхностных вен левого предплечья.

## **Дифференциальный диагноз**

Дифференцировать с редкими генетическими заболеваниями в рамках гемо-бластоза, гемосидероза, идиопатической интерстициальной пневмонии, диффузных заболеваний соединительной ткани, патологий ЦНС.

## ЗАКЛЮЧЕНИЕ

по результатам секвенирования экзона от 03.09.2020 г.

|                       |                                |
|-----------------------|--------------------------------|
| Код пациента:         | P13                            |
| Вид материала:        | Кровь (венозная)               |
| Дата забора:          | 16.04.2020                     |
| Экзомная панель:      | SeqCap EZ MedExome Kit (Roche) |
| Количество прочтений: | 4,0 млн                        |
| Длина прочтений:      | 150 bp                         |
| Выравнивание:         | 95,8 %                         |
| Покрывание экзона:    |                                |
| – $\geq 10$ :         | 20,5 %                         |
| – среднее:            | 7 прочтений/позицию            |

**Заключение:** Значимых изменений, соответствующих критериям поиска, в данных экзомного секвенирования не обнаружено. По совокупности клинических сведений и генетического исследования рекомендована консультация онколога.

## Пациент Р14

**Пол:** Мужской

**Дата рождения:** 07.01.1993 (27 лет)

**Дата обследования:** неизвестна (2020)

### Анамнез заболевания

Рецидивирующая сезонная (январь-март) фебрильная лихорадка, связанная с пребыванием в условиях дикой природы. Першение в горле, кашель, арт-ралгии, оссалгии, миалгии без отёка в период лихорадки. Макуло-папулёзно-узловатые болезненные высыпания на стопах и голенях, некоторые с шелушением.

03.2020: Лимфаденопатия органов брюшной полости, гепатоспленомегалия.

### Анамнез жизни

Данных о бытовых условиях нет. Работает, место работы — республика Саха-Якутия.

**Эпидемиологический анамнез:** Паразитозы — отрицательно многократно.

**Прочие:** Данные по наследственности, вредным привычкам, оперативным вмешательствам, сопутствующим заболеваниям, семейно-половой, трансфузионный и аллергологический анамнез отсутствуют.

### Лабораторные исследования

**АТ к клещевому энцефалиту от 2016:** Положительно.

**АТ к ЦМВ от 01.2019:** IgG и IgM положительны.

**Биохимический анализ крови от 03.2020:** Повышение АСТ, АЛТ, ЩФ, ГГ-ТП, СРБ, фибриногена, холестерина, триглицеридов.

**ОАК от 03.2020:** Незначительная анемия, тромбоцитопения, повышение СОЭ, лейкоцитоз.

Отрицательные многократно: антиДНК, ревматоидный фактор, анти-ЦЦП-АТ, АНЦА, антистрептолизин-О.

Норма многократно: уровни антител, ферритин.

## **Диагноз**

[R50] Лихорадка неясного генеза, сезонного характера (январь-февраль-март), резистентное течение с тяжёлыми обострениями с признаками системного воспалительного ответа.

## **Дифференциальный диагноз**

Дифференцировать с генетическими болезнями в рамках гемобластоза, ревматологических заболеваний (воспалительные артропатии, системный васкулит, диффузные заболевания соединительной ткани), эндокринопатий.



## ЗАКЛЮЧЕНИЕ

**по результатам секвенирования экзона от 03.09.2020 г.**

**Код пациента:** P14

**Вид материала:** Кровь (венозная)

**Дата забора:** 16.04.2020

**Экзомная панель:** SeqCap EZ MedExome Kit (Roche)

**Количество прочтений:** 11,0 млн

**Длина прочтений:** 150 bp

**Выравнивание:** 91,0 %

**Покрывание экзона:**

- $\geq 10$ : 10,0 %
- среднее: 4 прочтений/позицию

| Ген  | Положение (hg19) | Замена | Генотип | Тип замены | Глубина | Ref/Alt |
|--|------------------|--------|---------|------------|---------|---------|
| NLRP12   | chr19:54304654   | G>T    | het     | intronic   | 7       | 3/4     |
| <p><b>OMIM фенотип:</b> [611762] Семейный холододовый аутовоспалительный синдром, АД</p> <p><b>HGMD:</b> —</p> <p><b>Экзон:</b> —</p> <p><b>Частота аллеля:</b> Неизвестна</p> <p><b>Критерии патогенности (согласно российским рекомендациям по интерпретации данных MPS [29]):</b> [PM2] Вариант отсутствует в контрольных выборках.</p> <p><b>Комментарий:</b> Вариант находится вплотную к каноническому сайту сплайсинга 5 экзона. Несмотря на то, что вариант не соответствует критерию PP3, имеющиеся алгоритмы предсказания патогенности интронных и сплайс-вариантов расценивают его как патогенный (dbscSNV, RegSNPItron).</p> |                  |        |         |            |         |         |

**Заключение:** Выявлен не описанный ранее вариант нуклеотидной последовательности в 4 интроне гена *NLRP12* (chr19:54304654G>T) в гетерозиготном состоянии, находящийся в непосредственной близости от канонического сайта сплайсинга. Гетерозиготные варианты в гене *NLRP12* описаны для пациентов с семейным холодовым аутовоспалительным синдромом, АД (OMIM: 611762). Выявленный вариант не зафиксирован в контрольных выборках ESP, 1000 Genomes и ExAC. Несмотря на то, что вариант не соответствует критерию PP3, имеющиеся алгоритмы предсказания патогенности интронных и сплайс-вариантов расценивают его как патогенный (dbSNV, RegSNPItron). Кроме того, вариант требует проверки из-за низкой глубины покрытия экзона. По совокупности сведений, выявленный вариант нуклеотидной последовательности следует считать вариантом с неопределённой клинической значимостью, который может иметь отношение к фенотипу пациента в случае получения дополнительных подтверждающих данных. Других значимых изменений, соответствующих критериям поиска, не обнаружено. Подтверждение по Сэнгеру не проводилось.

## Пациент Р15

**Пол:** Женский

**Дата рождения:** 24.09.2016 (3 года)

**Дата обследования:** 06.07.2020

### Жалобы

Со слов матери — при ходьбе шатается, широко расставляет ноги, перестала есть самостоятельно.

### Анамнез заболевания

*05.2018:* Поступила в неврологическое отделение с DS: [G82.1] Центральный тетрапарез. [R27.0] Атактический синдром. [F82] Задержка психомоторного развития. Последствия перинатального поражения ЦНС. Консультация окулиста: вены глазного дна расширены.

*09.2019:* Повторная госпитализация в неврологическое отделение.

*03.2020:* Поступила в отделение ревматологии с DS: [M13.1] Артрит тазобедренных суставов.

### Анамнез жизни

**Неонатальный период:** Беременность 2, роды 2. Беременность протекала с преэклампсией, хронической фетоплацентарной недостаточностью, угрозой самопроизвольного выкидыша, на фоне хронической урогенитальной инфекции, тазовое предлежание плода. Роды оперативные в срок 39 нед. Вес 3 040 г, рост 53 см, окружность головы 34 см, окружность груди 33 см, Апгар 7/8 баллов. Конъюгационная желтуха I степени до 1 мес.

**Формула развития:** Держит головку с 2 мес., сидит с 7 мес., ползает с 11 мес., ходит с 22 мес. Отдельные слова с 10–11 мес.

**Генеалогический анализ и наследственность:** Наследственность не отягощена.

**Прочие:** Данные по оперативным вмешательствам, сопутствующим заболеваниям, эпидемиологический, трансфузионный и аллергологический анамнез отсутствуют.

### **Клиническое обследование**

Температура тела 36,6 °C, ЧСС 122 уд/мин, ЧДД 29 /мин, АД 80/55 торр, SpO<sub>2</sub> 100 %. Сознание ясное. Инструкции выполняет. В контакт не вступает. Состояние средней степени тяжести. Задержка психоречевого развития. Телосложение нормостеническое.

**Кровеносная система:** Пульс удовлетворительного наполнения и напряжения. Верхушечный толчок в пятом межреберье. Аускультативно тоны сердца ясные, ритмичные. Патологические шумы не выслушиваются.

**Дыхательная система:** Одышки нет. Носовое дыхание свободно, отделяемого нет. Грудная клетка правильной формы. Границы лёгких в пределах нормы. Перкуторно лёгочный звук. Аускультативно везикулярное дыхание, проводится по всем полям. Хрипы не выслушиваются.

**Пищеварительная система:** Видимые слизистые розовой окраски, чистые. Миндалины не увеличены, налета нет. Зев спокоен, симметричен, налета нет. Живот при пальпации мягкий, безболезненный. Печень по краю реберной дуги. Аускультативно: выслушивается перистальтика кишечника. Стул регулярный, оформленный.

**Мочевыделительная система:** Мочеиспускание свободное, безболезненное. Почки не пальпируются.

**Нервная система:** Череп правильной формы, при перкуссии безболезненный. Лицо симметричное. Мимика симметрична. ЧМН: глазные щели D=S, зрачки округлой формы D=S, фотореакции сохранены, косоглазие и нистагм отсутствуют, объем движения глазных яблок полный. Слух сохранен. Глоточный, небный рефлексы сохранены. Язык по средней линии. Сила в конечностях 4–5 баллов. Мышечный тонус повышен. Сухожильные рефлексы средней живости, D=S. Патологические знаки, менингеальные знаки отсутствуют. Расстройств чувствительности нет. Функция тазовых органов не нарушена. Опора на полную стопу. Походка атактическая, широко расставляет ноги, пошатывается. Легкие когнитивные нарушения.

**Кровотворная система:** Селезенка не пальпируется.

**Эндокринная система:** Щитовидная железа не увеличена.

**Костно-мышечная система:** Без особенностей.

**Местные изменения:** Без особенностей.

## Лабораторные исследования

**Биохимический анализ крови от 05.2020:** КФК 2 282 Ед/л.

**ОАК от 07.07.2020:** RBC  $5,42 \cdot 10^{12}$  кл/л, HGB 128 г/л, PLT  $332 \cdot 10^9$  кл/л, WBC  $6,09 \cdot 10^9$  кл/л, нейтрофилы (п/я) 1 %, нейтрофилы (с/я) 43 %, лимфоциты 47 %, моноциты 9 %.

**Биохимический анализ крови от 07.07.2020:** Общий белок 65 г/л, СРБ 0, мочевины 4,3 ммоль/л, креатинин 57,0 мкмоль/л, ЛДГ 407,0 Ед/л, КФК 104,0 Ед/л,  $K^+$  4,9 ммоль/л,  $Ca^{2+}$  2,4 ммоль/л,  $Na^+$  137 ммоль/л, общий билирубин 11,0 ммоль/л, АЛТ 14 Ед/л, АСТ 30 Ед/л, глюкоза 4,5 ммоль/л, ЩФ 644,0 Ед/л.

**ОАМ от 07.07.2020:** Удельный вес 1 030 г/л, белок, сахар, ацетон отрицательно, плоский эпителий 4–5 в п/з, лейкоциты 15–20 в п/з, эритроциты отрицательно, цилиндры гиалиновые 0–1 в п/з, слизь положительно, соли отрицательно, бактерии положительно.

**Соскоб на энтеробиоз от 07.07.2020:** Отрицательно.

**Кал на яйца гельминтов от 09.07.2020:** Отрицательно.

**ЭКГ от 09.07.2020:** Умеренная синусовая брадикардия (ЧСС 85–93 уд/мин).  
Неполная блокада правой ножки пучка Гиса.

**ЭЭГ от 09.07.2020:** Умеренные диффузные изменения биоэлектрической активности головного мозга. Межполушарной асимметрии и патологической активности не выявлено.

**ЭНМГ конечностей от 09.07.2020:** : поражения мотонейронов спинного мозга не выявлено. Легкая аксонопатия *n. peroneus* с обеих сторон.

**МСКТ головного мозга от 07.2020:** Данные за патологию отсутствуют.

**МРТ головного мозга от 07.2020:** Киста эпифиза.

### **Консультация специалистов**

**Окулист:** Патологии глазного дна не выявлено.

**Физиотерапевт:** Массаж обеих ног и пояснично-крестцовой области, СМП-терапия на речевые зоны, магнитотерапия на ПОП.

**Педиатр:** Острой соматической патологии не выявлено.

**Травматолог-ортопед:** Плосковальгусная постановка стоп.

### **Диагноз**

[G82.1] Центральный тетрапарез.

[R27.0] Атактический синдром.

[F82] Задержка психомоторного развития.

### **Дифференциальный диагноз**

Дифференцировать с редкими генетическими заболеваниями.

## **ЗАКЛЮЧЕНИЕ**

**по результатам секвенирования экзона от 03.09.2020 г.**

|                              |                                |
|------------------------------|--------------------------------|
| <b>Код пациента:</b>         | P15                            |
| <b>Вид материала:</b>        | Кровь (венозная)               |
| <b>Дата забора:</b>          | 18.08.2020                     |
| <b>Экзомная панель:</b>      | SeqCap EZ MedExome Kit (Roche) |
| <b>Количество прочтений:</b> | 12,8 млн                       |
| <b>Длина прочтений:</b>      | 150 bp                         |
| <b>Выравнивание:</b>         | 95,49 %                        |
| <b>Покрывание экзона:</b>    |                                |
| – $\geq 10$ :                | 60,63 %                        |
| – среднее:                   | 15 прочтений/позицию           |

| Ген   | Положение (hg19) | Замена | Генотип | Тип замены | Глубина | Ref/Alt |
|---|------------------|--------|---------|------------|---------|---------|
| SCN4A   | chr17:62019110   | G>A    | het     | nonsyn SNV | 18      | 12/6    |
| <p><b>OMIM фенотип:</b> [170500] Гиперкалемический периодический паралич. [613345] Гипокалемический эпизодический паралич 2. [608390] Калий-зависимая миотония. [168300] Врожденная миотония фон Эйленбурга.</p> <p><b>HGMD:</b> —</p> <p><b>Экзон:</b> 24/24</p> <p><b>Частота аллеля:</b> [rs377548921] 0.000008</p> <p><b>Критерии патогенности (согласно российским рекомендациям по интерпретации данных MPS [29]):</b> [PM1] Вариант расположен в «горячей» точке и/или важных и хорошо исследованных функциональных доменах белка (например, активный сайт фермента), в которых не описаны доброкачественные изменения (Ion transport domain). [PM2] Вариант отсутствует в контрольных выборках 1000 Genomes, ESP, ExAC. [PP3] Результаты не менее трёх программ предсказания патогенности <i>in silico</i> подтверждают патогенное действие варианта на ген или продукт гена (MetaSVM, GERP++, dbSNV).</p> <p><b>Комментарий:</b> —</p> |                  |        |         |            |         |         |

**Заключение:** Выявлен описанный ранее вариант нуклеотидной последовательности в 24 экзоне гена *SCN4A* (chr17:62019110G>A) в гетерозиготном состоянии, приводящий к замене аланина на аспарат в 1511 кодоне (NM\_000334c.4532C>T (p.S1511L)). Гетерозиготные варианты в гене *SCN4A* описаны для пациентов со следующими заболеваниями: гиперкалемический периодический паралич (OMIM: 170500), гипокалемический эпизодический паралич 2 (OMIM: 613345), калий-зависимая миотония (OMIM: 608390), врожденная миотония фон Эйленбурга (OMIM: 168300). Выявленный вариант не зафиксирован в контрольных выборках ESP, 1000 Genomes и ExAC, данная позиция является частью важного функционального домена (Ion transport domain). Алгоритмы предсказания патогенности расценивают его как патогенный (MetaSVM,



GERP++, dbSCSNV), алгоритмы предсказания консервативности и базы данных консервативности считают вариант консервативным (PhyloP), доброкачественные замены аминокислоты в паралогах млекопитающих в данной позиции не описаны. По совокупности сведений, выявленный вариант нуклеотидной последовательности следует считать вариантом с неопределённой клинической значимостью, который может иметь отношение к фенотипу пациента в случае получения дополнительных подтверждающих данных. Других значимых изменений, соответствующих критериям поиска, не обнаружено. Подтверждение по Сэнгеру не проводилось.

## Пациент Р16

**Пол:** Женский

**Дата рождения:** 30.09.2018 (1 год)

**Дата обследования:** 11.11.2019

### Жалобы

Не держит голову, не переворачивается, не сидит, не ползает. Активно двигает ногами. Низкий аппетит.

### Анамнез жизни

**Неонатальный период:** Беременность 4, роды 2. Сибс здоров. Беременность протекала на фоне угрозы самопроизвольного выкидыша, хронической урогенитальной инфекции, хронической внутриутробной гипоксии плода, хронической фетоплацентарной недостаточности, синдрома задержки развития плода. Роды в срок 39 нед., самопроизвольные. Вес 2 860 г, длина тела 49 см, окружность головы 33 см, окружность грудной клетки 30 см. Оценка по шкале Апгар 7/8 баллов. DS: [P39.2] Внутриамниотическая инфекция плода неуточненной этиологии [P91.0] Ишемия мозга новорожденного легкой степени. [P59.9] Конъюгационная желтуха II ст..

**Генеалогический анализ и наследственность:** Брак 1, родственность брака отрицается.

**Сопутствующие заболевания:** 2020: Буллезное импетиго средней степени тяжести.

**Аллергологический анамнез:** Лекарственная аллергия на пантогам.

**Прочие:** Данные по оперативным вмешательствам, эпидемиологический и трансфузионный анамнез отсутствуют.

## Клиническое обследование

Сознание ясное. Положение активное. Микросомия. Вес 6 кг, рост 68 см.

**Нервная система:** Сон спокойный. Не закатывается, обмороков, замираний нет. Черепные нервы: без особенностей. Зрачки симметричные, фотореакция живая, симметричная, нистагм отсутствует. Носогубные складки симметричные. Язык по средней линии. Глотание, фонация, голос не нарушены. Мышечный тонус повышен в сгибателях рук, понижен в нижних конечностях. Сила в руках и ногах 3 балла. Сухожильные рефлексы с рук и ног высокие, симметричные. Симптом Бабинского с двух сторон. При тракции за руки удерживает голову лучше. Группировки нет. В положении на животе опирается на предплечья, голову удерживает недолго. Опора на полную стопу с вальгусным компонентом. Менингеальные знаки отсутствуют. Гиперкинезы отсутствуют.

**Костно-мышечная система:** Голова правильной формы, окружность 41 см, швы закрыты. Рекурвация коленных суставов.

**Органы чувств:** Страбизм.

**Прочие:** Данные по кровеносной, дыхательной, пищеварительной, мочевыделительной, кроветворной и эндокринной системам отсутствуют.

## Лабораторные исследования

**Нейросонография от 02.2019:** Структурные и гемодинамические нарушения отсутствуют.

**МСКТ головного мозга от 02.2019:** Без патологий.

**ЭЭГ от 02.2019:** Умеренные диффузные изменения биоэлектрической активности головного мозга, эпилептиформная активность не обнаружена.

**Электронейромиография от 02.2019:** Поражения мотонейронов спинного мозга не обнаружено, миелопатия умеренной степени моторных и сенсорных волокон верхних и нижних конечностей.

**Тандемная масс-спектрометрия от 29.07.2019:** Данных за наследственные аминокислотопатии, органические ацидурии, дефекты митохондриального  $\beta$ -окисления не выявлено.

**Энзимодиагностика от 30.08.2019:** Активность лизосомальных ферментов в пределах референсных значений.

**FISH-анализ от 17.10.2019:** Микроделеции не обнаружены.

**Кариотипирование от 01.2020:** Хромосомной патологии не обнаружено.

### **Диагноз**

[G93.0] Органическое поражение ЦНС генетической этиологии.

[G82.1] Центральный тетрапарез.

[F82] Задержка психомоторного развития средней степени тяжести.

### **Дифференциальный диагноз**

Дообследование у генетика.

## ЗАКЛЮЧЕНИЕ

**по результатам секвенирования экзона от 03.09.2020 г.**

**Код пациента:** P16

**Вид материала:** Кровь (венозная)

**Дата забора:** 29.04.2020

**Экзомная панель:** SeqCap EZ MedExome Kit (Roche)

**Количество прочтений:** 23,0 млн

**Длина прочтений:** 150 bp

**Выравнивание:** 97,3 %

**Покрывание экзона:**

- $\geq 10$ : 79,35 %
- среднее: 26 прочтений/позицию

| Ген  | Положение (hg19) | Замена | Генотип | Тип замены | Глубина | Ref/Alt |
|--|------------------|--------|---------|------------|---------|---------|
| TUBB4A   | chr19:6495332    | G>A    | het     | nonsyn SNV | 52      | 24/28   |
| <p><b>OMIM фенотип:</b> [612438] Гипомиелинизирующая лейкодистрофия 6, AD</p> <p><b>HGMD:</b> —</p> <p><b>Экзон:</b> 4/4</p> <p><b>Частота аллеля:</b> Неизвестна</p> <p><b>Критерии патогенности (согласно российским рекомендациям по интерпретации данных MPS [29]):</b> [PP3] Результаты не менее трёх программ предсказания патогенности <i>in silico</i> подтверждают патогенное действие варианта на ген или продукт гена (MetaSVM, GERP++, dbSNV). [PM1] Вариант расположен в «горячей точке» и/или важных и хорошо исследованных функциональных доменах белка (например, активный сайт фермента), в которых не описаны доброкачественные изменения (С-терминальный конец тубулина). [PM2] Вариант отсутствует в контрольных выборках (ESP, 1000 Genomes и ExAC).</p> <p><b>Комментарий:</b> —</p> |                  |        |         |            |         |         |

**Заключение:** Выявлен ранее не описанный вариант нуклеотидной последовательности в 4 экзоне гена *TUBB4A* (chr19:6495332) в гетерозиготном состоянии, приводящий к замене аланина на аспарат в 393 кодоне (NM\_006087.4:p.Ala393Asp). Гетерозиготные варианты в гене *TUBB4A* описаны для пациентов с гипомиелинизирующей лейкодистрофией 6, AD, (OMIM: 612438). Выявленный вариант не зафиксирован в контрольных выборках ESP, 1000 Genomes и ExAC, данная позиция является частью важного функционального домена (С-терминальный конец тубулина), мутации в данном кодоне не описаны. Алгоритмы предсказания патогенности расценивают его как патогенный (MetaSVM, GERP++, dbSNV), алгоритмы предсказания консервативности и базы данных консервативности считают вариант консервативным (PhyloP), доброкачественные замены аминокислоты в паралогах позвоночных в данной позиции не описаны. По совокупности сведений, выявленный вариант нуклеотидной последовательности следует считать вариантом с неопределённой клинической значимостью, который может иметь отношение к фенотипу пациента в случае получения дополнительных подтверждающих данных. Других значимых изменений, соответствующих критериям поиска, не обнаружено. Подтверждение по Сэнгеру не проводилось.

## Пациент Р17

**Пол:** Мужской

**Дата рождения:** 28.05.2015 (4 года)

**Дата обследования:** 06.03.2020

### Жалобы

Кожные высыпания с зудом, заложенность носа, снижение слуха, влажный утренний кашель.

### Анамнез заболевания

*10.07.2015:* Госпитализация с DS: [L20.8] Атопический дерматит, младенческая форма, распространенная, острое течение. Подозрение на ихтиоз. Обострения 1–2 раза в год.

*10.09.2015:* Подозрение на дисхолию.

*2016:* Осмотр ортопеда, DS: [Q68.4] Варусное искривление костей нижних конечностей.

*30.03.2016:* Осмотр невролога, DS: [P21.9] Аноксическое поражение головного мозга, гипоксически-ишемическая энцефалопатия. [P91.3] Церебральная возбудимость новорожденного.

Заложенность носа в течение более 10 дней на фоне излеченного конъюнктивита, лечат деконгестантами.

### Анамнез жизни

Проживает в квартире. Домашние животные — кошка неуточненная. Декоративные растения отсутствуют. Наличие плесени отрицает. Наличие курящих в семье отрицает.

**Неонатальный период:** Беременность 1, роды 1. Родился в сроке 41 нед. Из роддома выписан с DS: [P83.5] Врожденное двустороннее гидроцеле. [P59.9] Неонатальная желтуха.

Вес 4 630 г, длина тела 57 см, окружность головы 36 см, окружность грудной клетки 37 см. Оценка по шкале Апгар 8/8 баллов. Нормы психического развития по возрасту.

**Генеалогический анализ и наследственность:** Прабабушка по линии отца — бронхиальная астма, неуточненная пищевая аллергия. Дядя — бронхиальная астма.

**Оперативные вмешательства:** Не было.

**Сопутствующие заболевания:** Фарингомикоз.

**Эпидемиологический анамнез:** От вакцинации отказались по идейным соображениям.

**Трансфузионный анамнез:** Данные отсутствуют.

**Аллергологический анамнез:** Пищевая аллергия (лесной орех, арахис), аллергия на препараты — аскорбиновая кислота, сенсibilизация на пылевого клеща, собачью шерсть.

## **Клиническое обследование**

Состояние удовлетворительное. Температура 36,6 °С, ЧСС 95 уд/мин, ЧДД 23 /мин.

**Кровеносная система:** Тоны сердца ясные, ритмичные.

**Дыхательная система:** Носовое дыхание свободное. В легких дыхание везикулярное, проводится по всем полям, хрипов нет. Грудная клетка участвует в акте дыхания равномерно, обычной формы, перкуторно легочный звук.

**Пищеварительная система:** Миндалины рыхлые, увеличены, нитевидный налет. Живот при пальпации мягкий, безболезненный. Стул в норме.



**Мочевыделительная система:** Мочеиспускание в норме.

**Костно-мышечная система:** Телосложение правильное.

**Органы чувств:** Инъецирования сосудов склер нет. Пробка наружного слухового прохода справа.

**Местные изменения:** Кожные покровы нормальной окраски, выраженная диффузная сухость кожи с диффузным среднепластинчатым шелушением, в области ВЧГ с отрубевидным шелушением. Периферические лимфоузлы не увеличены.

**Прочие:** Данные по нервной, кроветворной и эндокринной системам отсутствуют.

### **Лабораторные исследования**

**Аллергопробы от 2017:** IgE положительны на все пищевые аллергены.

### **Диагноз**

[Q80.9] Врожденный ихтиоз неуточненный.

[B49] Фарингомикоз неуточнённый.

### **Дифференциальный диагноз**

Дифференцировать с прочими редкими генетическими заболеваниями.

## ЗАКЛЮЧЕНИЕ

**по результатам секвенирования экзона от 03.09.2020 г.**

**Код пациента:** P17

**Вид материала:** Кровь (венозная)

**Дата забора:** 29.04.2020

**Экзомная панель:** SeqCap EZ MedExome Kit (Roche)

**Количество прочтений:** 20,8 млн

**Длина прочтений:** 150 bp

**Выравнивание:** 96 %

**Покрывание экзона:**

- $\geq 10$ : 64 %
- среднее: 16 прочтений/позицию

| Ген  | Положение (hg19) | Замена | Генотип | Тип замены | Глубина | Ref/Alt |
|--|------------------|--------|---------|------------|---------|---------|
| ALOXE3   | chr17:8015495    | G>A    | hom     | stopgain   | 5       | 1/4     |
| <p><b>OMIM фенотип:</b> [606545] Врождённый ихтиоз 3, AR</p> <p><b>HGMD:</b> [CM020010] Небуллёзная ихтиозиформная эритродерма 1 типа</p> <p><b>Экзон:</b> 7/16</p> <p><b>Частота аллеля:</b> [rs121434233] 0,000 38</p> <p><b>Критерии патогенности (согласно российским рекомендациям по интерпретации данных MPS [29]):</b> [PM2] Вариант отсутствует в контрольных выборках (или встречается с крайне низкой частотой). [PVS1] LOF-варианты, если таковые являются известной причиной заболевания. [PP3] Результаты не менее трёх программ предсказания патогенности <i>in silico</i> подтверждают патогенное действие варианта на ген или продукт гена. [PP5] Источники с хорошей репутацией указывают на патогенность варианта, но независимая оценка не проводилась. [PP4] Фенотип пациента или семейная история высоко специфичны для заболевания с данной наследственной этиологией.</p> <p><b>Комментарий:</b> —</p> |                  |        |         |            |         |         |

**Заключение:** Выявлен описанный ранее у больных вариант нуклеотидной последовательности в 7 экзоне гена *ALOXE3* (chr17:8015495G>A, dbSNP: rs121434233, HGMD: CM020010) в гомозиготном состоянии, приводящий к обрыву трансляции на 366 кодоне (NP\_001159432.1:p.Arg366Ter). Гомозиготные варианты, приводящие к обрыву трансляции в гене *ALOXE3*, описаны для пациентов с врождённым ихтиозом 3, AR (OMIM: 606545). Частота выявленного варианта нуклеотидной последовательности в контрольной выборке gnomAD составляет 0,038 % (12 мутантных аллелей на 31 394 хромосом). По совокупности сведений, выявленный вариант нуклеотидной последовательности следует расценивать как патогенный. Патогенность варианта подтверждается следующими публикациями (PubMed ID): 11773004, 16116617, 19131948. Других значимых изменений, соответствующих критериям поиска, не обнаружено. Подтверждение по Сэнгеру не проводилось.

## Пациент Р18

**Пол:** Женский

**Дата рождения:** 15.10.1988 (31 год)

**Дата обследования:** 17.03.2020

### Жалобы

Сухость и шелушение кожи. Постепенное снижение зрения. Повторяющиеся вывихи крупных суставов.

### Анамнез жизни

**Неонатальный период:** Родоразрешение в 7 мес. После рождения кома в течение 3 суток. Неонатальная пневмония. Пластинчатое шелушение кожи. DS: [Q80.9] Ихтиоз.

**Генеалогический анализ и наследственность:** У матери — сахарный диабет II типа.

**Оперативные вмешательства и травмы:** Частые вывихи коленных суставов (3 вывиха правого, 5 — левого). Вывих локтевого сустава. Оперативные вмешательства отрицает.

**Сопутствующие заболевания:** Хронический гастрит, дуоденогастральный рефлюкс (со слов).

**Гинекологический анамнез:** Беременностей не было, менструальный цикл регулярный, 28 дней.

**Прочие:** Эпидемиологический, трансфузионный и аллергологический анамнез отсутствуют.

## Клиническое обследование

Общее состояние удовлетворительное.

**Пищеварительная система:** Видимые слизистые без особенностей. Язык без налета. Живот нормального размера, симметричен, при пальпации мягкий, болезненность при пальпации в эпигастрии и околопупочной области. Перистальтика активная. Край печени плотно-эластический, безболезненный, по краю реберной дуги. Стул регулярный, оформленный, без патологических примесей.

**Мочевыделительная система:** Диурез в норме.

**Кроветворная система:** Селезенка не пальпируется.

**Эндокринная система:** Щитовидная железа пальпаторно не увеличена, неоднородная справа.

**Органы чувств:** Конъюктива глаз не изменена, роговица прозрачная, радужная оболочка структурная, хрусталик прозрачный, ДЗН бледно-розовый, границы четкие. Сетчатка розовая, без патологии. Острота зрения: D 0,5, S 0,6.

**Местные изменения:** кожные покровы сухие, гиперкератоз. Подкожный жировой слой выражен умеренно. Отеков нет.

**Прочие:** Данные по кровеносной, дыхательной, нервной и костно-мышечной системам отсутствуют.

## Лабораторные исследования

**Биохимия крови от 02.2018:** Билирубин 46,1 мкмоль/л.

**УЗИ брюшной полости от 15.01.2020:** Диффузные изменения поджелудочной железы.

**Биохимия крови от 02.03.2020:** глюкоза натощак 4,4 ммоль/л, через 2 ч. 3,5 ммоль/л.

**Биопсия кожи от 23.04.2020:** Морфологические изменения кожи характерны для генодерматозов, в том числе ламеллярного ихтиоза. Морфологические признаки ихтиозиформной эритродермии отсутствуют.

### **Диагноз**

[K21] Гастроэзофагальная рефлюксная болезнь.

[K30] Функциональная диспепсия.

[K58.9] Синдром раздраженного кишечника.

[E80.4] Синдром Жильбера.

[E07.9] Болезнь щитовидной железы неуточненная.

[H52.2] Астигматизм обоих глаз смешанный.

[Q80.9] Ихтиоз неуточненный.

### **Дифференциальный диагноз**

Консультация генетика.

## ЗАКЛЮЧЕНИЕ

по результатам секвенирования экзона от 20.10.2020

|                       |                                |
|-----------------------|--------------------------------|
| Код пациента:         | P18                            |
| Вид материала:        | Кровь (венозная)               |
| Дата забора:          | 29.04.2020                     |
| Экзомная панель:      | SeqCap EZ MedExome Kit (Roche) |
| Количество прочтений: | 20,9 млн                       |
| Длина прочтений:      | 150 bp                         |
| Выравнивание:         | 97,35 %                        |
| Покрывание экзона:    |                                |
| – $\geq 10$ :         | 72,46 %                        |
| – среднее:            | 21 прочтений/позицию           |

**Заключение:** Значимых изменений, соответствующих критериям поиска, в данных экзомного секвенирования не обнаружено.

## Пациент Р19

**Пол:** Мужской

**Дата рождения:** Неизвестна

**Дата обследования:** 25.09.2019

### Анамнез жизни

**Неонатальный период:** DS: [P70.1] Диабетическая фетопатия.

**Генеалогический анализ и наследственность:** Со слов матери — в семье по линии матери пробанда 5 инвалидов мужского пола, DS: [G80.9] Детский церебральный паралич. Старший сибс — DS: [G80.9] Детский церебральный паралич.

**Прочие:** Данные по оперативным вмешательствам, сопутствующим заболеваниям, эпидемиологический, трансфузионный и аллергологический анамнез отсутствуют.

### Клиническое обследование

**Нервная система:** Самостоятельно голову не удерживает. Гипертонус верхних и нижних конечностей.

**Костно-мышечная система:** Короткая, широкая шея.

**Прочие:** Данные по кровеносной, дыхательной, пищеварительной, мочевыделительной, кроветворной и эндокринной системам, а также органам чувств отсутствуют.

### Лабораторные исследования

**УЗИ головного мозга от 17.10.2019:** Расширение задних рогов боковых желудочков, субарахноидальное и межполушарное пространства не расширены.



Субэпендимальная псевдокиста слева. Постгипоксические изменения — перивентрикулярная ишемия, зернистость. Кисты сосудистых сплетений. Минерализирующая васкулопатия. Периферические сосудистые сопротивления в бассейне передней ПМА и ОА в пределах возрастных нормативов. Венозный отток по *v. galeni* не нарушен.

## **Диагноз**

[Q04.9] Врожденная аномалия головного мозга неуточненная.

## **Дифференциальный диагноз**

Провести генетический поиск редких заболеваний, сцепленных с X-хромосомой.

## ЗАКЛЮЧЕНИЕ

**по результатам секвенирования экзона от 28.08.2020 г.**

**Код пациента:** P19

**Вид материала:** Кровь (венозная)

**Дата забора:** 29.04.2020

**Экзомная панель:** SeqCap EZ MedExome Kit (Roche)

**Количество прочтений:** 18,6 млн

**Длина прочтений:** 150 bp

**Выравнивание:** 97,5 %

**Покрывание экзона:**

- $\geq 10$ : 68,3 %
- среднее: 17 прочтений/позицию

| Ген   | Положение (hg19) | Замена | Генотип | Тип замены | Глубина | Ref/Alt |
|---|------------------|--------|---------|------------|---------|---------|
| PLP1  | chrX:103044331   | A>G    | hem     | intronic   | 30      | 4/26    |
| <p><b>OMIM фенотип:</b> [312080] Болезнь Пелизеуса—Мерцбахера, XLR (PMD). [312920] Спастическая параплегия 2, XLR (SPG2)</p> <p><b>HGMD:</b> —</p> <p><b>Экзон:</b> 6/7</p> <p><b>Частота аллеля:</b> Неизвестно</p> <p><b>Критерии патогенности (согласно российским рекомендациям по интерпретации данных MPS [29]):</b> [PM2] Вариант отсутствует в контрольных выборках (или встречается с крайне низкой частотой) [PP3] Результаты не менее трёх программ предсказания патогенности <i>in silico</i> подтверждают патогенное действие варианта на ген или продукт гена.</p> <p><b>Комментарий:</b> Вариант находится на расстоянии 1 bp от канонического сплайс-сайта. Тем не менее, программы предсказания патогенности <i>in silico</i> считают его повреждающим сплайсинг. Анализ референсных прочтений показал, что они являются артефактом.</p> |                  |        |         |            |         |         |

**Заключение:** Выявлен ранее не описанный вариант нуклеотидной последовательности у границы 4 экзона в гене *PLP1* (chrX:103044331A>G) в гемизиготном состоянии, предположительно приводящий к нарушению процесса сплайсинга транскрипта. Мутации в гене *PLP1* в гомо- или гемизиготном состоянии описаны у пациентов с болезнью Пелизеуса—Мерцбахера (OMIM: 312080). Выявленный вариант не зафиксирован в контрольных выборках 1000Genomes и gnomAD. Алгоритмы предсказания патогенности расценивают его как патогенный (dbSNP, RegSNPintron). По совокупности сведений, выявленный вариант нуклеотидной последовательности следует считать вариантом с неопределённой клинической значимостью, который может иметь отношение к фенотипу пациента в случае получения дополнительных подтверждающих данных. Других значимых изменений, соответствующих критериям поиска, не обнаружено. Подтверждение по Сэнгеру не проводилось.

## Пациент Р20

**Пол:** Женский

**Дата рождения:** 02.01.1977 (43 года)

**Дата обследования:** 04.01.2020

### Жалобы

Слабость, потливость. Психогенная дисфония и дизартрия. Чередование запоров и поносов, кровавые прожилки в кале.

### Анамнез заболевания

30.01.2020: Консультация ревматолога, DS: [M35] Недифференцированное заболевание соединительной ткани.

### Анамнез жизни

Часто задыхалась в детстве, бронхоэктазы.

**Неонатальный период:** Данные отсутствуют.

**Генеалогический анализ и наследственность:** У обоих детей схожая симптоматика (пациенты Р21 и Р22).

**Оперативные вмешательства:** 1988 — удалена  $\frac{1}{3}$  легкого.

**Сопутствующие заболевания:** Синдром раздраженного кишечника. Психогенная дисфония и дизартрия.

**Эпидемиологический анамнез:** Перенесла краснуху в детском возрасте.

**Трансфузионный анамнез:** Данные отсутствуют.

**Аллергологический анамнез:** Отеки Квинке на неизвестный аллерген. Непереносимость неуточненных антибиотиков.

### **Клиническое обследование**

Вес 70 кг.

Данные клинического обследования по системам отсутствуют.

### **Лабораторные исследования**

**Биохимический анализ крови от 01.02.2018:** Фолиевая кислота 1,5 нг/мл, витамин В12 139 пг/мл.

**МРТ ГМ от 04.01.2020:** МР картина наружной гидроцефалии. Немногочисленные очаговые изменения вещества лобных долей, вероятно, дисциркуляторно-дистрофического характера.

### **Диагноз**

[G98] Неуточненные нарушения нервной системы, не классифицированные в других рубриках.

[M35] Недифференцированное заболевание соединительной ткани.

### **Дифференциальный диагноз**

Дифференцировать с другими редкими генетическими заболеваниями, связанными с нарушением метаболизма и патологиями нервной системы.

## ЗАКЛЮЧЕНИЕ

по результатам секвенирования экзона от 08.09.2020 г.

|                       |                                |
|-----------------------|--------------------------------|
| Код пациента:         | P20                            |
| Вид материала:        | Кровь (венозная)               |
| Дата забора:          | 29.04.2020                     |
| Экзомная панель:      | SeqCap EZ MedExome Kit (Roche) |
| Количество прочтений: | 22,6 млн                       |
| Длина прочтений:      | 150 bp                         |
| Выравнивание:         | 97,47 %                        |
| Покрывание экзона:    |                                |
| – $\geq 10$ :         | 61,92 %                        |
| – среднее:            | 16 прочтений/позицию           |

**Заключение:** При индивидуальном и семейном анализе (пациенты P20, P21, P22) значимых изменений, соответствующих критериям поиска, не обнаружено.

## **Пациент Р21**

**Пол:** Женский

**Дата рождения:** Неизвестна

**Дата обследования:** Неизвестна (2020)

### **Анамнез заболевания**

Со слов лечащего врача — симптомы схожи с сибсом (пациент Р22) и матерью (пациент Р20). Клинической информации по пациенту нет.

## ЗАКЛЮЧЕНИЕ

по результатам секвенирования экзона от 08.09.2020 г.

|                       |                                |
|-----------------------|--------------------------------|
| Код пациента:         | P21                            |
| Вид материала:        | Кровь (венозная)               |
| Дата забора:          | 29.04.2020                     |
| Экзомная панель:      | SeqCap EZ MedExome Kit (Roche) |
| Количество прочтений: | 25,7 млн                       |
| Длина прочтений:      | 150 bp                         |
| Выравнивание:         | 97,27 %                        |
| Покрывание экзона:    |                                |
| – $\geq 10$ :         | 72,85 %                        |
| – среднее:            | 21 прочтений/позицию           |

**Заключение:** При индивидуальном и семейном анализе (пациенты P20, P21, P22) значимых изменений, соответствующих критериям поиска, не обнаружено.



## Пациент Р22

**Пол:** Мужской

**Дата рождения:** 06.06.2012 (7 лет)

**Дата обследования:** 04.01.2020

### Жалобы

Задержка развития речи. Стресс-индуцированная дизартрия. Ацетон в моче на фоне беспричинной рвоты.

### Анамнез жизни

**Генеалогический анализ и наследственность:** У матери (пациент Р20) и сибса (пациент Р21) схожие симптомы.

**Эпидемиологический анамнез:** Неутонченная детская инфекция с последствиями в виде регресса речи и практических навыков.

**Аллергологический анамнез:** Поливалентная аллергия — лекарственные препараты, пищевые продукты, укусы насекомых.

**Прочие:** Данные по неонатальному периоду развития, оперативным вмешательствам, сопутствующим заболеваниям, а также трансфузионный анамнез отсутствуют.

### Клиническое обследование

**Костно-мышечная система:** Умеренная гипермобильность суставов верхних конечностей.

**Прочие:** Данные по кровеносной, дыхательной, пищеварительной, мочевыделительной, нервной, кроветворной и эндокринной системам, а также органам чувств отсутствуют.

## Лабораторные исследования

**Электромиография от 13.07.2016:** Снижение показателей электрогенеза мышц голени и, более выражено, мышц бедра. Структура записи не изменена.

**Биохимия крови от 09.08.2017:**  $\text{Na}^+$  135 ммоль/л.

**ОАК от 09.08.2017:** Эритроциты  $5,07 \cdot 10^{12}$  кл/л, MCV 73,8 фл, нейтрофилы 27,8 %, лимфоциты 62,4 %.

**УЗИ брюшной полости от 07.10.2017:** Эхоскопические признаки увеличения и незначительной деформации желчного пузыря, незначительного усиления сосудистого рисунка, незначительного уплотнения стенок отдельных внутрипеченочных протоков. Архитектоника почек не нарушена. Гепатоспленомегалия.

**Биохимический анализ мочи от 04.02.2018:** Органические кислоты в моче пациента в пределах нормы.

**Таргетная панель на миопатии от 18.06.2019:** Клинически значимых вариантов не обнаружено.

## Диагноз

[G98] Неуточненные нарушения нервной системы.

## Дифференциальный диагноз

Дифференцировать с другими редкими генетическими заболеваниями, связанными с нарушением метаболизма и патологиями нервной системы.

## ЗАКЛЮЧЕНИЕ

по результатам секвенирования экзона от 08.09.2020 г.

|                       |                                |
|-----------------------|--------------------------------|
| Код пациента:         | P22                            |
| Вид материала:        | Кровь (венозная)               |
| Дата забора:          | 29.04.2020                     |
| Экзомная панель:      | SeqCap EZ MedExome Kit (Roche) |
| Количество прочтений: | 19,5 млн                       |
| Длина прочтений:      | 150 bp                         |
| Выравнивание:         | 97,39 %                        |
| Покрывание экзона:    |                                |
| – $\geq 10$ :         | 54,7 %                         |
| – среднее:            | 14 прочтений/позицию           |

**Заключение:** При индивидуальном и семейном анализе (пациенты P20, P21, P22) значимых изменений, соответствующих критериям поиска, не обнаружено.

## Пациент Р23

**Пол:** Мужской

**Дата рождения:** 17.04.2017 (2 года)

**Дата обследования:** 02.09.2019

### Жалобы

Грубая задержка психомоторного развития. Формирование контрактур в конечностях.

### Анамнез заболевания

Фебрильные приступы в период новорожденности.

07.2017: Оперирован по поводу окклюзионной гидроцефалии с установкой вентрикулоперитонеального шунта.

01.2018: Поступил в нейрохирургическое отделение, назначен Конвулекс.

03.02.2018: Извлечение абдоминального катетера, установка наружных дренажей. Реконвалесценция вентрикулита.

01.2019: Консультация пульмонолога, DS: [P27.1] Бронхолегочная дисплазия, классическая форма, вне обострения, ДН 0. Оформлен паллиативный статус.

### Анамнез жизни

**Неонатальный период:** Роды в сроке 26 нед.

**Прочие:** Данные по наследственности, оперативным вмешательствам, сопутствующим заболеваниям, эпидемиологический, трансфузионный и аллергологический анамнез отсутствуют.

### Клиническое обследование

Температура тела — 36,6 °С, ЧСС — 100 уд/мин, ЧДД — 24 /мин. Состояние тяжелой степени тяжести, обусловлено основным заболеванием. На осмотр

реагирует моторным возбуждением. Периодически плач.

**Кровеносная система:** Сердечные тоны ритмичны.

**Дыхательная система:** Аускультативно дыхание жёсткое, хрипов нет.

**Пищеварительная система:** Отмечается зернистость и гиперемия зева. Печень не увеличена. Живот мягкий. Дефекация в норме.

**Мочевыделительная система:** Мочеиспускание в норме.

**Нервная система:** Не сидит, множественные контрактуры.

**Кроветворная система:** Селезенка не увеличена.

**Эндокринная система:** Данные отсутствуют.

**Костно-мышечная система:** Данные отсутствуют.

**Органы чувств:** Склерит.

**Местные изменения:** Кожа бледной окраски. Лимфоузлы не увеличены.

## **Лабораторные исследования**

**Кариотипирование от 17.04.2017:** Мужской кариотип. Хромосомная патология не выявлена.

**ОАК от 17.04.2017:** RBC  $5,02 \cdot 10^{12}$  кл/л, HGB 135 г/л, PLT  $249 \cdot 10^9$  кл/л, WBC  $11,6 \cdot 10^9$  кл/л, нейтрофилы (п/я) 4 %, нейтрофилы (с/я) 34 %, лимфоциты 55 %, моноциты 6 %, эозинофилы 1 %.

**Биохимия крови от 17.04.2017:** Глюкоза 3,3 ммоль/л.

**МРТ головного мозга от 23.11.2017:** Асимметричная внутренняя гидроцефалия, состояние после ВПШ, признаки аномалии Денди—Уокера.

**КТ головного мозга от 05.03.2018:** Наружно-внутренняя асимметричная гидроцефалия, вентрикулит, состояние после ВПШ с двух сторон.

**ЭЭГ от 07.2018:** Региональная эпилептическая активность в лобно-височных отделах в фазу медленного сна.

**Rg ОГК от 11.09.2019:** Задние отрезки IV–V ребер слева несколько сближены и утолщены. Легкие прозрачны. Корни фиброзные, малоструктурные. Диафрагма ровная и четкая, синусы свободны с обеих сторон. Средостение не расширено. Сердце несколько расширено влево.

### **Диагноз**

[Q04.8] Другие уточнённые врожденные аномалии мозга: синдром Денди—Уокера, внутренняя окклюзионная гидроцефалия, состояние после ВПШ с двух сторон, центральный тетрапарез, псевдобульбарный синдром. Симптоматическая фокальная эпилепсия.

[P27.1] Бронхолегочная дисплазия недоношенных тяжелой степени.

[Q24.8] Врождённые пороки сердца: открытый аортальный проток, состояние после лигирования, открытое овальное окно, персистирующее фетальное кровообращение.

[H35.9] Оперированная ретинопатия новорожденных.

### **Дифференциальный диагноз**

Дифференцировать с прочими редкими генетическими синдромами.

## ЗАКЛЮЧЕНИЕ

по результатам секвенирования экзона от 03.09.2020

|                       |                                |
|-----------------------|--------------------------------|
| Код пациента:         | P23                            |
| Вид материала:        | Кровь (венозная)               |
| Дата забора:          | 19.05.2020                     |
| Экзомная панель:      | SeqCap EZ MedExome Kit (Roche) |
| Количество прочтений: | 147,0 млн                      |
| Длина прочтений:      | 150 bp                         |
| Выравнивание:         | 99,59 %                        |
| Покрывание экзона:    |                                |
| – $\geq 10$ :         | 98,74 %                        |
| – среднее:            | 280 прочтений/позицию          |

**Заключение:** Значимых изменений, соответствующих критериям поиска, в данных экзомного секвенирования не обнаружено.

# **Приложение Д. Медицинские учреждения, предоставившие пациентов**

1. Томский национальный исследовательский медицинский центр Российской академии наук.

Юридический адрес: 634009 г. Томск, пер. Кооперативный, дом 5. Тел.: +7 (3822) 51-33-06, +7 (3822) 46-95-66. E-mail: [center@tnimc.ru](mailto:center@tnimc.ru).

Сайт: [www.tnimc.ru](http://www.tnimc.ru). Лицензия № ФС-72-01-001224 от 31.10.2019 г.

2. Научно-исследовательский клинический институт педиатрии имени академика Ю. Е. Вельтищева ФГАОУ ВО РНИМУ им. Н. И. Пирогова Минздрава России.

Юридический адрес: 117997 г. Москва, ул. Островитянова, дом 1. Тел.: +7 (495) 484-02-92. E-mail: [niki@pedklin.ru](mailto:niki@pedklin.ru). Сайт: [pedklin.ru](http://pedklin.ru). Лицензия № ФС-99-01-009646 от 16.05.2019 г.

3. Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр детской гематологии, онкологии и иммунологии имени Дмитрия Рогачева» Министерства здравоохранения Российской Федерации.

Юридический адрес: 117997 г. Москва, ул. Саморы Машела, дом 1. Тел.: +7 (495) 287-65-70. E-mail: [info@fnkc.ru](mailto:info@fnkc.ru). Сайт: [fnkc.ru](http://fnkc.ru). Лицензия № ФС-99-01-009603 от 25.12.2018 г.

4. Федеральное государственное бюджетное научное учреждение «Медико-генетический научный центр имени академика Н. П. Бочкова».

Юридический адрес: 115522 г. Москва, ул. Москворечье, дом 1. Тел.: +7 (499) 612-86-07, +7 (499) 612-00-37. E-mail: [mgnc@med-gen.ru](mailto:mgnc@med-gen.ru). Сайт: [med-gen.ru](http://med-gen.ru). Лицензия № ФС-99-01-009668 от 23.07.2019 г.



5. Научно-исследовательский институт клинической и экспериментальной лимфологии — филиал ИЦиГ СО РАН.

Юридический адрес: 630090, г. Новосибирск, Проспект Академика Лаврентьева, дом 10. Тел.: +7 (383) 291-14-81, +7 (383) 336-07-08.

Е-mail: [niikel@niikel.ru](mailto:niikel@niikel.ru). Сайт: [niikel.ru](http://niikel.ru). Лицензия № ФС-54-01-002231 от 18.09.2020 г.

6. ГБУЗ НСО «Клинический Центр охраны здоровья семьи и репродукции».

Юридический адрес: 630136, г. Новосибирск, ул. Киевская, дом 1. Тел.: +7 (383) 341-81-00 (приемная). Е-mail: [m3418100@yandex.ru](mailto:m3418100@yandex.ru).

Сайт: [csp.mznso.ru](http://csp.mznso.ru). Лицензия № ЛО-54-01-004181 от 13.12.2016 г.

# Тезисы

1. Генетические патологии являются одной из основных причин младенческой и детской смертности в развитых странах. Взрослые люди с генетическими патологиями требуют от государства вложения огромного количества денежных средств.

*Прочими причинами смертности являются недоношенность (половина) и родовые травмы. В развивающихся странах на первом месте стоят инфекции.*

2. Одним из наиболее перспективных и стремительно развивающихся методов детекции генетических заболеваний является секвенирование нового поколения. Изначально использовалось финансово затратное полногеномное секвенирование. Когда были получены данные, что 99 % значимых мутаций расположены в экзонах и рядом с ними, развитие получило более дешевое полноэкзомное секвенирование.

Тем не менее способность этих методов в определении хромосомных перестроек оставалась ограниченной. Было предложено множество методов для поиска хромосомных перестроек, среди которых я хочу отметить технологию захвата хромосом. Однако эти методы либо имеют низкую разрешающую способность (от этой проблемы страдает, например, классическое каритипирование), либо слишком дороги для рутинного использования (как, например, вышеупомянутая технология Hi-C).

В нашей лаборатории был предложен метод Echo-C, соединяющий Hi-C с экзомным обогащением. В теории он пригоден для поиска как точковых вариантов, так и хромосомных перестроек, при этом его себестоимость значительно ниже, чем у полногеномного секвенирования. Тем не менее, обработка результатов Echo-C требует специальных биоинформационных инструментов.

3. Таким образом, наша цель — сравнение эффективности методов Echo-C, полногеномного и экзомного секвенирования для поиска точечных полиморфизмов в геномах клеток человека. Поставленные нами зада-

чи — разработать биоинформационный протокол анализа данных Ехо-С, проверить эффективность метода Ехо-С на находящихся в открытом доступе геномных данных клеточной линии K562, а затем применить этот метод к пациентам и оценить его эффективность применительно к клиническим случаям.

4. Несмотря на то, что составляющие метода Ехо-С — Hi-C и экзомное обогащение — в настоящее время отработаны, их сочетание породило некоторое количество проблем. Нами были разработаны две вариации протокола (19 и 20), и их необходимо было проверить на соответствие критериям качества Hi-C и экзомного секвенирования, описанных в российских рекомендациях по обработке данных MPS. Библиотеки контрольного образца (клеточная линия K562) для обеих вариаций протокола Ехо-С соответствуют большинству критериев качества, за исключением таргетного обогащения (оно оказалось менее чем 10-кратным при заявленном 75-кратном). Кроме того, библиотека ЕхоС-20 пригодна только для поиска гомозиготных вариантов, а её доступное анализу покрытие экзонов на 20 % ниже рекомендуемого.
5. Далее решено было проверить, насколько метод Ехо-С подходит для идентификации генетических вариантов. В качестве контроля мы также взяли иммортализованную клеточную линию K56 и отобрали 8 библиотек из открытых данных секвенирования линии K562. Генетические варианты, найденные во всех восьми, были взяты как «золотой стандарт» экзомных вариантов. Далее две библиотеки, имеющие наименьший процент пересечения с прочими, были исключены из выборки. Получившаяся выборка надёжных геномных вариантов была взята нами как «серебряный стандарт».
6. Общее количество найденных в геноме и экзOME вариантов согласуется с литературными данными.
7. Геномные варианты, найденные в наших библиотеках 19 и 20, с вариантами контрольных выборок. Видно, что мы обнаруживаем в обеих библиотеках выше 80—90 % вариантов золотого и серебряного стандар-

та.

8. Таким образом, мы выяснили, что метод Ехо-С подходит для поиска полиморфизмов в экзOME. Далее мы решили проверить, подходит ли наш метод для анализа клинических случаев. Для анализа мы взяли библиотеки Ехо-С, приготовленные из образцов 10 пациентов. Также в качестве контроля для биоинформационного пайплайна мы использовали данные полноэкзомного секвенирования ещё 11 пациентов. Спектр предполагаемых патологий включал психоневрологические синдромальные нарушения, иммунодефициты, лихорадку неясного генеза, ихтиоз и нейрофиброматоз I типа. При анализе пациентов Ехо-С клинически значимые варианты, соответствующие критериям поиска, были выявлены в 4 случаях из 10 (из них 2 патогенных и 2 варианта неопределенной клинической значимости), в одном случае мой коллега идентифицировал по данным Ехо-С хромосомную перестройку.
9. При анализе экзомных данных клинически значимые варианты, соответствующие критериям поиска, были выявлены в 5 случаях из 11, из них 1 патогенный и 4 неопределённой клинической значимости.