

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ
УЧРЕЖДЕНИЕ «ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
ИНСТИТУТ ЦИТОЛОГИИ И ГЕНЕТИКИ СИБИРСКОГО ОТДЕЛЕНИЯ
РОССИЙСКОЙ АКАДЕМИИ НАУК»

На правах рукописи

ЗЛОБИН АЛЕКСАНДР СЕРГЕЕВИЧ

**Изучение генетического контроля мясной
продуктивности овец с использованием современных
методов количественной генетики**

Генетика - 1.5.7.

Диссертация на соискание ученой степени кандидата
биологических наук

Научный руководитель:

к.б.н. Я.А. Цепилов

НОВОСИБИРСК - 2022

Оглавление

Оглавление.....	2
Список сокращений	5
Введение.....	6
Актуальность темы исследования.....	6
Степень разработанности темы исследования.....	9
Цели и задачи	10
Научная новизна	11
Теоретическая и научно-практическая ценность	11
Методология и методы диссертационного исследования	12
Степень достоверности результатов	13
Положения, выносимые на защиту	13
Структура и объём работы.....	14
Личный вклад автора	14
Апробация результатов	14
Публикации по теме работы	15
1. Обзор литературы	17
1.1 Методы селекции	17
1.1.1 Классическая селекция.....	18
1.1.2 Маркер-ориентированная селекция.....	19
1.1.3 Геномная селекция	23
1.2 Овцеводство, современное состояние.	28
1.2.1 Овцеводство в России, общие положения	29

1.2.2	Российские мясные породы овец	30
1.2.3	Мясная селекция в овцеводстве	30
1.2.4	Геном овец, общие положения.....	32
1.3	Полногеномное исследование ассоциаций	33
1.3.1	Статистические модели ПГИА.....	34
1.3.2	Анализ результатов ПГИА	39
1.3.3	ПГИА в мясном овцеводстве.....	42
1.4	Решения для хранения и анализа результатов ПГИА.....	46
1.4.1	Платформа GWAS-MAP	49
1.5	Краткое заключение	50
2.	Материалы и методы	52
2.1	Устройство платформы GWAS-MAP ovis.....	52
2.2	Поиск и анализ статей для создания базы данных QTL и пополнения платформы GWAS-MAP ovis	52
2.3	Суммарные статистики для многомерного анализа.....	54
2.4	MANOVA и многомерные модели.....	54
2.5	Функциональная аннотация и приоритизация локусов, найденных при многомерном анализе	55
2.5.1	Функциональная сеть взаимодействий генов, ассоциированных с мясной продуктивностью	56
2.6	Выборка животных российской популяции для оценки племенной ценности признаков мясной продуктивности и подтверждения найденных локусов	56
2.6.1	Контроль качества генотипов.....	57
2.6.2	Контроль качества фенотипов.....	59

2.6.3 ПГИА на российской выборке	60
2.6.4. Подтверждение локусов из базы данных QTL	60
2.7 Анализ колокализации	61
3. Результаты	63
3.1 Создание платформы для комплексного анализа результатов полногеномных исследований ассоциаций	63
3.1.1 Создание каталога неравновесия по сцеплению	63
3.1.2 Формирование структуры платформы для хранения результатов ПГИА	65
3.2 Создание базы данных QTL.....	67
3.3 Многомерный анализ ассоциаций признаков мясной продуктивности из платформы GWAS-MAP ovis	70
3.3.1 Результаты многомерного анализа	71
3.3.2 Приоритизация генов в найденных локусах	72
3.3.3 Функциональная сеть взаимодействий ассоциированных локусов	74
3.3.4 Подтверждение найденных локусов на российской популяции овец.....	76
3.4 Анализ колокализации	77
3.5 Оценка племенной ценности признаков мясной продуктивности на российской выборке овец.....	79
4. Обсуждение	83
5. Заключение	89
6. Выводы.....	91
7. Библиографический список	93
Приложение	112

Список сокращений

ОНП, SNP – однонуклеотидный полиморфизм (Single Nucleotide Polymorphism)

КРС – крупный рогатый скот

МОС – маркер-ориентированная селекция

ГС – геномная селекция

ПЦ – племенная ценность (Breeding value, BV)

ПГИА – полногеномное исследование ассоциаций

QTL – локус количественного признака (Quantitative Trait Loci)

VEP - Variant Effect Predictor, программное обеспечение для функциональной аннотации локусов количественных признаков

БД – база данных

СУБД – система управления базами данных

Введение

Актуальность темы исследования

Селекция домашних животных занимает большое место в научных исследованиях различных государств. Создание новых и улучшение уже существующих пород животных является большой и важной задачей не только генетики, но и молекулярной биологии, зоотехнологии и других смежных наук. Как правило, необходимые для улучшения пород признаки контролируются не одним, а несколькими генами. Например, количество получаемого молока, его жирность и содержание белка у коров Голштинской породы контролируется более чем тридцатью локусами [1]. Изучение таких сложных и комплексных признаков позволяет получать новые знания о биологических системах, которые участвуют в формировании этих признаков, тем самым увеличивая потенциал и возможности для селекции.

На данный момент в селекции используют два ключевых подхода – маркер-ориентированная селекция (МОС) и геномная селекция (ГС) [2, 3]. Термин МОС применяется к методам, которые используют разнообразные молекулярные маркеры, в том числе ДНК-маркеры для достижения результата. [4]. ГС ориентирована на использование генома родителей для предсказания фенотипа потомства с помощью оценки племенной ценности (ПЦ) [5]. Для реализации этих подходов необходимы знания о генах и маркерах, ассоциированных с хозяйственно-полезными признаками у различных сельскохозяйственных животных.

По интенсивности генетических исследований существует заметный сдвиг в сторону отдельных видов. Например, для коров известны сотни тысяч маркеров, которые ассоциированы с тысячами экономически важных признаков, связанных с мясной и молочной продуктивностью, фертильностью, здоровьем животных и др. [6]. Благодаря этому на коровах эффективно используется ГС для увеличения количественных показателей [7].

Однако, для многих других видов сельскохозяйственных животных, таких как свиньи, козы и овцы, о генетическом контроле полезных для сельского хозяйства признаков известно в разы меньше.

Одной из ключевых задач в селекции сельскохозяйственных животных является увеличение мясной продуктивности. Эта задача особенно актуальна для овец, так как долгое время основным продуктом овцеводства была шерсть. В настоящее время для увеличения мясной продуктивности у овец используется несколько десятков маркеров, самые известные из которых *MSTN*, *GH*, *CLPN1*, *CAST* и др. Необходимы новые исследования для расширения списка маркеров, которые потенциально могут использоваться в МОС, так как по многим известным маркерам возможно достижение селекционного плато в ближайшие десятилетия.

Одним из самых популярных методов для картирования генов и локусов количественных признаков (QTL), ассоциированных с хозяйственно полезными признаками, является полногеномное исследование ассоциаций (ПГИА). Как правило, каждое такое исследование проводится на ограниченном числе животных определенной породы. Для того, чтобы сделать адекватные выводы о генетическом контроле того или иного признака, необходимо аккумулировать все знания, полученные в разрозненных исследованиях.

Несмотря на большие успехи применения ПГИА с целью поиска QTL хозяйственно полезных признаков, широкомасштабный анализ и использование ПГИА в селекции имеют ряд сложностей и ограничений. Во-первых, результаты ПГИА зачастую занимают большое количество места на физических носителях, что предполагает наличие развитой инфраструктуры для хранения и работы с такими объемами информации. Во-вторых, на данный момент существует множество протоколов проведения ПГИА. Это делает необходимым проведение контроля качества результатов и перевода данных в

универсальный формат. Третья сложность касается исследований в области генетики животных, где, в отличие, от генетики человека, не принято выкладывать в открытый доступ результаты ПГИА в виде суммарных статистик. В области генетики животных не существует платформ или баз данных, которые бы хранили полные результаты ПГИА, а также предоставляли различные модули анализа для работы с ними. Существуют базы данных, в которых хранятся списки локусов с показанной ассоциацией, например, «AnimalQTLdb» [6], «iSheep» [8] и др. Такая усеченная информация не позволяет сопоставлять различные породы, проводить мета-анализ, а также использовать новые методы анализа такие как региональный анализ ассоциаций [9, 10].

В области исследований генетики человека для решения подобных задач разрабатывают специальные платформы и базы данных. Одной из таких является платформа GWAS-MAP [11], которая позволяет хранить полные результаты ПГИА различных признаков человека в виде суммарных статистик, визуализировать их, анализировать как совместно, так и по отдельности. Создание подобной платформы для овец не только позволит аккумулировать все имеющиеся знания о генетике различных признаков, но и обеспечит проведение фундаментальных исследований и создаст базу для проведения МОС по хозяйственно важным признакам.

Помимо канонических методов проведения ПГИА, на данный момент в генетике человека и крупного рогатого скота широкое распространение получили так называемые многомерные модели анализа. Такие модели, основанные на совместном анализе нескольких признаков одновременно, обладают большей мощностью по сравнению с обычным анализом ассоциаций. Одним из преимуществ многомерных моделей является возможность их применения на данных результатов ПГИА, находящихся в открытом доступе [12], что широко используется в генетике человека [13, 14] и животных [15, 16]. Использование многомерных моделей на уже

накопленных данных по ПГИА различных признаков у овец позволит увеличить количество новых QTL.

Увеличение количества известных QTL благодаря применению разнообразных современных методов, валидация уже известных вариантов, а также наличие специализированной базы данных и платформы для признаков мясной продуктивности у овец – все это необходимые шаги для создания алгоритмов проведения МОС и увеличения фенотипических показателей различных российских и зарубежных пород овец.

Степень разработанности темы исследования

В настоящее время для увеличения мясной продуктивности у овец используется несколько десятков маркеров, самые известные из которых *MSTN*, *GH*, *CLPNI*, *CAST* и др. Существует несколько десятков работ, выполненных с использованием современных методов количественной генетики (например, полногеномные исследования ассоциаций), изучающих генетический контроль мясной продуктивности у овец. Крупнейшей на данный момент является работа S. Volortmaa с соавторами [16], в которой были представлены 56 признаков, связанных с мясной продуктивностью для 10613 животных различных пород. Также эта работа на данный момент является единственной для признаков мясной продуктивности у овец, в которой были опубликованы суммарные статистики ПГИА. В остальных работах зачастую не содержится информации о направлении эффекта, о референсном и эффекторном аллелях, что делает использование на практике результатов данных работ ограниченным или невозможным.

В настоящее время основным решением для работы с локусами количественных признаков животных является база данных “AnimalQTLdb”, которая позволяет искать локусы, ассоциированные с заданными признаками. Однако, с помощью “AnimalQTLdb” нельзя получить всю доступную информацию для найденных локусов, например, информацию о хромосоме,

позиции, размере эффекта, направлении эффекта и др. Пользователю показывается лишь набор публикаций, откуда взята информация. Пользователю необходимо самостоятельно искать в каждой статье информацию о направлении эффекта, аллелях и т.д., если эта информация в статье вообще присутствует. “AnimalQTLdb” эффективна для поиска генов, находящихся в локусах, ассоциированных с искомыми признаками, но не для поиска полной информации о локусе, которая может быть напрямую применена для дальнейших анализов и МОС.

Цели и задачи

Целью настоящей работы является поиск новых локусов, контролирующих признаки, относящиеся к мясной продуктивности овец, с использованием современных методов количественной генетики

Для достижения данной цели были поставлены следующие задачи:

- 1) Создание базы данных, аккумулирующей информацию о генах и QTL, связанных с показателями мясной продуктивности овец.
- 2) Создание платформы для унификации и хранения результатов полногеномного исследования ассоциаций различных признаков овец и для получения новых знаний о генетике этих признаков с использованием современных генетических подходов.
- 3) Поиск новых локусов, ассоциированных с признаками мясной продуктивности, овец, с помощью полногеномного исследования ассоциаций многомерных признаков.
- 4) Выявление плейотропных эффектов локусов, ассоциированных с различными признаками мясной продуктивности овец, с помощью анализа колокализации.
- 5) Подтверждение известных локусов и создание моделей оценки племенной ценности животных по признакам мясной продуктивности на российской выборке овец.

Научная новизна

В данной работе была создана платформа GWAS-MAP_{ovis} которая является уникальным на данный момент решением для хранения, унификации и анализа результатов ПГИА различных признаков овец. Веб-интерфейс платформы находится в открытом доступе по адресу <https://pheligoovis.icgbio.ru/>

С помощью созданной платформы был проведен многомерный анализ ранее опубликованных данных, в результате которого было обнаружено 12 локусов (из них 8 новых), ассоциированных с признаками мясной продуктивности у овец. Для новых 8 локусов было приоритизировано 13 генов. Один из 8 новых локусов (расположенный в районе генов *FAM3C* и *WNT16*) был подтвержден с использованием российской выборки овец. Для 12 найденных локусов было выявлено два связанных графа функционального взаимодействия. Первый состоит из 3 ранее известных локусов (гены *LCORL*, *SLC16A11* и *GHR*) и одного нового (*FAM3C/WNT16*). Второй состоит из ранее известного локуса *MEG8_2* и нового *SHISAL1*. Это позволяет выдвинуть гипотезу о функциональной и биологической связанности этих генов через механизмы зависимой регуляции экспрессии.

Была показана эффективность платформы для проведения различных фундаментальных генетических анализов, включая анализ колокализации, а также применимость для поиска потенциальных маркеров для создания моделей маркер-ориентированной селекции. Впервые в мире был проведен анализ колокализации для овец.

Теоретическая и научно-практическая ценность

Обнаруженные с помощью многомерного анализа локусы расширяют представление о генетической регуляции признаков мясной продуктивности у овец. Полученная в результате сеть взаимодействия генов позволяет выдвинуть гипотезу о функциональной и биологической связанности этих

генов через механизмы зависимой регуляции экспрессии. Этот результат может быть применен в дальнейших исследованиях путей и механизмов увеличения фенотипических показателей признаков мясной продуктивности у овец.

На российской популяции овец удалось подтвердить 2 локуса: rs401990068 (гены *FAM3C* и *WNT16*) и rs401834107 (ген *LCORL*). Оба этих локуса могут быть напрямую использованы для проведения МОС на российских популяциях овец.

Как один из необходимых инструментов, создана платформа GWAS-MAP|ovis, которая может быть использована для хранения и анализа результатов ПГИА различных признаков овец. Платформа позволяет всем заинтересованным исследователям проводить различные типы анализов, таких как визуализация результатов, совместный анализ (мета-анализ), анализ колокализации, многомерный анализ и другие. Далее, информация из платформы GWAS-MAP|ovis может быть напрямую применена селекционерами для создания моделей оценки племенной ценности и моделей маркер-ориентированной селекции.

В дополнение к созданной платформе GWAS-MAP|ovis была создана база данных локусов и генов, ассоциированных с признаками мясной продуктивности у овец. Данная база будет полезна как исследователям для подтверждения полученных результатов анализов картирования на локальных популяциях, так и селекционерам для выбора ключевых маркеров для проведения МОС.

Методология и методы диссертационного исследования

В рамках изучения генетического контроля мясной продуктивности овец были использованы различные современные методы количественной генетики, такие как многомерный анализ, анализ колокализации, построение

функциональной сети взаимодействия генов. Для многомерного анализа использовалась группировка фенотипов.

Биоинформатический анализ полученных результатов проводился с использованием метода приоритизации функциональных вариантов VEP, а также с помощью использования литературных данных.

В качестве основы для создания платформы GWAS-MAP|ovis в данной работе использовалось программное обеспечение и информационное окружение платформы GWAS-MAP для человека [11].

Степень достоверности результатов

Степень достоверности результатов многомерного анализа подтверждается согласованностью с ранее опубликованными данными и подтверждением найденных генетических ассоциаций с помощью многомерного анализа на независимой выборке российской популяции овец. Созданная в рамках исследования платформа GWAS-MAP|ovis была апробирована на признаках мясной продуктивности у овец. Это позволило показать эффективность использования созданной платформы для унификации, хранения и анализа результатов ПГИА и в перспективе различных экономически важных признаков овец.

Положения, выносимые на защиту

- 1) Локусы, расположенные на хромосомах 1 (rs193632759), 2 (rs428034699), 3 (rs428034699, rs403766990, rs399851221) 4 (rs401990068), 5 (rs408893215) и 23 (rs418394153) овец, ассоциированы с признаками мясной продуктивности.
- 2) Гены SLC16A11, GHR, MEG8_2 и SHISAL1, расположенные в локусах на хромосомах 11, 16, 18 и 3, соответственно, и гены LCORL, FAM3C/WNT16 на хромосоме 6, приоритезированные в качестве генов-кандидатов, вовлеченных в генетический контроль признаков мясной продуктивности овец, функционально связаны.

- 3) Ген LCORL (ОНП rs401834107) имеет плейотропный эффект на признаки, связанные с показателями мясной продуктивности и содержанием жировой прослойки у овец.

Структура и объём работы

Работа состоит из введения, обзора литературы, материалов и методов, результатов, обсуждения, заключения, выводов, списка литературы (137 источников) и приложения. Объём работы составляет 111 страниц. В работе представлено 6 таблиц и 17 рисунков.

Личный вклад автора

Основные результаты, изложенные в диссертации, получены и проанализированы автором лично. Программное обеспечение для создания платформы GWAS-MAP|ovis было предоставлено коллегами из МФТИ. Создание веб-интерфейса и формирование программной структуры платформы было проведено коллегами из лаборатории рекомбинационного и сегрегационного анализа ИЦиГ СО РАН. Создание каталога неравновесия по сцеплению, наполнение платформы и ее тестирование, создание базы данных QTL и генетический анализ проводились автором лично.

Апробация результатов

Материалы настоящей работы вошли в отчет по гранту Российского Научного Фонда № 18-16-00079 «Полногеномный анализ ассоциаций показателей роста, развития и частоты рекомбинации в ресурсных популяциях рода *Ovis*, полученных при межвидовом и межпородном скрещивании». Результаты работы были представлены лично автором на 6 международных научных конференциях в виде 1 устного и 5 стендовых докладов:

1. Alexander S. Zlobin, Natalia A. Volkova, Natalia A. Zinovieva, Baylar S. Iolchiev, Vugar A. Bagirov, Pavel M. Borodin, Tatiana I. Aksenovich, Yakov A. Tsepilov, "Genome-wide association study on the Russian population of

- sheep identifies eight novel loci associated with meat productivity traits”, 11-16 October 2021, First ESDAR-ECAR Online Conference
2. Alexander S. Zlobin et al. , «A platform for storage and analysis of results of genome-wide association studies of sheep», 10th Moscow Conference on Computational Molecular Biology MCCMB21 30 July- 02 August 2021, Moscow, Russia
 3. А. С. Злобин и др., «Поиск новых локусов, ассоциированных с признаками мясной продуктивности у овец (ovis aries), с помощью многомерного анализа», «Генетика, селекция и биотехнология животных: на пути к совершенству», 13-15 октября 2020 г., г. Пушкин
 4. Zlobin Alexander S., et al. «Multivariate analysis identify new loci associated with meat productivity and carcass traits in sheeps (ovis aries)», Bioinformatics of genome regulation and structure/systems biology (bgrs/sb-2020) 06-10 July 2020, Novosibirsk, Russia
 5. Zlobin Alexander S., et al. «A platform for storage and analysis of results of genome-wide association studies of sheep», Bioinformatics of genome regulation and structure/systems biology (bgrs/sb-2020) 06-10 July 2020, Novosibirsk, Russia
 6. А. С. Злобин и др., «Создание базы данных генетических вариантов, ассоциированных с признаками роста, мясной продуктивности и формирования каркаса у овец (Ovis aries)», Достижения в генетике, селекции и воспроизводстве сельскохозяйственных животных, 28-29 мая 2019, Санкт-Петербург

Публикации по теме работы

По материалам диссертации опубликовано три научные работы в изданиях, индексируемых в базах данных «Скопус» (Scopus) и «Сеть науки» (Web of Science) и зарегистрировано две базы данных:

1. Анатолий В. Кириченко, **Александр С. Злобин**, Татьяна И. Шашкова, Наталья А. Волкова, Байлар С. Иолчиев, Вугар А. Багиров, Павел М.

- Бородин, Леннарт С. Карссен, Яков А. Цепилов & Юрий С. Аульченко
Платформа GWAS-MAP|ovis для хранения и анализа результатов
полногеномных ассоциативных исследований овец. Вавиловский
журнал генетики и селекции, Том 26, № 4 (2022)
2. **Zlobin, A.S.**; Nikulin, P.S.; Volkova, N.A.; Zinovieva, N.A.; Iolchiev, B.S.;
Bagirov, V.A.; Borodin, P.M.; Aksenovich, T.I.; Tsepilov, Y.A. Multivariate
Analysis Identifies Eight Novel Loci Associated with Meat Productivity
Traits in Sheep. *Genes* 2021, 12, 367. <https://doi.org/10.3390/genes12030367>
 3. **Zlobin, A.S.**; Volkova, N.A.; Borodin, P.M.; Aksenovich, T.I.; Tsepilov,
Y.A. Recent advances in understanding genetic variants associated with
growth, carcass and meat productivity traits in sheep (*Ovis aries*): An update.
Arch. Anim. Breed. 2019, 62, 579–583
 4. Зарегистрирована база данных в государственном реестре российских
программ для электронных вычислительных машин и баз данных
«Генетические маркеры признаков роста, каркаса и мясной
продуктивности у овец» (ГМРиПО) с присвоенным номером
2019621453
 5. Платформа GWAS-MAP|ovis зарегистрирована в государственном
реестре российских программ для электронных вычислительных машин
и баз данных под названием «База данных генетических ассоциаций
признаков овец (ГАПО)» с присвоенным номером 2021620564

1. Обзор литературы

1.1 Методы селекции

С внедрением сельского хозяйства и одомашнивания животных человечество стало искать пути улучшения показателей урожайности, удоев и т.д. В основе первых представлений о селекции лежал искусственный отбор. Люди отбирали лучшие, по их мнению, плоды для дальнейшей посадки, лучших особей животных для дальнейшего воспроизводства, считая, что произойдет воспроизведение ранее полученного хорошего результата. В дальнейшем, еще до открытия основных законов генетики, люди стали использовать более осознанный и целенаправленный отбор, не просто пытаясь воспроизвести ранее полученный положительный результат, но и делая попытки скрещивать различные организмы с необходимыми свойствами. Одним из первопроходцев в области становления селекции, как научной дисциплины, стал Роберт Бейкуэлл [17]. В рамках британской сельскохозяйственной революции он во второй половине XVIII века занимался разведением овец. С помощью отбора лучших он смог улучшить породу овец Линкольн. Затем на основе данной породы Бейкуэлл вывел совершенно новую породу Лейстер.

С открытием Менделем первых законов генетики и дальнейшим развитием генетики, как науки в целом, произошло и развитие селекции, и ее обособление в отдельную науку. В основе селекции лежат идеи эволюции, предложенные Чарлзом Дарвином во второй половине XIX века, такие как естественный и искусственный отбор, а также наследственность и изменчивость. На основе этих идей и с использованием знаний о генетике, произошло бурное развитие селекции, начиная с 20-х годов XX века. Огромный вклад в это развитие внес британский ученый Роналд Фишер, который в своих работах смог соединить математические основы открытий Менделя и теорию естественного отбора. Его идеи впоследствии стали

основой популяционной генетики, которая широко используется в рамках селекции [18]. Также нельзя не отметить Николая Ивановича Вавилова. Он не только открыл закон гомологических рядов, но и написал «Теоретические основы селекции», в которых описал методы, так называемой, «классической селекции» [19].

1.1.1 Классическая селекция

Если рассматривать классическую селекцию в широком смысле, то, несомненно, основным методом будет являться искусственный отбор. Его идея состоит в отборе лучших особей по выбранным показателям и затем скрещивание отобранных особей. Однако, если рассматривать классическую селекцию как более узкий термин, возникший во времена становления селекции как научной дисциплины, то, несомненно, наряду с искусственным отбором, основным методом также является и гибридизация. Гибридизация - скрещивание особей с отобранными фенотипами для получения особей с улучшенными характеристиками. Гибридизация бывает межвидовая и внутривидовая. Классическими примерами межвидовой гибридизации являются гибриды пшеницы и ржи у растений, мулы (гибрид кобылы и осла) и лошаки (гибриды коня и ослицы). Гибриды, полученные путем межвидового скрещивания, зачастую бесплодны, однако у растений для преодоления этой проблемы используют полиплоидию [20]. При полиплоидии происходит увеличение набора хромосом, что позволяет не только избежать бесплодия у гибридов, но и зачастую повысить урожайность. Также среди растений зачастую распространено близкородственное скрещивание, при котором, например, происходит самоопыление растения и повышается гомозиготность, однако часто такое скрещивание снижает жизнеспособность и продуктивность полученного гибрида.

Проблема получения межвидовых гибридов у животных широко известна – в большинстве случаев гибриды остаются бесплодны. Это происходит вследствие разного числа хромосом у родительских форм. Однако

при скрещивании близких к друг другу в систематическом плане видов, зачастую, потомство плодовито. В целом наблюдается прямая зависимость от филогенетической близости родительских форм и плодовитости полученных гибридов. Существует также ряд проблем цитогенетического характера, влияющих на плодовитость гибридов. Например, несовместимость цитоплазмы яйцеклетки и ядра сперматозоида при оплодотворении или несовместимость антигенов сперматозоида с антителами самок при спаривании [21].

1.1.2 Маркер-ориентированная селекция

Активное применение гибридизации началось более 100 лет назад, однако, в связи с отсутствием представления об устройстве генома в целом, генов и локусов в частности, анализ хозяйственно важных признаков был весьма затруднен. В качестве простейшей модели использовали соотношение морфологических признаков с исследуемыми признаками. Например, у растений таким признаком является моногенно наследуемая окраска цветка. Но вскоре было обнаружено, что таких признаков немного, и селекция с использованием вспомогательных морфологических маркеров не получила дальнейшего развития. Вместе с тем, развитие генетики позволило получить информацию о том, что такое ген, локус, и дать толчок новому витку в селекции растений и животных – маркер-ориентированной селекции (МОС) [22].

Термин МОС применяется к методам селекции, которые используют разнообразные молекулярные маркеры (синоним ДНК-маркеры) для отбора особей. В настоящее время существует более 10 разнообразных типов молекулярных маркеров. Наиболее используемые из них разделяют на 3 группы в зависимости от основного метода анализа, с помощью которого данные маркеры исследуют: 1) Маркеры блот-гибридизации 2) Маркеры ПЦР 3) Маркеры с ДНК-чипов (Рисунок 1) [23]. В свою очередь, каждая из представленных групп делится также на монолокусные и мультилокусные.

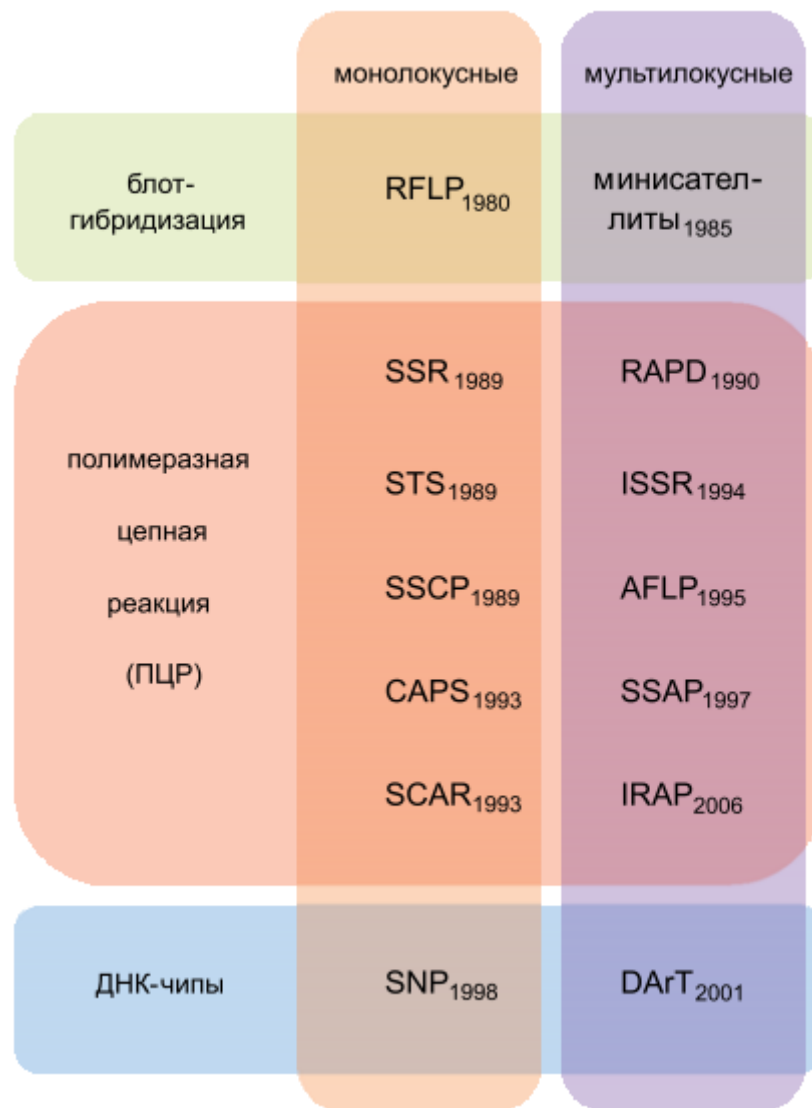


Рисунок 1. Схематическая классификация различных групп ДНК-маркеров и год их первого упоминания в литературе. Пояснения основных групп, показанных на рисунке, представлено в тексте. Взято из [23].

Эволюция использования молекулярных маркеров начинает свою историю с началом использования первой группы маркеров, а именно маркеров блот-гибридизации. Данный тип маркеров получил широкое распространение в 80-х годах двадцатого века. Применение RFLP (restriction fragment length polymorphism – полиморфизм длины рестрикционных фрагментов) было впервые показано Д. Ботштэйном в 1980 году и применено для построения RFLP-генетической карты человека [24]. В целом, использование RFLP позволяет находить различия в нуклеотидных

последовательностях, если они расположены в сайтах рестрикции. Широкое распространение данный метод получил в селекции растений, для получения клонов генов устойчивости к различным заболеваниям и устойчивости к различным видам стресса [25]. Также, с помощью использования RFLP-карт было выполнено сравнительное картирование различных геномов, что позволило установить эволюционные преобразования структуры геномов отдельных видов. Эти результаты стали основой современных исследований в сравнительной геномике [26]. В свою очередь, использование мультилокусных исследований с помощью блот-гибридизации, а именно минисателлитов (коротких повторяющихся последовательностей ДНК от 9 до приблизительно 100 нуклеотидов) получило широкое распространение как, например, в селекции крупного рогатого скота [27], так и в селекции растений [28].

Несмотря на большой прорыв с момента открытия и начала использования RFLP, данный метод не получил широкого использования из-за дороговизны и необходимости использования радиоактивно меченных проб. Методы с использованием ПЦР-маркеров напротив получили широкое распространение из-за своей дешевизны и простоты эксперимента. Микросателлитные маркеры (SSR) являются классическим примером ПЦР-маркеров. Впервые использование данных гипервариабельных последовательностей в качестве маркеров было предложено немцем Дю Таутцом в 1989 году [29]. Выход на рынок микрочипов для генотипирования по десяткам тысяч ОНП в геноме дал толчок к широкому использованию ДНК-чипов для проведения МОС. В настоящий момент — это технологически простой и дешевый способ обнаружения ОНП маркеров для проведения МОС [30].

Представленные выше методы позволили ДНК-маркерам стать одними из основных инструментов селекционеров. Произошла смена источника

информации для отбора - если раньше отбор проводился по фенотипу, то с появлением ДНК-маркеров стал возможен отбор по генотипу.

Имеется целый ряд преимуществ отбора по генотипу в сравнении с отбором по фенотипу: 1) процесс генотипирования зачастую либо частично, либо полностью автоматизирован, в то время как автоматический анализ фенотипов применяется достаточно редко и в очень конкретных случаях; 2) в отличие от фенотипов информацию о генотипах можно отбирать в разное время жизни животного и растений (от пренатальной диагностики до взрослых особей); 3) фенотипы сильно подвержены влиянию окружающей среды, в то время как информация о генотипах неизменна в процессе жизни организма. Благодаря данным преимуществам использование ДНК-маркеров стало основой селекционной работы по всему миру.

МОС может быть проведена различными способами и применена для решения широкого спектра селекционных задач. Используются различные комбинации маркеров для предсказания фенотипов, например, возможно использование не одного маркера, а нескольких, которые расположены по разные стороны от интересующего нас гена (фланкирующие маркеры). Зачастую маркеры расположены на значительном расстоянии от интересующего нас гена, тогда используют сочетание маркер-ориентированной селекции и классического отбора по фенотипу. В целом, МОС может быть применена в случаях линейной селекции, беккроссной селекции и при создании пирамид генов. В случае линейной селекции с помощью ДНК-маркеров отбираются организмы непосредственно перед скрещиваниями, что позволяет исключить нежелательные генотипы и получить необходимый результат гораздо быстрее. В случае возвратного (беккроссного) скрещивания происходит скрещивание гибридов первого поколения с родительскими формами. Благодаря использованию ДНК-маркеров возможен отбор необходимых организмов на каждом из этапов скрещиваний, что позволяет закрепить необходимый ген в гибриде и вернуть

генотип родительской формы в оставшейся части генома. Создание пирамид генов - это один из методов селекции, позволяющий с использованием ДНК-маркеров вести отбор по схожим признакам, за которые ответственны несколько генов: например, отбор по устойчивости к патогенам, поражающим один и тот же орган растения [31].

1.1.3 Геномная селекция

Использование ДНК-маркеров для МОС стало возможно с открытием новых методов анализа генетической информации, начиная с RFLP и заканчивая высокопроизводительными ДНК-чипами. В последнее десятилетие широкое распространение стали получать методы высокопроизводительного секвенирования, которые позволили реализовать множество программ по полногеномному секвенированию различных организмов. В частности, начиная с 2003 года, когда была получена информация о полной последовательности генома собаки, стало возможно создание таких комбинаций ДНК-маркеров (чаще всего ОНП), которые плотно и равномерно покрывают геном выбранного организма. На основе информации о множестве ОНП стал возможен переход от маркер-ориентированной к геномной селекции.

В отличие от применяемого в последнее время в МОС метода полногеномного анализа ассоциаций (ПГИА), целью которого является установить связь маркер-признак и найти так называемый локус количественного признака (Quantitative trait loci, QTL), геномная селекция применяется для предсказания фенотипа потомства на основе полученной генетической информации [32]. В основе ГС лежит предположение, что многие QTL находятся в неравновесии по сцеплению с хотя бы одним другим маркером [33]. С теоретической точки зрения геномная селекция может быть рассмотрена как регрессионный анализ всего генома, когда эффекты всех маркеров оцениваются одновременно [34]. Впервые метод ГС был применен в

молочном скотоводстве [3], а также метод ГС был успешно опробован на растениях [35].

Одной из основных целей ГС является предсказание племенной ценности. Племенная ценность (Breeding value, BV, ПЦ) – это уровень генетического потенциала племенного животного и влияние данного генетического потенциала на хозяйственно полезные признаки потомства. ПЦ определяется как сумма двух компонент: среднего значения ПЦ родителей и отклонение от этого среднего, которое обусловлено случайным распределением половины генетического материала от родителей к потомку. Однако не всегда удается получить значение ПЦ потомков напрямую, исходя из значений ПЦ родителей. Поэтому используют другие подходы для оценки ПЦ. В частности, были разработаны методы, позволяющие оценить так называемую геномную оценочную племенную ценность (genetic estimated breeding value, GEBV, ГОПЦ). Суть подхода заключается в наборе тренировочной (или референсной) популяции (training population, TRN), для которой известны данные, как исследуемого признака, так и полногеномного генотипирования особей. Далее, используя методы машинного обучения (например, best linear unbiased prediction – BLUP [36]), обучают алгоритм на этой выборке для оценки ГОПЦ. Этот алгоритм с обучением далее может быть уже применен на тестовых популяциях (test population, TST), для которых известна только лишь генетическая информация [37].

ГС позволяет значительно ускорить процесс отбора животных, уменьшив не только время для достижения результата, но и стоимость. Например, при исследовании любой крупной популяции становится возможным использовать 50% имеющегося материала в качестве тренировочной популяции, получить для данной популяции полную генетическую и фенотипическую информацию и рассчитать ГОПЦ, а на оставшейся половине материала использовать лишь генетическую информацию и предсказать ПЦ.

ГС это, прежде всего, подход, для которого важен относительно быстрый и доступный результат. В частности, у растений, фокусируясь на прогнозировании аддитивных эффектов, геномная селекция способна дать результат уже во втором поколении [38]. Стоит отметить, что геномная селекция используется для прогнозирования аддитивных значений ПЦ, а не общей генетической ценности. В свою очередь, существует подход полного предсказания генетической ценности (genomic-enabled prediction, GP, ПГЦ), который включает в себя не только аддитивные эффекты маркеров, но и неаддитивные эффекты (например, доминирование или эпистаз). Таким образом, становится возможным оценить экономическую выгоду отдельного сорта или породы.

Существуют разнообразные сложности предсказания ПГЦ, которые связаны с несколькими факторами. Например, когда признак комплексный и в его контроль вовлечено много локусов, точность определения ПГЦ зависит от нескольких факторов: 1) Размер и генетическое разнообразие тренировочной популяции, и связь с тестовой популяцией (т.е. являются ли особи из тренировочной популяции родственными особям из тестовой популяции) [39]; 2) Наследуемость выбранных признаков для селекции (сложные признаки с низкой наследуемостью и небольшим эффектом от маркеров подходят, как для ГС, так и для ПГЦ, в то время как признаки с высокой наследуемостью могут быть предсказаны и с помощью маркер-ориентированной селекции); 3) Зависимость от неравновесия по сцеплению сложных признаков с каким-либо QTL: в таком случае точность геномного предсказания напрямую зависит от размера тренировочной популяции и наследуемости [40]. Также точность геномного предсказания зависит от размера тренировочной популяции и количества маркеров и имеет тенденцию выходить на плато при достижении достаточного количества особей и маркеров [41, 42].

В свою очередь, статистические модели ГС также имеют свои сложности. Они связаны, прежде всего, с тем, что количество маркеров

значительно превышает количество исследуемых особей в популяции, и маркеры зачастую сильно коррелируют. Это приводит к невозможности использования каноничных моделей линейной регрессии. Для решения такого рода проблем существует несколько путей: 1) Выбор определенных переменных (маркеров) для анализа; 2) уменьшение размерности (например, методом главных компонент), что позволяет сгенерировать новый набор предикторов, которые не скоррелированы с исходными предикторами (маркерами) [43]; 3) Использование усовершенствованных методов регрессии, например PLS-регрессии (регрессия частично наименьших квадратов) [44].

Одним из классических методов в ГС для оценки ПЦ является метод сравнения со сверстницами (СС). Рассмотрим его на примере молочной продуктивности КРС. В данном методе продуктивность дочерей быка (D) сравнивается с продуктивностью лактировавших в аналогичных условиях сверстниц (C). Благодаря этому методу становится возможным отобрать быков, чье потомство имеет лучшие показатели лактации, чем в среднем по стаду. ПЦ оценивается по следующей формуле:

$$EBV_{CC} = 2b \frac{\sum(D - \bar{C})}{n} \quad (1)$$

$$b = \frac{w}{w+k} \quad (2)$$

$$k = \frac{4-h^2}{h^2} \quad (3)$$

$$w = \frac{n \times n_c}{n+n_c} \quad (4)$$

где $(D - \bar{C})$ – отклонение продуктивности дочерей быка от средней продуктивности сверстниц, n – число дочерей, w – число эффективных дочерей, n_c – число сверстниц, h^2 – коэффициент наследуемости.

Несмотря на относительную простоту применения, на практике метод СС имеет несколько конкретных условий для получения несмещенной оценки

среднего ПЦ. Прежде всего, необходимо чтобы все оцениваемые животные (отцы) были из одной популяции. Также важно отсутствие генетического тренда в выбранной популяции, то есть выбираются те популяции, где эффективность селекции равна нулю. Исходя из этих факторов, а также с учетом применения спермы животных из других стран, а, следовательно, популяций, для искусственного оплодотворения, метод СС является неоптимальным решением для оценки ПЦ животных.

Самым широко используемым в данный момент статистическим методом для использования в ГС является метод наилучшего линейного несмещенного прогноза BLUP (Best Linear Unbiased Prediction) [36]. Впервые данный метод был описан Чарльзом Роем Хендерсоном [45], однако речь шла лишь о теоретической модели, которая не имела практического применения. Началом использования метода BLUP для селекционных задач считается конец 70-х - начало 80-х годов XX века. По сравнению с описанным выше методом СС, метод BLUP не имеет ограничений и является более универсальным. В упрощенном виде ПЦ с помощью BLUP рассчитывается по следующей формуле:

$$EBV_{BLUP} = 2[g + b \frac{\Sigma(D-c')}{n}] \quad (5)$$

где в отличие от метода СС вместо \bar{C} в формуле присутствует C' , а также введена дополнительная переменная g . Эта переменная является оценкой эффекта генетической группы, к которой относится отец. Данная оценка рассчитывается как различие между средней ПЦ быков, рожденных в том же году что и исследуемый бык с ПЦ быков, составляющих выборку для сравнения. Выборки быков для сравнения могут быть использованы различные, в зависимости от задач и возможностей селекционера. Таким образом, учитывается генетический тренд, который не присутствует при оценке ПЦ с помощью метода СС. C' – это среднее по сверстникам, но скорректированное с учетом ПЦ отцов сверстниц из выборки быков для

сравнения. Это позволяет выявить генетические различия между производителями и отобрать наиболее продуктивных быков для дальнейшего проведения ГС.

1.2 Овцеводство, современное состояние.

В настоящее время самыми крупными отраслями в области разведения животных являются разведение крупного рогатого скота, разведение свиней и птиц. Все эти виды животных разводятся человеком для получения мяса, а также молока и яиц. Однако, в связи с ростом населения планеты и все более остро встающей проблемой обеспечения продовольственной безопасности, человечество в последние десятилетия развивает и другие направления животноводства. Одним из таких направлений является овцеводство. Основным продуктом овцеводства является шерсть, однако в последние годы идет смещение в сторону мясного овцеводства [46]. Также овцы используются человеком для получения молока и овчины, но доля этих продуктов незначительна. Глобально все породы овец подразделяют на следующие типы: 1) шерстные породы овец, которые, в свою очередь, подразделяются на тонкорунные, полутонкорунные, полугрубошерстные и грубошерстные; 2) мясные породы овец; 3) мясо-сальные породы; 4) мясо-шерстные; 5) молочные. Зачастую категории, связанные с мясной и молочной продуктивностью, относят к грубошерстным породам, однако это не всегда так. Например, южноафриканская порода овец Дорпер обладает отличными свойствами баранины, но практически не имеет пригодной для переработки шерсти. В настоящее время основными лидерами по поголовью овец являются Китай, Австралия, Индия, Судан и Иран. По производству баранины лидерами являются Китай, Австралия, Новая Зеландия, Судан и Турция [47]. В целом, если взглянуть на процентные соотношения поголовья овец в мире, то выясняется, что 44% разводят в Азии, 28.2% в Африке, 11.2% в Европе, 9.1% в Океании и 7.4% в Америке [48]. В странах СНГ и в России, в частности, также развито животноводство, но не на уровне мировых лидеров.

1.2.1 Овцеводство в России, общие положения

По данным Всероссийской сельскохозяйственной переписи 2016 года поголовье овец в России составило 24950,2 тыс. голов [49]. Данный показатель на 32% выше (~ 6000 тыс. голов), чем аналогичный в 2006. Вместе с увеличением поголовья овец, увеличилась и доля животных, находящихся в крестьянских (фермерских) хозяйствах и у индивидуальных предпринимателей (с 21.1% от общего поголовья овец до 35.9%). В то же самое время значительно снизилась доля овец, находящихся в личных подсобных и других индивидуальных хозяйствах граждан (с 55,8% от общего поголовья овец до 45,1%). Эти данные свидетельствуют о постепенном увеличении рынка овцеводства и продуктов овцеводства на территории России. Лидерами овцеводства в России являются республика Дагестан, республика Калмыкия, Ставропольский край, Астраханская область, Карачаево-Черкессия, Ростовская область, республика Тыва, Волгоградская область, республика Башкортостан и республика Алтай. Данные регионы находятся в различных климатических условиях, а значит разводят различные породы овец [50]. Племенной реестр Российской Федерации подразделяет овец на 4 группы:

- 1) Грубошерстные породы: эдильбаевская, тушинская, короткожирнохвостая тувинская, романовская, лезгинская, карачаевская, курдючная калмыцкая, бурятская, андийская и агинская;
- 2) Смушково-молочные: каракульская;
- 3) Полутонкорунные и тонкорунные породы: цигейская, северокавказская (относится к мясо-шерстным), горноалтайская, советская (сибирская) мясо-шерстная, тексель (голландская порода), куйбышевская, мясная южная и ташлинская.

Племенное овцеводство в России развито и практически не требует закупки чистопородных овец из других стран. Так, по состоянию на 2015 год, в Россию официально было ввезено всего 73 чистопородных животных [51]. Также стоит отметить, что, начиная с 1990 года, доля шерстяных пород в

России уменьшилась на 47%, в то время как доля мясных пород увеличилась на 34% [52].

1.2.2 Российские мясные породы овец

В России на данный момент разводятся несколько мясных пород. Стоит отметить, что данные породы не являются исключительно мясными, а скорее мясо-шерстными, и помимо мяса производители получают также шерсть и руно. Романовская порода овец была разведена более 2 веков назад и имеет широкое распространение. В разведении неприхотлива, быстро адаптируются к разным климатическим условиям. В тридцатые годы XX в. при помощи скрещивания племенных Гемпширов с аборигенными домашними баранами советским селекционером удалось получить Горьковскую мясную породу с грубой короткой шерстью и высокой продуктивностью. Благодаря легкому костяку и быстрому росту мышечной массы, а также неприхотливости, выносливости и хорошему иммунитету, овцы Горьковской породы получили широкое распространение. Не менее известна, чем Романовская, Северокавказская порода овец. Ее вывели, используя в качестве основы Ставропольских баранов. Чтобы увеличить их мясные характеристики, овец скрещивали с Ромни-марш и Линкольнами. В результате получилась адаптированная к климатическим условиям юга России и севера Кавказа мясная порода. В то же время данная порода является полутонкорунной [53].

1.2.3 Мясная селекция в овцеводстве

На данный момент баранина является основным продуктом разведения овец. Селекционная работа по признакам мясной продуктивности у овец ведется уже много столетий. Баранина является одной из ключевых частей рациона питания кочевых тюркских, монгольских и арабских народов. Эти народы использовали методы классической селекции, а именно гибридизацию, отбирая лучших баранов и овец и скрещивая их между собой. С началом развития генетики как науки и появления маркер-ориентированной селекции, методы МОС стали применяться и для разведения овец с

улучшенными признаками мясной продуктивности [54]. Однако овцеводство никогда не являлось крупнейшей отраслью животноводства, уступая селекции крупного рогатого скота, свиноводству и разведению птицы. Также стоит отметить, что в середине прошлого века шерсть являлась главным продуктом овцеводства. Поэтому первые специализированные научные работы, связанные с селекцией овец по признакам мясной продуктивности, появились лишь в XXI веке. В одной из первых таких работ французские исследователи проанализировали генетическое разнообразие 6 пород французских овец, используя, в том числе, и генеалогическую информацию [55].

Развитие генетики, в частности молекулярной генетики, дало большой толчок для идентификации и аннотации функциональных генов, связанных с мясной продуктивностью у овец. Несмотря на то, что рост и размножение организма регулируются, прежде всего, гипофизом [56], существуют также транскрипционные факторы и индуцирующие сигналы, влияющие непосредственно на гипофиз, а, следовательно, на признаки мясной продуктивности. К таким факторам относятся *LHX3* и *LHX4*, влияющие на рост и развитие животного [57]. Одним из самых известных генов, который широко используется в МОС, является кальпаин (*CAPN*). С помощью применения методов ПЦР было показано, что данный ген связан не только с признаками роста и мясной продуктивностью, но и влияет на структуру мышечного волокна, делая его более нежным [58, 59]. Ген *CAPN* кодирует внутриклеточные цистеиновые протеазы, которые активируются кальцием и участвуют в физиологических процессах организма, а именно в перестройке скелетных мышц [60]. Говоря о скелетных мышцах, нельзя не упомянуть негативный регулятор развития мышечных волокон миостатин, кодируемый геном *MSTN*. Было показано, что развитие мышц у овец увеличивается до 2 раз, при нарушении в работе гена *MSTN* [61]. Не менее известным, чем ген миостатина, является ген гормона роста *GH*, некоторые мутации в котором приводят к ускоренному развитию животного и меньшим жировым

отложениям [62]. Также стоит отметить ген *MEF2B*, который также влияет на признаки роста и мясной продуктивности у овец и широко используются в МОС [63]. Для вариантов в гене *MEF2B* была показана положительная ассоциация с диаметром волокна миоцита и отрицательная ассоциация с плотностью волокон миоцита у коз [64].

1.2.4 Геном овец, общие положения

Геном овец состоит из 54 хромосом: 26 пар аутосом и пара половых хромосом. Впервые геном овцы был полностью секвенирован и собран в 2010 году. Данная сборка получила название OAR_v1.0. Для ее создания были использованы 6 овец различных пород, нуклеотидные последовательности которых (полученные с помощью технологии 454 FLX) были выравнены с использованием генома КРС [65]. В 2011 и 2012 годах соответственно были опубликованы усовершенствованные версии генома OAR_v2.0 и OAR_v3.1.

Версия генома OAR_v3.1 была собрана с использованием данных нуклеотидных последовательностей 2 животных породы Тексель (самца и самки). При этом геном самца был выравнен по геному самки, использовавшемуся в качестве референса. Часть пробелов в геноме была заполнена с использованием данных секвенирования животных с помощью технологии Illumina GAII. На основе версии OAR_v3.1 был создан микрочип для генотипирования Illumina Ovine SNP50 Beadchip, покрывающий 50000 точек в геноме [66].

Уточненная версия генома OAR_v3.1 получила название OAR_v4.0 и была опубликована в 2015 году. Для восполнения пробелов предыдущей версии была использована технология получения длинных ридов (прочтений нуклеотидных последовательностей) «Pacific Biosciences» [67]. В результате среднее покрытие генома в версии OAR_v4.0 составляет 166 прочтений на позицию [68].

В 2017 году была опубликована версия генома для породы Рамбулье - Oar_rambouillet_v1.0. Для ее создания, как и для сборки OAR_v4.0, использовалась технология получения длинных ридов «Pacific Biosciences», а также технологические решения компании Illumina [69]. В настоящее время большинство исследователей в своих работах используют в зависимости от задач и исследуемых выборок 3 сборки генома овец OAR_v3.1, OAR_v4.0 и Oar_rambouillet_v1.0.

1.3 Полногеномное исследование ассоциаций

Полногеномное исследование ассоциаций (ПГИА) в настоящее время является одним из самых популярных методов для картирования сложных признаков различных организмов. В основе метода лежит поиск ассоциаций между большим количеством точек в геноме (от нескольких тысяч до миллионов) и исследуемым признаком. Самым используемым маркером в ПГИА являются однонуклеотидные полиморфизмы (ОНП). Основным результатом ПГИА являются локусы, ассоциированные с исследуемым признаком. Локусы формируются из ОНП, находящихся в неравновесии по сцеплению (linkage disequilibrium, LD). Неравновесие по сцеплению между двумя ОНП в одном локусе предполагает, что частота встречаемости аллелей данного ОНП в популяции статистически значимо отличается от случайной частоты независимого наследования данных аллелей. Данный эффект возможен из-за наличия сцепления аллелей (совместной передачи в ряду поколений). Стоит отметить, что рекомбинация уменьшает степень сцепления аллелей и это уменьшение прямо пропорционально физическому расстоянию между аллелями. Например, сильно сцепленные аллели, находящиеся в непосредственной близости друг от друга (1-2 сантиморгана) сохраняют свое сцепление десятки поколений [70]. Таким образом, предполагается, что для найденного при ПГИА значимо ассоциированного ОНП необходимо изучить также все ОНП, находящиеся с ним в неравновесии по сцеплению. Это

необходимо для выявления причинного функционального маркера и картирования гена, контролирующего исследуемый признак.

Активное использование ПГИА началось с момента создания коммерческих ДНК-микрочипов для проведения генотипирования по большому числу маркеров. Первая работа об успешном применении метода ПГИА вышла в 2005 году. В этой работе была исследована ассоциация более чем 100 тысяч ОНП с возрастной дегенерацией желтого пятна у человека [71]. На данный момент в рамках генетики человека имеется информация о более чем 5700 проведенных ПГИА для более чем 3300 различных признаков [72]. Использование огромных выборок людей (более миллиона человек) позволяет значительно увеличить статистическую мощность анализа [73]. В рамках генетики животных количество найденных ассоциаций значительно меньше. Однако, также существуют исследования с выборками огромных размеров и в рамках исследования генетики животных [1]. Тем не менее, несмотря на значительные успехи в проведении ПГИА, существует проблема биологической интерпретации найденных вариантов [74]. Зачастую, на исследуемый признак влияет очень много локусов, каждый из которых вносит небольшой вклад в дисперсию признака. Каждый из этих локусов, в свою очередь, может быть сцеплен с разнообразными другими локусами, не являющимися значимо ассоциированными с исследуемым признаком. Поэтому накопление новых результатов ПГИА позволяет находить новые ассоциации, но в то же время и усложняет биологическую интерпретацию и нахождение прямых причинно-следственных связей (мутация-признак).

1.3.1 Статистические модели ПГИА

Для поиска ассоциаций между маркером (ОНП) и признаком чаще всего используют либо линейную, либо логистическую регрессию. Выбор типа регрессии зависит от типа исследуемого признака. Для количественных признаков (масса, количество производимого молока, жирность молока и т.д.) используется линейная регрессия. Для бинарных признаков (наличие или

отсутствие болезни и др.) используют логистическую регрессию. Линейная регрессия для ПГИА может быть представлена в виде

$$Y \sim W\alpha + X_s\beta_s + e \quad (6)$$

где Y - вектор значений признака, W – матрица ковариат (или независимых переменных, предикторов), α – вектор значений эффекта ковариат, X_s – вектор значений генотипа ОНП (чаще всего в формате 0,1,2), β_s – размер эффекта для каждого ОНП [75]. Стоит отметить, что использование ковариат таких как пол, возраст, происхождение и др. необходимо для предотвращения влияния возможного скрытого фактора и, как следствие, получения недостоверной оценки эффекта, несмотря на возможное снижение статистической мощности для бинарных признаков [76].

Стоит отметить, что одними из важнейших ковариат для предотвращения влияния скрытого фактора (конфаундера) и уменьшения ложноположительных результатов являются ковариаты, описывающие популяционную структуру. Существует 4 основных метода для учета популяционной структуры и уменьшения инфляции статистики анализа – геномный контроль (Genomic Control, GC), структурная ассоциация (Structured Association, SA), анализ с включением главных компонент (Principal Component Analysis, PCA) и анализ смешанных моделей (mixed models) [77]. Первые три метода подходят для уменьшения популяционной стратификации, однако не могут полностью оценить родство между особями. Линейные смешанные модели, в свою очередь, позволяют учитывать популяционную структуру выборки в ПГИА путем попарной оценки генетического сходства между исследуемыми особями.

В линейной смешанной модели генетическую компоненту дисперсии признака разделяют на аддитивную и полигенную компоненты. Для каждого ОНП проверяется, является ли доля объяснённой дисперсии генетического

эффекта в этом локусе значимо отличной от нуля. В общем виде уравнение для линейной смешанной модели может быть записано в виде

$$Y \sim \mu + a + g + e \quad (7)$$

где μ - фиксированный эффект, описывающий общее среднее признака, a - описывает аддитивный генетический эффект ОНП, g - полигенный эффект на признак, e - случайный остаточный эффект на признак. В более полном виде в данное уравнение необходимо, как и в случае с обычной линейной регрессией, добавить векторы ковариат, генотипов и размера эффекта. Также, g и e в полном виде выглядят как

$$g \sim N(0, \sigma_A^2 \Psi) \quad (8)$$

$$e \sim N(0, \sigma_e^2 I) \quad (9)$$

где σ_A^2 - аддитивный генетический эффект, Ψ - стандартная генетическая матрица родства, σ_e^2 - дисперсия остатков, I - единичная матрица.

Линейные смешанные модели демонстрируют значительное превосходство в контроле ошибки первого рода в случае наличия родственной структуры в выборке [78]. Было предложено множество эффективных программных решений, использующих линейные смешанные модели, такие как, например, «GEMMA» [79], «EMMAX» [80], «FaST-LMM» [81] и др.

Также стоит отметить другие современные методы, позволяющие увеличивать мощность анализа. Одним из таких является метод многомерного анализа (Multivariate ANalysis Of VAriance - MANOVA) [82], в котором признаки группируются в некоторые группы по различным сходным свойствам (общность проявления, близкий биохимический состав и т.д.). Модель MANOVA выглядит следующим образом (упрощенно):

$$Y_1 + \dots + Y_i \sim g \quad (10)$$

где Y_i – признаки, входящие в состав многомерного признака для анализа, g – генотип исследуемого ОНП (0,1,2). В случае одного генотипа модель эквивалентна модели:

$$g \sim Y_1 + \dots + Y_i \quad (11)$$

при которой происходит оценка эффектов признаков Y_i на генотип g . В таком случае, если оценены эффекты признаков на генотип s_i , можно составить многомерный признак $S = \sum(s_i * Y_i)$, и модель (12) будет аналитически эквивалентна модели $S \sim g$, или

$$S = \beta_{mv} * g + e + \mu \quad (12)$$

где β_{mv} – эффект генотипа на многомерный признак S . Таким образом, общие подходы к анализу многомерных признаков можно свести к анализу стандартных одномерных моделей. Стоит отметить, что по факту многомерный признак является линейной комбинацией признаков Y_i . Однако, в отличие от обычного одномерного анализа линейных комбинаций признаков в рамках ПГИА, в MANOVA происходит оценка линейных коэффициентов признаков s_i для каждого из ОНП по отдельности, т.е. для каждого генотипа оптимальной является своя линейная комбинация признаков. Далее мы будем называть такой анализ с использованием модели MANOVA многомерным.

Ранее было показано, что для успешного применения модели MANOVA достаточно использовать суммарные статистики результатов одномерного анализа ПГИА (p-value, размер эффектов, а также дополнительная информация для каждого исследованного ОНП), без учета индивидуальных данных генотипов и фенотипов исследуемых особей. Это делает применимость таких моделей универсальной, в том числе, и для результатов мета-анализа одномерных моделей ПГИА [12, 83]. Для такого частного случая модель выглядит следующим образом:

$$T_j^2 = Z_j^T V^{-1} Z_j \quad (13)$$

где Z - это вектор суммарных статистик (z-score), V – матрица корреляции между суммарными статистиками, T_j^2 - искомые значения t-score.

Несмотря на упомянутые ранее исследования с использованием огромных выборок (миллион образцов), зачастую проводимые ПГИА исследуют меньшие количества образцов (тысячи и десятки тысяч). В таком случае для увеличения мощности ПГИА применяется метод объединения результатов ПГИА полученных с использованием различных выборок в один общий мета-анализ [84]. Одними из основных решений для проведения мета-анализа на данный момент являются METAL [85] и N-GWAMA (или MA-GWAMA) [86].

При проведении мета-анализа необходимо проведение предварительного контроля качества. При использовании мета-анализа первым шагом является унификация используемых протоколов анализа для каждой из используемых выборок [87]. Это также включает в себя использование согласованных фенотипов (использование одного и того же формата и одних и тех же единиц измерения). Первый шаг может включать масштабирование размеров эффекта до стандартного нормального распределения, поскольку фенотипические измерения и их предполагаемые абсолютные размеры эффекта иногда невозможно сравнить между выборками. Далее два независимых исследователя проводят следующие необходимые шаги контроля качества, которые могут включать в себя фильтрацию по частоте минорного аллеля, выравнивание аллелей между выборками и так далее. Наконец, на сводной статистике выполняется мета-анализ. Мета-анализ может быть выполнен с использованием модели с фиксированным эффектом, которая предполагает, что дисперсия ошибок одинакова для всех выборок, или модели со случайным эффектом для проверки неоднородности результатов;

например, проверка того, явно ли одна или две выборки отличаются от остальных.

1.3.2 Анализ результатов ПГИА

Основным результатом ПГИА являются значения p -value, а также размер и направление эффектов для каждого исследованного ОНП. Для визуализации результатов ПГИА используют «Manhattan Plot» (график Манхэттен) (Рисунок 2) и «Quantatile-Quantatile (Q-Q) plot» (График квантиль-квантиль) (Рисунок 3) [88]. На «Manhattan Plot» каждая точка соответствует одному ОНП. На оси абсцисс указывается физическое положение генетического варианта на карте генома. Ось ординат представляет собой отрицательный десятичный логарифм p -value. Чем больше значение по оси ординат, тем выше значимость ассоциации данного варианта с исследуемым признаком. Зачастую на «Manhattan Plot» указывают линию порога значимости, за счет чего можно по графику установить, является ли ОНП значимо ассоциированным или нет. «Quantatile-Quantatile (Q-Q) plot» отображает распределение ожидаемого p -value при нулевой гипотезе (нет ассоциации) и распределение наблюдаемого p -value. Значения p -value также преобразуются в отрицательный десятичный логарифм. По оси абсцисс отображаются ожидаемые значения p -value, а по оси ординат - наблюдаемые значения p -value. Это позволяет оценить, насколько сильна ассоциация с исследуемым признаком. Сильное отклонение наблюдаемого распределения от ожидаемого может быть вызвано как наличием значимых генетических ассоциаций, так и сильной популяционной стратификацией. Однако, зачастую, при наличии значимых генетических ассоциаций отклонение наблюдаемого распределения от ожидаемого будет сконцентрировано в правой части графика, тогда как при популяционной стратификации отклонение будет ближе к началу координат. Но также при сильном полигенном эффекте может наблюдаться отклонение наблюдаемого

распределения от ожидаемого в начале координат, что может усложнить интерпретацию результатов.

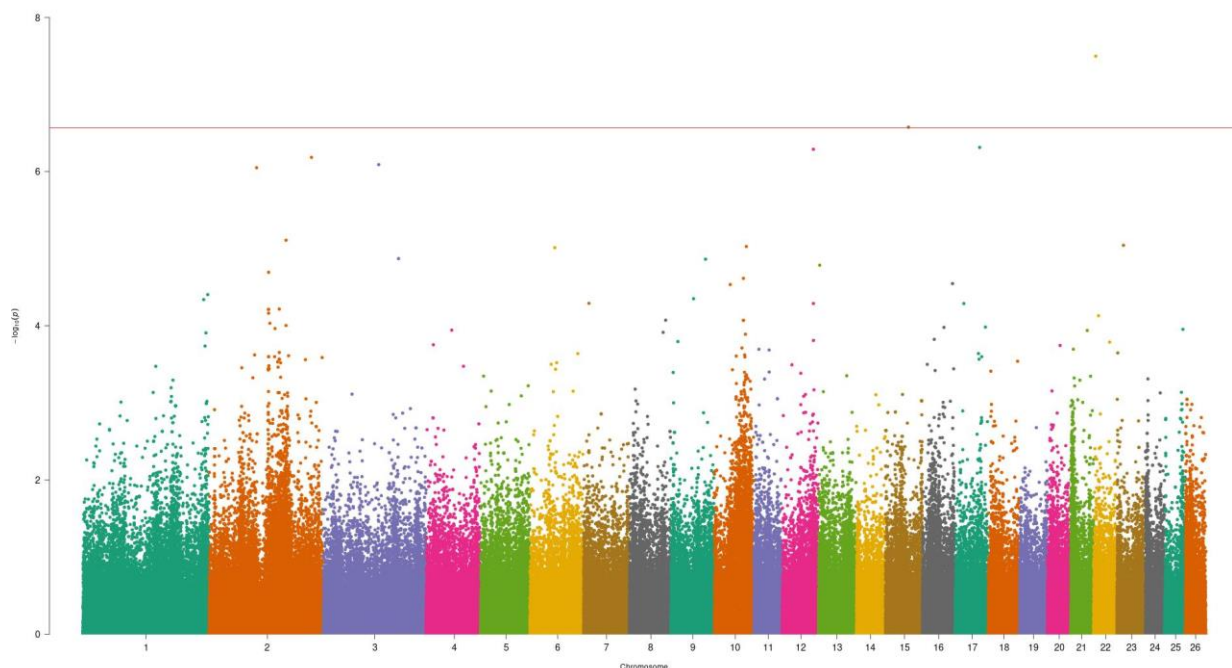


Рисунок 2. Пример представления результатов ПГИА в виде «Manhattan Plot». По оси абсцисс обозначены хромосомы, по оси ординат отрицательный десятичный логарифм p-value. Цветом обозначены ОНП, расположенные на одной хромосоме.

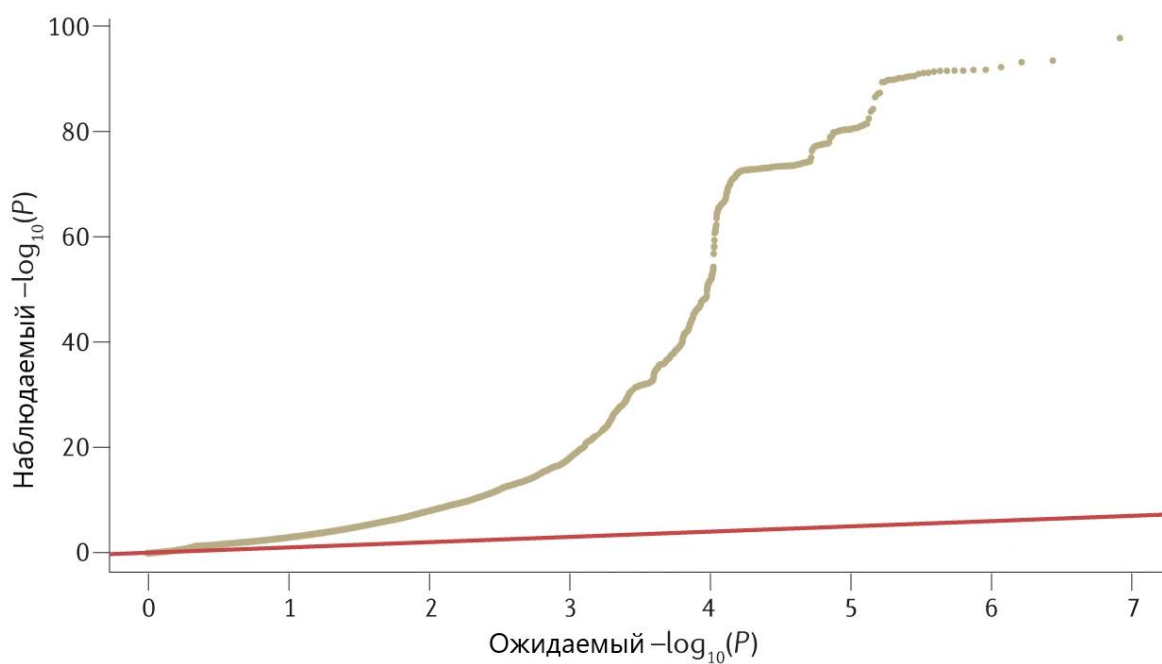


Рисунок 3. Пример представления результатов ПГИА в виде «Quantatitle-Quantatitle (Q-Q) plot». В данном случае большое отклонение наблюдаемого p-value от ожидаемого вызвано высокой полигенностью признака (индекс массы тела человека). Адаптировано из [75].

Так как ПГИА представляет собой, по сути, множественное тестирование ассоциаций ОНП с исследуемым признаком, необходимо использовать поправку на множественное тестирование, иначе высока вероятность получения ложноположительных результатов. При проведении ПГИА используют либо консервативную поправку Бонферрони [89], либо менее консервативный «FDR (False Discovery Rate)» [90]. В случае использования поправки Бонферрони для получения порогового значения p-value, при котором ошибка первого рода не превышает 5%, необходимо разделить номинальный порог $p\text{-value}=0.05$ на количество независимых тестов (в случае ПГИА это количество исследованных ОНП). Однако, стоит учитывать, что неравновесие по сцеплению делает количество тестов в ПГИА меньшим, чем количество ОНП. Это и приводит к высокой консервативности поправки Бонферрони, и, как следствие, к повышению вероятности получения ложноотрицательных результатов. Чтобы этого избежать, в качестве метрики количества тестов используют не общее количество ОНП, а количество условно независимых ОНП.

Основной идеей проведения ПГИА является поиск биологических причин проявления того или иного фенотипа. Для достижения данного результата необходимо провести функциональную аннотацию выявленных значимо ассоциированных с признаком локусов. Это понятие включает в себя несколько основных шагов. Во-первых, необходимо установить, расположен ли найденный локус в каком-либо гене, если да, то в каком участке гена он находится (энхансер, интрон, экзон). Если найденный локус лежит в межгенном пространстве, то определяют наиболее близкий к данной области ген и его указывают как функционально ответственный для данного локуса.

Далее, для самого ассоциированного ОНП в локусе, если он лежит в кодирующей части гена, необходимо установить, приводит ли замена аллеля к изменениям структуры белка за счет изменения кодируемой аминокислоты. Однако, из-за неравновесия по сцеплению, не всегда самый ассоциированный ОНП в локусе является функционально ответственным за проявление признака. Поэтому в дополнение к предыдущим шагам составляют список всех генов, находящихся на определенном расстоянии от локуса (обычно ± 500 тысяч пар нуклеотидов). Этот список затем используют для приоритизации найденных генов. Это позволяет установить, с какими физиологическими системами и биологическими путями связаны найденные гены, в каких клетках и тканях наблюдается наибольший уровень экспрессии продуктов данных генов. Для приоритизации генов в области генетики человека в большинстве случаев используют такие инструменты как «MAGMA» [91] и «DEPICT» [92].

Полученные в ходе проведения ПГИА результаты имеют широкое применение. В рамках генетики человека найденные локусы могут быть использованы как для установления причин какого-либо заболевания, так и для определения рисков заболеть тем или иным заболеванием. В рамках генетики животных и растений основным применением результатов ПГИА является использование информации о том, какие локусы ассоциированы с экономически важными признаками для проведения селекционных мероприятий.

1.3.3 ПГИА в мясном овцеводстве

С началом применения метода ПГИА для поиска QTL с целью проведения селекции, увеличилось число работ, связанных непосредственно с признаками мясной продуктивности, формирования каркаса и роста у овец. Одним из первых крупных ПГИА проведенных на выборке овец с целью поиска локусов, ассоциированных с признаками мясной продуктивности, является работа Zhang с соавторами 2013 года, в которой авторы исследовали

329 животных для поиска ассоциации с 11 признаками мясной продуктивности и признаками роста [93]. В результате авторы выделили 5 генов (*MEF2B*, *RFXANK*, *SAMKMT*, *TRHDE* и *RIPK2*), которые являются первоочередными генами-кандидатами, связанными с набором веса после отъема ягненка у матери. Начиная с этой работы и до конца 2018 года было проведено всего 9 ПГИА для признаков мясной продуктивности, формирования каркаса и роста [93–100], что относительно немного, по сравнению с работами по селекции КРС. Так, Riggio с соавторами обнаружили значимые ассоциации, связанные с массой тела животного на локусе, расположенном на 6 хромосоме [94]. В другой работе была также показана значимая ассоциация с массой тела локуса на 6-й хромосоме [95]. В дополнении к найденной ассоциации авторы выделили 3 гена-кандидата - *LAP3*, *NCAPG* и *LCORL*. Значимые ассоциации массы костей животного, а также плотности жировой прослойки, плотности мышечной ткани и веса жировой прослойки были также обнаружены на 6 хромосоме в районе 37 мегабазы [96]. Авторы другого исследования обнаружили 39 генов-кандидатов, связанных с размером тела у овец, причем 3 из них (*TP53*, *NTN1* и *ZNF521*) были показаны ранее, как ассоциированные с признаками мясной продуктивности [97]. В исследовании, связанном с поиском ассоциаций со среднесуточным приростом, было обнаружено 2 гена-кандидата *MAG11* и *ZNF770* [99]. В свою очередь, Ghasemi с соавторами выявили 3 гена кандидата (*RAB6B*, *Tf* и *GIGYF2*), связанных с весом ягнят при рождении [100].

Отдельно стоит отметить работу Volortaa с соавторами 2016, которая на данный момент является крупнейшим ПГИА у овец для признаков мясной продуктивности. В данной работе была исследована выборка из 10613 животных 9 различных пород (Рисунок 4). В качестве фенотипов были использованы 56 различных признаков, связанных с разными аспектами мясной продуктивности (масса различных частей тела животного, содержание жира, содержание различных веществ, цвет мяса после убоя и т.д.). Авторы

провели 56 исследований ассоциаций для каждого из признаков, а также применили подход, основанный на идее метода MANOVA, описанного выше в пункте 1.3.1. Для этого все 56 признаков были объединены в один большой многомерный признак. В результате, авторы выделили 23 локуса в геноме, не находящиеся в неравновесии по сцеплению друг с другом. Для этих 23 локусов авторы попарно сравнили корреляцию эффектов, для того чтобы обнаружить паттерны схожего влияния на исследуемые признаки. Затем авторы провели процедуру функциональной аннотации и приоритизации генов для всех найденных ОНП (687) и 23 самых значимо ассоциированных ОНП в 23 независимых локусах. В результате совмещения данных о паттернах схожего влияния, данных функциональной аннотации и данных приоритизации генов, авторы выделили 3 группы локусов, которые влияют на мясную продуктивность у овец (Рисунок 5).

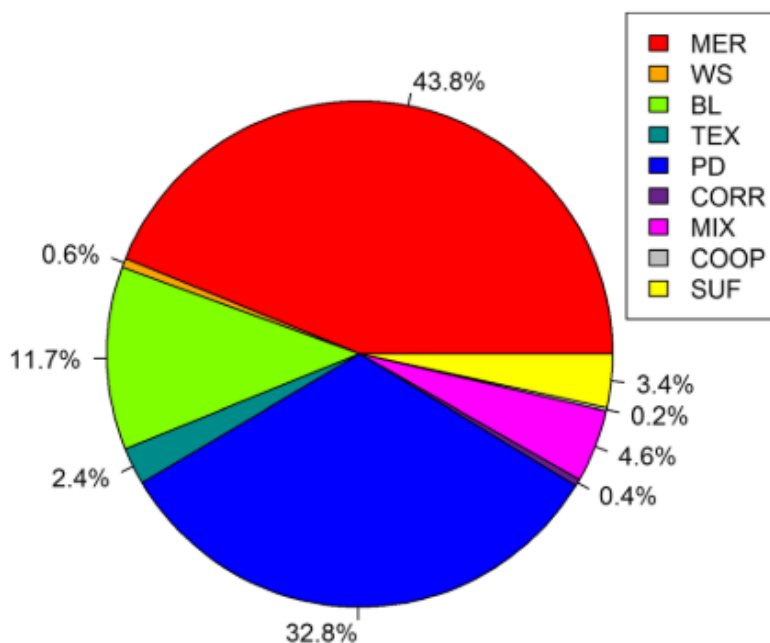


Рисунок 4. Распределение выборки животных по породам из статьи Volormaa с соавторами 2016 [16]. MER – Мериносы, WS – белый Суффолк, BS – Бордер-лейстер, TEX – Тексель, PD – Полл Дорсет, CORR – Корридейл, MIX – разнообразные гибриды, COOP – Купворт, SUF – Суффолк.

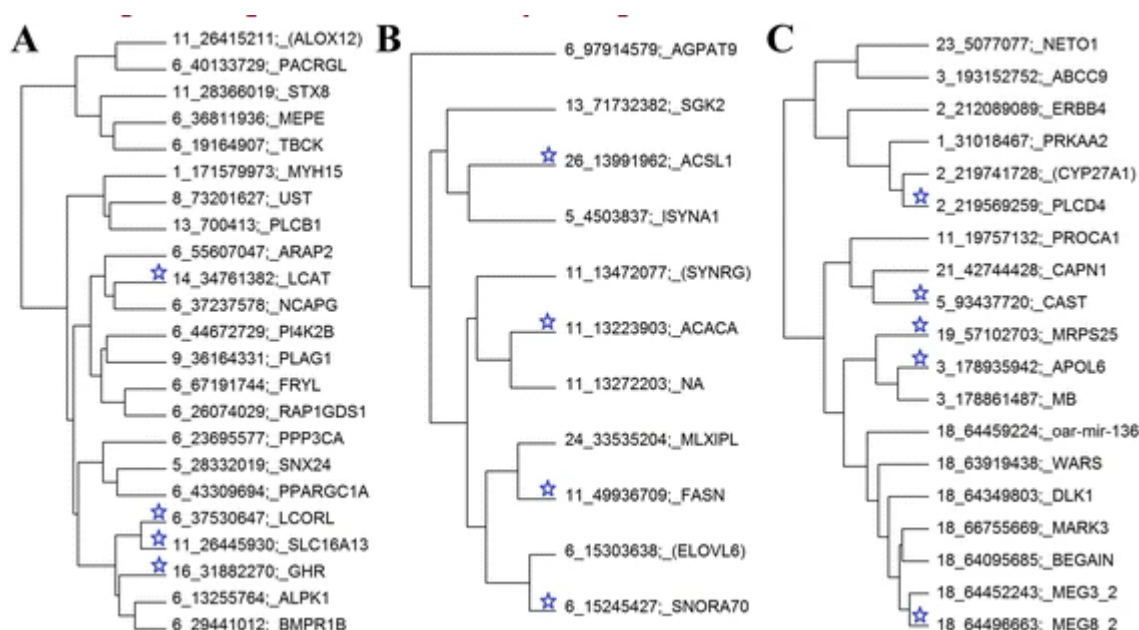


Рисунок 5. Дендрограмма 3 групп локусов из работы Volormaa с соавторами 2016. Звездочками отмечены ОНП из 23 независимых локусов. Названия генов в скобках обозначают наиболее вероятный функциональный ген для данного локуса. Каждая из групп связана с тем, какую функцию выполняют входящие в нее гены для признаков мясной продуктивности: А – рост и развитие, В – синтез и формирование жировых прослоек, С – цвет мяса.

Несмотря на наметившийся положительный тренд с началом использования методов ПГИА в области мясной селекции овец, доступная на данный момент информация очень разрознена и не может быть использована напрямую в селекционной работе, так как практически все описанные выше работы не содержат в себе информации не только об эффектах аллелей, но и о самих аллелях (какой из аллелей референсный, а какой эффекторный). Также невозможно использование напрямую информации и из крупнейшей на данный момент базы данных QTL для овец «AnimalQTLdb» по тем же причинам.

Также стоит отметить, что в сравнении с количеством исследований крупного рогатого скота, количество работ, связанных с экономически важными признаками овец, очень мало. Необходимы не только исследования,

объединяющие в себе всю известную на данный момент информацию, но и исследования для поиска новых маркеров и локусов, ассоциированных с признаками мясной продуктивности, формированием каркаса и ростом у овец. Более того, не менее важным является поиск, оценка размера эффекта и подтверждение найденных полиморфизмов на российских породах.

1.4 Решения для хранения и анализа результатов ПГИА

Хранение и удобные, в плане использования, интерфейсы доступа к результатам ПГИА необходимы для полной реализации потенциала ПГИА в рамках изучения природы проявления различных количественных и качественных признаков, таких как болезни человека, экономически важные признаки сельскохозяйственных животных и др. Результаты ПГИА хранят и обрабатывают в виде суммарных статистик. Суммарные статистики представляются в виде больших таблиц, где каждая строка соответствует одному исследованному ОНП. Одна строка содержит в себе информацию об идентификаторе ОНП в виде `rs_id`, хромосоме, позиции, референсном аллеле, эффекторном аллеле, частоте эффекторного аллеля, эффекте, стандартной ошибке и ассоциации с исследуемым признаком. В области генетики человека существует множество решений для хранения суммарных статистик, например, «GWAS Central» [101], «GWAS Catalog» [102], «Gene Atlas» [103], «PheWAS» [104], «PhenoScanner» [105].

Все существующие базы данных суммарных статистик обладают схожим функционалом. Пользователю предоставляется поиск вариантов с помощью идентификатора ОНП, хромосомы и позиции. Некоторые базы, такие как «PheWAS» и «PhenoScanner», применяют структуризацию загруженных суммарных статистик по категориям, таким как количественные признаки, заболевания, метаболиты, протеины и т.д., что помогает улучшить качество результатов поиска. Также в зависимости от политики использования данных, пользователь может либо загрузить полные суммарные статистики, либо, воспользовавшись поиском, получить информацию лишь об искомом

ОНП. Некоторые базы данных, например, «GWAS Central», хранят не полные суммарные статистики, а лишь информацию об ОНП с заданным порогом значимости $p\text{-value} < 1e-05$. В дополнение к этому для каждого из ОНП проводится поиск независимых ОНП в регионе ± 100 т.п.н..

В настоящий момент все базы данных проводят процедуру контроля качества перед загрузкой суммарных статистик. Контроль качества включает в себя проверку соответствия идентификатора ОНП хромосоме и позиции, проверку использованной сборки генома, конвертацию на использованную сборку при необходимости, выравнивание референсных и эффекторных аллелей, а также выравнивание частот. Таким образом, за счет проведения контроля качества, выбор базы суммарных статистик для пользователя обуславливается поставленной задачей, наличием прав доступа и личным предпочтением.

В области изучения генетики животных на данный момент не существует баз данных, хранящих в себе суммарные статистики. Существуют решения, которые хранят в себе информацию о локусах количественных признаков и то, в каких исследованиях они были получены. Крупнейшей такой базой на данный момент является «AnimalQTLdb» [6], которая содержит в себе информацию о самых ассоциированных локусах (но необязательно значимо-ассоциированных) для наиболее важных сельскохозяйственных видов животных, таких как коровы, овцы, свиньи и куры. Также существуют подобные решения для конкретных видов животных, например, база данных «iSheep» [8] для овец, которая помимо доступа к локусам количественных признаков предоставляет доступ к полным данным генотипирования и секвенирования животных.

Создание базы данных хранения суммарных статистик ПГИА животных должно учитывать существующие ограничения, которые во многом были решены в рамках баз данных суммарных статистик ПГИА человека. Во-

первых, результаты ПГИА зачастую имеют большой объем, что предполагает наличие развитой инфраструктуры для хранения и работы с такими объемами информации. Во-вторых, на данный момент существует множество протоколов проведения непосредственно ПГИА, что делает необходимым проведение контроля качества и перевода данных в универсальный формат обязательным шагом в каждом анализе. Третье ограничение касается непосредственно области генетики животных, где в от исследований, связанных с областью генетики человека, не является общей практикой выкладывать результаты ПГИА в виде суммарных статистик в открытый доступ.

Необходимость создания базы данных для работы с суммарными статистиками для животных очевидна. Это позволит находить новые локусы, ассоциированные с экономически важными признаками, а также подтверждать ранее найденные результаты. Несомненно, подобная платформа будет полезна и для селекционных нужд. Структурированная информация о генетическом контроле того или иного экономически важного признака поспособствует улучшению протоколов МОС и ГС, что, в свою очередь, также благоприятно отразится как на прибыли животноводческих хозяйств, так и на качестве жизни населения в целом за счет увеличения качества производимой продукции.

В качестве основы для создания базы суммарных статистик ПГИА целесообразно использовать платформу GWAS-MAP [11], разработанную для исследований в рамках генетики человека. Данная платформа не только хранит результаты ПГИА в виде суммарных статистик, но и предоставляет пользователю ряд инструментов по визуализации и анализу загруженных суммарных статистик.

1.4.1 Платформа GWAS-MAP

Данная платформа состоит из двух компонент. Первая обладает двумя вычислительными модулями (один – для добавления данных, второй – для анализа результатов ПГИА). Вторая содержит непосредственно саму базу данных. При добавлении новых данных суммарные статистики приводятся к одному общему, заранее установленному, формату. Затем данные проходят через процедуру контроля качества и добавляются непосредственно в базу данных. С помощью базы данных пользователь может запускать различные анализы с использованием суммарных статистик конкретных ПГИА, хранящихся внутри базы. Доступны для анализа расчеты, связанные с неравновесием по сцеплению, методами менделевской рандомизации, методами SMR (Summary-level Mendelian randomization) и HEIDI (heterogeneity in dependent instruments) [106], а также стандартные методы работы с ПГИА такие как мета-анализ и контроль качества. Схематично программная архитектура платформы GWAS-MAP представлена на рисунке 6.

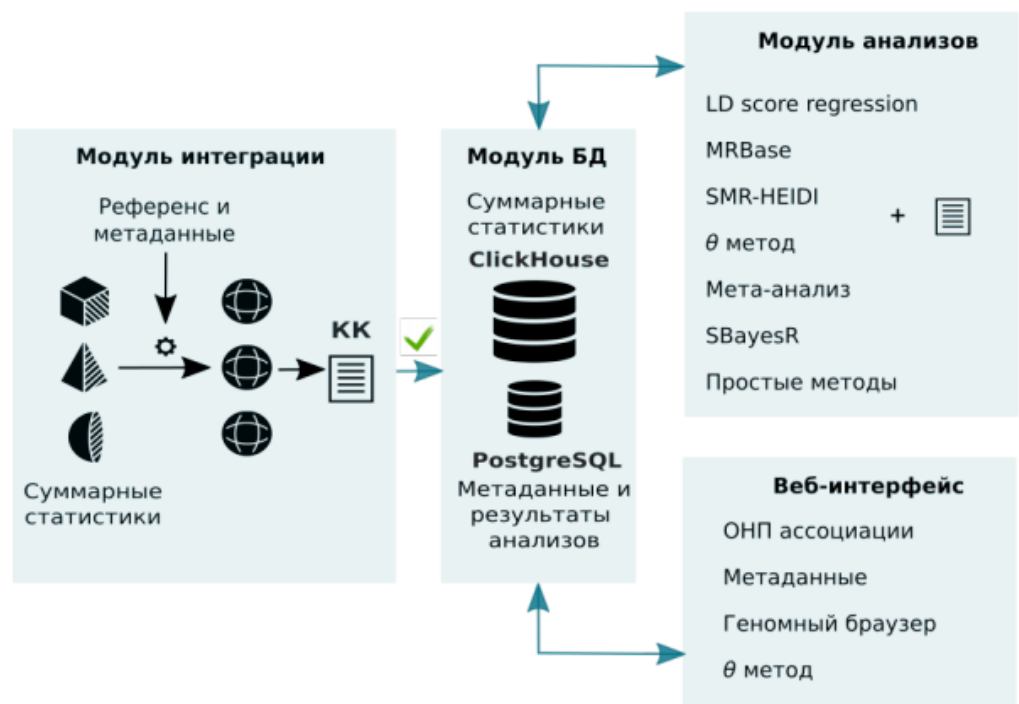


Рисунок 6. Схематичное изображение программной архитектуры платформы GWAS-MAP. Адаптировано из [11].

Стоит отметить, что GWAS-MAP также подходит для поиска необходимых исследований, например, для мета-анализа. ПГИА могут быть отфильтрованы по дате публикации, референсу, использовавшемуся для импутации (восстановление неизвестных генотипов с помощью информации о гаплотипах референса), по типу чипа, использовавшегося для генотипирования (Metabochip, Affymetrix, Illumina SNP arrays). Также, при необходимости, пользователь может искать ПГИА, сделанные с использованием данных, полученных путем полного секвенирования генома. На данный момент GWAS-MAP содержит в себе информацию о более чем 70 миллиардах ассоциаций, что делает данную платформу самой крупной базой данных результатов ПГИА человека.

Фенотипические данные, входящие в GWAS-MAP, могут быть разделены на несколько групп: комплексные признаки (включая болезни), метаболиты (mQTL), белки (pQTL), гликаны и данные по экспрессии генов. Благодаря такому широкому набору данных, пользователь может всесторонне исследовать интересующий его признак, узнав уровень экспрессии генов, белков и метаболитов, а также увидеть, связан ли данный признак с какими-либо другими признаками или болезнями.

1.5 Краткое заключение

Проведение селекционных мероприятий на животных является очень важной задачей для решения проблем продовольственной безопасности. В настоящее время самым широко используемым методом селекции является маркер-ориентированная селекция. Основным источником информации о том, какой маркер ассоциирован с исследуемым признаком, является полногеномное исследование ассоциаций. Существует заметный сдвиг в сторону изучения отдельных видов сельскохозяйственных животных. Также стоит отметить, что существующие стандарты исследований в генетике человека зачастую не применяются в исследованиях, связанных с животными (публикация полных суммарных статистик, информации о референсном и

эффекторном аллелях, размере и направлении эффекта и т.д.). Все вышеперечисленное связано, в том числе, и с отсутствием необходимого инструментария для хранения и анализа результатов ПГИА. Применение современных методов количественной генетики на овцах способно улучшить уже существующие и предоставить информацию для создания новых тест-систем для ведения МОС. Целью настоящей работы является поиск новых локусов, контролирующих признаки, относящиеся к мясной продуктивности овец, с использованием современных методов количественной генетики.

2. Материалы и методы

2.1 Устройство платформы GWAS-MAP|ovis

В качестве основы для создания платформы для сбора и хранения суммарных статистик полногеномных исследований признаков овец была использована платформа для хранения результатов ПГИА человека GWAS-MAP [11]. Кроме того, для представления и визуализации генетических данных был использован веб-интерфейс PheLiGe [107]. Принцип устройства платформы GWAS-MAP|ovis представлен на рисунке 7. Загружаемые в платформу суммарные статистики проходят процедуру контроля качества и унификации. Унифицированные суммарные статистики используются для поиска по ОНП и для проведения различных типов анализа.

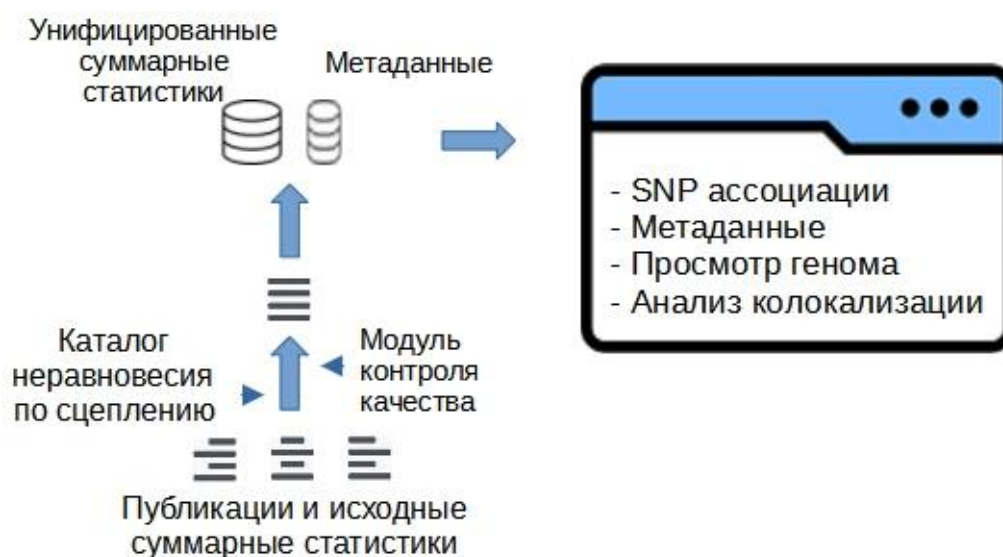


Рисунок 7. Схема устройства платформы GWAS-MAP|ovis.

2.2 Поиск и анализ статей для создания базы данных QTL и пополнения платформы GWAS-MAP|ovis

Для поиска и фильтрации статей были использованы сайты PubMed (<http://pubmed.com>) и Google Scholar (<http://scholar.google.com>). В качестве фильтра для поиска использовались следующие параметры: статьи на английском языке, опубликованные в рецензируемых журналах в 2013 году

или позже. Комбинации ключевых слов для поиска включали в себя слова, относящиеся к объекту исследования (sheep/ovine/Ovis aries), к исследуемому признаку (birth weight/weaning weight/6-month weight и т.д.) и к типу исследования (GWAS/gene candidate/association study). Пример комбинации ключевых слов - “sheep growth production traits GWAS”. Для всех используемых комбинаций были просмотрены и вручную отфильтрованы первые 20 результатов поиска. В рамках поиска статей мы также использовали дерево цитирования наиболее релевантных статей для расширения итогового списка найденных работ.

Все найденные статьи были вручную просмотрены и отфильтрованы. Были отобраны работы, связанные с мясной продуктивностью у овец. На первой стадии была извлечена следующая информация из статей: год публикации, название журнала, исследуемый признак, размер выборки, порода, тип исследования и др. С помощью полученной информации все статьи были оценены по релевантности, начиная с 1 (наиболее подходящая) до 3 (неподходящая или ошибка поиска). Статьи с оценками 1 и 2 составили окончательный список статей для дальнейшего анализа.

На следующем этапе была извлечена следующая информация: 1) список ассоциированных ОНП со всей доступной информацией (название ОНП, хромосома и позиция, эффекторный и референсный аллель, величина эффекта и стандартная ошибка, частота, ближайший приоритизированный ген; для ген-кандидатных исследований - прямые и обратные последовательности праймеров); 2) Список приоритизированных генов. Если статья содержала геномный уровень значимости и хромосомный уровень значимости для QTL, то был использован геномный уровень значимости. Для каждого гена из списка генов была произведена оценка количества публикаций (исключая обзоры), в которых данный ген упоминается.

2.3 Суммарные статистики для многомерного анализа

Для проведения многомерного анализа, а также для наполнения платформы GWAS-MAP^{ovis} были использованы суммарные статистики (z-score) из статьи Volormaa с соавторами 2016 [16]. В данной работе авторы провели ПГИА для выборки 10613 овец из 9 различных пород, которые были прогенотипированы с использованием Ovine Infinium® HD SNP Bead- Chip (Illumina Inc., San Diego, CA, USA) и Illumina 50k Ovine SNP chip (Illumina Inc., San Diego, CA, USA) по 56 различным признакам, связанным с мясной продуктивностью (Рисунок 4). В результате данного анализа авторы получили суммарные статистики для 510174 ОНП.

2.4 MANOVA и многомерные модели

Одним из методов количественной генетики является применение одномерного дисперсионного анализа (ANOVA – ANalysis Of VAriance) для поиска ассоциации одной или нескольких независимых переменных (пол, возраст, генотип и пр.) с зависимой переменной (признак, как качественный, так и количественный). Примером такой модели может являться поиск ассоциации веса животного с определенным ОНП. В общем случае модель записывается следующим образом:

$$Y = \beta * g + \sum(\alpha_i * z_i) + e \quad (13)$$

где β эффект генотипа g на признак Y , α_i эффекты ковариат z_i , e – средовой остаток.. В случае анализа такой модели оценивается эффект генотипа β и значимость его отличия от нуля (p-value). Такая модель является стандартной моделью для анализа ассоциаций. Далее будем называть такую модель одномерной (в противоположность многомерной).

Мы использовали метод “MANOVA”, описанный в разделе 1.3.1 обзора литературы для поиска новых локусов, ассоциированных с признаками мясной продуктивности. Значение p-value было рассчитано с помощью команды в программной среде R $pchisq(T_{score}^2, df = X, low = F)$, где T_score – это

полученные в виде вектора значения t критерия при использовании многомерного анализа, X – число степеней свободы (9 для признака MMass, 6 для признака MFat и 3 для признака MMeat (см. раздел 3.3).

В данной работе были отобраны 18 признаков из 56 из работы Volortmaa et al. 2016 и разбиты на 3 многомерных признака в зависимости от общности биологического происхождения признака (вес, жировая прослойка, мышцы). Каждый из полученных многомерных признаков был проанализирован с использованием суммарных статистик.

Также многомерный анализ использовался для анализа животных из российской выборки овец. 24 признака (описаны ниже) были объединены в 8 многомерных.

2.5 Функциональная аннотация и приоритизация локусов, найденных при многомерном анализе

Для проведения функциональной аннотации был использован Ensembl Variant Effect Predictor (VEP). Для анализа были отобраны ОНП, находящиеся в неравновесии по сцеплению с найденными ОНП ($r^2 > 0.7$). Анализ был осуществлен с использованием программного обеспечения, доступного онлайн [108].

Для приоритизации генов-кандидатов был использован список генов, расположенных в пределах ± 500 тыс. п.н. от ОНП с минимальным значением p -value. Для извлечения короткого резюме о гене, молекулярной функции и биологического процесса, в который вовлечен ген, мы использовали базу данных Uniprot (<http://www.uniprot.org>). Из-за недостаточной информации о генах у овец, в целом, по сравнению с информацией о генах для других видов, следующим шагом для выбранных генов-кандидатов были использованы база данных для человеческих генов OMIM [109] и база данных нокаутов мышинных генов [110].

2.5.1 Функциональная сеть взаимодействий генов, ассоциированных с мясной продуктивностью

Для тестирования гипотезы о функциональной связи найденных локусов был осуществлен попарный анализ корреляций между векторами эффектов ассоциированных ОНП для всех 56 признаков, так как все они связаны с мясной продуктивностью. Коэффициент корреляции Спирмана был посчитан для каждой пары из 12 значимо ассоциированных найденных локусов. В качестве порога значимости для корреляции Спирмана был использован $p\text{-value} \leq 7.6e-04$, который скорректирован на $\frac{12*11}{2} = 66$ тестов.

Для визуализации функциональной сети взаимодействия между генами был использован пакет «igraph» из среды R. Кластеризация и визуализация были осуществлены с помощью пакетов «corrplot» и «hclust» соответственно. Для кластеризации был оценен квадрат евклидовых расстояний путем вычитания абсолютных значений генетической корреляции из единицы. Был использован метод кластеризации Варда.

2.6 Выборка животных российской популяции для оценки племенной ценности признаков мясной продуктивности и подтверждения найденных локусов

Для проведения ПГИА была использована выборка животных, полученных коллегами из ВИЖ им. Л. К. Эрнста (г. Москва). Выборка состоит из 122 животных. В выборке представлены животные романовской породы и породы катадин, архары (*Ovis ammon*), гибриды различной кровности от скрещивания романовской породы овец и породы катадин, гибриды различной кровности романовской породы и дикого представителя рода *Ovis* архара (*Ovis ammon*), а также гибриды романовской породы и породы катадин с муфлоном (*Ovis gmelini*).

Животные были прогенотипированы с использованием чипа высокой плотности *Ovine Infinium® HD SNP Bead-Chip array* (606K), то есть для каждого животного имелась информация о более чем 600000 ОНП.

Для животных были доступны различные промеры туловища в разные промежутки жизни животного (6 дней, 42 дня, 3 месяца после рождения). Таблицы отражающие фенотипическое разнообразие исследуемой выборки представлены в Приложении (доп. таблицы 1-3).

2.6.1 Контроль качества генотипов

Для каждого набора данных все однонуклеотидные полиморфизмы (ОНП) были сопоставлены со сборкой генома OAR 3.1. Генотипы были закодированы в формате *IlluminaForward A / B*, а затем конвертированы в стандартный аддитивный числовой формат (по числу эффекторных аллелей, {0,1,2}), используя программную среду R. Далее эти данные были объединены в единый бинарный PLINK файл.

Контроль качества полученных генотипов был осуществлен с использованием программного обеспечения *plink 1.9* [111]. Был проведен анализ главных компонент генетической матрицы родства. Для этого генотипы были отфильтрованы по частоте минорного аллеля с использованием порогового значения 0.1, а также по качеству генотипирования с использованием порогового значения 0.05 (>95% успешно генотипированных особей по данному ОНП). График главных компонент представлен на рисунке 8. Первые 2 главные компоненты объясняют около 30 и 7 процентов общей дисперсии компонент соответственно. Показатель гетерозиготности среди всех животных в выборке составляет 0.42 (стандартное отклонение 0.01).

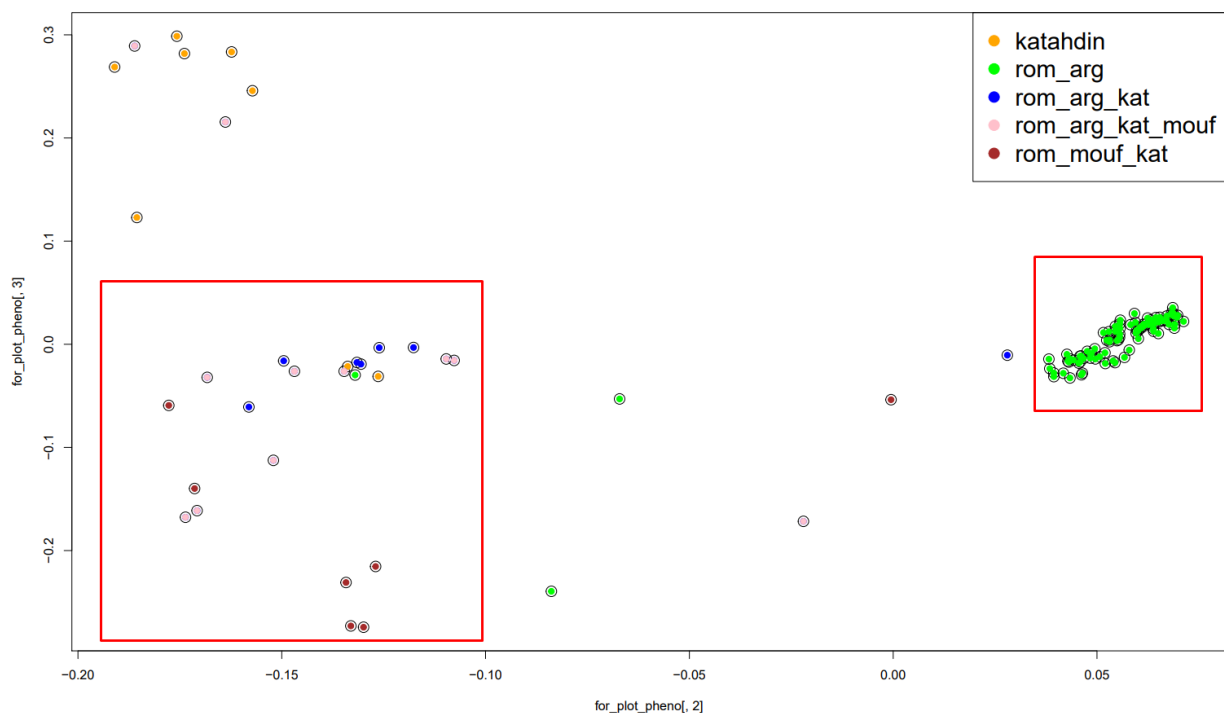


Рисунок 8. График главных компонент генетической матрицы родства для 122 животных. По оси ординат указана первая главная компонента, по оси абсцисс вторая главная компонента. Katahdin – катадины, rom_arg – гибриды романовской породы и архара, rom_arg_kat – гибриды романовской породы, катадинов и архара, rom_arg_kat_mouf – гибриды романовской породы, катадинов, муфлона и архара, rom_mouf_kat – гибриды романовской породы, муфлона и катадинов.

Согласно графику главных компонент, выборка разделилась на два генетических кластера. Первый состоит из гибридов романовской овцы и архара, полученных путем возвратного скрещивания (N=85). Второй состоит из гибридов различной кровности архаров, овец романовской породы, катадинов и муфлонов, а также чистокровных катадинов (N=23).

Для подтверждения результатов многомерного анализа были использованы животные из данных двух кластеров (N=108), а для оценки племенной ценности только гибриды, полученные путем скрещивания с архаром (N=94) для уменьшения влияния популяционной стратификации.

2.6.2 Контроль качества фенотипов

Замеры проводили во временных точках: 6 дней, 42 дня, 3 месяца и 6 месяцев после рождения. В последней временной точке количество животных было недостаточным для проведения ПГИА ($N < 10$), поэтому эти замеры далее не анализировались. Для каждой из трех оставшихся временных точек была собрана информация о 8 признаках: 7 индексов скорости роста и мясной продуктивности и масса животного. Информация об индексах представлена в Таблице 1. Таким образом, всего было доступно 24 признака (8 фенотипов в трех временных точках). Далее для каждого из 24 признаков было посчитано число животных с генотипами в каждом из двух генетических кластеров, определенных ранее. Для каждого фенотипа в каждом генетическом кластере мы провели поиск и удаление выбросов согласно правилу ± 3 IQR (interquartile range, межквартильное расстояние) [112]. Количество животных с доступными генотипами и фенотипами в первом и втором кластере отображено в Приложении (доп. таблица 4). Мы не проводили ПГИА для признака, если в группе было менее 20 животных.

Таблица 1. Индексы, использовавшиеся для анализа российской популяции овец

Название индекса	Сокращенное название
Длинноногости	Индекс 1
Растянутости	Индекс 2
Тазо-грудной	Индекс 3
Грудной	Индекс 4
Сбитости	Индекс 5
Перерослости	Индекс 6

Костистости	Индекс 7
-------------	----------

2.6.3 ПГИА на российской выборке

Для проведения ПГИА на имеющейся выборке животных была использована линейная смешанная модель, имплементированная в программном обеспечении GEMMA с использованием maximum likelihood теста (критерий максимального правдоподобия) [79]. Матрица родства была оценена с помощью EMMAX со стандартными параметрами [80]. Разные породы и гибриды анализировались отдельно друг от друга, а затем были объединены с использованием мета-анализа с фиксированными эффектами, взвешенными на обратную дисперсию. В итоге, всего было проведено 24 ПГИА для первого генетического кластера и 16 для второго, так как для признаков, измеренных через 3 месяца после рождения число животных во втором генетическом кластере было меньше 20 для каждого из 8 признаков.

Результаты ПГИА, полученные для разных генетических кластеров, были объединены с помощью мета-анализа. Для проведения мета-анализа было использовано программное обеспечение METAL [85].

Затем результаты мета-анализа были использованы для создания многомерных признаков и применения многомерной модели (MANOVA), описанной выше (см пункт 2.4). Было сформировано 8 многомерных признаков, 7 для индексов и 1 для массы. В каждый из признаков вошли объединенные данные по 3 временным точкам (6 дней, 42 дня, 3 месяца соответственно). Факторы инфляции тестовой статистики для всех одномерных и многомерных признаков приведены в Приложении (доп. таблица 5).

2.6.4. Подтверждение локусов из базы данных QTL

Локусы из базы данных QTL были подтверждены на российской выборке овец. Для этого с учетом пороговых значений по частоте минорного

аллеля $>1\%$ и качеству генотипирования $>95\%$ было отобрано 75 локусов (включая 6 обнаруженных в рамках данной диссертационной работы). В качестве порогового было использовано значение $\frac{0.05}{8*75} = 8.3e - 05$, где 75 – количество локусов, 8 – количество многомерных признаков.

2.7 Анализ колокализации

Одним из преимуществ созданной нами платформы является возможность проведения так называемого анализа колокализации (SMR- θ) [113]. Данный анализ позволяет сравнивать паттерны ассоциации конкретного геномного локуса с разными признаками и делать выводы о наличии или отсутствии плейотропного эффекта данного локуса на те или иные признаки. Вкратце, при справедливости гипотезы о плейотропном эффекте локуса, ожидается, что отношение оценок эффектов ОНП на два признака в изучаемом локусе должны отличаться между различными ОНП незначительно. Если паттерны ассоциаций в регионе разные – то есть отношение эффектов значительно варьируется от ОНП к ОНП – скорее всего, в локусе присутствуют разные для каждого признака функциональные полиморфизмы, находящиеся в неравновесии по сцеплению. Статистика θ – это взвешенная корреляция, при подсчёте которой используется информация о значениях p -value и направлении эффекта. Высокое абсолютное значение (например, $|\theta| > 0.7$) говорит в пользу плейотропного эффекта локуса на изучаемые признаки. Знак θ указывает на то, имеют ли ОНП в изучаемом локусе одинаковое направление эффектов на исследуемые признаки (положительное значение величины θ), или разные направления эффектов (отрицательное значение величины θ). Пример анализа колокализации представлен на рисунке 9.

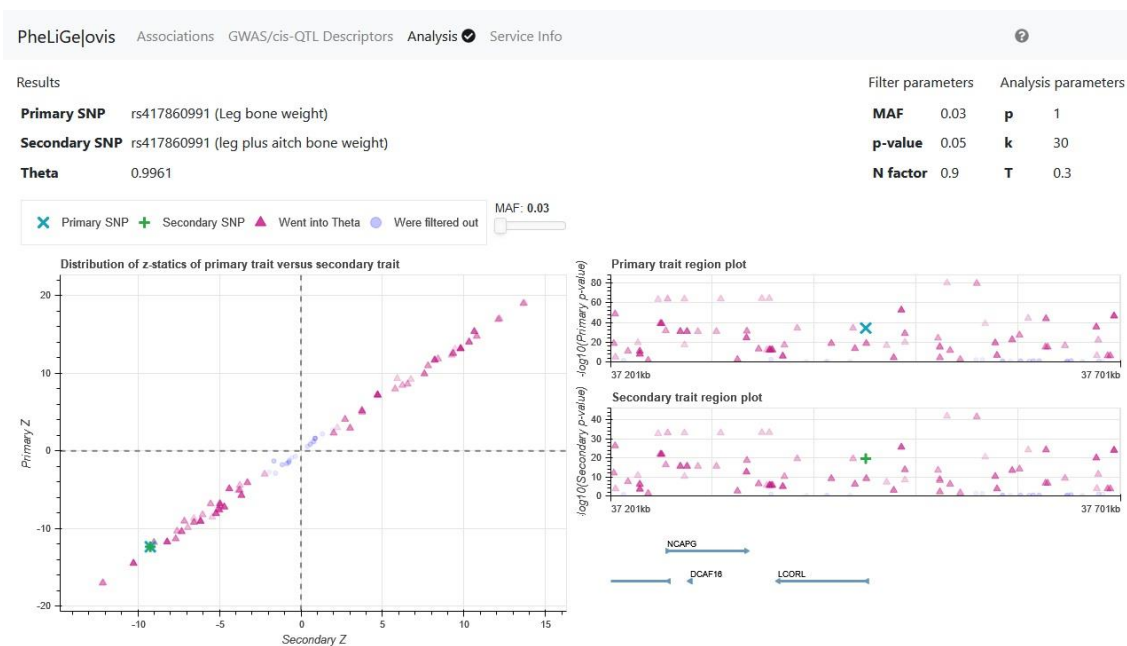


Рисунок 9. Пример выдачи результатов анализа колокализации, сделанного с использованием платформы GWAS-MAP|ovis. В левой части представлено совместное распределение Z-статистик для основного признака, вес костей голени (Leg bone weight, LEGBONE) (ось абсцисс) и вторичного признака, вес костей ноги с пяточной костью (leg + aitch bone weight, BONE) (ось ординат). В правой части представлены региональные графики ассоциаций для каждого из признаков (LEGBONE и BONE соответственно). Значение θ представлено в верхнем левом углу рисунка над графиком.

3. Результаты

3.1 Создание платформы для комплексного анализа результатов полногеномных исследований ассоциаций

3.1.1 Создание каталога неравновесия по сцеплению

На первом шаге был создан каталог неравновесия по сцеплению с использованием данных по 18 овцам романовской породы, 6 - породы катадин, 10 архарам, 48 гибридам F1 от скрещивания F1 (романовской овцы и архара) и романовской овцы, и 14 гибридам F1 от скрещивания романовской овцы и катадинов. Все животные были прогенотипированы с использованием чипа высокой плотности Infinium® HD SNP Bead-Chip (606060 ОНП, Illumina Inc., San Diego, CA, USA). Для каждого ОНП доступна информация о rs_id, хромосоме, позиции (на основе сборки генома OAR_V3.1), эффекторном и референсном аллеле, а также частоте, посчитанной на 96 животных. В ходе подготовки каталога неравновесия по сцеплению были удалены все дубликаты (по rs_id и хромосоме и позиции), аллели приведены в лексикографический порядок, удалены мономорфные ОНП, удалены половые хромосомы. На данный момент каталог неравновесия по сцеплению содержит в себе информацию о 523578 ОНП. Распределение ОНП по аутосомам показано в Таблице 2. Распределение частот минорных аллелей в каталоге неравновесия по сцеплению показано на Рисунке 10.

Таблица 2. Распределение ОНП по аутосомам в каталоге неравновесия по сцеплению

Хромосома	Количество ОНП	Среднее количество ОНП на 1 Мб
1	58334	211.64
2	52833	212.19

3	48225	215.05
4	25302	212.05
5	22911	213.79
6	24300	207.66
7	21462	214.48
8	19359	213.6
9	19775	208.36
10	17694	204.88
11	14062	226.24
12	17220	217.9
13	17907	215.49
14	13996	223.11
15	17363	214.63
16	15151	211.28
17	15283	211.89
18	14659	214.5
19	13335	220.77
20	11199	219.2
21	10557	210.96
22	11112	219.22
23	12990	208.13
24	9311	221.45
25	9882	217.64
26	9356	212.45

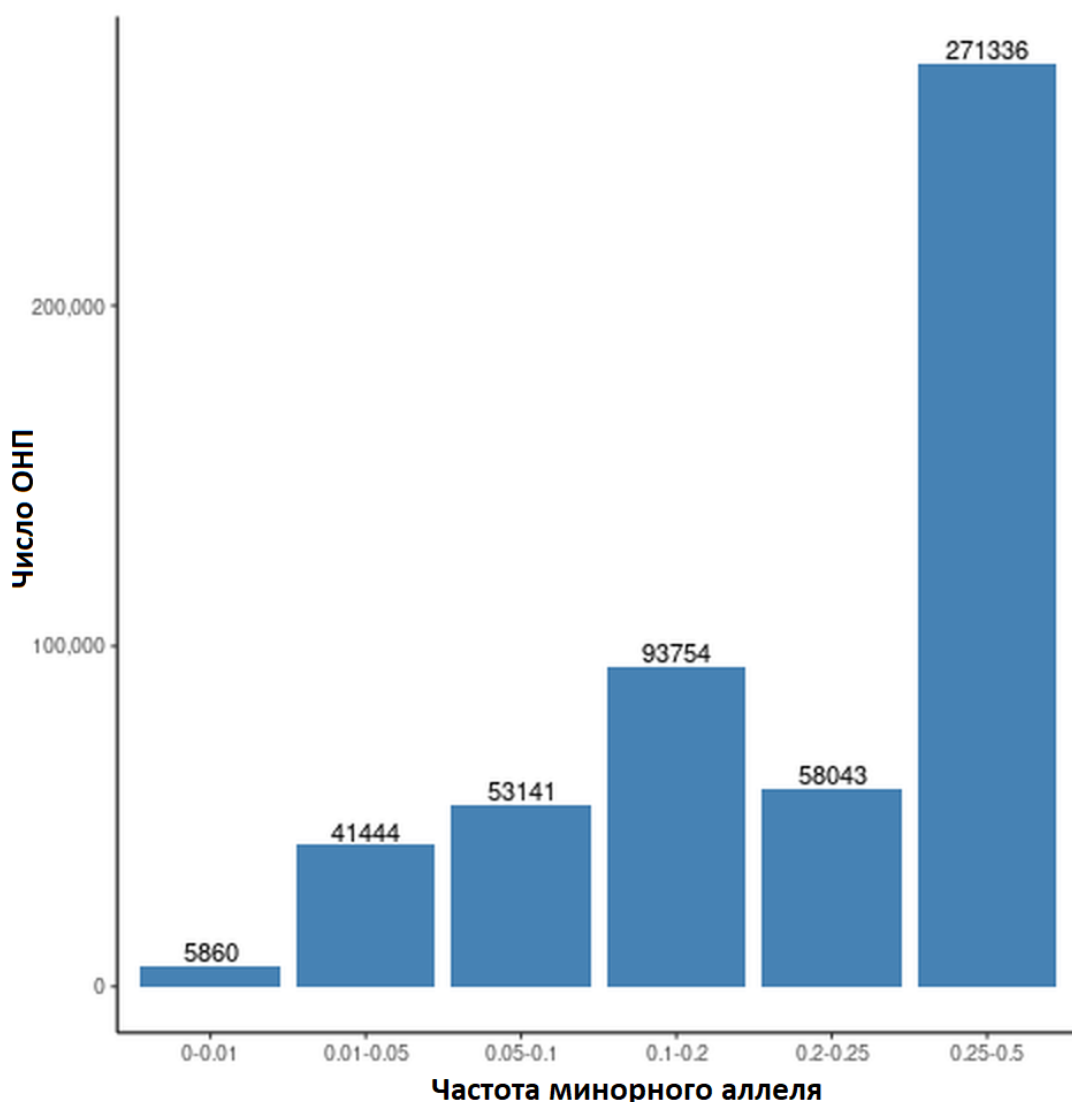


Рисунок 10. Количество ОНП в каталоге неравновесия по сцеплению с заданной частотой минорного аллеля.

3.1.2 Формирование структуры платформы для хранения результатов ПГИА

Созданная платформа состоит из двух компонент (Рисунок 11). Первая компонента БД хранит информацию об ассоциациях между генотипом и фенотипом – параметры однонуклеотидного полиморфизма (ОНП) и его эффекта на фенотип.

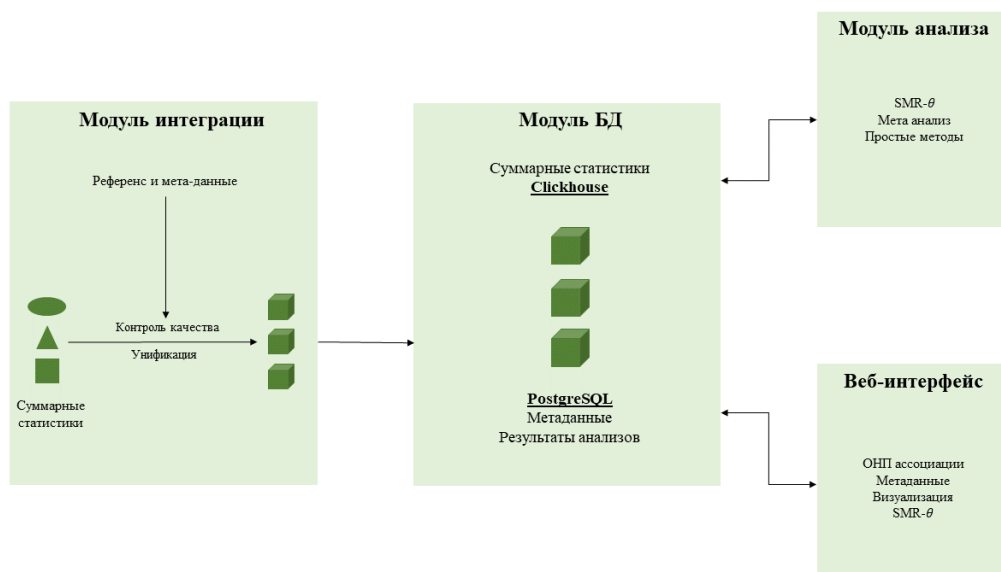


Рисунок 11. Устройство платформы GWAS-MAP|ovis. Стрелки между модулями обозначают возможные направления передачи данных.

Первая компонента платформы находится под управлением СУБД Clickhouse и состоит из двух физических таблиц с одинаковой структурой и виртуальной таблицы, позволяющей делать запрос одновременно из первых двух. Физическая таблица `ovis.gwas_snr` хранит данные результатов ПГИА. Данная компонента БД содержит 488074 записей в таблице, доступ к которым производится через единый запрос к виртуальной таблице `ovis.snr`.

Вторая компонента платформы хранит метаданные об исследовании – дескрипторы, и находится под управлением СУБД PostgreSQL, в которой содержится одна таблица (таблица дескрипторов). Таблица дескрипторов содержит 56 строк, по одной на каждый результат ПГИА. Таблицы платформы из разных её компонент связаны через идентификатор исследования (`gwas_id`). Одной записи в таблице дескрипторов соответствует множество записей в таблице ОНП.

Для удобства использования был создан веб-интерфейс существующей базы данных, который доступен по адресу <https://pheligeovis.icgbio.ru/>. В

колонке «rsid» указывается идентификатор необходимого ОНП, также можно задать необходимые пороги p-value и/или r^2 , специфицировать признак, для которого необходимо произвести поиск и выбрать коллекцию (Рисунок 10).

Разработанная платформа, состоящая из двух баз данных, хранящих суммарные статистики и метаданные, и веб-интерфейса для доступа к данным, доступна через общедоступный (веб) (Рисунок 12) и частный (на основе ssh) интерфейсы. Общие принципы работы веб-интерфейса и основных инструментов отражены в онлайн-помощнике при нажатии на красную кнопку под меню «Associations» в верхнем левом углу экрана на главной странице платформы.

Платформа результатов полногеномных и региональных исследований ассоциаций экономически-важных признаков овец (GWAS-MAP|ovis) на данный момент содержит в себе информацию о 80 результатах ПГИА), связанных с мясной продуктивностью, формированием каркаса и ростом у овец. Одни из результатов ПГИА были получены с использованием выборки из более чем 10000 овец 9 различных пород [16]. Для всех животных была доступна информация о генотипировании с использованием чипов высокой плотности (более 450000 ОНП). Таким образом, БД содержит в себе информацию о более чем 25 миллионах ассоциаций генетических полиморфизмов с признаками.

The screenshot shows the search interface of the GWAS-MAP|ovis platform. At the top, there is a navigation bar with tabs: 'PheLiGe-ovis', 'Associations', 'GWAS/cis-QTL Descriptors', 'Analysis' (which is active and highlighted in blue), and 'Service Info'. Below the navigation bar, there are several input fields for filtering results: 'rsID' (with a text input box and an example 'E.g. rs413799815, rs417860991'), 'P-value' (with a text input box), 'R^2' (with a text input box), 'Trait OR Trait Abbreviation' (with a text input box and an example 'E.g. Leg bone weight, Lean meat yield'), and 'Collection' (with a dropdown menu). To the right of these fields are two buttons: 'Add Filter' and 'Reset'. At the bottom left of the search area, there are two buttons: 'Search' (in blue) and 'Download CSV' (in grey).

Рисунок 12. Скриншот веб-интерфейса платформы GWAS-MAP|ovis

3.2 Создание базы данных QTL

Для создания базы данных генов и QTL, связанных с признаками мясной продуктивности, формированием каркаса и ростом у овец, был осуществлен

поиск статей с использованием Pubmed (pubmed.com) и Google Scholar (scholar.google.com). Мы использовали 15 различных комбинаций ключевых слов, с помощью которых было найдено 153 статьи (включая дубликаты). Затем мы отобрали наиболее релевантную работу, связанную с исследуемыми признаками, для того чтобы исследовать дерево цитирований данных работ для увеличения списка найденных статей. Мы выбрали работу Zhang et al. 2013 [93], и обнаружили 16 релевантных статей, 9 из которых уже были найдены нами на первом этапе поиска статей. Затем из этих найденных 16 статей мы отобрали 2 наиболее релевантных [97, 114] и путем исследования дерева цитирований обнаружили 1 дополнительную статью. По окончании данного шага после удаления всех дубликатов был составлен список из 46 статей для дальнейшего анализа.

Все статьи из полученного списка были вручную просмотрены и отфильтрованы по релевантности. После данного шага был получен список из 18 статей, на основе которых была создана база данных генов и QTL, ассоциированных с признаками мясной продуктивности, формирования каркаса и ростом у овец. Девять статей из 18 были сделаны с использованием ПГИА [93–100, 115], 7 статей описывали ген-кандидатные исследования [116–122] и 2 статьи были посвящены обзорам [123, 124].

Из 18 отобранных статей была извлечена информация о генах и QTL, ассоциированных с исследуемыми признаками. В результате был получен список из 198 генов (172 уникальных гена), ассоциированных с признаками мясной продуктивности, формирования каркаса и ростом у овец. Обнаруженная общая тенденция указывать ближайший ген к найденному ассоциированному ОНП может привести к ошибке интерпретации: функциональный ген (который действительно ассоциирован с исследуемым признаком) не обязательно является ближайшим к ОНП [125].

Также была извлечена вся доступная информация об ассоциированных ОНП с исследуемыми признаками. Был получен список из 163 ОНП (130 уникальных ОНП). Однако для найденных ОНП в статьях были использованы различные форматы обозначения: формат, связанный с данными микрочипирования (например, OAR21_7449077.1), формат названия ОНП с приставкой rs, установленный авторами формат имени (например, SNP23), формат обозначения ОНП через хромосому и позицию. В дополнение было обнаружено, что только в 5 из 18 работ содержится информация об эффекторном аллеле (1 работа ПГИА и 4 ген-кандидатных исследований) [116, 118–120, 122] и только 3 работы из 17 содержали в себе информацию о размере эффекта для ОНП [95, 98, 122]. Восемьдесят девять из найденных 130 ОНП представлены в платформе для генотипирования Ovine Infinium® HD SNP Bead-Chip array (606K).

В результате, была создана база данных QTL и генов, ассоциированных с признаками мясной продуктивности, формирования каркаса и ростом у овец, которая содержит в себе информацию об 198 (172 уникальных) генах и 163 (130 уникальных) ОНП на основе 17 отобранных статей. На данный момент пересечение с базой данных «AnimalQTLdb», в которой содержится информация о 390 локусах, связанных с мясной продуктивностью и формированием каркаса, составляет 38 локусов. Также стоит отметить, что по сравнению с «AnimalQTLdb созданная в рамках данной диссертационной работы база обладает рядом принципиальных отличий: 1) Наличие информации об эффекторном аллеле, размере и направлении эффекта (если такие данные были предоставлены авторами в статье). 2) Удобство поиска ОНП и/или генов, ассоциированных с признаками мясной продуктивности у овец (возможность быстро находить интересующий ОНП и/или ген). 3) Наличие точной физической позиции в геноме для ОНП. 4) Наличие только значимо-ассоциированных локусов. 5) Ручной отбор статей для пополнения

базы. Актуальная версия базы данных находится в открытом доступе по адресу: https://github.com/Defrag1236/Ovines_2018.

3.3 Многомерный анализ ассоциаций признаков мясной продуктивности из платформы GWAS-MAP|ovis

Пятьдесят шесть одномерных признаков из статьи Volormaa et al. 2016 были сгруппированы в 3 многомерных признака. Признаки были сгруппированы по принципу биологической близости. Признаки, связанные с весом, были объединены в признак MMass; признаки, связанные с жиром - в признак MFat; а признаки, связанные с мышцами - в признак MMeat. Для создания многомерных признаков было использовано 18 одномерных признаков из 56 представленных в статье, так как некоторые признаки не удалось присоединить ни к одному многомерному признаку. Группы, образующие многомерные признаки, представлены в таблице 3. Далее, с помощью данных, загруженных в платформу, была посчитана матрица фенотипических корреляций между одномерными признаками, входящими в состав многомерных признаков (см. Приложение, доп. рисунки 1-3).

Таблица 3. Состав многомерных признаков, использовавшихся для проведения многомерного анализа

Многомерный признак	Оригинальный признак из статьи	Название признака на русском языке	Аббревиатура
MMass	Slaughter weight	Убойная масса туши	PSWT (kg)
	Hot carcass weight	Масса освежеванной туши	HCWT (kg)
	Loin weight	Масса поясницы	LLWT (g)
	Topside weight	Масса огузка	TOP (g)
	Round weight	Масса задней ноги	RND (g)
	leg + aitch bone weight	Масса ноги и огузка с костью	BONE (g)
	Leg bone weight	Масса ноги с костью	LEGBONE (g)
	Lean meat yield	Процент выхода мяса	LMY (%)
	Dressing	Масса освежеванной туши в процентах	DRESSING (%)
MFat	Intramuscular fat	Внутримышечный жир	IMF (%)

	Loin fat weight	Масса жира в пояснице	LLFAT (g)
	Carcass fat score	Оценка содержания жира в туше	CFATSCORE
	Fat Depth C	Толщина жира C	CCFAT (mm)
	Fat depth GR	Толщина жира GR	HGRFAT (mm)
	Fat Depth 5th rib	Толщина жира в районе 5 ребра	CFAT5 (mm)
MMeat	Carcass eye muscle width	Ширина мышечных глазков	CEMW (mm)
	Carcass eye muscle depth	Глубина мышечных глазков	CEMD (mm)
	Eye muscle area	Площадь мышечных глазков	CEMA (cm ²)

3.3.1 Результаты многомерного анализа

Для 3 многомерных признаков было найдено 12 значимо ($p\text{-value} < 3.38 \times 10^{-8}$) ассоциированных локусов (Таблица 4). Девять из них ассоциированы с MMass и 3 с MFat. Из 12 найденных локусов, 4 были описаны ранее [16]. Результаты многомерного анализа в виде «Manhattan plot» представлены на рисунке 13.

Таблица 4. Двенадцать локусов, ассоциированных как минимум с одним многомерным признаком. Новые локусы отмечены звездочкой. Жирным шрифтом указаны значимые $p\text{-value}$. pa/эa – референсный и эффекторный аллель соответственно.

ОНП	Хромосома	Позиция	$p\text{-value}_{\text{MMass}}$	$p\text{-value}_{\text{MFat}}$	$p\text{-value}_{\text{MMeat}}$	pa/эa
rs193632759*	1	198273462	1.14E-08	0.95	0.11	T/G
rs420734786*	2	68158297	5.12E-09	0.56	0.74	A/C
rs428034699*	3	60513720	2.08E-09	0.3	0.38	T/C
rs403766990*	3	153924034	0.02	1.93E-08	0.05	A/G
rs428034699*	3	219082890	2.78E-09	0.25	0.41	T/C
rs401990068*	4	85985834	0.41	9.62E-09	0.9	T/G
rs408893215*	5	59475661	0.06	2.31E-08	0.01	T/C
rs401834107	6	37530647	6.04E-09	9.73E-24	5.11E-12	T/C
rs161042491	11	26445930	2.34E-04	2.64E-10	1.15E-06	A/G

rs405660596	16	31871071	3.28E-09	0.03	8.73E-04	A/C
rs408838557	18	62894338	4.64E-12	5.04E-08	1.29E-03	A/C
rs418394153 *	23	44492468	4.27E-09	0.6	0.28	A/G

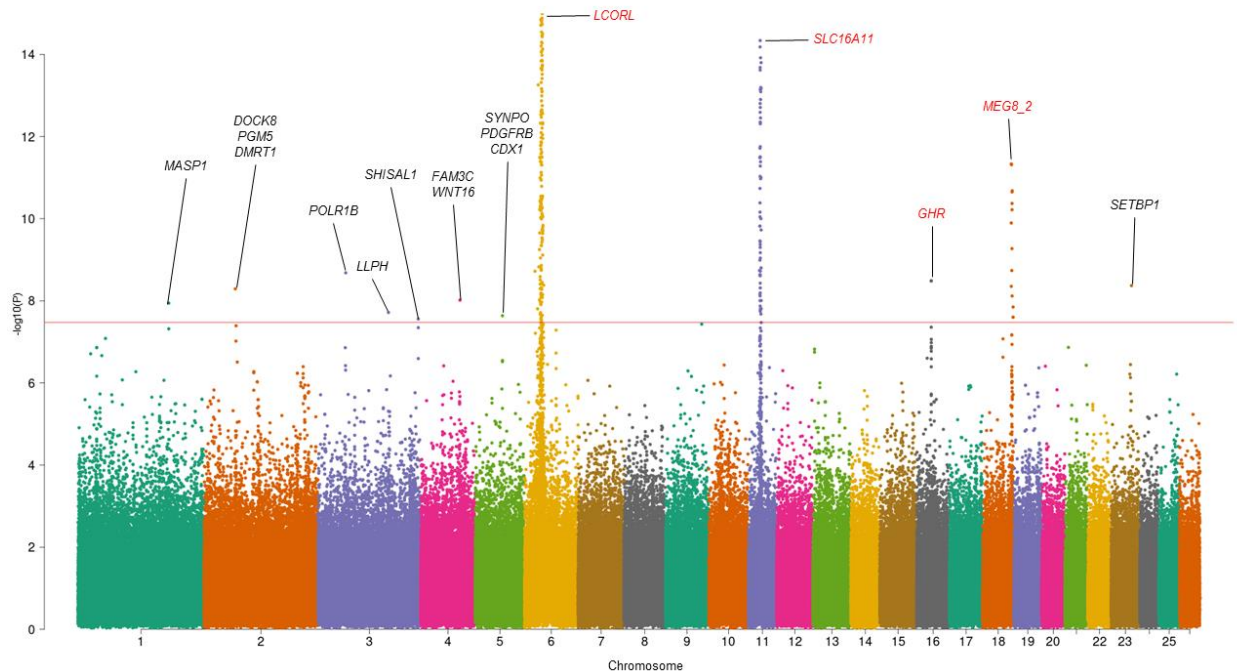


Рисунок 13. Для каждого значимо ассоциированного локуса представлены названия приоритизированных генов. Известные локусы и гены в них обозначены красным.

3.3.2 Приоритизация генов в найденных локусах

Для 8 новых найденных локусов мы провели ряд анализов для приоритизации генов. Прежде всего, мы использовали VEP (Variant Effect Predictor). Была обнаружена одна миссенс мутация (rs193632759, MET > ILE), расположенная в гене *MASP1*. Этот ген вовлечен в систему комплемента лектина, которая облегчает распознавание патогенов через олигосахаридные цепи и их последующее устранение [126].

Следующим шагом стало проведение приоритизации и аннотации генов на основе литературных данных (таблица 5). В результате было приоритизировано 13 генов для 8 новых локусов.

Таблица 5. Список приоритизированных генов.

Локус	Количество генов в регионе	Ближайший ген	Гены с миссенс мутацией	Приоритизация на основе литературы	Гены-кандидаты
rs193632759	14	<i>MASPI</i>	<i>MASPI</i>	<i>MASPI</i>	<i>MASPI</i>
rs420734786	5	<i>DOCK8</i>		<i>PGM5/DMRT1</i>	<i>DOCK8/PGM5/DMRT1</i>
rs399851221	12	<i>POLR1B</i>		<i>POLR1B</i>	<i>POLR1B</i>
rs403766990	4	<i>LLPH</i>		<i>LLPH</i>	<i>LLPH</i>
rs428034699	11	<i>SHISALI</i>			<i>SHISALI</i>
rs401990068	6	<i>FAM3C</i>		<i>WNT16</i>	<i>FAM3C/WNT16</i>
rs408893215	20	<i>SYNPO</i>		<i>PDGFRB/CDX1</i>	<i>SYNPO/PDGFRB/CDX1</i>
rs418394153	1	<i>SETBP1</i>		<i>SETBP1</i>	<i>SETBP1</i>

Для локуса rs420734786 не оказалось возможным приоритизировать какой-либо один конкретный ген. Самый ассоциированный ОНП расположен в гене *DOCK8*, который вовлечен в процессы иммунной реакции. Изменения в этом гене у человека приводят к комбинированной иммунной недостаточности [127]. Также в дополнение к этому гену имеются 2 близко расположенных гена: *PGM5*, который вовлечен в развитие мышечного корсета [128] и *DMRT1*, который вовлечен в процесс детерминации пола и имеет систематический эффект на развитие каркаса [129].

В локусе rs401990068 было приоритизировано 2 гена: *FAM3C* и *WNT16*. *FAM3C* – это ближайший ген к самому ассоциированному ОНП. Данный ген вовлечен в развитие многоклеточного организма. В свою очередь, ген *WNT16* в сигнальный путь Wnt/CTNNB1 стимуляции синовиального сустава. [130]

На пятой хромосоме в локусе rs408893215 внимание было сконцентрировано на 3 генах. Ближайший к самому ассоциированному ОНП ген *SYNPO* непосредственно участвует в связывании актина. Ген *PDGFRB* кодирует бета-рецептор фактора роста тромбоцитов. Несколько исследований указывают этот ген как потенциальную мишень для терапии у человека [131–133]. Третий из приоритизированных генов *CDX1* экспрессируется в толстой кишке и кишечнике взрослых людей [134].

Для локусов rs399851221, rs428034699, rs403766990 и rs418394153 ближайшие гены (*POLR1B*, *LLPH*, *SETBP1*, *SHISAL1*) были приоритизированы как возможные кандидаты для дальнейших исследований. Ген *POLR1B* вовлечен в развитие структуры костей [135]. Дефицит продукта гена *SETBP1* может являться причиной синдрома Шинцеля-Гидиона у людей. Один из симптомов этого состояния - это стагнация роста у детей. Продукт этого гена также ассоциирован с развитием костей, а также может вызывать рак костей или соединительной ткани [136].

3.3.3 Функциональная сеть взаимодействий ассоциированных локусов

После проведения приоритизации генов была протестирована гипотеза о функциональной связи генов в найденных локусах. Функциональная связь генов подразумевает, что гены, функционально связанные и/или вовлеченные в одну генную сеть, будут демонстрировать и аналогичную ассоциацию с исследуемыми признаками (в нашем случае 56 признаков). Идея данной гипотезы представлена в работах Volormaa с соавторами [15, 16] и в недавней работе по изучению уровней гликозилирования иммуноглобулина класса-G (IgG) человека [14]. Для проверки данной гипотезы мы рассчитали попарную корреляцию Спирмана между Z-статистиками ассоциированных ОНП для 56 признаков из оригинальной работы Volormaa с соавторами 2016. В данном случае знак корреляции не имеет значения из-за того, что направление эффекта зависит от того, какой аллель использовался как референсный. График корреляций Спирмана представлен на рисунке 14. Функциональный граф для генов как минимум с одной значимой корреляцией ($p\text{-value} \leq 7.6e-04$) представлен на рисунке 15.

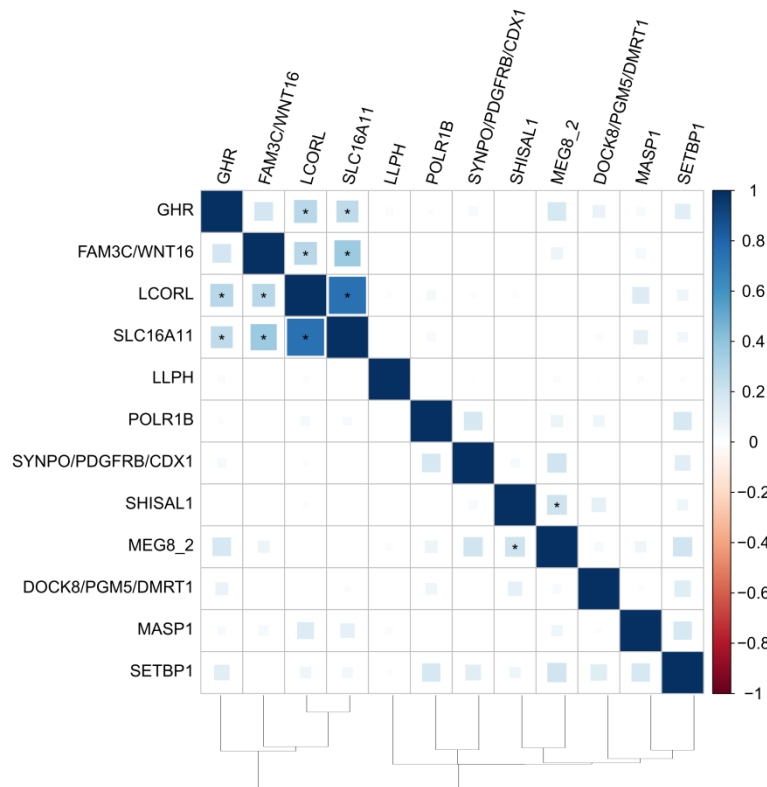


Рисунок 14. Иерархическая кластеризация попарных корреляций Спирмана. Каждая значимая корреляция обозначена звездочкой ($p\text{-value} \leq 7.6e-04$). Интенсивность цвета и размер квадрата зависят от величины коэффициента корреляции.

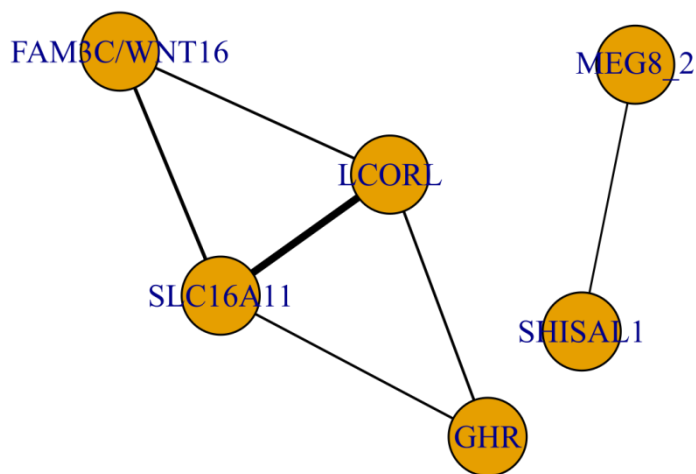


Рисунок 15. Функциональная сеть взаимодействий генов, связанных с мясной продуктивностью. Каждая линия представляет собой значимую

корреляцию Спирмана ($p\text{-value} \leq 7.6e-04$). Толщина линии прямо пропорциональна размеру корреляции.

Наибольшая корреляция наблюдается между 2 известными локусами на 6-й (*LCORL*) и 11-й (*SLC16A11*) хромосомах соответственно ($r = 0.74$). Также эти локусы связаны с генами *FAM3C/WNT16* и *GHR*. Другой из известных локусов на 18 хромосоме (*MEG8_2*) оказался связан с новым найденным локусом на 3 хромосоме (*SHISAL1*).

Для всех приоритизированных генов мы также проверили наличие функциональных связей между белками с помощью базы данных «String» [137]. В результате было выявлено функциональное взаимодействие только для 2 белков, кодируемыми генами *DOCK8* и *DMRT1*.

3.3.4 Подтверждение найденных локусов на российской популяции овец

С использованием российской популяции овец ($N=108$) было проведено подтверждение 6 из 8 новых найденных локусов и локусов из базы данных QTL. В качестве материала для подтверждения локусов были использованы результаты многомерного анализа для 8 признаков, связанных с мясной продуктивностью: 7 индексов (см. таблица 1) и показатель живой массы. Было подтверждено два локуса: rs401990068, найденный в данной диссертационной работе, около генов *FAM3C* и *WNT16* ($p\text{-value} < 6.08e-05$), ассоциированный с индексом длинноногости и rs401834107 (ген *LCORL*) ($p\text{-value} < 1.1e-06$), ассоциированный с тазо-грудным индексом (Рисунок 16). Для rs401990068 и rs401834107 мы подтвердили соответствие между аллелями (G для rs401990068 и C для rs401834107), повышающими мясную продуктивность в российской популяции овец и в оригинальном исследовании S. Volormaa с соавторами.

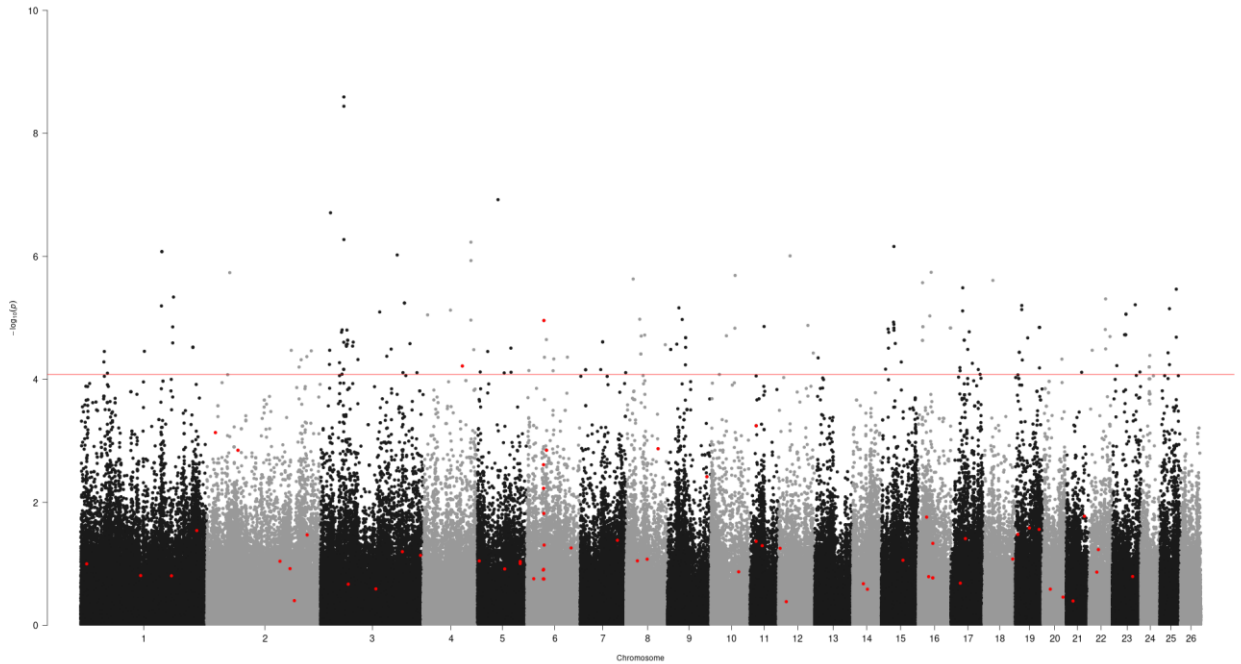


Рисунок 16. Manhattan plot для 8 многомерных признаков, связанных с мясной продуктивностью. Для каждого ОНП был использован минимальный p-value среди 8 многомерных признаков. Исследуемые ОНП обозначены красными точками.

3.4 Анализ колокализации

С помощью созданной нами ранее платформы для хранения, унификации и анализа результатов ПГИА для признаков мясной продуктивности у овец (см. раздел 4.2) был проведен анализ колокализации для самого значимо ассоциированного сигнала, полученного с помощью применения многомерного анализа на данных Volormaa et. al 2016 (см. раздел 4.3). Для ОНП rs401834107 (ген *LCORL*) были выбраны признаки, для которых значение p-value < 5e-07. Выбор такого порога обусловлен тем, что Volormaa с соавторами в своем исследовании использовали данный порог. В результате с 12 признаками данный ОНП был ассоциирован с заданным порогом значимости. ОНП rs401834107 из этих 12 признаков был сравнен попарно с использованием метода SMR- θ [113] для установления того, сходны ли региональные профили ассоциации. Результаты для анализа методом θ представлены на рисунке 17. В данном случае знак указывает на то, сходным

(+) или противоположным (-) образом, данный ОНП влияет на исследуемые в признаки. Как видно из тепловой карты, признаки образуют два кластера, отличающиеся направлением корреляции между паттернами, но при этом сильно связанные друг с другом. Первый кластер состоит из признаков, связанных с массой и содержанием жира в организме, второй из признаков связанных с массой отдельных частей туши. Отдельно стоит отметить два выпадающих признака - убойная масса туши и ширина мышечных глазков, для которых некоторые значения θ оказались низкими (отсутствие плеiotропии). Однако данные признаки по отдельности показывают плеiotропный эффект с другими признаками, поэтому мы считаем, что низкие значения θ в данном случае это либо статистический артефакт, либо следствие недостаточной мощности анализа. Таким образом, было установлено, что локус rs401834107 имеет плеiotропный эффект на двенадцать различных признаков, связанных с мясной продуктивностью у овец. Данный анализ является первым применением в мире метода колокализации на овцах.

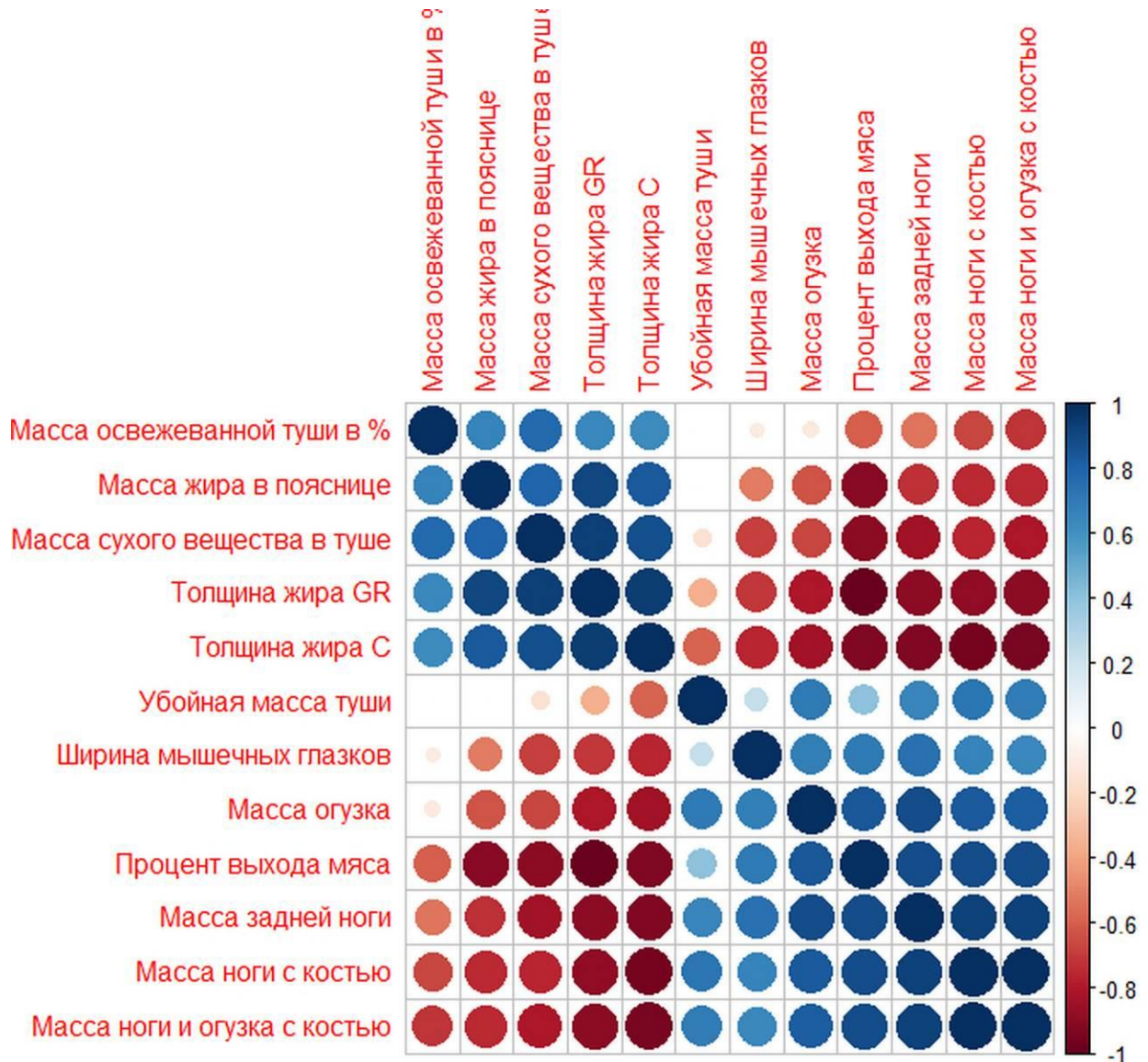


Рисунок 17. Результаты анализа колокализации для ОНП rs401834107 на 12 оригинальных признаках из статьи S. Volormaa с соавторами. Красным цветом показан отрицательный знак значения θ , голубым цветом показан положительный знак значения θ . Чем ярче цвет, тем ближе абсолютное значение к единице.

3.5 Оценка племенной ценности признаков мясной продуктивности на российской выборке овец

В дополнение к вышеописанному анализу плеiotропии, еще одним из преимуществ базы является возможность использования уже загруженных суммарных статистик для поиска маркеров для проведения маркер-

ориентированной селекции. Мы провели поиск маркеров для признаков, входящих в состав многомерного признака Mmass (см. Таблица 3) и оценили их предиктивный потенциал на российской выборке овец.

Мы провели процедуру поиска независимых локусов (клампинг). Для этого с помощью частного (на основе ssh) интерфейса мы выбрали интересующие нас суммарные статистики (признаки входящие в многомерный признак Mmass), а также установили порог значимости $p\text{-value} < 1e-07$. Для дальнейшего анализа были отобраны те признаки, для которых было обнаружено не менее 3 независимых (не находящихся в неравновесии по сцеплению) локусов (Таблица 6). Данный порог был выбран для увеличения мощности анализа при использовании небольшой выборки животных (N=94). Полученные 6 таблиц, содержащие в себе информацию об ОНП с наибольшей ассоциацией для каждого локуса), их хромосомах, позиции, эффекторном и референсном аллелях, эффекте и $p\text{-value}$ были использованы для дальнейшего анализа.

Таблица 6. Результаты оценки племенной ценности для 6 признаков, связанных с мясной продуктивностью. Разница в живой массе посчитана для животных из верхнего квартиля в сравнении со всеми животными. Расшифровка названий признаков указана в таблице 3.

Признак	ОНП	p-value лог.регрессия 6 дней	p-value критерий t 6 дней	Разница в средней массе 6 дней (в граммах)
HCWT	rs406365427 rs427891980 rs423487570 rs161042491	0.03	0.67	55
DRESSING	rs428034699 rs417971156 rs419905895 rs405560019	0.78	0.98	4

LEGBONE	rs401834107 rs161042491 rs428862267	0.05	0.12	-261
LMY	rs426448654 rs401834107 rs417149986	0.09	0.13	-265
PSWT	rs427891980 rs161042491 rs413684575	0.07	0.47	90
RND	rs429410401 rs402556633 rs161042491	0.81	0.3	-127

Используя найденные локусы, мы оценили племенную ценность (ПЦ) животных на выборке российской популяции овец. Мы использовали данные о 94 животных (гибриды от возвратного скрещивания романовской породы овец и архара) для уменьшения влияния популяционной стратификации. Племенная ценность была оценена по следующей формуле:

$$\text{ПЦ} = \sum b * g \quad (14)$$

где b – эффект эффекторного аллеля самого ассоциированного ОНП каждого локуса, g – генотипы 94 животных для данного ОНП в формате 0,1,2 (закодированные по количеству эффекторного аллеля).

Далее, с помощью линейной регрессии мы установили, насколько значимо оцененная ПЦ может предсказывать фенотип особи. В качестве признаков была выбрана живая масса животного через 6 и 42 дня после рождения. Модель выглядела следующим образом:

$$p \sim \text{ПЦ} + \text{cov} \quad (15)$$

где p – значения фенотипа, ПЦ – племенная ценность выбранных локусов, cov – ковариаты, которые включали в себя информацию о поле, количестве ягнят в приплоде, а также 2 первые главные компоненты матрицы

родства. В результате, оцененная ПЦ с использованием признака HCWT (вес туши после забоя) была номинально значимо ассоциирована с признаком живая масса через 6 дней после рождения ($p\text{-value} = 0.03$), что говорит о том, что отобранные нами локусы значимо влияют на показатель массы животного (Таблица 6). Остаточное стандартное отклонение для модели с использованием признака HCWT составило 0.61. Диаграмма рассеяния с регрессионной прямой представлена в приложении (доп. рисунок 4).

В дополнении к проведенному анализу мы оценили, какой прирост в массе ожидается при использовании животных со значением ПЦ, которое входит в четвертый верхний квартиль ($>75\%$). Используя t -критерий Стьюдента, мы сравнили средние фенотипа, поправленного на вышеотмеченные ковариаты, во всей выборке и выборке животных, попавших в верхний квартиль по ПЦ. Было установлено, что наибольший прирост в массе демонстрируется для ПЦ признака PSWT, 90 грамм, хотя отличия были статистически незначимы ($p\text{-value}=0.47$). Некоторые из оцененных значений ПЦ для других признаков демонстрируют снижение массы для животных из верхнего квартиля. В целом, $p\text{-value}$ получились незначимы, однако номинально значимые результаты для одного из признаков указывают, что это, скорее всего, является следствием недостаточной мощности анализа из-за небольшого размера выборки. Таким образом, маркеры, отобранные с использованием разработанной платформы и увеличивающие среднюю массу животного, потенциально могут быть применимы для разработки тест-систем для ведения маркер-ориентированной селекции у овец.

4. Обсуждение

В этой работе мы провели комплексный поиск генов и локусов, ассоциированных с показателями роста и развития у овец, а также создали платформу GWAS-MAP|ovis для унификации, хранения и анализа суммарных статистик результатов ПГИА. В первую очередь в качестве решения существующих задач для хранения и анализа крупных ПГИА мы создали платформу GWAS-MAP|ovis для хранения, унификации, визуализации и анализа ПГИА овец. Основой платформы стала ранее созданная платформа GWAS-MAP, которая была создана для ПГИА, проведенных в рамках генетики человека. Данная платформа обладает рядом преимуществ. Во-первых, при загрузке и унификации результатов ПГИА происходит контроль качества, для которого можно задать различные параметры, такие как минорная частота аллеля, размер эффекта и др. После данного контроля качества становится возможным анализировать ПГИА без дополнительных фильтров. Во-вторых, с помощью данной платформы можно осуществлять анализ колокализации для сравнения эффектов одного и того же локуса на разных признаках. В-третьих, платформа обладает собственным веб-интерфейсом, с помощью которого возможен быстрый поиск необходимых ОНП, а также построение «Manhattan Plot» для визуализации результатов ПГИА.

Непосредственно создание платформы — это необходимая техническая задача. Для изучения генетического контроля мясной продуктивности овец с использованием современных методов количественной генетики необходимо было наполнить платформу и провести исследования на существующих данных.

Для наполнения базы данных был произведен комплексный поиск работ, связанных с ПГИА по признакам мясной продуктивности у овец. Результатом

поиска стало создание базы данных QTL признаков мясной продуктивности у овец. Созданная база данных может быть использована, как учеными для поиска маркеров для изучения или подтверждения найденных результатов, так и селекционерами для выбора потенциальных маркеров для проведения МОС. Однако, из 18 статей, информация из которых легла в создание базы данных QTL, только в одной статье был предоставлен доступ к полным суммарным статистикам ПГИА. Эти суммарные статистики были использованы для наполнения созданной платформы и проведения сопутствующих анализов для апробации платформы GWAS-MAP|ovis.

Работа S. Volormaa с соавторами 2016 года [16] является крупнейшей на данный момент в области исследования генетики овец по размеру выборки и количеству найденных ассоциаций. Несмотря на это, авторы использовали не самый оптимальный подход к осуществлению многомерного анализа, объединив все 56 признаков в 1 многомерный. В свою очередь, в рамках изучения генетического контроля мясной продуктивности овец мы провели многомерный анализ на суммарных статистиках из работы S. Volormaa с соавторами.

В результате проведенного нами ПГИА на 3 многомерных признаках было найдено 12 значимо ассоциированных локусов, 4 из которых были найдены и в оригинальной работе S. Volormaa с соавторами. Для 8 новых локусов мы провели приоритизацию генов и воспользовались одним из современных подходов для построения функциональной сети взаимодействия локусов. В результате была построена сеть из 6 локусов (2 найденных в рамках данной диссертационной работы), которая подтверждает ранее описанные в литературе локусы, ответственные за мясную продуктивность у овец, а также позволяет выдвинуть гипотезу о возможной связи данных локусов через регуляцию механизмов зависимой регуляции экспрессии.

Стоит отметить, что мы обнаружили большие блоки неравновесия по сцеплению в 3 из 4 известных ранее локусах (между 22 Мб и 42 Мб на хромосоме 6, между 24.5 Мб и 28 Мб на хромосоме 11, и между 62 Мб и 70 Мб на хромосоме 18). Эти локусы имеют сильную ассоциацию с признаками мясной продуктивности и формированием каркаса (доля объясненной дисперсии для самого ассоциированного признака из оригинальных данных 3.42%, 1.42%, и 0.48% для локусов на хромосоме 6, 11 и 18 соответственно). Таким образом, можно предположить, что данные локусы имеют один или несколько гаплотипов с особенно сильным эффектом на мясную продуктивность, и они должны, в первую очередь, рассматриваться исследователями и селекционерами как потенциальные кандидаты для проведения МОС.

Найденные локусы мы добавили в ранее созданную нами базу данных QTL, ассоциированных с признаками мясной продуктивности, формирования каркаса и роста у овец, и провели поиск всех имеющихся в базе локусов на российской популяции овец с использованием платформы GWAS-MAP|ovis. В результате было подтверждено 2 локуса: один известный rs401834107 (ген *LCORL*) и один, впервые обнаруженный в данной диссертационной работе rs401990068 (*FAM3C/WNT16*). Ген *LCORL* входит в созданную нами сеть функционального взаимодействия, а также кодирует транскрипционный фактор, что позволяет выдвинуть гипотезу о том, что данный фактор напрямую воздействует на 3 других локуса в сети, гены в которых, вероятнее всего, ответственны за рецепторы, ферменты и ионные каналы. Стоит отметить, что и локус *FAM3C/WNT16* входит в данную сеть функционального взаимодействия, наряду с описанными ранее локусами на 6 и 11 хромосоме, а также локусом *GHR*. Таким образом, с помощью подтверждения генетической ассоциаций на достаточно небольшой выборке российских животных (N=108) мы смогли обнаружить 2 потенциальных маркера для проведения МОС на российской выборке овец. Исходя из ранее построенной сети

функционального взаимодействия генов, лучшим решением будет использовать в дополнении к 2 подтвержденным локусам еще 2, связанных с ними. Использование более крупных выборок для подтверждения генетических ассоциаций позволит увеличить мощность анализа и подтвердить дополнительные локусы, что благоприятным образом скажется на результатах МОС.

Одним из преимуществ созданной нами платформы является наличие веб-интерфейса доступного по адресу <https://pheligeovis.icgbio.ru/>. Данный веб-интерфейс позволяет проводить поиск ОНП, загруженных в платформу, визуализировать любой участок генома с помощью построения «Manhattan plot», проводить анализ колокализации для любого ОНП и ассоциированных с ним признаков. Поиск ОНП может быть применен для выбора потенциальных маркеров для проведения МОС. Мы продемонстрировали номинально значимую ассоциацию ПЦ для маркеров, ассоциированных с признаком НСWT (вес туши после забоя). Таким образом, возможно проводить оценку ПЦ для любого признака, загруженного в платформу. При этом исследователь или селекционер также может оценить на исследуемой выборке животных, имеется ли экономическая целесообразность применения выбранных маркеров для конкретных признаков.

Другим инструментом, который помогает выбирать маркеры для проведения МОС, является анализ колокализации. Данный анализ позволяет сравнивать паттерны ассоциации конкретного геномного локуса с разными признаками и делать выводы о наличии или отсутствии плейотропного эффекта данного локуса на те или иные признаки. Это может быть полезно при выборе потенциальных маркеров для проведения МОС. К примеру, при выборе между использованием 2 маркеров предпочтительнее будет выбрать тот, который имеет плейотропный эффект на несколько признаков мясной продуктивности, так как он потенциально принесет больше прибыли. В данной диссертационной работе мы провели первый в мире анализ

колокализации для самого значимо ассоциированного из найденных нами локусов с мясной продуктивностью rs401834107 (ген *LCORL*). Было продемонстрировано, что данный локус имеет плеiotропный эффект на 12 различных признаков, связанных с мясной продуктивностью. Стоит отметить, что направление эффекта данного локуса на признаки было разным. Таким образом, проведение анализа колокализации на потенциальных маркерах для внедрения в МОС позволяет лучше понять эффект выбранных маркеров и отсеять те из них, которые отрицательно влияют на один или несколько признаков, интересующих селекционера.

К сожалению, в отличие от ассоциативных исследований в области генетики человека, аналогичные исследования у животных не всегда придерживаются общепринятых стандартов. Например, при поиске работ для наполнения созданной платформы, только в нескольких работах были указаны эффекты, референсные и эффекторные аллели, тогда как эта информация является ключевой при выборе маркеров для МОС. В свою очередь, в области исследований генетики человека для каждого ОНП всегда публикуется полная информация (идентификатор ОНП, хромосома, позиция, значение эффекта, референсный и эффекторный аллель, p-value). Отсутствие такой же полной информации в исследованиях, связанных с ПГИА животных, связано, в первую очередь, с коммерческой тайной. Другой причиной является отсутствие удобного инструментария для анализа и хранения результатов ПГИА. Платформа GWAS-MAP_{ovis} создана на основе платформы GWAS-MAP и призвана помочь исследователям в хранении, унификации и анализе результатов ПГИА различных экономически важных признаков овец. Применение общепринятых стандартов из ПГИА человека в рамках ПГИА овец потенциально способно вывести овцеводство на новый уровень. Это поможет не только исследователям и селекционерам, но и всему человечеству в рамках решения проблем продовольственной безопасности. Мы призываем всех исследователей использовать общепринятые стандарты для проведения

ПГИА и выкладывать в открытый доступ результаты в виде суммарных статистик.

Созданная в рамках данной диссертационной работы платформа не имеет аналогов в рамках исследований в области генетики животных. Стоит отметить, что при появлении в открытом доступе суммарных статистик ПГИА, связанных с признаками мясной продуктивности у овец, мы планируем своевременное пополнение созданной нами платформы новыми данными. То же касается и созданной нами базы данных генов и QTL, которая также нуждается в своевременном пополнении новыми данными для актуализации использования содержащейся в базе данных информации для нужд фундаментальной науки и селекции.

Найденные в рамках изучения генетического контроля мясной продуктивности овец новые локусы, подтвержденные локусы, а также сеть функционального взаимодействия генов могут быть использованы для создания моделей МОС на любых породах овец и могут быть полезны для дальнейших исследований в области генетики овец.

5. Заключение

Данная работа посвящена изучению генетического контроля признаков мясной продуктивности овец с использованием современных методов количественной генетики. В качестве необходимой технической задачи была создана платформа GWAS-MAP|ovis для хранения, анализа и визуализации полногеномных исследований ассоциаций овец. Веб-интерфейс платформы находится в открытом доступе по адресу <https://pheligeovis.icgbio.ru/>. Созданная платформа была зарегистрирована в государственном реестре российских программ для электронных вычислительных машин и баз данных. Для наполнения и апробации созданной платформы был произведен комплексный поиск работ, связанных с ПГИА признаков мясной продуктивности, формирования каркаса и ростом у овец. В результате была создана база данных генов и QTL, ассоциированных с признаками мясной продуктивности у овец, а также найдены суммарные статистики для наполнения платформы GWAS-MAP|ovis.

С использованием загруженных в платформу данных был проведен многомерный анализ для поиска новых локусов, ассоциированных с мясной продуктивностью у овец. Было найдено 8 новых локусов, а также приоритизировано 13 генов. Была построена сеть функционального взаимодействия, состоящая из 6 локусов (2 новых). Данная сеть позволяет выдвинуть гипотезу о возможной связи данных локусов через механизмы зависимой регуляции экспрессии. Был проведен анализ колокализации для самого значимо ассоциированного ОНП rs401834107 (ген *LCORL*). Данный анализ является первым в мире применением метода колокализации на овцах. В дополнение была продемонстрирована возможность применения платформы GWAS-MAP|ovis для поиска и подбора потенциальных маркеров для проведения МОС и создания различных тест-систем.

Найденные в рамках данной работы локусы, сеть функционального взаимодействия генов, а также созданная платформа GWAS-MAP|ovis могут быть полезны исследователям для проведения фундаментальных исследований и селекционерам для подбора маркеров для проведения МОС.

6. Выводы

На основе проделанной работы можно сделать следующие выводы:

1) Создана база данных, содержащая информацию о 187 кандидатных локусах, ассоциированных с признаками, относящимися к мясной продуктивности различных мировых пород овец

2) Создана платформа GWAS-MAP|ovis для агрегации, унификации, визуализации результатов полногеномных исследований ассоциаций овец и для их дополнительного анализа с использованием современных методов количественной генетики.

3) Впервые была показана ассоциация восьми локусов на хромосомах 1 (rs193632759), 2 (rs428034699), 3 (rs428034699, rs403766990, rs399851221) 4 (rs401990068), 5 (rs408893215) и 23 (rs418394153) с признаками мясной продуктивности овец. Два из восьми локусов входят в сеть функционального взаимодействия генов, ассоциированных с мясной продуктивностью у овец (rs428034699 и rs401990068).

4) В результате первого в мире применения анализа колокализации у овец показан плейотропный эффект локуса rs401834107 (ген LCORL) на признаки мясной продуктивности овец.

5) Ассоциация двух ОНП rs401834107 (ген LCORL) и rs401990068 (FAM3C/WNT16) с признаками мясной продуктивности была верифицирована на российской выборке овец. Эти ОНП могут использоваться для создания тест-систем для маркер-ориентированной селекции.

6) Показана возможность создания модели оценки племенной ценности признаков мясной продуктивности. Для признака убойная масса туши была показана номинальная значимость предсказания на российской выборке.

Данный подход к созданию моделей может быть использован для разработки тест-систем для оценки племенной ценности животного.

7. Библиографический список

1. Jiang J. A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle. / Jiang J., Ma L., Prakapenka D., VanRaden P.M., Cole J.B., Da Y. // *Frontiers in genetics* – 2019. – Т. 10 – С.412.
2. Magalhães A.F.B. Genomic selection for meat quality traits in Nelore cattle / Magalhães A.F.B., Schenkel F.S., Garcia D.A., Gordo D.G.M., Tonussi R.L., Espigolan R., Silva R.M. de O., Braz C.U., Fernandes Júnior G.A., Baldi F., Carvalheiro R., Boligon A.A., Oliveira H.N. de, Chardulo L.A.L., Albuquerque L.G. de // *Meat Science* – 2019. – Т. 148 – С.32–37.
3. Hayes B.J. Invited review: Genomic selection in dairy cattle: Progress and challenges / Hayes B.J., Bowman P.J., Chamberlain A.J., Goddard M.E. // *Journal of Dairy Science* – 2009. – Т. 92 – № 2 – С.433–443.
4. Soller M. Marker Assisted Selection - An Overview / Soller M. // *Animal Biotechnology* – 1994. – Т. 5 – № 2 – С.193–207.
5. Boichard D. Genomic selection in domestic animals: Principles, applications and perspectives / Boichard D., Ducrocq V., Croiseau P., Fritz S. // *Comptes Rendus - Biologies* – 2016. – Т. 339 – № 7–8 – С.274–277.
6. Hu Z.-L. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era / Hu Z.-L., Park C.A., Wu X.-L., Reecy J.M. // *Nucleic Acids Research* – 2013. – Т. 41 – № D1 – С.D871–D879.
7. Mrode R. Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: Current status and future prospects / Mrode R., Ojango J.M.K., Okeyo A.M., Mwacharo J.M. // *Frontiers in Genetics* – 2019. – Т. 10 – № JAN – С.694.
8. Wang Z.H. iSheep: an Integrated Resource for Sheep Genome, Variant and

Phenotype / Wang Z.H., Zhu Q.H., Li X., Zhu J.W., Tian D.M., Zhang S.S., Kang H.L., Li C.P., Dong L.L., Zhao W.M., Li M.H. // *Frontiers in genetics* – 2021. – T. 12.

9. Li W. Identification of QTL regions and candidate genes for growth and feed efficiency in broilers / Li W., Zheng M., Zhao G., Wang J., Liu J., Wang S., Feng F., Liu D., Zhu D., Li Q., Guo L., Guo Y., Liu R., Wen J. // *Genetics Selection Evolution* – 2021. – T. 53 – № 1 – C.1–17.

10. Belonogova N.M. Region-based association analysis of human quantitative traits in related individuals / Belonogova N.M., Svishcheva G.R., Duijn C.M. van, Aulchenko Y.S., Axenovich T.I. // *PloS one* – 2013. – T. 8 – № 6.

11. Shashkova T.I. The GWAS-MAP platform for aggregation of results of genome-wide association studies and the GWAS-MAP|homo database of 70 billion genetic associations of human traits / Shashkova T.I., Gorev D.D., Pakhomov E.D., Shadrina A.S., Sharapov S.Z., Tsepilov Y.A., Karssen L.C., Aulchenko Y.S. // *Vavilovskii Zhurnal Genetiki i Seleksii* – 2021. – T. 24 – № 8 – C.876–884.

12. Stephens M. A Unified Framework for Association Analysis with Multiple Related Phenotypes / Stephens M. // *PLoS ONE* – 2013. – T. 8 – № 7 – C.e65245.

13. Shen X. Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation / Shen X., Klarić L., Sharapov S., Mangino M., Ning Z., Wu D., Trbojević-Akmačić I., Pučić-Baković M., Rudan I., Polašek O., Hayward C., Spector T.D., Wilson J.F., Lauc G., Aulchenko Y.S. // *Nature Communications* – 2017. – T. 8 – № 1 – C.447.

14. Klarić L. Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases / Klarić L., Tsepilov Y.A., Stanton C.M., Hayward C. // *Science Advances* – 2020. – T. 6 – № 8 – C.24.

15. Bolormaa S. A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle / Bolormaa S.,

Pryce J.E., Reverter A., Zhang Y., Barendse W., Kemper K., Tier B., Savin K., Hayes B.J., Goddard M.E. // PLoS Genetics – 2014. – Т. 10 – № 3 – С.e1004198.

16. Bolormaa S. Detailed phenotyping identifies genes with pleiotropic effects on body composition / Bolormaa S., Hayes B.J., Werf J.H.J. van der, Pethick D., Goddard M.E., Daetwyler H.D. // BMC Genomics – 2016. – Т. 17 – № 1.

17. Wykes D.L. Robert Bakewell (1725-1795) of Dishley: Farmer and Livestock Improver / Wykes D.L. // The Agricultural History Review – 2004. – Т. 52 – № 1 – С.38–55.

18. Grafen A. Fisher the evolutionary biologist / Grafen A. // Journal of the Royal Statistical Society: Series D (The Statistician) – 2003. – Т. 52 – № 3 – С.319–329.

19. Вавилов Н.И. Закон гомологических рядов в наследственной изменчивости: доклад на 3-ем Всероссийском селекционном съезде в г. Саратове 4 июня 1920 г. / Вавилов Н.И. – 1920.

20. Trojak-Goluch A. Polyploidy in Industrial Crops: Applications and Perspectives in Plant Breeding / Trojak-Goluch A., Kawka-Lipińska M., Wielgusz K., Praczyk M. // Agronomy 2021, Vol. 11, Page 2574 – 2021. – Т. 11 – № 12 – С.2574.

21. Gabryś J. Interspecific hybrids of animals in nature, breeding and science - A review / Gabryś J., Kij B., Kochan J., Bugno-Poniewierska M. // Annals of Animal Science – 2021. – Т. 21 – № 2 – С.403–415.

22. Reshma R.S. Molecular markers and its application in animal breeding / Reshma R.S., Das D.N. // Advances in Animal Genomics – 2021. – С.123–140.

23. Хлесткина Е.К. Молекулярные маркеры в генетических исследованиях и в селекции / Хлесткина Е.К. // Вавиловский журнал генетики и селекции – 2013. – Т. 17 – № 4/2 – С.1044–1054.

24. Botstein D. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. / Botstein D., White R.L., Skolnick M., Davis R.W. // *American journal of human genetics* – 1980. – T. 32 – № 3 – C.314–31.

25. Tanksley S.D. RFLP Mapping in Plant Breeding: New Tools for an Old Science / Tanksley S.D., Young N.D., Paterson A.H., Bonierbale M.W. // *Nature Biotechnology* – 1989. – T. 7 – № 3 – C.257–264.

26. Moore G. Cereal genome evolution. Grasses, line up and form a circle. / Moore G., Devos K.M., Wang Z., Gale M.D. // *Current biology : CB* – 1995. – T. 5 – № 7 – C.737–9.

27. Haberfeld A. Application of Multilocus Molecular Markers in Cattle Breeding. 1. Minisatellites and Microsatellites / Haberfeld A., Kalay D., Weisberger P., Gal O., Hilliel J. // *Journal of Dairy Science* – 1993. – T. 76 – № 2 – C.645–652.

28. Dallas J.F. Detection of DNA “fingerprints” of cultivated rice by hybridization with a human minisatellite DNA probe. / Dallas J.F. // *Proceedings of the National Academy of Sciences* – 1988. – T. 85 – № 18 – C.6831–6835.

29. Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. / Tautz D. // *Nucleic acids research* – 1989. – T. 17 – № 16 – C.6463–71.

30. Ragoussis J. Genotyping Technologies for Genetic Research / Ragoussis J. // <http://dx.doi.org/10.1146/annurev-genom-082908-150116> – 2009. – T. 10 – C.117–133.

31. Moose S.P. Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement / Moose S.P., Mumm R.H. // *Plant Physiology* – 2008. – T. 147 – № 3 – C.969–977.

32. Rabier C.-E. On the Accuracy of Genomic Selection / Rabier C.-E., Barre P., Asp T., Charmet G., Mangin B. // *PLOS ONE* – 2016. – T. 11 – № 6 –

C.e0156086.

33. Schulz-Streeck T. Genomic Selection using Multiple Populations / Schulz-Streeck T., Ogutu J.O., Karaman Z., Knaak C., Piepho H.P. // *Crop Science* – 2012. – T. 52 – № 6 – C.2453.

34. los Campos G. de Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding / los Campos G. de, Hickey J.M., Pong-Wong R., Daetwyler H.D., Calus M.P.L. // *Genetics* – 2013. – T. 193 – № 2 – C.327–345.

35. Jannink J.-L. Genomic selection in plant breeding: from theory to practice / Jannink J.-L., Lorenz A.J., Iwata H. // *Briefings in Functional Genomics* – 2010. – T. 9 – № 2 – C.166–177.

36. Robinson G.K. That BLUP is a good thing: The estimation of random effects / Robinson G.K. // *Statistical Science* – 1991. – T. 6 – № 1 – C.15–32.

37. Crossa J. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives / Crossa J., Pérez-Rodríguez P., Cuevas J., Montesinos-López O., Jarquín D., los Campos G. de, Burgueño J., González-Camacho J.M., Pérez-Elizalde S., Beyene Y., Dreisigacker S., Singh R., Zhang X., Gowda M., Roorkiwal M., Rutkoski J., Varshney R.K. // *Trends in Plant Science* – 2017. – T. 22 – № 11 – C.961–975.

38. Newell M.A. Genomic selection in plant breeding / Newell M.A., Jannink J.L. // *Methods in molecular biology* (Clifton, N.J.) – 2014. – T. 1145 – C.117–130.

39. Pszczola M. Reliability of direct genomic values for animals with different relationships within and to the reference population / Pszczola M., Strabel T., Mulder H.A., Calus M.P.L. // *Journal of Dairy Science* – 2012. – T. 95 – № 1 – C.389–400.

40. Daetwyler H.D. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods / Daetwyler H.D., Pong-Wong R., Villanueva B., Woolliams J.A. // *Genetics* – 2010. – T. 185 – № 3 – C.1021–1031.

41. Arruda M.P. Genomic Selection for Predicting Head Blight Resistance in a Wheat Breeding Program / Arruda M.P., Brown P.J., Lipka A.E., Krill A.M., Thurber C., Kolb F.L. // *The Plant Genome* – 2015. – Т. 8 – № 3 – С.0.

42. Lorenz A.J. Potential and Optimization of Genomic Selection for Fusarium Head Blight Resistance in Six-Row Barley / Lorenz A.J., Smith K.P., Jannink J.-L. // *Crop Science* – 2012. – Т. 52 – № 4 – С.1609.

43. Cuevas J. Bayesian genomic-enabled prediction as an inverse problem. / Cuevas J., Pérez-Elizalde S., Soberanis V., Pérez-Rodríguez P., Gianola D., Crossa J. // *G3 (Bethesda, Md.)* – 2014. – Т. 4 – № 10 – С.1991–2001.

44. Colombani C. A comparison of partial least squares (PLS) and sparse PLS regressions in genomic selection in French dairy cattle / Colombani C., Croiseau P., Fritz S., Guillaume F., Legarra A., Ducrocq V., Robert-Granié C. // *Journal of Dairy Science* – 2012. – Т. 95 – № 4 – С.2120–2131.

45. Henderson C.R. Best Linear Unbiased Estimation and Prediction under a Selection Model / Henderson C.R. // *Biometrics* – 1975. – Т. 31 – № 2 – С.423.

46. Kuzmin V.N. Sheep Breeding: State and Development Prospects / Kuzmin V.N., Marinchenko T.E., Korolkova A.P. // *Machinery and Equipment for Rural Area* – 2019. – № 12 – С.2–8.

47. А.И. Ерохин, Е.А. Карасев А.С.Е. Состояние и динамика производства мяса в мире и России / А.И. Ерохин, Е.А. Карасев А.С.Е. // *Овцы, козы, шерстяное дело* – 2014. – Т. №2 – С.37–40.

48. Gilbert M. Global sheep distribution in 2010 (5 minutes of arc) // – 2018.

49. Федеральная служба государственной статистикиТ. 1: Основные итоги Всероссийской сельскохозяйственной переписи 2016 года: кн. 1.: Основные итоги Всероссийской сельскохозяйственной переписи 2016 года по Российской Федерации. / Федеральная служба государственной статистики –

ИИЦ «Статистика России», 2018.– 458с.

50. Ерохин А. И. Разведение овец и коз / Ерохин А. И. – Эксмо-Пресс, Лик пресс, 2001.– 304с.

51. Овцеводство и козоводство России [Электронный ресурс]. URL: <https://ab-centre.ru/page/ovcevodstvo-i-kozovodstvo-rossii> (accessed: 15.09.2019).

52. Lescheva M. Current state and perspectives of sheep breeding development in Russian modern economic conditions / Lescheva M., Ivolga A. // *Ekonomika poljoprivrede* – 2015. – Т. 62 – № 2 – С.467–480.

53. Эрнст Л.К. Генетические ресурсы сельскохозяйственных животных в России и сопредельных странах / Л. К. Эрнст, Н. Г. Дмитриев, И. А. Паронян – СПб: ВНИИГРЖ, 1994.

54. Werf J.H.J. van der Marker-assisted selection in sheep and goats Rome: Food and Agriculture Organization of the United Nations (FAO), 2007. – 229–247с.

55. Huby M. Genetic variability of six French meat sheep breeds in relation to their genetic management. / Huby M., Griffon L., Moureaux S., Rochambeau H. De, Danchin-Burge C., Verrier E. // *Genetics, selection, evolution : GSE* – 2003. – Т. 35 – № 6 – С.637–55.

56. Kontogeorgos G. Hypophysis Academic Press, 2012. – 584–593с.

57. Zhao H. A novel 29bp insertion/deletion (indel) variant of the LHX3 gene and its influence on growth traits in four sheep breeds of various fecundity / Zhao H., He S., Zhu Y., Cao X., Luo R., Cai Y., Xu H., Sun X. // *Archives Animal Breeding* – 2017. – Т. 60 – № 2 – С.79–85.

58. Arora R. Identification of novel single nucleotide polymorphisms in candidate genes for mutton quality in Indian sheep / Arora R., Yadav H., Yadav D. // *Animal Molecular Breeding* – 2014.

59. Mahrous K.F. Association between single nucleotide polymorphism in ovine Calpain gene and growth performance in three Egyptian sheep breeds / Mahrous K.F., Hassanane M.S., Shafey H.I., Abdel Mordy M., Rushdi H.E. // Journal of Genetic Engineering and Biotechnology – 2016. – T. 14 – № 2 – C.233–240.

60. Huff-Lonergan E. Proteolysis of specific muscle structural proteins by mu-calpain at low pH and temperature is similar to degradation in postmortem bovine muscle. / Huff-Lonergan E., Mitsuhashi T., Beekman D.D., Parrish F.C., Olson D.G., Robson R.M. // Journal of Animal Science – 1996. – T. 74 – № 5 – C.993.

61. Sahu A.R. Polymorphism in exon 3 of myostatin (MSTN) gene and its association with growth traits in Indian sheep breeds / Sahu A.R., Jeichitra V., Rajendran R., Raja A. // Small Ruminant Research – 2017. – T. 149 – C.81–84.

62. McMahon C.D. Neuroregulation of growth hormone secretion in domestic animals / McMahon C.D., Radcliff R.P., Lookingland K.J., Tucker H.A. // Domestic Animal Endocrinology – 2001. – T. 20 – № 2 – C.65–87.

63. Zhang L. Identification of MEF2B and TRHDE Gene Polymorphisms Related to Growth Traits in a New Ujumqin Sheep Population / Zhang L., Ma X., Xuan J., Wang H., Yuan Z., Wu M., Liu R., Zhu C., Wei C., Zhao F., Du L. // PLOS ONE – 2016. – T. 11 – № 7 – C.e0159504.

64. Chen L. The molecular characterization and temporal–spatial expression of myocyte enhancer factor 2 genes in the goat and their association with myofiber traits / Chen L., Cheng B., Li L., Zhan S., Wang L., Zhong T., Chen Y., Zhang H. // Gene – 2015. – T. 555 – № 2 – C.223–230.

65. Ovis_aries_1.0 - oviAri1 - Genome - Assembly - NCBI [Электронный ресурс]. URL: https://www.ncbi.nlm.nih.gov/assembly/GCA_000005525.1/ (accessed: 14.07.2022).

66. Jiang Y. The sheep genome illuminates biology of the rumen and lipid

metabolism / Jiang Y., Xie M., Chen W., Dalrymple B.P. // *Science* – 2014. – Т. 344 – № 6188 – С.1168–1173.

67. English A.C. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology / English A.C., Richards S., Han Y., Wang M., Vee V., Qu J., Qin X., Muzny D.M., Reid J.G., Worley K.C., Gibbs R.A. // *PLOS ONE* – 2012. – Т. 7 – № 11 – С.e47768.

68. Oar_v4.0 - oviAri4 - Genome - Assembly - NCBI [Электронный ресурс]. URL: https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2/ (accessed: 14.07.2022).

69. International Sheep Genomics Consortium [Электронный ресурс]. URL: <https://www.sheepmap.org/> (accessed: 14.07.2022).

70. Ott J. Analysis of human genetic linkage / J. Ott – Johns Hopkins University Press, 1999. – 382с.

71. Klein R.J. Complement Factor H Polymorphism in Age-Related Macular Degeneration / Klein R.J., Zeiss C., Chew E.Y., Tsai J.-Y., Sackler R.S., Haynes C., Henning A.K., SanGiovanni J.P., Mane S.M., Mayne S.T., Bracken M.B., Ferris F.L., Ott J., Barnstable C., Hoh J. // *Science* – 2005. – Т. 308 – № 5720 – С.385–389.

72. Watanabe K. A global overview of pleiotropy and genetic architecture in complex traits / Watanabe K., Stringer S., Frei O., Umićević Mirkov M., Leeuw C. de, Polderman T.J.C., Sluis S. van der, Andreassen O.A., Neale B.M., Posthuma D. // *Nature genetics* – 2019. – Т. 51 – № 9 – С.1339–1348.

73. Jansen P.R. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways / Jansen P.R., Watanabe K., Stringer S., Skene N., Bryois J., Hammerschlag A.R., Leeuw C.A. de, Benjamins J.S., Muñoz-Manchado A.B., Nagel M., Savage J.E., Tiemeier H., White T., Agee M., Alipanahi B., Auton A., Bell R.K., Bryc K., Elson S.L., Fontanillas P., Furlotte N.A.,

Hinds D.A., Huber K.E., Kleinman A., Litterman N.K., McCreight J.C., McIntyre M.H., Mountain J.L., Noblin E.S., Northover C.A.M., Pitts S.J., Sathirapongsasuti J.F., Sazonova O. V., Shelton J.F., Shringarpure S., Tian C., Wilson C.H., Tung J.Y., Hinds D.A., Vacic V., Wang X., Sullivan P.F., Sluis S. van der, Polderman T.J.C., Smit A.B., Hjerling-Leffler J., Someren E.J.W. Van, Posthuma D. // *Nature genetics* – 2019. – T. 51 – № 3 – C.394–403.

74. Uffelmann E. Emerging Methods and Resources for Biological Interrogation of Neuropsychiatric Polygenic Signal / Uffelmann E., Posthuma D. // *Biological psychiatry* – 2021. – T. 89 – № 1 – C.41–53.

75. Uffelmann E. Genome-wide association studies / Uffelmann E., Huang Q.Q., Munung N.S., Vries J. de, Okada Y., Martin A.R., Martin H.C., Lappalainen T., Posthuma D. // *Nature Reviews Methods Primers* 2021 1:1 – 2021. – T. 1 – № 1 – C.1–21.

76. Pirinen M. Including known covariates can reduce power to detect genetic effects in case-control studies / Pirinen M., Donnelly P., Spencer C.C.A. // *Nature Genetics* 2012 44:8 – 2012. – T. 44 – № 8 – C.848–851.

77. Li G. Genetic Studies: The Linear Mixed Models in Genome-wide Association Studies / Li G., Zhu H. // *The Open Bioinformatics Journal* – 2014. – T. 7 – № 1 – C.27–33.

78. Amos C.I. A more powerful robust sib-pair test of linkage for quantitative traits / Amos C.I., Elston R.C., Wilson A.F., Bailey-Wilson J.E. // *Genetic epidemiology* – 1989. – T. 6 – № 3 – C.435–449.

79. Zhou X. Genome-wide efficient mixed-model analysis for association studies / Zhou X., Stephens M. // *Nature Genetics* – 2012. – T. 44 – № 7 – C.821–824.

80. Kang H.M. Variance component model to account for sample structure in genome-wide association studies / Kang H.M., Sul J.H., Service S.K., Zaitlen N.A.,

Kong S., Freimer N.B., Sabatti C., Eskin E. // Nature Genetics – 2010. – T. 42 – № 4 – C.348–354.

81. Lippert C. FaST linear mixed models for genome-wide association studies / Lippert C., Listgarten J., Liu Y., Kadie C.M., Davidson R.I., Heckerman D. // Nature Methods – 2011. – T. 8 – № 10 – C.833–835.

82. Tabachnick B.G. Multivariate Analysis of Variance (MANOVA) / Tabachnick B.G., Fidell L.S. // International Encyclopedia of Statistical Science – 2011. – C.902–904.

83. Ried J.S. A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape / Ried J.S., Jeff M. J., Chu A.Y., Loos R.J.F. // Nature Communications – 2016. – T. 7 – C.13357.

84. Egger M. Meta-analysis: Potentials and promise / Egger M., Smith G.D. // BMJ – 1997. – T. 315 – № 7119 – C.1371–1374.

85. Willer C.J. METAL: fast and efficient meta-analysis of genomewide association scans / Willer C.J., Li Y., Abecasis G.R. // Bioinformatics – 2010. – T. 26 – № 17 – C.2190–2191.

86. Baselmans B.M.L. Multivariate genome-wide analyses of the well-being spectrum / Baselmans B.M.L., Jansen R., Ip H.F., Dongen J. van, Abdellaoui A., Weijer M.P. van de, Bao Y., Smart M., Kumari M., Willemsen G., Hottenga J.J., Boomsma D.I., Geus E.J.C. de, Nivard M.G., Bartels M. // Nature genetics – 2019. – T. 51 – № 3 – C.445–451.

87. Winkler T.W. Quality control and conduct of genome-wide association meta-analyses / Winkler T.W., Day F.R., Croteau-Chonka D.C., Wood A.R., Locke A.E., Mägi R., Ferreira T., Fall T., Graff M., Justice A.E., Luan J., Gustafsson S., Randall J.C., Vedantam S., Workalemahu T., Kilpeläinen T.O., Scherag A., Esko T., Kutalik Z., Heid I.M., Loos R.J.F., Genetic Investigation of Anthropometric Traits (GIANT) Consortium // Nature Protocols – 2014. – T. 9 – № 5 – C.1192–

1212.

88. Gibson G. Hints of hidden heritability in GWAS / Gibson G. // *Nature Genetics* – 2010. – T. 42 – № 7 – C.558–560.

89. Hommel G. A Stagewise Rejective Multiple Test Procedure Based on a Modified Bonferroni Test / Hommel G. // *Biometrika* – 1988. – T. 75 – № 2 – C.383.

90. Murray M.H. False Discovery Rate Computation: Illustrations and Modifications / Murray M.H., Blume J.D.

91. Leeuw C.A. de MAGMA: Generalized Gene-Set Analysis of GWAS Data / Leeuw C.A. de, Mooij J.M., Heskes T., Posthuma D. // *PLOS Computational Biology* – 2015. – T. 11 – № 4 – C.e1004219.

92. Pers T.H. Biological interpretation of genome-wide association studies using predicted gene functions / Pers T.H., Karjalainen J.M., Chan Y., Westra H.-J., Wood A.R., Yang J., Lui J.C., Vedantam S., Gustafsson S., Esko T., Frayling T., Speliotes E.K., Boehnke M., Raychaudhuri S., Fehrmann R.S.N., Hirschhorn J.N., Franke L., Franke L. // *Nature Communications* – 2015. – T. 6 – № 1 – C.5890.

93. Zhang L. Genome-wide association studies for growth and meat production traits in sheep. / Zhang L., Liu J., Zhao F., Ren H., Xu L., Lu J., Zhang S., Zhang X., Wei C., Lu G., Zheng Y., Du L. // *PloS one* – 2013. – T. 8 – № 6 – C.e66569.

94. Riggio V. Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs / Riggio V., Matika O., Pong-Wong R., Stear M.J., Bishop S.C. // *Heredity* – 2013. – T. 110 – № 5 – C.420–429.

95. Al-Mamun H.A. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight / Al-Mamun H.A., Kwan P.,

Clark S.A., Ferdosi M.H., Tellam R., Gondro C. // *Genetics Selection Evolution* – 2015. – T. 47 – № 1 – C.66.

96. Matika O. Genome-wide association reveals QTL for growth, bone and in vivo carcass traits as assessed by computed tomography in Scottish Blackface lambs / Matika O., Riggio V., Anselme-Moizan M., Law A.S., Pong-Wong R., Archibald A.L., Bishop S.C. // *Genetics Selection Evolution* – 2016. – T. 48 – № 1 – C.11.

97. Kominakis A. Combined GWAS and ‘guilt by association’-based prioritization analysis identifies functional candidate genes for body size in sheep / Kominakis A., Hager-Theodorides A.L., Zoidis E., Saridaki A., Antonakos G., Tsiamis G. // *Genetics Selection Evolution* – 2017. – T. 49 – № 1 – C.41.

98. Garza Hernandez D. Analysis of single nucleotide polymorphisms variation associated with important economic and computed tomography measured traits in Texel sheep / Garza Hernandez D., Mucha S., Banos G., Kaseja K., Moore K., Lambe N., Yates J., Bunger L. // *animal* – 2018. – T. 12 – № 05 – C.915–922.

99. Pasandideh M. A genome scan for quantitative trait loci affecting average daily gain and Kleiber ratio in Baluchi Sheep / Pasandideh M., Rahimi-Mianji G., Gholizadeh M. // *Journal of Genetics* – 2018. – T. 97 – № 2 – C.493–503.

100. Ghasemi M. Genome-wide association study of birth weight in sheep / Ghasemi M., Zamani P., Vatankhah M., Abdoli R. // *animal* – 2019. – C.1–7.

101. Beck T. GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies / Beck T., Hastings R.K., Gollapudi S., Free R.C., Brookes A.J. // *European Journal of Human Genetics* 2014 22:7 – 2013. – T. 22 – № 7 – C.949–952.

102. Welter D. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations / Welter D., MacArthur J., Morales J., Burdett T., Hall P., Junkins H., Klemm A., Flicek P., Manolio T., Hindorff L., Parkinson H. // *Nucleic acids research* – 2014. – T. 42 – № Database issue.

103. Canela-Xandri O. An atlas of genetic associations in UK Biobank / Canela-Xandri O., Rawlik K., Tenesa A. // *Nature genetics* – 2018. – Т. 50 – № 11 – С.1593–1599.
104. Elsworth B. The MRC IEU OpenGWAS data infrastructure / Elsworth B., Lyon M., Alexander T., Liu Y., Matthews P., Hallett J., Bates P., Palmer T., Haberland V., Smith G., Zheng J., Haycock P., Gaunt T., Hemani G. // *bioRxiv* – 2020.
105. Staley J.R. PhenoScanner: a database of human genotype–phenotype associations / Staley J.R., Blackshaw J., Kamat M.A., Ellis S., Surendran P., Sun B.B., Paul D.S., Freitag D., Burgess S., Danesh J., Young R., Butterworth A.S. // *Bioinformatics* – 2016. – Т. 32 – № 20 – С.3207–3209.
106. Zhu Z. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets / Zhu Z., Zhang F., Hu H., Bakshi A., Robinson M.R., Powell J.E., Montgomery G.W., Goddard M.E., Wray N.R., Visscher P.M., Yang J. // *Nature Genetics* – 2016. – Т. 48 – № 5 – С.481–487.
107. Shashkova T.I. PheLiGe: An interactive database of billions of human genotype-phenotype associations / Shashkova T.I., Pakhomov E.D., Gorev D.D., Karssen L.C., Joshi P.K., Aulchenko Y.S. // *Nucleic Acids Research* – 2021. – Т. 49 – № D1 – С.D1347–D1350.
108. McLaren W. The Ensembl Variant Effect Predictor / McLaren W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R.S., Thormann A., Flicek P., Cunningham F. // *Genome Biology* – 2016. – Т. 17 – № 1 – С.122.
109. Home - OMIM - NCBI [Электронный ресурс]. URL: <http://www.ncbi.nlm.nih.gov/omim> (accessed: 19.05.2016).
110. Bult C.J. Mouse Genome Database (MGD) 2019 / Bult C.J., Blake J.A., Smith C.L., Kadin J.A., Richardson J.E., Anagnostopoulos A., Asabor R., Baldarelli R.M., Beal J.S., Bello S.M., Blodgett O., Butler N.E., Christie K.R., Corbani L.E.,

Creelman J., Dolan M.E., Drabkin H.J., Giannatto S.L., Hale P., Hill D.P., Law M., Mendoza A., McAndrews M., Miers D., Motenko H., Ni L., Onda H., Perry M., Recla J.M., Richards-Smith B., Sitnikov D., Tomczuk M., Tonorio G., Wilming L., Zhu Y. // *Nucleic Acids Research* – 2019. – T. 47 – № D1 – C.D801–D806.

111. Purcell S. PLINK: A tool set for whole-genome association and population-based linkage analyses / Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., Bakker P.I.W. De, Daly M.J., Sham P.C. // *American Journal of Human Genetics* – 2007. – T. 81 – № 3 – C.559–575.

112. Interquartile Range // *The Concise Encyclopedia of Statistics* – 2008. – C.266–267.

113. Momozawa Y. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes / Momozawa Y., Dmitrieva J., Théâtre E., Zhao Z.Z. // *Nature Communications* – 2018. – T. 9 – № 1 – C.1–18.

114. Wang H. Genome-Wide Specific Selection in Three Domestic Sheep Breeds / Wang H., Zhang L., Cao J., Wu M., Ma X., Liu Z., Liu R., Zhao F., Wei C., Du L. // *PLOS ONE* – 2015. – T. 10 – № 6 – C.e0128688.

115. Bolormaa S. Detailed phenotyping identifies genes with pleiotropic effects on body composition / Bolormaa S., Hayes B.J., Werf J.H.J. van der, Pethick D., Goddard M.E., Daetwyler H.D. // *BMC Genomics* – 2016. – T. 17 – № 1 – C.224.

116. Jalil-Sarghale A. Association of pituitary specific transcription factor-1 (POU1F1) gene polymorphism with growth and biometric traits and blood metabolites in Iranian Zel and Lori-Bakhtiari sheep / Jalil-Sarghale A., Moradi Shahrababak M., Moradi Sharbabak H., Sadeghi M., Mura M.C. // *Molecular Biology Reports* – 2014. – T. 41 – № 9 – C.5787–5792.

117. Cinar M.U. Ovar-DRB1 haplotypes *2001 and *0301 are associated with sheep growth and ewe lifetime prolificacy / Cinar M.U., Mousel M.R., Herrmann-

Hoising L.M., Taylor J.B., White S.N. // *Gene* – 2016. – T. 595 – № 2 – C.187–192.

118. Ma X. Effect of polymorphisms in the CAMKMT gene on growth traits in Ujumqin sheep / Ma X., Guan L., Xuan J., Wang H., Yuan Z., Wu M., Liu R., Zhu C., Wei C., Zhao F., Du L., Zhang L. // *Animal Genetics* – 2016. – T. 47 – № 5 – C.618–622.

119. Trukhachev V. MEF2B gene SNP markers of meat productivity in Severokavkazskaya sheep breed / Trukhachev V., Belyaev V., Kvochko A., Kulichenko A., Kovalev D., Pisarenko S., Volynkina A., Selionova M., Aybazov M., Shumaenko S., Omarov A., Mamontova T., Yatsyk O., Krivoruchko A., Petrovic M., Pantelic V., Petrovic-Caro V. // *Genetika* – 2016. – T. 48 – № 1 – C.97–108.

120. Zhang L. Identification of MEF2B and TRHDE Gene Polymorphisms Related to Growth Traits in a New Ujumqin Sheep Population / Zhang L., Ma X., Xuan J., Wang H., Yuan Z., Wu M., Liu R., Zhu C., Wei C., Zhao F., Du L. // *PLOS ONE* – 2016. – T. 11 – № 7 – C.e0159504.

121. Aali M. Association of the calpastatin genotypes, haplotypes, and SNPs with meat quality and fatty acid composition in two Iranian fat- and thin-tailed sheep breeds / Aali M., Moradi-Shahrbabak H., Moradi-Shahrbabak M., Sadeghi M., Yousefi A.R. // *Small Ruminant Research* – 2017. – T. 149 – C.40–51.

122. Armstrong E. Novel genetic polymorphisms associated with carcass traits in grazing Texel sheep / Armstrong E., Ciappesoni G., Iriarte W., Silva C. Da, Macedo F., Navajas E.A., Brito G., San Julián R., Gimeno D., Postiglioni A. // *Meat Science* – 2018. – T. 145 – C.202–208.

123. Sahu A.R. Advances in genomic strategies to improve growth and meat production traits in sheep: An overview / Sahu A.R., Nayak N., Panigrahi M., Kumar S. // *Indian Journal of Small Ruminants (The)* – 2017. – T. 23 – № 2 – C.139.

124. Xu S.-S. Recent advances in understanding genetic variants associated with economically important traits in sheep (*Ovis aries*) revealed by high-throughput

screening technologies / Xu S.-S., Li M.-H. // *Frontiers of Agricultural Science and Engineering* – 2017. – T. 4 – № 3 – C.279.

125. Pers T.H. Biological interpretation of genome-wide association studies using predicted gene functions / Pers T.H., Karjalainen J.M., Chan Y., Westra H.-J., Wood A.R., Yang J., Lui J.C., Vedantam S., Gustafsson S., Esko T., Frayling T., Speliotes E.K., Boehnke M., Raychaudhuri S., Fehrmann R.S.N., Hirschhorn J.N., Franke L. // *Nature Communications* – 2015. – T. 6 – C.5890.

126. Sirmaci A. MASP1 mutations in patients with facial, umbilical, coccygeal, and auditory findings of carnevale, malpuech, OSA, and michels syndromes / Sirmaci A., Walsh T., Akay H., Spiliopoulos M., Şakalar Y.B., Hasanefendioğlu-Bayrak A., Duman D., Farooq A., King M.C., Tekin M. // *American Journal of Human Genetics* – 2010. – T. 87 – № 5 – C.679–686.

127. Kearney C.J. DOCK8 regulates signal transduction events to control immunity / Kearney C.J., Randall K.L., Oliaro J. // *Cellular and Molecular Immunology* – 2017. – T. 14 – № 5 – C.406–411.

128. Moiseeva E.P. A novel dystrophin/utrophin-associated protein is an enzymatically inactive member of the phosphoglucomutase superfamily / Moiseeva E.P., Belkin A.M., Spurr N.K., Koteliansky V.E., Critchley D.R. // *European Journal of Biochemistry* – 1996. – T. 235 – № 1–2 – C.103–113.

129. Huang S. Sex determination and maintenance: The role of DMRT1 and FOXL2 / Huang S., Ye L., Chen H. // *Asian Journal of Andrology* – 2017. – T. 19 – № 6 – C.619–624.

130. Guo X. Wnt/ β -catenin signaling is sufficient and necessary for synovial joint formation / Guo X., Day T.F., Jiang X., Garrett-Beal L., Topol L., Yang Y. // *Genes and Development* – 2004. – T. 18 – № 19 – C.2404–2417.

131. Gilbertson R.J. PDGFRB is overexpressed in metastatic medulloblastoma // *Nat. Genet.* – 2003. – T. 35. – № 3. – 197–198c.

132. Nazarian R. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation / Nazarian R., Shi H., Wang Q., Kong X., Koya R.C., Lee H., Chen Z., Lee M.K., Attar N., Sazegar H., Chodon T., Nelson S.F., McArthur G., Sosman J.A., Ribas A., Lo R.S. // *Nature* – 2010. – T. 468 – № 7326 – C.973–977.

133. Lee J. Activation of PDGF pathway links LMNA mutation to dilated cardiomyopathy / Lee J., Termglinchan V., Diecke S., Itzhaki I., Lam C.K., Garg P., Lau E., Greenhaw M., Seeger T., Wu H., Zhang J.Z., Chen X., Gil I.P., Ameen M., Sallam K., Rhee J.W., Churko J.M., Chaudhary R., Chour T., Wang P.J., Snyder M.P., Chang H.Y., Karakikes I., Wu J.C. // *Nature* – 2019. – T. 572 – № 7769 – C.335–340.

134. Bonner C.A. Isolation, characterization, and precise physical localization of human CDX1, a caudal-type homeobox gene / Bonner C.A., Loftus S.K., Wasmuth J.J. // *Genomics* – 1995. – T. 28 – № 2 – C.206–211.

135. Sanchez E. POLR1B and neural crest cell anomalies in Treacher Collins syndrome type 4 / Sanchez E., Laplace-Builhé B., Mau-Them F.T., Richard E., Goldenberg A., Toler T.L., Guignard T., Gatinois V., Vincent M., Blanchet C., Boland A., Bihoreau M.T., Deleuze J.F., Olasso R., Nephi W., Lüdecke H.J., Verheij J.B.G.M., Moreau-Lenoir F., Denoyelle F., Rivière J.B., Laplanche J.L., Willing M., Captier G., Apparailly F., Wiczorek D., Collet C., Djouad F., Geneviève D. // *Genetics in Medicine* – 2020. – T. 22 – № 3 – C.547–556.

136. Makishima H. Somatic SETBP1 mutations in myeloid malignancies / Makishima H., Yoshida K., Nguyen N., Przychodzen B., Sanada M., Okuno Y., Ng K.P., Gudmundsson K.O., Vishwakarma B.A., Jerez A., Segui I.G., Takahashi M., Shiraishi Y., Nagata Y., Guinta K., Mori H., Sekeres M.A., Chiba K., Tanaka H., Muramatsu H., Sakaguchi H., Paquette R.L., McDevitt M.A., Kojima S., Sauntharajah Y., Miyano S., Shih L.Y., Du Y., Ogawa S., Maciejewski J.P. // *Nature Genetics* – 2013. – T. 45 – № 8 – C.942–946.

137. Szklarczyk D. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets / Szklarczyk D., Gable A.L., Lyon D., Junge A., Wyder S., Huerta-Cepas J., Simonovic M., Doncheva N.T., Morris J.H., Bork P., Jensen L.J., Mering C. Von // Nucleic acids research – 2019. – T. 47 – № D1 – C.D607–D613.

Приложение

Доп. таблица 1. Основные фенотипические показатели исследуемой выборки российских овец для временного промежутка 6 дней после рождения.

Признак	Минимальное значение	Максимальное значение	Среднее значение	Стандартное отклонение
высота в холке	27.5	43	34.95	3.09
высота до спины	27	44	34.38	3.49
высота в крестце	24	44	34.52	3.50
косая длина туловища	23.5	37	29.20	2.99
длина тела	21.5	38	27.46	3.56
глубина груди	9	16	12.07	1.81
ширина груди	3	8	5.56	1.12
ширина крестца	5	9	6.47	0.94
обхват груди	22	43.5	33.95	3.97
обхват пясти	4	7	5.14	0.65
живая масса	1.42	6.5	3.38	1.09

Доп. таблица 2. Основные фенотипические показатели исследуемой выборки российских овец для временного промежутка 42 дней после рождения.

Признак	Минимальное значение	Максимальное значение	Среднее значение	Стандартное отклонение
высота в холке	33	51	42.64	4.21
высота до спины	33	55	42.57	4.39
высота в крестце	33	52	43.04	4.26
косая длина туловища	32	49	39.91	3.72

длина тела	27	52	37.92	5.17
глубина груди	12	20.5	15.91	1.75
ширина груди	5	17	8.36	2.27
ширина крестца	6.5	14	9.51	1.55
обхват груди	36	64	47.01	5.72
обхват пясти	4.5	8	5.55	0.69
живая масса	3.29	18.5	7.74	2.53

Доп. таблица 3. Основные фенотипические показатели исследуемой выборки российских овец для временного промежутка 3 месяца после рождения.

Признак	Минимальное значение	Максимальное значение	Среднее значение	Стандартное отклонение
высота в холке	38	58	50.61	4.64
высота до спины	37	59	50.79	4.97
высота в крестце	38	62	51.48	5.11
косая длина туловища	38	63	50.34	5.56
длина тела	33	59	46.86	5.33
глубина груди	15	27	20.05	2.69
ширина груди	7	16	10.48	2.56
ширина крестца	7	18	11.91	2.52
обхват груди	43	73	57.79	7.09
обхват пясти	5	9	6.67	1.10
живая масса	9	32	17.87	8.31

Доп. таблица 4. Количество животных для ПГИА из каждого генетического кластера для каждого признака. Серым шрифтом показаны малочисленные группы, не вошедшие в дальнейший анализ.

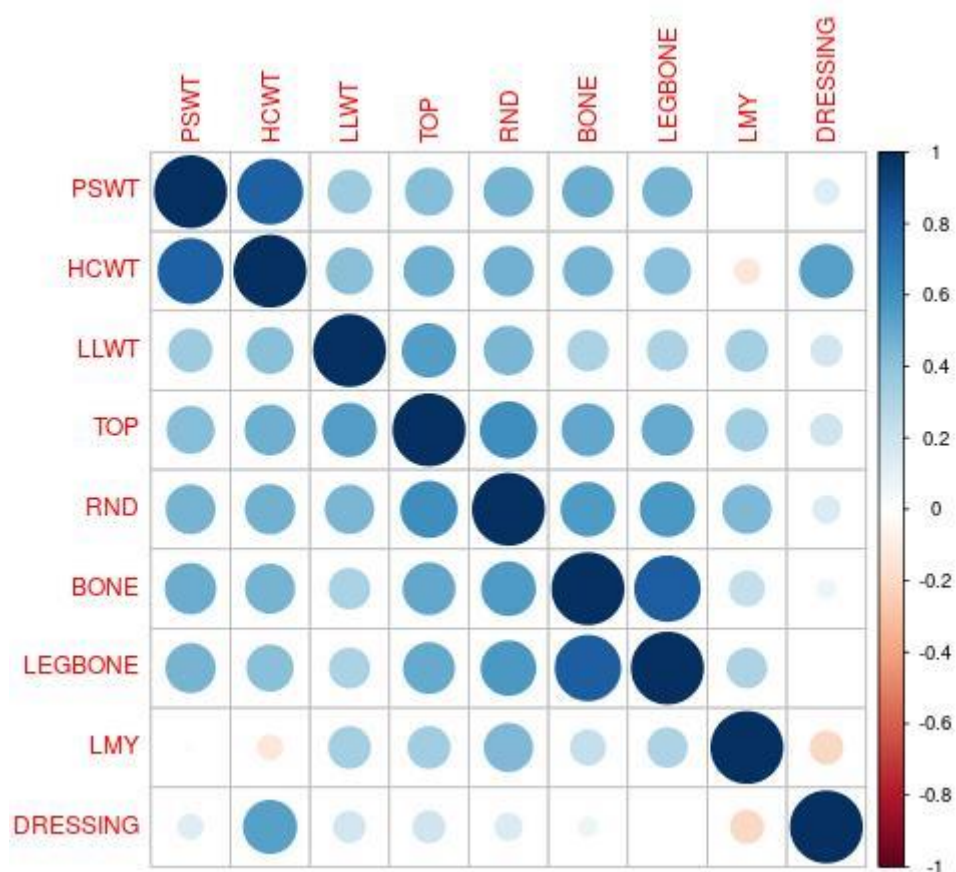
Признак	Количество животных кластер 1	Количество животных кластер 2
индекс 1 6 дней	84	23
индекс 2 6 дней	85	23
индекс 3 6 дней	84	23
индекс 4 6 дней	85	23
индекс 5 6 дней	82	23
индекс 6 6 дней	85	23
индекс 7 6 дней	84	23
индекс 1 42 дня	82	23
индекс 2 42 дня	82	23
индекс 3 42 дня	82	23
индекс 4 42 дня	81	23
индекс 5 42 дня	79	22
индекс 6 42 дня	82	23
индекс 7 42 дня	81	23
индекс 1 3 месяца	56	18
индекс 2 3 месяца	57	19
индекс 3 3 месяца	57	19
индекс 4 3 месяца	54	19
индекс 5 3 месяца	57	18
индекс 6 3 месяца	50	19
индекс 7 3 месяца	57	18
масса 6 дней	85	23
масса 42 дня	82	23
масса 3 месяца	56	18

Доп. таблица 5. Факторы инфляции тестовой статистики для результатов мета-анализа и многомерного анализа на ПГИА российской выборки овец.

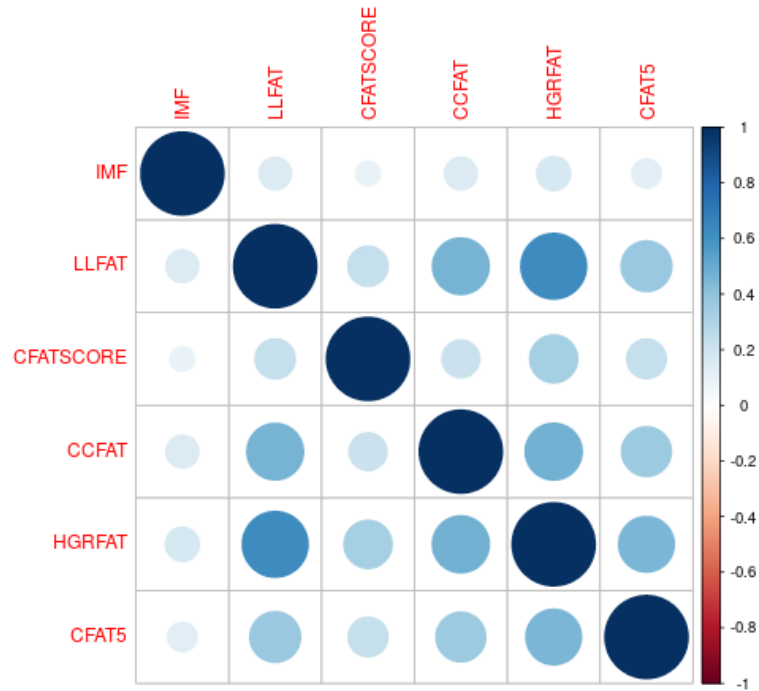
Признак	Лямбда
Индекс 1 6 дней	1.04649181829142
Индекс 2 6 дней	1.06249912290118

Индекс 3 6 дней	1.11363497330631
Индекс 4 6 дней	1.01926250086747
Индекс 5 6 дней	0.954578031709976
Индекс 6 6 дней	1.10607572287435
Индекс 7 6 дней	1.08561022828838
Масса 6 дней	1.04649181829142
Индекс 1 42 дня	1.07770102301269
Индекс 2 42 дня	1.18980716372825
Индекс 3 42 дня	1.10406645616895
Индекс 4 42 дня	1.0910728947139
Индекс 5 42 дня	1.06201140657259
Индекс 6 42 дня	1.11818915849148
Индекс 7 42 дня	0.993483723665425
Масса 42 дня	1.11717590612961
Индекс 1 3 месяца	0.996737991945094
Индекс 2 3 месяца	1.03591968997585
Индекс 3 3 месяца	1.02162998777294
Индекс 4 3 месяца	0.995807412081674
Индекс 5 3 месяца	1.06445164984693
Индекс 6 3 месяца	1.02970959370413
Индекс 7 3 месяца	1.00983265733743
Масса 3 месяца	1.11262483329924
Индекс 1 многомерный	1.08969292808687
Индекс 2 многомерный	1.24035482473994
Индекс 3 многомерный	1.23302462079166
Индекс 4 многомерный	1.08357033337675
Индекс 5 многомерный	1.08318260973951
Индекс 6 многомерный	1.17364488284329
Индекс 7 многомерный	1.08494796951025

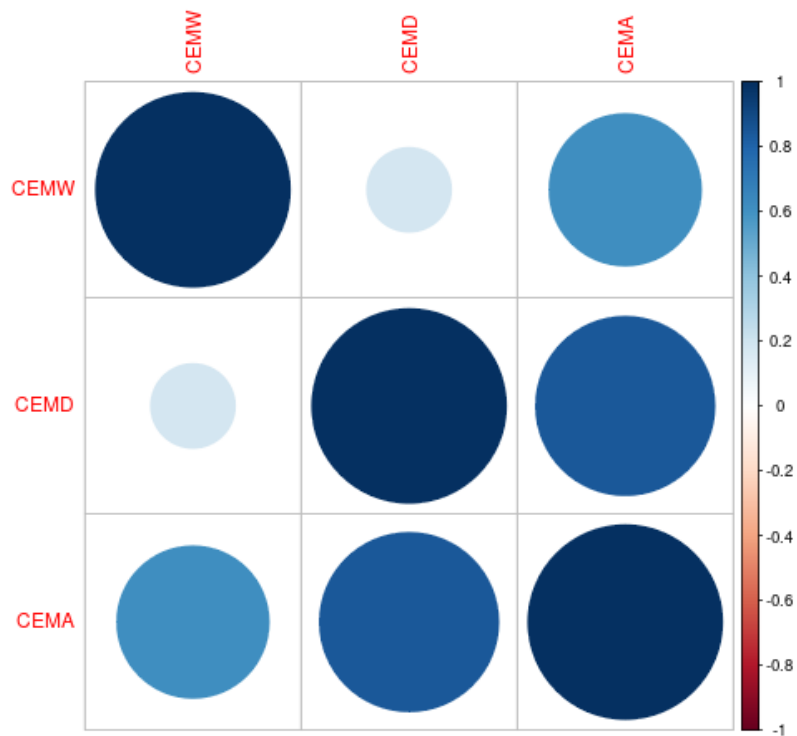
Масса многомерный	1.26512701719654
-------------------	------------------



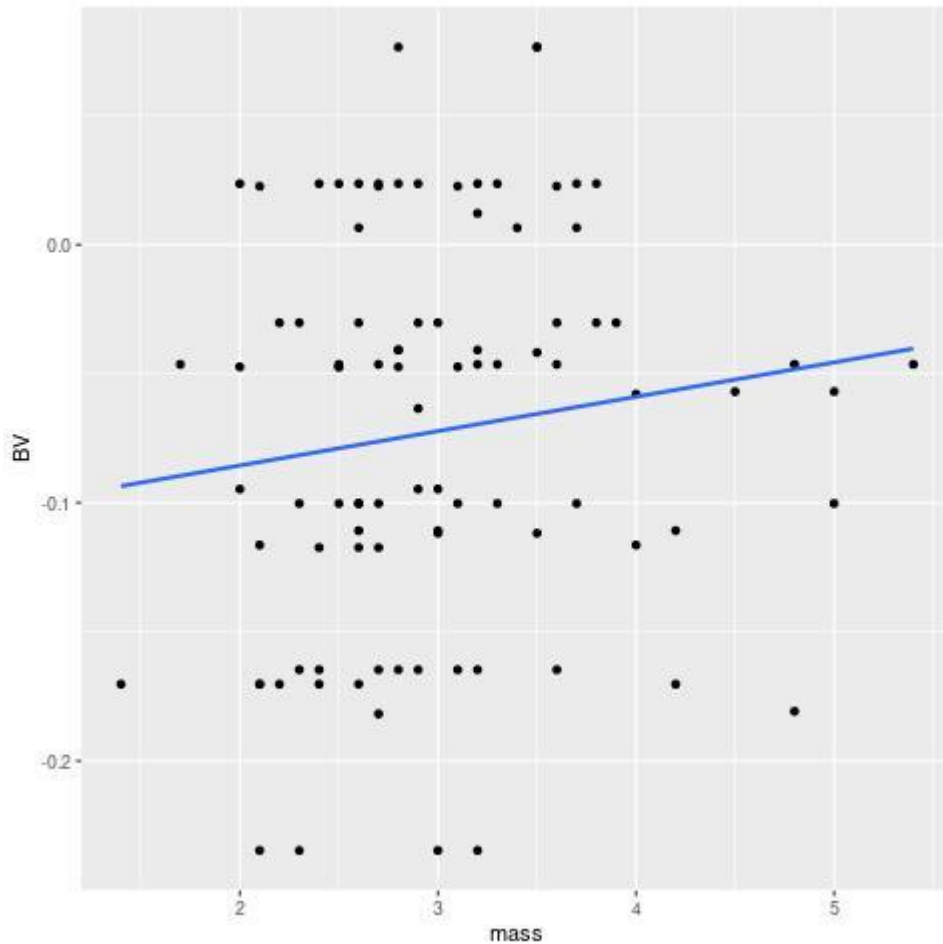
Доп. рисунок 1. Матрица корреляций для признаков, входящих в состав многомерного признака «ММасс». Расшифровка названий признаков дана в Таблице 3.



Доп. рисунок 2. Матрица корреляций для признаков, входящих в состав многомерного признака «MFat». Расшифровка названий признаков дана в Таблице 3.



Доп. рисунок 3. Матрица корреляций для признаков, входящих в состав многомерного признака «ММeat». Расшифровка названий признаков дана в Таблице 3.



Доп. рисунок 4. Диаграмма рассеяния для показателей массы и племенной ценности, российской популяции овец. По оси абсцисс указана масса животного через 6 дней после рождения, по оси ординат оцененная ПЦ с использованием признака HCWT (вес туши после забоя) из данных Volortmaa с соавторами 2016. Синим показана регрессионная прямая.