

Буян Андрей Игоревич

**Гетерогенность молекулярных фенотипов
тканей поперечно-полосатых мышц человека**

1.5.3. Молекулярная биология

Автореферат диссертации на соискание учёной степени
кандидата биологических наук

Пушино - 2026

Работа выполнена в Группе регуляции биосинтеза белка Федерального государственного бюджетного учреждения науки «Институт белка Российской академии наук»

Научный руководитель: **Кулаковский Иван Владимирович** – доктор биологических наук, ведущий научный сотрудник Группы регуляции биосинтеза белка Федерального государственного бюджетного учреждения науки «Институт белка Российской академии наук» (ИБ РАН)

Официальные оппоненты: **Скоблов Михаил Юрьевич** – доктор биологических наук, заведующий отделом функциональной геномики Федерального государственного бюджетного научного учреждения «Медико-генетический научный центр имени академика Н.П. Бочкова» (МГНЦ)

Фишман Вениамин Семенович – доктор биологических наук, ведущий научный сотрудник сектора геномных механизмов онтогенеза Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» (ИЦиГ СО РАН)

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт теоретической и экспериментальной биофизики Российской академии наук

Защита состоится 25 ноября 2026 года в 13:00 на заседании диссертационного Совета 24.1.081.01 (Д 002.235.01) при Федеральном государственном бюджетном учреждении науки Институте молекулярной биологии им. В.А. Энгельгардта Российской академии наук по адресу: 119991, Москва, ул. Вавилова, д. 32.

С диссертацией можно ознакомиться в кабинете 301 Федерального государственного бюджетного учреждения науки Института молекулярной биологии им. В.А. Энгельгардта Российской академии наук по адресу: 119991, Москва, ул. Вавилова, д. 32 и на сайте ИМБ РАН <http://eimb.ru>

Автореферат разослан «____» октября 2026 г.

Ученый секретарь диссертационного совета
кандидат физико-математических наук

А. А. Анашкина

Общая характеристика работы

Актуальность и степень разработанности темы

К 2003 году в рамках проекта «Геном Человека» было расшифровано более 90% референсной последовательности нашего генома [Collins и др., 2003]. Знание геномной последовательности стало отправной точкой в понимании реализации генетической программы человека как на уровне молекулярных фенотипов тканей и клеток, то есть представленности различных типов молекул, включая РНК и белки, так и на уровне наблюдаемых признаков целого организма. Дополнительный импульс дали технологии высокопроизводительного секвенирования, которые позволили за последние два десятилетия накопить значительный объем информации о молекулярной и генетической неоднородности индивидов, тканей и отдельных клеток. Сегодня активно расширяются каталоги распространенных в популяции полиморфизмов [Phan и др., 2025] и редких мутаций, ассоциированных с патологиями [Stenson и др., 2020], а данные по функциональной геномике различных тканей человека публикуются и систематизируются крупными международными консорциумами, такими как «Энциклопедия элементов ДНК» (от англ. Encyclopedia of DNA Elements, ENCODE) [ENCODE Project Consortium и др., 2020] и «Функциональная аннотация генома млекопитающих» (от англ. Functional ANnotation Of the Mammalian genome, FANTOM) [FANTOM Consortium and the RIKEN PMI and CLST (DGT) и др., 2014].

На основе накопленных данных успешно выявляются функциональные связи между изменениями последовательности генома, активности его регуляторных участков, промоторов и энхансеров, и экспрессии генов. Современные подходы, позволяющие установить связь генотипа с фенотипом, включают анализ локусов количественных признаков (QTL, quantitative trait loci) и полногеномный поиск ассоциаций (от англ. Genome-Wide Association Studies, GWAS), основанные на статистической оценке представленности генетических маркеров или вариантов генома в когортах людей с определенными признаками, и аллель-специфичный анализ, заключающийся в сравнении биологического сигнала между альтернативными аллелями гетерозиготных

вариантов, оцениваемого по результатам высокопроизводительных экспериментов функциональной геномики.

Связующим звеном, позволяющим перейти от последовательности генома к его транскрипционной активности, служат факторы транскрипции (ТФ) и их участки связывания. Выявление участков ДНК, распознаваемых данными белками, и последующее определение предпочитаемых ими нуклеотидных паттернов (мотивов связывания) – необходимое условие для дальнейшего развития регуляторной геномики, в частности при построении генных регуляторных сетей и интерпретации индивидуальных вариантов генома. Накопление данных высокопроизводительных экспериментов по определению ДНК-белковых взаимодействий, включая иммунопреципитацию хроматина с последующим секвенированием (от англ. Chromatin Immunoprecipitation Sequencing, ChIP-Seq) и систематическую эволюцию лигандов экспоненциальным обогащением (от англ. Systematic Evolution of Ligands by EXponential Enrichment, SELEX), позволили выявить мотивы связывания тысячи факторов транскрипции человека. Однако до недавнего времени около четверти потенциальных факторов транскрипции человека не имело охарактеризованного мотива связывания [Lambert и др., 2018]. Для решения этой проблемы был организован международный консорциум Codebook, поставивший своей целью определение ранее не охарактеризованных ДНК-мотивов связывания факторов транскрипции человека. Дополненный каталог мотивов позволит точнее реконструировать генные регуляторные сети на основе данных по активности промоторов и экспрессии генов в различных тканях и типах клеток человека.

Информация о тканеспецифичной экспрессии генов, влиянии на нее индивидуальных вариантов генома, и вовлеченных в регуляцию факторов транскрипции, является важной и для фундаментальной науки, и для развития персонализированной медицины. Подходы, разработанные путем сравнения молекулярных профилей здоровых и больных тканей, а также поиска вариантов генома, потенциально влияющих на индивидуальное состояние пациента, уже успешно применяются для лечения множества заболеваний, таких как синдром Лайелла [Nordmann и др., 2024] и альфа- и бета-талассемии [Sadiq и др., 2024]. Аналогичные методы сегодня ложатся в

основу терапии множества социально-значимых патологий, включая онкологические и аутоиммунные заболевания, а построение молекулярных атласов различных типов клеток человека продолжает оставаться в фокусе научного сообщества.

Тело человека включает множество поперечно-полосатых мышц, куда входит и сердечная мышца и более 600 скелетных мышц, составляющих около 40% его общей массы и необходимых для передвижения, поддержания позы, генерации тепла, защиты внутренних органов, запасаания метаболитов и эндокринной регуляции организма [Frontera, Ochala, 2014]. Помимо морфологических и анатомических различий, мышцы обладают уникальными молекулярными фенотипами, среди прочего отражающими особенности их эмбрионального развития и функциональные свойства [Noh, 2023]. Так, мышцы отличаются по скорости метаболизма, сокращения, доле медленных и быстрых волокон, ригидности, и резистентности к инсулину [Tirrell и др., 2012]. В случае патологий эти различия становятся критическими, поскольку разные группы мышц по-разному реагируют на миопатии, вызванные денервацией [Noh, 2023], аутоиммунными [Lundberg и др., 2021] и эндокринными нарушениями, митохондриальной дисфункцией [Beecher и др., 2024] и мышечной дистрофией [Ozisik и др., 2024]. Однако, систематическая оценка молекулярной гетерогенности скелетных мышц человека не проводилась, и в существующих молекулярных атласах чаще всего представлены биопсии избранных легкодоступных мышц ног. Например, база данных GTEx содержит лишь образцы икроножной мышцы [Melé и др., 2015], ENCODE также включает поясничную мышцу [Abascal и др., 2020], а максимальное число мышечных тканей, полученных и проанализированных в рамках одного проекта, составляет лишь 6 мышц ног в [Abbassi-Daloui и др., 2023].

Разнообразие профилей экспрессии генов у человека не ограничивается лишь типами клеток и тканей, но также зависит от индивидуальных особенностей организма, включающих пол, возраст и совокупность генетических вариантов, среди которых наиболее распространены однонуклеотидные варианты (ОНВ). Эффекты, оказываемые полом и возрастом на функциональность мышечной ткани, на сегодняшний день подробно изучаются в контексте сердечно-сосудистых заболеваний: у женщин до менопаузы наблюдается более благоприятный сосудистый

профиль, в то время как у мужчин риск развития подобных патологий, включающих ишемическую болезнь сердца и инфаркт, начинает расти раньше и повышен в среднем [Khoudayr и др., 2020]. Тем не менее, молекулярные изменения, происходящие в миокарде у представителей обоих полов с возрастом, до сих пор изучены лишь частично.

Чтобы охарактеризовать вклад ОНВ в фенотипы, связанные с мышцами, были проведены многочисленные полногеномные исследования ассоциаций (GWAS), включая анализ ассоциаций вариантов с массой мышц [Pei и др., 2020], силой сжатия кисти [Jones и др., 2021] и предрасположенностью к заболеваниям [Flanigan и др., 2023]. Они показали, что большинство локусов, ассоциированных с такими комплексными признаками, располагаются вне кодирующих участков генома и вероятно являются регуляторными, влияя на локальную структуру хроматина и экспрессию соседних генов [Alsheikh и др., 2022], что сильно усложняет приоритезацию, интерпретацию и дальнейшую валидацию таких вариантов. В частности, несмотря на успехи в применении высокопроизводительных методов молекулярной биологии, лишь считанные регуляторные ОНВ, связанные с наступлением юношеского идиопатического сколиоза [Ху и др., 2021] и тяжестью мышечной дистрофии Дюшенна [Weiss и др., 2018], были изучены в деталях, а их эффект – экспериментально подтвержден на молекулярном уровне.

Таким образом, общее молекулярное разнообразие скелетных мышц человека, а также вклад пола, возраста и генетической компоненты в наблюдаемую гетерогенность мышц между людьми, изучены только фрагментарно. Детальное изучение молекулярных фенотипов мышц, в том числе для создания цифрового ресурса, способного выступать в качестве основы при исследовании возрастных изменений и патологий мышц человека – важная и актуальная задача современной молекулярной биологии и биоинформатики. Каталогизация молекулярных фенотипов поперечно-полосатых мышц человека и изучение их гетерогенности в норме позволит лучше охарактеризовать изменения, происходящие при патогенезе, и поможет углубить понимание структуры тканеспецифичных регуляторных областей генома человека,

включая оценку эффектов, оказываемых индивидуальными вариантами генома на фенотипы.

Цель работы: охарактеризовать разнообразие активности регуляторных элементов, транскриптома и протеома поперечно-полосатых мышц человека.

Для достижения цели были поставлены следующие **задачи**:

1. По данным высокопроизводительного секвенирования 5'-концов кэпированных мРНК идентифицировать промоторы и транскрибируемые энхансеры, активные в скелетных мышцах человека. Сопоставить профили экспрессии генов с представленностью соответствующих белков по данным масс-спектрометрии и провести дифференциальный анализ молекулярных профилей скелетных мышц;
2. Разработать биоинформатический протокол для анализа активности мотивов связывания транскрипционных факторов на основе коллекции HOCOMOCO, дополненной мотивами ранее не охарактеризованных факторов из каталога Codebook; апробировать протокол на данных атласа промоторов и энхансеров человека FANTOM5. Определить транскрипционные факторы, активность мотивов которых объясняет различия молекулярных фенотипов скелетных мышц;
3. Оценить межиндивидуальные различия молекулярных профилей сердца и скелетных мышц; охарактеризовать различия активности транскрибируемых регуляторных элементов генома, связанные с полом и возрастом. Разработать биоинформатический протокол для проведения аллель-специфичного анализа на основе геномных картирований коротких прочтений, получаемых различными методами функциональной геномики, и идентифицировать с его помощью аллель-специфичные регуляторные варианты в промоторах и энхансерах сердца и скелетных мышц человека.

Научная новизна работы

В роли объекта исследования выступают молекулярные фенотипы поперечно-полосатых мышц человека, включая сердце и 75 различных скелетных мышц. Научная новизна работы заключается в систематизации и анализе гетерогенности наиболее

полного на сегодня набора молекулярных профилей скелетных мышц человека с применением мультиомиксного подхода, совмещающего данные по транскриптомике и протеомике мышц, а также в оценке межиндивидуальных различий активности промоторов сердца, зависящих от пола и возраста человека. В частности, в работе впервые была установлена доля транскрибируемых регуляторных элементов (ТРЭ) и генов, дифференциально экспрессируемых между скелетными мышцами, а с применением анализа активности мотивов были охарактеризованы факторы транскрипции, вносящие вклад в молекулярную гетерогенность данных тканей.

Также в рамках исследования был разработан протокол и проведен аллель-специфичный биоинформатический анализ результатов экспериментов по кэп-анализу экспрессии генов (от англ. Cap Analysis of Gene Expression followed by Sequencing, CAGE-Seq) и определению конформации хромосом с идентификацией энхансер-промоторных контактов (от англ. targeted High-throughput chromosome conformation Capture, HiCap) для установления возможных молекулярных причин наблюдаемых различий между отдельными мышцами и индивидами.

Теоретическая и практическая значимость

Теоретическая значимость работы заключается в установлении промоторов и транскрибируемых энхансеров, активных в различных скелетных мышцах человека, включая около десяти тысяч ранее не аннотированных регуляторных элементов, оценке их дифференциальной активности между тканями и согласованности транскриптома и протеома скелетных мышц. Результаты анализа дифференциального использования промоторов в образцах сердец людей разного пола и возраста расширяют знания о межиндивидуальных различиях молекулярных профилей сердца, в том числе проявляющихся в различных рисках развития сердечно-сосудистых заболеваний у мужчин и женщин. Разработанный в работе протокол аллель-специфичного анализа данных по активности ТРЭ и определенные в результате его применения регуляторные варианты генома найдут применение в дальнейших исследованиях по установлению связи генотипа с фенотипом.

Что касается практической ценности работы, атлас молекулярных фенотипов мышечных тканей человека и связанных с ними регуляторных вариантов генома позволит точнее охарактеризовать генетический компонент патологий поперечно-полосатых мышц, включая сердечно-сосудистые патологии и миопатии, и прогнозировать их избирательность в отношении определенных мышечных групп и индивидов.

Методология исследования

Основная часть результатов работы получена на основе данных CAGE-Seq – результатах экспериментов по высокопроизводительному секвенированию транскриптома, сфокусированных на 5'-концах кэпированных молекул РНК для точного определения участков инициации транскрипции. В работе используется широкий спектр биоинформатических инструментов, включая специализированные методы обработки данных CAGE-Seq, результатов масс-спектрометрии белков и других высокопроизводительных молекулярно-биологических экспериментов. Статистический анализ данных и визуализация результатов проведены с помощью языков Python и R с использованием среды разработки RStudio.

Основные положения, выносимые на защиту

1. По данным кэп-анализа экспрессии генов для 222 образцов 75 мышц и tandemной масс-спектрометрии для 60 образцов 22 мышц удалось идентифицировать 37001 транскрибируемый регуляторный элемент генома, оценить экспрессию 18329 генов и представленность 1804 белковых групп, включающих 1895 белков. Активность промоторов гена лишь частично определяет представленность кодируемого белка: коэффициент корреляции Пирсона составил 0,32. При этом совместный анализ транскриптома и протеома скелетных мышц выявил высокий уровень их молекулярной гетерогенности: более 80% генов экспрессируются дифференциально между мышцами. Наиболее выраженными собственными молекулярными фенотипами обладают экстраокулярные мышцы, язык и диафрагма; в частности, в них снижена экспрессия генов, вовлеченных в мышечную атрофию.

2. Анализ активности мотивов позволил оценить вклад 632 мотивов связывания транскрипционных факторов, включая 130 мотивов факторов из каталога Codebook, в регуляцию экспрессии генов в образцах 583 различных тканей и типов клеток человека. В применении к транскриптому скелетных мышц метод выявил 353 мотива с тканеспецифичной активностью. В частности, дифференциальная активность обнаружена для мотивов белков SIX и MEF2, вовлеченных в развитие быстрых и медленных мышечных волокон соответственно, факторов теплового шока, рецепторов тиреоидных гормонов, а также регуляторов транскрипции ALX, CASZ1 и ZBTB7A, мишени которых наиболее активны в экстраокулярных мышцах.

3. Анализ дифференциального использования 17 тысяч промоторов, активных в сердце человека, выявил изменения, зависящие от пола и возраста индивида: для 695 аутосомных генов использование альтернативных промоторов значимо ($FDR < 0,05$) различалось между мужчинами и женщинами, причем для 35 различия приводили к смене основной изоформы. В число таких случаев входит ген фактора транскрипции CREM, вовлеченного в поддержание функционирования сердца и регуляцию его ответа на β -адренергическую стимуляцию. Возраст-зависимое использование промоторов установлено для 1100 генов; часть изменений полоспецифична (в том числе для генов *UCKL1* и *HAND2*), часть согласована у мужчин и женщин. Так, вне зависимости от пола с возрастом растет относительное содержание короткой изоформы гена N-кадгерина *CDH2*, кодирующей белок, лишенный N-концевого сигнального пептида.

4. Аллель-специфичный анализ активности промоторов и энхансеров сердца и скелетных мышц человека выявил 2033 и 6653 однонуклеотидных варианта, связанных с различиями в активности транскрибируемых регуляторных элементов генома. Эти варианты значимо пересекались с локусами, ассоциированными с систолическим артериальным давлением и объемом мышц бедра соответственно, что указывает на вклад регуляторных вариантов генома в формирование комплексных признаков этих тканей. Аллель-специфичный анализ промотор-энхансерных взаимодействий в здоровых сердцах и в сердцах с двустворчатым аортальным клапаном выявил у

последних избыток вариантов со значимым аллельным дисбалансом, в том числе в локусах генов, вовлеченных в развитие патологии.

Степень достоверности данных

Достоверность полученных результатов была подтверждена путем их прямого сравнения со множеством существующих данных по эпигенетическим модификациям гистонов [Kundaje и др., 2015], активности транскрибируемых регуляторных элементов генома [Makhnovskii и др., 2022] и экспрессии генов в скелетных мышцах человека [GTEx Consortium, 2020]. Результаты аллель-специфичного анализа были подтверждены экспериментально для избранных вариантов в группе регуляции биосинтеза белка Института белка РАН с использованием люциферазного репортерного анализа.

Личный вклад автора

Личный вклад автора в представляемую работу заключается в разработке и применении биоинформатических протоколов анализа данных, включая аллель-специфичный анализ CAGE-Seq [Buyan и др., 2025; Deviatiiarov и др., 2023], 5'-концевого секвенирования РНК единичных клеток [Shevtsov и др., 2026] и HiCap [Zhigulev и др., 2026], анализ активности мотивов связывания факторов транскрипции [Jolma и др., 2026], определение активности ТРЭ и экспрессии генов скелетных мышц [Buyan и др., 2026], статистический анализ дифференциальной экспрессии генов и представленности белков, полоспецифичного и возраст-зависимого использования промоторов, а также интерпретацию и визуализацию результатов.

Публикации по теме работы

По теме диссертации опубликовано 5 статей (из них одна с совместным первым авторством) в рецензируемых журналах, индексируемых в базах данных Scopus, Web of Science и РИНЦ (все издания входят в первый квартиль согласно перечням «Белого списка» и SCImago Journal Rank), а также 5 тезисов в сборниках трудов научных конференций.

Апробация работы

Результаты диссертационной работы представлены в форме 1 постерного и 8 устных докладов на 9 российских и международных научных конференциях, включая «Биология и генетика мышц» (2022 г., Казань), 14-ю Международную школу молодых ученых «Системная биология и биоинформатика» (SBB, 2023 г., Новосибирск), 11-ю и 12-ю Московскую конференцию по вычислительной молекулярной биологии (МССМВ, 2023 и 2025 гг., Москва), 14-ю Международную мультиконференцию «Биоинформатика регуляции и структуры геномов / системная биология» (BGRS/SB, 2024 г., Новосибирск), 29-ю Ежегодную международную конференцию по исследованиям в области вычислительной молекулярной биологии (RECOMB и сателлитный симпозиум RECOMB-Genetics, 2025 г., Сеул), XXIV Конгресс Вельтищева «Инновационные технологии в педиатрии и детской хирургии» (2025 г., Москва) и Международную научно-практическую конференцию «Молекулярная диагностика» (2025 г., Москва).

Объем и структура работы

Диссертация состоит из Введения, Обзора литературы, Материалов и методов, Resultados, Заключение, Выводов и Списка литературы. Работа изложена на 126 страницах и иллюстрирована 39 рисунками. Список цитируемой литературы содержит 343 наименования.

Содержание работы

Введение

Во введении описывается актуальность темы и формулируется цель исследования, состоящая в оценке молекулярной гетерогенности мышц человека. Определяются задачи работы, подчеркивается ее новизна, теоретическая и практическая ценность, дается оценка личному вкладу автора и степени достоверности полученных результатов.

1. Обзор литературы

Глава 1 состоит из четырех подразделов и посвящена обзору литературы, касающейся строения и функции мышечных тканей человека, современным представлениям об их молекулярной гетерогенности и о влиянии индивидуальных вариантов генома на различные наблюдаемые и молекулярные признаки мышц.

2. Материалы и методы

Во второй главе описываются биоинформатические инструменты, применявшиеся в работе для анализа результатов экспериментов CAGE-Seq, аллель-специфичного анализа данных по активности ТРЭ и энхансер-промоторным ДНК-ДНК контактам.

2.1. Анализ данных CAGE-Seq

Обработка результатов CAGE-Seq 225 образцов 75 скелетных мышц (GEO GSE310878) включала картирование прочтений программой HISAT2 [Kim и др., 2019] на геном человека hg38 и контроль качества, позволивший отобрать 222 образца для дальнейшего анализа. Поиск транскрибируемых регуляторных элементов (ТРЭ) был проведен инструментом CAGEfightR [Thodberg и др., 2019]: в результате многоступенчатой фильтрации кластеров 5'-концов прочтений CAGE-Seq был идентифицирован 37001 ТРЭ, активный в скелетных мышцах. Проведенный отбор включал фильтрацию кластеров по минимальной доле прочтений, содержащих нематричные гуанины, добавляемые обратной транскриптазой при прочтении структуры кэпа. Данные CAGE-Seq 109 образцов 21 здорового сердца и 10 сердец с

кардиомиопатией, полученных от 17 мужчин и 14 женщин в возрасте от 26 до 71 года (GEO GSE150736) были предобработаны коллегами в работе Deviatiiarov и др.; для анализа дифференциальной активности ТРЭ использовалось 65298 отдельных транскрибируемых регуляторных элементов генома, из них 17219 соответствовали основным участкам инициации транскрипции, расположенным в промоторных областях генов, для которых изучалось дифференциальное использование промоторов (изменение относительного вклада в тотальную экспрессию гена). Для проверки результатов анализа активности промоторов на основе CAGE-Seq, с использованием программы kallisto [Bray и др., 2016] были обработаны результаты секвенирования РНК 20 образцов 7 здоровых сердец из того же набора для оценки представленности транскриптов (GEO GSE212695).

2.2. Оценка представленности белков по результатам масс-спектрометрии

В результате обработки данных тандемной масс-спектрометрии, полученных для 60 образцов 22 скелетных мышц человека (PRIDE PXD072528), были получены количественные оценки представленности 1895 белков (1804 белковых групп), которые после нормализации сравнивались с результатами CAGE-Seq. Для этого мы отобрали 1750 аутосомных генов, исключив влияние пола и митохондриальных генов, и оценили корреляцию между средним уровнем их экспрессии и общей представленностью соответствующих белков. Для сравнения протеома и транскриптома отдельных образцов мышц мы также вычислили z-оценки $\log_2$ интенсивностей репортерных ионов (по данным масс-спектрометрии) и $\log_2$ CPM (от англ. Counts Per Million, число прочтений, нормированное на размер библиотеки, по данным CAGE-Seq) отдельно для каждого гена, приведя их к единой шкале.

2.3. Определение тканеспецифичной экспрессии

Поиск дифференциально экспрессируемых ТРЭ и генов проводился путем анализа данных CAGE-Seq инструментом edgeR [Chen и др., 2025], а дифференциально представленных белков – R-пакетом limma [Ritchie и др., 2015].

В статистических моделях использовался дизайн «~ доноры + мышцы», чтобы учесть индивидуальные особенности доноров. Тесты включили попарные сравнения определенных групп мышц против остальных образцов, дисперсионный анализ, а также проверку на альтернативное использование промоторов и белковых изоформ.

Мышцы с уникальными профилями активности ТРЭ выявлялись с использованием пакета R fgsea [Korotkevich и др., 2016] для оценки обогащений тканей промоторами и энхансерами с наибольшей разницей между максимальным (по мышцам) и медианным $\log_2$ CPM.

Полоспецифичная и возраст-зависимая активность промоторов сердца оценивалась программой edgeR отдельно для желудочков и предсердий здоровых доноров и пациентов с кардиомиопатией с использованием в качестве биологической ковариаты категориальной переменной «пол» или количественной переменной «возраст».

2.4. Оценка относительного вклада ТРЭ в экспрессию генов

Вклад отдельных ТРЭ рассчитывался как доля их CPM относительно суммы по гену. Учитывая, что дифференциальная экспрессия основного промотора может искусственно создавать видимость переключения между альтернативными изоформами, из анализа были исключены слабо экспрессируемые ТРЭ и одиночные ТРЭ генов, не имеющие сопоставимых по покрытию альтернативных промоторов, в итоге было проанализировано 12928 регуляторных элементов.

2.5. Анализ активности мотивов

Для анализа активности мотивов связывания ТФ был использован Python-пакет MARADONER [Meshcheryakov и др., 2026], на вход которому подавались: матрица $\log_2$ CPM для 36120 однонаправленных ТРЭ, а также «суммарная занятость» последовательностей промоторов участками, схожими с мотивами связывания ТФ согласно базе данных HOCOMOCO v14 [Vorontsov и др., 2024]. Метод моделирует активность промотора как линейную функцию от оценок сходства его последовательности с мотивами связывания ТФ, где коэффициенты отражают тканеспецифичные активности этих мотивов, Рисунок 1. В результате были оценены

активности 478 репрезентативных мотивов связывания (соответствующих ТФ, экспрессирующимся в образцах скелетных мышц), их вариабельность между мышцами и вклад в тканеспецифичную регуляцию ТРЭ.

Запуск MARADONER на данных FANTOM5 [The FANTOM Consortium and the RIKEN PMI and CLST (DGT), 2014] по активности 209374 промоторов генов в 583 различных тканях и типах клеток человека проводился аналогичным образом с использованием 632 репрезентативных мотивов, соответствующих экспрессирующимся в образцах ТФ, включая 130 репрезентативных мотивов, впервые охарактеризованных в рамках проекта Codebook.

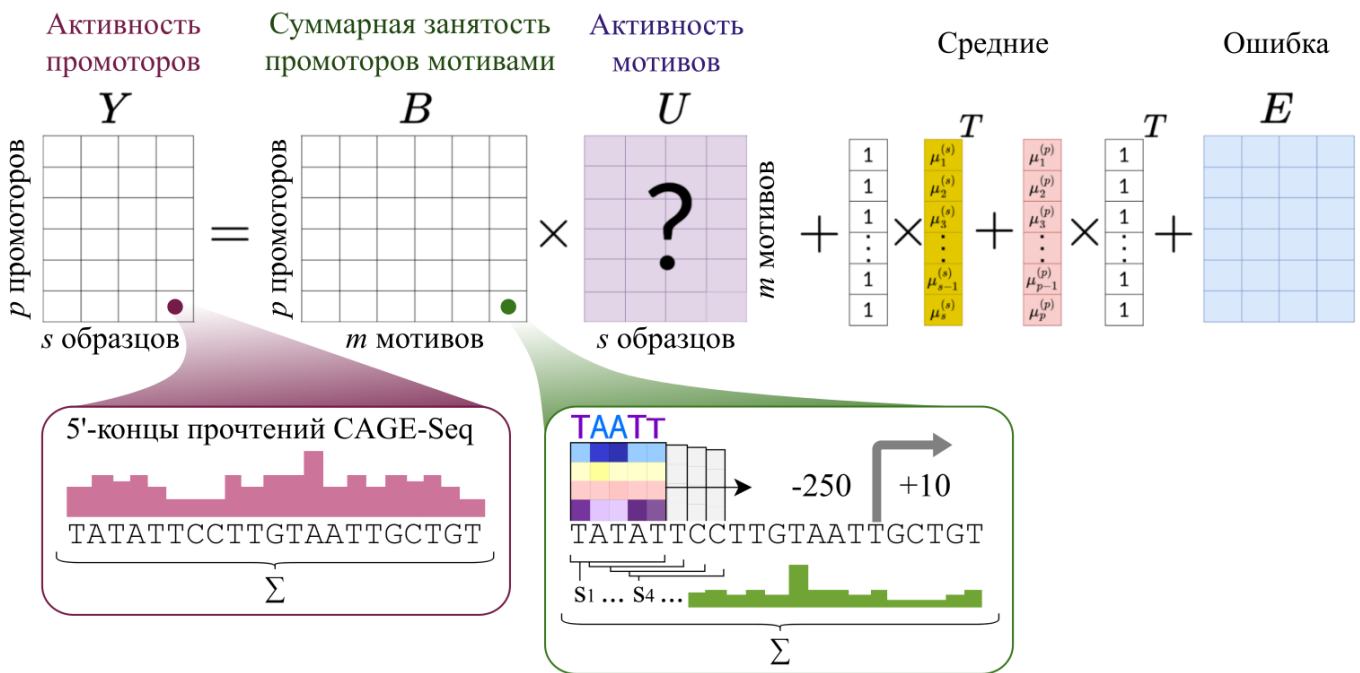


Рисунок 1 – Модель, используемая в MARADONER

2.6. Аллель-специфичный биоинформатический анализ

Для идентификации аллель-специфичных вариантов (ACB) был проведен поиск гетерозиготных ОНВ по данным CAGE-Seq скелетных мышц и сердца, а также HiCap 16 образцов аорты, с использованием программы bcftools [Genovese и др., 2024], после чего к первым двум применялся метод WASP [van de Geijn и др., 2015] для устранения ошибок картирования прочтений. Статистическая оценка аллельного дисбаланса выполнялась MIXALIME [Buayan и др., 2025] с использованием менее строгой отрицательной биномиальной (для скелетных мышц, 2-4 образца на мышцу) или более

строгой бета отрицательной биномиальной (для сердца и аорты, 109 и 16 образцов на мышцу, соответственно) моделей.

2.7. Анализ стратифицированной наследуемости

Вклад ACB в наследуемость признаков скелетных мышц был оценен с использованием результатов GWAS, полученных из баз данных UK Biobank [Karczewski и др., 2025] и NHGRI-EBI [Sollis и др., 2023], и стратифицированного регрессионного анализа по оценкам неравновесия по сцеплению (от англ. Stratified LD SCore regression, S-LDSC) [Finucane и др., 2015]. Из 58 мышечных фенотипов было отобрано 32 наследуемых признака, генетически не скоррелированных друг с другом.

2.8. Список внешних аннотаций, использованных в работе

Аннотация ТРЭ скелетных мышц проводилась с использованием данных по доступности хроматина и модификации гистонов из ENCODE [Zhang и др., 2020], Roadmap Epigenomics [Kundaje и др., 2015] и [Galle и др., 2022] (GEO GSE195854). Обнаруженные промоторы и энхансеры сопоставлялись с известными ТРЭ мышц, идентифицированными в FANTOM5 [FANTOM Consortium and the RIKEN PMI and CLST (DGT) и др., 2014] и [Makhnovskii и др., 2022] (GEO GSE164081). Экспрессия генов сопоставлялась с данными по секвенированию РНК из GTEx (v.10) [GTEx Consortium, 2020]. ACB сравнивались с локусами, ассоциированными с экспрессией генов (от англ. expression QTL, eQTL) в скелетных мышцах согласно GTEx (v.10), сайтами аллель-специфичного связывания ТФ из ADAstra (v.6.1) [Abramov и др., 2021] и вариантами, влияющими на доступность хроматина согласно UDACHA (v.1.0.3) [Buyan и др., 2025].

Для аннотации ACB сердца использовались eQTL желудочков и предсердий из базы данных GTEx (v.8), ACB из ADAstra (v.5.1.3) и локусы, ассоциированные с систолическим артериальным давлением по результатам GWAS из Pan-UK Biobank (идентификатор фенотипа 4080).

3. Результаты

Глава 3 диссертационной работы посвящена атласу активности ТРЭ и представленности белков в скелетных мышцах человека, результатам исследования их молекулярной гетерогенности и оценке вклада индивидуальных вариантов генома в активность промоторов и энхансеров мышц, а также в наблюдаемые мышечные признаки.

3.1. Создание транскриптомного и протеомного атласа скелетных мышц человека

Молекулярные профили мышц были объединены в атлас FANTOMUS (от англ. Functional ANnotation Of human skeletal MUScle genome), описывающий активность 37001 ТРЭ, включая 881 двунаправленный энхансер и промоторы 18329 генов, на основании данных CAGE-Seq для 222 образцов 75 мышц. Сравнение с базами данных по экспрессии генов (GTEx), и активности промоторов и энхансеров (FANTOM5, [Makhnovskii и др., 2022]) подтвердило высокую достоверность атласа, позволило выявить на 30% больше экспрессируемых в мышцах генов и обнаружить около 10 тысяч новых регуляторных элементов, Рисунок 2.

В дополнение к активности ТРЭ, нами был проведен анализ протеома скелетных мышц с использованием оценок представленности 1895 белков (1804 белковых групп), полученных по результатам тандемной масс-спектрометрии. Сравнение уровней РНК и белков показало слабую общую корреляцию ($r = 0,32$), Рисунок 3А.

Анализ согласованности транскриптома и протеома для каждого образца мышц в отдельности выявил наиболее высокую корреляцию у экстраокулярных мышц ($r_s = 0,34$), что может указывать на ослабленный посттранскрипционный контроль экспрессии в этих тканях, Рисунок 3Б.

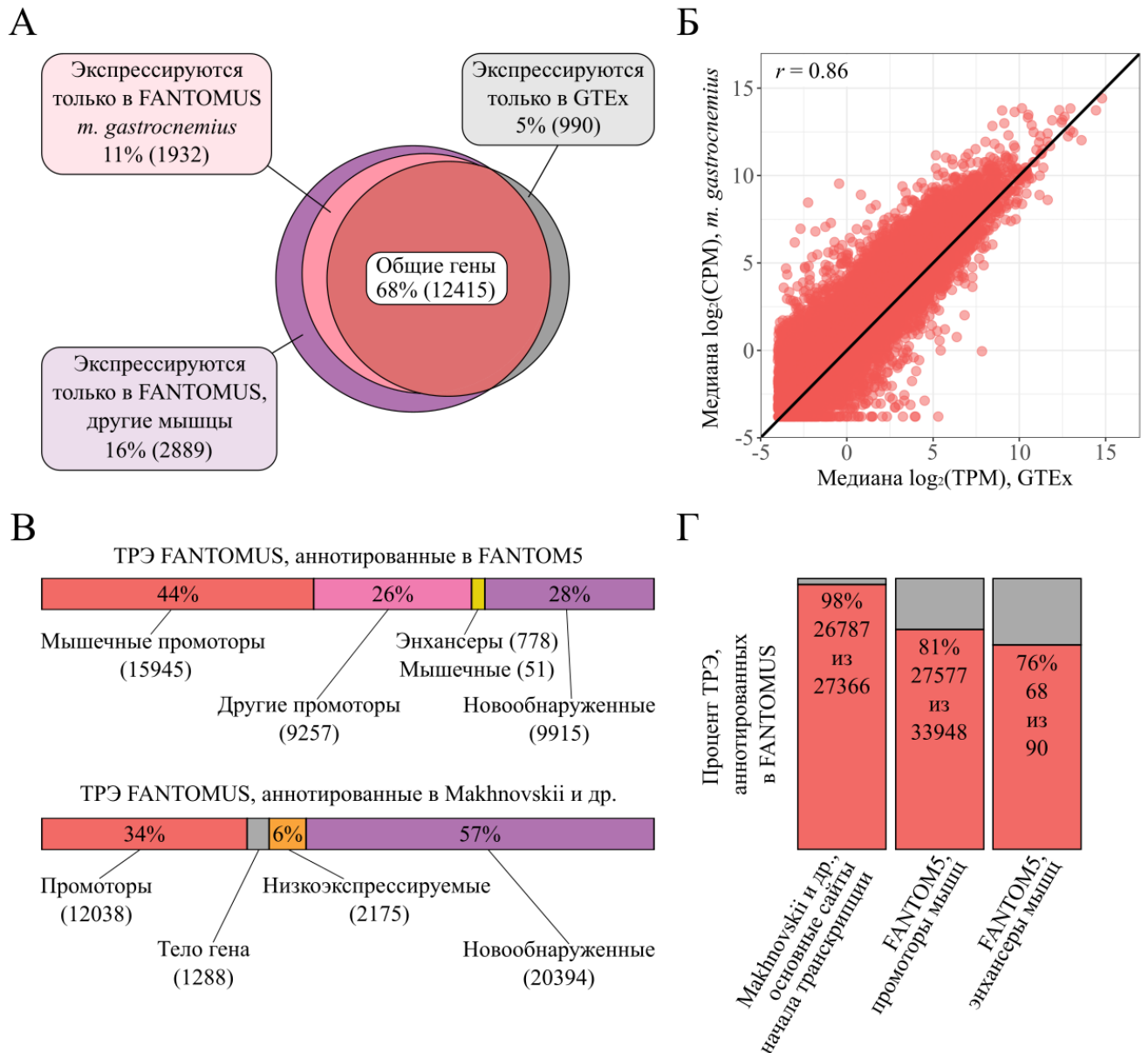
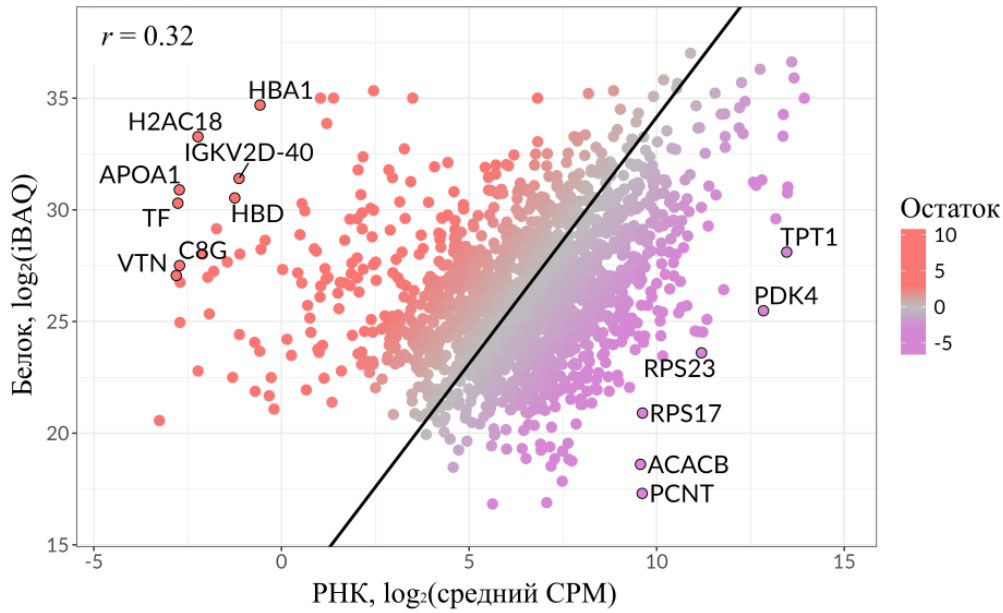


Рисунок 2 – **А.** Диаграмма Венна для белок-кодирующих генов и днРНК, экспрессирующихся в скелетных мышцах FANTOMUS и GTEx. **Б.** Корреляция между оценками экспрессии генов в скелетной мышце *m. gastrocnemius*, полученными в GTEx (ось X, $\log_2\text{TPM}$) и FANTOMUS (ось Y, $\log_2\text{CPM}$); линия соответствует диагонали $Y = X$. **В.** Процент ТРЭ FANTOMUS, аннотированных различными категориями из FANTOM5 (сверху) и Makhnovskii и др. (снизу). **Г.** Процент аннотированных в FANTOMUS ТРЭ из Makhnovskii и др. (слева) или FANTOM5 (справа)

А



Б

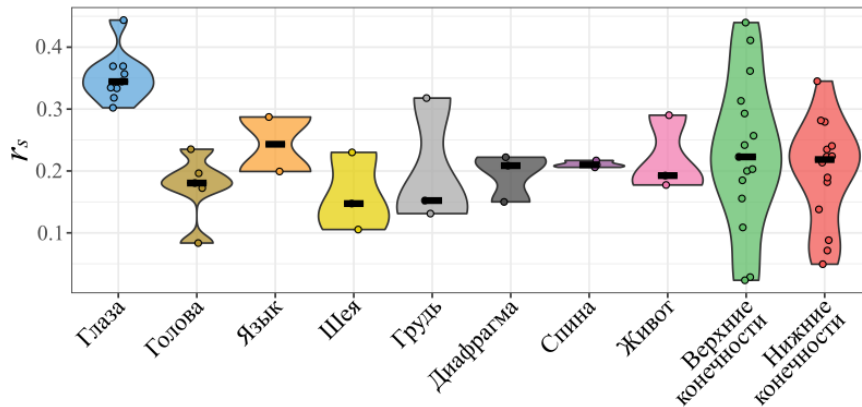


Рисунок 3 – А. Корреляция между оценками общего содержания РНК (ось X, log₂ среднего CPM) и белка (ось Y, log₂ iBAQ, от англ. intensity-Based Absolute Quantification value) в скелетных мышцах. Цвет отображает отклонение точек от общего тренда, оцененного с использованием регрессии Деминга (сплошная линия). Б. Скрипичные диаграммы, соответствующие коэффициентам корреляции Спирмена между z-оценками количества РНК (log₂ CPM) и белка (log₂ интенсивности репортерных ионов), вычисленным отдельно для каждого образца; черными штрихами выделены медианы

3.2. Характеристика молекулярных фенотипов скелетных мышц

Уникальные молекулярные профили были обнаружены у экстраокулярных мышц (ЭОМ), языка и диафрагмы как на уровне экспрессии генов, так и представленности белков, Рисунок 4. Молекулярный фенотип ЭОМ отличался повышенной экспрессией генов нейрогенеза и гомеостаза сетчатки, а также содержанием белков, участвующих

в митохондриальном дыхании. Для языка были характерны процессы эпидермального развития и гуморального иммунного ответа, а для диафрагмы – связанные с иммунным ответом, включая фагоцитоз и хемотаксис нейтрофилов, и окислительным фосфорилированием.

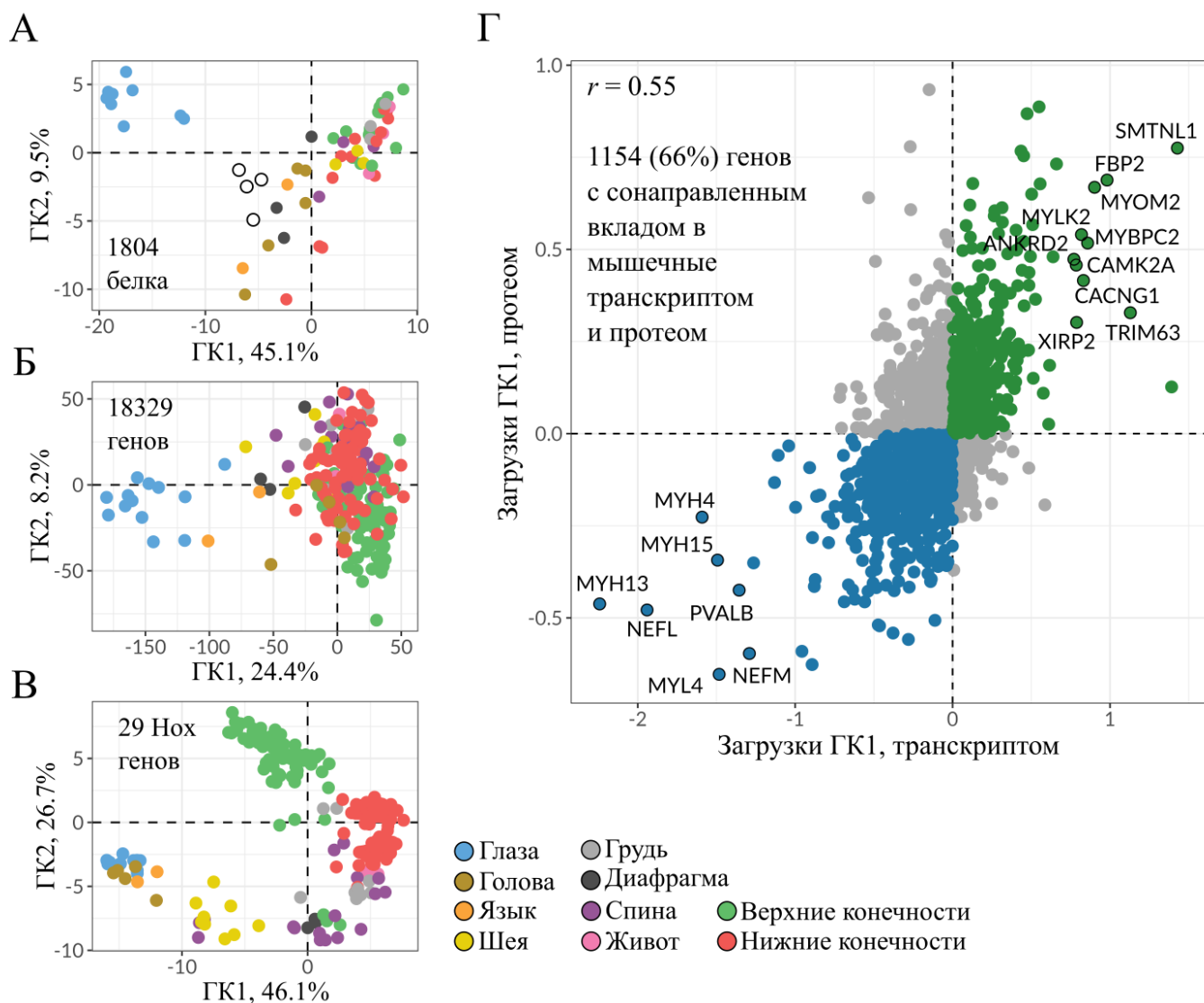


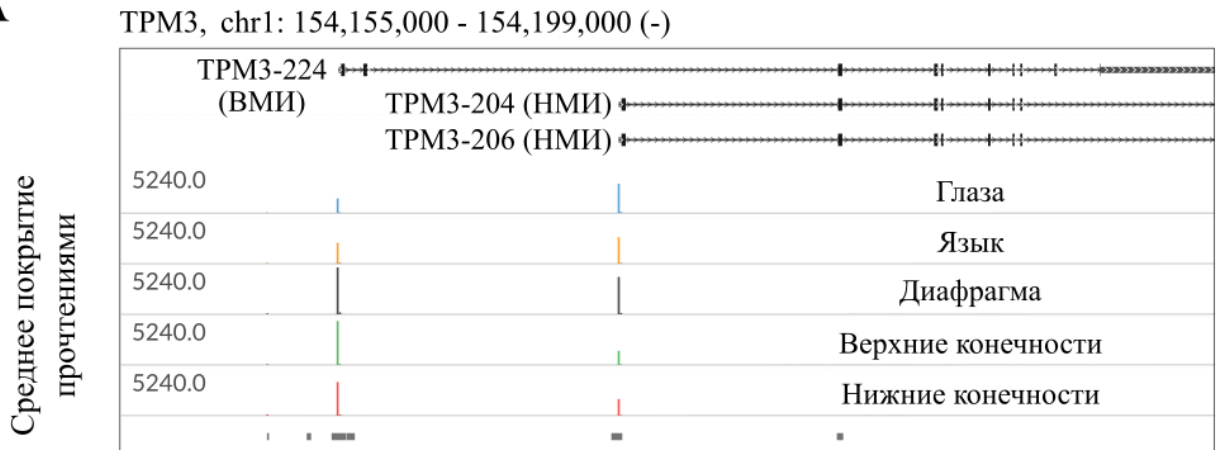
Рисунок 4 – Результаты анализа образцов мышц методом главных компонент с использованием оценок: представленности 1804 белковых групп (А), экспрессии 18329 генов (Б) и экспрессии 29 Нох-генов (В). Оси X и Y: первая и вторая главные компоненты, проценты соответствуют долям объясненной дисперсии; цветом обозначены части тела. Г. Диаграмма рассеяния для загрузок белков (ось Y) и генов (ось X) первых главных компонент анализов, проведенных на протеоме (А) и транскриптоме (Б) скелетных мышц, соответственно. Цвет отображает знак загрузок: отрицательные значения характерны для генов, чья представленность на уровне РНК и белков в ЭОМ, языке и диафрагме повышена (синий), а положительные – понижена (зеленый)

Дисперсионный анализ показал, что большинство генов и белков имеют значимую ($FDR < 0,05$) тканеспецифичную экспрессию и представленность, соответственно: 16033 из 18329 (87%) генов и 1469 из 1804 (81%) белковых групп. Среди них наиболее выраженные паттерны дифференциальной экспрессии демонстрировали: Нох-гены, участвующие в регуляции эмбрионального развития тканей [Terry и др., 2018], маркеры быстрых и медленных мышечных волокон [Abbassi-Daloi и др., 2023], а также гены, ассоциированные с атрофией мышц, включая *TRIM63* [Bodine и др., 2001] и *CACNG1* [Okada и др., 2021], экспрессия которых на транскриптомном и протеомном уровнях была заметно снижена в ЭОМ, языке и диафрагме.

Отдельно были изучены гены, мутации в которых ассоциированы с мышечными дистрофиями. В ЭОМ наблюдалось снижение экспрессии генов, связанных с мышечной дистрофией Дюшенна (*DMD*), дисферлино- (*DYSF*) и кальпаинопатией (*CAPN3*), и повышение компонентов альтернативных компенсаторных систем (функционального аналога дистрофина утрофина, *UTRN* [Porter и др., 1998], но только на уровне РНК, а также интегрин бета-1, *ITGB1*), что может объяснять их устойчивость к данным патологиям. Более того, мышцы ног, по-разному восприимчивые к *FKRP*-ассоциированной дистрогликанопатии согласно данным магнитно-резонансной томографии [Morishima, Schoser, 2023], демонстрировали значимую дифференциальную экспрессию гена *FKRP*: *P*-значение дисперсионного анализа, поправленное на общее число рассмотренных миопатий, составило 0,005.

Наконец, анализ альтернативных промоторов генов выявил 1648 случаев выраженного тканеспецифичного использования изоформ транскриптов, при котором разница относительных вкладов промоторов в тотальную экспрессию генов, рассчитанная между мышцами, превышает 25%. Среди наиболее ярких примеров: ключевой регулятор формирования экстраокулярных мышц *PITX2* [Kuriki и др., 2024], кардиоспецифичная атипичная протеинкиназа *ALPK2* [Hofsteen и др., 2018] и компонент цитоскелета *TRPM3* [Lambert, Gussoni, 2023]. Для последнего также удалось различить белковые изоформы, чья дифференциальная представленность соответствовала паттерну активности промоторов и свидетельствовала о сниженной экспрессии высокомолекулярной изоформы в ЭОМ, Рисунок 5.

А



Б

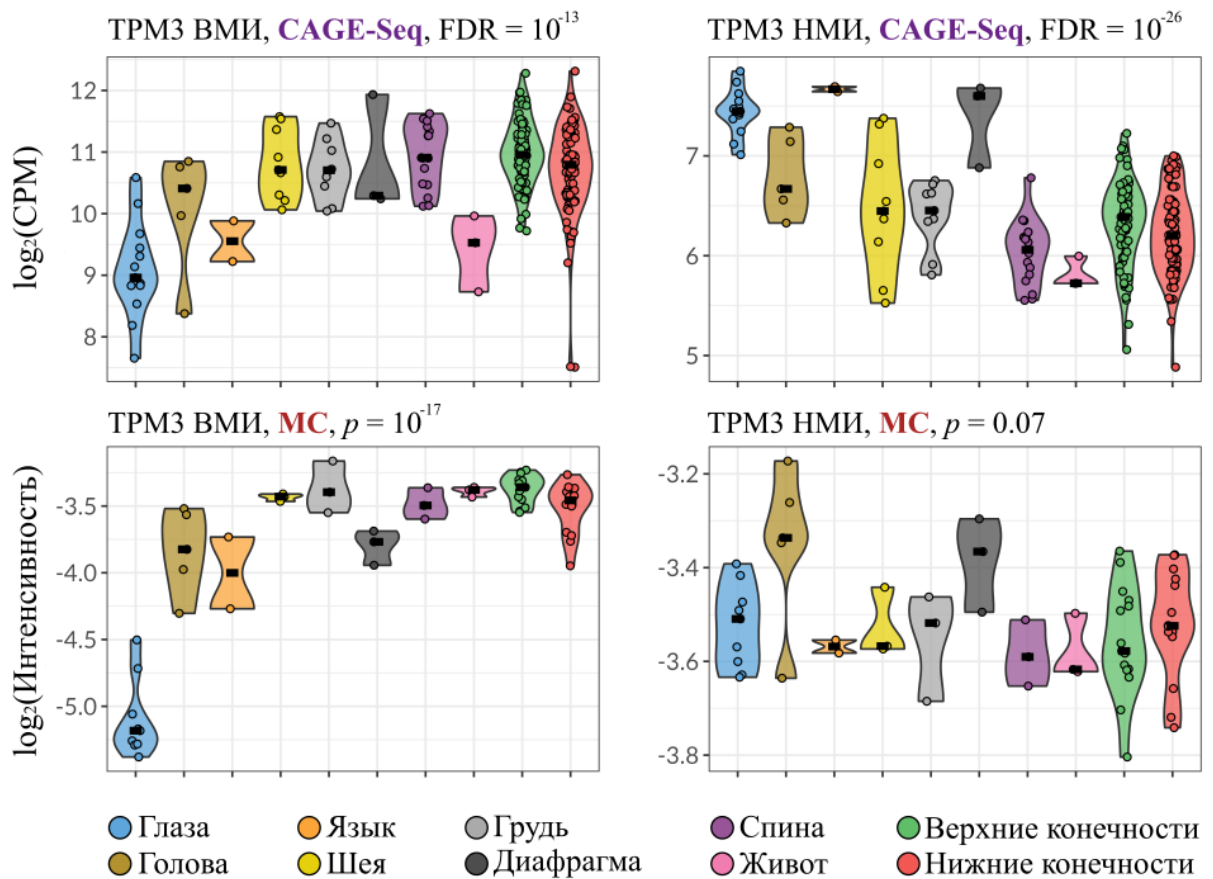


Рисунок 5 – А. Усредненные профили покрытия прочтениями CAGE-Seq гена *TPM3* для ЭОМ, языка, диафрагмы, верхних и нижних конечностей. Снизу отображены FANTOMUS ТРЭ для соответствующей гену цепи. Б. Распределения уровня активности ТРЭ и представленности белков, соответствующих альтернативно используемым изоформам *TPM3*. Сверху указаны значимости из дифференциального анализа (ANOVA), цветом обозначены части тела, черные штрихи соответствуют медианам

3.3. Применение анализа активности мотивов для выявления тканеспецифичных регуляторов транскрипции

Для выявления факторов транскрипции, определяющих молекулярную гетерогенность мышц и других тканей человека, необходима информация об их участках связывания в геноме и генах-мишенях, которую можно получить либо в результате прямого эксперимента, либо путем биоинформатического анализа последовательностей промоторов на их соответствие известным ДНК-мотивам связывания. В рамках проекта Codebook было охарактеризовано более ста новых мотивов связывания ТФ, для которых мы провели анализ дифференциальной активности с помощью MARADONER и выявили значимую тканеспецифичность. При этом 25 наиболее значимых мотивов формировали четыре группы: преимущественно активные в иммунных клетках, в иммунных и опухолевых клетках, в нервных тканях и в мезодермальных клетках человека, Рисунок 6.

Применительно к скелетным мышцам, мы провели анализ активности 478 мотивов HOCOMOCO v14, соответствующих 785 экспрессируемым в мышцах ТФ. В результате 353 мотива продемонстрировали значимую дифференциальную активность между различными группами мышц, Рисунок 7А. Среди них присутствовали мотивы связывания белков MEF2 и SIX, характерных для медленных [Noh, 2023] и быстрых [Sakakibara и др., 2014] мышечных волокон, соответственно, а также HOX9-13, активных в дистальных мышцах конечностей, Рисунок 7Б. Факторы теплового шока были более активны в мышцах бедер и спины, что может отражать эффекты сидячего образа жизни и саркопении [McArdle и др., 2006], а рецепторы гормонов щитовидной железы, участвующие в формировании быстрого мышечного фенотипа [Noh, 2023], – в дистальных мышцах рук, Рисунок 7В-Г. Последними рассмотренными мотивами стали сайты связывания белков ALX1/ALX4, ZBTB7A, MBD3 и CASZ1, специфически активные в ЭОМ и связанные с развитием периокулярных тканей [Puyanar и др., 2022], ассоциированные с макроцефалией и умственной отсталостью [Ohishi и др., 2020], регулирующие транскрипцию генов нейрогенеза [Mai и др., 2025] и дифференцировки скелетных мышц [Liu и др., 2020], соответственно, Рисунок 7Д.

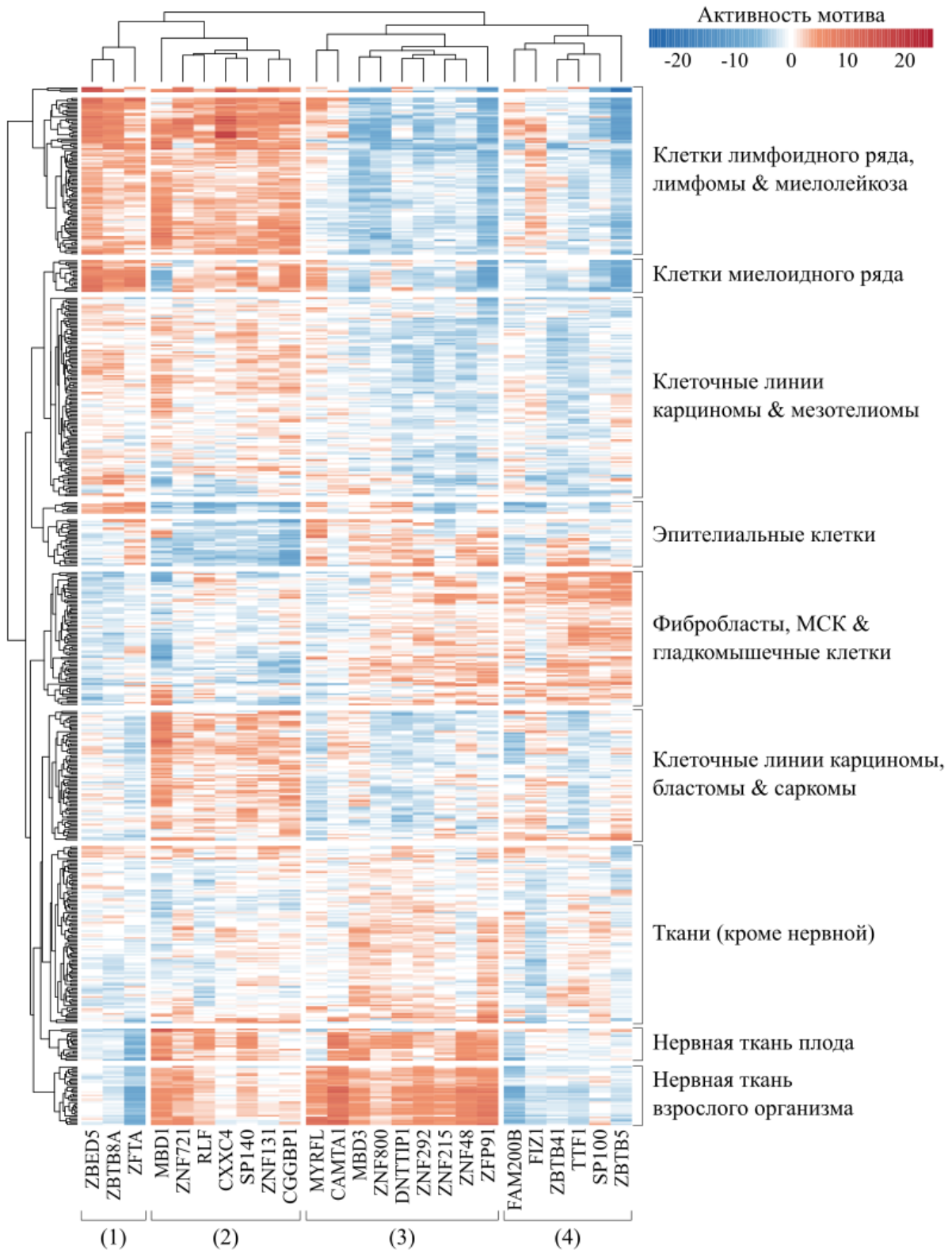


Рисунок 6 – Тепловая карта активностей новых мотивов ТФ проекта Codebook. По оси Y представлены 583 образца FANTOM5; кластеры образцов были аннотированы в соответствии с преобладающими типами клеток и тканей. По оси X представлены транскрипционные факторы, соответствующие 25 группам мотивов Codebook, демонстрирующим наиболее значимую дифференциальную активность между образцами FANTOM5

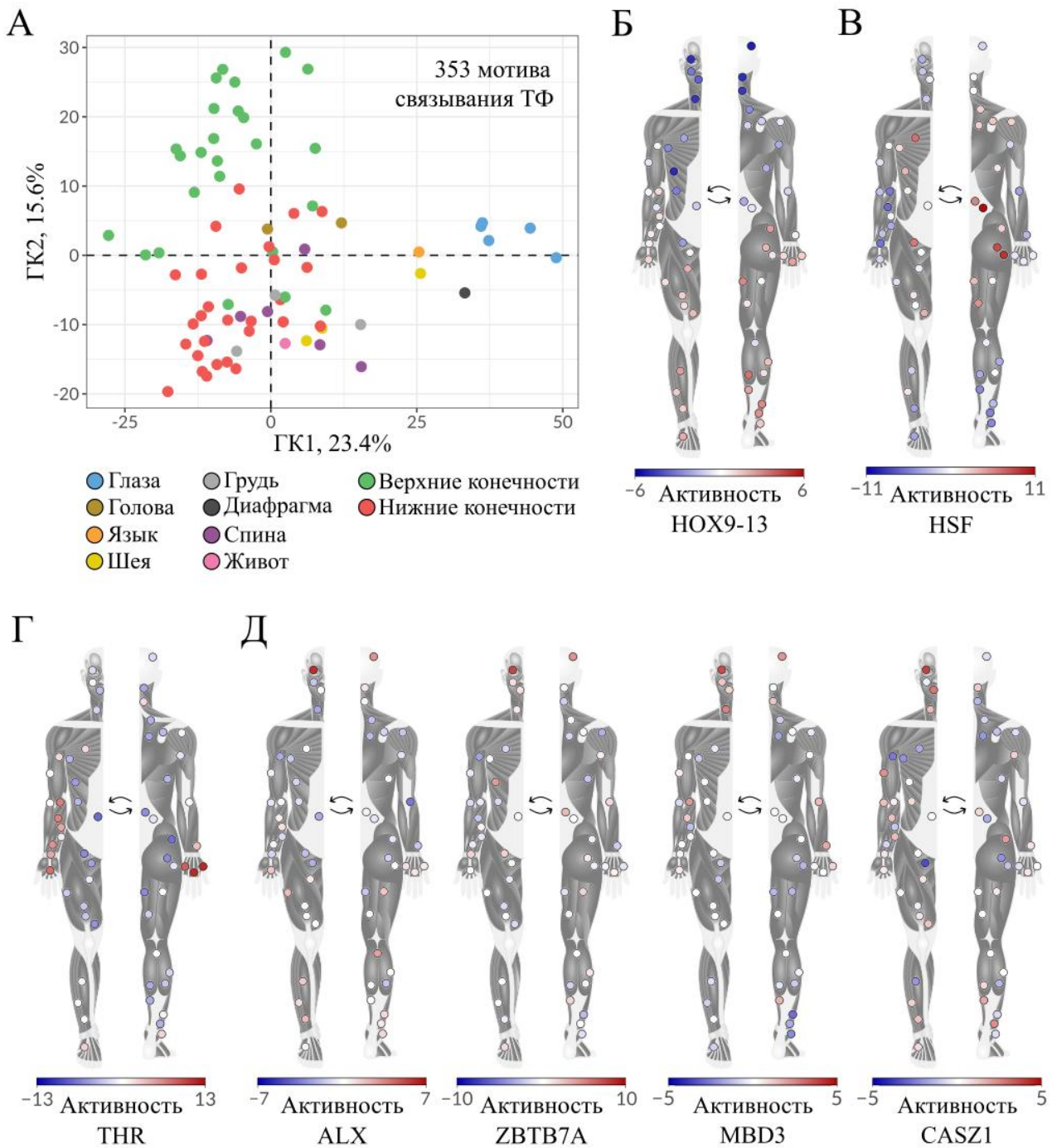


Рисунок 7 – А. Результаты анализа 75 мышц методом главных компонент с использованием оценок активности 353 мотивов, значимо (ANOVA FDR < 0,05) различающихся между образцами. Оси X и Y: первая и вторая главные компоненты, проценты соответствуют долям объясненной дисперсии, цветом обозначены части тела. Б. Распределение активности мотива связывания HOX9-13 в мышцах человека. Цвет отображает нормированные отклонения активности мотива в каждой группе образцов, оцененные с использованием MARADONER. В-Д. Распределение активности мотивов связывания факторов теплового шока (от англ. Heat Shock Factor, HSF), рецепторов гормонов щитовидной железы (от англ. Thyroid Hormone Receptor, THR), ALX, ZBTB7A, MBD3 и CASZ1, изображенные аналогично (Б)

3.4. Оценка межиндивидуальных различий активности ТРЭ в сердце, обуславливаемых возрастом и полом человека

Для оценки межиндивидуальной гетерогенности сердечной ткани был проанализирован атлас активности промоторов и энхансеров 109 образцов сердец 31 донора разного пола и возраста. В результате были выявлены 1563 полоспецифичных (1453 аутосомных) и 13890 возраст-зависимых транскрибируемых регуляторных элементов, при этом у мужчин была выше активность генов сократительного аппарата мышцы, у женщин – генов иммунного ответа, тогда как старение сопровождалось снижением активности генов окислительного фосфорилирования.

Альтернативное использование промоторов было обнаружено у 720 генов (695 аутосомных) при сравнении полов и у 1100 генов при оценке эффекта возраста. Наиболее воспроизводимым полоспецифичным эффектом является повышение у мужчин активности промоторов, с которых экспрессируются репрессорные изоформы транскрипционного фактора CREM, связанные с повышенной β -адренергической стимуляцией сердца и его аритмогенным ремоделированием [Seidl и др., 2020], Рисунок 8. Возрастные эффекты включали снижение (у женщин) относительной экспрессии одной из белок-кодирующих изоформ *UCKL1*, увеличение (у обоих полов) вклада короткой, лишенной сигнального пептида изоформы гена N-кадгерина в общую экспрессию *CDH2* и повышение доли прочтений (у мужчин), приходящихся на промотор нестабильной изоформы *HAND2*, подвергающейся нонсенс-опосредованному распаду. Эти результаты показывают, что пол- и возраст-зависимый выбор альтернативных промоторов генов может существенно влиять на ремоделирование, механическую целостность и функциональное состояние сердца, а также на риск развития сердечно-сосудистых заболеваний.

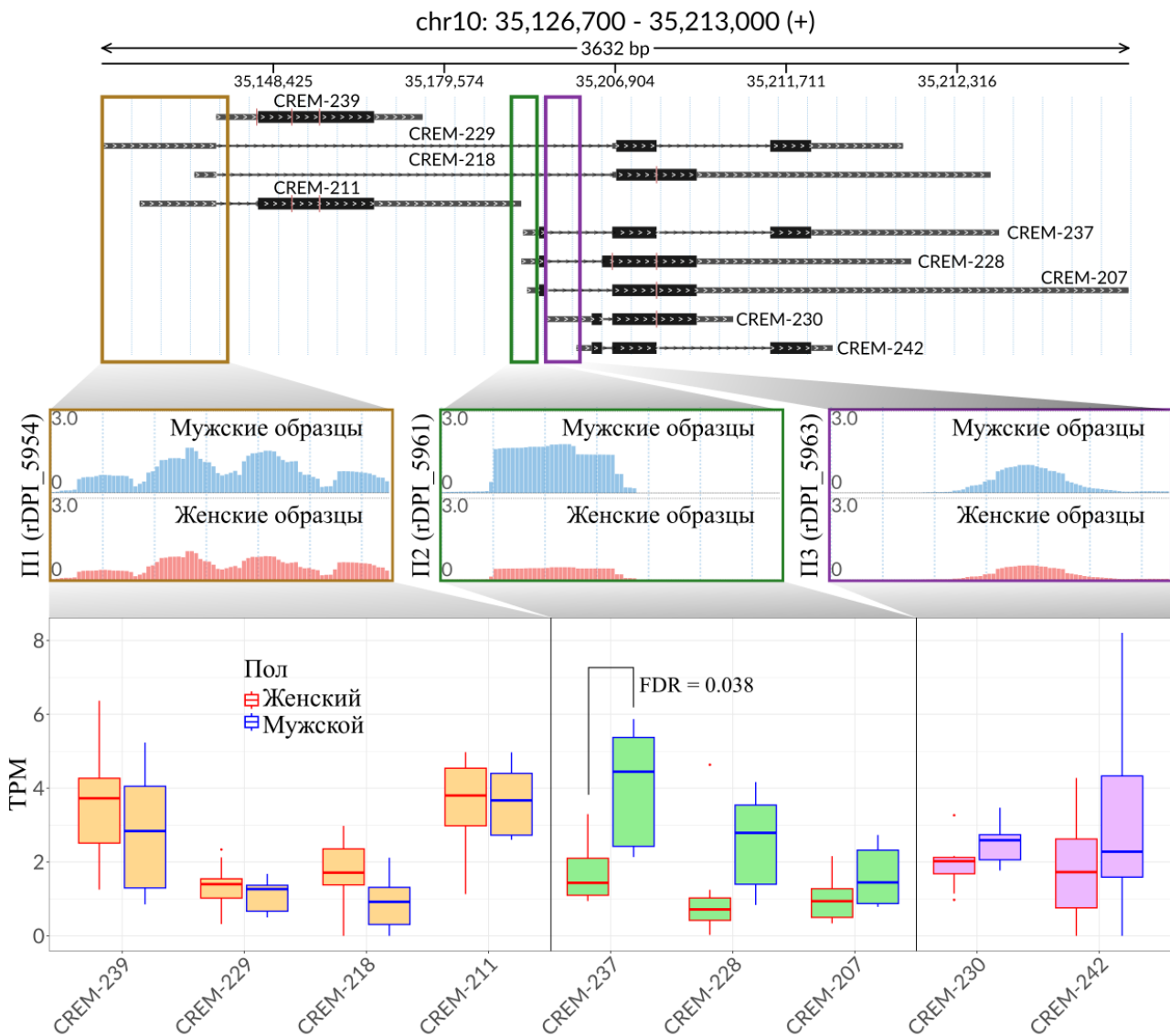


Рисунок 8 – Схематическое изображение структуры гена *CREM*, иллюстрирующее 9 белок-кодирующих изоформ, наиболее экспрессируемых в сердце, и профили покрытия соответствующих им промоторов прочтениями CAGE-Seq. Ось Y: нормированное число прочтений (CPM). Ниже представлен график типа «ящики с усами» (минимум, 25%, 50% и 75% квантиль, и максимум за исключением выбросов), иллюстрирующий представленность различных изоформ *CREM*; цвет соответствует стартам транскрипции, цвет обводки – полу, значимость приведена для результатов анализа дифференциальной экспрессии, проведенного с использованием edgeR

3.5. Аллель-специфичный анализ позволяет приоритезировать регуляторные варианты, влияющие на молекулярный фенотип мышц

Изучение влияния генетических вариантов на активность ТРЭ мышц было проведено с использованием аллель-специфичного анализа данных CAGE-Seq. В скелетных мышцах удалось выявить 6653 аллель-специфичных варианта (АСВ), значительная часть которых совпадала с eQTL из GTEx (в среднем 60%, P -значение одностороннего точного теста Фишера = 10^{-13} , отношение шансов = 1,3), сайтами аллель-специфичного связывания транскрипционных факторов из ADASTRA (54%, P -значение = 10^{-146} , отношение шансов = 2,5) и вариантами, влияющими на доступность хроматина согласно UDACHA (10%, P -значение = 10^{-25} , отношение шансов = 2,4), Рисунок 9А. Для ряда АСВ, расположенных в промоторах генов *MTOR*, *RPL32*, *SYPL2* и *ZDHHC5*, коллегам из Института Белка РАН удалось экспериментально подтвердить их влияние на активность ТРЭ с использованием люциферазного репортерного анализа. Один из таких примеров, rs2295079, расположенный в промоторе *MTOR*, приведен на Рисунке 9Б.

Аналогичный анализ данных CAGE-Seq сердца позволил выявить 2033 АСВ, также обогащенных eQTL и сайтами связывания ТФ. Для части сердечных вариантов, присутствующих у доноров в гомозиготных состояниях, было показано соответствие между аллель-специфичным эффектом и различиями активности ТРЭ у гомозиготных носителей разных аллелей, тем самым подтвердив функциональность обнаруживаемого аллельного дисбаланса.

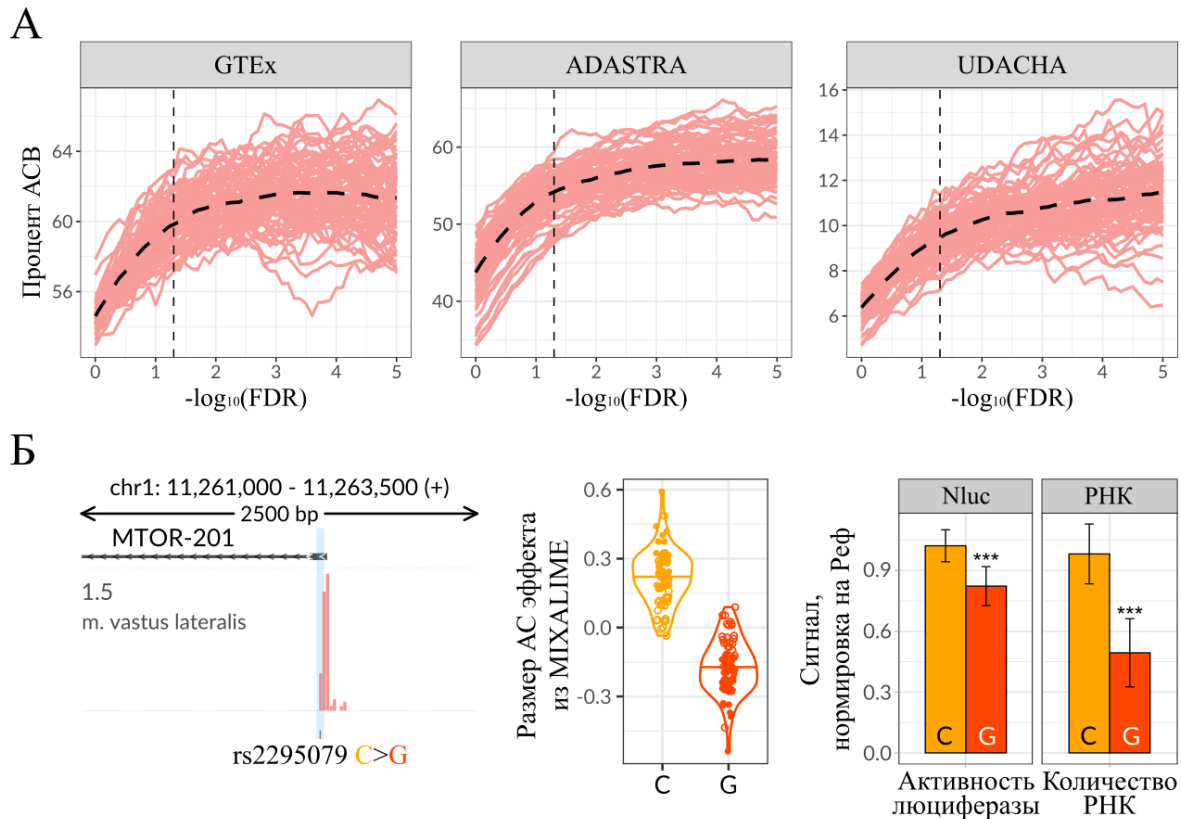


Рисунок 9 – **А.** Процент ACB из FANTOMUS, аннотированных как eQTL в скелетных мышцах согласно GTEx (слева), сайты аллель-специфичного связывания ТФ согласно ADAstra (в центре) и варианты, влияющие на доступность хроматина в мышцах согласно UDACHA (справа). Ось X: значимость аллельного дисбаланса ($-\log_{10} \text{FDR}$). Розовые линии соответствуют процентам аннотированных ACB, рассчитанным для каждой мышцы в отдельности; черным пунктиром обозначен усредненный тренд. Вертикальная линия соответствует $\text{FDR} = 0,05$. **Б.** Слева: профиль покрытия CAGE-Seq для *m. vastus lateralis* промоторной области гена *MTOR*, $\log_2$ среднего числа прочтений. Ниже отмечен локус, содержащий полиморфизм rs2295079. В центре: скрипичные диаграммы оценок аллельного дисбаланса rs2295079 в скелетных мышцах, оцененных в MIXALIME. Точки со значимым дисбалансом, $\text{FDR} < 0,05$, закрашены. Справа: результаты люциферазного репортерного анализа. Активность люциферазы и количество РНК были измерены для обеих аллелей и нормированы на референсный. Высота столбцов соответствует средним для трех реплик, полосы ошибок соответствуют стандартному отклонению. *P*-значения получены двусторонним *t*-критерием Стьюдента. ****P*-значение $< 0,001$

3.6. Идентификация вариантов генома, ассоциированных с наблюдаемыми мышечными фенотипами

Для изучения связи регуляторных генетических вариантов с заболеваниями в первую очередь был проведен статистический анализ представленности ОНВ в образцах сердца здоровых доноров и пациентов с кардиомиопатией: были выявлены полиморфизмы rs4256900 и rs3172920, значимо ($FDR < 0,05$) обогащенные у людей с заболеванием, Рисунок 10А. Эти варианты располагались в промоторах генов *UROS* и *HIGD2A*, являлись eQTL и были связаны с аллель-специфичным связыванием транскрипционных факторов EGR1 и ESR1, участвующих в регуляции ответа на ишемию [Yan и др., 2000] и кардиопротекции [Puzianowska-Kuźnicka, 2012], соответственно.

Привлечение результатов полногеномных исследований ассоциаций (GWAS) из базы данных Pan-UK Biobank позволило обнаружить обогащение сердечных ACB локусами, ассоциированными с систолическим артериальным давлением, Рисунок 10Б. Для учета эффекта сцепления вариантов был также применен метод стратифицированной регрессии по оценке неравновесия по сцеплению (S-LDSC), позволивший оценить вклад ACB скелетных мышц в наследуемость 32 мышечных фенотипов. В результате было выявлено значимое обогащение регуляторных вариантов локусами, ассоциированными с объемом мышц бедра согласно [van der Meer и др., 2022], Рисунок 10В. Таким образом, оба анализа, проведенных для ACB сердец и скелетных мышц, указывают на вклад регуляторных вариантов генома в формирование подобных полигенных признаков.

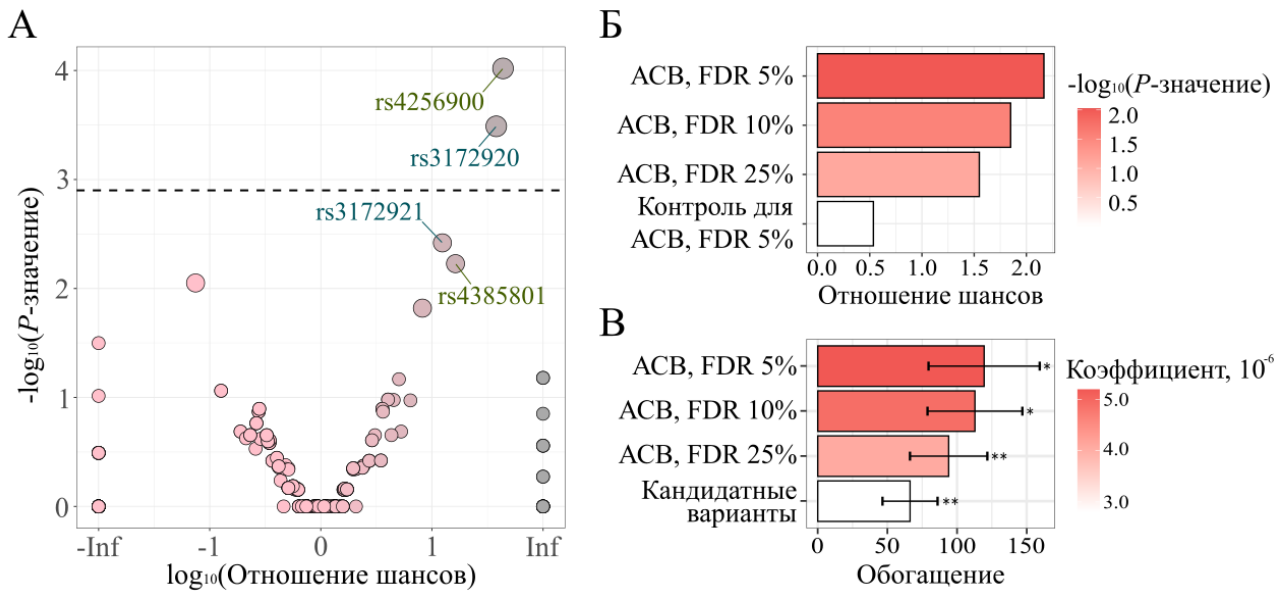


Рисунок 10 – **А**. Диаграмма рассеяния для результатов тестирования представленности вариантов в когорте людей с кардиомиопатией. Ось Y: $-\log_{10} P$ -значения двустороннего точного теста Фишера; ось X: $\log_{10}$ отношения шансов.

Пунктир соответствует отсечке FDR < 0,05. **Б**. Пересечение сердечных ACB с вариантами, значимо ассоциированными с систолическим артериальным давлением.

Ось Y: тестируемые наборы вариантов. Контрольный набор содержал незначимые варианты (FDR > 0,5) в том же количестве и примерно так же покрытых прочтениями, как и ACB FDR 5%. Ось X: отношение шансов, цвет отображает $-\log_{10} P$ -значения одностороннего точного теста Фишера. **В**. Обогащение ACB скелетных мышц локусами, ассоциированными с объемом мышц бедра, обнаруженное с использованием S-LDSC. Ось Y: тестируемые наборы вариантов, включая группу всех ОНВ, обнаруженных в образцах. Ось X: значение обогащения из S-LDSC; полосы ошибок соответствуют стандартным ошибкам. Цвет отображает коэффициент регрессии S-LDSC, звездочки соответствуют P-значениям обогащения из S-LDSC, поправленным на число тестируемых признаков: *FDR < 0,05, **FDR < 0,01

3.7. Аллельный дисбаланс как маркер редких потенциально патогенных мутаций, связанных с нарушением развития аортального клапана сердца человека

В работе [Zhigulev и др., 2026] мы использовали аллель-специфичный анализ для поиска редких регуляторных мутаций, связанных с развитием двустворчатого аортального клапана (ДАК) – врожденной патологии сердца с высокой наследуемостью, которая тем не менее плохо объясняется известными на текущий

момент ассоциированными ОНВ. Для этого были проанализированы данные HiCap по промотор-энхансерным контактам, полученные для образцов аорты 16 человек: 8 с ДАК и 8 с нормальным трехстворчатым клапаном.

Анализ выявил 2666 АСВ, отражающих различия в хроматиновых контактах промоторов генов с энхансерами между гомологичными хромосомами. Поскольку большинство таких вариантов были распространенными в популяции полиморфизмами, они рассматривались не как каузальные ОНВ, но как маркеры локусов с нарушенной регуляцией, потенциально содержащие патогенные мутации. Агрегация сигналов аллельного дисбаланса на уровне генов и образцов позволила выявить 1994 гена с ДАК-специфичными нарушениями энхансер-промоторных взаимодействий, Рисунок 11. Среди них обнаружены гены, участвующие в эмбриональном развитии сердца, включая *TBX20* [Chen и др., 2021] и *NFATC1* [Armstrong, Bischoff, 2004].

Последующий анализ, проведенный коллегами с использованием полногеномного секвенирования образцов и привлечением внешних данных по секвенированию РНК единичных клеток и пространственной транскриптомике сердца эмбриона [Lázár и др., 2025], позволил выявить редкие мутации, наблюдающиеся исключительно в образцах с или без определенных ДНК-ДНК контактов («конкордантные события»), и показал, что ключевую роль в развитии патологии играют мезенхимальные клетки аортального клапана. При этом изменения, вызванные ДАК-специфичными мутациями, чаще всего затрагивали процессы эпителиально-мезенхимального перехода и сигнального пути TGF- β . В итоге был сформирован список из 198 генов-кандидатов, экспрессирующихся в мезенхимальных клетках сердца, промоторы которых были вовлечены в конкордантные события создания или потери ДНК-ДНК контактов с энхансерами. Эти результаты позволили значительно расширить список генов, потенциально ассоциированных с развитием ДАК, подтвердить высокую генетическую гетерогенность заболевания и важность редких регуляторных мутаций в его патогенезе.

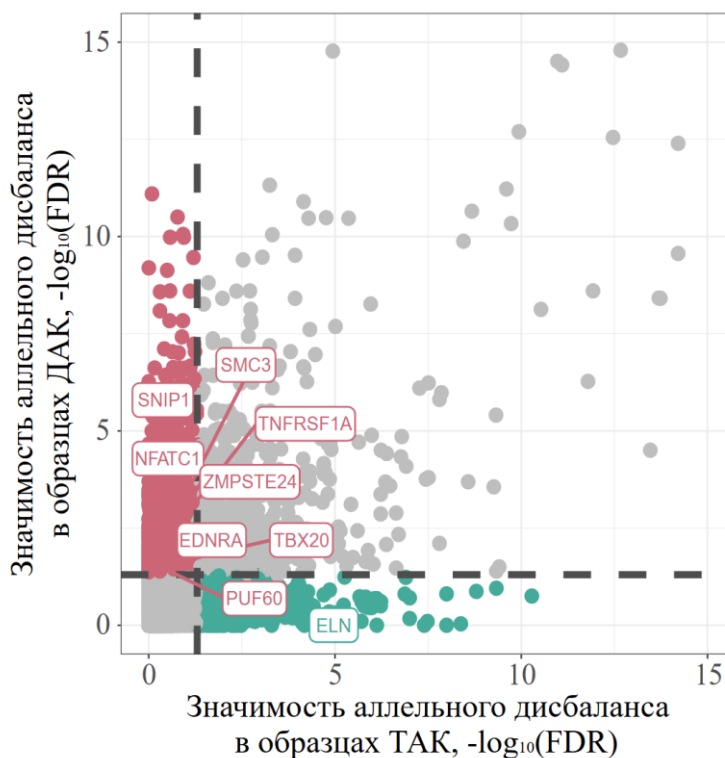


Рисунок 11 – Значимость агрегированного на уровне генов аллель-специфичного сигнала HiCap, $-\log_{10}$ FDR, для образцов с двустворчатым (ось Y) или трехстворчатым (ось X) аортальным клапаном. Пунктирные линии соответствуют порогу значимости $FDR = 0,05$; гены, ассоциированные с ДАК или участвующие в формировании и развитии аортального клапана согласно базе данных STRINGdb (v.12) [Szkarczyk и др., 2023], дополнительно подписаны

Заключение

В диссертационной работе с использованием данных CAGE-Seq был создан атлас молекулярных фенотипов 75 скелетных мышц человека FANTOMUS, продемонстрировавший высокую гетерогенность данных тканей: более 80% генов и белков демонстрируют тканеспецифичную экспрессию и представленность, соответственно. Наиболее уникальными молекулярными профилями при этом обладают экстраокулярные мышцы, язык и диафрагма, в частности, имеющие сниженную экспрессию генов, связанных с атрофией. Анализ активности ТРЭ выявил тканеспецифичное использование альтернативных промоторов, включая переключения изоформ генов *PITX2*, *ALPK2* и *TPM3*.

С использованием анализа активности мотивов удалось оценить тканеспецифичность 130 мотивов связывания ТФ, впервые охарактеризованных в рамках проекта Codebook. Применительно к скелетным мышцам, данный подход позволил обнаружить дифференциальную активность более 300 мотивов как ранее известных транскрипционных факторов, так и белков, чьи мотивы связывания были впервые описаны в Codebook.

Межиндивидуальная молекулярная гетерогенность мышц была охарактеризована с использованием данных по активности промоторов 31 сердца людей разного пола и возраста. В результате дифференциального анализа были выявлены сотни генов с полоспецифичным и возраст-зависимым использованием стартов транскрипции, включая *CREM*, кодирующий транскрипционный регулятор ответа сердца на β -адренергическую стимуляцию, и *CDH2*, кодирующий компонент адгезивных контактов кардиомиоцитов N-кадгерин.

Наконец, аллель-специфичный анализ позволил выявить около 6,6 и 2 тысяч генетических вариантов, влияющих на активность ТРЭ скелетных мышц и сердца, соответственно. Эти варианты были ассоциированы с экспрессией генов, объемом мышц и артериальным давлением. Оценка аллельного дисбаланса по данным HiCar выявила нарушения промотор-энхансерных взаимодействий при развитии ДАК и помогла определить 198 генов-кандидатов, мутации в промоторах и энхансерах которых вероятно ассоциированы с данной патологией.

Таким образом, проведенная работа подчеркивает молекулярное разнообразие скелетных мышц человека и вклад индивидуальных вариантов генома в транскриптом данных тканей и наблюдаемые мышечные фенотипы. Дальнейший анализ мышц, находящихся в определенных патологических состояниях, таких как мышечная дистрофия, атрофия в результате гипокинезии или саркопении, позволит лучше охарактеризовать молекулярные изменения, возникающие при патологии, и их генетические детерминанты.

Выводы

1. В скелетных мышцах человека обнаружена активность 37 тысяч промоторов и энхансеров, из которых более 9 тысяч ранее не были охарактеризованы. Активность промоторов слабо коррелирует с уровнем содержания соответствующих белков (коэффициент корреляции Пирсона 0,32). При этом экспрессия более 80% регуляторных элементов, генов и белков значимо ($FDR < 0,05$) различается между мышцами, а наиболее уникальными молекулярными профилями обладают экстраокулярные мышцы, диафрагма и язык. Среди генов и белков с выраженной тканеспецифичной экспрессией – регуляторы эмбриогенеза и дифференцировки мышечных клеток, включая Нох-гены, маркеры быстрых и медленных волокон, а также гены, ассоциированные с атрофией мышц и миопатиями.
2. Молекулярная гетерогенность тканей и типов клеток человека на уровне транскрипции в значительной мере объясняется дифференциальной активностью 632 ДНК-мотивов, узнаваемых различными факторами транскрипции, в том числе 130 мотивов факторов, впервые детально охарактеризованных в рамках проекта Codebook. Мотивы, активность которых различает скелетные мышцы человека, включают мотивы связывания факторов теплового шока, рецепторов гормонов щитовидной железы и регуляторов развития периокулярной мезенхимы ALX.
3. Экспрессия генов в сердце и скелетных мышцах человека зависит от индивидуальных особенностей организма, включая возраст и пол, а также от регуляторных однонуклеотидных вариантов, проявляющих аллельный дисбаланс в омиксных данных. Такие варианты часто находятся в сайтах аллель-специфичного связывания транскрипционных факторов и пересекаются с локусами количественных признаков, включая объем мышцы, систолическое артериальное давление и уровень экспрессии генов в рассматриваемых тканях, что указывает на вклад регуляторных вариантов генома в формирование мышечных фенотипов. Оценка аллельного дисбаланса может использоваться как маркер участков генома, содержащих вероятно патогенные мутации: анализ промотор-энхансерных взаимодействий в сердцах с трех- и двустворчатым аортальным клапаном выявил у последних избыток аллель-специфичных вариантов в локусах генов, вовлеченных в развитие патологии.

Список публикаций

Основные работы по теме исследования, опубликованные в изданиях, индексируемых в Web of Science, Scopus и РИНЦ:

1. Jolma A., Lavery K. U., Fathi A., Yang A. W. H., Yellan I., Vorontsov I. E., Gralak A. J., Kribelbauer-Swietek J. F., Inukai S., Razavi R., Albu M., Brechalov A., Patel Z. M., Nozdrin V., Meshcheryakov G., **Buyan A.**, et al. An expanded codebook of human transcription factor DNA-binding specificity // Nature. 2026. DOI: 10.1038/s41586-026-10798-9
2. Zhigulev A., **Buyan A.**, Lázár E., et al. Rare regulatory mutations disrupt mesenchymal molecular programs driving endocardial cushion formation in bicuspid aortic valve // Nature Communications. 2026. T. 17. № 3587. DOI: 10.1038/s41467-026-71758-5
3. Shevtsov A., **Buyan A.**, Nozdrin V., et al. NEXT-scASV: A Nextflow Pipeline for Allele-Specific Variants Calling from single cell RNA-seq data // GigaScience. 2026. giag042. DOI: 10.1093/gigascience/giag042
4. **Buyan A.**, Meshcheryakov G., Safronov V., et al. Statistical framework for calling allelic imbalance in high-throughput sequencing data // Nature Communications. 2025. T. 16. № 1. DOI: 10.1038/s41467-024-55513-2
5. Deviatiiarov R. M., Gams A., Kulakovskiy I. V., **Buyan A.**, Meshcheryakov G., Syunyaev R., et al. An atlas of transcribed human cardiac promoters and enhancers reveals an important role of regulatory elements in heart failure // Nature Cardiovascular Research. 2023. T. 2. № 1. С. 58–75. DOI: 10.1038/s44161-022-00182-x

Препринт:

1. **Buyan A.**, Gazizova G., Zgoda V. G., et al. A Joint Promoterome-Proteome Atlas Highlights the Molecular Diversity of Human Skeletal Muscles // bioRxiv. 2026. DOI: 10.64898/2026.04.08.717200

Тезисы конференций:

1. **Буян А. И.,** Газизова Г. Р., Ноздрин В. А., Мещеряков Г. А., Шагимарданова Е. И., Попов Д. В., Кулаковский И. В., Гусев О. А. Мультиомиксный атлас экспрессии и регуляции генов в скелетных мышцах человека // Сборник тезисов XII Международной научно-практической конференции «Молекулярная диагностика». 2025. сс. 528-529. ISBN: 978-5-6048946-9-9.
2. **Buyan A. I.,** Nozdrin V. A., Gazizova G. R., Meshcheryakov G. A., Hayashizaki Y., Shagimardanova E. I., Gusev O. A., Kulakovskiy I. V. Tissue-specific and allele-specific activity of transcribed gene regulatory elements in human skeletal muscle // Сборник тезисов 14-й Международной мультikonференции «Биоинформатика регуляции и структуры геномов / системная биология». 2024. сс. 120-121. DOI: 10.18699/bgrs2024-1.2-10. ISBN: 978-5-91291-067-8.
3. **Буян А.,** Мещеряков Г., Хаяшизаки Й., Кулаковский И.. Аллель-специфичная активность транскрибируемых регуляторных элементов генома человека в здоровых сердцах и больных кардиомиопатией // Сборник тезисов 11-й Московской конференции по вычислительной молекулярной биологии (MCCMB). 2023. URL: <https://mccmb.belozersky.msu.ru/2023/thesis/abstracts/50.pdf>. ISBN: 978-5-901158-33-3.
4. **Буян А. И.,** Мещеряков Г., Кулаковский И. В. Аллель-специфичная транскрипция регуляторных элементов генома в сердце человека в норме и патологии // Сборник тезисов 14-й Международной школы молодых ученых «Системная биология и биоинформатика». 2023. с. 5. DOI: 10.18699/SBB-2023-05. ISBN: 978-5-91291-062-3.
5. **Буян А. И.,** Мещеряков Г., Гамс А., Девятяров Р., Сюняев Р. и др. Аллель-специфичная экспрессия транскрибируемых регуляторных элементов генома в образцах сердца человека // Сборник тезисов международной конференции «Биология и генетика мышц». 2022. сс. 4-5. URL: <http://bgmconf.r-genomics.com/abstract/>.