

На правах рукописи

Злобин Александр Сергеевич

**ИЗУЧЕНИЕ ГЕНЕТИЧЕСКОГО КОНТРОЛЯ
МЯСНОЙ ПРОДУКТИВНОСТИ ОВЕЦ С
ИСПОЛЬЗОВАНИЕМ СОВРЕМЕННЫХ МЕТОДОВ
КОЛИЧЕСТВЕННОЙ ГЕНЕТИКИ**

1.5.7. - Генетика

АВТОРЕФЕРАТ

Диссертации на соискание учёной степени
кандидата биологических наук

Новосибирск
2023

Работа выполнена в лаборатории рекомбинационного и сегрегационного анализа ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», г. Новосибирск.

Научный руководитель: Цепилов Яков Александрович к.б.н., старший научный сотрудник лаборатории рекомбинационного и сегрегационного анализа ФГБНУ «ФИЦ Институт цитологии и генетики СО РАН», г. Новосибирск

Официальные оппоненты: Бекетов Сергей Валерьевич д.б.н., ведущий научный сотрудник лаборатории сравнительной генетики животных, ФГБУН Институт общей генетики им. Н.И. Вавилова РАН, г. Москва

Денискова Татьяна Евгеньевна к.б.н., ведущий научный сотрудник группы генетики и геномики мелкого рогатого скота, ФГБНУ ФИЦ ВИЖ им. Л.К. Эрнста, Городской округ Подольск, поселок Дубровицы

Ведущее учреждение: ФГБНУ «Северо-Кавказский федеральный научный аграрный центр», Ставропольский край, г. Михайловск

Защита диссертации состоится «__»_____ 2023г. на утреннем заседании диссертационного совета 24.1.239.01 на базе ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики СО РАН», в конференц-зале Института по адресу: пр. академика Лаврентьева 10, г. Новосибирск, 630090, тел.: (383) 363-49-06 (1321); факс: (383) 333-12-78;

e-mail: dissov@bionet.nsc.ru

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте института www.icgbio.ru

Автореферат разослан «__»_____ 2023г.

Ученый секретарь диссертационного совета

Доктор биологических наук

Т.М. Хлебодарова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

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 генетики, 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 [Jiang и др., 2019]. Изучение таких сложных и комплексных признаков позволяет получать новые знания о биологических системах, которые участвуют в формировании таких 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) [Hayes и др., 2009; Magalhães и др., 2019]. Термин МОС применяется к методам, которые используют разнообразные молекулярные маркеры, в том числе ДНК-маркеры для достижения результата [Soller, 1994]. ГС ориентирована на использование генома родителей для предсказания фенотипа потомства с помощью оценки племенной ценности (ПЦ) [Voichard и др., 2016]. 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. [Hu и др., 2013]. Благодаря этому на коровах эффективно используется ГС для увеличения количественных показателей [Mrode и др., 2019]. 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 всего около десяти маркеров (*MSTN*, *GH*, *CLPN1*, *CAST* и др.). Необходимы новые исследования для расширения списка маркеров, которые потенциально могут использоваться в МОС.

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 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» [Hu и др., 2013], «iSheep» [Wang и др., 2021] и др. Такая усеченная информация не позволяет сопоставлять различные породы, проводить мета-анализ, а также использовать новые методы анализа результатов ПГИА.

В области исследований генетики человека для решения подобных задач разрабатывают специальные платформы и базы данных. Одной из таких является платформа GWAS-MAP [Shashkova и др., 2021b], 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 [Stephens, 2013], что широко используется в генетике человека [Klarić и др., 2020; Shen и др., 2017] и животных [Bolormaa и др., 2014; Bolormaa и др., 2016]. 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.

Цель исследования - поиск новых локусов, контролирующих признаки, относящиеся к мясной продуктивности овец, с использованием современных методов количественной генетики.

Задачи исследования:

- 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 работы. 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://pheligeovis.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 реплицирован 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 создания моделей маркер-ориентированной селекции. Впервые в мире был проведен анализ колокализации для овец.

Методология и методы диссертационного исследования. В качестве основы для создания платформы GWAS-MAP|ovis в данной работе использовалось программное обеспечение и информационное окружение платформы GWAS-MAP для человека [Shashkova и др., 2021b].

В рамках апробации использования платформы GWAS-MAP|ovis для признаков мясной продуктивности у овец были использованы различные методы количественной генетики, такие как многомерный анализ, анализ колокализации, построение функциональной сети взаимодействия генов, репликация результатов ПГИА на независимых выборках.

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

Степень достоверности результатов. Степень достоверности результатов многомерного анализа подтверждается согласованностью с ранее опубликованными данными и репликацией 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 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, приоритизированные 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 объем работы. Работа состоит из введения, обзора литературы, материалов и методов, результатов, обсуждения, заключения, выводов, списка литературы (137 источников) и приложения. Объем 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: First ESDAR-ECAR Online Conference, 11-16 October 2021; 10th Moscow Conference on Computational Molecular Biology, Москва, Россия, 2021; «Генетика, селекция и биотехнология животных: на пути к совершенству», Пушкин, Россия, 2020; BGRS/SB-2020, Новосибирск, Россия, 2020; «Достижения в генетике, селекции и воспроизводстве сельскохозяйственных животных», Санкт-Петербург, Россия, 2019.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Глава 1. Обзор литературы

В разделе 1.1 дано краткое описание методов селекции в целом и ее основных направлений: классическая селекция, маркер-ориентированная селекция и геномная селекция. В разделе 1.2 представлено общее состояние овцеводства в России с упором на мясную селекцию в овцеводстве. В разделе 1.3 дано описание статистических моделей ПГИА, анализ результатов ПГИА и краткий обзор ПГИА в мясном овцеводстве. В разделе 1.4 рассмотрены основные решения для хранения и анализа результатов ПГИА в области исследований генетики человека и животных. В разделе 1.5 приведено краткое заключение, обосновывающее цели и задачи диссертации.

Глава 2. Материалы и методы

В качестве основы для создания платформы GWAS-MAP|ovis была использована платформа для хранения результатов ПГИА человека GWAS-MAP [Shashkova и др., 2021b]. Кроме того, для представления и визуализации генетических данных был использован веб-интерфейс PheLiGe [Shashkova и др., 2021a]. Принцип устройства платформы GWAS-MAP|ovis представлен на рисунке 1.

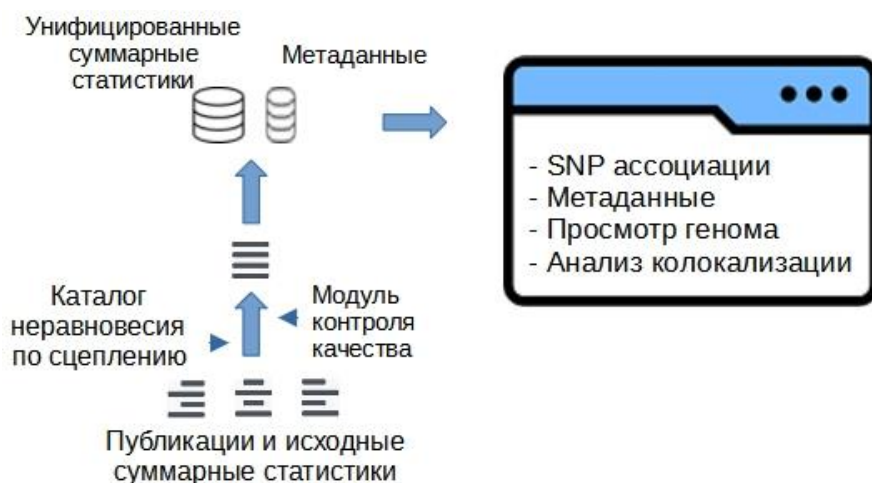


Рисунок 1. Схема устройства платформы GWAS-MAP|ovis.

Для создания базы данных QTL был проведен комплексный поиск работ, описывающих ПГИА на признаках мясной продуктивности у овец. Поиск и фильтрация статей проводились с использованием сайтов PubMed (<http://pubmed.com>) и Google Scholar (<http://scholar.google.com>).

Для наполнения платформы GWS-MAP|ovis были использованы суммарные статистики из статьи S. Bolormaa с соавторами [Bolormaa и др., 2016]. 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.

Многомерный анализ был проведен с использованием суммарных статистик, описанных выше. Для проведения функциональной аннотации результатов многомерного анализа был использован Ensembl Variant Effect Predictor (VEP). Для приоритизации генов-кандидатов был использован список генов, расположенных в пределах ± 500 тыс. п.н. от самого ассоциированного ОНП. Тестирование гипотезы о функциональной связи найденных локусов было осуществлено путем попарного анализа корреляций между векторами эффектов, ассоциированных ОНП для всех 56 признаков.

Для создания каталога неравновесия по сцеплению, репликации результатов многомерного анализа и оценки племенной ценности с помощью созданной платформы была использована выборка из 196 животных, полученная коллегами из ВИЖ им. Л. К. Эрнста (г. Москва). В выборке представлены животные 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 122 животных были доступны различные промеры туловища в разные промежутки жизни животного (6 дней, 42 дня, 3 месяца после рождения).

Глава 3. Результаты

3.1 Создание платформы для комплексного анализа результатов полногеномных исследований ассоциаций

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. На данный момент каталог неравновесия по сцеплению содержит в себе информацию о 523578 ОНП.

Следующим шагом стало создание платформы GWAS-MAP_{ovis} на основе платформы GWAS-MAP. Созданная платформа состоит из двух компонент. Первая компонента БД хранит информацию об ассоциациях между генотипом и фенотипом – параметры олигонуклеотидного полиморфизма (ОНП) и его эффекта на фенотип. Вторая компонента платформы хранит метаданные об исследовании – дескрипторы, и находится под управлением СУБД PostgreSQL, в которой содержится одна таблица (таблица дескрипторов). 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 2).

Рисунок 2. Скриншот веб-интерфейса платформы GWAS-MAP|ovis

На данный момент платформа GWAS-MAP|ovis содержит в себе информацию о 80 результатах полногеномных исследований ассоциаций (РПГИА), связанных с мясной продуктивностью, формированием каркаса и ростом у овец и содержит в себе информацию о более чем 25 миллионах ассоциаций генетический полиморфизм.

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). Все статьи из полученного списка были вручную просмотрены и отфильтрованы по релевантности. После данного шага был получен список из 17 статей, на основе которых была создана база данных генов и QTL, ассоциированных с признаками мясной продуктивности, формирования каркаса и ростом у овец.

База данных QTL содержит в себе информацию об 198 (172 уникальных) генах и 163 (130 уникальных) ОНП на основе 17 отобранных статей. Полученная база данных содержит значительно больше информации по признакам мясной продуктивности у овец, чем «AnimalQTLdb». 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 признака из статьи S. Volormaa с соавторами 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.

Для 3 многомерных признаков было найдено 12 значимо ($p\text{-value} < 3.38e-08$) ассоциированных локусов (Таблица 1). Девять из них ассоциированы с MMass и 3 с

MFat. Из 12 найденных локусов, 4 были описаны ранее. Результаты многомерного анализа в виде «Manhattan plot» представлены на рисунке 3.

Таблица 1. Двенадцать локусов, ассоциированных как минимум с одним многомерным признаком

ОНП	Хромосома	Позиция	p_value_MMmass	p_value_MFat	p_value_MMeat	ра/за
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-94	9.73E-24	5.11E-12	T/C
rs161042491	11	26445930	2.34E-41	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

Новые локусы отмечены звездочкой. Жирным шрифтом указаны значимые p-value. ра/за – референсный и эффекторный аллель соответственно.

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 [Sirmaci и др., 2010].

Следующим шагом стало проведение приоритизации и аннотации генов на основе литературных данных (Таблица 2). В результате было приоритизировано 13 генов для 8 новых 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 S.Bolormaa с соавторами [Bolormaa и др., 2014; Bolormaa и др., 2016] и в недавней работе по изучению уровней гликозилирования иммуноглобулина класса-G (IgG) человека [Klarić и др., 2020].

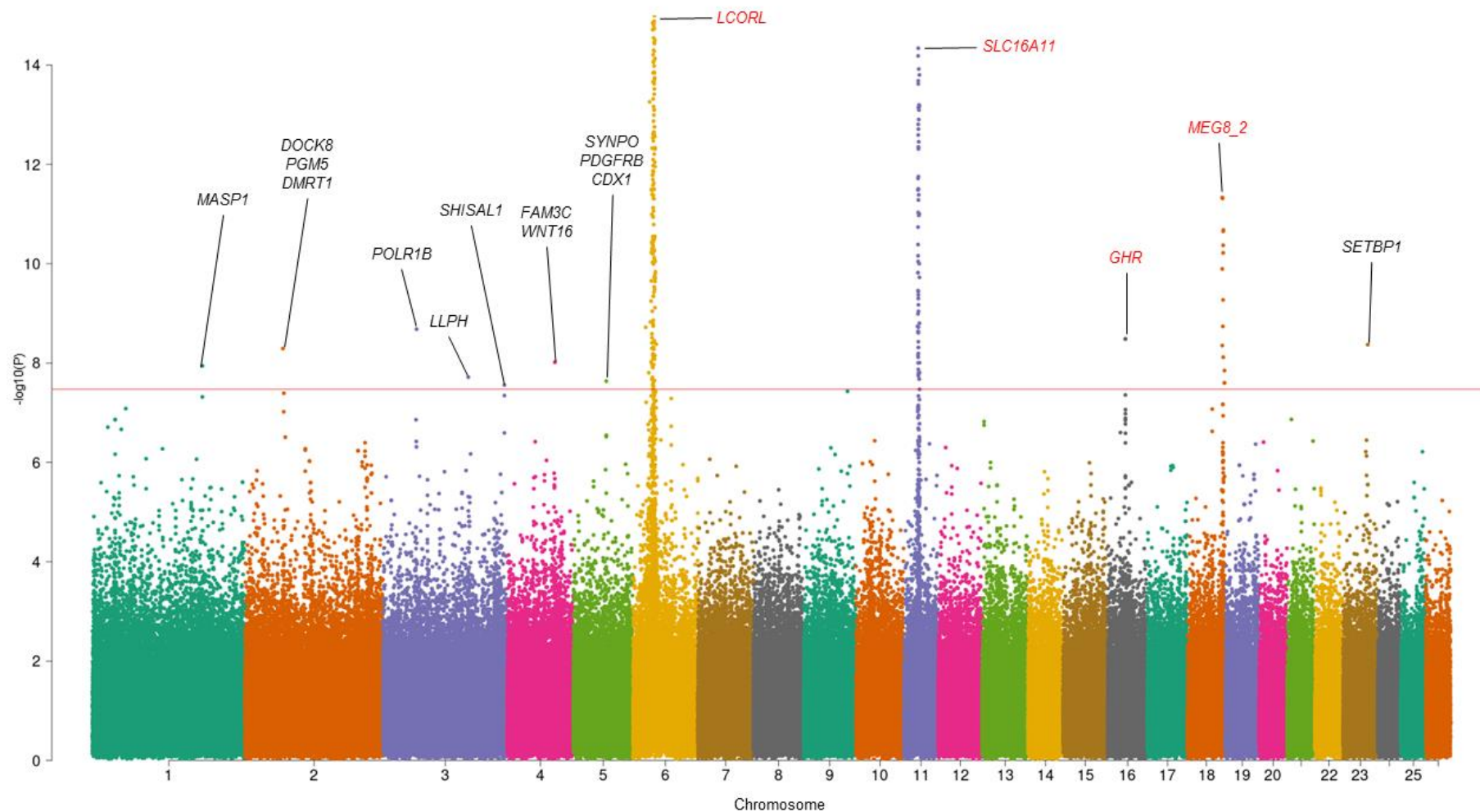


Рисунок 3. Manhattan plot для 3 многомерных признаков. Для каждого ОНП был использован минимальный p-value среди 3 многомерных признаков. Для каждого значимо ассоциированного локуса представлены названия приоритизированных генов. Известные локусы и гены в них обозначены красным.

Таблица 2. Список приоритизированных генов.

Локус	Количество генов в регионе	Ближайший ген	Гены с миссенс мутацией	Приоритизация на основе литературы	Гены-кандидаты
rs193632759	14	<i>MASP1</i>	<i>MASP1</i>	<i>MASP1</i>	<i>MASP1</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>SHISAL1</i>			<i>SHISAL1</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 проверки гипотезы мы рассчитали попарную корреляцию Спирмана между Z-статистиками ассоциированных ОНП для 56 признаков из оригинальной работы S. Volormaa с соавторами. В данном случае знак корреляции не имеет значения из-за того, что направление эффекта зависит от того, какой аллель использовался как референсный. Функциональный граф для генов как минимум с одной значимой корреляцией ($p\text{-value} \leq 7.6e-04$) представлен на рисунке 4.

Наибольшая корреляция наблюдается между 2 известными локусами на 6 (*LCORL*) и 11 (*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*).

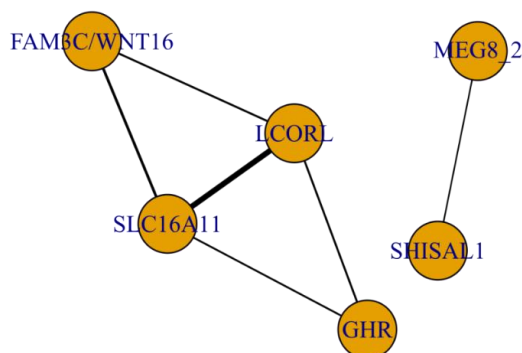


Рисунок 4. Функциональная сеть взаимодействий генов, связанных с мясной продуктивностью. Каждая линия представляет собой значимую корреляцию Спирмана ($p\text{-value} \leq 7.6e-04$). Толщина линии прямо пропорциональна размеру корреляции.

С использованием российской популяции овец (N=108) была проведена репликация 6 из 8 новых найденных локусов и локусов из базы данных QTL. В качестве материала для репликации были использованы результаты многомерного анализа для 8 признаков, связанных с мясной продуктивностью. Было реплицировано два локуса: rs401990068, найденный в данной диссертационной работе, около генов *FAM3C* и *WNT16* (p-value < 6.08e-05), ассоциированный с индексом длинноногости и rs401834107 (ген *LCORL*) (p-value < 1.1e-06), ассоциированный с тазо-грудным индексом. 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.

3.2 Анализ колокализации

Одним из преимуществ созданной нами платформы является возможность проведения так называемого анализа колокализации (SMR- θ) [Momozawa и др., 2018]. Данный анализ позволяет сравнивать паттерны ассоциации конкретного геномного локуса с разными признаками и делать выводы о наличии или отсутствии плеiotропного эффекта данного локуса на те или иные признаки. Вкратце, при справедливости гипотезы о плеiotропном 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 подсчете которой используется информация о значениях p-value и направлении эффекта. Высокое абсолютное значение (например, $|\theta| > 0.7$) говорит в пользу плеiotропного 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).

С помощью платформы GWAS-MAP|ovis был проведен анализ колокализации для самого значимо ассоциированного сигнала, полученного с помощью применения многомерного анализа на данных S. Volormaa с соавторами (см. раздел 3.3). Для ОНП rs401834107 (ген *LCORL*) были выбраны признаки, для которых значение p-value < 5e-07. Выбор такого порога обусловлен тем, что S. 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 методов

SMR и метода θ для установления того, сходны ли региональные профили ассоциации. Результаты для анализа методом θ представлены на рисунке 5. В данном случае знак указывает на то, сходным (+) или противоположным (-) образом, данный ОНП влияет на исследуемые признаки. Как видно из тепловой карты, признаки образуют два кластера, отличающиеся направлением корреляции между паттернами, но при этом сильно связанные друг с другом. Отдельно стоит отметить два выпадающих признака - убойная масса туши и ширина мышечных глазков, для которых некоторые значения θ оказались низкими (отсутствие плейотропии). Однако данные признаки по отдельности показывают плейотропный 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 плейотропный эффект на двенадцать различных признаков, связанных с мясной продуктивностью у овец. Данный анализ является первым применением в мире метода колоколизации на овцах.

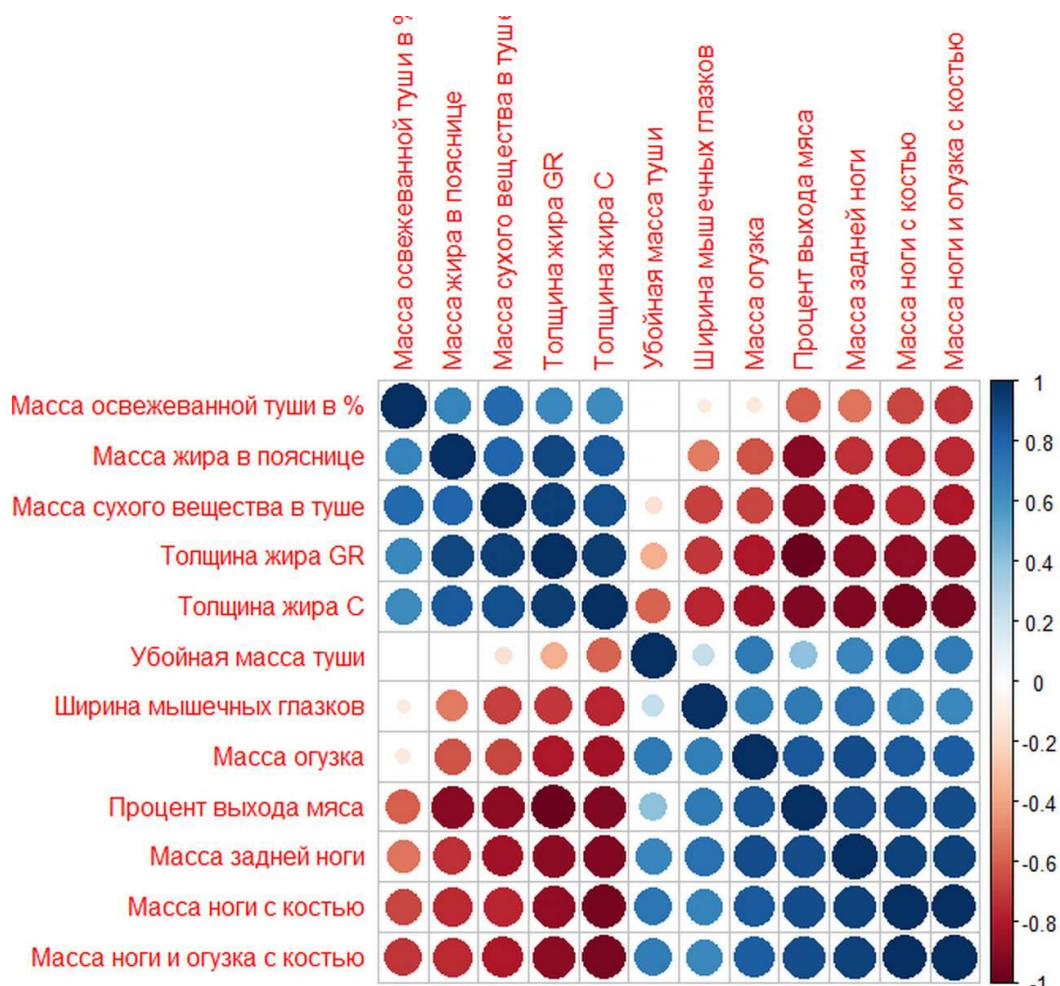


Рисунок 5. Результаты анализа колокализации для ОНП rs401834107 на 12 оригинальных признаках из статьи S. Volormaa с соавторами. Красным цветом показан отрицательный знак значения θ , голубым цветом показан положительный знак значения θ . Чем ярче цвет, тем ближе абсолютное значение к единице.

3.5 Оценка племенной ценности признаков мясной продуктивности на российской выборке овец

В дополнение к вышеописанному анализу плейотропии, 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 одного из признаков (НСWT, вес туши после забоя) 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.

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). Племенная ценность была оценена по следующей формуле $ПЦ = \sum(b * g)$, 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). В результате, оцененная ПЦ была номинально значимо ассоциирована с признаком НСWT (вес туши после забоя) ($p\text{-value} = 0.03$), что говорит о том, что отобранные нами локусы номинально значимо влияют на показатель массы животного через 6 дней после рождения на российской выборке.

ЗАКЛЮЧЕНИЕ

Диссертационная 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 теорию 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 применением в мире 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.

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.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

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