

Федеральное государственное автономное образовательное учреждение
высшего образования «Южный федеральный университет»

На правах рукописи

БАКОЕВ Сирождин Юсуфович

**АНАЛИЗ ГЕНЕТИЧЕСКОЙ АРХИТЕКТУРЫ
СЕЛЕКЦИОННО-ЦЕННЫХ ПРИЗНАКОВ СВИНЕЙ
С ПРИМЕНЕНИЕМ СОВРЕМЕННЫХ МЕТОДОВ ОБРАБОТКИ И
ИНТЕРПРЕТАЦИИ ГЕНОМНЫХ ДАННЫХ**

1.5.7. Генетика (биологические науки)

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
доктора биологических наук

Ростов-на-Дону – 2025

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования «Донской государственной аграрный университет» и на кафедре генетики Академии биологии и биотехнологии имени Д.И. Ивановского федерального государственного автономного образовательного учреждения высшего образования «Южный федеральный университет»

Научный консультант: Усатов Александр Вячеславович

доктор биологических наук, ФГАОУ ВО «Южный федеральный университет», Академия биологии и биотехнологии им. Д.И. Ивановского, профессор кафедры генетики

**Официальные
оппоненты:**

Куликов Алексей Михайлович

доктор биологических наук, заведующий лабораторией эволюционной генетики развития ФГБНУ «Институт биологии развития им. Н.К. Кольцова» РАН

Орлов Юрий Львович

доктор биологических наук, профессор кафедры информационных технологий и обработки медицинских данных ФГАОУ ВО Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет)

Ковалюк Наталья Викторовна

доктор биологических наук, заведующая лабораторией биотехнологии ФГБНУ "Краснодарский научный центр по зоотехнии и ветеринарии", заведующая лабораторией МГЭ ООО НПО "Юг-Плем"

Защита состоится «17» октября 2025 года в 14:00 на заседании диссертационного совета ЮФУ801.01.07 по биологическим наукам на базе Академии биологии и биотехнологии им. Д.И. Ивановского Южного федерального университета по адресу: 344090, г. Ростов-на-Дону, пр. Стачки, 194/1, к. 712.

С диссертацией можно ознакомиться в Зональной научной библиотеке им. Ю.А. Жданова Южного федерального университета по адресу: г. Ростов-на-Дону, ул. Зорге, д. 21-Ж, 2 этаж и на сайте Южного федерального университета <https://hub.sfedu.ru/diss/show/1345823/>

Автореферат разослан «__» _____ 2025 г.

Отзыв на автореферат в 2-х экз. (с указанием даты, полностью ФИО, ученой степени со специальностью, звания, организации, подразделения, должности, адреса, телефона, e-mail), заверенный печатью организации, просим направлять по адресу: 344090, г. Ростов-на-Дону, пр. Стачки, 194/1, к. 106, ученому секретарю диссертационного совета ЮФУ801.01.07 Бутенко Е.В., а также в формате .pdf на e-mail: evbutenko@sfedu.ru.

Ученый секретарь
диссертационного совета

Бутенко Елена Викторовна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Современные геномные технологии создают новые перспективы для повышения эффективности животноводства, обеспечивая возможность идентификации генетических маркеров, отвечающих за продуктивность, устойчивость к болезням и адаптацию к меняющимся условиям окружающей среды. Это играет ключевую роль в укреплении продовольственной безопасности и поддержании устойчивого развития сельского хозяйства в контексте климатических изменений и сохранения биоразнообразия. Методы биоинформатики позволяют анализировать большие объемы генетических данных, что значительно ускоряет процесс селекции и улучшения пород.

Геномы сельскохозяйственных животных формировались в процессе естественного и искусственного отбора, что позволило им приспособиться к различным условиям окружающей среды, пройти процесс доместикации и сформировать высокопродуктивные фенотипы. Одним из ключевых показателей геномного разнообразия является аутозиготность. Аутозиготность – это гомозиготность, возникающая вследствие наследственности (Gibson et al., 2006; Ceballos et al., 2018). Сегменты аутозиготности представлены непрерывными длинными сериями гомозиготных генотипов (RON, Runs Of Homozygosity). Протяженные гомозиготные участки в геноме возникают из-за передачи одинаковых гаплотипов от обоих родителей. Методы анализа RON позволяют точно идентифицировать аллели идентичности по происхождению (IBD, Identity By Descent) и широко применяются для оценки уровня аутозиготности у людей и животных (Ceballos et al., 2019; Vakoev et al., 2021; Kostyunina et al., 2022). Количество и размеры RON отражают демографическую историю популяции. Уникальная структура RON все больше привлекает внимание как показатель генетической архитектуры сложных признаков. В целом научное сообщество все больше приходит к выводу, что решающим фактором является не длина или количество сегментов RON, а уникальность ландшафта областей гомозиготности, что необходимо учитывать при разработке методов для исследования связей между RON и фенотипами.

Изменения в областях генома, спровоцированные отбором, стали называть «подписями отбора (селекции)» (selection signatures). Подписи указывают на генетические варианты, подвергшиеся отбору, и, следовательно, обозначают локусы значительной фенотипической изменчивости и являются объектом особого интереса при картировании сложных признаков (Druet, Gautier, 2017). Для идентификации локусов, подвергшихся отбору, были разработаны методы, позволяющие сканировать геном на наличие подписей селекции (Garud et al., 2015; Torres et al., 2018; Szpiech et al., 2021). Ключевая характеристика положительного отбора заключается в необычайно быстром увеличении частоты аллелей, происходящем в течение такого промежутка времени, чтобы рекомбинация не успевала существенно разрушить гаплотип, на котором возникла отобранная мутация. Как правило, тест на положительный отбор

включает в себя поиск основного гаплотипа с необычайно высокой гомозиготностью по расширенному гаплотипу (EHH, Extended Haplotype Homozygosity) и высокой частотой в популяции (Sabeti et al., 2002). На основе гомозиготности по расширенному гаплотипу разработаны методы iHS (integrated Haplotype Score) (Voight et al., 2006), nSI (number of Segregating sites by Length) (Ferrer-Admetlla et al., 2014), iHH12 (integrated Haplotype Homozygosity pooled) (Garud et al., 2015; Torres et al., 2018), которые получили широкое распространение в исследованиях, направленных на идентификацию подписей положительного отбора у человека и сельскохозяйственных животных.

Прогнозы на основе геномных данных, как правило, строятся в виде взвешенной линейной комбинации доступных SNP (single nucleotide polymorphisms). Для оценки весов SNP применяют полногеномный ассоциативный анализ (Genome-wide association study, GWAS), в котором интересующий результат по каждому SNP рассчитывается отдельно.

Варианты числа копий (Copy number variations, CNV) относятся к часто наблюдаемым структурным вариантам в форме делеций или дупликаций, превышающих 100 п.н. Относительно SNP они охватывают большую часть генома и имеют более высокую скорость мутаций. При этом, подобно SNP, CNV могут выступать предикторами в ассоциативных исследованиях и являться ценными источниками для анализа генетической изменчивости, способствующей различиям фенотипов селекционно-значимых признаков.

Распространенные подходы GWAS включают линейные модели, в том числе метод регрессии «лассо» (LASSO, Least Absolute Shrinkage and Selection Operator) и гребневую регрессию (Ridge regression) для выбора и регуляризации признаков, а также методы на основе алгоритмов машинного обучения и искусственных нейронных сетей. Эти методы направлены на повышение точности при выявлении значимых генетических маркеров, связанных с признаками или заболеваниями. При реализации ассоциативного анализа сложной задачей является оценка нелинейных взаимодействий между SNP и фенотипом (Behravan et al., 2018). Совмещение знаний биологии, геномики и статистики критически важно для создания эффективных инструментов, способных раскрыть механизмы, лежащие в основе сложных признаков и заболеваний.

Использование различных биоинформатических инструментов имеет огромный потенциал для научных исследований в генетике и геномике, в том числе сельскохозяйственных животных. На сегодняшний день не существует универсального метода, «золотого стандарта», применяемого для всех задач и наборов данных. Различные биологические контексты представляют уникальные характеристики данных, такие как различные уровни шума, размеры выборки и сложность признаков, которые могут повлиять на результат. В связи с этим необходимо проводить исследования, экспериментально сравнивать и определять наиболее подходящие решения для каждого уникального набора данных и исследовательских целей и

ограничений. Это позволит улучшить наше понимание генетической изменчивости и наследуемости сложных фенотипов, расширит границы биологических знаний, а также позволит более эффективно применять геномную информацию в практической работе, направленной на сохранение генетического разнообразия и создание высокопродуктивных фенотипов растений и животных, обеспечивающих продовольственную безопасность.

Цель исследования. Целью диссертационной работы является исследование генетической изменчивости свиней коммерческих пород крупная белая, ландрас и дюрок для оценки генетического разнообразия, идентификации подписей положительного отбора, отражающих селекционные процессы, а также ключевых генетических вариантов, определяющих селекционно-значимые признаки, с применением различных биоинформатических методов.

Задачи исследования:

1. Провести анализ генетической изменчивости и уровень аутозиготности диких кабанов, свиней пород крупная белая, дюрок и ландрас.
2. Провести анализ генетической изменчивости и уровень аутозиготности свиней породы крупная белая отечественной и импортной селекций.
3. Определить подписи отбора в геномах свиней породы крупная белая и провести их функциональную аннотацию относительно селекционно-значимых признаков.
4. Провести полногеномный ассоциативный анализ (GWAS) количества поросят при рождении с применением комплекса линейных моделей и алгоритмов машинного обучения. Выявить оптимальные модели, демонстрирующие наивысшую предсказательную способность в оценке генетической архитектуры репродуктивных признаков свиней (количество поросят при рождении).
5. Определить статистически значимые предикторы, ассоциированные с количеством поросят при рождении, на основе данных полногеномного ассоциативного анализа (GWAS) и идентифицированных подписей отбора.
6. Определить влияние вариаций числа копий геномных участков (CNV) на изменчивость селекционно значимых признаков свиней.
7. Разработать алгоритм визуализации и классификации областей гомозиготности (ROH) на основе данных полногеномного генотипирования с применением сверточных нейронных сетей (CNN). Провести оценку его применимости для изучения взаимосвязей между структурой ROH, происхождением и полигенными признаками.

Научная новизна и теоретическая значимость работы. Проведены крупномасштабные исследования генетической изменчивости свиней породы крупная белая на основе широкого спектра биоинформатических методов, что позволило получить информацию об эволюционных процессах, молекулярно-генетических основах адаптации, а также генетической архитектуре сложных признаков. Определены уровни аутозиготности дикого кабана и свиней пород крупная белая, ландрас и дюрок, относящихся к современному коммерческому поголовью, и получены новые сведения об истории инбридинга, недавних событиях отбора и их вкладе в адаптивные и продуктивные признаки.

Впервые изучено генетическое разнообразие и уровень аутозиготности свиней породы крупная белая, разводимых на территории РФ до и после процессов интенсификации. Идентифицированы подписи селекции, вызванные искусственным отбором. Определена генетическая дифференциация между современными свиньями породы крупная белая из разных селекционных центров и выявлены ключевые различия в их генетическом фоне. Предложена аннотация ключевых генетических локусов, отражающих подписи отбора, что проливает свет на молекулярно-генетические механизмы адаптации и селекционного процесса. Проведена оценка вклада CNV в фенотипическую изменчивость продуктивных признаков свиней, что расширяет знания о структурных вариантах генома и их возможностях для исследования генетической и фенотипической изменчивости.

Протестированы различные модели для реализации полногеномного ассоциативного анализа, показана применимость для этих задач алгоритмов машинного обучения и искусственных нейронных сетей. Для исследования генетической архитектуры репродуктивных признаков, а именно количества поросят при рождении, в качестве наилучших моделей предложены гребневая регрессия и алгоритм глубокого обучения.

Для исследования генетической архитектуры рассмотрены возможности применения синтеза методов поиска подписей отбора и полногеномных ассоциативных исследований. Этот подход был использован для поиска значимых предикторов SNP для фенотипа количества поросят при рождении, по результатам которого установлены генетические варианты, контролирующие такие процессы, как рост и развитие фолликулов, раннее эмбриональное развитие, восприимчивость эндометрия, имплантация, плацентация и эмбриональное развитие.

Предложен новый подход для анализа областей гомозиготности путем визуализации данных ROH в карты-изображения и последующего анализа с применением сверточных нейронных сетей (CNN). Эта стратегия представляет собой новый взгляд на анализ геномных данных, выходящий за рамки традиционных числовых или статистических данных и открывающий новые пути для точной дифференциации пород, улучшенного понимания структуры генома и расширенного исследования сложных признаков и заболеваний.

Практическая значимость работы. Результаты исследования вносят вклад в современное понимание генетических основ эволюции и селекции сельскохозяйственных животных в соответствии с текущими тенденциями мировых исследований в области геномики и биоинформатики. Они имеют высокую практическую значимость для научного сопровождения селекционных программ в свиноводстве. Проведённая апробация различных моделей полногеномного ассоциативного анализа (GWAS), включая алгоритмы машинного обучения и архитектуры глубоких нейронных сетей, демонстрирует реальные возможности интеграции современных биоинформатических подходов в текущие задачи прикладной генетики. На основании анализа генетической архитектуры репродуктивных признаков, в частности, количества поросят при рождении, в качестве наилучших инструментов предложены гребневая регрессия и алгоритмы глубокого обучения, которые показали устойчивую предсказательную способность на реальных данных.

Полученные модели и аналитические подходы могут быть непосредственно внедрены в практику селекционно-генетических центров для более точного отбора животных с желательными характеристиками, тем самым повышая эффективность генетического улучшения популяций. Визуализация и структурный анализ геномных данных, выполненные в работе, служат инструментами для мониторинга внутривидовой дифференциации и выявления потенциальных рисков снижения генетического разнообразия. Это обеспечивает обоснованную поддержку решений по управлению генетическими ресурсами и сохранению адаптивного потенциала пород. В настоящее время карты ROH представлены ведущим селекционно-племенным центрам и проходят апробацию. Диссертационная работа представляет собой научно-практическую основу для внедрения геномных технологий в отраслевую селекционную практику, способствуя переходу к более точной, управляемой и устойчивой системе разведения сельскохозяйственных животных.

Положения, выносимые на защиту:

1. Анализ аутозиготности позволяет получать сведения об истории инбридинга, недавних событиях отбора и их вкладе в адаптивные и продуктивные признаки свиней.
2. Подписи селекции в геноме отражают процессы адаптации в результате естественного и искусственного отбора и представляют интерес при исследовании генетической архитектуры сложных признаков.
3. Варианты копий (CNV) оказывают значимое влияние на изменчивость фенотипов селекционных признаков свиней.
4. Использование GWAS и методов идентификации подписей отбора позволяют определить предикторы, контролирующие процессы, связанные с генетической архитектурой плодовитости свиноматок.

5. Визуализация данных ROH в карты-изображения и анализ с применением сверточных нейронных сетей (CNN) представляют новый взгляд на анализ геномных данных и открывают новые возможности для точной дифференциации пород, улучшенного понимания структуры генома и расширенного исследования сложных признаков и заболеваний.

Степень достоверности и апробация результатов. Результаты исследования подтверждены использованием современных методов анализа и обработки данных, что обеспечивает их высокую степень достоверности. Основные положения и результаты исследования были представлены и обсуждены на международных и всероссийских научных конференциях, форумах и выставках: Всероссийская агропромышленная выставка «Золотая осень» (Москва, 2014–2019 гг.); Агропромышленных форумах юга России «ИнтерАгромаш» (Ростов-на-Дону, 2015–2017 гг.); Конкурсе стартапов агропромышленной тематики «АПК-Прорыв 2016» в рамках Всероссийского форума «Российское село – 2016», Москва, 2016 г.; IX Международном конгрессе «Биотехнология: состояние и перспективы», Москва, 2017 г.; Всероссийской конференции с международным участием «Генетика – фундаментальная основа инноваций в медицине и селекции», Ростов-на-Дону, 2019 г.; Международных биотехнологических форумах-выставках «РосБиоТех» (Москва, 2018–2020 гг.); Научно-практической конференции с международным участием «Генетика, селекция и биотехнология: на пути к совершенству», Пушкин, 2020 г.; Российской научно-практической конференции с международным участием «Фундаментальные основы технологического развития сельского хозяйства», г. Оренбург, ФНЦ биологических систем и агротехнологий РАН, 2019; XII Международной мультikonференции «Биоинформатика и системная биология» (BGRS/SB-2020), Новосибирск, 2020 г.; Ежегодной конференции американского и канадского общества наук о животных (2020 ASAS-CSAS-WSASAS Virtual Annual Meeting and Trade Show), г. Мэдисон, шт. Висконсин, 2020 г.; 71st Annual Meeting of European Federation of Animal Science (EAAP) Porto, Portugal, 2020 г., 72nd Annual Meeting of European Federation of Animal Science (EAAP), Davos, 2021 г.; Научно-практической конференции «Молекулярная генетика в животноводстве», Москва, 2024 г., XXVII Международном научно-производственном форуме «Аграрная наука – сельскохозяйственному производству СНГ и BRICS» посвященному 300-летию Российской академии наук, 95-летию создания ВАСХНИЛ, 55-летию Сибирского федерального научного центра агrobiотехнологий РАН, 70-летию Национальной академии наук Кыргызской Республики, 90-летию Кыргызского национального аграрного университета им. К.И. Скрябина, Кыргызская республика, Бишкек, 2024 г.

Публикации по теме диссертации. По материалам диссертации опубликовано 30 работ, из них 25 – в журналах, индексируемых в международных базах Scopus и Web of Science («Biology», «Animal Bioscience», «Genes», «Life», «PeerJ», «Цитология и генетика», «Экологическая генетика», «Animals»); 5 – в журналах, входящих в Перечень ведущих рецензируемых научных журналов РФ; 2 патента на изобретение: RU № 2634404 C2 (2017) и RU № 2822777 C1 (2024), 3 свидетельства о государственной регистрации: одна база данных (2015) и две программы для ЭВМ (2016, 2017).

Структура и объем диссертации. Диссертационная работа изложена на 323 страницах, содержит 21 таблицу, 79 рисунков, 10 приложений. Состоит из следующих разделов: введение, обзор литературы, материалы и методы исследования, собственные результаты и обсуждение, заключение, список использованных источников и приложения. Список литературы включает 385 источников.

Личный вклад автора. Автором самостоятельно был проведен анализ современного состояния проблемы, поставлены цели и задачи исследования, выбраны методы. Автор принимал участие во всех этапах работы, включая сбор данных, их обработку и подготовку текста диссертации. Личный вклад автора подтверждается публикациями по теме диссертации, а также участием в научных конференциях, форумах и выставках.

Благодарности. Автор выражает благодарность научному консультанту, доктору биол. наук Усатову А.В., ректору ФГБОУ ВО «Донской ГАУ», профессору, доктору с.-х. наук Федорову В.Х, генеральному директору ФГБУ «ЦСП» ФМБА России, доктору мед. наук Юдину С.М., заместителю генерального директора ФГБУ «ЦСП» ФМБА России, канд. мед. наук Кескинову А.А., директору Института синтетической биологии ФГБУ «ЦСП» ФМБА России, канд. биол. наук Юдину В.С, заместителю директора по науке Института синтетической биологии ФГБУ «ЦСП» ФМБА России, канд. биол. наук Макарову В.В., директору Федерального научного аграрного центра, академику РАН Клименко А.И., заведующей кафедры почвоведения и оценки земельных ресурсов ЮФУ, профессору, доктору биол. наук Минкиной Т.М., коллективу научной группы геномики и генетики сельскохозяйственных животных: доктору биол. наук Гетманцевой Л.В., канд. с.-х. наук Колосовой М.А., канд. биол. наук Шевцовой В.С., канд. с.-х. наук Колосову А.Ю., канд. с.-х. наук Романец Т.С., доктору с.-х. наук Колосову Ю.А.; сотрудникам ФГБНУ ФНЦ ВИЖ им. Л.К. Эрнста доктору биол. наук Костюниной О.В., канд. биол. наук Бакоеву Н.Ф.

Финансовая поддержка работы. Работа выполнена при финансовой поддержке грантов Российского научного фонда № 19-76-10012 «Определение генетической архитектуры репродуктивных признаков чистопородных и гибридных свиноматок с использованием полногеномных ассоциативных исследований», № 22-76-10015 «Исследование биологических механизмов формирования дефектов конечностей

свиней на основе мультимаркерного подхода» и № 23-76-10009 «Исследование генома свиней породы дюрок в аспекте породообразования, внутривидовой стратификации и интенсивного отбора по селекционно-ценным признакам».

СОДЕРЖАНИЕ РАБОТЫ

ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

В главе приведен анализ отечественной и зарубежной литературы по теме диссертации, который представлен в восьми подглавах: в первой подглаве рассмотрен процесс доместикации сельскохозяйственных животных; во второй подглаве выполнен анализ основных закономерностей генетической изменчивости процесса адаптации и отбора; в третьей подглаве представлен обзор исследования областей гомозиготности ROH; в четвертой подглаве рассмотрены основные методы идентификации областей гомозиготности ROH; в пятой подглаве представлен обзор современных исследований роли CNV с точки зрения определения связи между генетической изменчивостью и фенотипами; в шестой подглаве дана характеристика методов полногеномных ассоциативных исследований животных; в седьмой и заключительной подглавах подробно рассмотрено применение алгоритмов машинного обучения и нейронных сетей в генетике и геномике.

ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Для оценки генетического разнообразия и уровня аутозиготности *Sus scrofa* (дикий кабан) и *Sus scrofa domesticus* (домашняя свинья) были выбраны домашние свиньи трех коммерческих пород: крупная белая ($n = 68$), ландрас ($n = 73$), дюрок ($n = 70$), а также европейские дикие кабаны ($n = 87$). Данные по диким кабанам были взяты из международной базы данных DRYAD (datadryad.org). Для оценки генетического разнообразия и уровня аутозиготности были выбраны свиньи породы крупная белая отечественной (КБ1, $n = 99$) и импортной селекции (КБ2, $n = 207$; КБ3, $n = 170$). Были выбраны животные из одинаковых условий кормления и содержания. Для проведения полногеномного ассоциативного анализа с применением комплекса линейных моделей и алгоритмов машинного обучения были выбраны свиньи породы крупная белая в количестве 1200 голов, и путем синтезирования данных их количество было доведено до 5000 голов. Данные по фенотипу количества поросят при рождении (TNB) были собраны с 2018 по 2023 гг. Все свиноматки имели не менее трех опоросов. Для учета влияния дополнительных факторов (например, возрастных различий, различий в содержании, если таковые были зарегистрированы, эффектов помета, эффектов опороса, эффектов отбора, предпочтительного спаривания и т.д.) фенотипы были скорректированы в формате оценки племенной ценности (EBV) с использованием BLUPF90. Для определения влияния вариаций числа копий были выбраны свиньи породы крупная белая ($n = 100$). Для разработки алгоритма

визуализации и классификации областей гомозиготности (ROH) на основе данных полногеномного генотипирования с применением сверточных нейронных сетей (CNN) были выбраны свиньи пород крупная белая ($n = 568$) и дюрок ($n = 500$). Для оценки связи ROH с дефектами конечностей были выбраны свиньи породы крупная белая без дефекта ($n = 364$) и с дефектами ($n = 204$), дефекты были диагностированы визуально специалистами хозяйства. Генотипирование свиней проведено на основе биочипов GeneSeek® GGP PorcineHD Genomic Profiler v1 (Illumina Inc., США). Общий уровень генотипирования составил 0,99.

Плотность SNP для двух наборов данных была рассчитана с помощью пакета HandyCNV (Zhou et al., 2021). Для преобразования данных SNP в форматы PED и MAP использовали plink (Purcell et al., 2007). Управление данными SNP и их анализ выполняли в среде компьютерной математики R-4.1.0 (R Core Team, 2019) с использованием функций-оболочек пакета R SambaR (<https://github.com/mennodejong1986/SambaR>). После фильтрации и прореживания 28 915 из 60 тысяч SNP были сохранены для дальнейшего исследования. Анализ главных компонент (PCA) проводили с использованием функции «snpgdsPCA» пакета R SNPRelate-1.26.0 (Zheng et al., 2012). Коэффициенты родословной были рассчитаны с помощью программного обеспечения Admixture-1.3 (Alexander, Lange, 2011). Оценки неравновесия по сцеплению (LD) рассчитывали с использованием PLINK. Фазирование генотипов проводили с помощью программы Beagle (Browning et al., 2021). Аллель-специфическую расширенную гаплотипическую гомозиготность (EHH), интегрированную оценку гаплотипа (iHS), кросс-популяционную расширенную гаплотипическую гомозиготность (XP-EHH), nSL и интегрированную гомозиготность гаплотипа оценивали с помощью программы selscan (Szpiech, Hernandez, 2014).

Для идентификации CNV из данных массива SNP применили три программы: PennCNV v.1.0.4 (Wang et al., 2007), QuantiSNP v2.3 (Colella et al., 2007) и R-GADA 2.0.1 (Pique-Regi et al., 2010). PennCNV и QuantiSNP используют скрытую марковскую модель (HMM), тогда как GADA работает на основе модели разреженного байесовского обучения (SBL). Для объединения отдельных CNV в регионы была выполнена процедура Reciprocal Overlap (RO) с помощью пакета R CNVRanger v.1.18.0 (Silva et al., 2020). Данные были преобразованы в файл, содержащий номер хромосомы, начало и конец CNV, название образца и состояние CNV.

Для формирования карт гомозиготности сегменты ROH разделили на классы. Карты ROH были сформированы для каждого животного и сохранены в формате «JPEG» со следующими параметрами: разрешение 1200×1200 , ширина 1200 пикселей, высота 1200 пикселей, глубина цвета 24. Карты были сформированы для свиней породы дюрок и свиней породы крупная белая. На основе созданных карт гомозиготности была составлена модель сверточных нейронных сетей (CNN). По результатам модели с помощью Saliency map (Kadir T., Brady M., 2001) определили области важности и определили среднюю карту важности.

ГЛАВА 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1 Анализ генетического разнообразия у представителей рода *Sus*

Для изучения генетической структуры популяций использовали анализ главных компонент (РСА). РСА преобразует многомерные генетические данные, такие как SNP (однонуклеотидные полиморфизмы), в набор линейно-независимых компонент, что позволяет сократить размерность данных и выделить основные источники вариации. В контексте популяционной геномики с помощью РСА можно визуализировать генетические различия между популяциями, выявляя кластеры, которые отражают историю миграций, изоляции или смешения популяций. РСА помогает идентифицировать генетические маркеры, которые вносят наибольший вклад в различия между популяциями, что важно для понимания процессов адаптации и отбора. Метод также устраняет избыточность данных, связанную с высокой корреляцией между SNP, и фильтрует шум, улучшая качество анализа. Таким образом, РСА является мощным инструментом в популяционной геномике, предназначенным для анализа генетической структуры, визуализации данных и изучения эволюционных процессов.

На рисунке 1 представлена визуализация анализа главных компонент, согласно которому можно утверждать, что все особи четко дифференцированы в соответствии с породной принадлежностью. Первые две компоненты PC1 и PC2 содержат, соответственно, 20,4 и 13,9% дисперсии генетической дифференциации между свиньями различных пород и дикими кабанями.

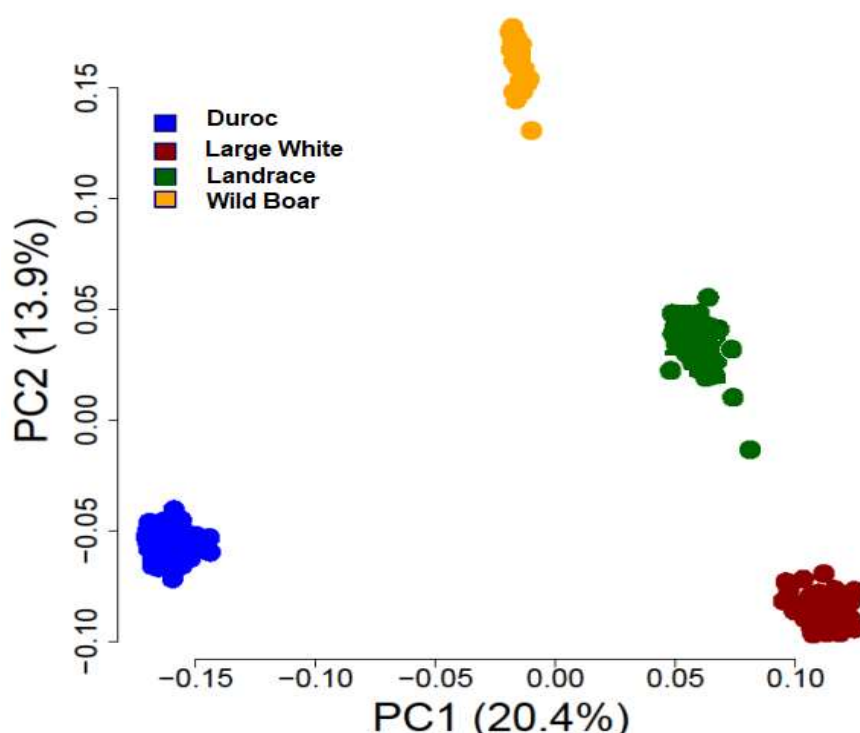


Рисунок 1 – Анализ главных компонент (PCA) для диких кабанов и свиней коммерческих пород. Представлены исследуемые популяции: дюрок (Duroc), крупная белая (Large White), ландрас (Landrace) дикие кабаны (Wild Boar)

Анализ главных компонент (PCA) выявляет генетическую дифференциацию между дикими кабанями и свиньями домашних пород дюрок, крупная белая и ландрас. Результаты анализа позволяют визуализировать генетические различия между группами, демонстрируя степень их схожести и различия. В частности, дикие кабаны образуют отдельный кластер, что указывает на значительные отличия в их генетическом составе. Домашние породы свиней, наоборот, располагаются ближе друг к другу, что отражает их общую селекционную историю и более тесную генетическую связь. Таким образом, полученные данные подтверждают существование генетической дифференциации между исследуемыми группами и свидетельствуют о воздействии факторов доместикации и селекции на их генетическую структуру.

Для оценки нуклеотидного разнообразия рассчитали долю однонуклеотидных полиморфизмов (SNP) относительно общего количества проанализированных нуклеотидных позиций (рисунок 2). Домашние породы свиней демонстрируют относительно низкое генетическое разнообразие. У домашних свиней значения гетерозиготности варьируются от 30 до 35%, а нуклеотидное разнообразие находится в диапазоне от 42 до 46%. Порода дюрок демонстрирует наименьшие показатели: гетерозиготность около 30% и нуклеотидное разнообразие около 42%, что может быть связано с интенсивной селекцией, направленной на закрепление желаемых признаков, что снижает генетическое разнообразие. У пород ландрас и крупная белая значения несколько выше: гетерозиготность составляет около 33–35%, а нуклеотидное разнообразие – 44–46%, что отражает умеренный уровень генетического разнообразия, характерный для этих пород. В целом, несмотря на небольшие различия, все три породы имеют относительно низкие уровни генетической изменчивости, что является следствием искусственного отбора.

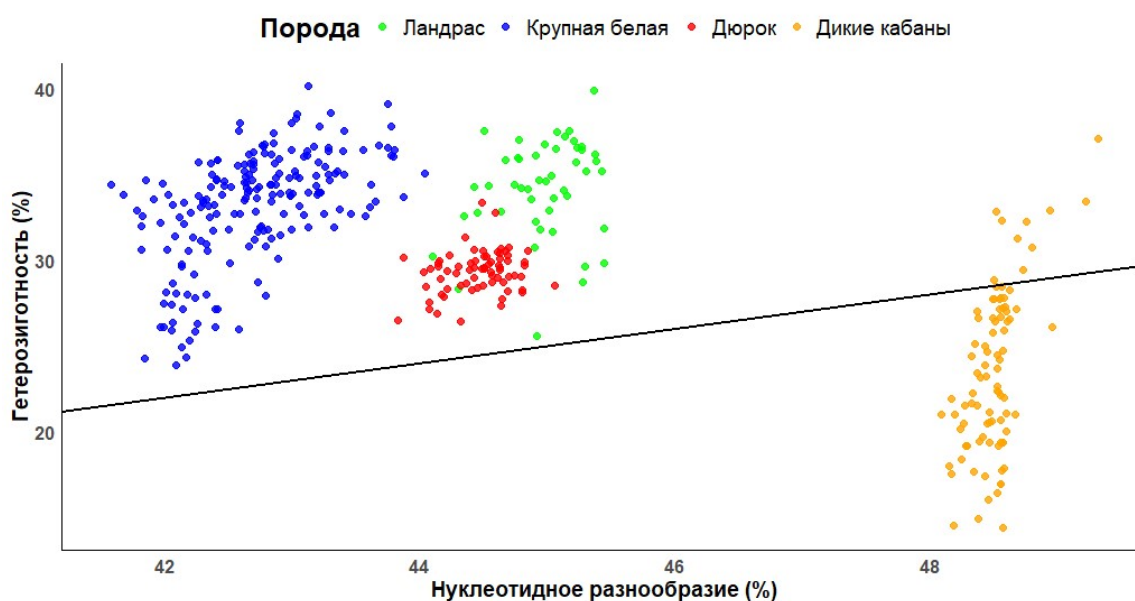


Рисунок 2 – Отношение нуклеотидного разнообразия и гетерозиготности в популяциях

Дикие кабаны, в отличие от домашних свиней, демонстрируют значительно более высокие показатели: гетерозиготность достигает 40%, а нуклеотидное разнообразие – 48%. При анализе плотности SNP были обнаружены неслучайные паттерны распределения SNP, при этом на всех хромосомах были обнаружены сильно различающиеся частоты SNP. Для кабанов SNP распределены в диапазоне 20–50 SNP в 1 Мб 1336 регионов, ≤ 20 соответственно 1120 регионов со средним значением 20 SNP/Мб, стандартным отклонением $+ SD = 28$ SNP/Мб и $- SD = 12$ SNP/Мб. Для пород ландрас, крупная белая и дюрок SNP распределены в диапазоне 20–50 SNP в 1 Мб 1729 регионов, ≤ 20 662 регионов и ≥ 50 соответственно 56 регионов со средним значением 27 SNP/Мб, стандартным отклонением $+ SD = 38$ SNP/Мб и $- SD = 16$ SNP/Мб.

Не менее важным аспектом исследования генетического разнообразия в популяциях является оценка неравновесия по сцеплению (LD, Linkage Disequilibrium). LD отражает неслучайное наследование аллелей, и его анализ позволяет оценить генетическую структуру, уровень разнообразия и эволюционные процессы, что важно для понимания адаптации и управления генетическими ресурсами. LD decay (распад неравновесия сцепления) – это мощный инструмент для изучения генетического разнообразия в популяциях. Проведенный анализ распада LD исследуемых пород показал, что породы дюрок и ландрас демонстрируют медленный распад LD (рисунок 3), порода крупная белая занимает промежуточное положение, а дикие кабаны, напротив, характеризуются быстрым распадом LD, что свидетельствует об интенсивном потоке генов и высокой скорости рекомбинации, типичных для крупных и менее изолированных популяций.

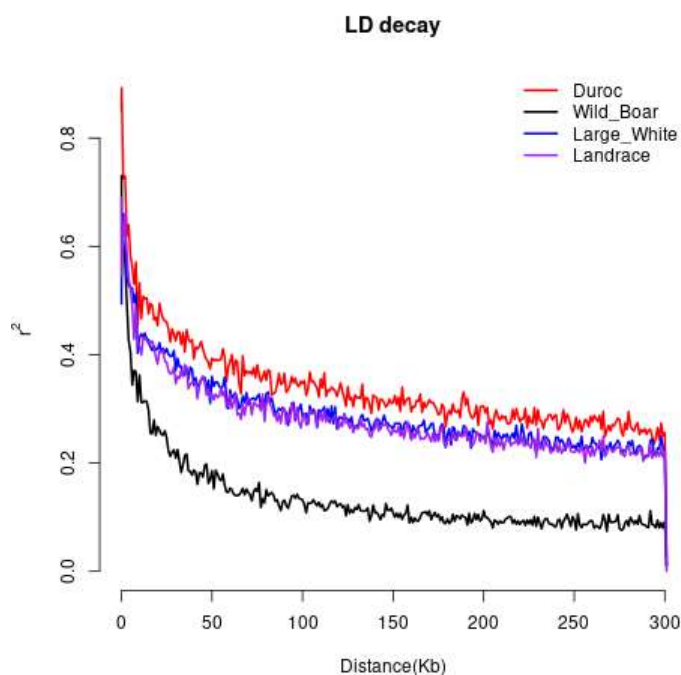


Рисунок 3 – График распада неравновесия по сцеплению (LD decay). Представлены исследуемые популяции: дюрок (Duroc), дикие кабаны (Wild Boar), крупная белая (Large White) и ландрас (Landrace)

Для оценки генетического разнообразия в контексте аутозиготности и распределения аутозиготных по происхождению сегментов в геномах кабанов и свиней основных коммерческих пород использовали модели, позволяющие не только определять сегменты аутозиготности (HBD, Homozygous-by-Descent), но и сопоставлять с возрастом предков, от которых были унаследованы эти сегменты, предложенные Druet с соавторами (Druet, Gautier, 2017). Идентификация областей HBD позволила определить коэффициенты геномного инбридинга у диких кабанов и свиней коммерческих пород. Средние значения коэффициента геномного инбридинга были практически на одном уровне во всех изучаемых группах и составили 0,23–0,29 (таблица 1).

Таблица 1 – Коэффициенты геномного инбридинга у диких кабанов и свиней коммерческих пород

Порода	Минимальное	Среднее	Максимальное
Крупная белая ($n = 68$)	0,2075	0,2518	0,3450
Ландрас ($n = 71$)	0,1779	0,2313	0,3251
Дюрок ($n = 70$)	0,2187	0,2887	0,3440
Дикие кабаны ($n = 87$)	0,0518	0,2782	0,5554

При оценке аутозиготности в соответствии с возрастом предков определено, что наибольший вклад у коммерческих пород свиней внес класс с коэффициентом $R_k = 128$ (т.е. предки около 64 лет назад) (рисунок 4).

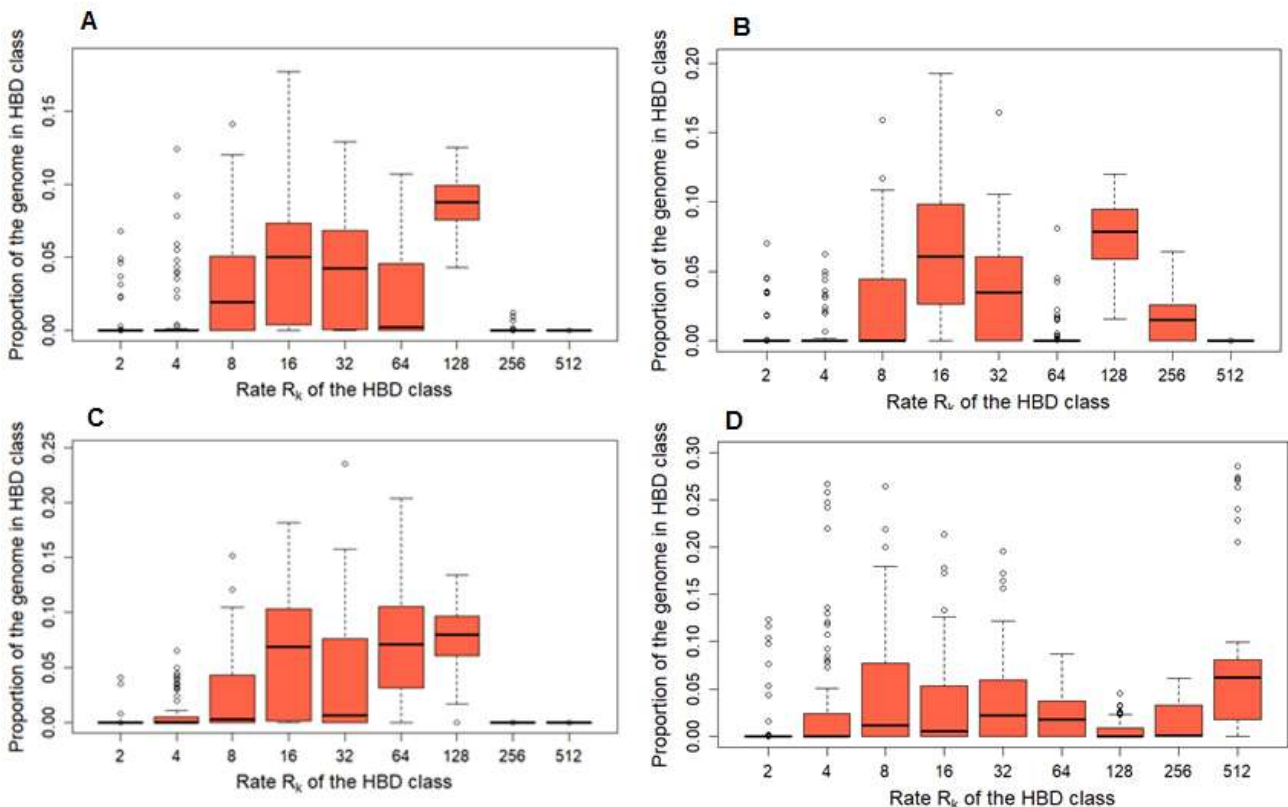


Рисунок 4 – Аутозиготность в соответствии с возрастом предков у диких кабанов и свиней коммерческих пород: А – крупная белая; В – ландрас; С – дюрюк и D – дикий кабан

У свиней породы ландрас можно выделить вклад класса $R_k = 16$ (около 8 лет назад), а у дюрюк классов $R_k = 16$ и $R_k = 64$ (около 8 и 32 лет назад). В отличие от коммерческих пород, дикие кабаны имели высокую долю аутозиготности (около 0,08 доли от генома), которую унаследовали примерно 206 лет назад.

Свиньи коммерческих пород в целом имеют общую линию роста до класса $R_k = 128$. В отличие от домашних свиней, дикие кабаны имеют очень большую вариабельностью на индивидуальном уровне и рост доли аутозиготности вплоть до класса $R_k = 512$. В целом у свиней коммерческих пород определено 30 416 сегментов HBD, в среднем 145 на одного животного; у диких кабанов – 11 598 сегментов, в среднем 133 на одного животного (таблица 2). Самый длинный сегмент размером 183,39 Мб (3 403 SNP) установлен в группе диких кабанов в первой хромосоме. Среди свиней коммерческих пород наибольшую длину имел сегмент размером 168,99 Мб (3 428 SNP) у свиней породы крупная белая, и он также был расположен в первой хромосоме.

Таблица 2 – Сводная статистика сегментов HBD у диких кабанов и свиней коммерческих пород

Параметр	Крупная белая ($n = 68$)		Ландрас ($n = 71$)		Дюрюк ($n = 70$)		Дикие кабаны ($n = 87$)	
	кол-во SNP	длина сегмента (Мб)	кол-во SNP	длина сегмен- та (Мб)	кол-во SNP	длина сегмен- та (Мб)	кол-во SNP	длина сегмен- та (Мб)
n	10489	10489	9494	9494	10433	10433	11598	11598
Mean	96,89	3,69	101,46	3,85	115,53	4,47	107,04	3,99
SE	1,31	0,06	1,62	0,07	1,39	0,07	2,07	0,09
Min	5,00	0,06	4,00	0,05	3,00	0,03	1,00	0,00
Max	3428,00	168,99	2668,00	120,74	2286,00	121,14	3403,00	183,39

Частоты SNP (%) в обнаруженных HBD были оценены для каждой группы (рисунок 5). Для каждой группы выбраны наиболее значимые HBD при условии, что частота HBD составляет не менее 60% и не менее 10 SNP. У свиней породы крупная белая данным условиям соответствовало 18 сегментов, локализованных в восьми хромосомах (SSC1, SSC2, SSC3, SSC4, SSC6, SSC8, SSC9, SSC12, SSC13, SSC15). Длина наиболее значимых HBD у свиней породы крупная белая составили от 0,54 Мб (SSC1) до 7,55 Мб (SSC15).

У свиней породы ландрас условиям наиболее значимых HBD соответствовали 9 сегментов, локализованные в SSC1, SSC4, SSC6–SSC8, SSC10, SSC14 и SSC15 (рисунок 5). Длина их варьировала от 0,30 Мб (SSC7) до 27,05 Мб (SSC1). У свиней породы дюрюк определено наибольшее количество ключевых HBD, 35 сегментов, расположенных во всех хромосомах, за исключением SSC3, SSC11, SSC16–SSC18. Минимальная длина этих сегментов составила 0,23 Мб (SSC10), а максимальная –

15,32 Мб (SSC14). У кабанов определено только 3 ключевых HBD в SSC1 (1,38 Мб), SSC4 (2,11 Мб) и SSC14 (0,31 Мб).

Для функциональной аннотации области топ HBD был проведен анализ обогащения этих областей локусами количественных признаков, полученных из базы QTLdb. Области топ HBD перекрываются с QTLs, среди которых во всех группах наиболее представлены признаки, связанные с мясной продуктивностью (Meat and Carcass traits). Однако представленность типов QTL отражает особенности селекционных стратегий для каждой из пород. Так, у свиней породы дюрок, относительно других пород и диких кабанов, QTL, связанные с мясной продуктивностью, наиболее представлены и составили более 80%. Наибольшую значимость показали признаки, ответственные за состав жирных кислот в мясе (содержание мононенасыщенных жирных кислот, стеариновой кислоты, насыщенных жирных кислот, пальмитолеиновой кислоты, пальмитиновой кислоты и олеиновой кислоты). Это соответствует целенаправленной селекционной работе с породой дюрок, ориентированной на улучшение показателей мясной продуктивности и качества мяса.

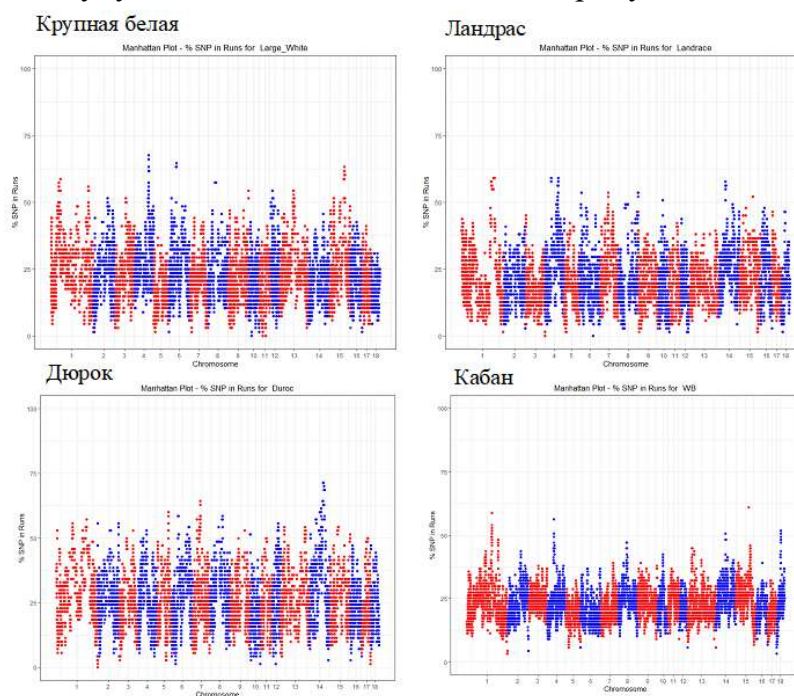


Рисунок 5 – Частоты SNP (%) в обнаруженных HBD у свиней коммерческих пород и диких кабанов

У диких кабанов значимое обогащение в областях топ HBD показали признаки, связанные с параметрами крови. Таким образом, анализ аутозиготности диких кабанов и домашних свиней позволил определить ключевые события формирования сегментов HBD. Сканирование генома для определения областей аутозиготности показало, что 0,08 доли генома диких кабанов покрыта сегментами HBD, предположительно происходящими от предков, живших около 206 лет назад. Дикие кабаны отличаются наибольшей вариабельностью HBD на индивидуальном уровне, что, вероятно, связано с тем, что в изучаемую группу вошли животные из различных областей и регионов.

Возможно, с этим также связано, что только 3 топ HBD были определены у диких кабанов. В областях топ HBD значимое обогащение показали признаки, связанные с параметрами крови, которые возможно интерпретировать в качестве предрасположенности диких кабанов к мышечному стрессу и повреждению печени. У коммерческих пород свиней (крупная белая, ландрас, дюрок) около 0,07–0,08 доли генома покрыта сегментами HBD, предположительно происходящими от предков, живших около 64 лет назад. В целом анализ сегментов HBD показал, что они отражают основные селекционные стратегии, направленные на формирование коммерческого поголовья свиней. У свиней породы дюрок более 80% QTL связаны с составом жирных кислот в мясе. Доля репродуктивных качеств у свиней пород крупная белая и ландрас в областях топ HBD составила около 5,5–6,0%. У свиней породы крупная белая, относительно других групп, значительно больше QTL (16,7%), связанных со здоровьем.

3.2 Анализ генетического разнообразия у представителей свиней породы крупная белая из различных селекционных центров

В течение длительного периода доместикиции и селекции у каждой породы сформировались свои породоспецифические, биологические и производственные особенности, проявляющиеся на геномном уровне. Одной из самых распространенных пород свиней сегодня является крупная белая. Генетическая структура этой породы достаточно пластична, что позволяет откликаться на селекционное давление, спровоцированное искусственным отбором, диктуемым различными потребностями и предпочтениями людей. Так, буквально 50 лет назад в нашей стране предпочтение отдавали свиньям с невысокой скоростью роста, но в целом не требовательным к условиям содержания и кормления и обладающим высокой устойчивостью к заболеваниям. В дальнейшем процессы интенсификации захлестнули все мировое животноводство, в том числе и Россию, и на смену низкопродуктивным свиньям пришли животные, обладающие высокой скоростью роста, высокими репродуктивными показателями, низким уровнем потребления корма, но крайне требовательные к условиям содержания и рационам питания, а также с более низкими качественными показателями мяса. Кроме того, на сегодняшний день производителями племенного поголовья свиней, в том числе и породы крупная белая, являются всего 5–6 мировых генетических центров, при этом каждый из них имеет собственные целевые показатели и селекционные стандарты. В связи с этим несмотря на то, что свиньи представляют одну породу, они имеют внутривидовую изменчивость, связанную с особенностями фенотипов селекционно-значимых признаков.

Для исследования внутривидовой изменчивости свиней породы крупная белая были созданы три группы: КБ1 – образцы свиней породы крупная белая отечественной селекции; КБ2 и КБ3 – образцы свиней современной селекции из двух селекционных центров. Максимальный показатель геномного инбридинга составил 0,38 и определен

в группе КБ2. Минимальный показатель 0,16 зафиксирован в группе КБ1. При этом можно отметить, что максимальные и минимальные показатели, как в группе КБ1, так и в группе КБ2, в большей степени являются выбросами. Значения геномного инбридинга в группах КБ1 распределены в диапазоне 0,19–0,31, практически так же, как в группе КБ3 (0,20–0,30), но в группе КБ2 диапазон значений сдвинут в правую сторону и составляет 0,25–0,35 (рисунок 6).

В целом уровень аутозиготности у свиней породы крупная белая находится практически на том же уровне, что и у свиней других пород и кабанов. Около 4–5% генома покрывают сегменты HBD классов $R_k = 16$ и $R_k = 32$ во всех группах КБ. Однако при оценке аутозиготности в соответствии с возрастом предков у свиней КБ1 наибольший вклад внес класс с коэффициентом $R_k = 256$ (т.е. предки около 128 лет назад), у современных свиней КБ наибольший вклад в аутозиготность внесли сегменты, унаследованные от предков около 64 лет назад. При этом у свиней КБ3 прослеживается аутозиготность от предков около 128 лет назад и составляет около 2% генома, в то время как у свиней КБ2 такого рода сегменты практически отсутствуют (около 0,2%).

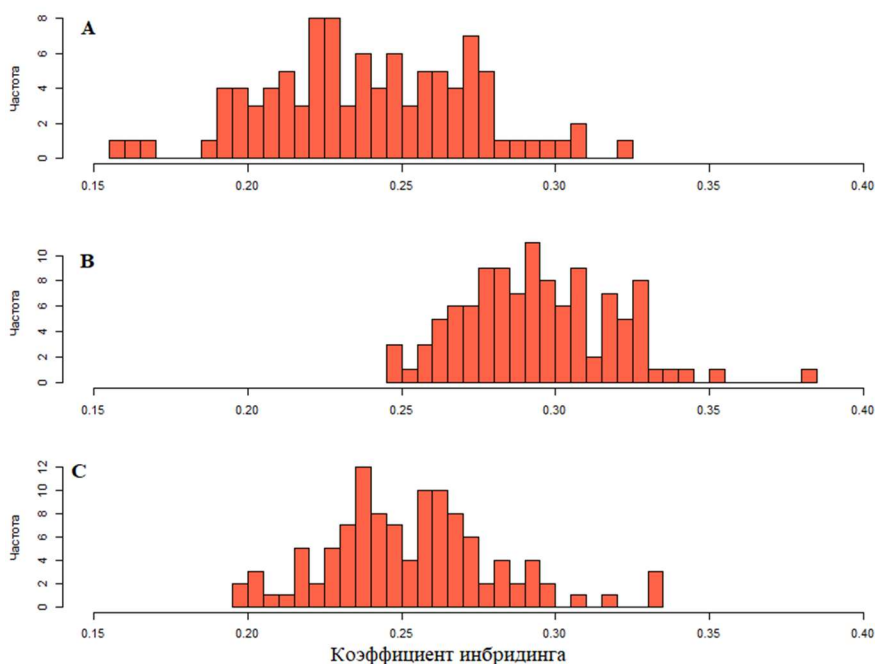


Рисунок 6 – Распределение коэффициента инбридинга у свиней породы крупная белая: А – в группе КБ1; В – в группе КБ2; С – в группе КБ3

Для каждой группы были выбраны топ HBD при условии, что частота HBD составляет не менее 60% и входят не менее 10 SNP. В результате у КБ1 были определены 4 области, расположенные в SSC1. У свиней группы КБ-2 обнаружено 10 областей, из которых 5 в SSC1, 2 в SSC6 и по одной в SSC10, SSC14 и SSC15. У свиней КБ-3 данным условиям соответствовало 7 сегментов, локализованных в хромосомах SSC4, SSC6 и SSC15 (таблица 3).

В процессе формирования отечественного поголовья одним из положительных факторов отбора выступал высокий уровень иммунитета и устойчивость к различного рода заболеваниям, что нашло отражение в сегментах НВД, наиболее представленных у свиней из группы КБ1. У свиней породы КБ1 гены, расположенные в наиболее значимых НВД, кодируют белки, которые выполняют функции модуляторов активности других белков и ингибиторов протеаз. Выполняемые ими биологические процессы и молекулярные функции сходятся на активности/ингибировании эндопептидаз, в том числе серинового типа. Сериновые эндопептидазы (протеазы) составляют почти треть всех протеаз и служат необходимыми компонентами, катализирующими гидролитические реакции, участвуют в различных физиологических процессах, таких как переваривание пищи, развитие эмбрионов и иммунная защита (Kanost et al., 2015).

Таблица 3 – Гены, локализованные в топ НВД у свиней породы крупная белая

КБ-1	
Хромосома:начало-конец (кол-во SNP)	Гены
1:63704733-64301394 (13)	<i>FHL5, GPR63, NDUFAF4, KLHL32</i>
1:116300444-117138985 (14)	<i>PYGO1, DNAAF4, C15orf65, CCPG1, PIGB, RAB27, RSL24D1</i>
1:149261646-150160033 (17)	<i>ZNF407, CNDP1, CNDP2, DIPK1C, C18orf63, CYB5A, FBX015, TIMM21</i>
1:154796451-158940596 (46)	<i>DSEL, CDH19, CDH7, SERPINB10, SERPINB7, SERPINB12, SERPINB8, SERPINB5, SERPINB2, SERPINB13, SERPINB11, VDS4B, PHLPP1, KDSR</i>
КБ-2	
1:41940391-53513159 (245)	<i>MAN1A1, FAM184A, ASF1A, CEP85L, SLC35F1, NUS1, DCBLD1, ROS1, VGLL2, RFX6, GPRC6A, FAM162B, KPNA5, ZUP1, RSPH4A, PTP4A1, PHF3, ADGRB3, LMBRD1, COL19A1, COL9A1, FAM135A, SMAP1, B3GAT2, OGFRL1, RIMS1, KCNQ5, DPPA5, OOEP, CYB5R4, MRAP2, CEP162</i>
1:56563956-57069421 (10)	<i>RNGTT</i>
1:61986232-62371709 (11)	–
1:225017887-225614369 (14)	<i>CEMIP2, ABHD17B, C9orf85, GDA</i>
1:227781084-229301982 (18)	<i>TRPM6, C9orf40, NMRK1, CARNMT1, OSTF1, PCSK5</i>
6:61261671-63135016 (20)	<i>PEG3, AURKC, ZNF304, ZNF772, ZNF773, ZNF550, ZNF606, ZNF135, ZNF329, ZNF274, ZNF8, RPS5, ZNF584, ZNF446, SLC27A5, ZBTB45, TRIM28, CHMP2A, UBE2M, MZF1</i>
6:107551091-113309905 (84)	<i>RBBP8, CABLES1, TMEM241, RIOK3, RMC1, NPC1, ANKRD29, LAMA3, TTC39C, CABYR, OSBPL1A, HRH4, ZNF521, SS18, PSMA8, TAF4B, KCTD1, AQP4, CHST9, CDH2</i>
КБ-3	
4:100967134-101855077 (27)	<i>FCRL5, ETV3, ARHGEF11, LRRC71, PEAR1, NTRK1, INSRR</i>
4:104518531-108045490 (112)	<i>NUP210L, RPS27, RAB13, JTB, CREB3L4, SLC39A1, CRTC2, GATAD2B, SLC27A3, INTS3, NPR1, S100A1, PPGRP-S, PGLYRP3, S100A12, SPRP, IVL, SMCP, CRNN, FLG2, THEM5, RORC, MRPL9, TUFT1, CELF3, LINGO4, SNX27, CGN, TDRKH, OAZ3, POGZ, SELENBP1, PSMB4, RFX5, PI4KB,</i>

	<i>ZNF687, VPS72, PSMD4, PIP5K1A, TMOD4, SCNMI, LYSMD1, SEMA6C, GABPB2, CDC42SE1, BNIPL, PRUNE, ANXA9, CERS2, SETDB1, CTSS, HORMAD1, GOLPH3L, ECM1, TARS2, ADAMTSL4, RPRD2, PRPF3</i>
6:20635220-21310322 (26)	Ограниченные элементы для 39 плацентарных млекопитающих EPO LOW COVERAGE
6:55261304-58859012 (53)	<i>NLRP13, ZNF787, ZNF667, GALP, ZNF583, ZNF582, ZNF71, ZNF134, PEG3, ZSCAN4, ZNF671, ZNF814, A1BG, HES4, B3GALT6, MIB2, CCNL2, TMEM52, GABRD, PANK4, PLCH2, FAM213B</i>
15:32168704-33013541 (27)	Ограниченные элементы для 39 плацентарных млекопитающих EPO LOW COVERAGE
15:105914299-113466125 (139)	<i>INPP1, GLS, MYO1B, NABP1, STAT4, TMEFF2, MFSD6, NAB1, STAT4, TMEFF2, STK17B, SF3B1, PLCL1, COQ10B, HSPE1, GTF3C3, HSPD1, MOBKL3, RFTN2, ANKRD44, BOLL</i>

У свиней КБ2 гены, локализованные в наиболее значимых HBD, выполняют биологические функции, связанные с процессами двухцепочного разрыва ДНК, с ядерно-цитоплазматическим транспортом, в том числе с сигналом ядерной локализации (англ. Nuclear localization signal, NLS) и расщеплением белков с помощью эндоплазматического ретикулума (англ. Endoplasmic-reticulum-associated degradation, ERAD). В аспекте адаптации к интенсивным технологиям производства можно предположить, что обогащение сегментов HBD генами, связанными с процессами двухцепочного разрыва ДНК, спровоцировано экзогенными агентами, такими как радиация и некоторые химические вещества, в связи с преобладанием закрытой системы содержания, использованием различных диагностических приборов, концентрированных кормовых смесей, что вполне может выступать в качестве повышенной нагрузки ионизирующей радиации. Концентрация в сегментах HBD генов, связанных с функционированием ядерно-цитоплазматического транспорта в ядре и пути ERAD, вполне обусловлена их ролью в физиологических функциях комплекса гормона роста свиньи (GH) и рецептора гормона роста (GHR).

Функции гормона роста и его рецептора очень обширны. Связывание гормона роста с его рецепторами активирует каскад сигнальных событий, который вносит значительный вклад, в том числе в признаки, характеризующие рост и развитие животных. Представляет интерес то, что, хотя процессы ядерно-цитоплазматического транспорта и ERAD-путь могут положительно влиять на продуктивные качества свиней, имеются данные об их возможной связи с репликацией цирковируса свиней типа 2 (Shuai et al., 2011) и вирусом репродуктивно-респираторного синдрома свиней (Lee et al., 2006). В данном контексте крайне важно отметить, что несмотря на высокий рост продуктивных показателей, современное коммерческое поголовье страдает от разрушительных последствий инфекционных заболеваний, т.к. за последнее время в мире возникло множество заболеваний, таких как репродуктивно-респираторный

синдром свиней (1991 г.), инфекция, вызванная цирковирусом свиней типа 2 (1994 г.), эпидемическая диарея свиней и др.

У свиней группы КБ3, в отличие от КБ2, в сегментах HVD выявлены гены, кодирующие белки классов фосфолипаз, киназ и мембраносвязанных сигнальных молекул. Эти белки участвуют в биосинтезе холестерина, сигнальном пути рецептора инсулиноподобного фактора роста (IGF), метаболизме ацил-КоА, а также в транспорте, активации и синтезе жирных кислот. Несмотря на то, что в целом считается, что программы селекции привели к снижению упитанности и содержанию внутримышечного жира у современного поголовья свиней, в группе КБ3 можно предположить, что одним из акцентов селекционной работы являлось повышение качества мяса. Здесь также следует отметить, что свиньи, представляющие группу КБ3, действительно отличаются высоким качеством мяса и высоким уровнем внутримышечного жира.

3.3 Идентификация подписей положительного отбора в геномах свиней породы крупная белая из различных селекционных центров

Для идентификации признаков отбора в популяции свиней породы крупная белая были использованы различные подходы к набору данных, включающему около 60 тысяч SNP-маркеров (после этапов фильтрации для анализа оставили 32 тысяч SNP). На первом этапе подписи отбора были установлены в каждой исследуемой популяции: свиней отечественной селекции (группа КБ1), импортной селекции двух центров (группа КБ2 и группа КБ3), с использованием трех методов расчета на основе паттернов гомозиготности гаплотипов iHS, nSI и iH12. В качестве значимых подписей рассматривали значения, показавшие критерий, равный 1 ($\text{crit} = 1$).

По результатам использования трех методов у свиней КБ1 определено 2739 уникальных вариантов, в том числе методом iHS – 2021, nSI – 1068 и iH12 – 431 соответственно. В группе КБ1 подписи отбора пересекаются с локусами, ответственными за упитанность и качество мяса, а также показатели здоровья. У свиней КБ2 определено 966 уникальных вариантов подписей отбора (iHS – 863, nSI – 316 и iH12 – 26). В группе КБ2 в генах под действием отбора оказались локусы, ответственные за мясо и тушу свиней, здоровье и продуктивность, но чуть в большей степени в группе КБ2, относительно КБ1, под отбором оказались признаки экстерьера и репродуктивности. У свиней КБ3 установлено 2778 уникальных вариантов подписей отбора (his – 1124, nSI – 563 и iH12 – 1623). В группе КБ3 в целом наблюдалась картина, аналогичная группе КБ2. Гены под действием отбора пересекаются с локусами, ответственными за мясо и тушу свиней, здоровье и продуктивность.

Общие варианты отбора для всех анализируемых групп свиней породы крупная белая определены в SSC4, 7, 8 и 14. Кроме того, 6 общих вариантов локализованы в SSC9, 2 из которых – в интронах гена *TENM4* (таблица 4).

Таблица 4 – Общие подписи отбора у свиней трех групп

SSC	Позиция	Номер rs	Вариант в геноме	Ген
14	25989860	rs80973471	Межгенный вариант	–
4	16486634	rs80809199	Интронный вариант	<i>ZHX2</i>
7	106025579	rs81295294	Межгенный вариант	–
8	18271949	rs81406415	Не кодирующий транскрипт экзонный вариант	<i>lncRNA</i>
9	13492371	rs81257618	Интронный вариант	<i>TENM4</i>
9	13580920	rs330708382	Интронный вариант	<i>TENM4</i>
9	13846118	rs81477175	Не кодирующий транскрипт экзонный вариант	<i>lncRNA</i>
9	13871291	rs81420099	Межгенный вариант	–
9	13961806	rs81420072	Вышестоящий генный вариант	<i>lncRNA</i>
9	14014071	rs3472119487	Нижестоящий генный вариант	<i>lncRNA</i>

Ген *TENM4* кодирует белок тенерин 4. Тенерины (Teneurins) являются эволюционно консервативными молекулами адгезии, их экспрессия отмечена во многих тканях, но на самом высоком уровне – в головном мозге. У свиней *TENM4* является критически важным онтогенетическим геном, т.к. участвует в развитии сомитов и, следовательно, скелета. Вследствие того, что ген *TENM4* оказался под давлением отбора во всех группах, можно предположить, что это связано с его ролью в развитии позвоночного столба у свиней, а также возможным влиянием на другие селекционно-значимые признаки.

Общие подписи отбора только для современного поголовья свиней породы крупная белая представлены 53 вариантами, из них 21 вариант локализован в межгенных областях, 4 – в длинных некодирующих РНК (*lncRNA*) и 31 вариант – в 26 генах (рисунок 7). В первой хромосоме определены три варианта отбора, но все они локализованы в межгенной области, что не позволяет сделать предположений об их возможном функциональном потенциале. Однако один из вариантов, rs81314465, в работе Vázquez-Gómez et al. (2020) был определен как вариант материнского феномена искажения передачи информации (Transmission ratio distortion, TRD). Предполагают, что варианты материнского феномена могут влиять на эмбриогенез у свиней, но на сегодняшний день информации по этому феномену и его связи с какими-либо фенотипами у свиней крайне мало. В раннем эмбриональном развитии также участвуют варианты положительного отбора, локализованные в третьей хромосоме.

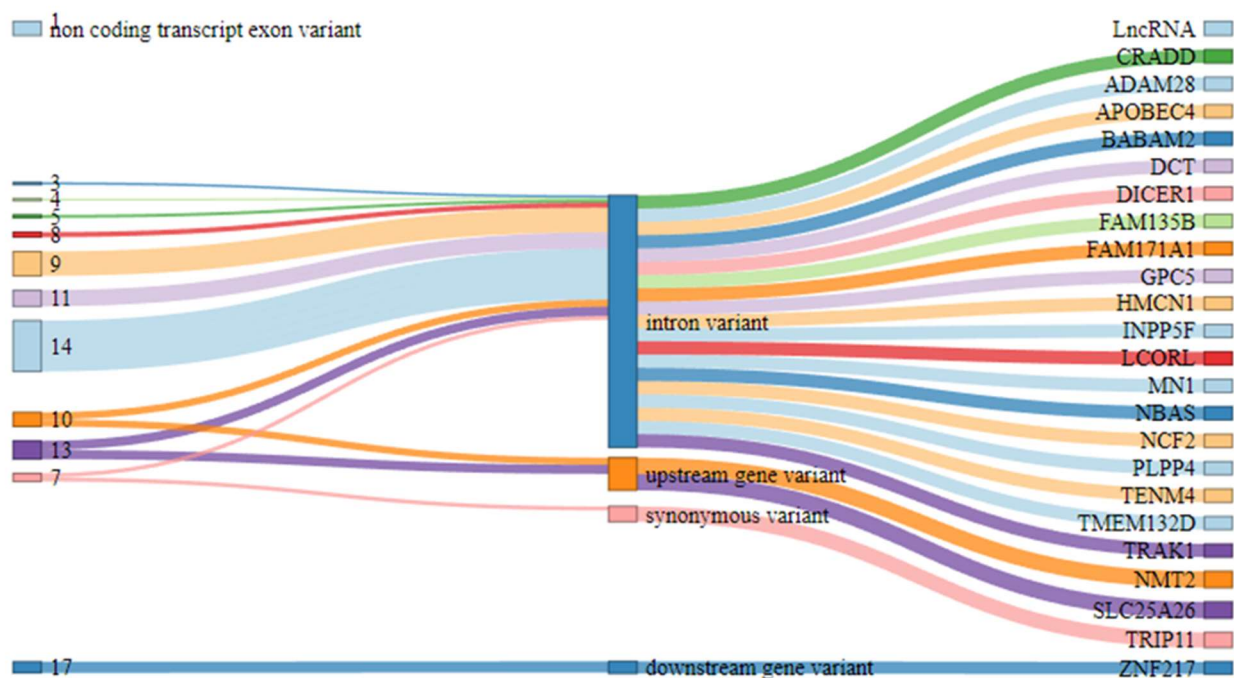


Рисунок 7 – Подписи отбора, локализованные в генах, идентифицированные у свиней импортной селекции

Со значительными изменениями мясной продуктивности и внутримышечного жира, произошедшими у современного коммерческого поголовья свиней, могут быть связаны варианты, локализованные в 5-й и 10-й хромосомах. Варианты положительного отбора, повлиявшие на скорость роста, длину и массу тела, определены в 8-й и 11-й хромосомах свиней. В 9-й хромосоме варианты отбора у современных свиней породы крупная белая в определенной степени связаны с иммунной системой, что представляет значительный интерес для дальнейших исследований, направленных на изучение генетической архитектуры устойчивости свиней к патогенам различного происхождения. В 13-й хромосоме положительный отбор затронул гены, ответственные за митохондриальный транспорт, что, вероятно, обусловлено необходимостью в обеспечении организма ресурсами для интенсивного роста и высоких репродуктивных показателей. В 7-й хромосоме варианты отбора принимают участие в регуляции метаболизма липидов и репродуктивной системы. В 14-й хромосоме варианты отбора задействованы в различных процессах, связанных с адаптационными изменениями экстерьера, сердечно-сосудистой системы, гематологических и клинико-биохимических параметров крови, репродуктивной системы и др.

В целом полученные результаты свидетельствуют о том, что варианты положительного отбора связаны со значительными физиологическими изменениями, произошедшими в современном поголовье свиней в связи с адаптацией к условиям кормления, содержания и требованиям к продуктивным показателям, обусловленным интенсивными технологиями, реализуемыми на сегодняшний день в ведущих

селекционно-генетических центрах. На основании полученных данных выявленные варианты положительного отбора могут рассматриваться в качестве потенциальных маркеров селекционно-значимых признаков, включая устойчивость к стрессовым факторам и различным заболеваниям.

3.4 Поиск значимых предикторов, ассоциированных с плодовитостью свиноматок, с применением методов машинного обучения

Поиск значимых предикторов для количества поросят при рождении был реализован с помощью 6 моделей, из них 4 алгоритма машинного обучения: глубокое обучение (Deep Learning), случайный лес (Random Forest (RF)), машина градиентного спуска (Gradient Boosting Machine (GBM)), экстремальный градиентный спуск (Extremal Gradient Boosting (XGBoost)), регрессия «лассо» (Lasso-regression (LR)) гребневая регрессия (Ridge-regression (RR)). Для оценки качества модели рассчитывали показатели метрики MAE, R2 (таблица 5). Глубокое обучение и гребневая регрессия превзошли другие модели по всем критериям оценки. Коэффициенты детерминации для глубокого обучения и гребневой регрессии составили 0,939 и 0,975 соответственно.

Таблица 5 – Метрики моделей

Метрики	Тренировочный набор		Тестовый набор
	среднее значение	стандарт. отклонение	среднее значение
Глубокое обучение			
mae	0,590020	0,071773	0,4750844
r ²	0,939155	0,011273	0,9618232
Экстремальный градиентный спуск			
mae	1,510068	0,071681	1,871898
r ²	0,604665	0,031571	0,3989991
Машина градиентного спуска			
mae	1,884282	0,075253	1,863851
r ²	0,393505	0,050699	0,4038164
Случайный лес			
mae	1,983754	0,106916	0,7669743
r ²	0,343705	0,041655	0,3938947
Лассо-регрессия (Lasso-regression (LR))			
mae	0,392218	0,018287	0,7721181
r ²	0,935614	0,008178	0,9015521
Гребневая регрессия (Ridge-regression (RR))			
mae	0,383172	0,017936	0,7669743
r ²	0,974526	0,002510	0,9717731

Генетические варианты, установленные в моделях DL и RR, были ранжированы по значениям важности. Результаты анализа обогащения генов, в которых локализованы значимые генетические варианты, определенные по модели DL,

показали в качестве основных путей деградацию лизина, сигнальные пути TGF- β и сигнальные пути, регулирующие плюрипотентность клеток.

Для определения статистически значимых предикторов, ассоциированных с количеством поросят при рождении, на основе данных полногеномного ассоциативного анализа (GWAS) и идентифицированных подписей отбора были выбраны общие варианты, определенные как минимум одной из моделей DL или RR и одним из методов подписей селекции, iHS, nSL и HH12 в этой популяции (рисунок 8).

Варианты локализованы в генах *NBAS*, *LRRC8C*, *LRFN3*, *BCL11B*, *CC2D2A*, *KCNK2*, *ESRRG*, *CDNF*, *NEBL*, *CCDC3*, *TNFRSF19*, *FAM117A*, *ULK4*, *SGCZ*, *TUSC3*, *GHRHR*, *MINDY4* и *INMT*. Эти гены играют значительную роль в период супоросности (беременности) свиноматок и принимают участие в росте и развитии фолликулов (*SGCZ*), раннем эмбриональном развитии (*CCDC3*, *LRRC8C*, *LRFN3*, *TNFRSF19*), рецептивности эндометрия и имплантации (*NEBL*), плацентации и эмбриональном развитии (*ESRRG*, *GHRHR*, *TUSC3*, *NBAS*).

Диаграмма Венна количества выявленных SNP методами GWAS и подписями отбора

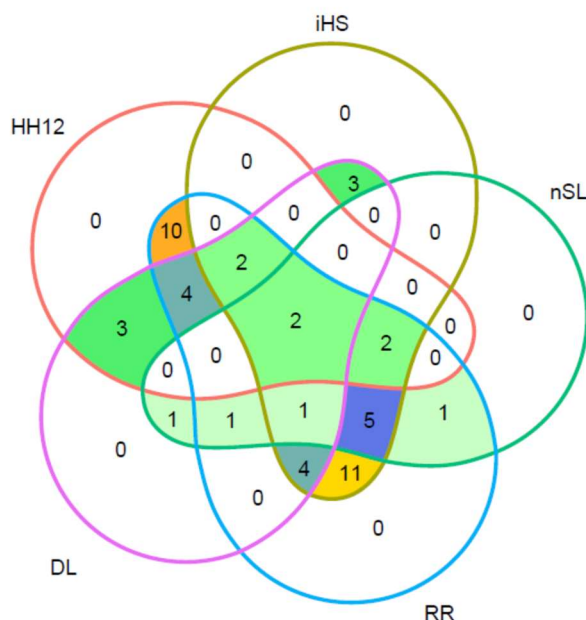


Рисунок 8 – Варианты, определенные по моделям DL и RR и методами идентификации подписей селекции

Интересно отметить, что целый ряд генов связан с риском нарушений развития нервной системы и головного мозга, что может привести к задержке эмбрионального развития и различным аномалиям (*BCL11B*, *CDNF*, *ULK4*, *CC2D2A*, *KCNK2*). Результаты исследования показали важные биологические пути, связанные с плодовитостью свиноматок, а также данные о ключевых генетических вариантах, которые могут выступать как потенциальные маркеры фертильности и использоваться в будущих программах селекции свиней.

3.5 Поиск значимых предикторов, ассоциированных с селекционно-значимыми признаками свиней, с помощью CNV-GWAS

С помощью алгоритмов PennCNV, QuantiSNP и GADA в исследуемом поголовье свиней породы крупная белая определили 3098 CNV: 429, 810 и 1859 CNV соответственно. CNV размером менее 50 Кб исключали, а для анализа отбирали только CNV, идентифицированные как минимум двумя алгоритмами. В результате исследований 775 CNV использовали для дальнейшего анализа. Длина делеций варьировала от 66 п.н. до 992 т.п.н. Длина дубликатов составляла 28,9–4,9 Мб. В среднем в геноме одного животного определено около 8 CNV. Максимальное количество CNV на одно животное в исследуемой выборке составило 52, минимальное – 2 CNV (рисунок 9).

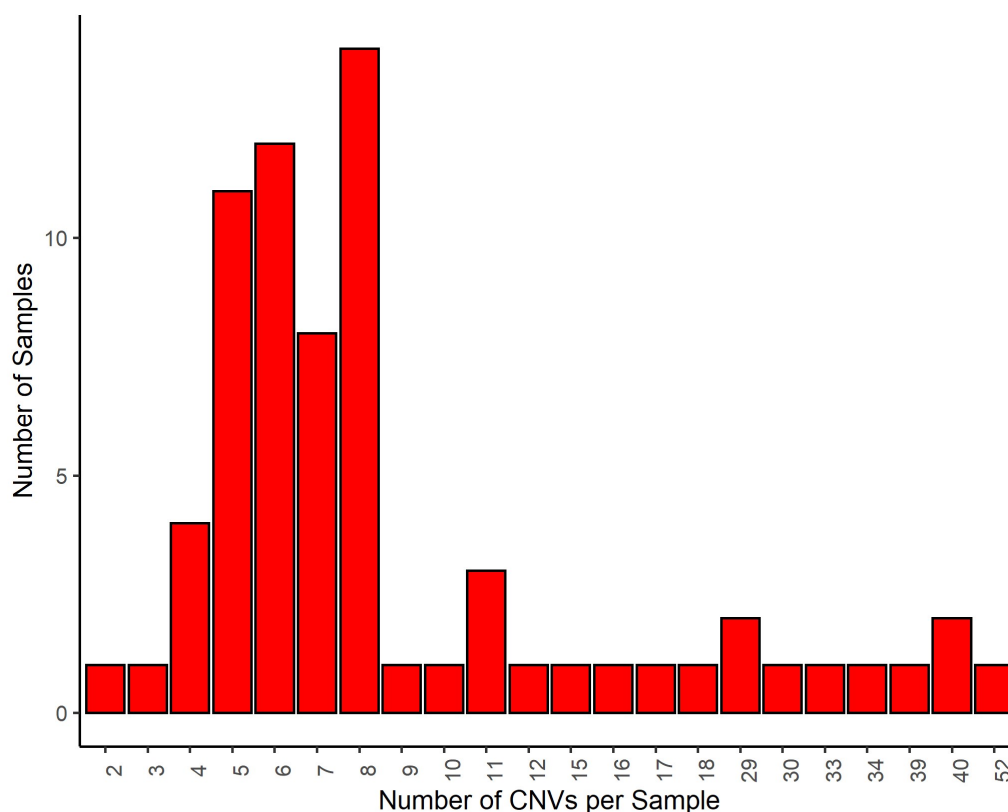


Рисунок 9 – Распределение количества CNV на образец

Объединение CNV в регионы привело к получению 246 CNVR, из которых 112 CNVR встречаются у двух или более свиней. Размер таких участков в исследованной выборке варьирует от 67 п.н. до 8,2 Мб. Геномный анализ выявил следующие вариации числа копий (CNV): потери копий (Loss) обнаружены в 12 регионах, приросты копий (Gain) – в 91 регионе, тогда как 9 регионов демонстрируют смешанный тип вариаций (Mixed), включающий как потери, так и приросты копий (рисунок 10).

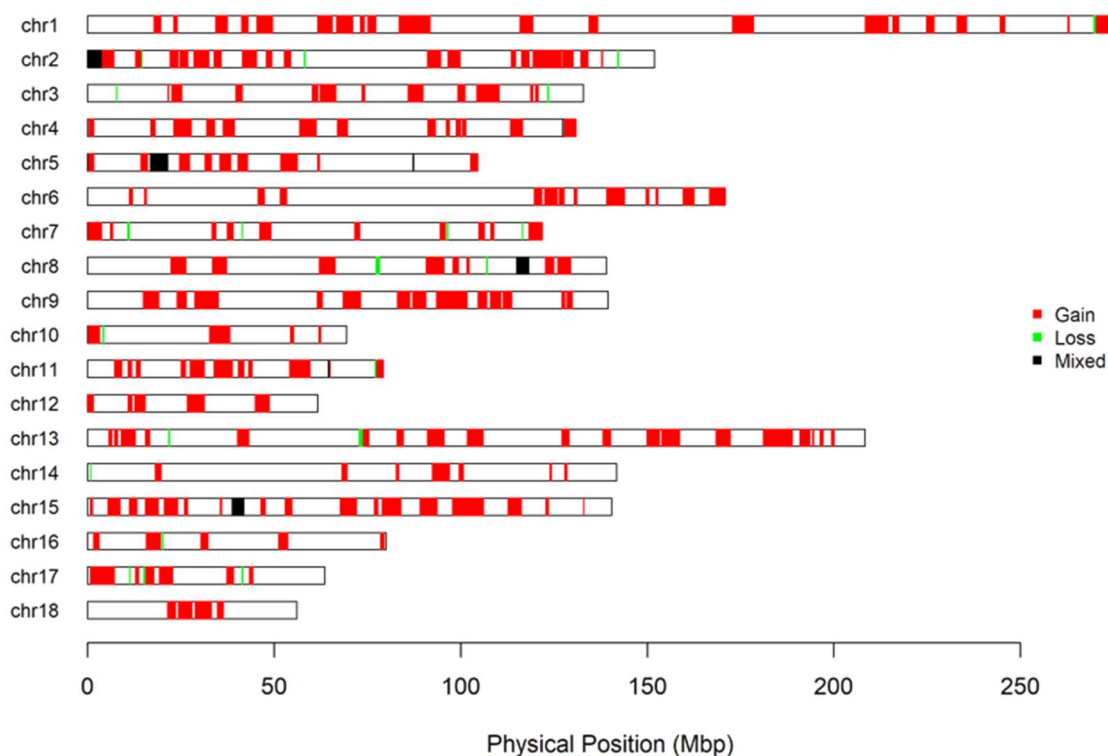


Рисунок 10 – Полногеномная карта CNVR у свиней породы крупная белая

Сопоставление локализации CNVR с локусами количественных признаков показало, что большинство регионов (около 57,59%) пересекаются с локусами, ответственными за мясные качества и тушу свиньи; 17,04% – с устойчивостью к болезням или чувствительностью; 12,04% – со свойствами, полезными для производственного цикла; 5,88% – с размножением; 5,42% – с внешним видом; 2,03 – с другими признаками.

Для проведения полногеномного ассоциативного анализа были выбраны признаки, представляющие наибольшую селекционную ценность в аспекте мясной и откормочной продуктивности свиней: скороспелость (возраст достижения 100 кг, дни), толщина шпика (мм), площадь постного мяса (прижизненная оценка, %). Кроме того, в анализ был включен признак, связанный с дефектами конечностей свиней (шишки/наросты в области скакательного сустава на задних конечностях), являющийся причиной значительных экономических потерь.

По результатам CNV-GWAS определили 3 CNV на уровне статистической значимости (рисунок 11). CNV локализованы в хромосомах 1 (1:41300464-42881624), 3 (3:123404335-123460528) и 6 (6:142358005-143865882). CNV 6:142358005-143865882 охватывает два гена: *CTH* и *SRSF11*. Ген *CTH* кодирует цистатионин- γ -лиазу, фермент, участвующий в биосинтезе цистеина через цистатионин. Мутации в этом гене способствуют нарушению метаболизма аминокислот.

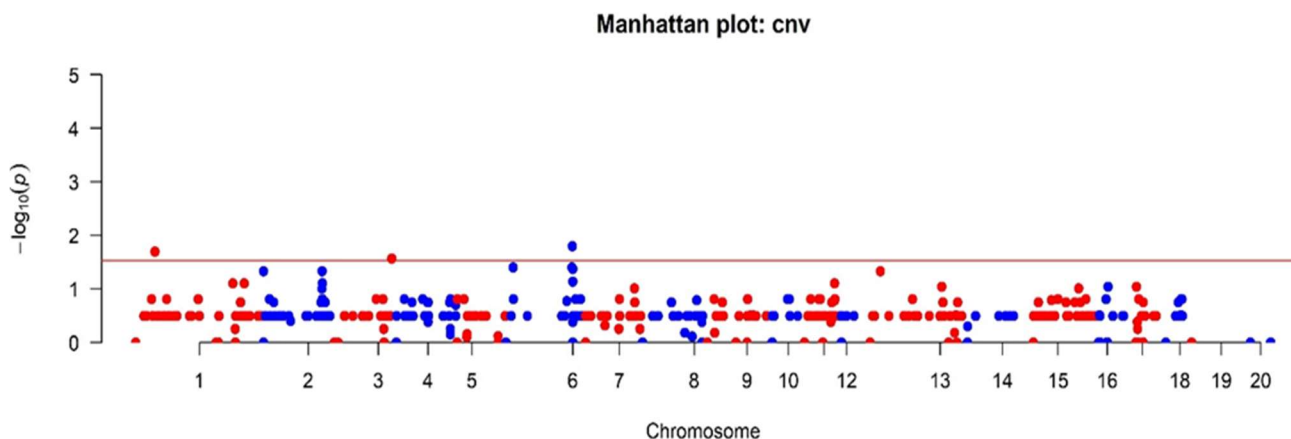


Рисунок 11 – Полногеномные ассоциации CNV (на уровне зонда) с дефектами конечностей свиней (FDR)

Ген *SRSF11* кодирует фактор сплайсинга пре-мРНК и является богатым аргинином-серином белком семейства SR. CNV, воздействующий на этот белок, может влиять на процесс адипогенеза, поскольку анализ транскриптома *SRSF11* у свиней показал, что этот белок предположительно может быть вовлечен в процесс дифференцировки адипоцитов в мышцах во время клеточного цикла. CNV 1:41300464-42881624 перекрывается с геном *MAN1A1*, который кодирует белок, участвующий в гликозилировании белка. CNV 3:123404335–123460528 локализован в межгенной области и не охватывает какие-либо гены. Однако на сегодняшний день становится все более очевидным, что в межгенных областях локализованы механизмы, играющие значимую роль в регуляции экспрессии генов. В связи с этим были рассмотрены близлежащие гены от CNV 3:123404335-123460528. В контексте исследования интерес вызывает ген *LPIN1* (3: 124940548-125074123), принадлежащий к семейству белков LPIN. У свиней основными функциональными путями, в которых задействован LPIN1, являются метаболизм глицеролипидов, переваривание и всасывание жиров, метаболические пути, метаболизм глицерофосфолипидов и сигнальный путь фосфолипазы D.

По результатам ассоциативного анализа между CNV и процентом мясности определен только один CNV (SSC10:3605230), соответствующий уровню статической значимости. Варианты расчетов анализа толщины шпика, где использовались поправки на множественное тестирование методом FDR, не показали значимых CNV, ассоциированных с толщиной шпика. По результатам анализа между CNV и скороспелостью выявлен только один CNV (SSC10:3605230), также ранее показавший связь и с процентом мясности.

CNV (SSC10:3605230) перекрывается с геном *BRINP3*, кодирующим нейроспецифический белок 3, индуцируемый костным морфогенетическим белком/ретиноевой кислотой. Таким образом, CNV-GWAS показал 4 CNV, ассоциированные с исследуемыми фенотипами продуктивности и дефектами

конечностей. Из них 3 CNV, локализованные в позициях: 6:142358005-143865882 (перекрывает гены *CTH* и *SRSF11*); 1:41300464-42881624 (перекрывает ген *MAN1A1*) и 3:123404335-123460528 (близлежащий ген *LPINI*), могут выступать как генетические маркеры предрасположенности к дефектам конечностей свиней. CNV локализованы в генах либо располагаются вблизи генов, ответственных за метаболизм аминокислот, жирных кислот, глицеролипидов и глицерофосфолипидов, тем самым связанных с иммунным ответом, функциями печени, содержанием внутримышечного жира и упитанностью животных. Результаты исследования соответствуют данным, представленным в работе Вакоев et al. (2021), где указывается, что предрасположенность к образованию шишек и наростов в области скакательного сустава конечностей может быть обусловлена генетическими вариантами, связанными с функциями печени и почек, гематологическими показателями, восприимчивостью к инфекциям, а также составом липидов и жирных кислот. Один из идентифицированных CNV (10:3605230) затрагивает ген *BRINP3*, регулирующий клеточные функции и влияющий на процессы роста. CNV (10:3605230) может быть перспективным генетическим маркером продуктивных признаков, связанных со скоростью ростом и мясностью свиней. В целом функциональный анализ позволил предположить, что идентифицированные CNV могут выступать в качестве потенциальных предикторов фенотипов признаков продуктивности и дефектов конечностей свиней.

3.6 Алгоритм визуализации и классификации областей гомозиготности (РОН)

Одним из возможных подходов для исследования РОН является создание карт РОН (визуализация геномной мозаики РОН) и использование инструментов для визуализации изображений. Так, сверточные нейронные сети (CNN) добиваются значительных прорывов в области компьютерного зрения, включая исследования генетической архитектуры. По сравнению с традиционными методами обработки изображений, CNN имеют более высокую сложность и могут извлекать информацию об изображении более высокого уровня. Вместо изучения сводной генетической статистики можно рассматривать изображения полученных результатов в качестве входных данных. Разумеется, эти данные будут чрезвычайно многомерными, однако структура CNN позволяет им неявно выполнять уменьшение размерности, одновременно фиксируя существенные структуры во входных данных, обеспечивая точную и эффективную классификацию.

Опираясь на уникальность ландшафта РОН в геноме и инструментарий глубокого обучения в области анализа изображений, разработали алгоритм для создания карт областей гомозиготности, позволяющий обнаружить и оценить уникальную архитектуру РОН как для отдельной особи, так и для определенной популяции. Предлагаемая стратегия состоит из трех основных этапов: обнаружение

РОН, их визуализация с построением карты РОН для каждой особи, классификация с использованием глубоких сверточных нейронных сетей (CNN), распознавание и дифференциация образцов в соответствии с заданными условиями.

Для обнаружения РОН существуют различные программы, включая RZooROH и Plink. Каждая имеет свои собственные методы для определения этих регионов. В программе RZooROH используется байесовский подход полного вероятностного моделирования для определения областей гомозиготности, что позволяет определять РОН, которые являются гомозиготными по происхождению (HBD – унаследованные области аутозиготности) от гомозиготных по состоянию (HBS). В программе Plink используется метод сканирования генома путем скользящего окна. Окно сдвигается по геному, и, если определенная доля последовательных SNP являются гомозиготными (выше предопределенного порога, обычно 90–100%), этот регион считается РОН. Идентификация РОН посредством Plink является одним из наиболее широко используемых инструментов для геномного анализа, т.к. быстро и достаточно эффективно позволяет обнаруживать РОН и обрабатывать различные типы геномных данных. Минусом этой программы является то, что она не различает гомозиготность по происхождению от гомозиготности по состоянию, т.е. идентифицирует любые длинные гомозиготные участки, независимо от того, произошли ли они от общего предка. В связи с тем, что каждая программа имеет свои плюсы и минусы, мы считаем, что для исследования демографической истории и генетического разнообразия, в частности свиней, более предпочтительна программа RZooROH.

В силу простоты использования и скорости для определения РОН с целью построения карт гомозиготности был протестирован алгоритм на основе программы Plink. В программе предусмотрены пороговые значения, настройки которых крайне важны. В нашем варианте настройки выглядят следующим образом. Минимальное количество SNP в гомозиготном участке (`homozyg-snp 50`) регулирует минимальное количество последовательных гомозиготных SNP, необходимых для определения областей гомозиготности. Учитывали 50 последовательных SNP, что позволило отфильтровать короткие фрагменты, скорее всего, не связанные с инбридингом или отбором, минимальная длина гомозиготного сегмента (`homozyg-kb 1000`) (регулирует минимальную длину гомозиготного сегмента). Учитывали сегменты длиной не менее 1000 Кб, что позволяет контролировать равномерность покрытия SNP при генотипировании); минимальная плотность (`homozyg-density 100`) (учитывает количество гомозиготных SNP на единицу геномного расстояния, например, 100 SNP на 1 Мб). Более высокое значение гомозиготной плотности указывает на то, что в анализ будут включены только регионы с более высокой концентрацией гомозиготных SNP, что помогает отфильтровывать менее надежные или более короткие сегменты); длина между двумя SNP из разных сегментов (`homozyg-gap 500`) представляет максимально допустимое расстояние (500 Кб) между последовательными гомозиготными SNP в пределах потенциального РОН. Если разрыв между двумя

гомозиготными SNP превышает этот порог, сегмент считается прерванным и больше не считается частью непрерывного ROH. Этот параметр помогает сбалансировать обнаружение длинных непрерывных регионов ROH, исключая сегменты с большими гетерозиготными разрывами; количество SNP в скользящем окне (homozyg-window-snp 50) (определяет минимальное количество последовательных гомозиготных SNP, необходимых для формирования допустимого окна в заданном геномном регионе. При значении 50 это означает, что в окне должно присутствовать не менее 50 последовательных гомозиготных SNP); максимальное количество гетерозиготных SNP в скользящем окне (homozyg-window-het 1) (при значении 1 в окне разрешен только один гетерозиготный SNP, прежде чем сегмент перестает считаться действительно ROH); максимальное количество гетерозиготных SNP во всем сегменте ROH (homozyg-het 1) (при значении 1, только один гетерозиготный SNP разрешен во всем сегменте гомозиготности).

Кроме того, при использовании геномных данных, полученных при помощи чипов (полногеномного генотипирования), нет необходимости удалять варианты, находящиеся в неравновесном сцеплении (фильтр по LD), т.к. сокращение по LD показало сильное влияние на количество обнаруженных участков, сокращая результат более чем на 90% посредством удаления длинных и коротких ROH, оставляя в анализе только фрагменты средней длины.

После определения областей гомозиготности для каждого животного была построена карта. На рисунке 12 показана карта ROH для одного животного. Каждая строка на рисунке соответствует хромосоме, располагаются они сверху вниз, от 1-й до 18-й хромосомы. Сегменты ROH разделены на классы в соответствии с их длиной в Мб и представлены разными цветами. Так, сегменты менее 2 Мб показаны зеленым цветом, 2–4 Мб – красным, 4–6 Мб – синим, 6–8 Мб – бирюзовым, 8–16 Мб – оранжево-желтым и более 16 Мб – серым.

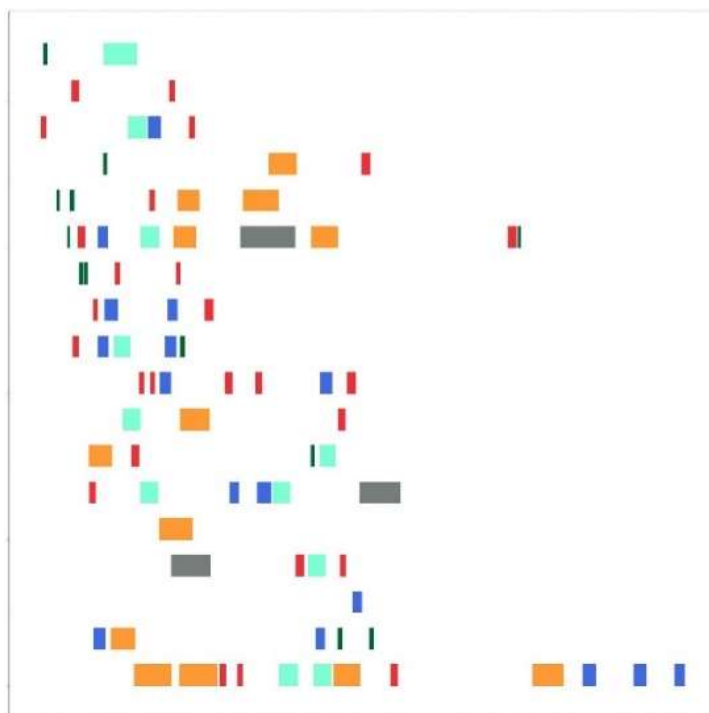


Рисунок 12 – Карта RON для одного животного

Далее применили метод сверточных нейронных сетей (CNN). Основными компонентами этой модели являются:

- подготовка данных, на этом этапе происходит преобразование изображений и загрузка их в память для случайной генерации выборок;
- подготовка набора данных, реализуется нормализация значений пикселей ($1/255$) и формируются наборы данных для обучения (тренировки модели) и валидации (оценки точности модели);
- развитие модели глубокого обучения, включает последовательность разнотипных слоев: сверточные слои с операциями предвыборки (pooling), глобальный усредняющий слой (global average pooling) и несколько полносвязных (dense) слоев. Модель компилируется с указанием функции потерь (loss function), оптимизатора с фиксированным коэффициентом обучения (learning rate) и метрик оценки точности;
- обучение модели, происходит на основе загруженных данных, на этом этапе задается количество эпох (количество запусков обучения модели), т.к. недостаточное количество эпох может привести к плохому обучению модели, а чрезмерное количество эпох – к переобучению. На этом же этапе устанавливают размер батча (части датасета), для несбалансированных классов в данных используют класс важности. Процесс обучения модели по породам представлен на рисунке 13;
- сохранение модели, происходит после обучения модели для дальнейшего использования или анализа;

- прогнозирование и оценка модели, генерируются метки для анализируемых данных, на основе метрик проводится оценка модели.

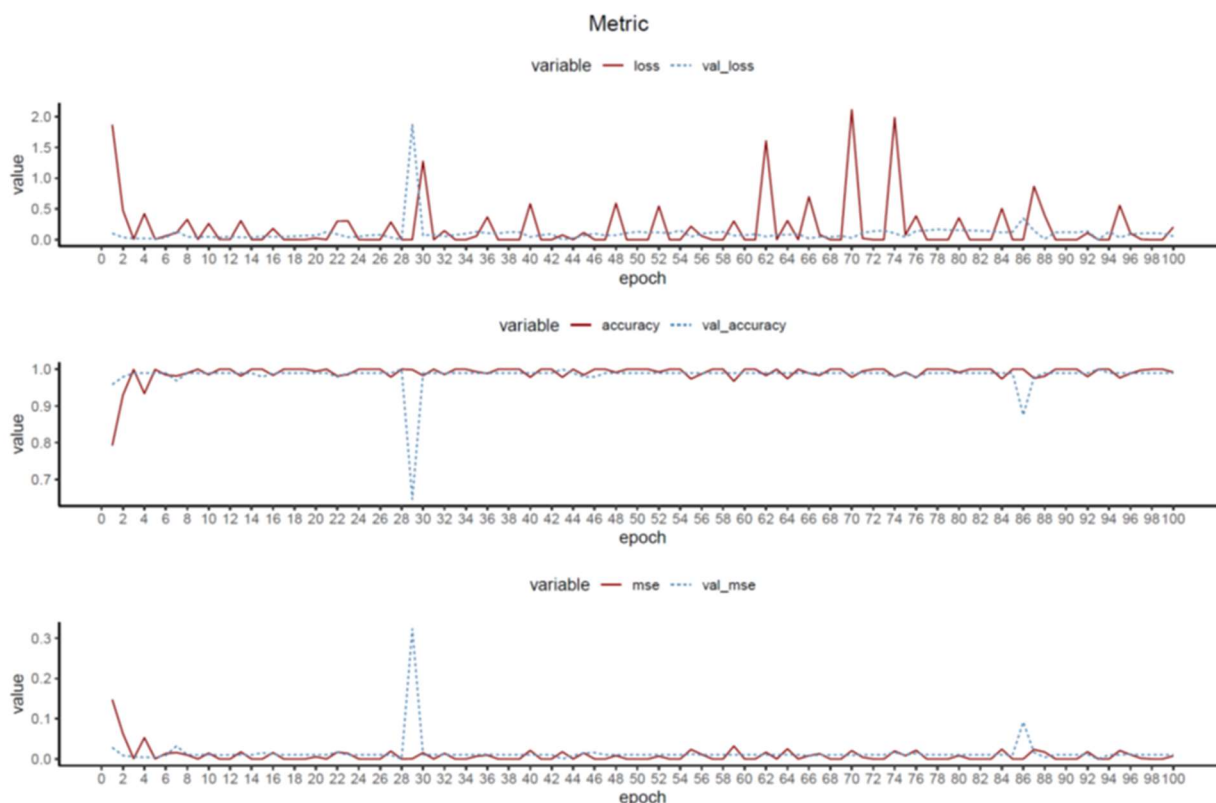


Рисунок 13 – Процесс обучения модели и валидации модели на 100 эпохах. В верхней части рисунка отражена динамика функции потерь (loss), в центре – динамика точности классификации (accuracy), в нижней части рисунка – динамика количественной меры расхождения предсказаний (mse) на обучающих данных и на валидационных данных

Преобразование генетических данных в изображения и их анализ с использованием CNN позволяет учитывать сложные паттерны гомозиготности, что недоступно традиционным методам. Этот метод может быть применен не только в сельском хозяйстве, но и в медицине для изучения связи между ROH и заболеваниями. В будущем мы планируем расширить применение этого подхода для анализа других типов генетических данных и их связи с фенотипами.

В результате модель для классификации карт ROH относительно межпородной дифференциации свиней породы крупная белая и дюрок имела все метрики, такие как точность, чувствительность, специфичность, положительная и отрицательная прогностическая ценность, равные 1,0 (Вакоеv et al., 2025). Это свидетельствует о безошибочной работе модели и 100%-ному определению породы свиней по геномному ландшафту сегментов ROH (таблица 6).

Таблица 6 – Матрица путаницы для модели межпородной дифференциации по картам ROH

Прогнозные значения	Наблюдаемые значения	
	LW	DR
LW	30	0
DR	0	37
Точность	Чувствительность	Специфичность
1,0 (95% CI (0.95))	1,0	1,0
Прогностическая ценность	Положительная	Отрицательная
	1,0	1,0

Разработанная модель для классификации на основе сверточных нейронных сетей с целью дифференциации наличия/отсутствия дефектов у свиней продемонстрировала умеренную точность (~78%), что свидетельствует о ее способности различать животных с дефектами и без них на основе карт ROH. Чувствительность и специфичность модели находятся на приемлемом уровне, что свидетельствует о том, что она может выявлять как положительные, так и отрицательные классы. Следует учитывать, что рассматриваемые дефекты, представляющие собой небольшие шишки, наросты в области скакательного сустава на задних конечностях свиней, – это сложный многофакторный процесс, на который оказывают влияние как генетика, так и внешние условия. Соответственно, невысокая положительная прогностическая ценность (78%) указывает на наличие дополнительных факторов, влияющих на фенотип. Это свидетельствует о том, что генетическая архитектура сложных признаков имеет сложные взаимодействия с окружающей средой, что следует учитывать при анализе связей между генетической и фенотипической изменчивостью.

Полученные результаты демонстрируют возможность использования карт ROH для исследования связей между геномом и фенотипами сложных признаков животных, однако модели для этих задач требуют дальнейшей настройки и оптимизации. Создание карт ROH с использованием данных полногеномного генотипирования и их дальнейший анализ с помощью глубоких нейронных сетей (CNN) могут служить важным инструментом для анализа связи между генетическими изменениями и фенотипами. Это открывает новые перспективы в исследованиях связи между ландшафтной мозаикой ROH и сложными признаками как у животных, так и у человека.

ЗАКЛЮЧЕНИЕ

1. Проведен анализ генетической изменчивости и уровня аутозиготности диких кабанов и свиней пород крупная белая, ландрас и дюрок. Коэффициенты геномного инбридинга у диких кабанов и свиней в среднем составили 0,23–0,29. Наибольший максимум геномного инбридинга, равный 0,6, зафиксирован у диких кабанов. При этом максимальное значение инбридинга у свиней коммерческих пород составило 0,3. Показано, что 0,08 доли генома диких кабанов покрыта сегментами HBD, предположительно происходящими от предков, живших около 206 лет назад. Дикие кабаны отличаются наибольшей вариабельностью HBD. В областях ключевых HBD диких кабанов значимое обогащение показали признаки, связанные с параметрами крови. У коммерческих пород свиней (крупная белая, ландрас, дюрок) около 0,07–0,08 доли генома покрыта сегментами HBD, предположительно происходящими от предков, живших около 64 лет назад. Анализ сегментов HBD показал, что они отражают основные селекционные стратегии, направленные на формирование коммерческого поголовья свиней.

2. Проведен анализ генетической изменчивости и уровня аутозиготности свиней отечественной и импортной селекции и выявлены значительные различия в генетической структуре и функциональной значимости сегментов HBD. У свиней КБ1 доля аутозиготных сегментов составила около 0,24 доли генома, при этом около 0,1 доли генома относилось к классу $R_k = 256$. У свиней КБ2 и КБ3 общая доля аутозиготных сегментов составляла около 0,25–0,29 доли генома, при этом около 0,1 доли генома относится к классу $R_k = 128$, т.е. на вклад предков около 64 лет назад. Этот период совпадает с этапом интенсификации свиноводства и формированием коммерческого поголовья с высокими производственными показателями. Анализ сегментов аутозиготности показал, что до интенсификации производства факторами отбора выступали высокий уровень иммунитета и устойчивость к различного рода заболеваниям, а у современного поголовья области гомозиготности связаны с адаптацией к условиям интенсификации и процессами, стимулирующими высокую продуктивность животных.

3. У свиней породы крупная белая определены варианты положительного отбора, связанные с ключевыми селекционно-значимыми признаками, включая эмбриогенез (хромосомы 1 и 3), мясную продуктивность и внутримышечный жир (хромосомы 5 и 10), скорость роста и массу тела (хромосомы 8 и 11), иммунную систему (хромосома 9), митохондриальный транспорт (хромосома 13), метаболизм липидов и репродуктивную систему (хромосома 7), а также адаптационные изменения экстерьера и сердечно-сосудистой системы (хромосома 14). Эти варианты могут служить кандидатами для дальнейшего изучения и использования в селекционных программах, направленных на улучшение устойчивости к стрессам, заболеваниям и повышение продуктивности.

4. Проведен полногеномный ассоциативный анализ (GWAS) для количества поросят при рождении с применением комплекса линейных моделей и алгоритмов машинного обучения. Были использованы четыре алгоритма машинного обучения (Глубокое обучение, Случайный лес, Машина градиентного спуска, Экстремальный градиентный спуск) и две линейные модели с регуляризаторами (Лассо-регрессия и Гребневая регрессия). Наибольшую эффективность продемонстрировали модели Глубокого обучения (Deep Learning) и Гребневой регрессии (Ridge-regression), которые превосходили другие методы по всем критериям оценки (MAE, MSE, RMSE, R², RMSLE). Коэффициенты детерминации для этих моделей составили 0,939 и 0,975 соответственно, что подтверждает их высокую точность в прогнозировании значимых генетических вариантов.

5. На основе результатов полногеномного ассоциативного анализа и подписей отбора установлены SNP как предикторы, контролирующие репродуктивные процессы, связанные с фенотипом количества поросят при рождении. Значимые SNP локализованы в генах, выполняющих значительную роль в период супоросности (беременности) свиноматок. Они принимают участие в росте и развитии фолликулов (*SGCZ*), раннем эмбриональном развитии (*CCDC3*, *LRRC8C*, *LRFN3*, *TNFRSF19*), рецептивности эндометрия и имплантации (NEBL), плацентации и эмбриональном развитии (*ESRRG*, *GHRHR*, *TUSC3*, *NBAS*). Ряд генов связан с риском нарушений развития нервной системы и головного мозга, что может привести к задержке эмбрионального развития и различным аномалиям (*BCL11B*, *CDNF*, *ULK4*, *CC2D2A*, *KCNK2*). Результаты представляют важные биологические пути, связанные с плодовитостью свиноматок, а также ключевые генетические варианты, которые могут выступать как потенциальные предикторы фенотипа количества поросят при рождении.

6. Определены варианты копий (CNV), ассоциированные с продуктивностью и дефектами конечностей свиней породы крупная белая. Три из них, локализованные в позициях 6:142358005-143865882 (гены *CTH* и *SRSF11*), 1:41300464-42881624 (ген *MAN1A1*) и 3:123404335-123460528 (близлежащий ген *LPIN1*), могут служить генетическими маркерами предрасположенности к дефектам конечностей. Четвертый CNV (10:3605230) затрагивает ген *BRINP3*, который регулирует клеточные функции и процессы роста. Этот CNV может быть перспективным маркером для продуктивных признаков, таких как скорость роста и мясность свиней.

7. Разработан алгоритм визуализации и классификации областей гомозиготности (ROH) на основе данных полногеномного генотипирования с применением сверточных нейронных сетей (CNN). Модель на основе разработанного алгоритма продемонстрировала 100%-ную точность в межпородной дифференциации свиней. Для оценки генетической изменчивости, связанной с дефектами конечностей, точность модели составила 0,76. Эти результаты демонстрируют эффективность

данного подхода и подчеркивают потенциал использования карт ROH и глубоких нейронных сетей для анализа селекционно-ценных признаков свиней.

Статьи, опубликованные в изданиях, индексируемых в международных базах Web of Science и Scopus

1. Visualization of runs of homozygosity and classification using convolutional neural networks / **S. Bakoev**, Kolosova M., Romanets T., Bakoev F., Kolosov A., Romanets E., Korobeinikova A., Bakoeva I., Akhmedli V., Getmantseva L. // *Biology*. – 2025. – Vol. 14, No 4. – P. 426. – DOI 10.3390/biology1404426 (Q1) K1
2. Identifying Significant SNPs of the Total Number of Piglets Born and Their Relationship with Leg Bumps in Pigs / **S. Bakoev**, L. Getmantseva, M. Kolosova [et al.] // *Biology*. – 2024. – Vol. 13, No. 12. – P. 1034. – DOI 10.3390/biology13121034. (Q1). K1
3. Evaluation of genetic differentiation and search for candidate genes for reproductive traits in pigs / E. Romanets, **S. Bakoev**, T. Romanets [et al.] // *Animal Bioscience*. – 2024. – Vol. 37, No. 5. – P. 832-838. – DOI 10.5713/ab.23.0297. (Q1). K1
4. Генетическое разнообразие и оценка аутозиготности в популяциях дикого европейского кабана и домашних свиней / **С. Ю. Бакоев**, Т. С. Романец, А. В. Коробейникова [и др.] // *Экологическая генетика*. – 2023. – Т. 21, № 4. – С. 343-355. – DOI 10.17816/ecogen569181. (K1)
5. Genome-wide analysis of genetic diversity and artificial selection in Large White pigs in Russia / S. Bakoev, L. Getmantseva, O. Kostyunina [et al.] // *PeerJ*. – 2021. – Vol. 9. – P. e11595. – DOI 10.7717/peerj.11595. Q1. K1
6. Detection of genomic regions associated malformations in newborn piglets: A machine-learning approach / **S. Bakoev**, A. Traspov, L. Getmantseva [et al.] // *PeerJ*. – 2021. – Vol. 9. – P. e11580. – DOI 10.7717/peerj.11580. (Q1). K1
7. Genomic regions and candidate genes linked to capped hock in pig / L. Getmantseva, M. Kolosova, F. Bakoev [et al.] // *Life*. – 2021. – Vol. 11, No. 6. – P. 510. – DOI 10.3390/life11060510. (Q1). K1
8. Analysis of homozygous-by-descent (Hbd) segments for purebred and crossbred pigs in Russia / **S. Bakoev**, A. Kolosov, F. Bakoev [et al.] // *Life*. – 2021. – Vol. 11, No. 8. – P. 861. – DOI 10.3390/life11080861. (Q1). K1
9. PigLeg: Prediction of swine phenotype using machine learning / **S. Bakoev**, L. Getmantseva, O. Kostyunina [et al.] // *PeerJ*. – 2020. – Vol. 8, No. 3. – P. e8764. – DOI 10.7717/peerj.8764. (Q1). K1
10. Survey of snps associated with total number born and total number born alive in pig / **S. Bakoev**, L. Getmantseva, F. Bakoev [et al.] // *Genes*. – 2020. – Vol. 11, No. 5. – P. 491. – DOI 10.3390/genes11050491. (Q2). K1
11. PSVII-26 Detection of selection signatures in Civil and Large White pig breeds / L. V. Getmantseva, **S. Yu. Bakoev**, O. V. Kostyunina [et al.] // *Journal of Animal Science*. – 2020. – Vol. 98, No. Suppl. 4. – P. 464-465. – DOI 10.1093/jas/skaa278.808. (Q1). K1

12. PSVII-28. Program Chair Poster Pick: Effects of SMAD6 and SUGCT genes on Litter traits in pigs / L. V. Getmantseva, **S. Yu. Bakoev**, V. S. Shevtsova [et al.] // Journal of Animal Science. – 2020. – Vol. 98, No. Suppl. 4. – P. 463-464. – DOI 10.1093/jas/skaa278.807. (Q1). K1
13. Finding Predictors of Leg Defects in Pigs Using CNV-GWAS / L. Getmantseva, M. Kolosova, K. Fede [et al.] // Genes. – 2023. – Vol. 14, No. 11. – P. 2054. – DOI 10.3390/genes14112054. (Q2). K1
14. Investigation of the Genetic Architecture of Pigs Subjected to Breeding Intensification / A. Kolosov, L. Getmantseva, M. Kolosova [et al.] // Genes. – 2022. – Vol. 13, No. 2. – P. 197. – DOI 10.3390/genes13020197. (Q2). K1
15. PSIX-23. Correlation between VRTN and PPARD gene variants and the phenotypic manifestation of productive traits and the hind limbs conformation in pigs / **S. Bakoev**, L. Getmantseva, F. Bakoev [et al.] // Journal of Animal Science. – 2021. – Vol. 99, No. Suppl. 3. – P. 499-500. – DOI 10.1093/jas/skab235.880. (Q1). K1
16. Assessing the Effect of SNPs on Litter Traits in Pigs / L. V. Getmantseva, **S. Y. Bakoev**, A. Y. Kolosov [et al.] // Scientifica. – 2020. – Vol. 2020. – P. 5243689. – DOI 10.1155/2020/5243689. (Q1). K1
17. Identification of «selection signatures» in pigs and wild boars (review) / L. V. Getmantseva, A.A. Traspov, N.F. Bakoev [et al.] // Agricultural Biology. – 2020. – Vol. 55, No. 2. – P. 243-256. – DOI 10.15389/agrobiology.2020.2.243eng. [*Русскоязычная версия: Поиск "отпечатков отбора" у домашних свиней и дикого кабана (обзор) / Л. В. Гетманцева, А. А. Траспов, Н. Ф. Бакоев [и др.] // Сельскохозяйственная биология. – 2020. – Т. 55, № 2. – С. 243-256. – DOI 10.15389/agrobiology.2020.2.243rus.*]. (Q3). K1
18. Associations between PRLR/AluI gene polymorphism with reproductive, growth, and meat traits in pigs / N. V. Mihailov, L. V. Getmantseva, **S. U. Bakoev**, A. V. Usatov // Cytology and Genetics. – 2014. – Vol. 48, No. 5. – P. 323-326. – DOI 10.3103/S0095452714050053. (Q4). K2
19. Effect of polymorphisms in intron 1 of the swine POU1F1 gene on growth and reproductive traits / L. Getmantseva, A. Kolosov, M. Leonova [et al.] // Turkish Journal of Veterinary and Animal Sciences. – 2017. – Vol. 41, No. 5. – P. 643-647. – DOI 10.3906/vet-1702-77. K1.
20. Polymorphisms in obesity-related leptin gene and its association with reproductive traits of sows / L. Getmantseva, A. Kolosov, M. Leonova [et al.] // Bulgarian Journal of Agricultural Science. – 2017. – Vol. 23, No. 5. – P. 843-850. K1.
21. Polymorphisms in several porcine genes are associated with growth traits / L. Getmantseva, A. Kolosov, M. Leonova [et al.] // American Journal of Animal and Veterinary Sciences. – 2016. – Vol. 11, No. 4. – P. 136-141. – DOI 10.3844/ajavssp.2016.136.141. K1.
22. Leukemia inhibitory factor (LIF) gene polymorphism and its impact on reproductive traits of pigs / L. Leonova, L. V. Getmantseva, V. N. Vasilenko [et al.] // American Journal

of Animal and Veterinary Sciences. – 2015. – Vol. 10, No. 4. – P. 212-216. – DOI 10.3844/ajavsp.2015.212.216. K1.

23. Genetic architecture of hock joint bumps in pigs: insights from ROH and GWAS analyses / Getmantseva L., Kolosova M., Shevtsova V., Kolosov A., Bakoev F., Romanets E., Romanets T., **Бакоев С.** // *Animals*. – 2025. – Vol. 15, №8. – P. 1178. – DOI 10.3390/ani15081178 (Q1). K1.

Статьи в научных изданиях, входящих в Перечень ВАК, RSCI

24. Разработка современных методов селекции свиней в ЗАО "Племзавод-Юбилейный" / С. Н. Мамонтов, Л. В. Гетманцева, М. А. Леонова [и др.] // *Свиноводство*. – 2015. – № 5. – С. 35-37. K1.

25. **Бакоев, С. Ю.** Методы оценки инбридинга и подписей селекции сельскохозяйственных животных на основе протяженных гомозиготных областей / С. Ю. Бакоев, Л. В. Гетманцева // *Достижения науки и техники АПК*. – 2019. – Т. 33, № 11. – С. 63-68. – DOI 10.24411/0235-2451-2019-11114. K2.

26. Полиморфизм SNP rs80867243, rs81379421 и rs81236069 у свиней материнских пород / Л. В. Гетманцева, М. А. Колосова, **С. Ю. Бакоев** [и др.] // *Аграрная наука Северо-Востока*. – 2018. – № 4(65). – С. 98-102. – DOI 10.30766/2072-9081.2018.65.4.98-102. K1.

27. Породная дифференциация желательных генотипов гена PRLR у свиней / А. И. Клименко, А. Ю. Колосов, М. А. Леонова [и др.] // *Сибирский вестник сельскохозяйственной науки*. – 2017. – Т. 47, № 4(257). – С. 32-37. K2.

28. Поиск новых локусов и генов-кандидатов, влияющих на репродуктивную эффективность свиней / Е. А. Романец, Л. В. Гетманцева, А. В. Радюк [и др.] // *Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки*. – 2023. – № 10. – С. 22-26. – DOI 10.37882/2223-2982.2023.10.32. K3.

Иные публикации

29. The single and combined effect of MC4R and GH genes on productive traits of pigs / Getmantseva L. V., Leonova M. A., Usatov A. V., **Бакоев С. Ю.**, Klimenko A. I., Vasilenko V. N., Kolosov A. Yu., Shirockova N. V., Makarenko M. S. // *American Journal of Agricultural and Biological Sciences*. – 2017. – Vol. 12, № 1. – P. 28–32. – DOI 10.3844/ajabssp.2017.28.32.

30. Effects of melanocortin-4 receptor gene on growth and meat traits in PIGS raised in Russia / A. Klimenko, L. Getmantseva, Y. Kolosov [et al.] // *American Journal of Agricultural and Biological Science*. – 2014. – Vol. 9, No. 2. – P. 232-237. – DOI 10.3844/ajabssp.2014.232.237.

Патенты на изобретения

31. Патент № 2634404 С Российская Федерация, МПК А01К 67/02, С12Q 1/68. Способ оценки плодовитости свиней пород ландрас и крупная белая : № 2015139635 : заявл. 17.09.2015: опубл. 26.10.2017, Бюл. № 30 / Л. В. Гетманцева, М. А. Леонова, С. Н. Мамонтов [и др.]; патентообладатель Федеральное государственное бюджетное образовательное учреждение высшего образования "Донской государственный аграрный университет". – 9 с.; ил.

32. Патент № 2822777 C1 Российская Федерация, МПК C12N 15/11, C12Q 1/68. Способ диагностики предрасположенности свиней породы дюрок к бурситам на основе вариантов SNP rs81327419 (SSC13) в гене ROBO2: № 2023113706 : заявл. 25.05.2023 : опубл. 12.07.2024, Бюл. № 20 / Т. С. Романец, С. Ю. Бакоев, Ф. С. Бакоев [и др.]; патентообладатель Федеральное государственное бюджетное образовательное учреждение высшего образования "Донской государственный аграрный университет". – 9 с.; ил.

Свидетельства о государственной регистрации

33. Свидетельство о государственной регистрации базы данных № 2015621623 Российская Федерация. База данных аутосомных ДНК-маркеров свиней : № 2015621146 : заявл. 09.09.2015: опубл. 02.11.2015 / Л. В. Гетманцева, А. Ю. Колосов, М. А. Леонова [и др.]; правообладатель Федеральное государственное бюджетное образовательное учреждение высшего образования «Донской государственный аграрный университет».

34. Свидетельство о государственной регистрации программы для ЭВМ № 2016661414 Российская Федерация. Оценка информативности ДНК-маркеров. Метод Шеннона : № 2016618776 : заявл. 12.08.2016: опубл. 07.10.2016 / А. Ю. Колосов, Л. В. Гетманцева, М. А. Леонова [и др.]; правообладатель Федеральное государственное бюджетное образовательное учреждение высшего образования «Донской государственный аграрный университет».

35. Свидетельство о государственной регистрации программы для ЭВМ № 2017660746 Российская Федерация. Оценка моделей прогнозирования племенной ценности свиней : № 2017617918 : заявл. 03.08.2017: опубл. 25.09.2017 / А. Ю. Колосов, Л. В. Гетманцева, М. А. Леонова, С. Ю. Бакоев; правообладатель Федеральное государственное бюджетное образовательное учреждение высшего образования «Донской государственный аграрный университет».