

# Деконволюционная оценка соотношения быстрых и медленных транскриптомных профилей и клеточного состава скелетных мышц человека по данным CAGE-seq

Ш.Р. Низамов<sup>1</sup>, А.И. Билялов<sup>2,3,4</sup>, Н.С. Филатов<sup>2,4</sup>, Р.М. Девятяров<sup>1,2,4,5,6</sup>,  
Е.И. Шагимарданова<sup>2,7</sup>, Г.Р. Газизова<sup>1,8</sup>, Ю.П. Сергеева<sup>1</sup>, С.С. Бровкин<sup>3</sup>,  
Д.Ю. Шестаков<sup>3</sup>, Н.А. Бодунова<sup>3</sup>, О.А. Гусев<sup>2,5,8,\*</sup>

- <sup>1</sup> ФГАОУ ВО «Казанский (Приволжский) федеральный университет», Институт фундаментальной медицины и биологии, 420008, г. Казань, ул. Кремлевская, д. 18, Россия
- <sup>2</sup> ООО «ЛИФТ Центр», Россия, 121205, г. Москва, вн.тер.г. муниципальный округ Можайский, тер. Инновационного центра Сколково, ул. Нобеля, д. 5
- <sup>3</sup> ГБУЗ «Московский клинический научно-практический центр им. А.С. Логинова» Департамента здравоохранения г. Москвы, Россия, 111123, г. Москва, Новогиреевская ул., д. 1, корп. 1,
- <sup>4</sup> Уфимский федеральный исследовательский центр Российской академии наук, Россия, 450054, г. Уфа, проспект Октября, д. 71
- <sup>5</sup> Центр исследований неизлечимых заболеваний, Высшая медицинская школа, Университет Джунтендо, 113-8421, Токио, Япония
- <sup>6</sup> ФГБУ «Национальный медицинский исследовательский центр эндокринологии им. академика И.И. Дедова» Министерства здравоохранения Российской Федерации, Институт персонализированной медицины, Россия, 117292, г. Москва, ул. Дмитрия Ульянова, д. 11
- <sup>7</sup> Центр коллективного пользования в области геномики и биовизуализации, Сколковский институт науки и технологий, Россия, 143026, г. Москва, Западный административный округ, Можайский район, Инновационный центр Сколково, Большой бул., д. 30, стр. 1,
- <sup>8</sup> ФГБУ «Федеральный центр мозга и нейротехнологий» ФМБА России, Россия, 117513, Москва, ул. Островитянова, д. 1, стр. 10

\* Адрес для корреспонденции: [o.gusev.fo@juntendo.ac.jp](mailto:o.gusev.fo@juntendo.ac.jp)

## Аннотация

Представлен вычислительный метод для оценки клеточного состава образцов скелетной мышцы человека по данным CAGE-seq с использованием деконволюции. Подход позволяет оценить относительный вклад профилей экспрессии, ассоциированных с быстрым и медленным фенотипами, без гистологической классификации. Входными данными служили транскриптомные профили CAGE, позволяющие точно определить уровень экспрессии и точку инициации транскрипции в промоторе. Для определения пропорций клеточных компонентов применяли метод деконволюции MuSiC с использованием данных секвенирования отдельных ядер мышц из проекта Heart Cell Atlas. На основе полученных оценок определен порог бинарного разделения по доле быстрых мышечных волокон (10%), показавший устойчивые характеристики (AUC = 0,934 и 0,828 для двух схем аннотации). Дальнейший анализ показал, что состав волокон и связанные с ним профили экспрессии отличаются

между различными анатомическими группами мышц. Эти различия легли в основу функциональной аннотации, выявившей обогащение по биологическим процессам, связанным с развитием, специализацией мышечной ткани и возможными ассоциациями с патологией. Метод обеспечивает количественную, автоматизированную и воспроизводимую оценку спектра скоростных фенотипов скелетных мышц, открывая путь к стандартизации транскриптомного профилирования в фундаментальных и прикладных задачах.

**Ключевые слова:** CAGE-seq, деконволюция MuSiC, транскриптомные профили, ROC-анализ

**Конфликт интересов:** авторы заявляют об отсутствии конфликта интересов.

**Финансирование:** выполнено при поддержке национального проекта «Новые технологии сохранения здоровья», грант № 388-00081-25-03 от 30.04.2025.

**Для цитирования:** Низамов Ш.Р., Билялов А.И., Филатов Н.С., Девятияров Р.М., Шагимарданова Е.И., Газизова Г.Р., Сергеева Ю.П., Бровкин С.С., Шестаков Д.Ю., Бодунова Н.А., Гусев О.А. Деконволюционная оценка соотношения быстрых и медленных транскриптомных профилей и клеточного состава скелетных мышц человека по данным CAGE-seq. *Регенерация органов и тканей*. 2025;3(3). <https://doi.org/10.60043/2949-5938-2025-3-82>

Поступила: 25.06.2025

Обработана: 10.08.2025

Online first: 14.09.2025

## Deconvolution-based assessment of fast and slow transcriptomic components and cellular composition of human skeletal muscles using CAGE-seq data

Shamil R. Nizamov<sup>1</sup>, Airat I. Bilyalov<sup>2,3,4</sup>, Nikita S. Filatov<sup>2,4</sup>,  
Ruslan M. Deviatiiarov<sup>1,2,4,5,6</sup>, Elena I. Shagimardanova<sup>2,7</sup>, Guzel R. Gazizova<sup>1,8</sup>,  
Iuliia P. Sergeeva<sup>1</sup>, Sergey S. Brovkin<sup>3</sup>, Dmitriy Yu. Shestakov<sup>3</sup>, Natalya A. Bodunova<sup>3</sup>,  
Oleg A. Gusev<sup>2,5,8,\*</sup>

<sup>1</sup> Institute of Fundamental Medicine and Biology, Kazan Federal University, 18 Kremlevskaya St., Kazan 420008, Russia

<sup>2</sup> Life Improvement by Future Technologies (LIFT) Center, Skolkovo Innovation Center, 5, Nobel Str., Mozhaysky Municipal District of the Skolkovo Innovation Center, Moscow 121205, Russia

<sup>3</sup> Moscow Clinical Scientific Center Named After Loginov, 1 Bldg., 1 Novogireevskaya St., Moscow 111123, Russia

<sup>4</sup> Ufa Federal Research Centre of Russian Academy of Sciences, Institute of Biochemistry and Genetics, Oktyabrya Ave., Ufa 450054, Russia

<sup>5</sup> Intractable Disease Research Center, Graduate School of Medicine, Juntendo University, 113-8421, Tokyo, Japan

<sup>6</sup> Endocrinology Research Centre, Institute of Personalized Medicine 11 Dmitry Ulyanov St., Moscow 117292, Russia

<sup>7</sup> Genomics and Bioimaging Core Facility, Skolkovo Institute of Science and Technology, 30 Bolshoy Blvd., Bldg. 1, Skolkovo Innovation Center, Mozhaysky District, Western Administrative Okrug, Moscow 143026, Russia

<sup>8</sup> Federal Center of Brain Research and Neurotechnology of the Federal Medical Biological Agency (FMBA) of Russia, 1 Bldg., 10 Ostrovityanova St., Moscow 117513, Russia

\* Correspondence: [o.gusev.fo@juntendo.ac.jp](mailto:o.gusev.fo@juntendo.ac.jp)

**Abstract**

The methodology for classifying human muscle fibers by contraction speed is presented, based solely on transcriptomic data and without the use of classical morphological methods. The input data consisted of CAGE transcriptomic profiles, which allow precise determination of expression levels and transcription start sites in the promoter. To estimate the proportions of cellular components, the MuSiC deconvolution method was applied, using single-nucleus muscle sequencing data from the Heart Cell Atlas project. Based on the obtained estimates a binary threshold for the proportion of fast muscle fibers (10 percent) was defined, demonstrating stable characteristics (AUC = 0.934 and 0.828 for two annotation schemes).

Further analysis showed that fiber composition and associated expression profiles differ across anatomical muscle groups. These differences formed the basis for functional annotation, which revealed enrichment for biological processes related to development, specialization of muscle tissue, and possible associations with pathology.

The method provides a quantitative, automated, and reproducible assessment of the spectrum of skeletal muscle speed phenotypes, opening the way to standardizing transcriptomic profiling in fundamental and applied research.

**Keywords:** CAGE-seq, MuSiC deconvolution, transcriptomic profiles, ROC analysis

**Conflicts of Interest:** The authors declare no conflicts of interest.

**Funding:** This research was supported by the National project “New health-saving technologies” funded by the Government of the Russian Federation under Grant No. 388-00081-25-03 30.04.2025.

**For citation:** Nizamov Sh.R., Bilyalov A.I., Filatov N.S., Deviatiiarov R.M., Shagimardanova E.I., Gazizova G.R., Sergeeva Iu.P., Brovkin S.S., Shestakov D.Yu., Bodunova N.A., Gusev O.A. Deconvolution-based assessment of fast and slow transcriptomic components and cellular composition of human skeletal muscles using CAGE-seq data. *Tissue and organ regeneration*. 2025;3(3). <https://doi.org/10.60043/2949-5938-2025-3-82>

Received: 25.06.2025

Revised: 10.08.2025

Online first: 14.09.2025

**Список сокращений:**

**кДНК** — комплементарная ДНК

**AUC** — Area Under the Curve — площадь под кривой

**CAGE** — Cap Analysis of Gene Expression — анализ экспрессии на основе 5'-кэпа

**DOSE** — Disease Ontology Semantic and Enrichment — пакет для семантического анализа и обогащения по онтологии заболеваний

**FDR** — False Discovery Rate — доля ложных обнаружений

**GENCODE** — собственное имя проекта — референсная аннотация генов человека и мыши

**GRCh38** — Genome Reference Consortium human build 38 — сборка 38 генома человека консорциума GRC

**BWA** — **Burrows-Wheeler Aligner** — Выравниватель Берроуза — Уилера

**HISAT2** — Hierarchical Indexing for Spliced Alignment of Transcripts 2 — иерархический индекс для сплайсированного выравнивания транскриптов

**MuSiC** — Multi-Subject Single Cell deconvolution — деконволюция одноклеточных данных с учетом нескольких доноров

**pROC** — package for ROC analysis — пакет R для ROC-анализа

**RNA-seq** — **RNA sequencing** — секвенирование РНК

**ROC** — **Receiver Operating Characteristic** — кривая чувствительности — специфичности

**snRNA-seq** — **single-nucleus RNA sequencing** — секвенирование РНК на уровне отдельных ядер

**TPM** — **transcripts per million** — транскриптов на миллион

## 1. Введение

Скелетные мышцы человека неоднородны по структуре и функциям. Классическое деление на медленные (окислительные) и быстрые (гликолитические) волокна отражает лишь часть разнообразия мышечных фенотипов, которое формируется не только за счет сократительных белков, но и за счет различий в метаболической и регуляторной активности. Соотношение быстрых и медленных волокон, а также локальные особенности регуляции экспрессии определяют устойчивые различия между разными мышцами [1].

Гетерогенность скелетных мышц приобретает высокую клиническую значимость при изучении нейродегенеративных заболеваний и мышечных дистрофий, демонстрирующих неравномерное поражение групп мышечной ткани [2, 3]. Одним из наиболее интересных в этом контексте примеров является феномен сохранности глазодвигательных мышц, что свидетельствует о наличии мышечно-специфических механизмов устойчивости [4, 5].

Современная транскриптомика позволила детально описать клеточную архитектуру скелетной мышцы и выявить транскрипционное разнообразие ядер в мышечных волокнах. Профилирование по данным sc/snRNA-seq демонстрирует не только вариативность клеточного состава скелетной мышцы, но и наличие специализированных транскрипционных доменов внутри многоядерных миоцитов. В частности, в работе на мышцах выявлена внутримышечная гетерогенность ядерных популяций в пределах одной мышцы — с идентификацией специфических ядер в зонах нервно-мышечного и сухожильного соединений [6]. Анализ включал мышцы *tibialis anterior* и *soleus*, тем самым охватывая не одну, а несколько мышц. В случае человека построен эталонный scRNA-seq атлас на материалах десяти доноров с охватом разных анатомических локализаций, внутри каждого образца определялись пропорции резидентных клеточных типов, а сами образцы не были ограничены одной конкретной мышцей [7]. Последняя мультиомная работа (2024) использовала биопсии мышц нижней конечности у 31 участника, сочетая snRNA-seq, scRNA-seq и snATAC-seq, и позволила количественно охарактеризовать внутривнутрипопуляционную гетерогенность клеток и специализированные миоядра [8]. При этом работа с разнотипными коллекциями образцов требует строгого контроля биологических и тех-

нических вариаций, которые могут затруднять определение биологических особенностей, специфичных для мышц. К числу таких факторов относятся возраст, пол, уровень физической активности, анатомическая локализация мышцы и пространственная неоднородность волокон внутри одного образца [1, 9–11]. Прямая оценка долей быстрых и медленных волокон по гистологическим срезам также является сложной задачей, результаты которой зависят в том числе от точки отбора, тогда как повторный забор ткани повышает инвазивность и стоимость исследований. Даже при применении классических подходов на основе АТФазной гистохимии и иммуномаркеров измерения чувствительны к протоколам и объему выборки, что ведет к систематическим расхождениям и высокой дисперсии оценок [10, 12, 13]. Сопоставление данных из разных коллекций и лабораторий дополнительно осложняется техническими сдвигами и артефактами, требующими нормализации и статистической коррекции при объединении матриц экспрессии и сравнении анатомических групп мышц [14].

В этих условиях оправдан вычислительный подход, который извлекает информацию о клеточном составе из транскриптомных профилей и тем самым позволяет получить операциональные оценки относительного вклада транскриптомных профилей быстрых и медленных мышечных волокон в суммарных транскриптомных профилях ткани bulk-профилей без выполнения морфологической типизации. Определение пропорций клеточных компонентов основано на референсных одноядерных и одноклеточных данных и демонстрирует хорошую переносимость между наборами данных и донорами [15]. Сопоставимые количественные оценки долей быстрых и медленных волокон позволяют упорядочивать мышцы по их клеточному составу и характеристикам скорости сокращения. Это снижает влияние межиндивидуальных и технических вариаций, а также создает основу для выбора порога бинарного разделения, отражающего непрерывный спектр скоростных фенотипов с учетом анатомических и функциональных различий между мышцами [1, 15].

В этих условиях оправдан вычислительный подход, позволяющий по данным bulk-транскриптомики оценивать относительный вклад референсных профилей, соответствующих оцененным долям быстрых и медленных типов

мышечных волокон, а также сопутствующих клеточных популяций. Получаемые оценки описывают состав исследуемого образца и предназначены для стратификации образцов по доле транскрипционного вклада, а не для вывода о физиологической скорости сокращения. Отсутствие морфологической верификации в настоящей работе рассматривается как ограничение и учитывается при интерпретации результатов [10, 13]. На этой основе можно сформировать пороговое правило классификации образцов на быстросокращающиеся и медленносокращающиеся группы. Оно будет опираться на оцененную долю быстрых мышечных волокон и обеспечивать объективную оценку скорости мышц с учетом их анатомической и функциональной неоднородности [15]. Такой подход позволяет оценивать доли быстрых и медленных волокон на основе транскриптомных данных, без необходимости дополнительного инвазивного гистологического анализа. Это особенно важно при выраженной пространственной неоднородности мышечной ткани, когда состав существенно зависит от зоны отбора образца и требуется стандартизованная количественная оценка, дополняющая гистологические методы [10, 13].

Таким образом, целью настоящей работы являлись разработка и апробация вычислительного метода, позволяющего по данным CAGE-seq оценивать клеточный состав и относительный вклад референсных сигнатур, ассоциированных с быстрым фенотипом в образцах скелетной мышцы человека, а также использовать полученные оценки для стратификации образцов.

## 2. Материалы и методы

### 2.1. CAGE-seq

В дополнение к анализу атласных и референсных данных в исследование включены образцы скелетной мышцы, полученные от 4 доноров. В выборку вошли 3 мужчины (56, 62 и 67 лет) и 1 женщина (63 года). Согласно доступным клиническим сведениям у 3 доноров в анамнезе отмечено острое нарушение мозгового кровообращения, у 1 донора — инфаркт миокарда. Эти сведения приведены исключительно для характеристики доноров, при интерпретации результатов учитывается, что возраст и сопутствующие заболевания могут вносить вклад в межиндивидуальные различия транскриптомных профилей. Из 225 образцов мышц была выделена тотальная РНК с помощью набора

реагентов RNeasy Fibrous Tissue Kit (Qiagen, Германия). Ткань в объеме 19–30 мг гомогенизировали в фарфоровой ступке с азотом, затем добавляли лизирующий буфер и отбирали лизат в чистые пробирки. Далее образцы инкубировались с раствором протеиназы К в течение 10 минут при температуре 55 °С и центрифугировались на максимальных оборотах. При преципитации к супернатанту добавляли равный объем 96% спирта и наносили на проточную колонку. В отличие от стандартного протокола повышенное содержание спирта (~50%) обеспечило сохранение в растворе коротких фрагментов РНК (малые РНК, некодирующие РНК, энхансерные РНК). Далее образцы промывались промыточным буфером RPE два раза, а конечные этапы выделения (сушка и элюция) проходили по стандартному протоколу производителя.

Показатели чистоты (A260/A230, A260/A280) выделенной тотальной РНК измерялись на спектрофотометре NanoPhotometer (Implen, Германия), концентрация РНК измерялась как на спектрофотометре, так и на флуориметре Qubit 4.0 (Invitrogen, США). Целостность полученной РНК оценивалась с использованием чипа Pico RNA 6000 на капиллярном электрофоретическом анализаторе Bioanalyzer 2100 (Agilent Technologies, США).

Библиотеки были подготовлены по стандартному протоколу nAnT-i CAGE-seq. В среднем 5 мкг тотальной РНК каждого образца было сконцентрировано в 5 мкл воды, свободной от нуклеаз, с помощью вакуумного концентратора SpeedVac (ThermoFisher Scientific, США). Далее на матрице тотальной РНК с участием случайных праймеров и обратной транскриптазы SuperScript III Reverse Transcriptase (Invitrogen, США) была синтезирована первая цепь кДНК, которая затем была биотинилирована на 5'-конце (Dyna-beads M-270 Streptavidin, ThermoFisher Scientific, США). Это позволило с помощью стрептавидиновых частиц осуществить захват РНК, содержащую кэп на 5'-конце, и элиминировать рРНК. Для полного удаления неспецифических нитей РНК образцы были обработаны РНКазой I и H и очищены с помощью магнитных частиц Agencourt AMPure XP (Beckman Coulter, США). Далее к 5'- и 3'-концам кДНК были последовательно лигированы линкеры, один из которых содержит баркод для мультиплексирования. На финальной стадии была синтезирована вторая цепь кДНК, и готовые библиотеки были

очищены с помощью Agencourt AMPure XP (Beckman Coulter, США). Концентрация готовых библиотек была измерена с помощью qPCR с использованием KAPA Library Quantification Kits Illumina (KAPA Biosystems, ЮАР). Библиотеки были секвенированы на платформах Illumina HiSeq 2500 и NovaSeq 6000 в режиме одноконцевых чтений длиной 60 и 75 п.н. соответственно.

Выравнивание выполняли на референсный геном человека GRCh38 с использованием комбинации программ bwa и HISAT2, что позволило точно определить точки инициации транскрипции [16, 17]. Для идентификации сайтов начала транскрипции проведена кластеризация ридов с расстоянием между соседними сайтами не более 20 п.н., к дальнейшему анализу относили только кластеры с уровнем экспрессии выше 10 TPM хотя бы в одном образце, обеспечивая фильтрацию низкоэкспрессируемых сигналов [18]. Исходные CAGE-данные приводили к формату OSC-таблиц [19]. Для обеспечения воспроизводимости фиксировали версии программного обеспечения, параметры пробоподготовки и секвенирования, модели оборудования и настройки. Следует подчеркнуть, что получаемые значения отражают относительный вклад транскриптомных профилей, определенных по референсу, и не являются прямым морфологическим измерением долей волокон.

В настоящей работе данные CAGE-seq использованы как источник bulk-транскриптомных профилей, а анализ выполнен на уровне генов. Специфические возможности метода CAGE-seq, связанные с точным картированием сайтов начала транскрипции и анализом промоторной активности, в рамках настоящего исследования не использовались. Это следует рассматривать как ограничение примененного подхода и как перспективное направление для дальнейших исследований.

Забор биоматериала скелетных мышц человека и его последующее использование в настоящем исследовании выполнялись в соответствии с постановлением Правительства Российской Федерации № 750 от 21.07.2012 «Об утверждении правил передачи невостробованного тела, органов и тканей умершего человека для использования в медицинских, научных и учебных целях, а также использования невостробованного тела, органов и тканей умершего

человека в указанных целях» и действующим законодательством Российской Федерации.

## 2.2. Анализ деконволюции

Для определения клеточного состава образцов на основе объединенных (bulk) данных CAGE-seq была выполнена деконволюция с использованием программы MuSiC (Multi-subject Single Cell deconvolution) и референсных данных из атласа Heart Cell Atlas версии 1 (подраздел Human Adult Skeletal Muscle), где представлен анализ 52 825 одиночных клеток и ядер скелетных мышц от пяти доноров [15, 20, 21]. Этот метод позволяет оценивать пропорции клеточных типов в суммарных профилях экспрессии, опираясь на референс, где каждый тип клетки описан характерным транскриптомным профилем. Перенос этой информации в данные целой ткани достигается регулированием вклада генов с учетом согласованности между донорами и внутри клеточных типов. Модель восстанавливает доли основных компонентов ткани и позволяет количественно оценивать соотношение медленных и быстрых мышечных клеток в каждом образце.

Референсные данные были импортированы с помощью программного пакета zellkonverter (версия 1.18.0) для R и преобразованы в объект SingleCellExperiment. Для каждой клетки были сохранены метаданные, включая идентификатор образца и клеточные типы из оригинальной классификации [15, 21]. Исходные данные CAGE-кластеров были аннотированы с использованием аннотации генома человека GENCODE версии 45. Для ассоциации кластеров CAGE с генами применялся программный пакет ChIPseeker с определением промоторных регионов в диапазоне  $\pm 3000$  п.н. относительно сайтов инициации транскрипции. Кластеры CAGE, аннотированные как промоторные, были преобразованы в матрицу экспрессии на уровне генов путем суммирования нормализованных значений экспрессии всех промоторов, ассоциированных с каждым геном. Для конвертации идентификаторов GENCODE версии 45 в символные названия генов использовалась база данных org.Hs.eg.db (версия 3.21.0). Полученная матрица экспрессии была преобразована в объект ExpressionSet с соответствующими метаданными образцов. Деконволюция клеточного состава выполнялась с помощью функции music\_prop из программного пакета MuSiC (версия 1.0.0). В качестве входных параметров

использовались: матрица экспрессии на основе объединенных данных CAGE-seq, объект SingleCellExperiment с референсными данными, а также указание переменных для типов клеток и идентификаторов образцов. Результатом деконволюции стали взвешенные оценки пропорций клеточных типов для каждого образца с суммарными транскриптомными профилями. Полученные пропорции использовали для формирования порогового правила классификации образцов с учетом возраста, пола и анатомической локализации как ковариат в многофакторных моделях.

### 2.3. ROC-анализ

ROC-анализ применяли для выбора порога, основанного на оцененной доле референсной сигнатуры, ассоциированной с быстрым фенотипом, а также для проверки согласованности результатов деконволюции с маркерными генами. Данный метод широко применяется для оценки дискриминативной способности биомедицинских классификационных моделей, позволяя охарактеризовать эффективность разделения двух биологических классов по заданному признаку — в данном случае по выраженности маркерных генов, ассоциированных с быстрыми и медленными фенотипами. В качестве предиктора использовали логарифм отношения экспрессии каждого гена, рассчитанный при разделении образцов по предполагаемому порогу доли быстрых волокон.

Бинарные метки для ROC формировали двумя независимыми способами. В первом варианте использовали список маркерных генов быстрого и медленного типа из литературных источников [22]. Во втором варианте проводили анализ дифференциальной экспрессии генов между волокнами быстрого и медленного типов из референсных данных Heart Cell Atlas [21]. Дифференциально активные и статистически значимые ( $FDR < 0,05$ ) гены из данного сравнения использовали как маркеры, характерные для соответствующего типа мышечных волокон. В обоих случаях положительным классом считались гены, ассоциированные с транскриптомным профилем быстрых мышечных волокон, а отрицательным — гены, ассоциированные с транскриптомным профилем медленных мышечных волокон. Параметры множественных сравнений быстрых волокон против медленных задавали по стандартной схеме контроля ложного открытия. Для каждого порога строили ROC-кривую

по значениям  $\log_{2}FC$  как переменной прогноза и бинарным меткам как истинному классу. Распределение числа образцов, отнесенных к быстрым и медленным на каждом пороге доли быстрых волокон, использовалось для проверки устойчивости калибровки.

ROC-кривые строились с помощью функции `roc()` из программного пакета `rROC` (версия 1.19.0.1), где в качестве истинных меток выступала бинарная принадлежность генов к маркерам, а в качестве предсказанных значений — величина  $\log_{2}FC$  между быстрыми и медленными мышечными волокнами в суммарных CAGE-данных. Площадь под кривой (AUC) вычислялась автоматически с использованием стандартного непараметрического подхода, реализованного в `rROC`. Для визуализации качества классификации каждая ROC-кривая представлялась графически с отображением соответствующего значения AUC. Все расчеты выполняли в среде R с использованием библиотеки (R) `rROC` [23].

### 2.4. Функциональный анализ

Для оценки функциональных различий, возникающих после исключения генов, связанных со скоростью сокращения мышечных волокон, был проведен анализ дифференциальной экспрессии по трем контрастам с помощью программного пакета `edgeR` [24] для R. Первый контраст сравнивал мышечные пробы головы с остальными скелетными мышцами тела (19 образцов против 206). Второй — образцы верхних конечностей у мужчин с нижними (56 образцов против 68), третий — аналогичное сравнение для женских образцов (24 образца против 31).

Функциональное обогащение определяли отдельно для каждого контраста и для каждой группы генов (с повышенной и пониженной экспрессией) с использованием базы `WikiPathways` [25] и программного пакета `clusterProfiler` (версия 4.16.0) для R. Для поиска ассоциаций с заболеваниями использовали пакет `DOSE` (версия 4.2.0) [26]. В качестве фона использовали набор всех генов, прошедших фильтрацию и допущенных к тестированию. Применяли тест Фишера с корректировкой на множественные сравнения (FDR). Для повышения биологической интерпретируемости ограничивали размеры функциональных наборов, исключая как слишком узкие, так и чрезмерно обобщенные термины. Идентификаторы генов предварительно

унифицировали в соответствии с референсными словарями.

Исходный код доступен в открытом репозитории GitHub: [https://github.com/JynxTorquilla/skeletal\\_muscle\\_classification.git](https://github.com/JynxTorquilla/skeletal_muscle_classification.git).

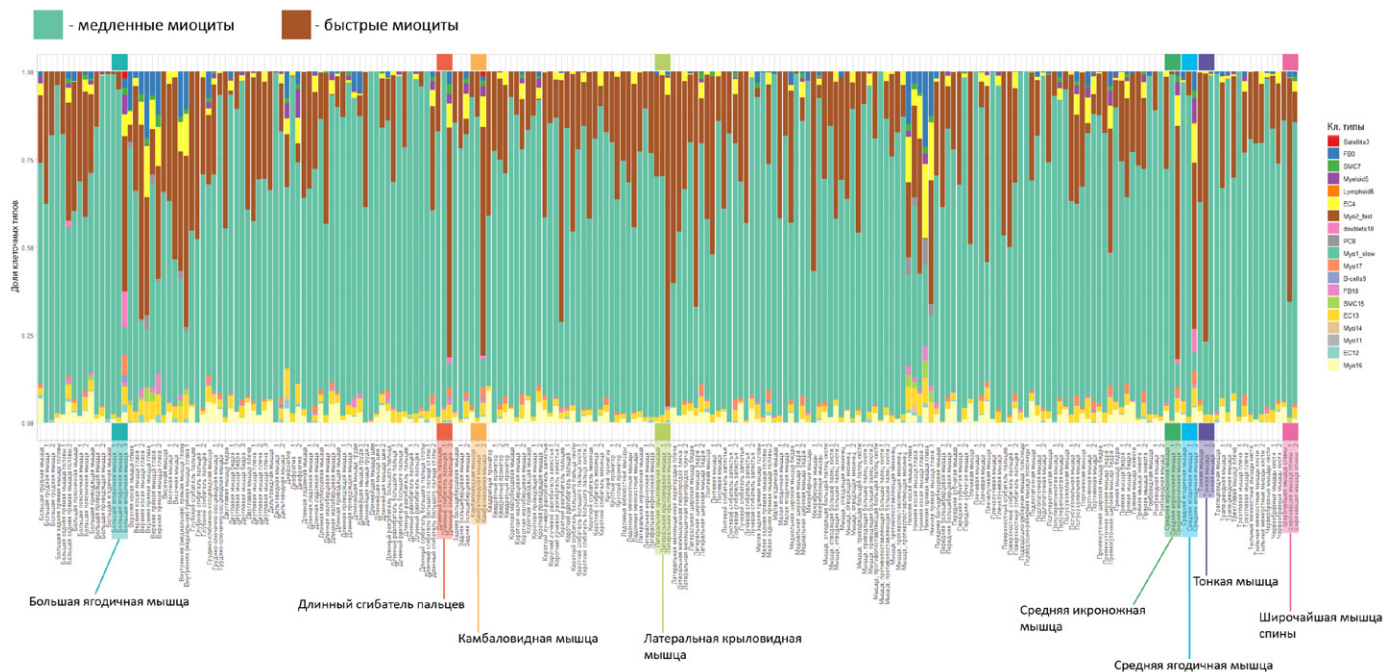
### 3. Результаты

#### 3.1. Деконволюция клеточного состава и внутримышечная вариабельность пропорций волокон

Для 225 образцов скелетных мышц человека, профилированных методом CAGE, выполнена деконволюция методом MuSiC [15] с использованием референса на основе одноядерных и одноклеточных транскриптомных данных скелетных мышц из проекта Heart Cell At-

las [21], агрегированных по типам клеток и по донорам (рис. 1). Во всех образцах доминируют миоциты, при этом доли с быстрым и медленным фенотипом демонстрируют выраженную вариабельность как между анатомическими группами, так и между биоптатами одной и той же мышцы. Наблюдаемая внутримышечная вариабельность интерпретируется как следствие отбора ткани из различных зон мышцы, в которых соотношение типов волокон может различаться.

Согласованность оценок MuSiC дополнительно проверена на 8 публично доступных образцах камбаловидной мышцы из эксперимента с сухой иммерсией. Этот набор использовали в качестве независимой валидационной выборки



**Рис. 1.** Оцененные методом MuSiC доли клеточных типов в 225 CAGE-профилированных образцах скелетной мышцы. Каждый столбец соответствует отдельному образцу, по оси X образцы сгруппированы по анатомическому наименованию мышц, группы выделены цветом под осью. Цветные сегменты отражают доли клеточных популяций, аннотированных в референсных single-cell и single-nucleus данных. Myo1\_slow – медленные миоциты, Myo2\_fast – быстрые миоциты, Myo11/Myo14/Myo16/Myo17 – подтипы миоцитов, Satellite3 – сателлитные клетки, EC4/EC12/EC13 – эндотелиальные клетки, FB0/FB18 – фибробласты, SMC7/SMC15 – гладкомышечные клетки, PC0/PC8 – перicyты, Lymphoid6 – лимфоидные клетки, Myeloid5 – миелоидные клетки, B-cells9 – B-клетки, doublets10 – техническая категория двойных профилей

**Fig. 1.** Cell type proportions estimated by the MuSiC method in 225 CAGE-profiled skeletal muscle samples. Each column represents an individual sample. Along the X-axis, samples are grouped according to anatomical muscle names, with groups color-coded beneath the axis. Colored segments indicate the proportions of cell populations annotated from reference single-cell and single-nucleus datasets. Myo1\_slow denotes slow-twitch myocytes, Myo2\_fast fast-twitch myocytes, Myo11, Myo14, Myo16, and Myo17 myocyte subtypes. Satellite3 corresponds to satellite cells. EC4, EC12, and EC13 represent endothelial cells. FB0 and FB18 denote fibroblasts. SMC7 and SMC15 correspond to smooth muscle cells. PC0 and PC8 indicate pericytes. Lymphoid6 represents lymphoid cells, Myeloid5 myeloid cells, B-cells9 B cells, and doublets10 a technical category of doublet profiles.

для проверки корректности работы конвейера и воспроизводимости оценок деконволюции в bulk-транскриптомных данных [15]. В группе иммерсии деконволюция показала отсутствие заметного вклада профиля, ассоциированного с быстрым фенотипом, тогда как в образцах сравнения оцененная доля соответствующего профиля составляла 4,7–29%. Эти результаты рассматриваются как внешняя проверка того, что метод воспроизводимо различает образцы с преобладанием медленного фенотипа и образцы с более выраженным быстрым вкладом и позволяет использовать полученные оценки для последующей калибровки порога стратификации по доле профиля, ассоциированного с быстрым фенотипом. Различия между образцами одной и той же мышцы интерпретируются как проявление тканевой гетерогенности и/или различий участка отбора, а не как измерение физиологической скорости сокращения.

### 3.2. ROC-анализ и выбор порога доли быстрых волокон

Образцы CAGE данных были распределены по процентному содержанию быстрых мышечных волокон, как показано в таблице 1.

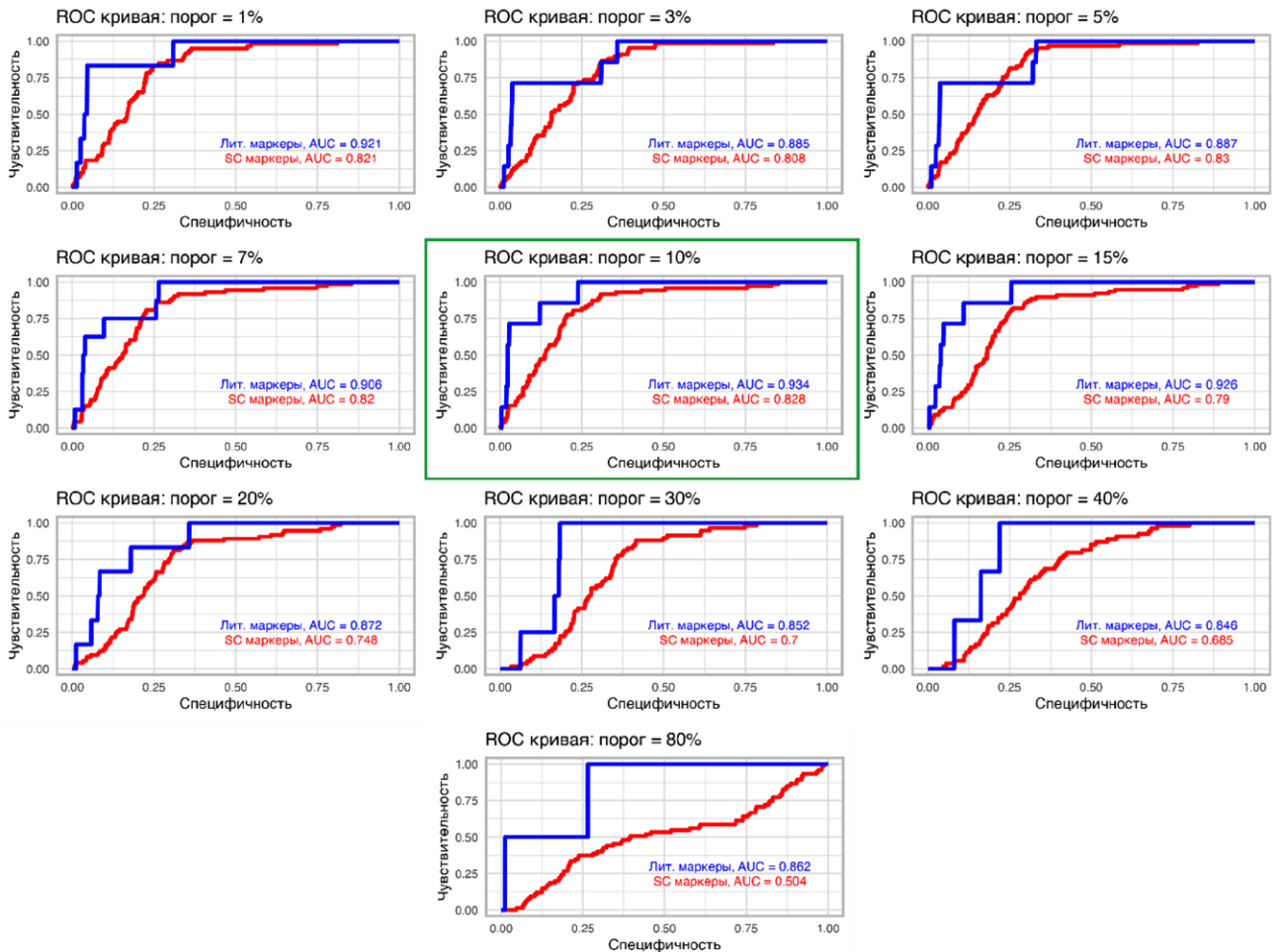
Калибровка порога по доле быстрых мышечных волокон показала максимум качества при значении 10% (рис. 2). На этом уровне получены наивысшие площади под кривой. По схеме с маркерными генами площадь равна 0,934, а с признаками из Heart Cell Atlas площадь равна 0,828. В диапазоне от 0 до 20% качество остается высоким. При 30% и выше наблюдается выраженное снижение качества. Эти результаты согласованы между двумя источниками меток и подтверждают выбор порога 10% для перевода непрерывных оценок в двухклассовое решение (рис. 2). Фиксация порога позволяет далее рассматривать межгрупповые и индивидуальные различия независимо от влияния скорости волокон как смешивающего фактора.

Далее выполнен анализ дифференциальной экспрессии с сопоставлением профилей быстрых мышц относительно медленных. Путем ассоциации CAGE-кластеров с генными моделями было определено 12 380 генов, активных в мышцах. Среди них выявлено 937 генов, ассоциированных со скоростью волокон, из которых 791 ген — белок-кодирующий. Из них: 447 генов (365 белок-кодирующих) характерны для быстрых и 490 (426 белок-кодирующих) —

**Таблица 1.** Пропорции образцов с высокой и низкой долей быстрых типов мышечных волокон, определенной по данным CAGE на основе деконволюции

**Table 1.** Proportions of samples with high and low fractions of fast muscle fiber types, as determined from CAGE data based on deconvolution analysis.

| Процентное содержание образцов с высокой долей быстрых типов мышечных волокон (CAGE данные) | Количество образцов с высокой долей быстрых типов мышечных волокон | Количество образцов с высокой долей медленных типов мышечных волокон |
|---|--|--|
| 0%  | 193  | 32   |
| 3%  | 181  | 44   |
| 5%  | 177  | 48   |
| 7%  | 168  | 57   |
| 10%   | 156  | 69   |
| 15%   | 140  | 85   |
| 20%   | 114  | 111  |
| 30%   | 81   | 144  |
| 40%   | 49   | 176  |
| 60%   | 15   | 210  |
| 70%   | 10   | 215  |
| 80%   | 6  | 219  |



**Рис. 2.** ROC-кривые для оценки качества пороговой классификации по доле быстрых мышечных волокон по различному процентному соотношению быстрых и медленных мышечных волокон. Панель с порогом в 10% отмечена зеленой рамкой и соответствует максимальным значениям площади под кривой. Синяя кривая соответствует классификации на основе литературных маркерных генов быстрых и медленных мышечных волокон, красная кривая — классификации на основе дифференциально экспрессированных генов из референсных одноклеточных данных

**Fig. 2.** ROC curves evaluating the performance of threshold-based classification according to the proportion of fast muscle fibers at different fast-to-slow fiber percentage ratios. The panel with a 10% threshold is highlighted with a green frame and corresponds to the maximum area under the curve. The blue curve represents classification based on literature-derived marker genes of fast and slow muscle fibers, whereas the red curve represents classification based on differentially expressed genes derived from reference single-cell data.

для медленных. Эти наборы могут использоваться как рабочие маркеры для биологической интерпретации и изучения вовлеченности соответствующих программ в патологию мышечной ткани.

В результате анализа дифференциальной экспрессии было выявлено:

- 7834 дифференциально активных гена для первой группы сравнения (мышцы головы против остальных);
- 3739 генов для группы сравнения образцов верхних конечностей у мужчин против нижних;

- 1716 генов для группы сравнения образцов верхних конечностей у женщин против нижних.

Выделенные гены, не ассоциированные с быстрыми или медленными мышцами и выявленные в результате межгрупповых сравнений, представляют особый интерес для дальнейших исследований. Их экспрессия может быть связана с фенотипическими особенностями образцов, статистически обогащенных по Disease Ontology для дифференциально экспрессируемых генов.

### 3.3. Функциональное обогащение и патологические ассоциации по контрастам экспрессии

После исключения генов, ассоциированных со скоростью сокращения волокон, и формирования контрастов для дифференциальной экспрессии получены значимые сигналы по базам WikiPathways и Disease Ontology. Наиболее представленные категории относятся к дифференциации предшественников сердечной мускулатуры и дифференциации плюрипотентных клеток (табл. 2).

**Таблица 2.** Основные биологические пути из базы WikiPathways, значимые для мышц головы (27 путей по FDR; 3889 дифференциально активных генов). Указаны пути, представляющие наибольший интерес. Полная таблица, включая списки использованных генов, представлена в дополнительном файле №1.

**Table 2.** Key biological pathways from the WikiPathways database that are significant for head muscles, including 27 pathways by FDR and 3,889 differentially active genes. The pathways of greatest relevance are shown. The complete table, including the lists of genes used, is provided in the Supplementary File №1.

| WikiPathways ID | Название пути (WikiPathways)   | FDR (adjusted <i>p</i> -value) |
|-----------------|--|--------------------------------|
| WP5574          | Oligodendrocyte development  | 1,44E-06                       |
| WP5417          | Cell lineage map for neuronal differentiation                                  | 2,48E-06                       |
| WP2406          | Cardiac progenitor differentiation   | 0,000329408                    |
| WP3896          | Dengue 2 interactions with complement and coagulation cascades                 | 0,000329408                    |
| WP558           | Complement and coagulation cascades  | 0,000329408                    |
| WP289           | Myometrial relaxation and contraction pathways                                 | 0,001305222                    |
| WP2328          | Allograft rejection  | 0,001764893                    |
| WP545           | Complement activation  | 0,002020305                    |
| WP5087          | Pleural mesothelioma   | 0,00246335                     |
| WP5568          | SOX genes and their influence on neurogenesis and neurodevelopmental disorders | 0,003209254                    |
| WP5090          | Complement system in neuronal development and plasticity                       | 0,004227253                    |
| WP5236          | Markers of kidney cell lineage   | 0,005203681                    |

Дальнейший анализ, включая анализ обогащения с использованием онтологии заболеваний (Disease Ontology Enrichment), помог установить потенциальные связи этих генов с определенными заболеваниями (табл. 3). Следует подчеркнуть, что указанные заболевания представлены как термины Disease Ontology (онтология заболеваний) и отражают результаты анализа обогащения. Наиболее представленные категории включают hypertension (артериальная гипертензия), cerebrovascular disease (цереброваскулярные заболевания), eye disease (заболевания глаза), connective tissue disease (заболевания соединительной ткани) и viral infectious disease (вирусные инфекционные заболевания).

**Таблица 3.** Заболевания из базы Disease Ontology для мышц головы (всего 56 значимых путей по FDR, 3889 дифференциально активных генов). Указаны пути, представляющие наибольший интерес. Полная таблица, включающая списки задействованных генов, приведена в дополнительном файле №2.

**Table 3.** Disease Ontology terms associated with head muscles, including a total of 56 significant terms by FDR and 3,889 differentially active genes. The most relevant terms are shown. The complete table, including the lists of genes involved, is provided in the Supplementary File №2.

| Disease Ontology ID | Термин Disease Ontology         | FDR (adjusted <i>p</i> -value) |
|---------------------|---------------------------------|--------------------------------|
| DOID:10763          | hypertension                    | 2,00E-08                       |
| DOID:934            | viral infectious disease        | 0,000116304                    |
| DOID:6713           | cerebrovascular disease         | 0,000208548                    |
| DOID:5614           | eye disease                     | 0,000208548                    |
| DOID:65             | connective tissue disease       | 0,000236722                    |
| DOID:6432           | pulmonary hypertension          | 0,000246053                    |
| DOID:1579           | respiratory system disease      | 0,000574866                    |
| DOID:557            | kidney disease                  | 0,001027703                    |
| DOID:77             | gastrointestinal system disease | 0,001172642                    |
| DOID:0050161        | lower respiratory tract disease | 0,001172642                    |
| DOID:14004          | thoracic aortic aneurysm        | 0,004163406                    |
| DOID:3118           | hepatobiliary disease           | 0,004368802                    |

#### 4. Обсуждение

Предложенная методология классификации скорости мышечных волокон основана исключительно на данных высокопроизводительного секвенирования и не опирается на классическую морфологическую оценку. CAGE-профилирование позволяет не только количественно оценивать транскрипционную активность, но и фиксировать точки старта транскрипции, что особенно важно для тканей с выраженной промоторной и изоформной вариативностью [27]. Для скелетной мышцы это означает возможность выявлять регуляторные сдвиги через изменение использования альтернативных промоторов. Подобные переключения могут оставаться незаметными при агрегации экспрессии на уровне гена [28]. В этой связи опора на крупные ресурсы промоторной активности и транскрибируемых энхансеров, представленные существующими атласами CAGE-промоторов и энхансеров человека, повышает сопоставимость результатов между исследованиями и снижает зависимость выводов от особенностей платформы, нормализации и последующих вычислительных процедур [28–30].

Ключевым элементом методики является деконволюция суммарных транскриптомных профилей по одноядерно-одноклеточным референсам. Это позволяет количественно восстанавливать состав клеточных компонентов, включая оценку относительного вклада транскриптомных профилей быстрых и медленных волокон в суммарном транскриптомном профиле ткани. Для деконволюции суммарных транскриптомных профилей использовался программный пакет MuSiC, учитывающий междонорную вариабельность экспрессии и тем самым обеспечивающий более точное восстановление клеточного состава при анализе тканей с родственными клеточными типами [15]. Использование транскриптомного подхода частично снижает влияние технических источников искажений, характерных для одноточечных морфологических измерений, включая зависимость результатов от толщины среза, выбора полей зрения, вариабельности окрашивания и ограниченности локальной выборки, при этом не устраняя присущую скелетной мышце биологическую пространственную неоднородность [31–33]. Наблюдаемая в литературе значительная изменчивость состава волокон вдоль одной мышцы и между биоптатами подчеркивает описываемую уязвимость одноточечных гистологических измерений и одно-

временно подтверждает ценность транскриптомных данных, усредняющих сигнал по ткани с последующей деконволюцией [10, 34]. Следует отдельно подчеркнуть, что bulk-сигнал скелетной мышцы включает вклад неммышечных компартментов.

Точечная оценка изоформ миозина тяжелой цепи (МуНС) представляет собой рациональную стратегию в специально спланированных исследованиях и широко применяется для количественного определения соотношения типов мышечных волокон, в том числе с использованием разделения МуНС-изоформ методом SDS-PAGE [35]. Кроме того, описан валидированный подход к типизации на уровне отдельных волокон, основанный на количественном электрофоретическом анализе МуНС, позволяющий напрямую оценивать вклад конкретных изоформ [36]. Вместе с тем при работе с уже существующими bulk-транскриптомными наборами данных, а также в задачах, где важно одновременно описывать транскриптомные программы ткани и вклад неммышечных клеточных популяций, секвенирование в сочетании с деконволюцией предоставляет информацию, принципиально недоступную при измерении ограниченного числа маркеров. В частности, метод MuSiC использует одноядерные или одноклеточные референсные данные для оценки долей клеточных типов в bulk-профилях, что позволяет интерпретировать межобразцовые различия с учетом тканевой композиции [15]. При этом практическая состоятельность деконволюционных подходов и чувствительность результатов к выбору референса, этапам предобработки и набору маркеров подробно обсуждены в бенчмаркинговом исследовании, где сравниваются различные конвейеры и формулируются рекомендации по повышению надежности оценок [37]. Колебания долей клеточных популяций могут систематически смещать результаты дифференциального анализа и функциональных обогашений [8, 38]. Это особенно важно при переносе оценок между наборами данных. [15]. Дополнительно важно учитывать, что даже внутри синцитиального мышечного волокна присутствуют специализированные подтипы миоядер, включая домены нервно-мышечного соединения и миотендинозного перехода. Поэтому интерпретацию долей клеточных компартментов как единственного измерения миофибриллярного фенотипа следует сопровождать оговоркой о внутриволоконной

гетерогенности [8, 39]. Практически это означает, что дальнейшее повышение точности потребует расширения референсных сигнатур за пределы упрощенной оси «быстрые против медленных» в сторону доменно-специфических и функциональных программ [39].

Выделенный нами порог в 10% по доле быстрых мышечных волокон, определенный по ROC-кривым, обеспечивает высокое качество бинарной классификации и устойчиво воспроизводится между независимыми источниками меток. Этот порог функционирует как операциональный критерий для перевода непрерывного индекса в клинически интерпретируемую двоичную оценку. При этом пороговое правило классификации по скорости волокон отражает непрерывный спектр «скоростных» характеристик, учитывая анатомическую и функциональную неоднородность мышцы. Принципы ROC-анализа и оценка точности по площади под кривой хорошо стандартизованы в современном биоинформатическом инструментарии и обеспечивают воспроизводимую калибровку пороговых решений [23].

Биологическая интерпретация результатов согласуется с концепцией непрерывности мышечных фенотипов и известной мозаичностью регуляторных схем в скелетной мускулатуре. Современные одноядерные транскриптомные карты показывают, что ядра даже внутри одного мышечного волокна могут иметь различные транскрипционные профили, соответствующие функциональным доменам. Это объясняет устойчивые локальные различия в пределах мышцы и наблюдаемую вариабельность «скоростных» характеристик [1, 6]. По сравнению с альтернативными подходами, использующими маркерные матрицы и линейную регрессию, наше решение более устойчиво и менее подвержено смещению за счет учета междонорной вариабельности и использования высокоточных данных CAGE. При этом методологически оно совместимо с существующими методиками в области деконволюции гетерогенных тканей, что расширяет возможности применения, включая сценарии с альтернативными референсами и независимую валидацию [40].

Область практического применения такой стратегии широка. В фундаментальных исследованиях она может быть использована для изучения анатомических, гендерных и возрастных различий, а также адаптационных процессов на уровне

регуляции транскрипции. В прикладной плоскости подход открывает возможность количественной стратификации мышц и мониторинга их состояния без необходимости инвазивной морфометрии. Кроме того, он может быть применим в задачах изучения избирательной уязвимости мышечных групп и разработки биомаркеров. Так, экстраокулярные мышцы, сохраняющие структуру при ряде дистрофий и в процессе старения, представляют интересную модель для изучения за счет своей уникальной молекулярной организации и регенераторного потенциала [5]. Упоминания заболеваний в настоящей работе связаны с результатами анализа онтологического обогащения с использованием Disease Ontology (онтология заболеваний) и не предполагают клинической классификации исследуемых образцов или выводов о причинно-следственных связях. Соответствующие термины применяются исключительно для биологической интерпретации наборов генов, выявленных в межгрупповых сравнениях.

При оценке переносимости предложенного порога и качества классификации необходимо учитывать неоднородность распределения типов волокон внутри одной мышцы и зависимость оценок от места и глубины биопсии. Данные по человеку показывают, что одна биопсия может не отражать фактический состав волокон на уровне всей мышцы. Вариабельность вдоль продольной оси и по глубине способна достигать величин, сопоставимых с межгрупповыми эффектами, что напрямую влияет на калибровку порога и интерпретацию моделей [34, 41]. Кроме того, в современной литературе подчеркивается риск упрощенного представления о «двух типах» волокон. Вместо этого фенотип целесообразно трактовать как непрерывный спектр состояний. Это согласуется со стратегией использования непрерывных оценок долей с последующей калибровкой порога [32, 42, 43]. В практическом плане рационально рассматривать стратификацию по анатомической локализации и стандартизацию зоны биопсии как приоритетные шаги, если метод применяется для сравнений между независимыми когортами [32].

Следует учитывать, что расчетные оценки долей профилей, ассоциированных с быстрым и медленным фенотипами, отражают состав конкретного биообразца и могут зависеть от участка отбора, глубины биопсии и локальной

гетерогенности ткани. В рамках настоящего набора образцов указанные преданалитические параметры не были стандартизированы, поэтому единичные экстремальные значения следует интерпретировать с осторожностью и не использовать как основание для обобщений о «типичном» профиле соответствующей мышцы. В настоящей работе такие наблюдения рассматриваются как частные случаи, требующие независимой верификации в специально спланированном исследовательском дизайне.

Ограничения настоящей работы в первую очередь связаны с характеристиками референсных данных секвенирования РНК отдельных ядер (snRNA-seq). В настоящей работе данные CAGE-seq использованы как источник bulk-транскриптомного профиля на уровне генов. Специфические возможности CAGE-seq, связанные с картированием сайтов начала транскрипции и промоторной активностью, в рамках анализа не использовались, поскольку последующая обработка выполнялась на геномном уровне. Поэтому выводы и предложенная процедура в целом применимы и к другим платформам bulk-транскриптомики, включая RNA-seq, при сопоставимой предобработке и выборе референса для деконволюции. Точность деконволюции зависит от качества аннотации клеточных типов, репрезентативности донорских профилей и соответствия анатомических источников исследуемым образцам. Хотя метод CAGE фокусируется только на промоторной активности и не дает информации

о структуре изоформ, именно это обеспечивает высокую точность картирования точек инициации транскрипции и делает его особенно подходящим для задач деконволюции клеточного состава скелетной мышцы. Оцененная по транскриптомным данным композиция типов волокон действительно чувствительна к таким факторам, как возраст, пол и уровень физической активности, однако в настоящем исследовании эти переменные были явно учтены как ковариаты в многофакторных моделях. При этом оценка остается стабильной вне зависимости от возраста, пола и физической активности. В дальнейшем подход может быть расширен за счет включения более разнообразных и популяционно-репрезентативных snRNA-seq референсов, интеграции с пространственной транскриптомикой и протеомикой, а также тестирования на панелях из различных анатомических групп мышц [1, 28].

## 5. Заключение

В совокупности стратегия, объединяющая промоторную транскриптомику и деконволюцию, формирует воспроизводимую, автоматизируемую и верифицируемую методологическую основу для количественной оценки скорости мышечных волокон. Она обеспечивает качественное описание межобразцовой неоднородности, остается совместимой с существующими атласами и алгоритмами и создает основу для дальнейшей стандартизации профилирования скелетной мускулатуры в исследовательских и прикладных задачах.

## Литература

1. Schiaffino S, Reggiani C. Fiber types in mammalian skeletal muscles. *Physiol Rev.* 2011;91:1447–1531. DOI: 10.1152/physrev.00031.2010
2. Nijssen J, Comley LH, Hedlund E. Motor neuron vulnerability and resistance in amyotrophic lateral sclerosis. *Acta Neuropathol.* 2017;133(6):863–885. DOI: 10.1007/s00401-017-1708-8
3. Scaricamazza S, Salvatori I, Ferri A, Valle C. Skeletal muscle in ALS: an unappreciated therapeutic opportunity? *Cells.* 2021;10(3):525. DOI: 10.3390/cells10030525
4. Khurana TS, Prendergast RA, Alameddine HS, Tome FM, Fardeau M, Arahata K, et al. Absence of extraocular muscle pathology in Duchenne’s muscular dystrophy: role for calcium homeostasis in extraocular muscle sparing. *J Exp Med.* 1995;182:467–475. DOI: 10.1084/jem.182.2.467
5. Titova A, Nikolaev S, Bilyalov A, Filatov N, Brovkin S, Shestakov D, et al. Extreme tolerance of extraocular muscles to diseases and aging: why and how? *Int J Mol Sci.* 2024;25:4985. DOI: 10.3390/ijms25094985
6. Petrany MJ, Swoboda CO, Sun C, Chetal K, Chen X, Weirauch MT, et al. Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. *Nat Commun.* 2020;11:6374. DOI: 10.1038/s41467-020-20063-w

7. De Micheli AJ, Spector JA, Elemento O, Cosgrove BD. A reference single-cell transcriptomic atlas of human skeletal muscle tissue reveals bifurcated muscle stem cell populations. *Skeletal Muscle*. 2020;10:19. DOI: 10.1186/s13395-020-00236-3
8. Lai Y, Ramírez-Pardo I, Isern J, An J, Perdiguero E, Serrano AL, et al. Multimodal cell atlas of the ageing human skeletal muscle. *Nature*. 2024;629(8010):154–164. DOI: 10.1038/s41586-024-07348-6
9. Lindholm ME, Huss M, Solnestam BW, Kjellqvist S, Lundeberg J, Sundberg CJ. The human skeletal muscle transcriptome: sex differences, alternative splicing, and tissue homogeneity assessed with RNA sequencing. *FASEB J*. 2014;28(10):4571–4581. DOI: 10.1096/fj.14-255000
10. Horwath O, Envall H, Röja J, Emanuelsson EB, Sanz G, Ekblom B, et al. Variability in vastus lateralis fiber type distribution, fiber size, and myonuclear content along and between the legs. *J Appl Physiol*. 2021;131(1):158–173. DOI: 10.1152/jappphysiol.00053.2021
11. Dahmane R, Djordjevic S, Simunic B, Valencic V. Spatial fiber type distribution in normal human muscle: histochemical and tensiomyographical evaluation. *J Biomech*. 2005;38(12):2451–2459. DOI: 10.1016/j.jbiomech.2004.10.020
12. Hintz CS, Coyle EF, Kaiser KK, Chi MM-Y, Lowry OH. Comparison of muscle fiber typing by quantitative enzyme assays and by myosin ATPase staining. *J Histochem Cytochem*. 1984;32(6):655–660. DOI: 10.1177/32.6.6202737
13. Vikne H, Strøm V, Pripp AH, Gjøvaag T. Human skeletal muscle fiber type percentage and area after reduced muscle use: a systematic review and meta-analysis. *Scand J Med Sci Sports*. 2020;30(8):1298–1317. DOI: 10.1111/sms.13675
14. Leek JT, Scharpf RB, Corrada Bravo H, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. 2010;11(10):733–739. DOI: 10.1038/nrg2825
15. Van de Castele F, Van Thienen R, Horwath O, Van der Stede T, Van Leemputte M, De Groote L, et al. Muscle fiber typing technologies: agreement between immunohistochemistry and single fiber SDS-PAGE in human skeletal muscle. *Physiol Genomics*. 2024;56(5):189–201. DOI: 10.1152/physiolgenomics.00196.2024
16. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2010;26(5):589–595. DOI: 10.1093/bioinformatics/btp698
17. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907–915. DOI: 10.1038/s41587-019-0201-4
18. Kouno T, Moody J, Kwon ATJ, Shibayama Y, Kato S, Huang Y, et al. C1 CAGE detects transcription start sites and enhancer activity at single-cell resolution. *Nat Commun*. 2019;10(1):360. DOI: 10.1038/s41467-018-08126-5
19. RIKEN. ZENBU Wiki: OSCtable (n.d.). <https://zenbu-wiki.gsc.riken.jp/zenbu/wiki/index.php/OSCtable> [Accessed October 6, 2025]
20. Heart Cell Atlas. Heart Cell Atlas v1: Cells of the adult human heart (n.d.). <https://www.heartcellatlas.org/v1.html> [Accessed October 6, 2025]
21. Litviňuková M, Talavera-López C, Maatz H, Reichart D, Worth CL, Lindberg EL, et al. Cells of the adult human heart. *Nature*. 2020;588(7838):466–472. DOI: 10.1038/s41586-020-2797-4
22. Li Y, Ma Q, Shi X, Yuan W, Liu G, Wang C. Comparative transcriptome analysis of slow-twitch and fast-twitch muscles in Dezhou donkeys. *Genes (Basel)*. 2022;13(9):1610. DOI: 10.3390/genes13091610
23. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77. DOI: 10.1186/1471-2105-12-77
24. Chen Y, Chen L, Lun ATL, Baldoni P, Smyth GK. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res*. 2025;53(2):gkaf018. DOI: 10.1093/nar/gkaf018
25. Wu T, Hu E, Xu S, Chen M, Guo P, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation*. 2021;2(3):100141. DOI: 10.1016/j.xinn.2021.100141

## ОРИГИНАЛЬНЫЕ СТАТЬИ

Низамов Ш.Р. и соавт.

Деконволюционная оценка соотношения быстрых и медленных транскриптомных профилей...

26. Wu T, Hu E, Xu S, Chen M, Guo P, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation*. 2021;2(3):100141. DOI: 10.1016/j.xinn.2021.100141
27. Forrest ARR, Kawaji H, Rehli M, Baillie JK, de Hoon MJL, Haberle V, et al. A promoter-level mammalian expression atlas. *Nature*. 2014;507(7493):462–470. DOI: 10.1038/nature13182
28. intus SS, Akberdin IR, Yevshin I, Makhnovskii P, Tyapkina O, Nigmatzyanov I, et al. Genome-wide atlas of promoter expression reveals contribution of transcribed regulatory elements to genetic control of disuse-mediated atrophy of skeletal muscle. *Biology*. 2021;10(6):557. DOI: 10.3390/biology10060557
29. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507(7493):455–461. DOI: 10.1038/nature12787
30. Noguchi S, Arakawa T, Fukuda S, Furuno M, Hasegawa A, Hori F, et al. FANTOM5 CAGE profiles of human and mouse samples. *Sci Data*. 2017;4:170112. DOI: 10.1038/sdata.2017.112
31. Baguet A, Everaert I, Hespel P, Petrovic M, Achten E, Derave W. A new method for non-invasive estimation of human muscle fiber type composition. *PLoS One*. 2011;6(7):e21956. DOI: 10.1371/journal.pone.0021956
32. Murach KA, Dungan CM, Kosmac K, Voigt TB, Tourville TW, Miller MS, et al. Fiber typing human skeletal muscle with fluorescent immunohistochemistry. *J Appl Physiol*. 2019;127(6):1632–1639. DOI: 10.1152/jappphysiol.00624.2019
33. Baguet A, Everaert I, Hespel P, Petrovic M, Achten E, Derave W. A new method for non-invasive estimation of human muscle fiber type composition. *PLoS One*. 2011;6(7):e21956. DOI: 10.1371/journal.pone.0021956
34. Van de Casteele F, Van Thienen R, Horwath O, Apró W, Van der Stede T, et al. Does one biopsy cut it? Revisiting human muscle fiber type composition variability using repeated biopsies in the vastus lateralis and gastrocnemius medialis. *J Appl Physiol*. 2024;137(5):1341–1353. DOI: 10.1152/jappphysiol.00394.2024
35. Kohn TA, Myburgh KH. Electrophoretic separation of human skeletal muscle myosin heavy chain isoforms: the importance of reducing agents. *J Physiol Sci*. 2006;56(5):355–360. DOI: 10.2170/physiolsci.RP007706
36. Tikunov BA, Sweeney HL, Rome LC. Quantitative electrophoretic analysis of myosin heavy chains in single muscle fibers. *J Appl Physiol*. 2001;90(5):1927–1935. DOI: 10.1152/jappphysiol.2001.90.5.1927
37. Avila Cobos F, Alquicira-Hernandez J, Powell JE, Mestdagh P, De Preter K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Commun*. 2020;11(1):5650. DOI: 10.1038/s41467-020-19015-1
38. Rubenstein AB, Smith GR, Raue U, Begue G, Minchev K, Ruf-Zamojski F, et al. Single-cell transcriptional profiles in human skeletal muscle. *Sci Rep*. 2020;10(1):229. DOI: 10.1038/s41598-019-57110-6
39. Dos Santos M, Backer S, Saintpierre B, Izac B, Andrieu M, Letourneur F, et al. Single-nucleus RNA-seq and FISH identify coordinated transcriptional activity in mammalian myofibers. *Nat Commun*. 2020;11(1):5102. DOI: 10.1038/s41467-020-18789-8
40. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–457. DOI: 10.1038/nmeth.3337
41. Lexell J, Henriksson-Larsén K, Sjöström M. Distribution of different fibre types in human skeletal muscles. 2. A study of cross-sections of whole m. vastus lateralis. *Acta Physiol Scand*. 1983;117(1):115–122. DOI: 10.1111/j.1748-1716.1983.tb07185.x
42. Zhang MY, Zhang WJ, Medler S. The continuum of hybrid IIX/IIB fibers in normal mouse muscles: MHC isoform proportions and spatial distribution within single fibers. *Am J Physiol Regul Integr Comp Physiol*. 2010;299(6):R1582–R1591. DOI: 10.1152/ajpregu.00402.2010
43. Blemker SS, Brooks SV, Esser KA, Saul KR. Fiber-type traps: revisiting common misconceptions about skeletal muscle fiber types with application to motor control, biomechanics, physiology, and biology. *J Appl Physiol*. 2024;136(1):109–121. DOI: 10.1152/jappphysiol.00337.2023

**Об авторах**

**Низамов Шамиль Ринатович** — аспирант, Высшая школа биологии, кафедра зоологии и общей биологии, ФГАОУ ВО «Казанский федеральный (Приволжский) университет». ORCID: <https://orcid.org/0000-0001-6944-1646>

**Билялов Айрат Ильдарович** — к.м.н., научный сотрудник, ООО «ЛИФТ Центр». ORCID: <https://orcid.org/0000-0002-8888-8395>

**Филатов Никита Сергеевич** — младший научный сотрудник, ООО «ЛИФТ Центр». ORCID: <https://orcid.org/0009-0000-9179-515X>

**Девятияров Руслан Мансурович** — биоинформатик, ФГБУ «Национальный медицинский исследовательский центр эндокринологии им. академика И.И. Дедова» ORCID: <https://orcid.org/0000-0003-0019-2076>

**Шагимарданова Елена Ильясовна** — к.б.н., старший научный сотрудник, ООО «ЛИФТ Центр». ORCID: <https://orcid.org/0000-0003-2339-261X>

**Газизова Гузель Рашитовна** — старший научный сотрудник, Научный центр «Регуляторная геномика», ФГАОУ ВО «Казанский федеральный (Приволжский) университет». ORCID: <https://orcid.org/0000-0003-4978-8495>

**Сергеева Юлия Петровна** — научный сотрудник, Научный центр «Регуляторная геномика», ФГАОУ ВО «Казанский федеральный (Приволжский) университет». ORCID: <https://orcid.org/0000-0003-0878-8528>

**Бровкин Сергей Сергеевич** — травматолог-ортопед, Отделение ортопедии и сложной травмы, ГБУЗ «Московский клинический научно-практический центр им. А.С. Логинова» Департамента здравоохранения г. Москвы. ORCID: <https://orcid.org/0009-0009-8014-2060>

**Шестаков Дмитрий Юрьевич** — к.м.н., заведующий отделением ортопедии и сложной травмы, ГБУЗ «Московский клинический научно-практический центр им. А.С. Логинова» Департамента здравоохранения г. Москвы. ORCID: <https://orcid.org/0000-0001-8365-6416>

**Бодунова Наталья Александровна** — к.м.н., руководитель Центра персонализированной медицины, ГБУЗ «Московский клинический научно-практический центр им. А.С. Логинова» Департамента здравоохранения г. Москвы. ORCID: <https://orcid.org/0009-0008-5135-4841>

**Гусев Олег Александрович** — к.б.н., заведующий лабораторией функциональной нейрогеномики, ФГБУ «Федеральный центр мозга и нейротехнологий» ФМБА России. ORCID: <https://orcid.org/0000-0002-6203-9758>

**Authors**

**Shamil R. Nizamov** — PhD student, Higher School of Biology, Department of zoology and general biology, Kazan (Volga Region) Federal University. ORCID: <https://orcid.org/0000-0001-6944-1646>

**Airat I. Bilyalov** — Cand. Sci. (Medicine), research fellow, LIFT Center LLC. ORCID: <https://orcid.org/0000-0002-8888-8395>

**Nikita S. Filatov** — junior research fellow, LIFT Center LLC. ORCID: <https://orcid.org/0009-0000-9179-515X>

**Deviatiarov Ruslan Mansurovich** — bioinformatician, Endocrinology Research Centre. ORCID: <https://orcid.org/0000-0003-0019-2076>

**Elena I. Shagimardanova** — Cand. Sci. (Biology), senior research fellow, LIFT Center LLC. ORCID: <https://orcid.org/0000-0003-2339-261X>

**Guzel R. Gazizova** — senior research fellow, Scientific center for regulatory genomics, Kazan (Volga Region) Federal University. ORCID: <https://orcid.org/0000-0003-4978-8495>

**Iuliia P. Sergeeva** — research fellow, Scientific center for regulatory genomics, Kazan (Volga Region) Federal University. ORCID: <https://orcid.org/0000-0003-0878-8528>

## ОРИГИНАЛЬНЫЕ СТАТЬИ

Низамов Ш.Р. и соавт.

Деконволюционная оценка соотношения быстрых и медленных транскриптомных профилей...

**Sergey S. Brovkin** — traumatologist-orthopedist, Department of orthopedics and complex trauma, A.S. Loginov Moscow Clinical Scientific and Practical Center, Moscow City Health Department. ORCID: <https://orcid.org/0009-0009-8014-2060>

**Dmitriy Yu. Shestakov** — Cand. Sci. (Medicine), Head of Department of orthopedics and complex trauma, A.S. Loginov Moscow Clinical Scientific and Practical Center, Moscow City Health Department. ORCID: <https://orcid.org/0000-0001-8365-6416>

**Natalya A. Bodunova** — Cand. Sci. (Medicine), Head of the Center for personalized medicine, A.S. Loginov Moscow Clinical Scientific and Practical Center, Moscow City Health Department. ORCID: <https://orcid.org/0009-0008-5135-4841>

**Oleg A. Gusev** — Head of the laboratory of functional neurogenomics, Federal Center of Brain Research and Neurotechnologies, FMBA of Russia. ORCID: <https://orcid.org/000-0002-6203-9758>

### Вклад авторов

**Ш.Р. Низамов** — концепция и дизайн работы, сбор данных, анализ и интерпретация данных, составление статьи, окончательное утверждение версии для публикации.

**А.И. Билялов** — концепция и дизайн работы, анализ и интерпретация данных, составление статьи, итоговая переработка статьи, окончательное утверждение версии для публикации.

**Н.С. Филатов** — концепция и дизайн работы, анализ и интерпретация данных, составление статьи, итоговая переработка статьи, окончательное утверждение версии для публикации.

**Р.М. Девятияров** — сбор данных, окончательное утверждение версии для публикации.

**Е.И. Шагимарданова** — анализ и интерпретация данных, итоговая переработка статьи, окончательное утверждение версии для публикации.

**Г.Р. Газизова** — сбор данных, окончательное утверждение версии для публикации.

**Ю.П. Сергеева** — сбор данных, окончательное утверждение версии для публикации.

**С.С. Бровкин** — сбор данных, окончательное утверждение версии для публикации.

**Д.Ю. Шестаков** — сбор данных, окончательное утверждение версии для публикации.

**Н.А. Бодунова** — сбор данных, окончательное утверждение версии для публикации.

**О.А. Гусев** — концепция и дизайн работы, анализ и интерпретация данных, итоговая переработка статьи, окончательное утверждение версии для публикации.

### Author contribution statement

**Shamil R. Nizamov** — concept and study design, data collection, data analysis and interpretation, drafting the manuscript, final approval of the version for publication.

**Airat I. Bilyalov** — concept and study design, data analysis and interpretation, drafting the manuscript, critical revision of the manuscript, final approval of the version for publication.

**Nikita S. Filatov** — concept and study design, data analysis and interpretation, drafting the manuscript, critical revision of the manuscript, final approval of the version for publication.

**Ruslan M. Deviatiiarov** — data collection, final approval of the version for publication.

**Elena I. Shagimardanova** — data analysis and interpretation, critical revision of the manuscript, final approval of the version for publication.

**Guzel R. Gazizova** — data collection, final approval of the version for publication.

**Iuliia P. Sergeeva** — data collection, final approval of the version for publication.

**Sergey S. Brovkin** — data collection, final approval of the version for publication.

**Dmitriy Yu. Shestakov** — data collection, final approval of the version for publication.

**Natalya A. Bodunova** — data collection, final approval of the version for publication.

**Oleg A. Gusev** — concept and study design, data analysis and interpretation, critical revision of the manuscript, final approval of the version for publication.

**Дополнительный файл №1.** Полный перечень значимых путей WikiPathways для мышц головы и списки дифференциально активных генов

**Supplementary File 1.** Complete list of significant WikiPathways for head muscles and associated differentially active gene sets

| ID     | Описание пути  | FDR         | Гены, задействованные в пути  |
|--------|--|-------------|---|
| WP5574 | Oligodendrocyte development  | 1,44E-06    | PTN/DUSP15/FGF2/TMEM98/MAG/CNTN1/SOX10/BMP4/TCF7L2/ID2/PDGFR/BMP1/EGR2/SOX9/PTPRZ1/SOX2/RELN  |
| WP5417 | Cell lineage map for neuronal differentiation  | 2,48E-06    | NEFM/NEFL/S100B/NGFR/GFAP/SOX10/BMP4/NLGN1/CD34/LIN7A/GABBR1/SHANK3/GLRB/MAP2/SYP/SNAI2/CDH1/SOX9/PCLO/DLG3/FABP7/ROBO1/SOX2/TUBB3/CD24   |
| WP2406 | Cardiac progenitor differentiation   | 0,000329408 | TNNT2/FGF2/KDR/BMP4/NKX2-5/PDGFR/KIT/BMP1/ACTC1/NCAM1/MYLK3/SOX2/CXCR4/MYH6   |
| WP3896 | Dengue 2 interactions with complement and coagulation cascades                         | 0,000329408 | CLU/F3/PROS1/CFH/C3/C1S/PLAT/CFI/C1R/SERPING1/C7/MASP1/C4B/F8/CFD   |
| WP558  | Complement and coagulation cascades  | 0,000329408 | CLU/F3/PROS1/CFH/C3/C1S/PLAT/CFI/C1R/SERPING1/C7/MASP1/C4B/F8/CFD   |
| WP289  | Myometrial relaxation and contraction pathways   | 0,001305222 | MYL4/RYR2/RAMP2/GNB5/IGFBP5/IGFBP6/GRK5/RGS4/CALD1/GNG11/ITPR1/ITPR3/GUCY1A1/ADCY4/ADCY3/GNG2/GNB4/LPAR1/PRKACB/GNG7/IGFBP2/RAMP3/CNN2/ACTC1/PLCG2/NOS3/IGFBP4/PLCD1/RGS1/SFN   |
| WP2328 | Allograft rejection  | 0,001764893 | ABCB1/CCL19/HLA-B/FAS/C3/HLA-DPA1/CXCL12/HLA-DOA/C7/VIM/HLA-DRB5/C4A/HLA-F/C4B/PDGFR/HLA-DMA/HLA-DPB1/HLA-DQA1  |
| WP545  | Complement activation  | 0,002020305 | C3/C1S/C1R/C7/MASP1/C4A/C4B/CFD   |
| WP5087 | Pleural mesothelioma   | 0,00246335  | LAMA4/CDH19/SFRP5/CCND2/CDH18/FGF2/FGF1/ITGB4/PAK3/NGFR/PDGFR/HGF/KDR/FZD10/SLC2A1/FGF7/CDH20/HMGN1/CXCL12/MDK/SPARC/NTRK2/FGF10/BCL2/AKT3/TEK/TCF7L2/FZD2/ITPR3/VEGFB/MAPK10/CDH9/PDGFR/FZD1/SFRP2/CCND1/KIT/NDRG1/AJUBA/PRSS23/COL4A4/PODXL/AMOT/KITLG/ACTC1/CDH1/RASSF2/CASP1/MMP14/VEGFD/CDH11/CCL5/PLCB4/FZD7/SRC/RASSF4/SFRP1/LAMC2/WIF1/ACTG2/SELE/ITGA2/FGF5/CCL4/SFRP4 |
| WP5568 | SOX genes and their influence on neurogenesis and neurodevelopmental disorders         | 0,003209254 | SOX15/H4C5/H4C4/H4C1/H4C6/H4C12/H4C9/SOX2/H4C11/H4C2  |
| WP5090 | Complement system in neuronal development and plasticity                               | 0,004227253 | CLU/PROS1/CFH/FAS/C3/C1S/ATP8B2/PLSCR4/CFI/C1R/SERPING1/C7/SUSD4/MASP1/C4A/C4B/CFD/PLSCR1/MARK1/FCN1/TGFB2  |
| WP5236 | Markers of kidney cell lineage   | 0,005203681 | KDR/FOXD1/FGF7/BMP4/ALDH1A2/TBX18/FAT4/NT5E/DCN/KIT/OSR1/DACT1/SFRP1  |
| WP2848 | Pluripotent stem cell differentiation pathway  | 0,006524726 | CNTF/FGF2/FGF1/HGF/BMP4/FGF10/NT5E/KIT/TPO/KITLG  |
| WP4541 | Hippo Merlin signaling dysregulation   | 0,021180714 | CDH19/CDH18/ITGB4/PAK3/NGFR/KDR/CDH20/NTRK2/TEK/ITGB8/CDH9/PDGFR/PRKACB/ITGA9/CCND1/KIT/AJUBA/AMOT/CDH1/CDH11/PLCB4/ITGA2   |
| WP5078 | T cell modulation and desmoplasia in pancreatic cancer                                 | 0,022386333 | VSIR/PDGFR/FAS/CXCL12/VEGFB/NT5E/VEGFD/ENTPD1/CXCR4/FAP/TGFB2   |
| WP4830 | GDNF RET signaling axis  | 0,027732685 | IFT25/BMP4/ALDH1A2/FAT4/FOXC1/SLIT2/ROBO1   |
| WP5564 | Human Fact complex and DNA double strand break repair                                  | 0,03822848  | H4C5/H4C4/H4C1/H4C6/TP53BP1/H4C12/H4C9/H4C11/H4C2/H1-1  |
| WP4304 | Oligodendrocyte specification and differentiation leading to myelin components for CNS | 0,038463485 | PLP1/CNTF/FGF2/MAG/SOX10/BMP4/SOX9  |
| WP4172 | PI3K Akt signaling   | 0,043614719 | LAMA4/CCND2/FGF2/FGF1/ITGB4/NGFR/GNB5/PDGFR/HGF/KDR/FGF7/NTRK2/FGF10/LPAR6/BCL2/GNG11/AKT3/TEK/ITGB8/CREB3L1/VEGFB/IL3RA/GNG2/GNB4/LPAR1/SGK2/PDGFR/ITGA9/PCK2/GNG7/CCND1/KIT/COL4A4/KITLG/VEGFD/OSMR/NOS3/RELN/LAMC2/ITGA2/FGF5/PRL  |
| WP5548 | Blood based neuroinflammation markers  | 0,049985141 | NEFL/S100B/GFAP/TAC1/CCL5/DBI/CHI3L1  |

**Дополнительный файл №2.** Полный перечень значимых терминов Disease Ontology для мышц головы и списки дифференциально активных генов

**Supplementary File 2.** Complete list of significant Disease Ontology terms for head muscles and associated differentially active gene sets

| ID         | Описание пути            | FDR         | Гены, задействованные в пути  |
|------------|--------------------------|-------------|---|
| DOID:10763 | hypertension             | 2,00E-08    | PTGDS/S100B/LEPR/CLU/VCAM1/CCND2/DIO2/FGF2/HPGD/ADD3/RYR2/RAMP2/RARRES2/F3/NGFR/GRK5/AHR/HGF/FAS/CD01/C3/KDR/MYO6/RGS4/SLC2A1/EPAS1/SMAD6/HIF3A/ACE2/GSTM5/STS/BCL2/TEK/ABAT/ITPR1/ROBO4/F11R/VIM/ITPR3/SLC1A5/SOD3/GABBR1/CTNND1/VEGFB/HYAL2/B2M/PDGFR/TLR4/NR2F2/ITGA9/APP/ENG/IGFBP2/FGFR1/ACVRL1/PEG10/PTGR1/ATP2B4/OGG1/GSN/HDAC1/TP53BP1/HLA-DQA1/S1PR1/PECAM1/PDGFRB/ACTC1/G6PD/NOX4/VWF/NQO1/LAMA2/NISCH/CYP1B1/TAC1/CDH5/HSD11B1/CCL5/NHERF1/CYP4F12/LGALS3/LIPE/NR1H3/ENTPD1/HTRA1/EGFR/GJA5/PKM/CAV1/MIF/LINC02210-CRHR1/TGFB2/HLA-A/ROBO1/PML/BAD/NHERF2/SLC9A1/APLN/CREB1/GNAI2/TGFB1/NOS3/ATP6AP2/CAVIN1/APOE/EDNRA/GRK4/KCNJ8/CD40/CXCR4/ANXA1/ERG/PTGIS/PLCD1/SRC/CYBA/ATP1A1/NPR2/TIMP4/EPHX1/NOTCH2/IL1RN/NPR1/HYAL1/ABO/EDNRB/ITPR2/PDGFA/COL6A3/MME/HLA-DQB1/ALAS1/GATA2/CCN2/PDGFB/F2R/FMO3/HIF1A/GJA1/ECE1/PTPRC/BMP2/AQP1/PDE5A/CASP9/LCN2/PAPPA/CD82/GLCE/MYBPH/GNAI3/FBN1/XDH/ARRB2/AR/FADD/FLT1/RGS5/ID1/SOX17/HSPG2/CD209/ARNT/SERPINA1/ADRA2A/SK1/SLC12A2/MMP2/NPR3/BMP1B/INVS/FADS1/RBFOX2/CHST12/ANGPT1/TIMP1/ALOX5/CHI3L1/ALAD/CX3CL1/MAPK3/GPX3/LDHA/PDIA3/SMAD9/ATP5F1A/AGTR1/MTR/PRL/TNFRSF1B/EDN1/LOX/ELN/H19/EIF2AK2/AGT/CD36/AVPR1A  |
| DOID:934   | viral infectious disease | 0,000116304 | NEFM/NEFL/MT3/S100B/ABC1/LEPR/TNNT2/CCND2/FGF2/F3/NGFR/STING1/HGF/HLA-B/CFH/FAS/C3/ALDOC/GFAP/HLA-DPA1/FGF7/IL33/IRF9/CXCL12/SELP/PLAT/ACE2/NTRK2/ATM/SERPING1/A2M/PREX2/HLA-DRB5/IFITM3/CD9/B2M/TLR4/ADAR/HLA-DPB1/HLA-DQA1/PECAM1/NRP1/TAC1/HLA-DRA/CCL5/LGALS3/STAT2/CTNNB1/EGFR/MIF/HLA-A/SOCS1/SERPINF1/STAT1/TGFB1/TAP2/MX1/APOE/PLA2G2A/CD40/MYD88/ABO/CXCL16/HLA-DQB1/PDGFB/IGF2/IL32/GJA1/BST2/MUC1/S100A8/CASP9/LCN2/ACTB/SELE/ANPEP/CD209/SERPINA1/CCL4/IFNAR2/MMP2/LDLR/ITGB2/RFC1/RSAD2/PTPN6/IRF3/HLA-DRB1/HLA-C/PLAU/CX3CL1/RAB31/CXCL9/AGTR1/PRL/TNFRSF1B/EIF2AK2/GJB6  |
| DOID:6713  | cerebrovascular disease  | 0,000208548 | PTGDS/MT3/S100B/VCAM1/SCN7A/FGF2/GRP/F3/NGFR/MAG/PROS1/HGF/HLA-B/CFH/C3/KDR/PRCP/SLC2A1/GHRL/KALRN/MDK/PLAT/ACE2/SNAP91/BCL2/PON2/SERPING1/PROCR/GPNMB/A2M/VIM/GUCY1A1/PKD2/PAWR/CDC42BPA/B2M/PDGFR/TLR4/GAP43/MGP/ITGA9/APP/ENG/TIMP2/CALCRL/MAP2/OGG1/GSN/STK26/EGR2/HLA-DQA1/PDGFRB/ACTC1/G6PD/VWF/NQO1/MMP14/PTPRZ1/THSD1/NR1H3/NCAM1/HLA-A/SERPINF1/BAD/SLC9A1/CREB1/TGFB1/NOS3/APOE/ANXA1/PTGIS/ARHGAP17/IL1RN/HNMT/CNP/ACTA2/CTSK/ACTG1/PDGFB/MEIS2/IGF2/HIF1A/ACTG2/HVCN1/CASP9/ACTB/SELE/PAPPA/GLCE/XDH/ARRB2/FADD/FLT1/SOX17/HSPG2/CASP6/SLC12A2/MMP2/CTSS/HLA-DRB1/TIMP1/MAPK3/GPX3/MAP3K5/AGTR1/MTR/PRL/EDN1/PDCD10/LUM/H19/AGT/COL1A2/UPP1   |
| DOID:5614  | eye disease              | 0,000208548 | BEX3/TSPAN12/TLCD3B/PCARE/CLU/WFS1/CNTF/VCAM1/CCND2/FGF2/TMEM98/F3/RDH11/TCF4/NGFR/ARHGAP6/IGFBP6/AHR/HGF/CYP4V2/HLA-B/CFH/AFAP1/FAS/C3/KDR/GFAP/RARB/EYA2/CLDN1/PGRMC1/BMP4/CXCL12/ACE2/NTRK2/CD34/CFI/ATM/CALD1/BCL2/SERPING1/EFEMP1/CPAMD8/GPNMB/A2M/PIK3C2A/VIM/CHN1/GUCY1A1/SOD3/PKD2/SNCB/FOXC1/ARL3/ADAMT5-AS2/B2M/LAMB2/TLR4/CAV2/RFTN1/WWOX/ENG/FBLN5/TIMP2/DCN/IGFBP2/FGFR1/HGSNAT/UNC119/IFT122/OGG1/SNCG/RAB28/HDAC1/ELOVL4/HLA-DQA1/PECAM1/PDGFRB/G6PD/MARK1/VWF/NQO1/LAMA2/CYP1B1/LCA5/OAT/CDH5/CCL5/PLCB4/CTNNB1/HTRA1/VPS13B/EGFR/PRIMPOL/MAB21L2/SLIT2/CAV1/MIF/IFT172/YAP1/HLA-A/SERPINF1/ROBO1/BAD/MYOC/SLC9A1/PDHA1/ALDH18A1/PISD/TGFB1/NOS3/LYST/GALK1/APOE/PLA2G2A/SOX2/ASNS/ANXA1/WRN/CYBA/SLC6A6/NOTCH1/IL1RN/GCNT2/SCO1/C1QTNF5/IL4R/COL6A3/MDM1/WDR19/TUBB6/RDH5/HLA-DQB1/ADAMTSL4/CCN2/LAMB1/PDGFB/EPHA2/MEIS2/CDHR1/CA4/GMDS/TACSTD2/HIF1A/ASS1/TIAM1/PTPRC/SH3PXD2B/TIMP3/ANXA3/TUBB2A/MUC1/RP2/COL18A1/AQP1/ALDH1A3/CLEC3B/CASP9/IDH3A/LCN2/AGK/LRP5/FAM161A/MERTK/FBN1/XDH/ITGA2/TAP1/CEPT1/ANXA5/FLT1/ID1/DLL4/TNFAIP6/JAK2/CCL4/MMP2/IMPDH1/MAP3K8/SELENBP1/CFB/TUBB3/CEP290/NPR3/RDH13/TUBB4A/AGBL5/VEGFC/VLDLR/ANGPT1/HLA-DRB1/TIMP1/P4HA2/DHDDS/ADAM10/CX3CL1/IFT43/EDN1/STAU2/SORD/ELN/LUM/AGT/PLTP/CD36/NMNAT1/CS |

Продолжение дополнительного файла №2

| ID        | Описание пути              | FDR         | Гены, задействованные в пути  |
|-----------|----------------------------|-------------|---|
| DOID:65   | connective tissue disease  | 0,000236722 | ALX4/ABC1/LEPR/MIA/FGF2/AZGP1/DDR2/PSMB9/NGFR/HGF/HLA-B/FAS/C3/EBF1/ABCG2/EPAS1/IL33/C1S/MESD/BMP4/EVC/SELP/PLAT/SMAD6/SPARC/DYNC2H1/C1R/BCL2/A2M/IFT52/FZD2/VIM/CREB3L1/C4B/FLNB/PLS3/F8/TNFRSF10D/IFT80/RAB8A/CIITA/FKBP10/B2M/TLR4/JAG1/PHLDB1/KIF5A/DNM3/PDLIM4/TIMP2/DCN/EXOC6B/SPRED2/FGFR1/ACVRL1/FRZB/TENT5A/COL4A4/AEBP1/BMP1/SUCO/PRDM5/HLA-DQA1/NOTCH4/PECAM1/SHOX/ANKH/XYLT2/VWF/MMP14/SOX9/RAB33B/EVC2/LIPE/IRAK3/DSE/ADGRG6/EFNB1/LFNG/TNFRSF10/CAV1/SEC24D/LMBR1/RAB13/IFT172/INPP5D/TGFBR2/SERPINF1/ERF/TCIRG1/BAD/PTK7/DYNLT2B/EFEMP2/DNMT1/TGFB1/NOS3/MBTPS1/TAP2/CAVIN1/KCNJ8/CD40/LTF/DNM1/CXCR4/ERG/NPR2/EPHX1/NOTCH2/IL1RN/FLNA/DYNC2LI1/KIF22/CTSK/NEK1/IL4R/ADAMTS5/WDR19/TNXB/DYNC2I2/HLA-DQB1/PLOD1/CCN2/PDGFB/TNFRSF10C/LTBP3/IGF2/IL32/PPIB/CRTAP/WDR35/TNFRSF10A/HIF1A/GJA1/PITX1/PTGER4/SH3PXD2B/PENK/TGDS/TGFB2/FKBP14/RAB8B/KIF3A/BGN/EXT1/ANTXR2/CLEC3B/CASP9/TRAPP2/P3H1/CCDC134/LRP5/PAPPA/ZMPSTE24/FBN1/COL5A1/AR/SEMA3C/TAP1/FLT1/IRF6/HSPG2/TCF12/PAPSS2/SERPINA1/JAK2/ADRA2A/CCL4/MMP2/LRRK2/EXT2/MTAP/NPR3/BMPR1B/HLA-DRB1/TIMP1/SFRP4/LIFR/CHI3L1/CX3CL1/IFT43/ADAMTS4/COL4A3/TNFRSF1B/EDN1/MYH6/LOX/ELN/LUM/CD36/COL1A2         |
| DOID:6432 | pulmonary hypertension     | 0,000246053 | CLU/VCAM1/FGF2/HPGD/RYP2/RAMP2/AHR/EPAS1/SMAD6/HIF3A/ACE2/BCL2/TEK/VIM/SLC1A5/CTNND1/VEGFB/HYAL2/FGFR1/ACVRL1/HDAC1/TP53BP1/PECAM1/VWF/TAC1/CDH5/CCL5/EGFR/CAV1/MIF/TGFBR2/APLN/TGFB1/NOS3/CAVIN1/EDNRA/CD40/CXCR4/ERG/PTGIS/CYBA/IL1RN/HYAL1/EDNRB/PDGFA/ALAS1/CCN2/PDGFB/HIF1A/GJA1/ECE1/BMPR2/PDE5A/LCN2/ID1/SOX17/CD209/SERPINA1/BMPR1B/ANGPT1/ALOX5/CX3CL1/MAPK3/PDIA3/SMAD9/ATP5F1A/AGTR1/EDN1/EIF2AK2/AGT  |
| DOID:1579 | respiratory system disease | 0,000574866 | S100B/FAIM2/ABC1/GRIK5/LEPR/CCL19/GRP/RAMP2/F3/PSMB9/NGFR/ERBB3/IGFBP5/HGF/HLA-B/FAS/C3/KDR/HLA-DPA1/EPAS1/IL33/ADAM33/SELP/SMAD6/ACE2/NTRK2/FUT8/BCL2/HLA-DOA/CRHR2/NKX2-5/ITPR1/A2M/HLA-E/SOD3/STAT6/PDGFR/TLR4/JAG1/ITGA9/ELK3/DCN/CPT1A/MAP2/KIT/HLA-DPB1/PRMT2/GSN/HLA-DQA1/PECAM1/VWF/NQO1/CDH1/CASP1/TAC1/GLS/HLA-DRA/NOD1/CCL5/LGALS3/IRAK3/EGFR/MIF/SCGB1A1/TGFBR2/MYLK/BAD/STAT1/FAM13A/FSTL1/TGFB1/NOS3/TAP2/EDNRA/PLA2G2A/CD40/ANXA1/CYBA/PTMA/NOTCH1/EPHX1/PRDX5/IL1RN/HNMT/ABO/EDNRB/IL4R/HLA-DQB1/CCN2/PDGFB/IL32/HIF1A/CSF2RB/BMPR2/MUC1/ORMDL1/MCAM/KIF3A/CASP9/SGMS1/PTCH1/TAP1/FLT1/GALC/CD209/SERPINA1/CCL4/MMP2/CFB/ITGB2/BMPR1B/SCNN1A/HLA-DRB1/TIMP1/ALOX5/BAMBI/PDPN/HLA-C/CHI3L1/ACTR2/PLAU/CX3CL1/DNM1L/GPX3/PDIA3/CXCL9/AGTR1/TNFRSF1B/EDN1/ELN/AGT/CD36/COL1A2  |
| DOID:557  | kidney disease             | 0,001027703 | AK4/STK39/ABC1/LEPR/CLU/VCAM1/CCND2/ADD3/AZGP1/GRP/RAMP2/EMP2/PDGFD/CFH/FAS/C3/KDR/EPAS1/GHRL/PLCE1/PLAT/APOL3/SMAD6/SPARC/APOL1/CFI/BCL2/SERPING1/C7/PROCR/TEK/A2M/VIM/CYB5A/PKD2/TBC1D8B/CYGB/NR3C2/ARL3/STAT6/EHHADH/B2M/APOL4/LAMB2/KIRREL1/TLR4/AXL/MS4A1/ENG/CPT1A/FGFR1/KIT/ALPL/TPO/COL4A4/TULP3/HLA-DQA1/WWTR1/ACTN4/G6PD/XYLT2/VWF/NQO1/GFRA1/CASP1/MMP14/NDUF88/FYN/CCL5/NHERF1/LGALS3/EGFR/INSIG2/GJA5/SLIT2/MIF/CTSH/LAMA5/IFT172/YAP1/ACTN1/SCGB1A1/TGFBR2/SERPINF1/ROBO1/TCIRG1/BAD/NHERF2/MYOC/SLC9A1/EPHB4/AHI1/TGFB1/MPV17/NOS3/ZEB2/APOE/EDNRA/KANK2/CD40/CXCR4/ANXA1/SRC/CYBA/PTMA/IL1RN/GANAB/NEK1/C1GALT1C1/CEP164/COL6A3/MME/WDR19/HLA-DQB1/TNFAIP2/CCN2/PDGFB/MGAT5/SDC1/MYH9/HIF1A/COQ2/ECE1/PTGER4/SLC4A1/B9D1/ITGA3/MUC1/YES1/KIF3A/SCLT1/AKAP12/STOM/AQP1/PDE5A/CASP9/LCN2/LRP5/GREB1L/SELE/SPRY2/CD2AP/RNF40/CFL1/RHCG/FBN1/MAGED2/XDH/FGR/RAC2/STIM1/AR/SELL/LTBP1/MAPKB1/ANXA5/CDK2/FLT1/LATS1/SH3KBP1/LYN/DLL4/SERPINA1/WAS/ITGA8/MMP2/CFB/ITGB2/CEP290/INVS/GOT2/PIAS1/SCNN1A/VLDLR/ANGPT1/HLA-DRB1/ALOX5/PDPN/PLAU/ALAD/CX3CL1/DNM1L/GPX3/HR/COL4A3/ACTR3/MAP3K5/AGTR1/COQ6/MTR/TNFRSF1B/EDN1/ELN/AGT/CD36/NUP107/AVPR1A/HCK/HPSS/KAT2A |

| ID           | Описание пути                   | FDR         | Гены, задействованные в пути   |
|--------------|---------------------------------|-------------|--|
| DOID:77      | gastrointestinal system disease | 0,001172642 | PTN/GREM2/ABCB1/LEPR/VCAM1/CCND2/FGF2/HPGD/ADD3/AZGP1/DDR2/F3/PSMB9/NGFR/ERBB3/PDGFD/ATP10D/HLA-B/CFH/FAS/KDR/HLA-DPA1/SOX10/SLC2A1/FGF7/EPB41L2/CLDN1/IL33/BMP4/CXCL12/SELP/PLAT/ACE2/NTRK2/FGF10/ACLS5/ATM/ARPC1B/CP/BCL2/CRHR2/ST6GALNAC1/TEK/A2M/F11R/FMOD/PDZD2/CYB5A/CYGB/CIITA/DPYD/B2M/TLR4/ADAR/JAG1/APP/ENG/CPT1A/RHOC/FGFR1/KIT/HLA-DPB1/GJB2/OGG1/ARRB1/HLA-DQA1/PECAM1/PDGFRB/ACTC1/G6PD/GPC3/NQO1/GFRA1/TAC1/EPB41L1/CENPC/MYO5B/CCL5/LGALS3/ENTPD1/HTRA1/EGFR/ATP8B1/EFS/MIF/TRMU/VASH1/LINC02210-CRHR1/SOCS1/SERPINF1/SPINT2/BAD/APLN/DNMT1/GNAI2/TGFB1/NOS3/MX1/ZEB2/APOE/PLA2G2A/CD40/LTF/CXCR4/ANXA1/PLVAP/NOTCH2/RELN/IL1RN/CARD8/FLNA/HNMT/CLMP/PRKAA1/SCO1/ABO/EDNRB/CNP/ACTA2/GANAB/IL4R/COL6A3/MME/WDR19/HLA-DQB1/ACTG1/CCN2/PDGFB/HMGCS2/IGF2/IL32/WDR35/PLCG1/PTGER4/NTHL1/POT1/TGFB2/AEBP2/ACTG2/COL18A1/AKAP12/PDE5A/ACTB/GDF1/LRP5/CTHRC1/SPRY2/FBN1/XDH/ACO1/ARRB2/SELL/LRP6/SETD7/FLT1/ID1/SOX17/SERPINA1/CPEB1/JAK2/MMP2/HSD3B7/EFNB2/LRRK2/VEGFC/KRT18/BMPR1B/PTGES/AZIN1/EPB41L3/ANGPT1/HLA-DRB1/NUP35/TIMP1/ALG8/PLAU/ALAD/CX3CL1/DNM1L/MAPK3/GPX3/PDIA3/SMAD9/MAP3K5/TNFRSF1B/EDN1/ELN/PDCD10/AGT/USP53/CD36/DUOX2 |
| DOID:0050161 | lower respiratory tract disease | 0,001172642 | FAIM2/ABCB1/GRIK5/LEPR/CCL19/GRP/RAMP2/F3/PSMB9/ERBB3/IGFBP5/HGF/HLA-B/FAS/C3/KDR/HLA-DPA1/EPAS1/IL33/ADAM33/SELP/SMAD6/ACE2/NTRK2/FUT8/HLA-DOA/CRHR2/NKX2-5/ITPR1/A2M/HLA-E/SOD3/STAT6/PDGFR/TLR4/JAG1/ITGA9/ELK3/DCN/CPT1A/KIT/HLA-DPB1/PRMT2/GSN/HLA-DQA1/PECAM1/VWF/NQO1/CDH1/CASP1/TAC1/GLS/HLA-DRA/NOD1/CCL5/LGALS3/IRAK3/EGFR/MIF/SCGB1A1/TGFB2/MYLK/BAD/STAT1/FAM13A/FSTL1/TGFB1/NOS3/TAP2/EDNRA/PLA2G2A/CD40/ANXA1/CYBA/PTMA/NOTCH1/EPHX1/IL1RN/HNMT/ABO/EDNRB/IL4R/HLA-DQB1/CCN2/PDGFB/IL32/HIF1A/CSF2RB/BMPR2/MUC1/ORMDL1/MCAM/KIF3A/CASP9/SGMS1/TAP1/FLT1/GALC/CD209/SERPINA1/CCL4/MMP2/ITGB2/BMPR1B/SCNN1A/HLA-DRB1/TIMP1/ALOX5/BAMBI/PDPN/HLA-C/CHI3L1/ACTR2/PLAU/CX3CL1/DNM1L/GPX3/CXCL9/AGTR1/TNFRSF1B/EDN1/ELN/AGT/CD36/COL1A2  |
| DOID:14004   | thoracic aortic aneurysm        | 0,004163406 | THSD4/TIMP2/ACTC1/MMP14/MYLK/FLNA/ACTA2/COL6A3/ACTG1/MFAP5/TIMP3/ACTG2/ACTB/MMP2/TIMP1/MYH11/ADAM10/CXCL9/LOX  |
| DOID:3118    | hepatobiliary disease           | 0,004368802 | PTN/ABCB1/LEPR/VCAM1/CCND2/FGF2/ADD3/AZGP1/DDR2/F3/PDGFD/ATP10D/CFH/FAS/KDR/HLA-DPA1/CLDN1/IL33/ACE2/CP/CYGB/DPYD/TLR4/ADAR/JAG1/ENG/CPT1A/HLA-DPB1/GJB2/OGG1/HLA-DQA1/PECAM1/PDGFRB/G6PD/GPC3/NQO1/CENPC/MYO5B/CCL5/LGALS3/ENTPD1/EGFR/ATP8B1/TRMU/VASH1/SOCS1/SERPINF1/SPINT2/DNMT1/TGFB1/NOS3/MX1/APOE/NOTCH2/RELN/PRKAA1/SCO1/ABO/EDNRB/CNP/GANAB/WDR19/HLA-DQB1/CCN2/PDGFB/HMGCS2/IGF2/WDR35/POT1/TGFB2/AKAP12/PDE5A/GDF1/LRP5/FBN1/XDH/ARRB2/SETD7/FLT1/ID1/SOX17/SERPINA1/CPEB1/JAK2/MMP2/HSD3B7/VEGFC/KRT18/AZIN1/HLA-DRB1/TIMP1/ALG8/PLAU/ALAD/CX3CL1/PDIA3/EDN1/ELN/PDCD10/AGT/USP53/CD36  |
| DOID:633     | myositis                        | 0,00458876  | MYH13/MYH4/CLU/HGF/HLA-B/FAS/BMP4/FMOD/HLA-E/HLA-F/TLR4/HLA-DQA1/SLIT2/HLA-A/ROBO1/CD40/IL1RN/HLA-DQB1/CCN2/LRP5/MYH3/HLA-DRB1/HLA-C/CX3CL1/MYH8/TNFRSF1B/MYH6/ELN/PMS1/CD36   |
| DOID:0080001 | bone disease                    | 0,004872547 | ALX4/ABCB1/LEPR/MIA/FGF2/AZGP1/DDR2/PSMB9/NGFR/HGF/HLA-B/FAS/C3/ABCG2/EPAS1/IL33/MESD/EVC/SELP/PLAT/SMAD6/SPARC/DYNC2H1/BCL2/A2M/IFT52/FZD2/VIM/CREB3L1/C4B/FLNB/PLS3/F8/TNFRSF10D/IFT80/RAB8A/CIITA/FKBP10/B2M/TLR4/JAG1/PHLDB1/KIF5A/PDLIM4/EXOC6B/SPRED2/FGFR1/FRZB/TENT5A/BMP1/SUCO/HLA-DQA1/NOTCH4/PECAM1/SHOX/ANKH/VWF/MMP14/SOX9/RAB33B/EVC2/IRAK3/ADGRG6/EFNB1/LFNG/TNFSF10/CAV1/SEC24D/LMBR1/RAB13/IFT172/INPP5D/SERPINF1/ERF/TCIRG1/BAD/PTK7/DYNLT2B/EFEMP2/DNMT1/TGFB1/NOS3/MBTPS1/TAP2/KCNJ8/CD40/CXCR4/ERG/NPR2/EPHX1/NOTCH2/IL1RN/FLNA/DYNC2L1/KIF22/CTSK/NEK1/IL4R/ADAMTS5/WDR19/DYNC2I2/HLA-DQB1/CCN2/PDGFB/TNFRSF10C/LTBP3/IGF2/IL32/PPIB/CRTAP/WDR35/TNFRSF10A/HIF1A/GJA1/PITX1/PTGER4/SH3PX2B/PENK/TGDS/TGFB2/RAB8B/KIF3A/BGN/EXT1/CLEC3B/CASP9/TRAPPC2/P3H1/CCDC134/LRP5/PAPPA/ZMPSTE24/FBN1/AR/SEMA3C/TAP1/FLT1/IRF6/HSPG2/TCF12/PAPSS2/SERPINA1/JAK2/CCL4/MMP2/LRRK2/EXT2/MTAP/NPR3/BMPR1B/HLA-DRB1/TIMP1/SFRP4/LIFR/CHI3L1/CX3CL1/IFT43/ADAMTS4/TNFRSF1B/EDN1/MYH6/ELN/CD36/COL1A2  |

Окончание дополнительного файла №2

| ID         | Описание пути          | FDR         | Гены, задействованные в пути  |
|------------|------------------------|-------------|---|
| DOID:854   | collagen disease       | 0,006644005 | FAS/C1S/C1R/TLR4/DNM3/DCN/COL4A4/AEBP1/PRDM5/PECAM1/DSE/TGFB1/DNM1/TNXB/PLOD1/FKBP14/FBN1/COL5A1/HLA-DRB1/COL4A3/LOX/LUM/COL1A2   |
| DOID:1459  | hypothyroidism         | 0,007582968 | NEFM/NEFL/MT3/DIO2/DPYSL2/TSHR/NGFR/SLC2A1/MAP1A/ARPC5/NKX2-5/IGFBP2/TPO/GSN/NRGN/G6PD/ENTPD1/CAV1/NCAM1/MBP/BAD/SLC9A1/NOS3/APOE/ANXA1/RELN/DUOX1/HLA-DQB1/PLOD1/PLOD3/MAP1B/ANXA5/HLA-DRB1/ALAD/MYH6/PCSK1/SRD5A1/COL1A2/DUOX2/CS   |
| DOID:2841  | asthma                 | 0,007582968 | ABCB1/CCL19/HLA-B/FAS/C3/KDR/HLA-DPA1/IL33/ADAM33/HLA-DOA/CRHR2/ITPR1/HLA-E/STAT6/TLR4/DCN/HLA-DPB1/PRMT2/HLA-DQA1/NQO1/CDH1/CASP1/TAC1/HLA-DRA/NOD1/CCL5/LGALS3/IRAK3/EGFR/MIF/SCGB1A1/MYLK/STAT1/TGFB1/NOS3/TAP2/CD40/CYBA/EPHX1/IL1RN/HNMT/ABO/EDNRB/IL4R/HLA-DQB1/PDGF/IL32/CSF2RB/BMPR2/MUC1/ORMDL1/TAP1/CD209/SERPINA1/CCL4/BMPR1B/HLA-DRB1/ALOX5/HLA-C/CHI3L1/PLAU/CX3CL1/AGTR1/EDN1 |
| DOID:576   | proteinuria            | 0,00861152  | LEPR/CLU/ADD3/EMP2/CFH/C3/PLCE1/CFI/A2M/VIM/TBC1D8B/STAT6/B2M/LAMB2/KIRREL1/HLA-DQA1/LAMA5/SLC9A1/TGFB1/MPV17/ZEB2/APOE/EDNRA/KANK2/CXCR4/IL1RN/HLA-DQB1/MGAT5/SDC1/MYH9/HIF1A/COQ2/ECE1/STOM/CD2AP/CFL1/ANXA5/SH3KBP1/SERPINA1/MMP2/SCNN1A/VLDLR/ANGPT1/ALOX5/PDPN/CX3CL1/GPX3/HR/AGTR1/COQ6/EDN1/CD36/NUP107/HPS5   |
| DOID:13359 | Ehlers-Danlos syndrome | 0,009455972 | C1S/C1R/DCN/AEBP1/PRDM5/DSE/TGFB1/TNXB/PLOD1/FKBP14/COL5A1/LOX/LUM/COL1A2   |