

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НОВОСИБИРСКИЙ НАЦИОНАЛЬНЫЙ  
ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ» (НОВОСИБИРСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ, НГУ)  
Институт медицины и психологии В. Зельмана НГУ

## КУРСОВАЯ РАБОТА

Валеев Эмиль Салаватович

Группа 12452

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

# 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.

# 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.) [?]. 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 [?]. 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 — ?).

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 [?]. 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) [?]. 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 [?].

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 [?]. 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 [?]. 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. [?]. 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®) [?];
3. Orphanet [?].

Около 80 % редких болезней имеют генетическую природу и начинаются в раннем детстве [?]. 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-кодона 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 [?].

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 [?].

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 [?]. Изменение ацетилирования гистонов некоторых генов в клетках головного мозга связано с развитием такого заболевания, как шизофрения [?].

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) [?]. 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*) [?].

Несмотря на то, что в развитии наследственных заболеваний эпигенетика 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 — сбалансированные (без изменения количества генетической информации) и несбалансированные (с изменением количества генетической информации).

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

- Анэуплоидии — изменение числа хромосом. Примерами анэуплоидий 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). Частичная моносомия — синдром кошачьего крика (связан с утратой плеча 5 хромосомы). Прочие анэуплоидии ведут к несовместимым с жизнью нарушениям эмбрионального развития и, как следствие, спонтанным абортам.
- 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 — ?) и служить онкогенами [?].
- Вариации числа копий (англ. *Copy Number Variations, CNV*) — дубликации (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 мо-

гут носить как положительный характер, так и отрицательный — в частности, дупликации в гене *CCL3L1* способны увеличить устойчивость к ВИЧ [?], а крупные CNV в разных частях генома ассоциированы с расстройствами аутистического спектра [?].

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 загираания 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 негомолотичными хромосомами) и Робертсоновские (слияние акроцентрических хромосом с образованием метацентрической или субметацентрической). Сбалансированные транслокации могут как не проявляться в фенотипе (сказываясь только на фертильности — ?), так и приводить к серьёзным последствиям — например, синдрому Дауна (робертсоновская транслокация является причиной синдрома Дауна в 2–4 % случаев — ?).

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 ?, 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 [?].
- 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 [?]. 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 [?]. Генетические варианты могут как разрушать канонические сплайс-сайты, так и способствовать образованию новых внутри интронных участков [?]. 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 [?]:

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 отдифференцировать 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. С-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, С-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 %) [?].

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 различных 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 [?]. 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 участок 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 [?].

**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 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 агау-CGH (aCGH), или хромосомного микроматричного анализа (ХМА), при котором CGH комбинируется с микрочиповой гибридизацией [?]. 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 неизвестных или часто повторяющихся

последовательностей, эти методы не подходят [?].

**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 [?].

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 [?]. 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 прочитать 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 клинической практике (поиск патогенных генетических вариантов). Секвенирование длинными прочтениями хорошо подходит для сборки новых геномов и изучения отдельных изоформ генов [?]. 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 увели-

чением 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 [?].

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

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 транскриптома одной клетки (англ. *SMART*) используют длинные пробы, содержащие как праймер, так и регион для остановки 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 [?].

**Полноэкзомное секвенирование** (англ. *Whole Exome Sequencing, WES*). Техника заключается в секвенировании обогащённого экзома — 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 [?].

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 [?], 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 [?]. 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 [?].

**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 [?]. При обработке большого количества 3C-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 3C. Самым известным и широко применяемым является метод Hi-C, сочетающий 3C с методами массового параллельного секвенирования. С его помощью можно подсчитать количество контактов во всём геноме — как внутри-, так

и межхромосомные контакты [?].

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.

**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 [?].



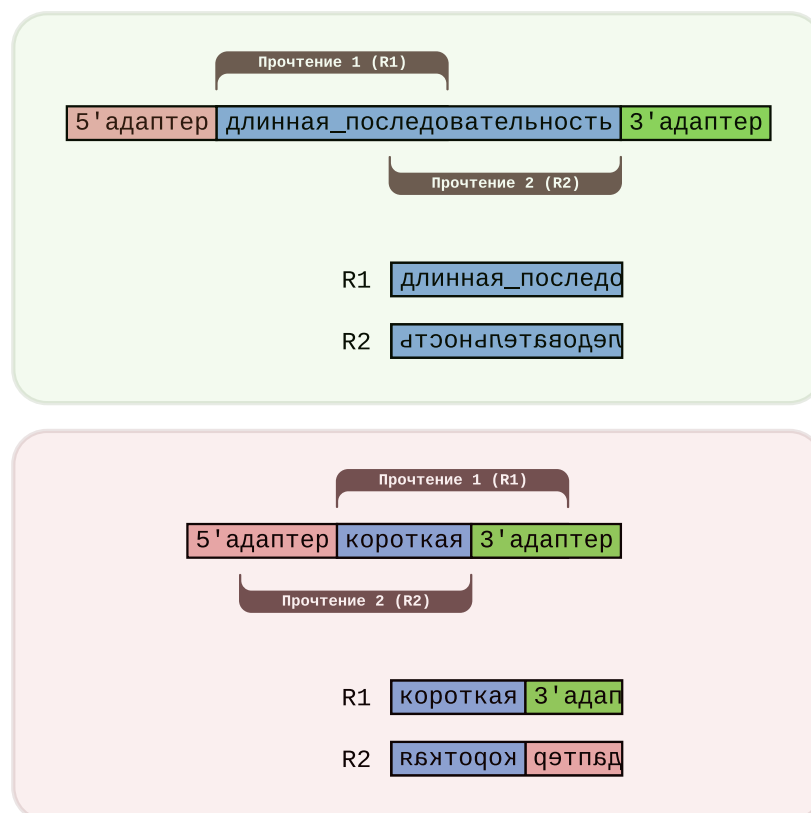


Рис. 1: Результаты секвенирования библиотек, содержащих короткие последовательности, могут быть контаминированы адаптерами. На рисунке показаны результаты секвенирования последовательностей, длина которых превышает количество циклов секвенирования (сверху), либо значительно меньше количества циклов (снизу).

**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 приво-

дится 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 [?].

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 выравни-

ниваний 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 невыравненного 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*).
- 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.

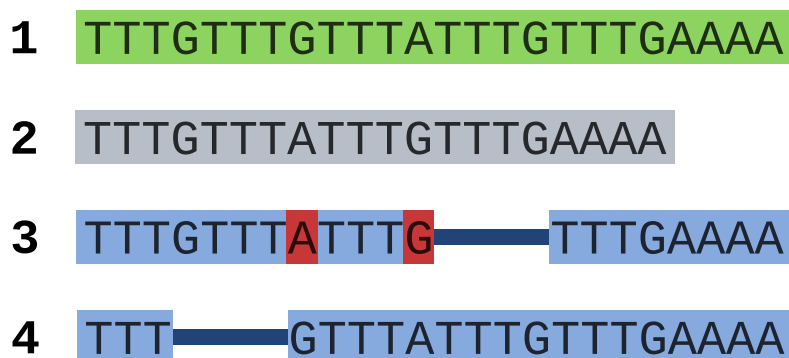


Рис. 2: Неоптимальное картирование прочтения, содержащего делецию. (1) — референсная последовательность, (2) — последовательность прочтения, (3) — картирование, произведённое алгоритмом, включающее две SNV и одну делецию, (4) — оптимальное местоположение делеции

**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 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 [?].

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 [?]. 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 [?]. 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 [?].

**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 неправиль-

ная 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) [?], 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 аннотация 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 [?], 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 [?].
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 [?] и OrphaData [?].

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.

Основные проблемы pLI [?]:

- Плохо приспособлен к распознаванию аутосомно-рецессивных вариантов (из-за того, что частота повреждающих вариантов в популяции может быть высокой) и X-сцепленных рецессивных вариантов (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 pLI 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 pLI с гаплонедостаточностью конкретного гена в принципе сложно объяснить.

Таким образом, высокое значение pLI 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 [?], 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 [?],

в которой отражены частоты по популяции Ближнего Востока, ABraOM [?], предоставляющая частоты генетических вариантов среди практически здорового пожилого населения Бразилии. Также для анализа берутся популяции, в которых велика доля близкородственных связей, например, пакистанская [?].

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 [?] и CLINVAR [?]. Тем не менее, CLINVAR считается лишь дополнительным источником, так как часто содержит информацию низкого качества [?].
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 *BRCA1* и *BRCA2* это 33 % и 50 % соответственно. 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 [?]. 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 [?].

**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. 3С-методы на сегодняшний момент являются наиболее перспективным способом обнаружения хромосомных перестроек [?], но при небольшой глубине секвенирования в них обнаружение точечных полиморфизмов затруднительно [?]. WGS способно обнаруживать большую часть SNV, небольших инсерций и делеций, но требует большую глубину секвенирования [?]; WES, с другой стороны, позволяет выявить генетические варианты при небольшой глубине секвенирования, но только в

экзомах. Возможности обнаружения хромосомных перестроек для последних двух методов ограничены.

Компромиссом между ценой и возможностями поиска генетических вариантов может служить новейший метод Echo-C, сочетающий технологии таргетного обогащения с ЗС. 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) [?].

Тем не менее, как выяснилось, уже существующие биоинформационные методы следует модифицировать для корректной обработки данных Echo-C. Это связано в первую очередь с особенностями протокола Hi-C, к примеру, наличием технических последовательностей (бридж-адаптеров), которые приводят к появлению ложных SNV в экзонных 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 точечных генетических вариантов в Echo-C-данных и последующего сравнения Echo-C с методами полногеномного и полноэкзомного секвенирования.

## Глава 2. Материалы и методы

**Данные секвенирования.** Поиск данных секвенирования производился в базах данных NCBI (GEO DataSets, SRA, PubMed) и ENCODE с использованием ключевых слов “K562”, “K562+WGS”, “K562+WES”, “K562+Hi-C”.

**Контроль качества NGS-данных.** Для контроля качества прочтений мы использовали утилиту FastQC [?], 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 [?].

**Удаление адаптерных последовательностей.** Удаление адаптерных последовательностей производилось с помощью утилиты cutadapt [?]. В ? 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.

**Картирование.** Картирование производилось с помощью инструментов Bowtie2 [?] и BWA [?]. BWA показал лучшие результаты; кроме того, он значительно более эффективно работает с химерными ридями, 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 (некартиро-

ванные/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 [?]:

- 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 [?].

**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 [?], интегрированным в GATK. Оптимальные показатели скорости MarkDuplicates достигаются при запуске Java с параллелизацией сборщиков мусора и количеством сборщиков мусора равным двум [?]. Также, согласно рекомендациям разработчиков, прочтения были предварительно отсортированы по именам, чтобы удалению подверглись не только первичные, но и добавочные выравнивания [?].

**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 [?]. 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 ? 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 про-

чений, 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 [?].

**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 [?]. 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 [?].

Используемые базы данных:

1. Human Gene Mutation Database (HGMD®) [?]
2. Online Mendelian Inheritance in Man (OMIM®) [?]
3. GeneCards®: The Human Gene Database [?]



4. CLINVAR [?]
5. dbSNP [?]
6. Genome Aggregation Database (gnomAD) [?]
7. 1000Genomes Project [?]
8. Great Middle East allele frequencies (GME) [?]
9. dbNSFP: Exome Predictions [?]
10. dbSNV: Splice site prediction [?]
11. RegSNPIntron: intronic SNVs prediction [?]

Проблемные регионы:

1. ENCODE Blacklist [?] — 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 [?] — содержит регионы, на которых сложно обнаружить генетические варианты (из-за низкого покрытия прочтениями, систематических ошибок секвенирования и локальных проблем картирования).
3. NCBI GeT-RM [?] — 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 % [?].

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 [?];
- 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 [?];
- 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 [?].

# Глава 3. Результаты

На сегодняшний день были выполнены следующие этапы работы:

1. Создание контрольной выборки генетических вариантов, с помощью которой будет проведена оценка пригодности Ехо-С-библиотек к поиску генетических вариантов;
2. Проверка качества данных, полученных в результате массового параллельного секвенирования Ехо-С-библиотек;
3. Разработка, отладка и тестирование автоматизированного инструмента для обработки данных секвенирования Ехо-С-библиотек.

## 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 [???]. 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 экзома, 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	Доля дубликатов, %	Доля экзема с глубиной покрытия более 10, %	Среднее покрытие в экземе	Среднее покрытие вне экзема	Обогащение экзема, раз	Доля непокрытых регионов в экземе, %	Доля непокрытых регионов вне экзема, %
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.

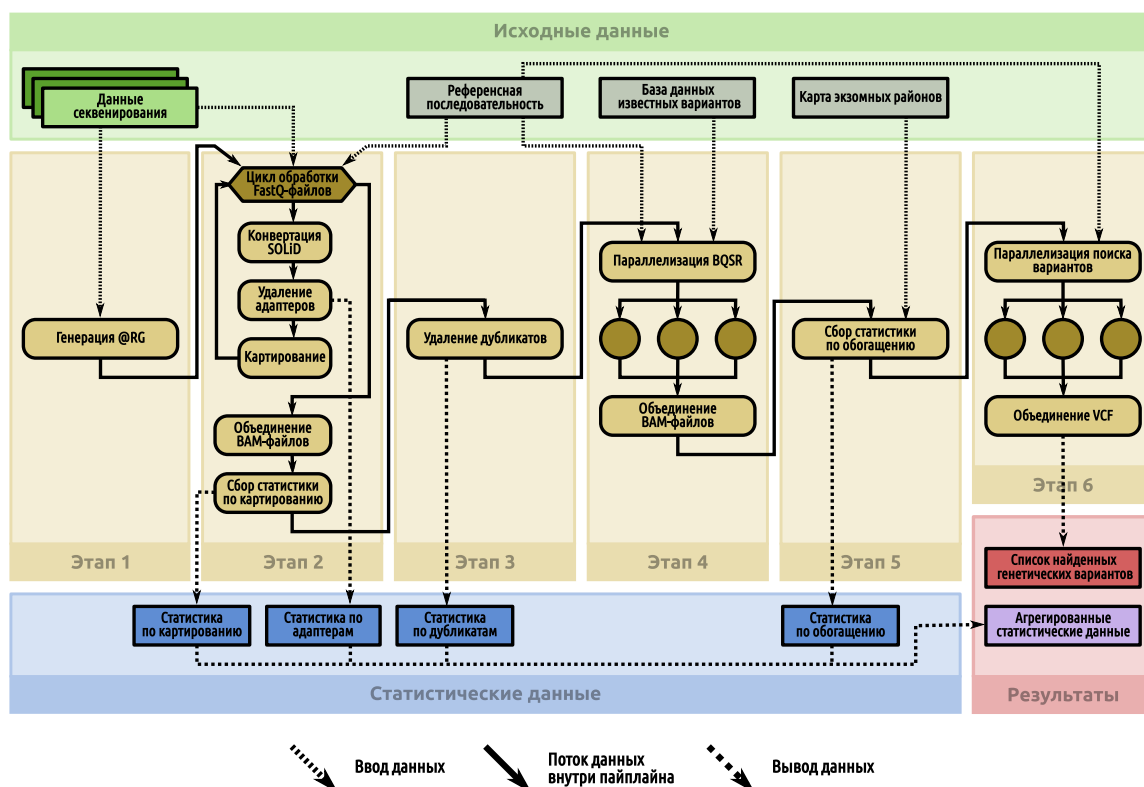


Рис. 3: Принципиальная схема пайплайна для обработки Eho-C-данных

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.
- 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

на: автоматизировано индексирование референсной последовательности, настройки вынесены в специальный конфигурационный файл, есть возможность обработки пула данных, используя один короткий сценарий;

Код пайплайна доступен на GitHub [?].

### 3.3. Сравнение данных секвенирования клеточной линии K562

Следующим важным этапом работы была проверка эффективности поиска генетических вариантов в Echo-C-библиотеках. Было решено использовать для этого распространённую иммортализованную клеточную линию K562, полученную от пациентки с хроническим миелолейкозом [?]. 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 [???????]. 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 chr2:25455845-25565459 с фланкированием 1 kbp (ген *DNMT3A*), так как в одной из работ использовали генетически модифицированную линию с вариантами в данном гене [?]. В качестве тестовых Echo-C-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 Табл. 7 и Табл. 8.

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 вариантов, встречающихся в одной библиотеке и не встречающихся в остальных (Табл. 2). Наибольший процент уникальных вариантов найден в данных ?

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, %
?	WES	254 983 225	408 008	41 830	10,25
?	Hi-C	72 914 268	1 399 457	27 365	1,95
?	WGS	366 291 496	4 649 012	327 184	7,03
?	Hi-C	256 500 659	2 365 361	67 678	2,86
?	Hi-C	1 366 228 845	4 218 233	320 508	7,59
?	Hi-C	428 306 794	1 789 324	89 624	5,00
?	Repli-seq	301 663 640	2 207 451	37 578	1,70
?	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.

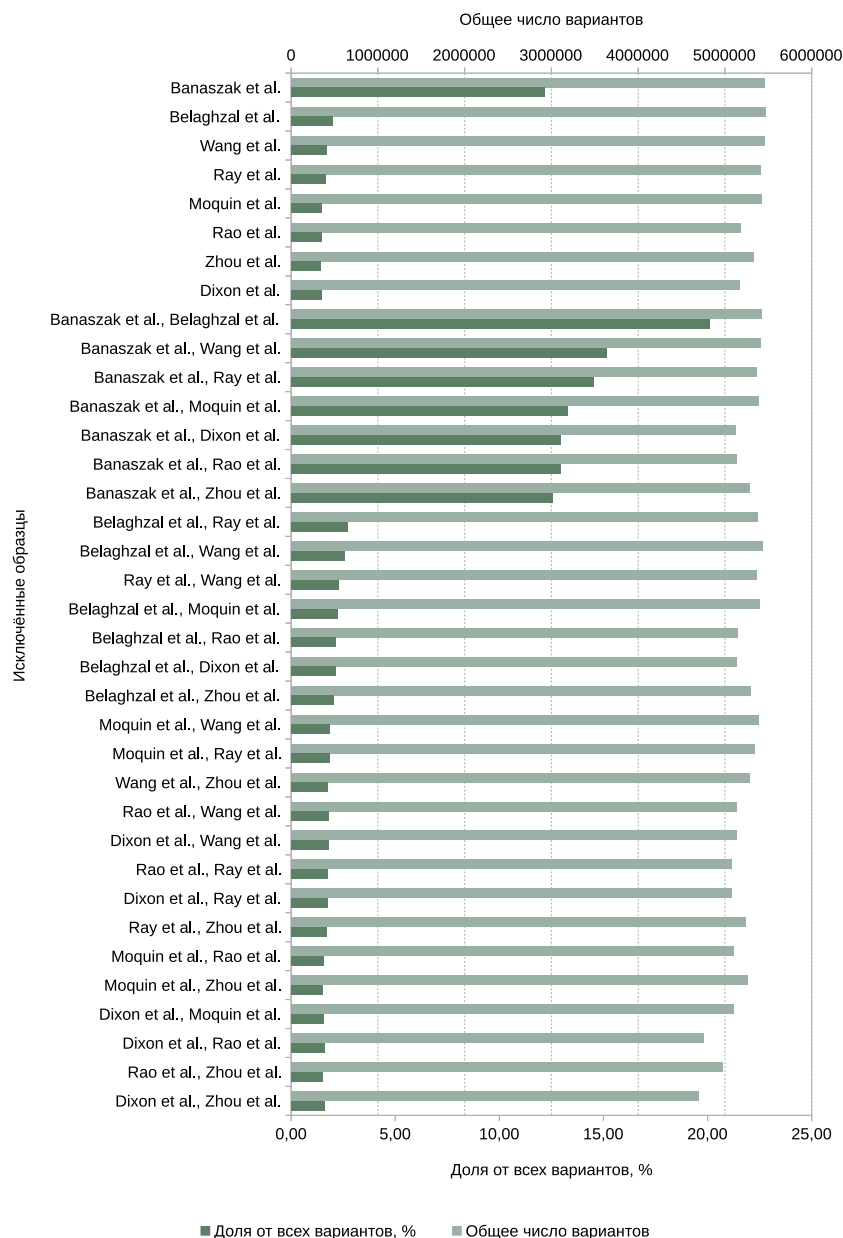


Рис. 4: Исключение отдельных образцов из контрольной выборки позволило увеличить количество генетических вариантов, которые можно использовать как стандарт. На рисунке показано суммарное количество вариантов в выборке (светло-зелёный) и процент общих для этой выборки вариантов (тёмно-зелёный) для выборок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

При исключении из выборки данных ? и ? 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 % библиотек в выборке. Таких вариантов в «золотом стандарте» — 62 518 (82,99 %),

в «серебряном стандарте» — 937 797 (85,93 %).

Далее мы использовали варианты «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 «серебряного» и «золотого» стандартов и наших Ехо-С-библиотек, их долю от общего числа вариантов в Ехо-С-библиотеках, а также количество и долю ложноположительных (отсутствующих в контрольных образцах) генетических вариантов.

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.

Поиск вариантов «серебряного» и «золотого» стандартов в наших библиотеках был произведён до и после фильтрации. Результаты показаны в Табл. 5.

Таблица 5: 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	—	—	—	—	—	—

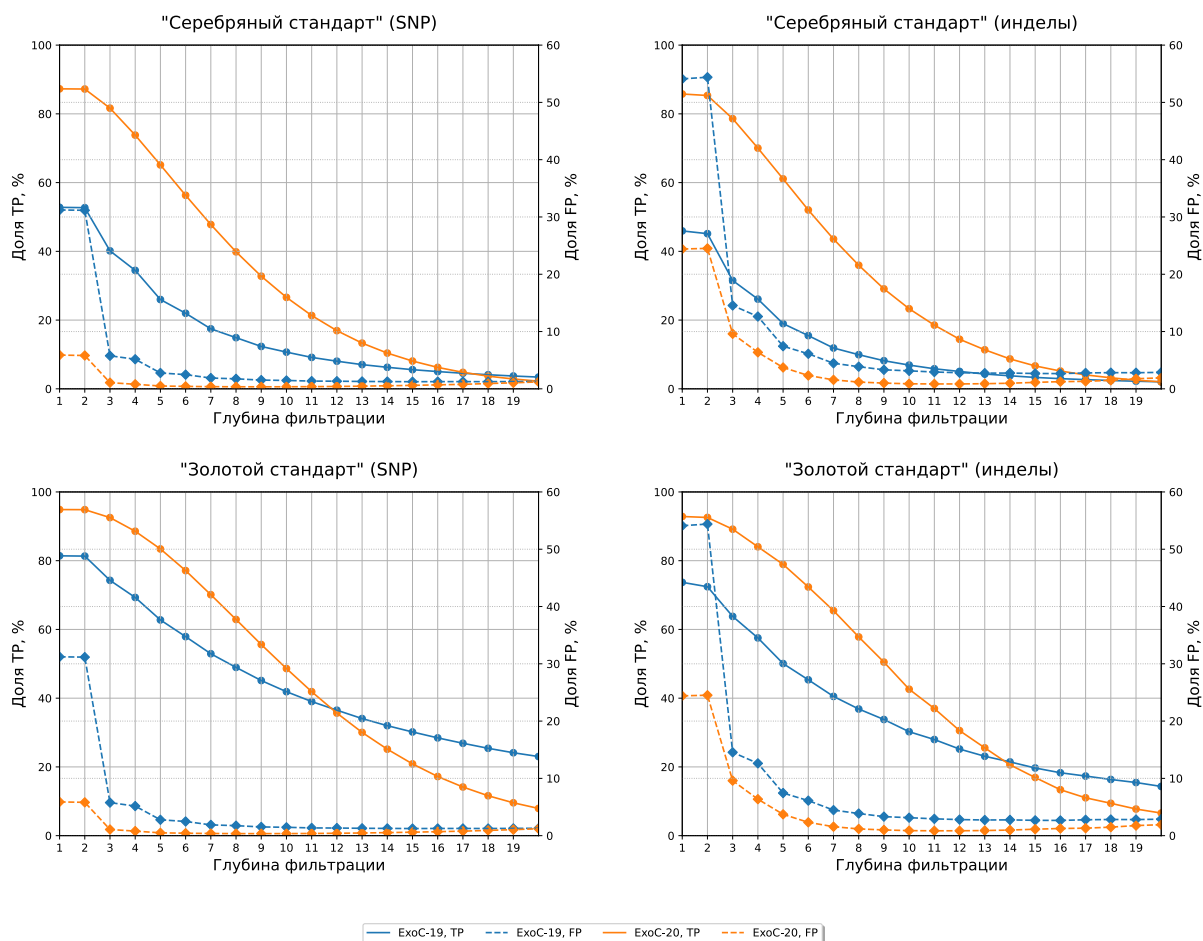


Рис. 5: Чувствительность и специфичность при фильтрации вариантов с разной глубиной альтернативного аллеля. Доля ложноположительных результатов вычислялась относительно всего образца, доля истинно положительных — относительно размера контрольной выборки. Варианты «серой зоны», которые присутствуют в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

# Глава 4. Обсуждение результатов

## 4.1. Контрольные образцы

«Золотой стандарт» с учётом подбора библиотек скорее всего является набором генетических вариантов, относящихся к экзомным регионам, так как одна из библиотек представляла собой результаты WES. Их было обнаружено 75 тыс., что соответствует оценкам среднего количества генетических вариантов в кодирующих регионах у человека — 100 тыс. [?]. Общее число несоответствий с референсным геномом у среднего человека составляет 4.1–5 млн [?], 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 ? 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 покрытие в экзоне — 60,51 и 14,88 прочтений для ЕхоС-19 и ЕхоС-20 соответственно. Глубину покрытия более 10 прочтений имеют 91,68 % и 72,58 % экзона для ЕхоС-19 и ЕхоС-20 соответственно. Согласно ?, для репрезентативных

результатов экзомного секвенирования необходима глубина секвенирования не менее чем в  $10^{10}$  bp, 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 [?]. Тем не менее, можно утверждать, что ExoC-19 отвечает требованиям для поиска SNV, а ExoC-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 ExoC-19 и ExoC-20 соответственно). Различия объясняются протоколами приготовления: у библиотеки ExoC-20 выше глубина покрытия в экзоме, но в 6 раз ниже обогащение в экзомных районах. Кроме того, в библиотеке ExoC-19 были использованы адаптерные последовательности, дающие большое количество шума.

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

- В библиотеке ExoC-19 потеряна большая доля вариантов, чем в ExoC-20 — как относительно общего числа, так и относительно вариантов «золотого» и «серебряного» стандартов.
- Доля ложноположительных (отсутствующих в контрольных образцах) генетических вариантов снизилась в 5 раз.

Всё это можно объяснить наличием в библиотеке ExoC-19 большого количества регионов с низким покрытием, генетические варианты в которых были отсеяны фильтрацией по глубине. То есть, фильтрация по глубине является эффективным способом улучшения данных низкого качества.

# Предварительные выводы

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

1. Пайплайн, созданный нами с учётом актуальных рекомендаций для биоинформационной обработки, позволяет обрабатывать данные Ехо-С-секвенирования, а также находить в этих данных SNV.
2. Использование разработанного конвейера биоинформационных инструментов позволило обнаружить около 5,5 млн генетических вариантов в контрольных данных клеточной линии K562 (что сопоставимо со средним количеством точечных полиморфизмов в геноме человека), из которых наличие 75 тыс. подтвердилось в восьми независимых исследованиях, а 1 млн — в шести независимых исследованиях, не включающих экзомные данные.
3. 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.

# План работы

В следующем семестре мы планируем:

1. Произвести анализ генетических вариантов в контрольных и наших образцах по следующим параметрам:

- тип;
- количество альтернативных аллелей;
- распределение в геноме — попадает ли золотой стандарт в экзомную панель;
- точность определения: определяет ли Ехо-С вариант, правильно ли определяет букву, правильно ли определяет зиготность.

Таблица 6: Валидация контрольных вариантов в Ехо-С-библиотеках

Параметр	ЕхоС-19				ЕхоС-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

- глубина покрытия;
- зиготность.

2. Произвести анализ данных Ехо-С на предмет систематических ошибок поиска генетических вариантов.

3. Произвести анализ результатов секвенирования Ехо-С-библиотек у реальных пациентов.



# Глава А. Краткое клиническое описание пациентов

Таблица 9: Список пациентов, чьи образцы были использованы для приготовления Ехо-С-библиотек

Код	Медицинское учреждение	Пол	Дата рождения	Предварительный диагноз	Родственные связи	Тип библиотеки
P1	Томский НИМЦ <sup>1</sup>	Ж	13.07.2018	[E84.9] Муковисцидоз неуточненный	—	Ехо-С
P2	Томский НИМЦ <sup>1</sup>	М	07.12.2012	[F83] Задержка психо-речевого развития	сибс P4	Ехо-С
P3	Томский НИМЦ <sup>1</sup>	Ж	27.04.2013	[F83] Задержка психо-речевого развития	—	Ехо-С
P4	Томский НИМЦ <sup>1</sup>	М	13.10.2014	[F83] Задержка психо-речевого развития	сибс P2	Ехо-С
P5	Томский НИМЦ <sup>1</sup>	М	25.07.2014	[F83] Задержка психо-речевого развития	—	Ехо-С
P6	НИКИ Педиатрии <sup>2</sup>	М	16.02.2007	[Q85.0] Нейрофиброматоз I типа	ребенок P7	Ехо-С
P7	НИКИ Педиатрии <sup>2</sup>	Ж	—	[Q85.0] Нейрофиброматоз I типа	мать P6	Ехо-С
P8	НМИЦ им. Рогачева <sup>3</sup>	Ж	01.04.2004	[D84.9] Первичный иммунодефицит неуточненный, осложненный иммунной нейтропенией и тромбоцитопенией	—	Ехо-С
P9	НМИЦ им. Рогачева <sup>3</sup>	Ж	19.02.2013	[Q87.8] Синдром повреждения Неймегена	—	Ехо-С
P10	МГНЦ <sup>4</sup>	Ж	—	[Q89.9] Комплекс врожденных аномалий	—	Ехо-С
P13	НИИКЭЛ <sup>5</sup>	Ж	10.03.2002	[D61.3] Идиопатическая апластическая анемия	—	WES
P14	НИИКЭЛ <sup>5</sup>	М	1993	[R50] Лихорадка неясного генеза, сезонного характера	—	WES
P15	КЦОЗСиР <sup>6</sup>	Ж	24.09.2016	[R27.0] Атактический синдром	—	WES
P16	КЦОЗСиР <sup>6</sup>	Ж	30.09.2018	[G93.0] Органическое поражение ЦНС генетической этиологии	—	WES
P17	КЦОЗСиР <sup>6</sup>	М	28.05.2015	[Q80.9] Ихтиоз неуточненный	—	WES
P18	КЦОЗСиР <sup>6</sup>	Ж	15.10.1988	[Q80.9] Ихтиоз неуточненный	—	WES
P19	КЦОЗСиР <sup>6</sup>	М	11.09.2019	[Q04.9] Врожденная аномалия головного мозга, предположительно X-сцепленная	—	WES
P20	КЦОЗСиР <sup>6</sup>	Ж	02.01.1977	[G98] Неуточненные нарушения нервной системы	мать P21 и P22	WES
P21	КЦОЗСиР <sup>6</sup>	Ж	—	[G98] Неуточненные нарушения нервной системы	ребенок P20, сибс P22	WES
P22	КЦОЗСиР <sup>6</sup>	М	—	[G98] Неуточненные нарушения нервной системы	ребенок P20, сибс P21	WES
P23	КЦОЗСиР <sup>6</sup>	М	17.04.2017	[Q04.8] Синдром Денди—Уокера	—	WES

Таблица 7: Библиотеки данных секвенирования клеточной линии K562

Библиотека	Статья	Рецензиратор	Код доступа	Тип данных	Тип прочтений	Глубина, прочтений	Общее число прочтений	Доля картированных, % от общего числа	Доля дополнительных, % от общего числа	Картированные PE прочтения	Дубликаты PE прочтений	Дубликаты синглетов	Доля дубликатов, %	Оценка размера библиотеки
Контрольные данные														
GSM1551618_HIC069	?	GEO	SRR1658693	Hi-C	PE	456757799	1 001 169 246	96,57	8,755	424 945 100	29 290 805	17 848 021	13 182 626	491 611 432
GSM1551619_HIC070	?	GEO	SRR1658694	Hi-C	PE	591 854 553	1 314 487 935	98,7	9,949	575 565 379	15 452 072	98 778 796	8811 132	1 479 944 337
GSM1551620_HIC071	?	GEO	SRR1658695	Hi-C	PE	79 905 895	173 931 529	98,81	8,118	77 880 938	1 975 600	486 883	269 138	6 202 732 721
GSM1551621_HIC072	?	GEO	SRR1658697	Hi-C	PE	79 578 049	159 160 116	98,38	0,003	77 155 821	2 265 995	364 805	285 395	8 088 955 029
GSM1551622_HIC073	?	GEO	SRR1658698	Hi-C	PE	77 353 816	154 710 364	98,33	0,002	74 866 287	2 383 970	240 304	293 115	11 637 260 975
GSM1551623_HIC074	?	GEO	SRR1658702	Hi-C	PE	80 778 733	175 291 763	98,65	7,835	78 467 394	2 254 814	644 986	321 965	4 746 870 162
ENC8R025CPQ	?	ENCODE	ENCFF574YL.G	WGS	SE	258 022 356	260 044 021	85,39	0,777	—	220 029 156	—	50 689 083	—
ENC8R03AXS	?	ENCODE	ENCFF921AXL	WGS	SE	1 472 492 722	1 592 540 515	91,19	7,538	—	1 332 175 586	—	496 237 198	—
ENC8R71UNY	?	ENCODE	ENCFF471WSA	WGS	SE	880 796 215	889 473 769	98,72	0,965	—	888 239 055	—	203 498 352	—
GSE148362_G1	?	GEO	SRR6251264	Replic-seq	PE	366 291 496	737 534 099	98,72	0,671	364 794 328	923 254	730 18 048	406 066	785 091 005
GSE148362_G2	?	GEO	SRR11518301	Replic-seq	SE	24 804 095	24 804 396	96,39	0,001	—	23 909 072	—	921 353	—
GSE148362_S1	?	GEO	SRR11518308	Replic-seq	SE	33 032 314	33 033 010	97,61	0,002	—	32 241 907	—	3881 591	—
GSE148362_S2	?	GEO	SRR11518302	Replic-seq	SE	30 884 788	30 885 298	98,39	0,002	—	30 481 936	—	2 156 480	—
GSE148362_S3	?	GEO	SRR11518303	Replic-seq	SE	45 359 273	45 360 305	98,39	0,002	—	44 630 884	—	1 939 846	—
GSE148362_S4	?	GEO	SRR11518304	Replic-seq	SE	49 807 076	49 807 988	98,79	0,002	—	49 205 535	—	2 889 464	—
GSE148362_S5	?	GEO	SRR11518305	Replic-seq	SE	44 149 029	44 149 770	98,46	0,002	—	43 469 002	—	2 678 091	—
INSITU_HS1	?	GEO	SRR11518306	Replic-seq	SE	38 424 060	38 424 835	97,96	0,002	—	37 640 056	—	3 600 560	—
INSITU_HS2	?	GEO	SRR11518307	Replic-seq	SE	35 203 005	35 203 676	97,51	0,002	—	34 224 742	—	4 177 438	—
INSITU_NHS1	?	GEO	SRR9019504	Hi-C	PE	86 294 895	172 589 790	93,3	0	75 521 119	9 982 274	1 841 061	1 615 286	1 523 677 153
INSITU_NHS2	?	GEO	SRR9019506	Hi-C	PE	127 093 919	254 187 638	93,36	0	111 730 240	13 658 195	19 231 46	30 48 273	3 208 280 367
PDDE_TRANSIENT	?	GEO	SRR5470533	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
PD_STABLE_REP1	?	GEO	SRR5470534	Hi-C	PE	128 472 386	256 944 772	93,27	0	112 615 319	14 417 076	1 961 996	3 196 535	2 597 3 194 317 878
PD_STABLE_REP2	?	GEO	SRR5470536	Hi-C	PE	55 158 049	110 319 638	95,6	0,003	51 158 920	3 140 556	391 7 308	721 938	3 167 780 447
PD_TRANSIENT	?	GEO	SRR5470538	Hi-C	PE	67 172 619	134 347 099	97,58	0,001	64 767 511	1 565 427	55 73 966	376 260	354 373 851
GSM2536769_WT	?	GEO	SRR5470534	Hi-C	PE	52 872 167	105 745 908	98,23	0,001	51 442 087	983 483	20 58 449	217 598	6 25 522 723
GSM2536770_WT_TF	?	GEO	SRR5470539	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	361 336 652
GSM2536771_MT2	?	GEO	SRR5470538	Hi-C	PE	72 914 268	172 533 452	98,39	15,478	72 067 575	648 294	9 694 590	2 10 273	243 264 112
GSM2536772_MT3	?	GEO	SRR5470533	Hi-C	PE	39 211 303	78 464 649	98,46	0,054	38 914 993	171 253	7 821 960	91 145	83 342 746
GSM2536773_MT4	?	GEO	SRR5470533	Hi-C	PE	49 384 206	98 820 633	98,54	0,033	49 068 605	193 565	10 478 814	114 795	97 869 629
GSM2536774_MT5	?	GEO	SRR5470533	Hi-C	PE	42 020 936	84 083 776	98,63	0,062	41 772 436	189 177	8 755 216	104 927	85 177 326
GSM2536775_MT6	?	GEO	SRR5470533	Hi-C	PE	43 669 612	87 375 385	99,6	0,041	43 414 109	164 448	9 489 133	93 001	84 242 110
GSM2536776_MT7	?	GEO	SRR5470533	Hi-C	PE	39 879 263	79 798 847	98,53	0,038	39 609 943	166 651	8 590 165	90 009	77 577 055
GSM2536777_MT8	?	GEO	SRR5470536	Hi-C	PE	40 807 904	81 649 292	99,59	0,041	40 559 469	163 957	8 801 283	91 545	79 383 290
Тестовые данные														
FG_ExcBcl-001	—	—	—	Exo-C	PE	136 600 179	359 215 777	98,31	23,940	135 150 334	443 409	25 453 568	159 152	319 784 450
FG_Quarantine-A	—	—	—	Exo-C	PE	53 598 130	140 214 460	99,79	23,150	53 598 130	259 561	7 809 282	68 779	193 853 459
FG_Quarantine-B	—	—	—	Exo-C	PE	55 279 173	144 641 130	99,76	23,108	55 279 173	310 369	8 808 307	90 489	177 375 163

«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 дублированных, % от общего числа	FR PE прочтения, % от картированных	Картированные PE прочтения	Картированные синглеты	Картированные пары, % от картированных	Картированные на разные хромосомы % от картированных на разные хромосомы
Контрольные данные											
?	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
?	WGS	SE	2 621 311 283	2 752 058 305	93,43	4,751	—	—	—	—	—
?	WGS	PE	366 291 486	737 534 099	99,72	0,671	97,16	729 588 656	923 254	1,25	51,22
?	Repli-seq	SE	301 663 640	301 669 278	98,09	0,002	—	—	—	—	—
?	Hi-C	PE	428 306 794	856 613 588	93,33	0	35,92	751 519 632	47 995 392	22,77	76,00
?	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
?	Hi-C	PE	72 914 268	172 533 452	99,39	15,478	24,77	144 135 150	648 294	34,02	88,02
?	WES	PE	254 983 225	510 192 582	99,56	0,044	99,41	506 689 110	1049 051	0,11	81,38
Тестовые данные											
ExoC-19	Exo-C	PE	136 609 179	339 215 777	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

# 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.

---

<sup>1</sup>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.

<sup>2</sup>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.

<sup>3</sup>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.

<sup>4</sup>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: [mngnc@med-gen.ru](mailto:mngnc@med-gen.ru). Сайт: [med-gen.ru](http://med-gen.ru). Лицензия № ФС-99-01-009668 от 23.07.2019 г.

<sup>5</sup>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. E-mail: [niikel@niikel.ru](mailto:niikel@niikel.ru). Сайт: [niikel.ru](http://niikel.ru). Лицензия № ФС-54-01-002231 от 18.09.2020 г.

<sup>6</sup>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). E-mail: [m3418100@yandex.ru](mailto:m3418100@yandex.ru). Сайт: [cspz.mznso.ru](http://cspz.mznso.ru). Лицензия № ЛО-54-01-004181 от 13.12.2016 г.

## **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.

## 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.



## 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.

## 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.

## 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.

## 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, моче́вая 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, альбуминово-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, pH 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.

## **Пациент Р7**

**Пол:** Женский

**Дата рождения:** Неизвестна

**Дата обследования:** Неизвестна

### **Анамнез заболевания**

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

## 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 лейкостимом 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 (интратект 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.

## 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 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:** ОУ-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*  $10^3$  КОЕ/мл, *Corynebacterium pseudodiphthericum*  $10^3$  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 tuberculostrictum* <10<sup>3</sup> КОЕ/мл, *Staphylococcus aureus* 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 мицелля 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.

## 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.

## 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:** Hb 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.

## 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, артралгии, 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.

## 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, объём 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.

## 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.



## 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 неучтенная. 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.

## 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**

**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:** Достоверных вариантов, соответствующих описанному фенотипу, при экзомном секвенировании не выявлено.

## 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

**Вид материала:** Кровь (венозная)

**Дата забора:**

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

**Количество прочтений:** 5 млн

**Длина прочтений:** 150 bp

**Выравнивание:** 5 %

**Покрывание экзона:**

- $\geq 10$ : 5 %
- среднее: 5 прочтений/позицию

**Заключение:**

## 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.

## **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.

## 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.

## 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.