

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НОВОСИБИРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ» (НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ, НГУ)

Факультет естественных наук

Кафедра информационной биологии

Направление подготовки 06.04.01 Биология (магистр)

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Елгаевой Елизаветы Евгеньевны

Тема работы «Выявление общей генетической компоненты хронических скелетно-мышечных болей человека разной локализации»

«К защите допущена»

Заместитель заведующего кафедрой,

д. б. н., в. н. с.

Фурман Д. П./.....

«.....».....20...г.

Научный руководитель

к. б. н., с.н.с.

Лаборатория рекомбинационного и сегрегационного анализа

ИЦиГ СО РАН

Цепилов Я. А./.....

«.....».....20...г.

Новосибирск, 2020

Оглавление

Список сокращений.....	4
Введение.....	5
Обзор литературы.....	7
1. Природа боли.....	7
1.1. Классификация боли.....	9
1.2. Описание физиологической основы боли.....	12
1.2.1. Болевые рецепторы.....	12
1.2.2. Задние рога спинного мозга.....	13
1.2.3. Воротный контроль.....	14
1.2.4. Опиоидные рецепторы.....	14
1.2.5. Восходящие ноцицептивные пути.....	15
1.2.6. Нисходящая модуляция боли.....	16
1.3. Физиологическая основа хроническая боли.....	16
2. Методы анализа генетических ассоциаций.....	19
2.1. Ген-кандидатные исследования.....	20
2.2. Полногеномный анализ ассоциаций.....	20
2.2.1. Биоинформатические методы обработки результатов полногеномного анализа ассоциаций.....	23
3. Метод главных компонент и его применение в биологических исследованиях.....	25
3.1. Математический аппарат метода главных компонент.....	25
3.2. Примеры использования метода главных компонент в генетике.....	29
4. Генетика скелетно-мышечных болей человека.....	32
5. Аналитическое заключение.....	37
Материалы и методы.....	39
Материалы.....	39
Исследуемая выборка и фенотипирование.....	39
Генотипирование и импутация геномных данных.....	41
Полногеномный анализ ассоциаций.....	41
Методы.....	42
Главные компоненты матрицы генетических ковариаций.....	42
Мета-анализ.....	47
База данных GWAS-MAP.....	47

Поиск локусов	47
Условный и совместный анализ COJO	48
Анализ представленности функциональных групп генов DEPICT.....	48
Литературная аннотация генов	48
Предсказание эффектов локусов	48
Генетические корреляции и наследуемость	49
Результаты.....	50
Краткий обзор дизайна эксперимента	50
Главные компоненты матрицы генетических ковариаций	50
Полногеномный анализ ассоциаций ГКМГК.....	52
Функциональная аннотация найденных сигналов.....	55
Предсказание эффектов локусов	55
Приоритизация генов с помощью DEPICT	56
Результаты приоритизации генов с использованием разных инструментов	56
Анализ представленности функциональных групп генов	57
Генетические корреляции ГКМГК с другими сложными признаками	58
Обсуждение.....	60
Заключение.....	66
Выводы	67
Список использованной литературы	68
Приложения	75

Список сокращений

- БДР – большое депрессивное расстройство
- ГАМК – γ -аминомасляная кислота
- ГКМГК – главная компонента матрицы генетических ковариаций
- и др. – и другие
- ИМТ – индекс массы тела
- и т. д. – и так далее
- кб – килобаза
- ЛПВП – липопротеины высокой плотности
- Мб – мегабаза
- млн – миллион
- МКБ – Международная Классификация Болезней
- мкм - микрометр
- ПГАА – полногеномный анализ ассоциаций
- пн – пар нуклеотидов
- сМ – сантиморган
- СТБ – отношение обхвата талии к обхвату бедер
- ТЦА – трициклические антидепрессанты
- ЦНС – центральная нервная система
- АМРА – амино-3-гидрокси-5-метил-4-изоксазол-пропионовая кислота
- D' – коэффициент неравновесия по сцеплению (коэффициент Левонтина)
- FDR – false discovery rate
- h^2 – наследуемость признака
- HRC – Haplotype Reference Consortium
- IgG – иммуноглобулин G
- MHC (major histocompatibility complex) – главный комплекс гистосовместимости
- NMDA – N-метил-D-аспартат
- OR (odds ratio) – отношение шансов
- r^2 – коэффициент корреляции Пирсона
- r_g – генетическая корреляция
- SNP (single nucleotide polymorphism) – однонуклеотидный полиморфизм

Введение

Боль – это неприятное сенсорное переживание, вызванное повреждением тканей или угрозой его возникновения. Ее можно рассматривать как адаптивную реакцию организма, обеспечивающую защиту от воздействия повреждающих факторов. Однако с переходом боли в хроническую форму (хронической считается боль, продолжающаяся три и более месяцев подряд) она утрачивает свое адаптивное значение и приобретает черты самостоятельного заболевания.

Хроническая боль как клинический синдром отличается высокой гетерогенностью из-за разнообразия патогенетических механизмов и множества причин возникновения. Она влечет за собой структурно-анатомические изменения в организме, а также изменения в поведении и психическом состоянии человека. Хроническая боль часто встречается в популяции. В частности, скелетно-мышечные боли – самые распространенные среди хронических болей, поражают до 30% взрослого населения по всему миру [1]. В то же время, существующие терапевтические подходы к лечению хронической боли характеризуются низкой эффективностью. Это делает хроническую боль масштабной социально-экономической проблемой. По результатам исследования Глобального бремени болезней [2] на протяжении почти 30 последних лет хроническая скелетно-мышечная боль является самой распространенной причиной нетрудоспособности в мире. Решение этой проблемы требует всестороннего изучения основ хронической боли.

К настоящему моменту широкое распространение получила биопсихосоциальная модель хронической боли, описывающая хроническую боль как результат взаимодействия биологических, психических и социально-демографических факторов. Долгое время изучение биологической составляющей боли ограничивалось исследованиями в области физиологии и молекулярной биологии, однако постепенно стало накапливаться все больше доказательств генетических основ хронической боли. Текущие оценки наследуемости признака варьируют от 30 до 60% в различных источниках в зависимости от типа боли. На данный момент значительное количество работ посвящено исследованию генетики отдельных типов хронической боли, однако, число найденных и подтвержденных ассоциаций по-прежнему остается небольшим. Попытки решить эту проблему, с использованием классических подходов, путем многократного увеличения размера выборки не принесли желаемого результата. Причинами, по которым традиционные подходы имеют низкую мощность анализа при изучении хронической боли, являются гетерогенность признака, его мультифакторность и отсутствие объективных критериев

оценки его выраженности. Следовательно, для достижения лучшего результата, необходимо применение особых методов исследования, направленных на решение обозначенных проблем.

Одним из подходов, позволяющих преодолеть проблему потери мощности из-за гетерогенности признака, является изучение его в составе группы схожих с ним признаков, имеющих нечто общее. Эта концепция реализована, в частности, в методе главных компонент, широко применяемом в различных областях науки. В данной работе мы решили использовать метод главных компонент для изучения генетики хронических скелетно-мышечных болей человека разной локализации (боли в спине, шее, бедре и колене). Мы основывались на результатах работы 2014 года [3], в которой в ходе близнецовых исследований и анализа генетических корреляций между признаками было показано, что различные типы хронических скелетно-мышечных болей имеют общие наследственные факторы. Поскольку предметом нашего интереса является генетическая составляющая разных типов хронической скелетно-мышечной боли, мы приняли решение применить метод главных компонент на матрице генетических ковариаций между признаками интереса (вместо классического подхода, в котором метод главных компонент применяется на матрице фенотипических ковариаций). Применив метод главных компонент на выборке, состоящей из здоровых людей и людей, страдающих разными типами хронической скелетно-мышечной боли, мы сможем выделить факторы, по которым больные тем или иным типом хронической скелетно-мышечной боли максимально отличаются от здоровых людей.

Целью данной работы является поиск общих наследственных факторов четырех хронических скелетно-мышечных болей человека.

Для достижения данной цели нами были поставлены следующие задачи:

1. выделить главные компоненты матрицы генетических ковариаций (ГКМГК) хронических скелетно-мышечных болей человека;
2. выделить локусы, ассоциированные с ГКМГК, и провести их репликацию;
3. провести функциональный анализ полученных результатов анализа ассоциаций ГКМГК *in silico*:
 - а) рассчитать генетические корреляции ГКМГК с другими заболеваниями и признаками человека;
 - б) провести анализ представленности функциональных групп генов, ассоциированных с ГКМГК, в тканях и биохимических путях.

Обзор литературы

1. Природа боли

Согласно определению, предложенному Международной ассоциацией по изучению боли (The International Association for the Study of Pain, ISAP), боль – это неприятное ощущение и эмоциональное переживание, связанное с текущим или потенциальным повреждением тканей или описываемое в терминах такого [4].

Выделяют пять основных компонентов боли [4]:

- перцептуальный – позволяет определить локализацию очага боли;
- эмоционально-аффективный – создает неприятное психоэмоциональное переживание;
- вегетативный – отражает рефлекторные изменения в работе симпато-адреналовой системы и внутренних органов;
- двигательный – характеризует активность, направленную на устранение повреждающего воздействия;
- когнитивный компонент – обеспечивает формирование субъективного отношения к переживаемой в настоящий момент боли на основе предыдущего опыта.

Важно различать острую и хроническую боль. Острая боль (боль продолжительностью до трех месяцев) – это прежде всего сигнал о повреждении ткани (реальном или потенциальном), для нее характерно выраженное адаптивное значение. Основу острой боли составляет местный патологический процесс, вызванный травмой, воспалительным или инфекционным процессом и другими причинами [4]. Хроническая боль или боль, персестирующая три и более месяцев подряд, отличается от острой не только продолжительностью, но и особенностями патогенеза и клиническими проявлениями. Например, одними из отличительных черт генеза хронической боли являются выраженная вовлеченность психических процессов [5] и генетическая обусловленность – наследуемость признака варьирует от 30 до 60% по разным данным [6]. В случае хронической боли роль первичного повреждающего фактора нивелируется, патологический процесс становится более обособленным и принимает диффузный характер. Хроническая боль теряет сигнальное значение и приобретает черты самостоятельного заболевания, требующего особых терапевтических подходов к лечению пациентов [7].

Так, на протяжении больше чем десяти последних лет развивается тенденция рассмотрения хронической боли как отдельного заболевания. В течение этого времени хроническая боль (в первую очередь, наиболее распространенная – скелетно-мышечная хроническая боль) остается одной из самых частых проблем со здоровьем в мире. Она встречается у 20 - 30%

взрослого населения планеты и является наиболее распространенной причиной нетрудоспособности [8]. В частности, боль в пояснице – самое часто встречающееся хроническое заболевание опорно-двигательного аппарата – по данным исследования Глобального бремени болезней (Global Burden of Disease Study) за 2017 год стабильно лидирует среди других заболеваний по количеству лет, прожитых с инвалидностью (years lived with disability, YLDs), за период с 1990 года [2]. А по данным начала 2018 года боль в пояснице стала причиной 21.7 млн потерянных лет трудоспособной жизни в мировом масштабе [9].

Известно множество причин хронической боли. Согласно масштабному эпидемиологическому исследованию среди европейцев [1], в котором респондентам предлагалось перечислить причины испытываемой ими хронической боли, самой распространенной причиной является остеоартрит (артрит) – о нем заявило 34% людей в выборке. Грыжи и дегенерация межпозвонковых дисков были заявлены в качестве причины хронической боли 15% респондентов. В ряде случаев испытуемые сообщали, что хроническая боль являлась следствием травмы (12% выборки) или хирургического вмешательства (3% опрошенных). Хроническая боль неустановленной этиологии встречалась у 12% выборки. Среди более редких причин хронической боли отмечены ревматоидный артрит (о нем сообщило 8% респондентов), мигрень (7% выборки), переломы позвоночника (у 6% испытуемых), повреждения нервов и хлыстовая травма (по 4% выборки каждый), рак (1% опрошенных) и др. Стоит добавить, что респонденты имели возможность указать несколько наиболее вероятных причин хронической боли одновременно.

Представления о природе боли значительно менялись с течением времени. В XIX – XX столетиях были распространены биомедицинские модели боли, в которых ведущая роль отводилась ноцицептивным процессам [10]. Одной из наиболее популярных моделей была специфическая теория боли (М. вон Фрей, 1895 год), которая постулировала существование уникальных рецепторов, проводящих путей и нервных центров для восприятия боли. Альтернативой ей служила неспецифическая теория (А. Гольдшейдер, 1894 год) боли, которая отрицала наличие специальных структур для восприятия, проведения и анализа болевых сигналов, полагая, что боль возникает в рецепторе любой модальности при чрезмерной его стимуляции [11]. Позднее распространение получила теория паттернов (Д. П. Нейф, 1929 год), согласно которой всякое соматосенсорное восприятие формируется на основе специфического паттерна нервных стимулов, при этом тип и интенсивность стимула определяется пространственно-временным профилем возбуждения. Наконец, в

1965 году Рональдом Мелзаком и Патриком Д. Уоллом была предложена теория воротного контроля (см. пункт 1.2.), в которой изложены современные представления о физиологических основах боли.

Однако по мере накопления знаний о боли (во многом благодаря пристальному изучению хронической боли) стало очевидно, что она имеет комплексную природу, основанную на совокупности биологических, психических и социальных факторов. Эти представления сформировали биопсихосоциальную модель боли [5,10], имеющую важнейшее значение для разработки актуальных методов терапии хронической боли. Согласно данной концепции боль представляется как результат взаимодействия биологических (физиологических, биохимических и молекулярно-генетических), психических (эмоциональных, когнитивных, поведенческих) и социально-демографических факторов, таких как: общественная поддержка, медицинское страхование, семейное окружение, культурные особенности и т. д. Биопсихосоциальная модель учитывает субъективность восприятия боли и влияние на него множества показателей: пола (боль чаще диагностируется у женщин), возраста, этнической принадлежности, индивидуального опыта и др. [1,12] Она также способствует лучшему пониманию взаимосвязей хронической боли с другими заболеваниями, например, депрессией и тревожным расстройством (известно, что пациенты с хронической болью часто страдают от депрессии [8], и на настоящий момент найдены доказательства в пользу генетической схожести этих заболеваний [5]). Таким образом, всестороннее рассмотрение боли формирует индивидуальный подход к пациентам с хронической болью, зачастую имеющей неустановленную причину и плохо поддающуюся лечению стандартными методиками [5]. Как можно заключить из всего вышеописанного, боль, в особенности хроническая, является сложным признаком с высокой гетерогенностью. По этой причине классификация боли представляется крайне непростой задачей, требующей последовательного и систематического подхода. О трудностях классификации боли речь пойдет в следующем разделе.

1.1. Классификация боли

Существует большое многообразие подходов к классификации боли. Мы остановимся лишь на некоторых из них.

Как уже было отмечено ранее, по продолжительности боль принято подразделять на острую (продолжительностью до 3 месяцев) и хроническую (продолжительностью от 3 месяцев). Иногда также выделяют подострую боль, продолжительностью от 6 до 12 недель [9].

В соответствии с предполагаемыми патофизиологическими механизмами, боль подразделяют на:

- ноцигенную – то есть боль, вызванную стимуляцией болевых рецепторов;
- нейрогенную – боль, обусловленную повреждениями и патологическими процессами в периферической или центральной нервной системе;
- психогенную – боль, вызванную психическими факторами.

По некоторым классификациям также принято выделять воспалительную боль (боль вследствие воспаления) и функциональную боль (гиперчувствительность к болевым ощущениям, вызванная нарушением обработки сигнала в центральной нервной системе – ЦНС) [13].

Ноцигенную боль, которую часто называют ноцицептивной, делят на соматическую поверхностную (при повреждении кожных покровов), соматическую глубокую (в результате повреждения костно-мышечной системы) и висцеральную (в следствие повреждения внутренних органов) [4].

Нейрогенную боль подразделяют на центральную и периферическую (или нейропатическую) боль.

По характеру локализации различают местную (непосредственно в очаге болевого воздействия), проекционную (по ходу нерва и на дистальных его участках при локализации болевого воздействия в проксимальном участке нерва), отраженную (в участках кожи, иннервируемых из того же сегмента спинного мозга, что и внутренние органы, в которых расположен источник болевого воздействия), иррадирующую (в области иннервации одной ветви нерва при болевом воздействии в зоне иннервации другой ветви того же нерва) и генерализованную (диффузно распределенную) боль. Отдельно выделяют фантомные боли, возникающие после удаления (ампутации) органа или его деафферентации [14].

Широко применяется классификация боли по топологии: боль в спине, боль в колене, бедре и т.д.

Боль также классифицируют на основе этиологии, например: боль, вызванная остеоартритом, боль в следствие травмы, идиопатическая боль (боль неизвестной природы).

Наконец, по характеру болевых ощущений, испытываемых пациентами, выделяют тупую, ноющую, колющую, схваткообразную боль и т. д.

На данный момент в клинической практике действует классификация боли (<https://icd.who.int/browse10/2016/en#/R52>), представленная в 10-ой редакции Международной Классификации Болезней (МКБ-10). Согласно ей, принято разделять сайт-

специфичную боль (боль в груди, зубную боль, боль в суставах и др.), головную боль и боль, не относящуюся к определенным органам и участкам тела. К последней относят острую боль, постоянную некупирующуюся боль и, так называемые, «другую постоянную боль» и «неуточненную боль».

Как видно из приведенного описания, классификация боли по МКБ-10, является далеко не полной, содержит значительную долю неопределенности и не отвечает современным представлениям в области алгологии (отрасли медицины, изучающей острую и хроническую боль). Международная ассоциация по изучению боли предложила свой вариант классификации, который реализован в новой редакции МКБ-11 (<https://icd.who.int/dev11/f/en#/http%3a%2f%2fid.who.int%2f11%2f01%2f01%2f01>), которая вступит в силу в 2022 году [15]. Новая классификация опирается на три главных параметра боли: топологию, этиологию и физиологию (ноцигенное, нейрогенное, психогенное происхождение боли). При этом классификация проводится, в первую очередь, по этиологии боли, затем по патофизиологии и, наконец, по топологии. Внедрение мультидоменной системы диагностических кодов в МКБ-11, облегчило создание такой многоступенчатой классификации боли, поскольку благодаря этой системе, один и тот же диагноз может быть отнесен к более чем одной категории.

В соответствии с МКБ-11, сайт-специфичная боль также рассматривается отдельно от боли как таковой, которая четко подразделяется на острую (включает в себя острую лицевую, головную и постоперационную боль) и хроническую боль, приравниваемую к самостоятельному заболеванию. Новая категория «хроническая боль» объединяет наиболее распространенные клинически значимые расстройства, которые разделены на семь групп: хроническая первичная боль, хроническая онкогенная боль, хроническая посттравматическая и послеоперационная боль, хроническая нейропатическая боль, хроническая головная и орофациальная боль, хроническая висцеральная боль и хроническая скелетно-мышечная боль. Ожидается, что введение новой медицинской классификации обеспечит более точную постановку диагноза и адекватное лечение хронической боли, отличающейся сложным патогенезом.

Для понимания патофизиологии хронической боли необходимо знание механизмов, лежащих в основе болевого ощущения. Рассмотрению данных аспектов будет посвящена следующая глава.

1.2. Описание физиологической основы боли

В данном разделе описаны основные пути проведения боли (компонентов рефлекторной нервной дуги) и ее механизмов. Особое внимание уделено модели воротного контроля, как основной концепции в физиологии боли.

1.2.1. Болевые рецепторы

В качестве первичного звена ноцицептивного пути, осуществляющего трансдукцию – то есть перевод воздействующего стимула в электрический импульс, могут выступать специализированные болевые рецепторы (ноцицепторы), локализованные в коже, глубоких тканях или внутренних органах. Ноцицепторы расположены по всему телу, за исключением головного мозга. Существует два основных типа ноцицепторов: А δ и С-ноцицепторы [4].

А δ ноцицепторы представляют собой свободные нервные окончания, разветвленные в виде дерева (миелиновые волокна). Они реагируют на сильные механические (высокопороговые А δ механоноцицепторы, 5 - 50 м/с) и, в некоторых случаях, на термические раздражители (А δ механо-термоноцицепторы, 3 - 20 м/с). А δ ноцицепторы локализируются, главным образом, в коже, однако они также встречаются в мышцах, суставах и внутренних органах. Передачу возбуждения от А δ ноцицепторов к восходящим путям в спинном мозге – трансмиссию – осуществляют слабомиелинизированные быстропроводящие А δ волокна (скорость 6 - 30 м/с, $d = 2 - 6$ мкм), проводящие информацию об острой, резкой, режущей боли, температурной и тактильной чувствительности [16]. Основным транзиттером А δ волокон является глутамат [17].

Второй тип ноцицепторов – это плотные некапсулированные гломерулярные тельца, афференты немиелинизированных С – волокон ($d = 0.4 - 1.2$ мкм), со скоростью проведения 0.5 – 2 м/с. С-ноцицепторы полимодальны, то есть они реагируют и на механические, и на температурные, и на химические (например, медиаторы воспаления) раздражители. С-волокна представлены во всем организме, кроме ЦНС [4]. Они передают информацию о тупой, ноющей боли, тактильных и тепловых ощущениях. Главным нейротранзиттером С-волокон является субстанция Р. Также транзиттерами выступают кальцитонин-ген-связанный пептид и вазоактивный интестинальный пептид.

Волокна А δ и С работают синергично, проводя основную массу болевых ощущений. Однако помимо них в ноцицепции участвуют быстрые миелиновые А β волокна (30 – 70 м/с, $d = 5 - 12$ мкм), которые не связаны с ноцицепторами и в норме отвечают за проприоцептивную чувствительность [18]. Тем не менее, повреждение и раздражение волокон А β распространено в случае нейропатической боли и наблюдается при аллодении

(восприятию неболевого стимула как болевого) и дизестезии (неприятных или патологических ощущениях при действии стимула или при его отсутствии).

Кроме Аδ и С-ноцицепторов за восприятие болевых стимулов отвечают рецепторы «интенсивности» и «молчащие» рецепторы, расположенные в тканях внутренних органов, на рассмотрении которых мы не будем останавливаться.

1.2.2. Задние рога спинного мозга

Болевые волокна входят в спинной мозг через спинномозговые нервы (если они отходят от шеи, туловища, конечностей), либо в составе тройничного нерва входят непосредственно в продолговатый мозг [16].

Перед вхождением в спинной мозг задний корешок спинномозгового ганглия разделяется на медиальную и латеральную часть, в которой проходят Аδ и С-волокна. Примерно в 30% случаев проксимальные аксоны волокон типа С после выхода из спинномозгового ганглия направляются к канатику – месту совместного хода афферентных и эфферентных нервных волокон – и входят в спинной мозг через передние корешки. Внутри спинного мозга ноцицептивные волокна разделяются на восходящую и нисходящую ветвь, и до вхождения в серое вещество задних рогов спинного мозга они могут направляться к нескольким сегментам спинного мозга, формируя густую сеть (в совокупности эта нейроанатомическая структура носит название заднерогового комплекса) [4]. Дальнейшую передачу болевого импульса осуществляют три типа нейронов, расположенных во II – III пластинках серого вещества задних рогов спинного мозга (желатинозной субстанции): специфические ноцицептивные нейроны, воспринимающие сигналы от Аδ и С-волокон; нейроны широкого динамического диапазона (wide dynamic range, WDR нейроны), принимающие импульсы от Аδ, Аβ и С-волокон; и неноцицептивные нейроны, в норме не проводящие болевые сигналы, но передающие ноцицептивные импульсы в случае повреждений или сильной стимуляции по Аβ и Аδ волокнам. В большинстве случаев передача возбуждения от первичных афферентных нейронов на нейроны желатинозной субстанции осуществляется через вставочные нейроны, синапсы которых могут облегчать передачу импульса или напротив, препятствовать его проведению [4].

На уровне задних рогов спинного мозга осуществляется процесс модуляции – то есть изменение сигнала перед передачей его в ЦНС. Существует два механизма модуляции: тормозной и возбуждающий.

Тормозной механизм может действовать на локальном (сегментарном) уровне, посредством действия глицина и γ-аминомасляной кислоты (ГАМК), или нисходящем (надсегментарном) уровне, за счет действия серотонина и норадреналина.

Возбуждающая медиация осуществляется через глутамат, который высвобождается первичными афферентными волокнами. Он действует на AMPA (α -амино-3-гидрокси-5-метил-4-изоксазол-пропионовая кислота) рецепторы, обеспечивающие быструю синаптическую передачу, и NMDA (N-метил-D-аспартат) рецепторы, усиливающие и поддерживающие ее на протяжении длительного времени после воздействия стимула [16].

1.2.3. Воротный контроль

Теория воротного контроля – это физиологическая концепция, описывающая контроль над проведением болевых импульсов от первичных афферентов к высшим отделам ЦНС. Согласно данной теории, задние рога спинного мозга представляют собой ворота, через которые ноцицептивные сигналы периферических болевых волокон попадают в ЦНС и далее направляются в головной мозг. В общем случае эти ворота открыты, однако существует два фактора, способных их закрыть, – это импульсы, поступающие в спинной мозг по быстрым миелиновым волокнам, передающим тактильные возбуждения, и определенные нисходящие импульсы из ЦНС [4]. Первый механизм основан на действии стороннего раздражителя (механическое растирание кожи, разогревающие мази и т. д.), блокирующем проведение болевых импульсов по восходящим путям спинного мозга. Второй механизм (закрытие ворот изнутри) реализуется при активации нисходящих тормозных волокон ствола мозга: в случае их прямой стимуляции или за счет гетеросегментарной акупунктуры (низкочастотной высокоинтенсивной периферической стимуляции). Нисходящие волокна, в свою очередь, активируют вставочные нейроны (в поверхностных слоях задних рогов спинного мозга), которые обеспечивают постсинаптическое торможение клеток желатинозной субстанции и, тем самым, блокируют передачу болевого сигнала в вышележащие отделы нервной системы.

1.2.4. Опиоидные рецепторы

Известно, что эндогенные опиоидные пептиды, такие как: энкефалины, эндорфины, динарфины – связываются с особыми опиоидными рецепторами в случае, когда ноцицептивный сигнал преодолевает болевой порог. Антиноцицептивный эффект такого взаимодействия достигается путем блокады распространения потенциалов действия и снижения выброса возбуждающих нейромедиаторов (например, субстанции P). Этот принцип реализуется в медицинской практике при применении опиатных анальгетиков. Среди всех типов опиоидных рецепторов клинически значимыми являются μ -, κ - и δ -рецепторы (Г. В. Костерлиц, С. Д. Патерсон, 1985). Наибольшее количество опиоидных рецепторов обнаружено в задних рогах спинного мозга, в среднем мозге и таламусе. Спинальные опиоидные рецепторы расположены преимущественно вблизи окончаний C-

волокон в I пластинке задних рогов спинного мозга [4]. Синтез этих рецепторов осуществляется в телах малых клеток спинных ганглиев, откуда они транспортируются по аксонам в проксимальном и дистальном направлениях.

Однако опиатные рецепторы и их лиганды были найдены не только в ЦНС, но и на поверхности клеток иммунной системы, в периферической нервной системе, в воспаленных тканях (активация рецепторов наступает в течение нескольких часов с момента начала воспалительной реакции). Показано, что порядка 70% рецепторов к энкефалинам и эндорфину локализовано в пресинаптической мембране ноцицепторов, таким образом, в большинстве случаев болевой сигнал подавляется, не достигнув дорсальных рогов спинного мозга [4]. Динорфин провоцирует выделение ГАМК, путем активации рецепторов и ингибирования интернейронов. Это приводит к гиперполяризации нейронов дорсальных рогов и блокированию передачи сигнала.

1.2.5. Восходящие ноцицептивные пути

Большинство восходящих болевых путей расположено в переднебоковых канатиках белого вещества спинного мозга, с противоположной стороны от заднерогового комплекса. Однако некоторая часть волокон спиноталамического и спиноретикулярного трактов, проводящих ноцицептивные сигналы, находится в заднебоковом канатике [4].

Спиноталамический тракт делится на [16]:

1. Неоспиноталамический тракт – передает сигнал от А-волокон, характеризуется моносинаптической передачей и быстрым проведением возбуждения, связан с хорошо локализованной резкой болью. Неоспиноталамический тракт восходит к специфическим латеральным ядрам таламуса (вентрозаднелатеральному и вентрозаднемедиальному ядрам).
2. Палеоспиноталамический тракт – передает сигнал от С-волокон, характеризуется полисинаптической передачей и медленным проведением, связан с плохо локализованной ноющей болью. Палеоспиноталамический тракт направляется к неспецифическим медиальным таламическим ядрам (медиальному и интраламинарному ядрам, срединному центру), часть волокон восходит к ретикулярной формации.

По некоторым данным [4], нарушение баланса между медиальным и латеральным таламическими ядрами приводит к их сверхторможению ретикулярным ядром таламуса и парадоксальной активации участков коры головного мозга, связанных с ноцицепцией. Проведение медиальной стереотаксической таламотомии в более чем 60% случаев облегчает хроническую, терапевтически резистентную нейрогенную боль (как периферическую, так и центральную) на 50 - 100%.

Возбуждение, проводимое по неоспиноталамическому тракту, передается на волокна, проводящие импульсы через заднее бедро внутренней капсулы к первичной и вторичной соматосенсорной коре и постцентральной извилине [4]. Пространственная локализация боли обеспечивается высоким уровнем топической организации бокового таламического ядра.

Ноцицептивные сигналы, проводимые по палеоспиноталамической системе, передаются на срединное таламическое ядро и диффузно проецируются на новую кору головного мозга. Показано, что за эмоциональное восприятие боли отвечает лобная кора, а для цингулярной извилины и орбитальной фронтальной коры головного мозга доказана активация при ноцицептивном воздействии [4]. На этих наблюдениях основаны успешные хирургические методы лечения хронической онкогенной боли, такие как цингулотомия и префронтальная лоботомия.

Из вышеописанного следует, что в головном мозге не существует единого болевого центра, восприятие боли и реакция на нее – это функция ЦНС в целом.

1.2.6. Нисходящая модуляция боли

Префронтальная и островковая зоны коры головного мозга осуществляют контроль над гипоталамусом, в аркуатном (дугообразном) ядре которого находятся нейроны, трансмиттером которых выступает В-эндорфин. Эти клетки оказывают тормозное влияние на вставочные нейроны периакведуктального серого вещества (PAG) головного мозга. Ингибирование данных интернейронов приводит к активации клеток, ведущих свои аксоны к большому ядру шва ретикулярной формации. Нейроны большого ядра шва являются в большинстве своем серотонинергическими, некоторые – норадреналинергическими, как и подавляющее число нейронов ретикулярной формации. Аксоны всех этих клеток идут вниз по дорсолатеральному канатику спинного мозга и заканчиваются в поверхностных слоях заднего рога. Через них осуществляется тормозная модуляция ноцицептивных стимулов, поступающих в спинной мозг от первичных афферентов.

Действие трициклических антидепрессантов (ТЦА) вызывает анальгезию, блокируя обратный захват серотонина и норадреналина, усиливая, тем самым, ингибирующее влияние на нейроны спинного мозга. Помимо ТЦА для лечения болевого синдрома используют прямую электростимуляцию перивентрикулярного и периакведуктального серого вещества головного мозга [16].

1.3. Физиологическая основа хроническая боли

Хроническая боль может иметь как ноцигенное, так и нейрогенное происхождение [19]. Однако необходимо принять во внимание, что разделение хронической боли на

ноцигенную и нейрогенную является в значительной степени условным. В частности, по причине того, что на фоне длительного течения ноцигенной боли наблюдается развитие нейрогенных болевых процессов. Демаркация также представляется весьма затруднительной в случае ряда заболеваний. Так например, при онкогенных болях опухоль одновременно вызывает повреждение тканей – ноцигенный компонент боли – и компрессию нервов – нейрогенная составляющая патогенеза [4]. Кроме того, все хронические боли имеют выраженную психогенную компоненту, по этой причине более корректно определять хроническую боль как смешанную. Тем не менее, классификация хронической боли на ноцигенную и нейрогенную зачастую бывает полезна на ранних этапах терапии при выборе методики лечения, поскольку хроническая ноцигенная и хроническая нейрогенная боль имеют различный патогенез.

Как было отмечено ранее, ноцигенная боль – это боль, вызванная активацией ноцицепторов. Продолжительное или сильное воздействие болевого стимула, например, медиатора воспаления, приводит к повышению чувствительности нервных волокон [19]. Этот процесс носит название периферической сенситизации, следствием которой является первичная гипералгезия – то есть повышенная чувствительность к ноцицептивным импульсам [4]. Первичная гипералгезия распространяется на поврежденные ткани и характеризуется снижением болевого порога и толерантности к ноцицептивному воздействию механических и термических стимулов. Периферическая сенситизация выражается в снижении порога активации ноцицепторов, увеличении рецептивных полей, частоты и длительности разрядов, результатом чего является усиление потока болевых импульсов [4]. Нейроны дорсальных рогов спинного мозга реагируют на усиление входящего афферентного потока повышением возбудимости и реактивности – возникает так называемая центральная сенситизация (в дальнейшем она распространяется и на ноцицептивные нейроны таламуса и соматосенсорной коры головного мозга). Усиленная деполяризация клеток приводит к активации NMDA и нейропептидных рецепторов, что влечет за собой резкое увеличение концентрации внутриклеточного Ca^{2+} . Ионы кальция запускают сигнальные пути и активируют экспрессию ряда генов, чьи белковые продукты вовлечены в регуляцию долговременной возбудимости мембраны [19]. В частности, показано, что устойчивое повышение возбудимости ноцицептивных нейронов опосредовано экспрессией генов раннего ответа, таких как *c-fos*, *c-jun*, *junB* и др. [4] Однажды возникшая сенситизация ноцицептивных нейронов способна к самоподдержанию в течение длительного времени, независимо от наличия или отсутствия дополнительных импульсов из первичного очага повреждения.

Центральная сенситизация сопровождается развитием вторичной гипералгезии, которую отличают нормальный болевой порог, снижение болевой толерантности по отношению к механическим раздражителям и локализация за пределами очага поражения. Центральная сенситизация также приводит к развитию ноцицептивной реакции в ответ на стимулы, в норме не являющиеся болевыми (аллодиния) [4]. Кроме того, клинические исследования свидетельствуют о том, что центральная сенситизация приводит к структурным изменениям в мозге, которые, впрочем, являются обратимыми в случае устранения боли [10,19].

Патофизиологический механизм нейрогенной боли (далее речь пойдет о нейропатической боли, как наиболее частой и хорошо изученной) принципиально отличается от такого для ноцицептивной боли. Повреждение периферических нервов вызывает не зависящую от стимуляции (эктопическую) активность нервных волокон. Эктопические разряды отличаются большой амплитудой и продолжительностью. Их источниками выступают участки демиелинизации и регенерации нервов, невромы, и нейроны спинномозговых ганглиев, связанные с аксонами поврежденных клеток. Эти зоны обладают самоподдерживающейся активностью и способны к перекрестному возбуждению соседних волокон, что является основой для дизестезии [4]. В ответ на повреждение развивается иммунная реакция: непосредственно в месте самого повреждения, в спинальных ганглиях, где находятся тела соматосенсорных нейронов, а также в задних рогах спинного мозга [19]. Активная микроглия дорсальных рогов выделяет химические медиаторы, модулирующие активность рядом расположенных нейронов. Один из таких медиаторов – нейротрофический фактор мозга – уменьшает ингибирующие эффекты от воздействия ГАМК и глицина. Растворивание нейронов открывает полисинаптические связи в дорсальном роге спинного мозга, что приводит к дополнительному усилению входящего сигнального потока из поврежденного нерва. Как и в случае ноцигенной боли, возникает центральная сенситизация [4]. При нейропатической боли она выражается в увеличении спонтанной импульсной активности нейронов задних рогов спинного мозга, вспышках высокочастотных разрядов, увеличении рецептивных полей, повышенной реактивности нейронов и более продолжительном времени послеразрядов. Интенсивная глутаматергическая передача сигнала усугубляется относительным дефицитом нейротрансмиттера и вызывает эксайтотоксическую гибель клеток дорсальных рогов спинного мозга с последующим их замещением глиальными клетками. Гибель тормозных интернейронов, действие нейромедиаторов и нейромодуляторов (возбуждающих аминокислот, нейрокининов и оксида азота), дефицит опиоидов и рецепторов к ним, а также

повышенная чувствительность клеток к субстанции Р и нейрокинину А [4] приводят к устойчивой деполяризации нейронов и глубокому дисбалансу между процессами торможения и возбуждения в нервной ткани [19]. Агрегаты взаимодействующих сенситизированных нервных клеток с повышенной возбудимостью и нарушенными механизмами торможения развивают продолжительную самоподдерживающуюся патологическую активность, не зависящую от периферической афферентной стимуляции. Сенситизация нейронов регистрируется и в высших отделах ЦНС – вентробазальном и парафасцикулярном таламических комплексах и соматосенсорной коре головного мозга. Независимо от основного патофизиологического процесса, определяющего развитие хронического болевого синдрома, во всех случаях отмечается значительное влияние комплекса психических и социальных факторов на течение заболевания [5]. Зачастую хроническая боль сопровождается повышенной тревожностью и депрессивными расстройствами [10], что вызвано общностью механизмов патогенеза [20]. Однако причинно-следственные взаимосвязи между этими состояниями до сих пор не установлены [21], и вопрос о том, может ли хроническая боль иметь исключительно психогенное происхождение, до настоящего времени остается открытым.

Подводя итог, можно сделать вывод о том, что центральная сенситизация является ключевым понятием при описании патофизиологических механизмов хронической боли. Патогенез хронической боли отличается сложностью и недостаточной изученностью. По этой причине лечение хронической боли является крайне затруднительной терапевтической проблемой. Поиск биомаркеров хронической боли представляется важной задачей биомедицинской отрасли, направленный на разработку новых методик по устранению хронического болевого синдрома и предотвращению его развития.

2. Методы анализа генетических ассоциаций

Центральной проблемой генетики как науки, имеющей важное фундаментальное и прикладное значение, является установление взаимосвязей между генотипом и фенотипом. Значительную роль в изучении генетического контроля признаков играют различные методы картирования генов.

Выделяют три основных подхода генетического картирования: функциональное картирование, исследование генов-кандидатов и позиционное картирование. Функциональное картирование позволяет установить связь между признаком и геном через функцию белка, имеющую определяющее значение для развития признака (признак - функция - ген - карта). Данный метод подразумевает высокую степень изученности биологических основ формирования признака и имеет наибольшее применение при

исследовании менделевских признаков. Для изучения сложных полигенных признаков чаще используются другие подходы. О них и пойдет речь в данной главе.

2.1. Ген-кандидатные исследования

Подход, основанный на тестировании генов-кандидатов, предполагает наличие представлений о биологической природе признака, опираясь на которые, можно было бы выдвинуть гипотезу о вовлеченности тех или иных генов (генов-кандидатов), в процесс развития данного признака. Для проверки этой гипотезы определяют наличие связи между полиморфизмами генов-кандидатов и фенотипической изменчивостью. Необходимость понимания биологии признака является ограничивающим фактором, зачастую препятствующим применению этого метода в генетике. С другой стороны, наличие гипотезы о связи гена с фенотипом повышает уровень доверия к статистически значимым результатам исследования и облегчает их интерпретацию.

Вопреки имеющимся ограничениям, до недавнего времени этот подход был настолько популярен, что фактически являлся синонимом генетического тестирования ассоциаций для количественных и бинарных признаков, ввиду своей относительной простоты и доступности. Однако с развитием и распространением технологии ДНК-микрочипов, данный метод практически утратил свою актуальность, уступив место подходам, позволяющим работать с полногеномными данными.

2.2. Полногеномный анализ ассоциаций

В ситуациях, когда биохимические основы формирования признака мало изучены и нет предположений о генах-кандидатах, применяется метод позиционного картирования генов. Первым шагом данного подхода является анализ совместной встречаемости признака и генетических маркеров с заранее известным положением в геноме. Определение локусов, ассоциированных с признаком, позволяет связать его с генами, расположенными в данном регионе, и установить их роль в развитии признака через выполняемые ими функции (признак - карта - ген - функция).

Позиционное картирование генов основывается на ряде биологических явлений, таких как: сцепление генов, их рекомбинация в мейозе и полиморфизм генов. Сцепление генов обуславливает их совместное наследование в виде так называемых блоков сцепления, а феномен генетической рекомбинации обеспечивает разбиение этих блоков на более мелкие фрагменты при наследовании в ряду поколений. Явление полиморфизма генов объясняет существование нескольких вариантов блоков сцепления, несущих один и тот же участок генома, за счет разного набора аллелей в них. Такие полиморфные локусы выступают в качестве генетических маркеров. Для позиционного картирования гена достаточным

является установить связь признака с маркерными аллелями. Решение этой задачи имеет две реализации: с помощью анализа сцепления и с помощью анализа ассоциаций. В данной главе мы ограничимся рассмотрением только второго метода.

Наиболее распространенным методом картирования, применяемым при исследовании сложных полигенных признаков, является полногеномный анализ ассоциаций (ПГАА). Данный метод был впервые применен в 2005 году для исследования возрастной дегенерации жёлтого пятна у человека [22]. С того момента использование метода ПГАА стало принимать все больший размах. Так, например, уже к 2011 году благодаря методу ПГАА в более чем 3000 исследований найдено свыше 60 тысяч уникальных SNP (single nucleotide polymorphism – однонуклеотидный полиморфизм), ассоциированных с более чем тысячей признаками и заболеваниями человека [23]. Преимуществом ПГАА перед методом анализа сцепления является то, что он не требует анализа родословных и позволяет работать с выборками неродственных организмов, исходя из положений об общности их происхождения и отсутствии аллельной гетерогенности в популяции.

В основе метода ПГАА лежит феномен неравновесия по сцеплению, который заключается в том, что в популяции вероятность совместной встречи сцепленных локусов оказывается выше таковой при условии их независимой сегрегации. Данный подход к картированию генов имеет разрешающую способность до 0.1 сМ в случае наиболее близко расположенных локусов. Генетические варианты, имеющие вклад в формирование признака, находятся в определенном генетическом контексте благодаря неравновесию по сцеплению с соседними локусами. За счет этого наличие ассоциации между признаком и генотипированным маркером дает возможность сделать предположение о нахождении в ближайшем окружении с ним функционального аллеля, определяющего развитие признака [24] (или же о том, что сам маркер является функциональным вариантом, что наблюдается значительно реже).

Метод ПГАА применяется на выборках размером до нескольких сотен тысяч фенотипированных индивидов, генотипированных по десяткам миллионов полиморфных локусов. Это стало возможным благодаря использованию технологии ДНК-микрочипов. Наиболее подходящими генетическими маркерами для ПГАА являются SNP-маркеры, ввиду высокой плотности их распределения по геному.

Процедура проведения ПГАА предполагает проведение множественного тестирования. Независимость тестов позволяет использовать в качестве поправки на множественное сравнение поправку Бонферрони [25]. Таким образом, порог статистической значимости вычисляется как α/N , где α – это общая экспериментальная ошибка первого рода (обычно

0.05), а N – число независимых экспериментов (количество блоков сцепления в геноме, для человека имеет порядок $1e-06$ при тестировании распространенных SNP с частотой редкого аллеля ≥ 0.05). Следовательно, уровень полногеномной значимости результатов ПГАА для человека определяется как $p\text{-value} < 5.0e-08$.

Нередко перед исследователями встает задача повышения мощности метода, одним из способов решения которой является объединение результатов ПГАА различных выборок путем мета-анализа [26]. Объединение суммарных статистик в ходе мета-анализа осуществляется исключительно по маркерам, общим для всех объединяемых исследований. Поскольку в разных исследованиях генотипирование может проводиться с помощью разных микрочипов, для унификации набора полиморфизмов проводят процедуру импутации генотипов [27]. Импутация – это восстановление неопределенных генотипов по известным, генотипированным SNP с использованием референсных генотипов. Чаще всего для этого используют данные о референсных гаплотипах, с помощью которых определяется вероятность наблюдения того или иного генотипа в негенотипированных участках. Кроме того, для импутации также используются данные секвенирования проекта «1000 Геномов» [28] и Haplotype Reference Consortium [29]. Проведение импутации в сочетании с мета-анализом дает возможность объединения результатов нескольких независимых исследований, повышая мощность анализа и достоверность результатов ПГАА, и позволяет обнаружить ассоциации признаков с генетическими вариантами со слабыми эффектами, что может быть недостижимо на выборках малого объема.

Важным этапом проведения ПГАА является репликация – подтверждение статистически значимых результатов анализа на данных независимой выборки. Лocusы считаются реплицированными, если на репликационной выборке для них получено значение тестовой статистики, соответствующее экспериментальной ошибке первого рода $\alpha < 0.05$ [30]. Благодаря проведению репликации при ПГАА, найденные locusы редко оказываются ложноположительными, и результаты подтверждаются другими исследованиями.

Установление ассоциаций между генетическими маркерами и признаком не является достаточным условием для понимания его генетической архитектуры и молекулярно-генетических механизмов контроля над его развитием. Последующие этапы анализа направлены на определение функционального варианта, описание механизмов его действия и оказываемого эффекта на некоторый ген или группу генов. Важными инструментами, применяемыми на этой стадии ПГАА, являются *in silico* методы, о которых речь пойдет далее.

2.2.1. Биоинформатические методы обработки результатов полногеномного анализа ассоциаций

В настоящее время разработано множество методов анализа результатов ПГАА. К ним относятся методы тестирования ассоциаций отдельных полиморфизмов (single-variant association tests), в частности, методы мета-анализа, условного анализа и импутации на основе суммарных статистик; также существует круг методов, проверяющих ассоциацию на уровне гена (gene-based association tests), к ним относятся подходы, использующие транскриптомные данные, и методы анализа редких вариантов; еще одна группа методов – это методы тонкого картирования (fine-mapping), интегрирующие данные функциональной аннотации и/или данные по различным этническим группам; применяются и методы предсказания полигенного риска развития заболевания и определения полигенной архитектуры признака, а также методы совместного анализа нескольких признаков [31]. В этой главе мы коротко рассмотрим некоторые из перечисленных подходов, использующихся для решения задач данного исследования.

Методы условного анализа применяются для того, чтобы определить число независимых сигналов внутри ассоциированного локуса. Они опираются на данные о неравновесии по сцеплению и проверяют каждый полиморфизм в локусе на наличие ассоциации с признаком интереса, проводя оценку его эффекта при условии эффекта ведущего (то есть наиболее значимого) SNP в локусе [31]. Это позволяет отличить наведенные сигналы, которые возникают из-за сильного неравновесия по сцеплению с ведущим SNP, от независимых сигналов ассоциации. Методы условного анализа традиционного предполагали использование персональных геномных данных, однако современные подходы позволяют проводить анализ с использованием суммарных статистик ПГАА. В частности, эта методология реализована в программном пакете COJO (Conditional and joint analysis) [32]. Для отбора наиболее вероятных функциональных вариантов в ассоциированном локусе, то есть их приоритизации, можно использовать методы функциональной аннотации SNP. Одним из наиболее часто применяемых для этого инструментов является программный пакет VEP (Variant Effect Predictor, <https://www.ensembl.org/Tools/VEP>). Он использует данные об экспрессии генов в разных тканях, сведения о регуляторных районах генов и частотах встречаемости полиморфизмов в популяциях для аннотации замен в кодирующих и не кодирующих областях. VEP позволяет определить положение полиморфизма относительно гена (сайт сплайсинга, интрон, 5'-нетранслируемая область и т. д.) и оценить его влияние на транскрипцию, сплайсинг, экспрессию, а так же на функцию конечного белка.

Для приоритизации генов (поиска наиболее вероятно ассоциированных с признаком генов) и функциональной аннотации ассоциированных локусов проводится анализ обогащения, то есть анализ перепредставленности групп генов в ассоциированных локусах. Для этого используется заранее оцененные вероятности принадлежности каждого локуса к определенной генной сети, молекулярному каскаду и т. д. Этот метод позволяет определить наиболее вероятные биологические пути и процессы, в которых участвуют ассоциированные гены, а также установить, в каких тканях они экспрессируются. Такой анализ проводится, в частности, с помощью программы DEPICT (Data-driven Expression Prioritized Integration for Complex Traits) [33]. В основе этого подхода лежит предположение о том, что ассоциированные с определенным признаком гены вовлечены в одни и те же биохимические пути или ко-экспрессируются в одних и тех же тканях, то есть имеют сходную функциональную аннотацию. Таким образом, в ассоциированных локусах приоритизируются функционально близкие гены.

Многие сложные признаки имеют общую генетическую природу, которая может быть обусловлена наличием общих SNP с ненулевым эффектом (плейотропия), либо корреляцией между размерами эффектов функциональных SNP (генетическая корреляция) [31]. Для поиска плейотропных эффектов может быть использован метод Менделевской рандомизации (Summary data-based Mendelian Randomization, SMR) в сочетании с тестом на гетерогенность (Heterogeneity in Dependent Instruments, HEIDI) [34]. Генетические корреляции могут быть оценены с использованием персональных данных с помощью метода максимального правдоподобия [35] или на суммарных статистиках. Например, было показано, что генетические корреляции можно вычислить на основании полигенного риска признаков, однако этот подход может иметь завышенные оценки [31]. Другой способ предполагает использование метода LD Score регрессии [36], который представляет z-статистику (отношение размера эффекта SNP на признак к стандартной ошибке эффекта) для каждого SNP как функцию от параметра, отражающего неравновесие по сцеплению данного варианта с другими полиморфизмами в геноме. Оценка генетической корреляции между двумя признаками осуществляется следующим образом: сначала вычисляется произведение z-статистик SNP для пары признаков, а затем с помощью метода LD Score регрессии производится оценка коэффициентов линейной модели (угла наклона регрессионной прямой и интерсепта – точки пересечения регрессионной прямой с осью Oy). Угол наклона регрессионной прямой пропорционален генетической ковариации между признаками, что позволяет оценить их генетическую корреляцию. Показано, что результаты, полученные этим методом, с высокой точностью

совпадают с оценками, полученными на персональных данных [35]. Кроме того, отмечено, что методы вычисления генетических корреляций на основе полногеномных данных эффективны как при работе с признаками, имеющими набор ассоциированных локусов, так и с признаками, не имеющими статистически значимых сигналов [37]. Примечательно также, что суммарные статистики для некоторого признака могут быть получены без непосредственного проведения ПГАА, путем преобразования суммарных статистик генетически коррелированных с ним признаков [38].

3. Метод главных компонент и его применение в биологических исследованиях

Многие отрасли науки оперируют большими объемами данных, анализ которых требует особых подходов. Одной из задач при работе с такими массивами данных является уменьшение их размерности с сохранением максимума информации. Метод главных компонент, впервые опубликованный в 1901 году в работе Карла Пирсона [39] и развитый позднее Гарольдом Хотеллингом [40], является одним из старейших и наиболее широко используемых подходов к решению этой проблемы. Кроме того, метод главных компонент, помимо перехода к некоррелированным переменным, позволяет визуализировать сложный набор данных, определить наиболее изменчивые переменные и увидеть особенные наблюдения в выборке. В то же время, данный метод может существенно затруднить интерпретацию результатов, что осложняет его применение на практике.

Идея метода главных компонент заключается в переходе к новому набору линейно независимых переменных. При этом каждая новая переменная – главная компонента – является линейной комбинацией исходных переменных и вычисляется таким образом, чтобы ее дисперсия была максимальной. Задача о нахождении главных компонент может быть сведена к вычислению собственных чисел и собственных векторов матрицы ковариаций или матрицы корреляций исходного набора данных.

Метод главных компонент, как описательный инструмент, не нуждается в предположениях о распределении набора данных и потому является адаптивным исследовательским методом, применимым к числовым данным различных типов [41]. Данный подход имеет множество модификаций, позволяющих использовать его для решения задач в различных отраслях науки. Далее мы остановимся на стандартном методе главных компонент и подробно рассмотрим его математический аппарат.

3.1. Математический аппарат метода главных компонент

Стандартным контекстом применения метода главных компонент является набор данных с наблюдениями p числовых параметров для n объектов или индивидов. Эти данные определяют p n -мерных векторов x_1, \dots, x_p или, что одно и то же, матрицу X размера $n \times p$,

j -ый столбец которой представляет собой вектор x_j наблюдений j -ой переменной. Задача – найти линейную комбинацию столбцов матрицы X с максимальной дисперсией. Линейная комбинация определяется как $\sum_{j=1}^p a_j x_j = Xa$, где a – это вектор констант a_1, a_2, \dots, a_p . Дисперсия каждой такой линейной комбинации вычисляется по формуле: $var(Xa) = a'Sa$, где S – это выборочная ковариационная матрица для исходного набора данных, а символ $'$ обозначает операцию транспонирования [41]. Таким образом, вычисление линейной комбинации с максимальной дисперсией эквивалентно получению p -мерного вектора a , максимизирующего квадратичную форму $a'Sa$. Для того чтобы эта задача имела единственное решение, необходимо наложить дополнительное ограничение. Наиболее распространенное ограничение подразумевает работу с нормированными (единичными) векторами, то есть требуется, чтобы $a'a = 1$. Задача нахождения a при этом равносильна задаче максимизации выражения $a'Sa - \lambda(a'a - 1)$, где λ – это множитель Лагранжа [41]. Дифференцирование по вектору a и приравнивание к нулевому вектору дает уравнение вида $Sa - \lambda a = 0 \Leftrightarrow Sa = \lambda a$ (1).

Отсюда a – это единичный собственный вектор, а λ – соответствующее ему собственное число матрицы ковариаций S . В частности, интерес представляет наибольшее собственное число λ_1 (и соответствующий ему собственный вектор a_1), поскольку собственное число равно дисперсии линейной комбинации, определяемой соответствующим собственным вектором a : $var(Xa) = a'Sa = \lambda a'a = \lambda$ [41]. Равенство (1) также сохранится, если собственные вектора умножить на -1 , что означает, что абсолютные знаки при всех коэффициентах произвольны и важным является только их отношение и относительные величины коэффициентов.

Любая действительная симметричная матрица размера $p \times p$, такая как ковариационная матрица S , имеет ровно p действительных собственных чисел λ_k ($k = 1, \dots, p$), и соответствующие им собственные вектора образуют набор ортонормированных векторов, то есть $a_k'a_{k'} = 1$ при $k = k'$ и $a_k'a_{k'} = 0$ в противном случае. Метод множителей Лагранжа с добавочным ограничением на ортогональность векторов при разных коэффициентах служит демонстрацией того, что полный набор собственных векторов матрицы S является решением проблемы получения p новых некоррелированных друг с другом линейных комбинаций $Xa_k = \sum_{j=1}^p a_{jk} x_j$ с максимальной дисперсией [42]. Условие некоррелированности происходит из того факта, что ковариация между двумя линейными комбинациями Xa_k и $Xa_{k'}$ определяется как $a_{k'}'Sa_k = \lambda_k a_{k'}'a_k = 0$ при $k' \neq k$.

Линейные комбинации Xa_k являются главными компонентами исходного набора данных, однако некоторые ошибочно называют главными компонентами собственные вектора a_k . В стандартной терминологии метода главных компонент, элементы собственных векторов a_k принято называть «нагрузками» (loadings) главных компонент, а элементы линейных комбинаций Xa_k – «счетами» (scores) главных компонент.

Распространенной практикой при работе со стандартным методом главных компонент является определение главных компонент как линейных комбинаций центрированных переменных x_j^* , таких, что $x_{ij}^* = x_{ij} - \bar{x}_j$, где \bar{x}_j – это среднее значение наблюдений для j -ой переменной. Это преобразование не влияет на решение, так как матрица ковариаций для центрированных и нецентрированных переменных одинакова, однако оно оказывается удобным для альтернативного, более геометричного подхода к вычислению главных компонент [41].

Определив X^* , как матрицу размера $n \times p$ со столбцами центрированных переменных x_j^* , получаем: $(n - 1)S = X^{*'}X^*$ (2).

Уравнение (2) связывает разложение ковариационной матрицы S на собственные числа вектора с сингулярным разложением матрицы X^* центрированных значений [41]. Любая произвольная матрица Y размера $n \times p$ и рангом r ($r \leq \min\{n, p\}$) может быть записана как $Y = ULA'$ (3), где U и A – это матрицы размера $n \times r$ и $p \times r$ соответственно, с ортогональными столбцами ($U'U = I_r = A'A$, где I_r – единичная матрица размера $r \times r$), а L – диагональная матрица размера $r \times r$ [41]. Столбцы матрицы A называют правыми сингулярными векторами матрицы Y . Они являются собственными векторами матрицы $Y'Y$ размера $p \times p$, соответствующими ее ненулевым собственным числам. Столбцы матрицы U называют левыми сингулярными векторами матрицы Y . Они являются собственными векторами матрицы YY' размера $n \times n$, связанными с ее ненулевыми собственными числами. Диагональные элементы матрицы L носят название сингулярных чисел матрицы Y , и они являются неотрицательными квадратными корнями из общих для матриц $Y'Y$ и YY' ненулевых собственных чисел. Предположим, что диагональные элементы матрицы L расположены по убыванию, что однозначно определяет порядок столбцов матриц U и A (кроме случая равных сингулярных чисел). Тогда, приняв $Y = X^*$, получим, что правые сингулярные вектора матрицы X^* являются векторами a_k нагрузок главных компонент. Ввиду ортогональности столбцов матрицы A , столбцы матрицы, полученной перемножением $X^*A = ULA'A = UL$, будут главными компонентами матрицы X^* . Дисперсии этих главных компонент будут рассчитываться как квадраты сингулярных чисел

матрицы X^* , деленные на $(n - 1)$. С учетом (2) и перечисленных свойств, получаем: $(n - 1)S = X^{*'}X^* = (ULA')'(ULA') = ALU'ULA' = AL^2A'$ (4), где L^2 – это диагональная матрица с квадратами сингулярных чисел (то есть собственными числами матрицы $(n - 1)S$) [41]. Уравнение (4) – это спектральное разложение (разложение на собственные вектора и числа) матрицы $(n - 1)S$. Таким образом, метод главных компонент равнозначен сингулярному разложению матрицы центрированных переменных X^* .

Свойства сингулярного разложения предполагают интересные геометрические интерпретации метода главных компонент [41]. Присвоив некоторый ранг r матрице Y размера $n \times p$, получим, что матрица Y_q , такого же размера, но с рангом $q < r$, элементы которой минимизируют сумму квадратов отклонений от соответствующих элементов матрицы Y , будет раскладываться как $Y_q = U_q L_q A_q'$ (5), где L_q – это диагональная матрица размера $q \times q$ первых (наибольших) q диагональных элементов матрицы L , а U_q и A_q – это матрицы размера $n \times q$ и $p \times q$ соответственно, полученные путем сохранения q соответствующих столбцов матриц U и A .

В нашем контексте, n строк ранга r матрицы центрированных переменных X^* дают диаграмму рассеяния n точек в r -мерном подпространстве действительных чисел \mathbb{R}^p , с началом координат в качестве центра тяжести диаграммы рассеивания. Из приведенного выше результата следует, что «лучшее» n -точечное приближение к этой диаграмме рассеяния в q -мерном подпространстве задается строками X_q^* , определенными из уравнения (5), где «лучшее» означает, что сумма квадратов расстояний между соответствующими точками в каждой диаграмме рассеяния минимизируется, как и в оригинальном методе Пирсона [39]. Система из q осей в данном контексте, заданных первыми q главными компонентами, образуют подпространство главных компонент. Следовательно, метод главных компонент – это метод понижения размерности пространства, посредством которого набор из p исходных переменных может быть заменен оптимальным набором из q производных переменных – главных компонент. В случае, когда $q = 2$ или $q = 3$, графическая аппроксимация n -точечной диаграммы рассеяния также возможна и часто проводится для визуализации исходного набора данных. Важно заметить, что данный результат является адаптивным в том смысле, что «лучшим» подпространством размерности $q + 1$ для того же набора данных может быть получено добавлением еще одного столбца к тем, которые образуют «лучшее» q -мерное решение.

Качество любой q -мерной аппроксимации может быть оценено по изменчивости отобранных главных компонент. В действительности, сумма дисперсий p исходных

переменных равна сумме диагональных элементов ковариационной матрицы S . Также эта сумма равна сумме дисперсий всех p главных компонент. Отсюда, стандартным показателем качества аппроксимации пространства с помощью q главных компонент служит отношение их суммарной дисперсии к общей дисперсии: $\pi_j = \frac{\lambda_j}{\sum_{j=1}^p \lambda_j} = \frac{\lambda_j}{tr(S)}$ (6), где $tr(S)$ – это сумма диагональных элементов матрицы S (от английского «track» – след) [41]. Адаптивный характер главных компонент позволяет говорить о доле дисперсии набора S главных компонент (обычно, но необязательно, первых q главных компонент), которая измеряется в процентах от общей дисперсии: $\sum_{j \in S} \pi_j \times 100\%$.

На практике распространенным подходом является использование предварительно вычисленных долей общей объясненной дисперсии для принятия решения о том, сколько главных компонент оставить в анализе (70% – часто применяемый порог). Однако необходимость графической визуализации набора данных зачастую приводит к тому, что оставляют только первые две или три главные компоненты. Даже для таких случаев определение доли общей объясненной дисперсии является базовым критерием оценки качества графического представления данных после уменьшения размерности. В методе главных компонент акцент почти всегда идет на первые несколько главных компонент, однако в некоторых случаях последние главные компоненты также могут быть интересны, например, для определения выбросов [42].

Главные компоненты также могут быть подходящим решением для широкого круга других задач в разных областях. Оптимальные критерии для метода главных компонент подробно описаны в различных источниках (среди прочего, ознакомиться с ними можно здесь [42–44]). В то же время McCabe [45] применяет некоторые из этих критериев для формирования подвыборки из исходных переменных, так называемых, главных переменных. Такой подход представляет отдельную, вычислительно более сложную, задачу анализа данных [46].

3.2. Примеры использования метода главных компонент в генетике

Появление технологий массового секвенирования и их стремительное распространение, привело к колоссальному росту объемов информации в биологии и в генетике, в частности. Как и в других областях науки, оперирующих большими объемами данных, в генетике актуальной проблемой является снижение их размерности. Особенно остро этот вопрос стоит при работе с полногеномными данными. Как было отмечено ранее, одним из способов решения данной задачи является метод главных компонент.

Применимость метода главных компонент при проведении ПГАА была неоднократно показана в ряде работ. Так например, в исследовании 2014 года [47] описано применение данного метода при поиске ассоциаций между SNP и коррелированными признаками (с помощью симуляций было рассмотрено до 100 коррелированных признаков). Метод был применен к данным ПГАА пяти признаков, связанных со свертываемостью крови человека, и было найдено два новых SNP, ассоциации с которыми не были показаны при помощи стандартных подходов. Авторами был сделан вывод, что для описания всего набора данных комбинирование главных компонент является более оптимальным решением, чем отбор только нескольких первых из них. Кроме того, в работе был проведен сравнительный анализ метода комбинированных главных компонент с другими методами многомерного анализа данных, такими как: MANOVA (multivariate analysis of variance), порядковая логистическая регрессия (Multiphen), смешанная линейная модель множества признаков (multitrait linear mixed model, MTMM) и метод коррекции *p-value* TATES. По результатам анализа, метод комбинированных главных компонент показал мощность, близкую к оптимальной, и проявил большую гибкость и надежность, чем другие методы.

Преимущества метода комбинированных главных компонент так же были показаны в работе 2016 года, где была реализована его модификация [48] для поиска ассоциаций гена *TRAF1-C5* с ревматоидным артритом.

Несмотря на то, что метод главных компонент и ранее применялся в ПГАА, первым масштабным (размер выборки > 170,000 человек) примером его использования для работы с полногеномными данными можно считать работу 2016 года [49], посвященную изучению антропометрических признаков. В ней данный подход был применен для вычисления главных компонент, описывающих морфологию тела, как производную от шести антропометрических показателей: индекса массы тела, роста, веса, обхватов талии и бедра и их отношения. Метод главных компонент применялся к исходным антропометрическим данным 65 различных исследований. Для каждого из них были получены корреляционные матрицы, которые были использованы для вычисления комбинированной матрицы нагрузок, которая вычислялись как взвешенная сумма всех матриц, деленная на размер выборки. Из этой матрицы были получены взвешенные нагрузки главных компонент (рис. 1, *a*), с помощью которых для каждого исследования были вычислены главные компоненты. Для них был проведен ПГАА с последующим мета-анализом для всех исследуемых когорт. Данный подход показал свою эффективность, поскольку дисперсия первых четырех главных компонент составила > 99% (рис. 1, *b*) от общей объясненной

дисперсии признаков, а в результате исследования удалось выявить шесть новых локусов, для которых ассоциация с антропометрическими чертами была показана впервые.

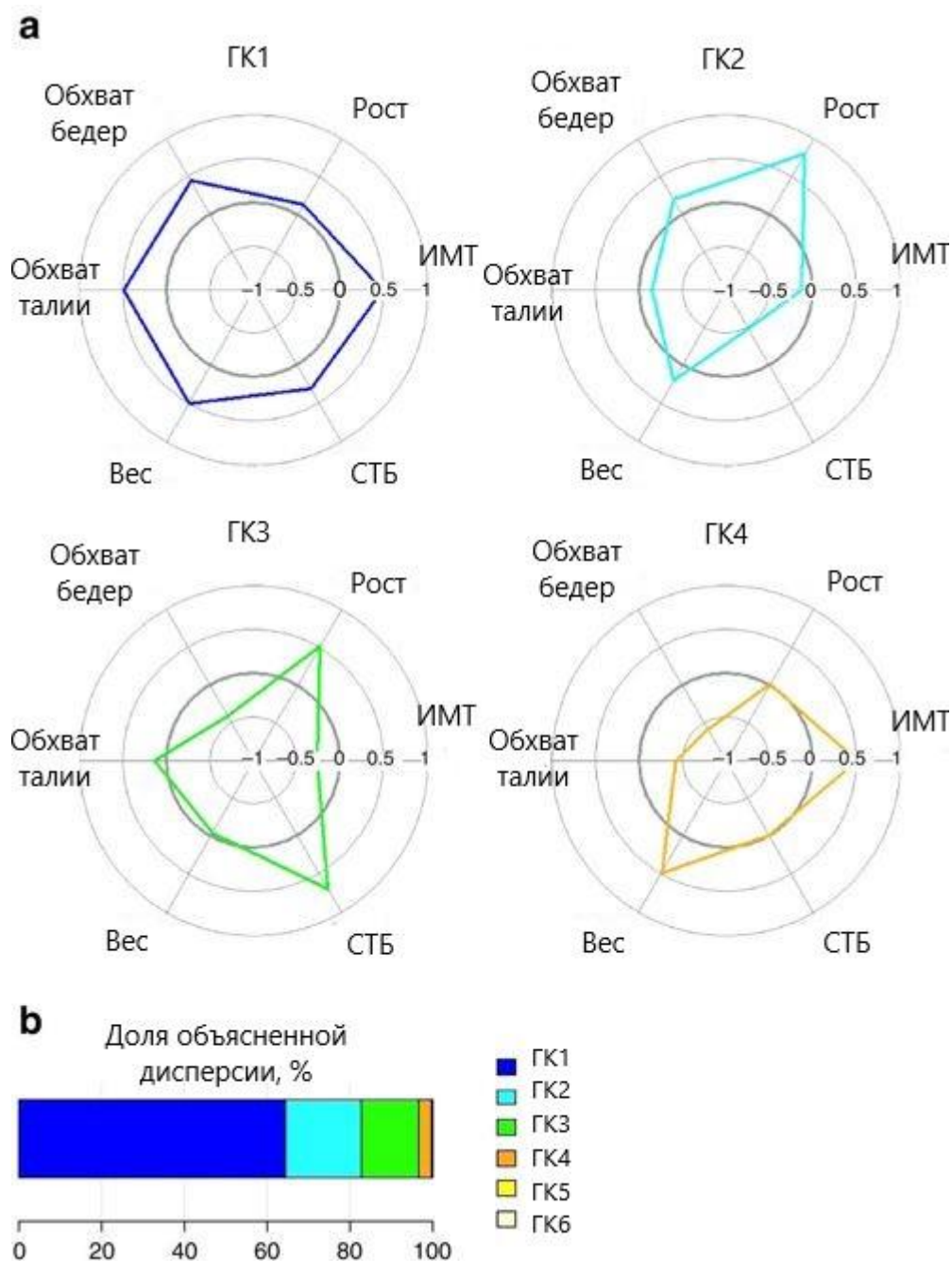


Рисунок 1 – Характеристики главных компонент для шести антропометрических признаков [49]:

a – взвешенные нагрузки главных компонент; b – доли общей объясненной дисперсии.

ГК – главная компонента со взвешенными нагрузками; ИМТ – индекс массы тела;

СТБ – отношение обхвата талии к обхвату бедра

Еще одним интересным примером использования метода главных компонент в генетике является более раннее исследование 2009 года [50], в котором метод главных компонент в сочетании с LASSO (least absolute shrinkage and selection operator) был применен для выявления ген-генных взаимодействий в рамках ПГАА ревматоидного артрита. В работе было выбрано 28 генов, расположенных на одной хромосоме с геном *HLA-DRB1*, доказано связанным с данным заболеванием. Метод главных компонент был применен для снижения размерности набора SNP в каждом из этих генов, а отобранные маркеры анализировались с помощью LASSO для определения межгенных взаимодействий. Итогом работы стало обнаружение двух статистически значимых пар ген-генных взаимодействий.

Широкую распространенность метод главных компонент получил и в области анализа экспрессии генов. Например, в 2009 году [51] данный подход был применен для определения сетей генных взаимодействий при обработке транскриптомных данных. А в недавнем исследовании 2017 года [52] была предпринята попытка решить проблему сложной интерпретации результатов метода главных компонент. С этой целью исследователями был предложен измененный метод разреженных главных компонент, и его эффективность была проверена на наборе данных об экспрессии генов. Также описано использование метода главных компонент для решения задачи кластеризации при работе с транскриптомами [53]. Однако оптимальность применения данного метода в этом контексте не была показана.

Таким образом, метод главных компонент имеет применение в геномном и транскриптомном анализе и позволяет не только снизить размерность данных, но также выявить уникальные наблюдения и выделить нечто общее среди коррелированных признаков. В совокупности это позволяет приблизиться к решению фундаментальной задачи поиска механизмов генетической регуляции признаков и к решению прикладных задач в области биомедицинских и фармакологических исследований [54].

4. Генетика скелетно-мышечных болей человека

Скелетно-мышечные боли, к которым относят расстройство височно-нижнечелюстного сустава, боль в пояснице, фибромиалгию, боль во всем теле, а также ряд других расстройств, имеют выраженную генетическую компоненту и обладают значимой наследуемостью [55]. Так, например, наследуемость расстройства височно-нижнечелюстного сустава оценивается в 3 – 12%, для боли во всем теле и фибромиалгии этот показатель достигает 54%, наследуемость боли в шее составляет 24 – 58%, а для боли в пояснице – до 68% [56]. Более того, были получены доказательства общности наследственных факторов, определяющих разные типы скелетно-мышечных болей [57]. В

данном случае речь идет о полигенной модели наследования признака. Однако стоит заметить, что науке известен ряд редких врожденных заболеваний, связанных с повышенной болевой чувствительностью или полным ее отсутствием, наследование которых описывается законами Менделевской генетики. Многие из этих состояний вызваны мутациями в генах ионных каналов (потенциал-зависимых Na^+ каналов, Ca^{2+} каналов, Na^+/K^+ -АТФазы), а также мутациями в генах различных киназ и связанных с ними белков (*IKBKAP-1*, *NTRK1*, *trkA* и др.) [58].

К настоящему времени известно множество полиморфизмов, потенциально ассоциированных с различными типами боли. Так по данным базы Human Pain Genes Database (<https://humanpaingenetics.org/hpgdb/>, дата обращения: 10.04.2020) выявлено 1,237 SNP, предположительно связанных с разными формами болевого синдрома у человека.

Для скелетно-мышечных болей счет ассоциированных генов идет на десятки. Самым известным среди них является ген *COMT* (catechol-O-methyltransferase), кодирующий катехоламин-О-метилтрансферазу – фермент, осуществляющий катаболизм катехоламинов (в том числе нейротрансмиттеров адреналина, норадреналина, дофамина). Гипофункция этого фермента приводит к повышению уровня адреналина, потенцирующего передачу болевого сигнала через β -адренергические рецепторы [55]. Это обуславливает ассоциацию *COMT* гена с расстройством височно-нижнечелюстного сустава, болью в пояснице, фибромиалгией, болью во всем теле [59], хроническими болями [5] и другими заболеваниями. Ген *COMT* является высоко полиморфным и содержит множество функциональных SNP. Наиболее изученной является несинонимичная замена Вал158Мет (rs4680), снижающая активность фермента [5,8] и ассоциированная со многими болевыми синдромами, аффективными расстройствами и нарушениями когнитивной деятельности [55]. В рамках более обширных исследований функциональный локус данного гена был сведен к трем основным гаплотипам, которые влияют на экспрессию и активность фермента, меняя риск развития острой болевой чувствительности и хронической боли [5]. Помимо *COMT*, ассоциация с фибромиалгией и болью во всем теле была показана для гена *MAOA*, кодирующего моноаминоксидазу-А [55], выполняющую схожую с катехоламин-О-метилтрансферазой функцию, а также гена *ADRB2*, кодирующего β_2 -адренергический рецептор (функциональными являются несинонимичные мутации Арг16Гли (rs1042713) и Глн27Глу (rs1042714)). Также установлена связь *ADRB2* с хронической болью и связанными с ней состояниями [8].

Известно, что гены *SLC6A4* и *HTR2A*, кодирующие белок-транспортер серотонина (5HT transporter, 5HTT) и рецептор серотонина (hydroxytryptamine receptor 2A, 5HT-2A),

соответственно, ассоциированы со скелетно-мышечными болями человека [58]. Мутации в этих генах влияют на серотонинергическую передачу, а значит, на тормозную модуляцию болевых сигналов. Так, индел длиной 44 пн в промоторе гена *SLC6A4* ассоциирован с хронической болью и сопутствующими ей психическими состояниями – депрессией, тревожным расстройством и т. д. Кроме того, короткий аллель данного гена связан с повышенным риском развития фибромиалгии, а длинный аллель ассоциирован с расстройством височно-нижнечелюстного сустава. Синонимическая замена S34S (rs6313, 102T > Ц) в гене *HTR2A* также ассоциирована с повышенным риском развития обоих заболеваний [8].

Множество противоречивых данных свидетельствует о связи полиморфизмов в генах опиоидных рецепторов со скелетно-мышечными болями человека. В частности, в гене μ -опиоидного рецептора *OPRM1* обнаружена несинонимичная замена N40D (118A > Г, rs1799971), влияющая на афинность связывания, а также, возможно функциональный, полиморфизм rs563649. Согласно одним источникам, связь этих вариантов с болью во всем теле не была подтверждена в ряде популяционных исследований [59]. В то же время, есть данные в пользу ассоциации полиморфизма rs1799971 с восстановлением пациентов, страдающих радикулопатией пояснично-крестцового отдела [60]. Описано снижение чувствительности к опиоидам у гомозигот по Г аллелю полиморфизма rs1799971 и высказано предположение о совместном влиянии генов *OPRM1* и *COMT* на этот показатель [58]. Есть также информация об ассоциации гена *OPRD1*, кодирующего δ -опиоидный рецептор, с ответом на болевые стимулы (полиморфизмы rs2234918, rs419335 и rs533123) [61].

На вовлеченность иммунной системы в хронический болевой процесс могут указывать данные по генам *HLA-B58*, *HLA-DR4*, *HLA-DR5* и *HLA-DR8*, однако их ассоциация с этим заболеванием не была подтверждена в более поздних исследованиях [8]. Также о роли иммунного ответа свидетельствует ассоциация с геном *SERPINA1*, кодирующем α 1-антитрипсин: сообщается, что Z аллель этого гена чаще встречается у больных фибромиалгией [8]. В пользу этой гипотезы выступают и данные о том, что гены, кодирующие интерлейкин 1 (*IL-1 α*), агонист рецептора к интерлейкину 1 (*IL-1RN*) и интерлейкин 6 (*IL-6*), связаны с риском радикулопатии пояснично-крестцового отдела [60]. Помимо перечисленных генов в литературе по генетике скелетно-мышечных болей человека часто упоминаются гены *CCT5* (chaperonincontaining-TCP1-complex-5 gene) и *FAM173B* [55,62,63], поскольку в локусе этих генов расположен вариант rs13361160, аллель Ц которого повышает риск боли во всем теле на 30% [64]. Также часто встречается ген

GCH1, кодирующий ГТФ-циклогидролазу 1 – фермент, вовлеченный в синтез нейротрансмиттеров и, соответственно, влияющий на ноцицептивную чувствительность. Неоднозначные данные приводятся по гаплотипу данного гена, снижающему выработку фермента и уменьшающему болевую чувствительность. Так, было показано, что данный гаплотип более распространен среди пациентов с болью во всем теле [59] и не связан с этим заболеванием [20], однако он же ассоциирован с пониженным риском боли в пояснице [58]. Наконец, нередко в источниках упоминается ген *ESR1*, кодирующий эстрогеновый рецептор. Для него показана ассоциация с расстройством височно-нижнечелюстного сустава [55]. Интересно также, что в литературе описаны ген-генное взаимодействие между *COMT* и *OPRM1*, сложное взаимодействие ген-пол-среда для гена аргинин-вазопрессин рецептора 1A (*AVPR1A*) и влияние микроРНК и модификаций гистонов на ноцицепцию [65].

Большинство из перечисленных генов были определены как ассоциированные со скелетно-мышечными болями человека благодаря ген-кандидатным исследованиям и на данный момент требуют подтверждения с помощью ПГАА. Далее мы осветим наиболее актуальные сведения, полученные из результатов последних ПГАА для скелетно-мышечных болей человека.

В работе 2017 года [66] был проведен ПГАА для расстройства височно-нижнечелюстного сустава. В ходе этого исследования была обнаружена и реплицирована (на выборке лиц женского пола) ассоциация локуса на 13 хромосоме (ведущий SNP rs60249166, odds ratio [OR] = 0.65, $p\text{-value} = 3.6 \times 10^{-8}$), расположенного в 5'-нетранслируемой области гена *RXP2* (relaxin/insulin-like family peptide receptor 2). Так же была показана и подтверждена (на выборке из обоих полов) ассоциация локуса, расположенного на хромосоме 17 (rs1531554, OR = 0.68, $p\text{-value} = 2.9 \times 10^{-8}$).

В рамках крупнейшего ПГАА для боли в спине, опубликованного нашей командой в 2019 году [67], было обнаружено и реплицировано три ассоциированных локуса: rs12310519 ($p\text{-value} = 5.00 \times 10^{-5}$), rs7814941 ($p\text{-value} = 5.32 \times 10^{-5}$) и rs3180 ($p\text{-value} = 6.59 \times 10^{-3}$). Двое из перечисленных локусов уже были описаны ранее, как ассоциированные с различными болевыми расстройствами. В частности, локус на 12 хромосоме с ведущим SNP rs12310519 располагается в интроне гена *SOX5*, связь которого с болью в спине была показана в более раннем исследовании консорциума CHARGE и PainOMICS [68]. Локус, расположенный на 8 хромосоме (ведущий SNP rs7814941), находится в участке между генами *GSDMC* и *CCDC26*. Для него также была описана ассоциация с болью в спине в исследовании консорциума CHARGE и PainOMICS и, кроме того, показана его связь с ишиасом [69].

Новым в данном исследовании стал локус на 10 хромосоме (rs3180 SNP), расположенный в 3'-нетранслируемой области генов *SPOCK2* и *CHST3*. Этот локус ранее был описан как ассоциированный с дегенерацией межпозвонковых дисков, с функциональным вариантом SNP rs4148941, влияющим на экспрессию гена *CHST3* в позвоночных дисках [70]. Данный ген кодирует фермент, катализирующий сульфирование хондроитина. Другим геном в этом локусе является *SPOCK2*, кодирует протеогликан SPARC/Остеонектин (Cwcv And Kazal Like Domains Proteoglycan 2), вовлеченный в формирование внеклеточного матрикса и активно экспрессирующийся в ЦНС. Помимо приоритизации перечисленных генов в данной работе была проведен анализ плейотропных эффектов наследственных факторов, лежащих в основе возникновения боли в спине. Было показано наличие плейотропии по отношению к таким признакам, как рост и проблемы межпозвонковых дисков. Установлены генетические корреляции с депрессией, невротизмом, нарушениями сна, ожирением и курением. Анализ представленности групп генов показал обогащение терминами, относящимися к ЦНС и развитию скелетно-мышечной системы. Результаты этой работы свидетельствуют о наличии двух основных составляющих, определяющих развитие боли в спине – структурно-морфологической компоненты, связанной с антрометрическими признаками и проблемами межпозвонковых дисков, и психической компоненты, влияющей на ноцицепцию.

В недавней работе 2019 года был проведен ПГАА для хронической боли разной локализации [71] на выборке UK Biobank. Фенотип определялся как число сайтов хронической боли, которое измерялось от 0 до 7. Пациенты, сообщившие о боли во всем теле, были исключены из анализа, поскольку этот фенотип не относится к локализованной хронической боли и является отдельным состоянием. В ходе этого исследования было выявлено 76 независимых сигналов, расположенных в 39 локусах, ассоциированных с хронической болью разной локализации, а наследуемость признака была оценена в 10.2%. Было приоритизировано большое количество генов, среди которых ген *DCC* (Deleted in Colorectal Cancer a.k.a. DCC netrin 1 receptor), кодирующий рецептор нетрина-1, играющего важную роль в развитии нервной системы; ген *MAML3* (Mastermind-Like Transcriptional coactivator 3), кодирующий основной элемент сигнального пути Notch, регулирующего развитие клеток разных типов тканей у метазоев, и ген *NUMB* (Endocytic Adaptor Protein), кодирующего белок, ингибирующий Notch-сигнальный путь и влияющий на дифференциацию нейронов; ген *SOX6*, потенциально вовлеченный в развитие скелетных мышц, поддержание нервных стволовых клеток и развитие вставочных нейронов коры головного мозга, а также ассоциированный с минеральной плотностью костей; ген *FOXP2*,

кодирующий транскрипционный фактор из семейства FOX, регулирующих экспрессию сотен генов на разных этапах развития организма, в том числе генов, экспрессирующихся в мозге, предполагается, что этот транскрипционный фактор может влиять на развитие мозга, нейрогенез, синаптическую пластичность и передачу сигнала; наконец, гены *GABRB2* и *GABRB3* (кодирующий одну из трех β -субъединиц рецептора ГАМК), для которых показаны ассоциации с шизофренией и фибромиалгией, соответственно. Для ассоциированных генов выявлено обогащение терминами, относящимися к нейрогенезу, синаптической пластичности, развитию ЦНС и регуляции клеточного цикла. Были обнаружены генетические корреляции данного признака с различными психиатрическими, аутоиммунными и антропометрическими признаками, в частности, с большим депрессивным расстройством, астмой и индексом массы тела. Кроме того, исследование показало, что полигенная модель риска хронической боли разной локализации является хорошей предсказательной моделью возникновения боли во всем теле, что говорит о наличии общих наследственных факторов, определяющих развитие обоих состояний. Данная работа также свидетельствует в пользу значительной роли нервной системы в возникновении хронической боли. Стоит отметить, что в данной работе не было проведено репликации найденных локусов на независимой выборке.

Подводя итог всему выше изложенному, можно заключить, что для различных типов скелетно-мышечных болей человека показаны ассоциации с генами, вовлеченными в пути передачи нервного импульса, дифференциацией и развитием нервной системы, генами, вовлеченными в иммунный ответ, а также генами, имеющими отношения к морфологии и анатомии опорно-двигательного аппарата.

5. Аналитическое заключение

Хронические скелетно-мышечные боли – это группа распространенных заболеваний, формирующихся под воздействием биологических, психических и социальных факторов. Большинство заболеваний из этой группы имеют полигенную модель наследования, отличаются высокой наследуемостью и имеют общую генетическую природу. Сложная этиология, многообразие механизмов и проявлений хронической боли затрудняют разработку подходов к ее лечению. Это делает необходимым исследование основ хронической боли с целью поиска новых мишеней для терапевтических средств, однако сложность и высокая гетерогенность признака осложняет его изучение. В частности, эти факторы приводят к снижению мощности анализа и вносят шум. Следствием этого является то, что к настоящему моменту известно небольшое количество подтвержденных ассоциаций между генами и разными типами хронических скелетно-мышечных болей. Все

это указывает на необходимость применения особых подходов к изучению хронической боли. Метод главных компонент рассматривается как один из самых распространенных подходов, позволяющих увеличить мощность анализа и снизить гетерогенности признака, путем рассмотрения группы схожих признаков совместно и поиска их общих генетических факторов.

Материалы и методы

Материалы

В данной работе использовались результаты ПГАА, проведенного нашим коллегой Фрейдиным М. Б. (Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences, King's College London, Лондон, Великобритания). Ниже описана методика получения данных:

Исследуемая выборка и фенотипирование

Выборка для данного исследования была сформирована на основе данных UK Biobank [72] (более подробную информацию о социально-демографических и физических характеристиках, а также об особенностях образа жизни и показателях здоровья участников когорты можно найти в источнике [73]). В выборку вошли люди в возрасте от 40 до 69 лет, менее склонные к ожирению, курению, употреблению алкоголя и имеющие меньше проблем со здоровьем по сравнению с населением в целом. Все участники проекта дали письменное информированное согласие на обработку персональных данных, исследование было одобрено North West Multi-centre for Research Ethics Committee (11/NW/0382).

Разрешение на проведение данного исследования было дано исследовательской группой UK Biobank в рамках проекта №18219 «Genetic and epidemiological analyses of low back pain».

Фенотипирование участников исследования проводилось методом анкетирования. Прежде всего, испытуемым предлагалось ответить на вопрос о том, какой(ие) из типов боли они испытывали в последние месяцы. Далее необходимо было уточнить, продолжалась ли эта боль в течение трех и более месяцев подряд. Люди, сообщившие о боли в спине, колене, шее (или плечевом отделе) или о боли в бедре, продолжающейся три и более месяца подряд, были отнесены к случаям (cases) хронической боли в спине, колене, шее и бедре, соответственно. Участники, сообщившие о нескольких типах боли, длившейся три и более месяца подряд, учитывались как случаи каждого заявленного типа хронической боли по отдельности. Индивиды, не отметившие ни один из типов боли сохранявшимся на протяжении трех и более последовательных месяцев, были классифицированы как контроли (controls). Люди, воздержавшиеся от ответа или сообщившие о боли во всем теле, продолжающейся три и более месяца подряд, были исключены из анализа.

В исследовании приняло участие 456,580 людей (табл. 1) с импутированными (то есть восстановленными с использованием референсной выборки) генотипами, среди которых 265,000 людей европейского происхождения (этническая принадлежность была

установлена методом главных компонент). Решение о включении в поисковую выборку людей только с европейским происхождением обусловлено их большей представленности в когорте UK Biobank. Репликационная выборка (191,580 человек) представлена людьми африканского происхождения (7,541 человек), людьми южно-азиатского происхождения (уроженцы Индии, Пакистана и Бангладеш, 9,208 человек) и другими людьми европейского происхождения (174,831 человек). Каждая из подвыборок анализировалась независимо.

Таблица 1 – Характеристики выборки

	Частота встречаемости в выборке	Размер выборки	Возраст (среднее \pm SD), годы	ИМТ (среднее \pm SD), кг/м ²	Доля женщин, %
Поисковая выборка (N = 265,000)					
Хроническая боль в спине	17.9%	Случаи (N = 47,507)	57.65 (7.99)	28.33 (5.18)	53.88
		Контроли (N = 217,493)	57.26 (8.03)	27.15 (4.61)	54.32
Хроническая боль в шее	16.3%	Случаи (N = 43,287)	57.73 (7.79)	27.90 (5.02)	53.84
		Контроли (N = 221,713)	57.25 (8.07)	27.25 (4.68)	54.32
Хроническая боль в бедре	9.2%	Случаи (N = 24,300)	59.15 (7.44)	28.91 (5.40)	54.35
		Контроли (N = 240,700)	57.15 (8.06)	27.20 (4.64)	54.23
Хроническая боль в колене	17.5%	Случаи (N = 46,292)	58.61 (7.59)	29.18 (5.37)	54.12
		Контроли (N = 218,708)	57.06 (8.09)	26.97 (4.50)	54.27
Репликационная выборка (N = 191,580)					
Выборка людей африканского происхождения (N = 7,541)					
Хроническая боль в спине	21.0%	Случаи (N = 1,586)	53.77 (8.24)	30.62 (5.79)	54.50
		Контроли (N = 5,955)	52.04 (8.00)	29.27 (5.13)	54.19
Хроническая боль в шее	16.1%	Случаи (N = 1,217)	54.38 (7.98)	30.06 (5.52)	54.353
		Контроли (N = 6,324)	52.02 (8.04)	29.45 (5.25)	54.24
Хроническая боль в бедре	8.5%	Случаи (N = 641)	55.00 (7.91)	31.30 (6.14)	54.37
		Контроли (N = 6,900)	52.16 (8.05)	29.39 (5.19)	54.25
Хроническая боль в колене	20.4%	Случаи (N = 1,539)	54.67 (8.30)	31.64 (6.11)	54.49
		Контроли (N = 6,002)	51.82 (7.92)	29.01 (4.93)	54.20
Выборка людей европейского происхождения (N = 174,831)					
Хроническая боль в спине	18.0%	Случаи (N = 31,428)	57.62 (7.96)	28.36 (5.22)	54.05
		Контроли (N = 143,403)	57.26 (8.02)	27.14 (4.58)	54.28
Хроническая боль в шее	16.3%	Случаи (N = 28,482)	57.82 (7.76)	27.92 (5.02)	54.27
		Контроли (N = 146,349)	57.23 (8.06)	27.25 (4.66)	54.24
Хроническая боль в бедре	9.2%	Случаи (N = 16,022)	59.26 (7.40)	28.86 (5.41)	54.61
		Контроли (N = 158,809)	57.13 (8.05)	27.21 (4.63)	54.20

Продолжение таблицы 1

	Частота встречаемости в выборке	Размер выборки	Возраст (среднее \pm SD), годы	ИМТ (среднее \pm SD), кг/м ²	Доля женщин, %
Репликационная выборка (N = 191,580)					
Выборка людей европейского происхождения (N = 174,831)					
Хроническая боль в колене	17.3%	Случаи (N = 30,173)	58.71 (7.54)	29.24 (5.41)	54.27
		Контроли (N = 144,658)	57.04 (8.08)	26.97 (4.47)	54.23
Выборка людей южно-азиатского происхождения (N = 9,208)					
Хроническая боль в спине	21.6%	Случаи (N = 1,993)	54.66 (8.51)	27.76 (4.58)	54.29
		Контроли (N = 7,215)	53.87 (8.47)	26.92 (4.23)	54.22
Хроническая боль в шее	20.2%	Случаи (N = 1,864)	54.65 (8.24)	27.43 (4.56)	54.31
		Контроли (N = 7,344)	53.88 (8.53)	27.01 (4.25)	54.22
Хроническая боль в бедре	6.6%	Случаи (N = 610)	56.61 (8.21)	28.30 (4.90)	54.07
		Контроли (N = 8,598)	53.86 (8.47)	27.01 (4.26)	54.25
Хроническая боль в колене	20.1%	Случаи (N = 1,850)	55.97 (8.23)	28.52 (4.86)	54.10
		Контроли (N = 7,358)	53.55 (8.47)	26.74 (4.09)	54.27

ИМТ – индекс массы тела; N – размер выборки; SD (standard deviation) – стандартное отклонение

Генотипирование и импутация геномных данных

Генотипирование и импутация были осуществлены на данных UK Biobank от марта 2018 года. Генотипирование индивидов было проведено с применением чипов Affymetrix UK BiLEVE и Affymetrix UK Biobank Axiom. Импутация реализована в программе IMPUTE4 (<https://jmarchini.org/impute-4/>) [74] с использованием данных Haplotype Reference Consortium (HRC) [29] и объединенных референсных панелей третьей фазы проекта «1000 Геномов» и UK10K (подробную информацию о выделении и подсчете ДНК [75], контроле качества генотипирования, структуре популяций, фазировании и импутации геномных данных можно найти здесь [76]). Общее количество SNP после проведения импутации и контроля качества составило порядка 90 миллионов.

Полногеномный анализ ассоциаций

ПГАА был проведен с использованием статистического программного обеспечения BOLT-LMM v.2.3.2 [77]. Линейная модель со смешанными эффектами была применена для определения аддитивных эффектов SNP на фенотипы интереса с поправками на пол, возраст, систематические экспериментальные ошибки генотипирования (бэтч-эффекты) и на первые десять главных компонент матрицы родства. В ходе анализа были применены следующие фильтры: частота минорного аллеля $> 2.0e-04$ для европейских популяций и $> 5.0e-03$ для африканских и южно-азиатских популяций; качество импутации > 0.7 ; процент генотипирования по SNP и по индивиду > 0.98 . В анализ были включены только

биаллельные SNP и инделсы. Для запуска BOLT-LMM использовались известные показатели неравновесия по сцеплению (LD Score) для европейской популяции. Оценка LD Score для исследуемых SNP для африканской и южно-азиатской популяций проводилась при помощи метода LD Score регрессии [36] на 500 случайным образом выбранных индивидах из каждой популяции, соответственно.

Результаты ПГАА были скорректированы на остаточную инфляцию с помощью интерсепта (фактора инфляции тестовой статистики) LD Score регрессии [36].

Методы

Общая схема исследования представлена на рисунке 2.

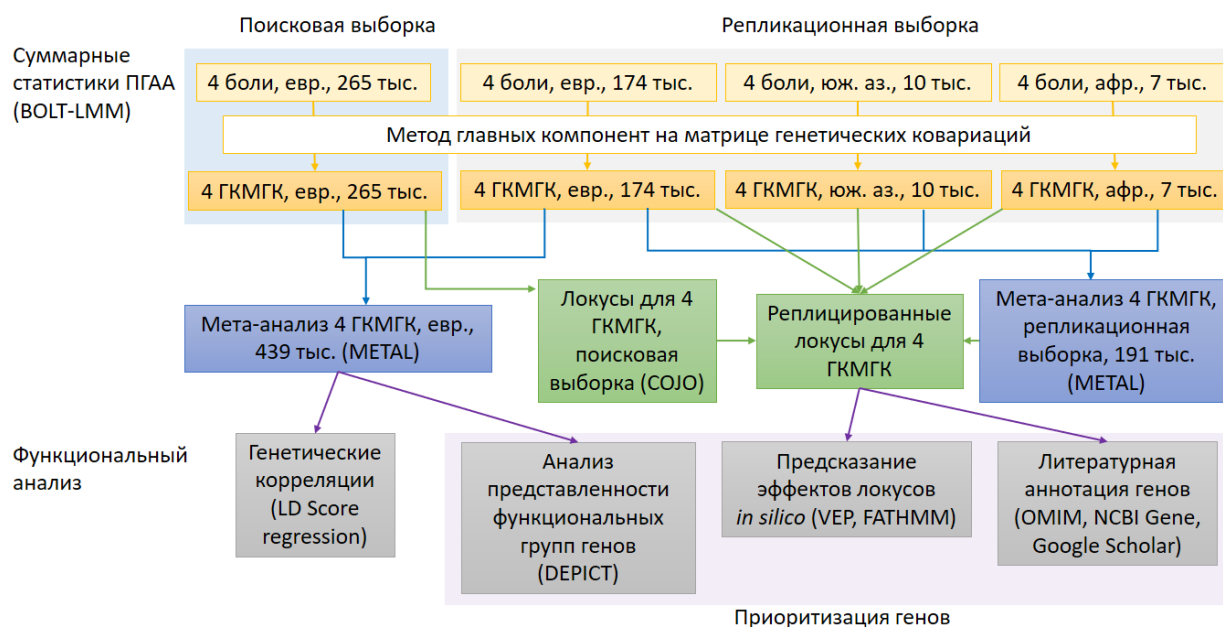


Рисунок 2 – Дизайн эксперимента

евр. – выборка людей европейского происхождения; юж. аз. – выборка людей южно-азиатского происхождения; афр. – выборка людей африканского происхождения; тыс. – тысяч (человек)

Главные компоненты матрицы генетических ковариаций

В данном исследовании для выделения общей генетической компоненты четырех хронических скелетно-мышечных болей человека нами был применен модифицированный метод главных компонент. Метод главных компонент – это метод статистического анализа, который использует ортогональную трансформацию для преобразования набора потенциально коррелированных переменных в набор линейно независимых переменных –

главных компонент. Каждая главная компонента является линейной комбинацией исходных переменных. Первая главная компонента вычисляется таким образом, что она объясняет наибольшую долю дисперсии, то есть имеет максимально возможную дисперсию. Вычисление каждой последующей главной компоненты проводится так, чтобы она объясняла наибольшую долю остаточной дисперсии и была ортогональна уже найденным главным компонентам. Вектора, полученные из коэффициентов ортогональной трансформации (a_i), образуют ортогональный базис пространства. Метод главных компонент чувствителен к шкалированию исходных переменных. В случае положительно определенной матрицы ковариаций исходных переменных (Σ), a_i – это собственные вектора матрицы Σ . Каждое соответствующее собственное число пропорционально доле дисперсии, объясненной этой главной компонентой.

Для того чтобы разложить признаки интереса на генетически независимые компоненты – выделить a_i – мы использовали матрицу генетических ковариаций Ω (вместо матрицы фенотипических ковариаций, традиционно используемую при анализе биологических признаков). Полученные главные компоненты мы назвали ГКМГК – главными компонентами матрицы генетических ковариаций.

ГКМГК характеризуются следующими особенностями:

- каждая ГКМГК является линейной комбинацией исходных признаков, по этой причине соответствующие результаты ПГАА могут быть получены для всех ГКМГК и каждая ГКМГК может быть проанализирована как самостоятельный признак при помощи *in silico* методов;
- ГКМГК генетически независимы: парные генетические корреляции между ними равны нулю (фенотипические корреляции исходных признаков могут быть ненулевыми);
- доверительные интервалы для каждой a_i могут быть оценены с помощью оценки стандартных ошибок матрицы генетических ковариаций.

Ниже приведены детали реализации метода главных компонент на матрице генетических ковариаций.

Переменные:

Ω – матрица генетических ковариаций (квадратная матрица размера $m \times m$, где m – количество исходных признаков)

Ω_{SE} – матрица стандартных ошибок генетических ковариаций (размер $m \times m$)

Σ – матрица фенотипических ковариаций (размер $m \times m$); в случае стандартизованных данных она эквивалентна матрице фенотипических корреляций

B – матрица размеров эффектов (β) для m фенотипов (размер $M \times m$, где M – количество SNP в анализе); b_i – это i -ый столбец матрицы B

SE – матрица стандартных ошибок β для m фенотипов ($M \times m$). SE_i – это i -ый столбец матрицы SE

$varY_i$ – дисперсия i -го признака (после стандартизации $varY_i = 1$)

SD_i – стандартное отклонение i -го признака ($SD_i = \sqrt{varY_i}$)

B_s – матрица стандартизованных β для m фенотипов ($M \times m$), b_i^s – это i -ый столбец матрицы B_s

SE_s – матрица стандартизованных стандартных ошибок для m фенотипов ($M \times m$), SE_i^s – это i -ый столбец матрицы SE_s

GPC_i – i -ая ГКМГК

A – матрица собственных векторов матрицы Ω ($m \times m$), каждый столбец – это a_i , а $\{a_1 \dots a_m\}$ – это вектор коэффициентов ортогональной трансформации исходных m признаков в m ГКМГК $\{GPC_1 \dots GPC_m\}$

A_s – матрица шкалированных собственных векторов матрицы Ω ($m \times m$), каждый столбец – это a_i^s , $\{a_1^s \dots a_m^s\}$ – это вектор коэффициентов ортогональной трансформации исходных m признаков в m шкалированных ГКМГК с дисперсией равной 1

L – вектор собственных чисел $\{l_1 \dots l_m\}$

Для расчета ГКМГК четырех типов скелетно-мышечных болей были реализованы следующие шаги:

1. Вычисление Ω и Ω_{SE} с помощью метода LD Score регрессии (<https://github.com/bulik/ldsc/>) (см. далее)*.
2. Вычисление $varY_i$ и матрицы корреляций Пирсона для четырех фенотипов хронической боли.
3. Стандартизация суммарных статистик ПГАА четырех типов боли ($\beta_i^s = \beta_i / SD_i$ и $SE_i^s = SE_i / SD_i$).
4. Проверка собственных чисел матрицы Ω на положительность.
5. Вычисление вектора собственных чисел (L) и матрицы собственных векторов (A) для матрицы Ω .
6. Проверка знаков коэффициентов хронической боли в спине в собственных векторах и смена знаков всех коэффициентов в соответствующем собственном векторе на противоположные ($a_i = -a_i$) в случае, если $a_{i,back\ pain} < 0$

7. Расчет дисперсии ГКМГК по формуле $var(GPC_i) = \sum[(a_i \otimes a_i) \circ \Sigma]$, где \otimes – внешнее произведение векторов.
8. Шкалирование коэффициентов ГКМГК $a_i^s = a_i / SD(GPC_i)$.
9. Оценка 95%-го доверительного интервала для a_i (см. далее) **.
10. Получение суммарных статистик ПГАА для каждой ГКМГК (см. далее)***.

*Вычисление парных генетических ковариаций между признаками по методу LD Score регрессии [37] проводится по формуле $E[z_{1j} * z_{2j}] = \frac{\sqrt{N_1 * N_2} * \rho_g}{M} * l_{gj} + \frac{\rho * N_s}{\sqrt{N_1 * N_2}}$, где z_{1j} и z_{2j} – z-статистика j-го SNP ($z_{ij} = \frac{\beta_{ij}}{se_{ij}}$, где β_{ij} – эффект j-го SNP на i-ый признак, se_{ij} – стандартная ошибка β_{ij}) для первого и второго признаков соответственно; N_1 и N_2 – размер выборки для первого и второго признака соответственно; ρ_g – генетическая ковариация между парой признаков; M – количество SNP; l_{gj} – показатель неравновесия по сцеплению (LD Score), который представляет собой сумму квадратов корреляций Пирсона j-го SNP с другими SNP: $l_{gj} = \sum_{k=1}^M r_{jk}^2$, где $r_{jk} = E[X_j * X_k]$ для нормированных и центрированных генотипов (X_i – вектор генотипов i-го SNP); ρ – фенотипическая корреляция между признаками; N_s – количество людей, попавших в выборки обоих признаков. Параметры ρ и N_s могут быть не известны (при отсутствии индивидуальных данных), однако это не влияет на оценку ρ_g . Вектор значений l_g вычисляются по данным референсной выборки (в данном случае проект «1000 Геномов»), M также определяется референсом. Остальные переменные, за исключением ρ_g , известны из данных по первому и второму признакам. Таким образом, чтобы найти искомое значение ρ_g для пары признаков, необходимо провести регрессию $z_1 * z_2$ (поэлементное перемножение векторов z-статистик) на l_g и полученное значение умножить на $\frac{M}{\sqrt{N_1 * N_2}}$. Для оценки стандартной ошибки ρ_g используется метод «складного ножа» (jackknife). *Примечание:* метод LD Score регрессии также позволяет оценить наследуемости [36] признаков по формуле $E[z^2] = \frac{N * h^2}{M} * l_g + N * a + 1$, где z^2 – вектор $M \times 1$ квадратов z-статистик для некоторого признака, N – размер выборки, h^2 – наследуемость признака, M – количество SNP, l_g – вектор $M \times 1$ значений LD Score, a – показатель стратификации выборки. Зная ρ_g для пары признаков и их наследуемости h_1^2 и h_2^2 , можно оценить генетическую корреляцию между признаками по формуле $r_{g12} = \frac{\rho_g}{\sqrt{h_1^2 * h_2^2}}$.

**Для оценки 95%-го доверительного интервала для ГКМГК был применен метод Монте-Карло. Была проведена 1000 раундов симуляции. На каждом этапе симулировалась шумовая компонента для матрицы генетических ковариаций Ω – так называемая Ω^{noise} ($m \times m$). Каждый i,j -ый элемент ($i > j$) матрицы Ω^{noise} выбирался из нормального распределения со средним значением равным нулю и стандартным отклонением равным i,j -ому элементу матрицы стандартных ошибок ($\Omega_{i,j}^{SE}$). Результирующая ковариационная матрица была получена как сумма Ω и Ω^{noise} . Стандартизованные собственные числа были вычислены по описанному выше методу, формируя эмпирическое распределение для элементов матрицы A_s . Для каждого элемента матрицы A_s , 95%-ый доверительный интервал был определен как абсолютная разность значений 0.975 и 0.025 квантилей, деленная пополам.

***Суммарные статистики ПГАА для ГКМГК были получены следующим образом:

1. Размеры эффектов M SNP были вычислены по формуле $\beta_{GPC_i} = B_s \times a_i^s$, где \times – это внутреннее произведение.
2. Дисперсии ГКМГК были рассчитаны как $varGPC_i = \sum[(a_i \otimes a_i) \circ \Sigma]$. Матрица фенотипических корреляций Σ была получена на исследуемой популяции.
3. Стандартные ошибки эффектов M SNP были вычислены по формуле

$$SE_{GPC_i} = \sqrt{varGPC_i * \left(SE_1^{s^2} + \frac{b_1^{s^2}}{N} \right) - \frac{\beta_{GPC_i}^2}{N}}, \text{ где } N \text{ – это размер выборки.}$$

4. Шкалирование размеров эффектов и стандартных ошибок было проведено по ранее описанному алгоритму.
5. Соответствующие значения p -value были оценены с помощью Вальд-теста:

$$(Z - score = \beta_{GPC_i} / SE_{GPC_i}).$$

Общая генетическая дисперсия, объясненная каждой ГКМГК была вычислена по формуле:

$$R_{GPC_i}^2 = l_i / \sum_{i=1}^m l_i.$$

Наследуемость каждой ГКМГК определена как $h_{GPC_i}^2 = \frac{\sum[(a_i \otimes a_i) \circ \Omega]}{\sum[(a_i \otimes a_i) \circ \Sigma]}$.

Генетические корреляции между ГКМГК и исходными признаками вычислены в соответствии с формулой:

$$r_g\{c_1 | c_2\} = \frac{\sum[(c_1 \otimes c_2) \circ \Omega]}{\sqrt{\sum[(c_1 \otimes c_1) \circ \Omega] \times \sum[(c_2 \otimes c_2) \circ \Omega]}}$$

где c_1 и c_2 – линейные коэффициенты для признака 1 и 2 соответственно (c_1 и c_2 , в случае GPC_i , определяется как $c_j = a_i^s$).

Вклад каждой ГКМГК в генетическую составляющую исходных признаков (объясненная генетическая дисперсия) был оценен как квадрат коэффициента генетической корреляции ГКМГК с исходным признаком.

Мета-анализ

Суммарные статистики, полученные в результате ПГАА ГКМГК трех репликационных когорт, были объединены в ходе мета-анализа. То есть ГКМГК1 африканцев, ГКМГК1 азиатов и ГКМГК1 европейцев из репликационной выборки мета-анализировались вместе. То же было повторено и для остальных ГКМГК. Мета-анализ проводился с применением метода обратной взвешенной дисперсии (модель фиксированных эффектов) в программе METAL [78].

Также был проведен мета-анализ ГКМГК двух европейских когорт: из поисковой и репликационной выборок (суммарный размер выборки составил 439,831 человек). Результаты данного мета-анализа были использованы в дальнейшем для проведения функционального анализа.

База данных GWAS-MAP

GWAS-MAP представляет собой базу данных, хранящую суммарные статистики ПГАА на 673 сложных признака когорты UK Biobank, 123 метаболомных признака, 1,206 циркулирующих белка, 41 цитокинов и факторов роста, 190 белков плазмы крови и IgG N-гликаны, воспалительные заболевания кишечника (включая болезнь Крона) и 8 признаков, связанных с ишемической болезнью сердца, инфарктом миокарда и факторами, ассоциированными с ними. Суммарные статистики признаков UK Biobank были предоставлены Neale Lab (<http://www.nealelab.is/>) и Gene ATLAS (<http://geneatlas.roslin.ed.ac.uk/>) [79]. В ходе исследования база также была пополнена суммарными статистиками ПГАА 12 признаков, связанных с хроническими скелетно-мышечными болями человека.

В базу данных GWAS-MAP также интегрировано программное обеспечение для проведения контроля качества геномных данных, LD Score регрессии [36], парной менделевской рандомизации (MR-Base package [80]) и SMR/HEIDI анализа [34]. Данные инструменты, за исключением парной менделевской рандомизации (2-sample Mendelian randomization) и SMR/HEIDI, использовались нами при работе над проектом в рамках GWAS-MAP.

Поиск локусов

Локусы, ассоциированные с ГКМГК, были определены как участки генома в районе наиболее значимых (ведущих) SNP ($p\text{-value} < 5.0e-08$), найденных на поисковой выборке, с

окном ± 250 кб. Затем была проведена их репликация с использованием данных репликационной выборки. Критериями репликации являлись: 1) статистическая значимость ассоциации на данных мета-анализа для репликационной выборки с учетом поправки Бонферрони на множественное сравнение; 2) совпадение направления эффектов SNP для поисковой выборки и данных мета-анализа репликационных когорт.

Условный и совместный анализ COJO

Условный и совместный анализ SNP (Conditional and joint analysis, COJO) [32], ассоциированных с ГКМГК, проводился с помощью программного обеспечения GCTA [81]. Матрица неравновесия по сцеплению была вычислена при помощи PLINK 1.9 (<https://www.cog-genomics.org/plink2>) на данных 100,000 человек, случайным образом отобранных из поисковой выборки. В случае, если ни один из сигналов в отдельном локусе, за исключением ведущего SNP, не преодолевал порог статистической значимости $p\text{-value} < 5.0\text{e-}08$ полагалось, что данный локус содержит только один сигнал.

Анализ представленности функциональных групп генов DEPICT

Анализ представленности функциональных групп генов, ассоциированных с ГКМГК, в различных типах тканей и клеток, а также приоритизация генов была проведена при помощи DEPICT v1.1 rel194 [33]. Запуск программы был осуществлен с настройками по умолчанию (<https://data.broadinstitute.org/mpg/depict/>) с двумя порогами статистической значимости для SNP: $p\text{-value} < 5.0\text{e-}08$ и $p\text{-value} < 1.0\text{e-}05$. Участок главного комплекса гистосовместимости (major histocompatibility complex, МНС) был исключен из анализа. При анализе результатов DEPICT был установлен порог значимости $\text{FDR} < 0.05$ (false discovery rate, поправка Бенджамини-Хохберга [82]).

Литературная аннотация генов

Для генов, расположенных рядом с наиболее значимыми SNP (± 250 кб), была проведена литературная аннотация с помощью баз данных OMIM (<https://www.omim.org/>), Google Scholar, NCBI Gene (<https://www.ncbi.nlm.nih.gov/gene>) и PubMed с целью определения их функции для отбора (приоритизации) генов, наиболее вероятно вовлеченных в патогенез хронической боли. Список генов, находящихся в исследуемых регионах, был составлен с помощью региональных графиков ассоциаций (см. онлайн-приложение А) и приведены в таблице В.1 онлайн-приложения В. Сводная информация о генах, наиболее вероятно связанных с хронической болью, приведена в таблице С.1 онлайн-приложения С.

Предсказание эффектов локусов

Для реплицированных локусов нами было проведено предсказание функциональных эффектов. В анализ, помимо SNP с наиболее сильными сигналами ассоциации в

реплицированных локусах, были также включены SNP и инделы, находящиеся в высоком неравновесии по сцеплению ($r^2 > 0.8$) с ними. Неравновесие по сцеплению было вычислено с помощью Plink 1.9 [83] (опция --show-tags) на данных генотипирования людей европейского происхождения из проекта «1000 Геномов» (3 фаза, версия 5, объем выборки 503 человека). Кроме того, из реплицированных локусов были отобраны SNP, ассоциированные с ГКМГК с порогом $p\text{-value} \leq T$, где $\log_{10}(T) = \log_{10}(P_{\min}) + 1$, и P_{\min} – это значение $p\text{-value}$ для ведущего SNP в регионе. Эти SNP были включены в анализ, поскольку данные генотипирования UK Biobank были импутированы с использованием HRC панели, и часть SNP из HRC могла не входить в панель проекта «1000 Геномов». Все отобранные SNP были аннотированы при помощи Ensembl Variant Effect Predictor (VEP) [84], а также инструментов FATHMM-XF [85] и FATHMM-INDEL [86]. Для последних двух методов предсказание эффектов SNP было проведено в соответствии со шкалой от 0 до 1, где значения > 0.5 соответствуют вредным мутациям, а значения < 0.5 указывают на возможный полезный или нейтральный эффект замены.

Генетические корреляции и наследуемость

Наследуемость признаков, обусловленная исследуемыми SNP (h^2), и генетические корреляции ГКМГК с различными признаками человека были оценены методом LD Score регрессии [36]. В совокупности, было проанализировано 209 признаков человека, не описанных в данных UK Biobank и доступных в базе данных LD hub (<http://ldsc.broadinstitute.org/ldhub/>). Предварительно из анализа были удалены дубликаты и отобраны только самые последние исследования для данных признаков по номеру ID в PubMed. Порог статистической значимости был задан как $5.98e-05$ (то есть $0.05/209*4$, где 209 – это общее число признаков, а 4 – количество ГКМГК).

Для 39 признаков, статистически значимо коррелированных хотя бы с одной из ГКМГК, была построена матрица генетических корреляций. Кластеризация и визуализация результатов были реализованы в пакете «corrplot» на языке программирования R (стандартная функция «hclust»). Для кластеризации была проведена предварительная оценка квадратов Евклидовых расстояний путем вычитания абсолютных значений генетических корреляций из 1 и применения метода кластеризации ward.D2.

Результаты

Краткий обзор дизайна эксперимента

Данное исследование направлено на выделение и описание общих наследственных факторов, лежащих в основе четырех типов скелетно-мышечных болей человека: боли в спине, шее, бедре и колене (рис. 3). Каждый из перечисленных типов боли имеет статистически значимую SNP-наследуемость (2 – 4%, табл. D.1 онлайн-приложение D) и генетически коррелирован с остальными типами хронической боли (рис. 3, *d*). Для достижения поставленной цели на матрице генетических ковариаций был применен метод главных компонент.

Каждая из хронических болей была разложена на четыре генетические главные компоненты (ГКМГК1, ГКМГК2, ГКМГК3, ГКМГК4) на данных поисковой и репликационной выборок. Для каждой ГКМГК были получены суммарные статистики ПГАА. Ассоциации, преодолевшие статистический порог полногеномной значимости на данных поисковой выборки, рассматривались как реплицированные при условии статистической значимости для данных мета-анализа репликационных когорт. Для реплицированных локусов была проведена приоритизация генов с использованием нескольких подходов. Более того, также был проведен функциональный биоинформатический анализ представленности групп генов (DEPICT) и исследование генетических корреляций ГКМГК с другими комплексными признаками и заболеваниями человека. Весь функциональный анализ проведен на данных выборки людей европейского происхождения с целью достижения большей мощности анализа за счет большего размера выборки.

Главные компоненты матрицы генетических ковариаций

Графическое представление нагрузок (loadings), то есть коэффициентов с которыми каждая из четырех хронических скелетно-мышечных болей входит в ГКМГК, а также отображение вклада каждой из ГКМГК в разные типы боли представлено на рисунке 3, *a*, *b*.

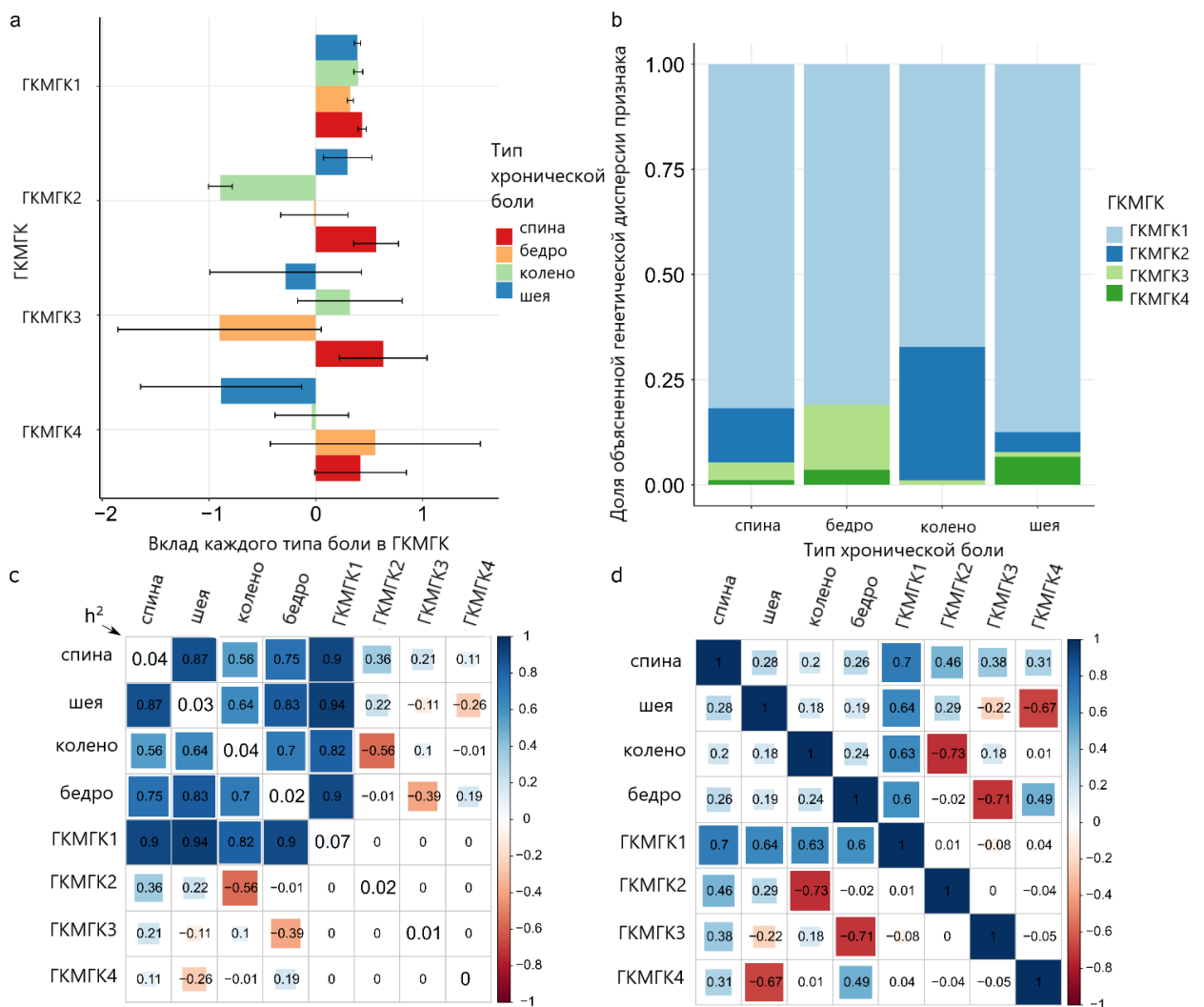


Рисунок 3 – ГКМГК четырех хронических скелетно-мышечных болей человека:

а – вклад разных типов боли в ГКМГК; б – генетическая дисперсия разных типов боли, объясненная ГКМГК; в – тепловая карта генетических корреляций между четырьмя типами боли и ГКМГК; д – тепловая карта фенотипических корреляций четырех типов боли и ГКМГК

На графике *а* отражен 95% доверительный интервал коэффициентов ортогональной трансформации. На рисунке *в* по диагонали отражены коэффициенты наследуемости признаков, h^2

Из рисунка 3, *а* видно, что нагрузки при всех четырех болях в составе ГКМГК1 имеют одинаковые знаки и примерно равны по модулю, что не наблюдается для других ГКМГК. Так, в ГКМГК2 наибольший по модулю коэффициент имеет боль в колени, а в ГКМГК3 и ГКМГК4 – боль в бедре и боль в шее, соответственно. Различный характер взаимосвязи между разными типами боли и ГКМГК можно также проследить на рисунке 3, *б*. Он

отражает схожесть в распределении вкладов ГКМГК в хроническую боль в спине и в шее и уникальность паттернов для боли в колене и бедре. Из графика также видно, что ГКМГК1 имеет наибольший вклад во все фенотипы, что согласуется с тем, что ГКМГК1 объясняет наибольшую долю дисперсии признаков. Кроме того, для ГКМГК1 была показана наибольшая стабильность на основе 95% доверительного интервала.

Полногеномный анализ ассоциаций ГКМГК

При анализе данных поисковой выборки было найдено девять локусов, преодолевших порог статистической значимости $p\text{-value} < 1.3e-08$ (то есть $5.0e-08/4$, где 4 – количество ГКМГК) после коррекции на фактор инфляции тестовой статистики (1.016 для ГКМГК1, 1.001 для ГКМГК2, 1.013 для ГКМГК3 и 1.021 для ГКМГК4). Шесть из них ассоциированы с ГКМГК1, три остальных – с ГКМГК2 (табл. 2). Условный и совместный анализ с помощью SOJO показал наличие только одного сигнала в каждом из найденных локусов (см. онлайн-приложение E).

Таблица 2 – Ведущие SNP, ассоциированные с ГКМГК

ГКМГК**	Ведущий SNP	Хромосома:позиция†	RefA/EffA‡	Поисковая выборка (N = 265,000)					Мета-анализ репликационных когорт*				
				β	SE	P	$P_{(GC)}^{\ddagger}$	EAF	β	SE	P	EAF	N
ГКМГК2	rs143384	20:34025756	Ц/Т	-0.020	0.003	4.87e-13	7.40e-13	59.8%	-0.022	0.003	1.65e-10	58.5%	191,580
ГКМГК1	rs7628207	3:49754970	Т/Ц	-0.023	0.004	1.71e-10	2.37e-10	82.3%	-0.012	0.004	4.92e-03	81.8%	191,580
ГКМГК1	rs13107325	4:103188709	Т/Ц	-0.032	0.005	8.78e-10	1.19e-09	92.6%	-0.035	0.007	4.21e-08	92.6%	191,580
ГКМГК1	rs3737240	1:150483355	Т/Ц	0.017	0.003	2.01e-09	2.69e-09	60.4%	0.010	0.003	3.17e-03	61.1%	191,580
ГКМГК1	rs73581580	9:140251458	Г/А	0.025	0.004	3.89e-09	5.15e-09	12.4%	0.030	0.005	9.54e-09	12.3%	174,831
ГКМГК1	rs12705966	7:114248851	Г/А	0.018	0.003	5.71e-09	7.52e-09	66.7%	0.012	0.004	1.70e-03	67.2%	191,580
ГКМГК2	rs4985445	16:69867835	Г/А	0.017	0.003	1.56e-09	2.09e-09	54.3%	0.007	0.003	0.0371	53.2%	191,580
ГКМГК2	rs548227718	5:175902724	Г/А	-0.283	0.048	3.02e-09	4.01e-09	0.1%	0.096	0.060	0.1056	0.1%	174,831
ГКМГК1	rs111368900	1:53084695	Г/А	0.242	0.041	5.01e-09	6.60e-09	0.2%	0.089	0.048	6.55e-02	0.2%	174,831

Примечание – Реплицированные локусы выделены полужирным шрифтом

EAF – частота эффекторного аллеля (effective allele frequency); P – значение *p-value*; SE – стандартная ошибка β (standard error); N – размер выборки

*Репликационные когорты людей африканского, южно-азиатского и европейского происхождения (3.9%, 4.8% и 91.3% от общего размера репликационной выборки)

**ГКМГК, ассоциированная с локусом

†Хромосома: позиция – позиция (в пн) на хромосоме в соответствии с геномной сборкой GRCh37.p13

‡Референсный аллель/эффекторный аллель (reference allele/effective allele)

§Значение *p-value*, скорректированное на фактор инфляции тестовой статистики с помощью интерсепта LD Score регрессии

В результате репликации была подтверждена ассоциация шести локусов с ГКМГК (пять из них ассоциированы с ГКМГК1 и один – с ГКМГК2). Порог репликации был определен как $p\text{-value} < 5.6e-03$ (то есть $0.05/9$, где 9 – это количество локусов, найденных на поисковой выборке). Информация о реплицированных локусах приведена в таблице В.1 онлайн-приложения В.

Два из шести реплицированных локусов ассоциированы не только с ГКМГК1 и ГКМГК2, но и статистически значимо ассоциированы с хронической болью по данным поисковой выборки ($p\text{-value} < 5.0e-08$, онлайн-приложение В). К ним относится ГКМГК1-ассоциированный локус с сигналом в SNP rs73581580 (ассоциирован с хронической болью в спине) и ГКМГК2-ассоциированный локус с сигналом в rs143384 (ассоциирован с хронической болью в шее). Ассоциации с хронической болью для этих и двух дополнительных локусов также достигли порога статистической значимости на данных мета-анализа для европейских когорт. К этим дополнительным локусам относятся два ГКМГК1-ассоциированных локуса с сигналами в rs13107325 и rs3737240, ассоциированные с хронической болью в шее и в бедре, соответственно (см. онлайн-приложение В).

Для SNP rs143384, rs13107325 и rs3737240 функциональные эффекты и ассоциации с другими сложными признаками и заболеваниями человека были описаны в ряде других работ (см. онлайн-приложение F). В частности, полиморфизм rs13107325 является миссенс-мутацией в гене транспортера бивалентных катионов *SLC39A8*. Он также является одним из самых полиморфных участков в человеческом геноме, ассоциированным с множеством признаков (см. онлайн-приложение F табл. F.1). Так, например, показана ассоциация аллеля Т со снижением роста, увеличением кривизны позвоночника, повышенным риском развития тяжелого подросткового идиопатического сколиоза [87], болезнью Крона [88] и шизофренией [89]. Миссенс-вариант rs3737240 находится в гене *ECM1* (extracellular matrix protein 1, внеклеточный матричный белок 1), кодирующем негативный регулятор энхондрального окостенения [90]. В более ранних исследованиях описана ассоциация с аллелем Т (или аллеля А близко расположенного к нему полиморфизма rs13294, также являющегося миссенс-вариантом в гене *ECM1*) с язвенным колитом [90–93]. Показано также, что полиморфизм rs143384 в 5'-нетранслируемой области гена фактора роста и дифференциации *GDF5* находится в сильном неравновесии по сцеплению с SNP rs143383 (аллель Т rs143384 сцеплен с аллелем Т полиморфизма rs143383). По данным предыдущих исследований, Т аллель rs143383 ассоциирован с пониженным уровнем экспрессии гена *GDF5* в суставах [94], повышенным риском развития остеоартрита [94–98] и врожденным вывихом бедра [99,100]. В данном исследовании ассоциация rs143383 с ГКМГК2 имеет

такой же размер эффекта, как и ассоциация rs143384 с ГКМГК2, и обе эти ассоциации являются статистически значимыми (для данных поисковой выборки $p\text{-value} = 8.53e-12$ после коррекции на остаточную инфляцию тестовой статистики).

Функциональная аннотация найденных сигналов

В результате анализа литературных источников, было найдено 13 генов (см. онлайн-приложение С), которые потенциально могут быть ответственными за возникновение хронической скелетно-мышечной боли. К ним относятся: гены *GDF5* и *MMP24* (для полиморфизма rs143384), *AMIGO3* и *BSN* (для rs7628207), *SLC39A8* (для rs13107325), *MIR6878*, *ECM1* и *CTSS* (для SNP rs3737240); гены *MIR7114*, *NSMF*, *NOXA1* и *GRIN1* (для полиморфизма rs73581580) и ген *FOXP2* (для rs12705966). Для некоторых из них ранее была установлена связь с нейропатической болью (*MMP24*, *CTSS*, *NOXA1*) и восстановлением ЦНС после повреждения (*AMIGO3*). Гены *GDF5*, *SLC39A8*, *ECM1*, *MIR6878* и *MIR7114* также ассоциированы с заболеваниями скелетно-мышечного аппарата (остеоартритом, анкилозирующим спондилитом) и/или развитием скелета. Часть генов (*BSN*, *NSMF*, *GRIN1* и *FOXP2*) оказались вовлечены в процессы развития нервной системы и синаптическую передачу. В частности, продукт гена *FOXP2* является необходимым для правильного развития речевой и языковой зон головного мозга в процессе эмбриогенеза.

Предсказание эффектов локусов

С помощью VEP было определено четыре миссенс-варианта: rs13107325 в гене *SLC39A8*, rs3737240 и rs13294 в гене *ECM1* и rs79140116 в гене *EXD3*. С помощью инструментов SIFT и PolyPhen предсказаны вредные (possibly damaging/deleterious) эффекты только для полиморфизмов rs13107325 и rs13294, остальные SNP были определены как нейтральные (benign/tolerated) (см. онлайн-приложение G табл. G.1).

Для триаллельного SNP rs13107325 (Ц > Т, А) потенциально вредные (possibly damaging) эффекты были предсказаны для обоих минорных (с меньшей частотой в популяции) аллелей Т и А. Аллель А является очень редким, поэтому он не был включен в данное исследование. Аллель Т оказался ассоциирован с ГКМГК1 с положительным знаком при β (размер эффекта).

Полиморфизм rs13294 также является триаллельным (Г > А, Т). Как и в предыдущем случае, очень редкий аллель Т не был включен в ПГАА. Инструментами SIFT и PolyPhen предсказан вредный эффект только для наиболее редкого аллеля Т, в то время как аллель А (ассоциированный с ГКМГК1 с отрицательным знаком при β) был определен как нейтральный. Однако, возможно, что в случае сильного эффекта редкого аллеля rs13294 Т на ГКМГК1, сильный сигнал SNP rs3737240 является наведенным (rs3737240 и rs13294

расположены на расстоянии 1.6 кб, что отвечает сильному неравновесию по сцеплению). С помощью FATHMM-XF и FATHMM-INDEL были предсказаны возможно патогенные SNP rs28535523 в интроне гена *UBA7*, а также интронный индел rs34291892 в гене *FOXP2* (онлайн-приложение G табл. G.2, 3). Потенциально вредные замены rs28535523 Т и rs34291892 (инсерция А) оказались положительно ассоциированными с ГКМГК1.

Данные соотнесения описанных аллелей с эффектами на ГКМГК, прогнозируемые последствия нуклеотидных замен, аминокислотных замен (где это возможно) и аллелей ведущих SNP представлены в таблице G.4 онлайн-приложения G.

Приоритизация генов с помощью DEPICT

Статистически значимые результаты приоритизации генов с помощью DEPICT ($FDR < 0.05$) были получены только для ГКМГК1 и только при пороге значимости $p\text{-value} = 1.0e-05$. Список приоритизированных генов приведен в таблице H.1 онлайн-приложения H. Все гены, отобранные на основе анализа литературных источников, были проверены на вхождение в этот список.

Результаты приоритизации генов с использованием разных инструментов

Полный список генов, отобранных на основе литературной аннотации и приоритизации в DEPICT представлен в таблице 3. Для локусов с сигналами в SNP rs143384, rs13107325, rs3737240 и rs12705966 были получены указания на вовлеченность в патогенез хронической боли следующих генов: *GDF5*, *SLC39A8*, *ECM1* и *FOXP2*, соответственно. Однако для полиморфизмов rs7628207 и rs73581580 не было отобрано ни одного гена, поскольку разные методы дали разные результаты. Так, ближайшим к rs7628207 геном является *AMIGO3*, для которого было описано участие в ингибировании регенерации аксонов ЦНС после повреждения [101]. В то же время, в данном регионе расположено два других гена, которые были приоритизированы с *in silico* методов и литературной аннотации (в частности, ген *BSN*, кодирующий пресинаптический белок цитоматрикса Bassoon). Ведущий SNP rs73581580 располагается в интроне гена *EXD3* и является ортологом *mut-7* гена *C. elegans*, участвующего в сайленсинге транспозонов и РНК-интерференции. Тем не менее, литературные данные указывают еще на четыре гена, с более вероятной вовлеченностью в развитие хронической скелетно-мышечной боли (см. онлайн-приложение С).

Таблица 3 – Результаты приоритизации генов

Ведущий SNP в локусе	Локус*	Количество генов в локусе [†]	Приоритизированный ген	Является ли ген ближайшим к ведущему SNP	Основание для приоритизации
rs143384	20:34025756	15	<i>GDF5</i>	да (5'-НТО)	Л
rs7628207	3:49754970	18	<i>AMIGO3</i>	да (интронный)	Л
			<i>BSN</i>	нет	Л, D
			<i>UBA7</i>	нет	V
rs13107325	4:103188709	3	<i>SLC39A8</i>	да (миссенс)	Л, V
rs3737240	1:150483355	19	<i>ECM1</i>	да (миссенс)	Л, V
rs73581580	9:140251458	32	<i>MIR7114</i>	нет	Л
			<i>NSMF</i>	нет	Л
			<i>NOXA1</i>	нет	Л
			<i>GRIN1</i>	нет	Л
rs12705966	7:114248851	2	<i>FOXP2</i>	да (интронный)	Л, V, D

Примечание – Гены с наиболее вескими основаниями для приоритизации выделены полужирным шрифтом. Для генов, ближайших к ведущему SNP, в скобках указано его положение относительно данного гена

Л – литературная аннотация; D – анализ с помощью DEPICT; V – анализ с помощью VEP/FATHMM; НТО – нетранслируемая область

*Номер хромосомы: позиция (в пн) на хромосоме (в соответствии со сборкой GRCh37.p13)

[†]Вычислено на основе региональных графиках ассоциаций, полученных при помощи LocusZoom (<http://locuszoom.org/>, онлайн-приложение A)

Анализ представленности функциональных групп генов

Анализ представленности функциональных групп генов в различных тканях и типах клеток показал статистически значимые результаты только для ГКМГК1 (см. онлайн-приложение Н табл. Н.3-6). Для наборов SNP, ассоциированных с ГКМГК1 с порогом значимости $p\text{-value} < 5.0e-08$, обогащение с $FDR < 0.05$ было обнаружено для двух терминов: «нейтральные стволовые клетки» и «сетчатка». После снижения порога вхождения SNP до $p\text{-value} < 1.0e-05$, было обнаружено обогащение генами в 24 дополнительных типах тканей, каждый из которых связан с ЦНС. Сходный паттерн наблюдался и при анализе обогащения исследуемого набора SNP функциональными группами генов, который выявил 462 термина (порог статистической значимости для SNP $p\text{-value} < 1.0e-05$) в большинстве своем связанных с функционированием нервной системы,

ее развитием и морфологией (например, «регуляция развития нервной системы», «аксоногенез», «синапс», «регуляция передачи нервного импульса» и другие).

Генетические корреляции ГКМГК с другими сложными признаками

Анализ генетических корреляций ГКМГК показал статистически значимые корреляции ГКМГК1 с 39 сложными признаками человека (рис. 4, онлайн-приложение J табл. J.1). Среди них 11 признаков, напрямую связанных с избыточным весом (индекс массы тела, лишний вес, ожирение, обхват талии), что согласуется с эпидемиологическими наблюдениями об ассоциации хронической боли в спине и ожирении [102]. Пять других признаков также попали в данный кластер: холестерин липопротеинов высокой плотности или ЛПВП-холестерин (отрицательная генетическая корреляция с ГКМГК1), триглицериды, индекс инсулинорезистенции НОМА-IR, лептин и инсулин натощак. Сильные генетические корреляции ($|r_g|$ в диапазоне от 0.31 до 0.54) были показаны также для ГКМГК1 и кластера психиатрических и личностных черт (большое депрессивное расстройство, депрессивные симптомы, субъективное благополучие и невротизм). Эти наблюдения подтверждают данные более ранних близнецовых и семейных исследований с применением смешанных моделей, свидетельствующих об общности генетических основ боли и депрессии [103–105]. Также ГКМГК1 оказалась генетически коррелирована с различными социально-демографическими и репродуктивными признаками, обучением и курением, ревматоидным артритом и ишемической болезнью сердца, а также с продолжительностью сна. Наиболее сильные генетические корреляции с ГКМГК1 были обнаружены для: возраста матери при рождении первого ребенка ($r_g = -0.56$), депрессивными симптомами ($r_g = 0.54$) и окончанием колледжа ($r_g = 0.54$). в целом, наблюдаемый паттерн генетических корреляций ГКМГК1 сходен с таковым для хронической боли в спине [67].

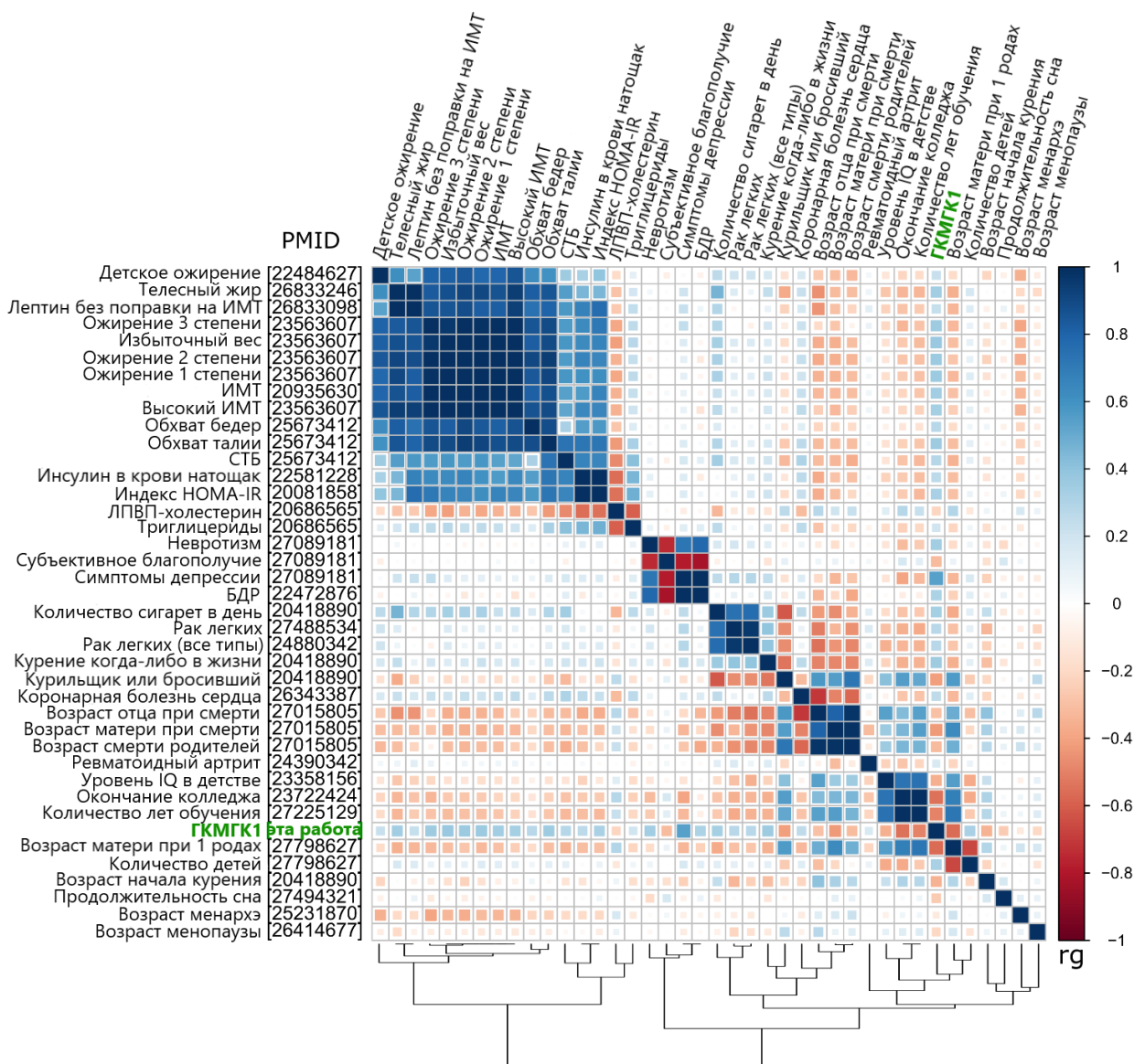


Рисунок 4 – Тепловая карта генетических корреляций ГКМГК1 с другими комплексными признаками и заболеваниями человека

Цветовая шкала сбоку отображает величину коэффициента генетической корреляции (r_g).

ИМТ – индекс массы тела; СТБ – отношение обхвата талии к обхвату бедер; HOMA-IR – индекс инсулинорезистентности; ЛПВП – липопротеины высокой плотности; БДР – большое депрессивное расстройство; PMID – PubMed ID литературного источника, предоставляющего суммарные статистики ПГАА для данного признака

Анализ генетических корреляций для других ГКМГК показал, что ГКМГК2 генетически коррелирована только с признаками, связанными с ожирением, а ГКМГК4 – с обхватом талии. Для ГКМГК3 статистически значимых корреляций с другими сложными признаками человека выявлено не было (онлайн-приложение J табл. J.2-4).

Обсуждение

Хронические скелетно-мышечные боли человека отличаются сложной генетической архитектурой, в которую вовлечено множество генетических вариантов с малыми эффектами. Эта особенность затрудняет изучение данных заболеваний, поскольку даже масштабные исследования с применением метода ПГАА позволяют найти и подтвердить лишь небольшое количество ассоциаций и существенно занижают оценки наследуемости признаков по SNP. В настоящее время накапливается все больше доказательств того, что хроническая боль разных локализаций имеет общую генетическую основу [3,57]. Рассмотрение нескольких признаков совместно в рамках одного исследования может помочь выявить общие гены и биологические пути хронических скелетно-мышечных болей.

В этой работе был применен метод главных компонент, позволивший выявить общие наследственные факторы для хронических скелетно-мышечных болей разной локализации: боли в спине, шее, колене и бедре. Предложенный подход предполагает получение наследственной компоненты группы генетически коррелированных признаков путем перехода к новым генетически независимым признакам – главным компонентам матрицы генетических ковариаций. ГКМГК представляют собой взвешенную сумму исходных признаков, веса которых подбирались таким образом, чтобы ГКМГК1 объясняла наибольшую долю генетической дисперсии признаков интереса и генетической ковариации между ними, а каждая последующая ГКМГК (ГКМГК2-4) объясняла все меньшую долю описываемых параметров. Веса, с которыми четыре признака, рассматриваемые в данном исследовании, вошли в ГКМГК1, оказались практически одинаковыми (рис. 3, а). Это указывает на то, что ГКМГК1 – наследственная компонента, объясняющая большинство случаев хронической скелетно-мышечной боли изучаемых нами локализаций, влияет на риск развития данных заболеваний в практически равной мере, не зависимо от локализации. В отличие от ГКМГК1, ГКМГК2 сайт-специфична и оказывает влияние на генетическую предрасположенность к хронической боли в колене, но не к хронической боли в спине или шее.

В данной работе были картированы и реплицированы шесть геномных локусов (пять из них ассоциировано с ГКМГК1, один – с ГКМГК2). Важно заметить, что на данных поисковой выборки, только два из шести реплицированных локусов имели полногеномный уровень значимости ассоциации с исходными признаками: rs73581580 ассоциирован с хронической болью в спине, а rs143384 – с хронической болью в колене. Помимо этого, оценка

наследуемости ГКМГК1 по SNP оказалась значительно выше таковой для любого из исходных признаков (7% против 2-4% на исследуемой выборке, оценка получена методом LD Score регрессии). Эти результаты свидетельствуют о том, что метод главных компонент увеличивает мощность анализа при поиске генетических факторов, влияющих на риск возникновения хронической боли. В сравнении с другими методами многомерного анализа: методом максимизации наследуемости признака (Maximization of heritability) [106], MANOVA [107] и MTAG (Multi-Trait Analysis of GWAS) [108], – позволяющими увеличить мощность анализа при изучении генетики признаков, метод главных компонент на матрице генетических ковариаций имеет ряд преимуществ. Во-первых, перечисленные подходы решают только одну задачу – увеличивают мощность анализа, то есть увеличивают количество найденных локусов, ассоциированных хотя бы с одним признаком. Эти методы не направлены на поиск факторов, общих для всего набора признаков. Во-вторых, из названных альтернативных подходов только метод максимизации наследуемости позволяет получить суммарные статистики, пригодные для дальнейшего *in silico* биоинформатического анализа. Следовательно, применение MANOVA и MTAG ограничивает возможности дальнейшего анализа и интерпретации выходных данных. Из всего выше сказанного можно заключить, что метод главных компонент на матрице генетических ковариаций является наиболее оптимальным подходом для решения задач данного исследования.

Среди шести реплицированных в данной работе локусов, три содержат хорошо изученные полиморфизмы, ассоциированные с рядом других признаков и состояний (rs13107325, rs3737240 и rs143384, онлайн-приложение F). Эти результаты иллюстрируют многообразие эффектов ГКМГК-ассоциированных локусов и свидетельствуют о наличии общих механизмов, лежащих в основе хронической скелетно-мышечной боли и множества различных признаков человека.

Биологические пути и такни, ассоциированные с ГКМГК1, в большинстве своем имеют отношение к развитию и функционированию ЦНС, что позволяет выдвинуть предположение о том, что ГКМГК1 отражает неврологические и психические аспекты хронической скелетно-мышечной боли. С этой гипотезой согласуется тот факт, что один из приоритизированных генов из ГКМГК1-ассоциированного локуса – *FOXP2* – кодирует транскрипционный фактор, который экспрессируется в мозге в эмбриональном периоде и во взрослом организме и играет важную роль в развитии структур мозга, ответственных за речь [109,110]. На вовлеченность психической компоненты дополнительно указывает сильная положительная генетическая корреляция между ГКМГК1 и депрессивными

симптомами. Наряду с этим важно, что ГКМГК1 также ассоциирована с признаками, отражающими общее состояние организма, и с факторами риска скелетно-мышечных болей: социально-демографическими и репродуктивными характеристиками, образованием и курением, а также с продолжительностью сна. Роль морфологических факторов в развитии хронической скелетно-мышечной боли также подтверждается ассоциацией ГКМГК1 с генами *SLC39A8* и *ECM1*, которые участвуют в развитии и функционировании опорно-двигательной системы человека. Ген *ECM1* кодирует негативный регулятор минерализации костей и хондрогенеза [90,111]. Показано, что ассоциированный с ГКМГК1 (повышающей риск хронической скелетно-мышечной боли) полиморфизм из этого гена ассоциирован с повышенной концентрацией белка ECM1. Кроме того, ГКМГК1-ассоциированный аллель rs3737240 Ц в гене *ECM1* находится в неравновесии по сцеплению с аллелем rs12040949 Ц ($r^2 = 0.94$ для европейской популяции), ассоциированным с повышенным риском остеоартрита тазобедренного сустава [112]. Продукт гена *SLC39A8* также вовлечен в разрушение хрящей при остеоартрите [113]. Для мутантов *Slc39a8* рыбы данио-рерио описаны аномалии развития позвоночника, нарушение процессов роста и снижение двигательной активности. ГКМГК1-ассоциированная миссенс-мутация rs13107325 в гене *SLC39A8* ранее была описана как ассоциированная с повышенным риском остеоартрита [112] и тяжелым подростковым идиопатическим сколиозом [87]. Таким образом, сходно с результатами более раннего исследования боли в спине, проведенного нашей исследовательской группой [114], генетические факторы, обуславливающие хроническую скелетно-мышечную боль, отображают биологическую, социальную и психическую компоненты боли.

Поскольку данная работа посвящена исследованию хронической скелетно-мышечной боли в тех анатомических сайтах, которые наиболее часто поражаются остеоартритом, нами ожидаемо были обнаружены локусы и гены, ассоциированные с этим заболеванием, а также были показаны генетические корреляции между ГКМГК1 и остеоартритом ($r_g = 0.65$). Здесь важно заметить, что анализ обогащения в генах и тканях для ГКМГК1 показал преобладание терминов, связанных с ЦНС. В то время как, в недавнем исследовании генетики остеоартрита анализ обогащения не выявил терминов, напрямую относящихся к ЦНС («морфогенез анатомических структур», «транспорт через ионные каналы», «метаболизм гистидина» и др.) [112]. Это указывает на генетические различия между ГКМГК1 и остеоартритом и подтверждает, что генетическая общность хронических скелетно-мышечных болей не является опосредованной, в первую очередь, остеоартритом.

Для подсчета генетических корреляций нами были использованы суммарные статистики ПГАА остеоартрита, полученные на данных UK Biobank. Фенотипы в этих данных были определены в соответствии с кодами МКБ-9 и МКБ-10 (из электронных медицинских карт пациентов), следовательно, о том, была ли проведена радиография суставов для пациентов, не было известно, а значит, диагноз «остеоартрит» мог быть поставлен без нее, на основе внешних изменений и симптоматики (в первую очередь, болевой). Такой подход к диагнозе этого заболевания обоснован тем, что боль считается основным симптомом и клиническим следствием развития остеоартрита [115]. Таким образом, возможно, что наблюдаемая генетическая корреляция между ГКМГК1 и остеоартритом – это генетическая корреляция ГКМГК1 с болью при остеоартрите, а не с остеоартритом как таковым. Между тем, остеоартрит не всегда сопровождается болью, что наблюдается при корректной постановке диагноза с использованием метода радиографии. Так, имеются данные популяционных исследований, которые сообщают о слабой взаимосвязи между тяжестью поражений при остеоартрите, определенном в ходе радиографического обследования, и интенсивностью боли. В частности, порядка 30-40% пациентов с остеоартритом коленного сустава III и IV степени вообще не испытывают боли [20].

В то время как ГКМГК1 является компонентой, по которой пациенты с любым из исследуемых типов хронической скелетно-мышечной боли максимально отличаются от здоровых людей, другие ГКМГК могут отражать факторы, определяющие хроническую боль конкретной локализации. В частности, доказательство в поддержку этого предположения можно наблюдать на примере ГКМГК2, относящейся к боли в колене (рис. 3, *b*). Единственным ассоциированным с ней геном (на полногеномном уровне значимости) является *GDF5*, для которого показана ассоциация с остеоартритом периферических суставов и дегенерацией межпозвонковых дисков [94,95,97,112]. Данный факт согласуется с тем, что коленный сустав является одним из самых частых сайтов остеоартрита. Интерпретация биологического смысла ГКМГК3 и ГКМГК4 оказалась осложнена отсутствием значимых сигналов ассоциации, однако из рисунка 3, *b* можно предположить, что они также отражают аспекты хронической боли, специфичные для определенных локализаций. Так ГКМГК3 вносит существенный вклад в хроническую боль в бедре, а ГКМГК4 – в боль в шее.

Нельзя не отметить, что в недавней работе Джонстона и др. [71] был предложен альтернативный способ для изучения общих наследственных факторов хронической боли разных локализаций. Для совместного рассмотрения группы заболеваний в этом исследовании был также осуществлен переход к новому признаку, который отражал

количество сайтов хронической боли. Новый фенотип, сформированный в этой работе, может показаться схожим с ГКМГК1. Однако стоит заметить, что простое суммирование сайтов хронической боли может несколько исказить результаты, поскольку такой подход допускает совместное рассмотрение признаков с потенциально разной генетической основой. В частности, этот метод предполагает, но не проверяет, эквивалентность генетических факторов, влияющих на скелетно-мышечные боли (например, боль в спине и колене), мигрени (в случае головной боли), зубной или нейропатической боли (при орофациальной боли), а также висцеральной боли (например, боли в животе). Это является очень сильным утверждением, требующим дополнительного доказательства. В отличие от этого, подход, реализованный в нашей работе, является эмпирическим, и выделение общих наследственных факторов для группы признаков возможно только при условии их генетической общности (мерой которой является генетическая ковариация).

Тем не менее, наше исследование и работа Джонстона и др. являются одними из первых, использующих ПГАА для изучения генетики хронических болей разных локализаций. Несмотря на отличия в методологии и исследуемых фенотипах, полученные в обеих работах результаты хорошо согласуются между собой. В частности, оба исследования указывают на существенную роль ЦНС в развитии хронической скелетно-мышечной боли и на ее взаимосвязь с депрессией. Кроме того, четыре из приоритизированных нами генов были также обнаружены в работе Джонстона и др.: *AMIGO3* (rs7628207), *SLC39A8* (rs13135092), *ECM1* (rs59898460) и *FOXP2* (rs12537376). Стоит добавить, что в нашей работе перечисленные локусы были реплицированы, в отличие от исследования Джонстона и др. Однако использование данных UK Biobank для формирования поисковой и репликационной выборки в данном исследовании, также как и отсутствие репликации в работе Джонстона, является ограничением, поскольку полученные результаты необходимо подтвердить на независимой выборке.

Наше исследование имеет и другие ограничения. Первое, и самое важное из них, – это установление фенотипов по данным анкетирования. Данный способ может привести к увеличению гетерогенности исследуемых признаков. Чтобы преодолеть эту проблему, мы использовали ГКМГК, которые аппроксимируют генетические основы группы признаков и, по-видимому, отображают общую болевую компоненту хронических скелетно-мышечных болей. Следующим ограничением является то, что в данной работе была рассмотрена боль в анатомических сайтах, связь между которыми может быть обусловлена остеоартритом (поскольку они являются наиболее частыми сайтами локализации этого заболевания). Это означает, что необходимо быть осторожным при обобщении полученных

результатов на другие хронические боли человека. Наконец, для двух из шести найденных локусов (с ведущими SNP rs7628207 и rs73581580) мы не смогли приоритизировать определенный ген, а для двух локусов, содержащих полиморфизмы rs143384 и rs73581580 соответственно, все потенциально функциональные гены были определены только на основе литературных данных (табл. 3).

Заключение

С использованием метода главных компонент выявлены общие наследственные факторы для хронических скелетно-мышечных болей человека: боли в спине, шее, колене и бедре – и определены функциональные локусы и гены, ассоциированные с их общей генетической составляющей. Установлена ведущая роль неврологических и психических факторов в развитии хронической скелетно-мышечной боли. Это хорошо согласуется с биопсихосоциальной моделью данного заболевания. Предложенный в рамках этой работы подход – совместное рассмотрение группы схожих признаков и применение метода главных компонент на матрице генетических ковариаций между ними – позволил изучить то общее (с точки зрения генетики), что есть у исходных признаков, и справиться с проблемами, характерными для классического подхода к изучению хронических болей по отдельности (низкая мощность анализа и шум в данных из-за высокой гетерогенности признака). Таким образом, было показано, что данная методология может быть использована для изучения механизмов хронической боли и выявления новых ассоциаций.

Выводы

1. С использованием метода главных компонент на матрице генетических ковариаций были выделены ГКМГК четырех хронических скелетно-мышечных болей человека.
2. Показана ассоциация шести локусов с ГКМГК четырех скелетно-мышечных болей человека, для четырех из них ассоциация с хронической болью показана впервые.
3. Выявлено 11 генов, локализованных в шести ассоциированных с ГКМГК районах генома и наиболее вероятно вовлеченных в генетический контроль хронической скелетно-мышечной боли.
4. Показано, что первая ГКМГК обогащена генами, связанными с развитием нервной системы, и экспрессирующимися преимущественно в стволовых клетках, клетках ЦНС и сенсорной системы. Она генетически коррелирована с антропометрическими, социально-демографическими и психиатрическими признаками человека, и, по-видимому, отражает биопсихические основы хронической скелетно-мышечной боли.
5. Показано, что вторая, третья и четвертая ГКМГК гетерогенны, генетически коррелированы с морфологическими и антропометрическими признаками человека, и, по-видимому, отражают уникальные генетические компоненты хронических скелетно-мышечных болей.

Список использованной литературы

1. Breivik H. et al. Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment // *Eur. J. Pain*. 2006.
2. Institute for Health Metrics and Evaluation (IHME). Findings from the Global Burden of Disease Study 2017. Seattle, WA: IHME, 2017.
3. Vehof J. et al. Shared genetic factors underlie chronic pain syndromes // *Pain*. 2014.
4. Тактика и неотложная терапия болевого синдрома: научно-образовательный материал. М.: Ассоциация московских вузов, изд-во Рос. гос. мед. ун-та Фед. агентства по здравоохранению и социальному развитию, 2010. С. 265.
5. Crofford L.J. Chronic Pain: Where the Body Meets the Brain // *Transactions of the American Clinical and Climatological Association*. 2015.
6. Junqueira D.R.G. et al. Heritability and lifestyle factors in chronic low back pain: Results of the Australian Twin Low Back Pain Study (The AUTBACK study) // *Eur. J. Pain (United Kingdom)*. 2014.
7. Loeser J.D. Chapter 2 Pain as a disease // *Handb. Clin. Neurol*. 2006.
8. Diatchenko L. et al. The phenotypic and genetic signatures of common musculoskeletal pain conditions // *Nature Reviews Rheumatology*. 2013.
9. Парфенов В.А., Иванова М.А. Ведение пациентов с болью в спине врачами общей практики // *Медицинский совет*. 2018. Т. 135, № 1. С. 112–117.
10. Gatchel R.J. The biopsychosocial model of chronic pain // *Futur. Med*. 2013. P. 5–17.
11. Chen J. History of pain theories // *Neurosci. Bull*. 2011. Vol. 27, № 5. P. 343–350.
12. McBeth J., Jones K. Epidemiology of chronic musculoskeletal pain // *Best Practice and Research: Clinical Rheumatology*. 2007.
13. Woolf C.J. Pain: Moving from Symptom Control toward Mechanism-Specific Pharmacologic Management // *Annals of Internal Medicine*. 2004.
14. Черепкина Л.П., Таламова И.Г. Избранные лекции по физиологии человека (нервная и сенсорные системы): учебное пособие. Омск: Изд-во Сиб. гос. ун-та физ. культуры и спорта, 2013. С. 112.
15. Treede R.D. et al. A classification of chronic pain for ICD-11 // *Pain*. 2015.
16. Морган Д.Э., Мэгид С.М. Клиническая анестезиология. Т. 2 / Ред. Бунятыян А.А. М., СПб.: БРШОМ, Невский Диалект, 2000. С. 366.
17. Dudley P. Pain Pathways [Electronic resource]. 2018. Vol. 0, № September 2018. P. 1–4.
18. Русинов А. Молекулярная биология боли // *Биомолекула*. 2019. P. 1–15.

19. Scholz J. Mechanisms of chronic pain // *Mol. Pain*. 2014. № 10. P. (Suppl 1):O15.
20. Phillips K., Clauw D.J. Central pain mechanisms in chronic pain states - Maybe it is all in their head // *Best Practice and Research: Clinical Rheumatology*. 2011.
21. Sheng J. et al. The Link between Depression and Chronic Pain: Neural Mechanisms in the Brain // *Neural Plasticity*. 2017.
22. Edwards A.O. et al. Complement factor H polymorphism and age-related macular degeneration // *Science* (80-.). 2005.
23. Yu W. et al. GWAS Integrator: A bioinformatics tool to explore human genetic associations reported in published genome-wide association studies // *Eur. J. Hum. Genet.* 2011.
24. Аульченко Ю.С., Аксенович Т.И. Методологические подходы и стратегии картирования генов, контролирующих комплексные признаки человека // *Вестник ВОГиС*. 2006.
25. Hommel G. A stagewise rejective multiple test procedure based on a modified bonferroni test // *Biometrika*. 1988.
26. Egger M. et al. Meta-analysis: Principles and procedures // *BMJ*. 1997.
27. Li M. et al. An Imputation Approach for Oligonucleotide Microarrays // *PLoS One*. 2013.
28. MARTIN A.R. et al. IMPUTATION-BASED ASSESSMENT OF NEXT GENERATION RARE EXOME VARIANT ARRAYS. 2013.
29. McCarthy S. et al. A reference panel of 64,976 haplotypes for genotype imputation // *Nat. Genet.* 2016.
30. Clarke G.M. et al. Basic statistical analysis in genetic case-control studies // *Nat. Protoc.* 2011.
31. Pasaniuc B., Price A.L. Dissecting the genetics of complex traits using summary association statistics // *Nature Reviews Genetics*. 2017.
32. Yang J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits // *Nat. Genet.* 2012.
33. Pers T.H. et al. Biological interpretation of genome-wide association studies using predicted gene functions // *Nat. Commun.* 2015.
34. Zhu Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets // *Nat. Genet.* 2016.
35. Lee S.H. et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs // *Nat. Genet.* 2013.
36. Bulik-Sullivan B. et al. LD score regression distinguishes confounding from polygenicity

- in genome-wide association studies // *Nat. Genet.* 2015.
37. Bulik-Sullivan B. et al. An atlas of genetic correlations across human diseases and traits // *Nat. Genet.* 2015.
 38. Nieuwboer H.A. et al. GWIS: Genome-Wide Inferred Statistics for Functions of Multiple Phenotypes // *Am. J. Hum. Genet.* 2016.
 39. Pearson, K. On lines and planes of closest fit to systems of points in the space // *Philosophical Magazine.* 1901.
 40. Hotelling H. Analysis of a complex of statistical variables into principal components // *J. Educ. Psychol.* 1933.
 41. Jolliffe I.T., Cadima J. Principal component analysis: A review and recent developments // *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.* 2016.
 42. Jolliffe I.T. *Principal Component Analysis. Second Edition* // Springer Series in Statistics. 2002.
 43. Berrendero J.R., Justel A., Svarc M. Principal components for multivariate functional data // *Comput. Stat. Data Anal.* 2011.
 44. Franklin S.B. et al. Parallel Analysis: a method for determining significant principal components // *J. Veg. Sci.* 2006.
 45. McCabe G.P. *Principal variables* // *Technometrics.* 1984.
 46. Cadima J., Cerdeira J.O., Minhoto M. Computational aspects of algorithms for variable selection in the context of principal components // *Comput. Stat. Data Anal.* 2004.
 47. Aschard H. et al. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies // *Am. J. Hum. Genet.* 2014.
 48. Li Z. et al. Power Calculation of Multi-step Combined Principal Components with Applications to Genetic Association Studies // *Sci. Rep.* 2016.
 49. Ried J.S. et al. A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape // *Nat. Commun.* 2016.
 50. D'Angelo G.M., Rao D., Gu C.C. Combining least absolute shrinkage and selection operator (LASSO) and principal-components analysis for detection of gene-gene interactions in genome-wide association studies // *BMC Proc.* 2009.
 51. Ma S., Kosorok M.R. Identification of differential gene pathways with principal component analysis // *Bioinformatics.* 2009.
 52. Liu M. et al. Integrative sparse principal component analysis of gene expression data // *Genet. Epidemiol.* 2017.

53. Yeung K.Y., Ruzzo W.L. Principal component analysis for clustering gene expression data // *Bioinformatics*. 2001.
54. Giuliani A. The application of principal component analysis to drug discovery and biomedical data // *Drug Discovery Today*. 2017.
55. Zorina-Lichtenwalter K. et al. Genetic predictors of human chronic pain conditions // *Neuroscience*. 2016.
56. Nielsen C., Knudsen G., Steingrimsdóttir Ó. Twin studies of pain // *Clinical Genetics*. 2012.
57. Williams F.M.K., Spector T.D., MacGregor A.J. Pain reporting at different body sites is explained by a single underlying genetic factor // *Rheumatology*. 2010.
58. James S. Human pain and genetics: some basics // *Br. J. Pain*. 2013.
59. Holliday K.L., Thomson W., McBeth J. Genetics of chronic musculoskeletal pain // *Oxford Textbook of Rheumatology*. 2014.
60. Bjorland S. et al. Genetic predictors of recovery in low back and lumbar radicular pain // *Pain*. 2017.
61. Hilgemeier A.T. et al. Pain Perception and the Opioid Receptor Delta 1 // *Cureus*. 2018.
62. Parisien M. et al. Effect of Human Genetic Variability on Gene Expression in Dorsal Root Ganglia and Association with Pain Phenotypes // *Cell Rep*. 2017.
63. Docampo E. et al. Genome-wide analysis of single nucleotide polymorphisms and copy number variants in fibromyalgia suggest a role for the central nervous system // *Pain*. 2014.
64. Peters M.J. et al. Genome-wide association study meta-analysis of chronic widespread pain: Evidence for involvement of the 5p15.2 region // *Ann. Rheum. Dis*. 2013.
65. Mogil J.S. Pain genetics: Past, present and future // *Trends in Genetics*. 2012.
66. Sanders A.E. et al. GWAS Identifies New Loci for Painful Temporomandibular Disorder: Hispanic Community Health Study/Study of Latinos // *J. Dent. Res*. 2017.
67. Freidin M.B. et al. Insight into the genetic architecture of back pain and its risk factors from a study of 509,000 individuals // *Pain*. 2019.
68. Suri P. et al. Genome-wide meta-analysis of 158,000 individuals of European ancestry identifies three loci associated with chronic back pain. // *PLoS Genet*. 2018.
69. Bjornsdottir G. et al. Sequence variant at 8q24.21 associates with sciatica caused by lumbar disc herniation // *Nat. Commun*. 2017.
70. Song Y.Q. et al. Lumbar disc degeneration is linked to a carbohydrate sulfotransferase 3 variant // *J. Clin. Invest*. 2013.

71. Johnston K.J.A. et al. Genome-wide association study of multisite chronic pain in UK biobank // *PLoS Genet.* 2019.
72. Sudlow C. et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age // *PLOS Med.* 2015.
73. Fry A. et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population // *Am. J. Epidemiol.* 2017.
74. Herzig A.F. et al. Strategies for phasing and imputation in a population isolate // *Genet. Epidemiol.* 2018.
75. Welsh S. et al. Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort // *BMC Genomics.* 2017.
76. Bycroft C. et al. The UK Biobank resource with deep phenotyping and genomic data // *Nature.* 2018.
77. Loh P.R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts // *Nat. Genet.* 2015.
78. Willer C.J., Li Y., Abecasis G.R. METAL: Fast and efficient meta-analysis of genomewide association scans // *Bioinformatics.* 2010.
79. Canela-Xandri O., Rawlik K., Tenesa A. An atlas of genetic associations in UK Biobank. // *Nat. Genet.* 2018.
80. Hemani G. et al. The MR-Base platform supports systematic causal inference across the human phenome // *Elife.* 2018.
81. Yang J. et al. GCTA: a tool for genome-wide complex trait analysis. // *Am. J. Hum. Genet.* 2011.
82. Benjamini Y., Hochberg Y. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc.* 1995;B57:289–300. // *J. R. Stat. Soc. B.* 1995.
83. Chang C.C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets // *Gigascience.* 2015.
84. McLaren W. et al. The Ensembl Variant Effect Predictor // *Genome Biol.* 2016.
85. Rogers M.F. et al. FATHMM-XF: Accurate prediction of pathogenic point mutations via extended features // *Bioinformatics.* 2018.
86. Ferlino M. et al. An integrative approach to predicting the functional effects of small indels in non-coding regions of the human genome // *BMC Bioinformatics.* 2017.
87. Haller G. et al. A missense variant in SLC39A8 is associated with severe idiopathic scoliosis // *Nat. Commun.* 2018.

88. Li D. et al. A Pleiotropic Missense Variant in SLC39A8 Is Associated With Crohn's Disease and Human Gut Microbiome Composition // *Gastroenterology*. 2016.
89. Carrera N. et al. Association study of nonsynonymous single nucleotide polymorphisms in schizophrenia // *Biol. Psychiatry*. 2012.
90. Kong L. et al. Interaction between cartilage oligomeric matrix protein and extracellular matrix protein 1 mediates endochondral bone growth // *Matrix Biol*. 2010.
91. Fisher S.A. et al. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease // *Nat. Genet*. 2008.
92. Festen E.A.M. et al. Genetic analysis in a dutch study sample identifies more ulcerative colitis susceptibility loci and shows their additive role in disease risk // *Am. J. Gastroenterol*. 2010.
93. Adali G. et al. Extracellular matrix protein 1 gene rs3737240 single nucleotide polymorphism is associated with ulcerative colitis in Turkish patients // *Turkish J. Gastroenterol*. 2017.
94. Egli R.J. et al. Functional analysis of the osteoarthritis susceptibility-associated GDF5 regulatory polymorphism // *Arthritis Rheum*. 2009.
95. Miyamoto Y. et al. A functional polymorphism in the 5' UTR of GDF5 is associated with susceptibility to osteoarthritis // *Nat. Genet*. 2007.
96. Southam L. et al. An SNP in the 5'-UTR of GDF5 is associated with osteoarthritis susceptibility in Europeans and with in vivo differences in allelic expression in articular cartilage // *Hum. Mol. Genet*. 2007.
97. Chapman K. et al. A meta-analysis of European and Asian cohorts reveals a global role of a functional SNP in the 5' UTR of GDF5 with osteoarthritis susceptibility // *Hum. Mol. Genet*. 2008.
98. Evangelou E. et al. Large-scale analysis of association between GDF5 and FRZB variants and osteoarthritis of the hip, knee, and hand // *Arthritis Rheum*. 2009.
99. Dai J. et al. Association of a single nucleotide polymorphism in growth differentiate factor 5 with congenital dysplasia of the hip: A case-control study // *Arthritis Res. Ther*. 2008.
100. Rouault K. et al. Evidence of association between GDF5 polymorphisms and congenital dislocation of the hip in a Caucasian population // *Osteoarthr. Cartil*. 2010.
101. Almutiri S. et al. Non-viral-mediated suppression of AMIGO3 promotes disinhibited NT3-mediated regeneration of spinal cord dorsal column axons // *Sci. Rep*. 2018.
102. Okifuji A., Hare B.D. The association between chronic pain and obesity // *Journal of Pain Research*. 2015.

103. Gasperi M. et al. Genetic and Environmental Influences on Sleep, Pain, and Depression Symptoms in a Community Sample of Twins // *Psychosom. Med.* 2017.
104. REICHBORN-KJENNERUD T. et al. Back–neck pain and symptoms of anxiety and depression: a population-based twin study // *Psychol. Med.* 2002.
105. McIntosh A.M. et al. Genetic and Environmental Risk for Chronic Pain and the Contribution of Risk Variants for Major Depressive Disorder: A Family-Based Mixed-Model Analysis // *PLoS Med.* 2016.
106. Zhou J.J. et al. Integrating multiple correlated phenotypes for genetic association analysis by maximizing heritability // *Hum. Hered.* 2015.
107. Ning Z. et al. Beyond power: Multivariate discovery, replication, and interpretation of pleiotropic loci using summary association statistics // *bioRxiv.* 2019.
108. Turley P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG // *Nat. Genet.* 2018.
109. Vargha-Khadem F. et al. FOXP2 and the neuroanatomy of speech and language // *Nature Reviews Neuroscience.* 2005.
110. Fisher S.E., Scharff C. FOXP2 as a molecular window into speech and language // *Trends in Genetics.* 2009.
111. Deckers M.M.L. et al. Recombinant human extracellular matrix protein 1 inhibits alkaline phosphatase activity and mineralization of mouse embryonic metatarsals in vitro // *Bone.* 2001.
112. Tachmazidou I. et al. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data // *Nature Genetics.* 2019.
113. Kim J.H. et al. Regulation of the catabolic cascade in osteoarthritis by the zinc-ZIP8-MTF1 axis // *Cell.* 2014.
114. Freidin M.B. et al. Insight into the genetic architecture of back pain and its risk factors from a study of 509,000 individuals // *Pain.* 2019.
115. Zhou W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies // *Nat. Genet.* 2018.

Приложения

Онлайн приложения включают в себя следующие материалы:

- Приложение А_Региональные графики ассоциаций;
- Приложение В_Статистики ПГАА ассоциированных с ГКМГК локусов;
- Приложение С_Результаты приоритизации на основе литературной аннотации;
- Приложение D_Признаки в GWAS-MAP;
- Приложение E_Результаты условного и совместного анализа COJO;
- Приложение F_Информация о хорошо изученных SNP, ассоциированных с ГКМГК;
- Приложение G_Предсказание эффектов локусов, ассоциированных с ГКМГК;
- Приложение H_Результаты DEPICT для ГКМГК1;
- Приложение J_Анализ генетических корреляций ГКМГК.

Перечисленные материалы будут доступны онлайн в течение года по ссылке <https://www.dropbox.com/sh/h22yckk5f80o8kr/AAAN6z2VLuQQS1yLZQFRNSOa?dl=0>. По завершении указанного срока действие ссылки будет прекращено, и ознакомиться с сопроводительными данными можно будет отправив запрос автору работы.