

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)

**ДИССЕРТАЦИЯ**

на соискание ученой степени  
доктора биологических наук

Научный консультант:

д-р биол. наук, проф.

Усатов Александр Вячеславович

Ростов-на-Дону – 2025

## СОДЕРЖАНИЕ

|                                                         |    |
|---------------------------------------------------------|----|
| СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ.....           | 5  |
| ВВЕДЕНИЕ.....                                           | 7  |
| Актуальность темы исследования.....                     | 7  |
| Объект и предмет исследования.....                      | 13 |
| Цель работы .....                                       | 14 |
| Задачи исследования.....                                | 14 |
| Научная новизна и теоретическая значимость работы ..... | 15 |
| Практическая значимость работы .....                    | 16 |
| Положения, выносимые на защиту .....                    | 17 |
| Степень достоверности и апробация результатов .....     | 17 |
| Публикации по теме диссертации .....                    | 19 |
| Объем и структура работы .....                          | 19 |
| Личный вклад автора .....                               | 19 |
| Благодарности .....                                     | 19 |
| Финансовая поддержка работы .....                       | 20 |
| 1. ОБЗОР ЛИТЕРАТУРЫ.....                                | 21 |
| 1.1 Доместикация сельскохозяйственных животных.....     | 21 |
| 1.2 Подписи отбора .....                                | 26 |
| 1.3 Протяженные гомозиготные области .....              | 35 |
| 1.4 Основные методы определения R <sup>2</sup> ОН ..... | 37 |
| 1.5 Варианты числа копий (CNV) .....                    | 41 |
| 1.6 Исследования ассоциаций по всему геному (GWAS)..... | 44 |
| 1.7 Алгоритмы машинного обучения.....                   | 46 |
| 1.8 Искусственные нейронные сети.....                   | 53 |
| 2 МАТЕРИАЛЫ И МЕТОДЫ.....                               | 62 |
| 2.1 Материалы .....                                     | 62 |
| 2.2 Подготовка данных .....                             | 64 |
| 2.3 Методы.....                                         | 66 |
| 2.3.1 Анализ структуры .....                            | 67 |
| 2.3.2 Анализ генетического разнообразия.....            | 68 |

|                                                                                                                                                                          |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 2.3.3 Анализ выбора.....                                                                                                                                                 | 69  |
| 2.3.5 Определение отбора с помощью статистики гаплотипа.....                                                                                                             | 69  |
| 2.3.6 Вызов CNV .....                                                                                                                                                    | 71  |
| 3 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ .....                                                                                                                                          | 73  |
| 3.1 Анализ генетического разнообразия у представителей рода <i>Sus</i> .....                                                                                             | 73  |
| 3.2 Анализ генетического разнообразия у свиней крупной белой породы из различных селекционных центров .....                                                              | 113 |
| 3.3 Идентификация сигналов положительного отбора в геномах свиней крупной белой породы из различных селекционных центров .....                                           | 142 |
| 3.4 Поиск значимых предикторов, ассоциированных с плодовитостью свиноматок, с применением методов машинного обучения .....                                               | 167 |
| 3.5 Поиск значимых предикторов, ассоциированных с селекционно-значимыми признаками свиней с помощью CNV-GWAS.....                                                        | 190 |
| 3.6 Алгоритм визуализации и классификации областей гомозиготности (RON).....                                                                                             | 201 |
| 3.6.1 Модель межпородной дифференциации .....                                                                                                                            | 211 |
| 3.6.2 Модель дифференциации шишек на ногах свиней .....                                                                                                                  | 216 |
| ЗАКЛЮЧЕНИЕ .....                                                                                                                                                         | 225 |
| СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....                                                                                                                                   | 228 |
| Приложения .....                                                                                                                                                         | 277 |
| Приложение 1. Скрипт для расчета и визуализации <i>nsi</i> , <i>ihs</i> и <i>ihh12</i> .....                                                                             | 277 |
| Приложение 2. Скрипт для расчета и визуализации CNV .....                                                                                                                | 287 |
| Приложение 3. Скрипт для расчета RON и визуализации для модели .....                                                                                                     | 291 |
| Приложение 4. Скрипт для расчета HBD и для графиков.....                                                                                                                 | 302 |
| Приложение 5. Скрипт для модели с наличием и отсутствием шишек на ногах свиней .....                                                                                     | 307 |
| Приложение 6. Патент на изобретение «Способ оценки плодовитости свиней пород ландрас и крупная белая».....                                                               | 319 |
| Приложение 7. Патент на изобретение «Способ диагностики предрасположенности свиней породы дюрок к бурситам на основе вариантов SNP rs81327419 (SSC13) в гене ROBO2»..... | 320 |
| Приложение 8. Свидетельство о государственной регистрации программ для ЭВМ «База данных аутомных ДНК-маркеров свиней» .....                                              | 321 |

|                                                                                                                                              |     |
|----------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Приложение 9. Свидетельство о государственной регистрации программ для ЭВМ «Оценка информативности ДНК-маркеров. Метод Шенона»....           | 322 |
| Приложение 10. Свидетельство о государственной регистрации программ для ЭВМ «Оценка моделей прогнозирования племенной ценности свиней» ..... | 323 |

## СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

AIC (англ. Akaike Information Criterion) – информационный критерий Акаике;  
BIC (англ. Bayesian Information Criterion) – информационный критерий Байеса;  
CNN (англ. Convolution Neural Network) – сверточные нейронные сети;  
CNV (англ. Copy Number Variation) – вариация числа копий;  
CNVR (англ. Copy Number Variation Regions) – регионы вариаций числа копий;  
DL (англ. Deep Learning) – глубокое обучение;  
FST (англ. fixation index) – индекс фиксации;  
GAN (англ. Generative adversarial network) – генеративно-сопоставительные сети;  
GBM (англ. Gradient Boosting Machine) – машина градиентного спуска;  
GHR (англ. Growth hormone receptor) – рецептор гормона роста;  
GWAS (англ. Genome-wide Association Studies) – полногеномный поиск ассоциаций;  
HBD (англ. Homozygous-by-Descent) – сегменты аутозиготности;  
LD (англ. Linkage Disequilibrium) – неравновесное сцепление генов;  
LR (англ. Lasso-regression) – регрессия «лассо»;  
LW (англ. Large White) – порода свиней крупная белая;  
MAE (англ. Mean Absolute Error) – средняя абсолютная ошибка;  
Mb (англ. Megabase) – мегабаза, единица длины последовательности ДНК (генома);  
MSE (англ. Mean Squared Error) – средняя квадратическая ошибка;  
NA (англ. Not Available) – отсутствующие значения, представленные зарезервированным (специальным) значением в R;  
NBA (англ. Number born alive) – количество поросят, родившихся живыми;  
PCA (англ. Principal component analysis) – метод главных компонент;  
QTL (англ. Quantitative Trait Loci) – локус количественных признаков;  
RF (англ. Random Forest) – метод случайный лес;  
RMSE (англ. Root Mean Squared Error) – корень из среднеквадратической ошибки;

ROH (англ. Run of homozihosity) – пробег гомозиотности;

RR (англ. Ridge-regression) – гребневая регрессия;

SNP (англ. Single Nucleotide Polymorphism) – однонуклеотидный полиморфизм;

SSC (англ. *Sus scrofa* Chromosome) – обозначение хромосом свиньи;

TNB (англ. Total number born) – общее количество рожденных поросят;

XGBoost (англ. Extremal Gradient Boosting) – экстремальный градиентный спуск;

WB (англ. Wild Boars) – дикие кабаны;

АСТ – аспартатаминотрансфераза;

КБ – порода свиней крупная белая.

# ВВЕДЕНИЕ

## Актуальность темы исследования

Современные высокопроизводительные геномные технологии вместе с развитием биоинформатических методов позволяют проводить крупномасштабные исследования и получать беспрецедентную информацию, способствующую пониманию эволюционных процессов, молекулярно-генетических основ адаптации, а также генетической архитектуры сложных признаков.

Геномы сельскохозяйственных животных формировались в процессе естественного и искусственного отбора, что позволило им приспособиться к различным условиям окружающей среды, пройти процесс domestikации и сформировать высокопродуктивные фенотипы. Анализ генома сельскохозяйственных животных проливает свет на сложную историю domestikации, потока генов и адаптации, показывая, как естественный отбор и вмешательство человека сформировали генетическую структуру современных популяций сельскохозяйственных животных. Свиньи являются одними из распространенных сельскохозяйственных животных, и исследование их генома представляет большой интерес как в области генетики и селекции, в аспекте сохранения, развития племенных ресурсов и продовольственной безопасности, так и в качестве идеальных моделей для изучения патогенеза и эффективных методов лечения заболеваний у людей.

Одним из ключевых показателей геномного разнообразия является аутозиготность. Аутозиготность – это гомозиготность, возникающая вследствие наследственности (Gibson et al. 2006; McQuillan et al. 2008; Druet & Gautier 2017; Ceballos et al. 2018). Сегменты аутозиготности представлены непрерывными длинными сериями гомозиготных генотипов (ROH, Runs of Homozygosity). Протяженные гомозиготные участки в геноме возникают из-за передачи одинаковых гаплотипов от обоих родителей. Если два одинаковых аллеля происходят от общего предка в результате инбридинга, генотип

считается аутозиготным (идентичным по происхождению, IBD), в противном случае – аллозиготным (идентичным по состоянию, IBS). Методы анализа ROH позволяют точно идентифицировать аллели IBD и широко применяются для оценки уровня аутозиготности у людей и животных (Ceballos et al., 2019; Eusebi et al., 2020; Stoffel et al 2021; Bakoev et al 2021; Kostyunina et al., 2022). Количество и размеры ROH отражают демографическую историю популяции: длинные участки указывают на недавнего общего предка, а короткие – на более далеких предков. Одним из преимуществ ROH является способность надежно идентифицировать длинные сегменты даже при низкой плотности маркеров, что делает метод привлекательным в качестве инструмента для оценки демографической истории популяций и коэффициентов геномного инбридинга у животных. Более того, можно выявить историю инбридинга и локализовать инбридинг на уровне генома (Druet, Gautier 2017). Последнее позволяет исследовать высокоинбридинговые геномные регионы в популяции, которые являются популяционно-специфическими селекционными подписями и позволяют определить гены, которые подверглись дивергентному отбору между популяциями, а также использоваться для связи определенного генотипа с фенотипом (Lander & Botstein 1989; Ceballos et al., 2018). Уникальная структура ROH все больше привлекает внимание как показатель генетической архитектуры сложных признаков. Сегодня применяются попытки для разработки методов, способных реализовать данные исследования, но в большинстве случаев они основаны на анализе связи между длиной (долей) областей гомозиготности и общим количеством сегментов с признаками. Существуют подходы на основе ассоциативных анализов, но в этом случае достаточно сложно выбрать анализируемые сегменты, а также необходимы достаточно большие выборки, что создает большие сложности при исследовании связей между ROH и фенотипами. В целом научное сообщество все больше приходит к выводу, что решающим фактором является не длина или количество сегментов ROH, а уникальность ландшафта областей гомозиготности, что необходимо



учитывать при разработке методов для исследования связей между RОН и фенотипами.

Для удовлетворения продовольственных потребностей людей сельскохозяйственные животные, в частности свиньи, длительный период времени подвергались сильному искусственному отбору. Так, буквально 60 лет назад в нашей стране предпочтение отдавали свиньям с невысокой скоростью роста, но в целом нетребовательным к условиям содержания и кормления и обладающих высокой устойчивостью к заболеваниям. В дальнейшем процессы интенсификации затронули все мировое животноводство, в том числе Россию, и на смену низкопродуктивным свиньям пришли животные, обладающие высокой продуктивностью, но крайне требовательные к условиям содержания и рационам питания.

Изменения в областях генома, спровоцированные отбором, стали называть «подписями 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. Для идентификации локусов, подвергшихся отбору, были разработаны методы, позволяющие сканировать геном на наличие сигналов селекции. Некоторые из методов основаны на паттернах гомозиготности гаплотипов (Chen et al., 2018, Liu et al., 2024). Ключевая характеристика положительного отбора заключается в исключительно 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 (ЕНН) и высокой частотой в популяции. На основе гомозиготности по расширенному гаплотипу разработаны методы iHS (integrated haplotype score), nSl (number of segregating sites by length), iHH12 (integrated haplotype homozygosity pooled), которые получили широкое распространение в

исследованиях, направленных на идентификацию сигналов положительного отбора у людей и сельскохозяйственных животных (Voight et al., 2006, Sabeti et al., 2007, Vakoev et al., 2021).

Геномика, как генетический анализ с использованием маркеров, распределенных по всему геному, набирает все большую популярность в научных исследованиях и практическом применении в таких областях, как здравоохранение и сельское хозяйство. Термин геномика был предложен в 1986 году Томасом Родериком для описания зарождающейся дисциплины секвенирования, картирования, аннотирования и анализа геномов.

Существует по крайней мере два подхода к пониманию/применению геномики: с точки зрения биологии и с точки зрения математики. Условно их можно назвать статистика (матрицы) и последовательность (молекулы). С точки зрения биологии, геномные данные являются источником причинных вариантов и основной целью является идентификация и функциональная аннотация этих вариантов. Потенциал этого направления заключается в понимании механизмов и возможности манипулировать ими. С точки зрения статистики геномные данные представляют собой большие наборы генетических маркеров. Целью является оценка генетического риска или предрасположенности к заболеваниям для людей или оценка племенной ценности для животных и растений. Потенциал этого направления сводится к оптимизации предсказательной способности путем объединения всех маркеров в одну статистическую модель. В данном случае информация о функции конкретных вариантов не рассматривается как значимая. В сельском хозяйстве эта информация в дальнейшем используется для геномного отбора и повышения эффективности селекции. Геномная селекция прошла удивительно короткий путь от теоретического обоснования до практической реализации. Для создания высокоурожайных сортов растений и высокопродуктивных пород животных, неотягощенных различными аномалиями и дефектами, необходимо понимать молекулярно-генетические механизмы и по возможности регулировать генетическую изменчивость в

заданном направлении. Таким образом, только синтез биологии и статистики, использование современных биоинформатических методов позволяет получить необходимую информацию и провести интерпретацию, связывающую генетическую и фенотипическую изменчивость.

Подавляющее большинство исследований генетической архитектуры сложных признаков и заболеваний являются поисковыми и тестируют влияние десятков тысяч однонуклеотидных полиморфизмов (SNP) на интересующий результат. Прогнозы на основе геномных данных, как правило, строятся как взвешенная линейная комбинация доступных SNP. Для оценки весов SNP применяют метод наименьших квадратов. Однако при условии, что количество SNP предикторов ( $P$ ) намного больше, чем количество анализируемых образцов ( $N$ ), метод наименьших квадратов не может быть применен. Решением этой проблемы является 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. За последнее десятилетие было проведено множество исследований для изучения CNV и их ассоциаций с фенотипами у сельскохозяйственных животных. Тем не менее все еще наблюдается недостаточный уровень знаний о CNV как у свиней, так и у многих других видов. Из всех тем, связанных с CNV, знания об их ассоциативных связях с селекционными признаками являются наиболее ограниченными.

Модели для полногеномных ассоциативных исследований (GWAS) должны обрабатывать многомерные данные и учитывать сложность

генетических взаимодействий. Распространенные подходы включают линейные модели, такие как регрессия Лассо и гребневая регрессия, 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). В данном контексте явные преимущества в способности обнаруживать более сложные закономерности с более высокой точностью прогнозирования имеют алгоритмы машинного обучения (Chowdhury et al., 2023). Несмотря на то, что модели машинного обучения обладают большим потенциалом для извлечения шаблонов из наборов данных однонуклеотидного полиморфизма, уровень их внедрения для исследования генетической архитектуры сложных признаков все еще очень низок. Это подчеркивает необходимость проведения исследований, поиска новых подходов, оптимизации моделей, проверки их надежности в условиях больших и разнородных данных, а также решения проблем, связанных с интерпретацией результатов. Модели машинного обучения могут быть объединены с GWAS и популяционной геномикой для выявления генетических вариантов и биологических путей, связанных с определенными фенотипическими признаками. Исследование генетической изменчивости, сформированной под действием отбора с учетом ассоциативного анализа, позволит получить ценную информацию о генетических механизмах, связанных с признаками продуктивности. 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 области генетики и геномики, 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 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.

### **Объект и предмет исследования**

Объектом исследования являются представители вида *Sus scrofa* домашние свиньи (*Sus scrofa domesticus*). Свиньи обладают способностью адаптироваться к различным типам окружающей среды, а также достаточно гибко реагировать на изменения, связанные с селекционным давлением. Кроме того, свиньи являются идеальными моделями для изучения патогенеза и разработки эффективных методов лечения заболеваний у человека благодаря сходству физиологических особенностей с человеческим организмом.

Предметом исследования выступают геномные данные как источник информации о демографических изменениях, адаптационных способностях, обусловленных искусственным и естественным отбором, позволяющие прогнозировать возможности селекционного отбора и моделировать желательные признаки продуктивности и устойчивости к заболеваниям.

## **Цель работы**

Целью диссертационной работы является исследование генетической изменчивости свиней 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 уровня аутозиготности диких кабанов, свиней пород крупная белая, дюрок и ландрас.
2. Провести анализ генетической изменчивости и уровня аутозиготности свиней крупной белой породы отечественной и импортной селекций.
3. Определить подписи отбора в геномах свиней крупной белой породы 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 количества 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 крупной белой породы 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 крупной белой породы, разводимых на территории России до и после процессов интенсификации. Идентифицированы подписи селекции, вызванные искусственным отбором. Определена генетическая дифференциация между современными свиньями крупной белой породы 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.

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 количества поросят при рождении, в результате чего были идентифицированы 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 последующим анализом с применением сверточных нейронных сетей (CNN). Эта стратегия представляет новый взгляд на анализ геномных данных, выходящий за рамки традиционных числовых и статистических методов 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. 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.

Полученные данные открывают новые перспективы для оптимизации селекционных программ, направленных на повышение продуктивности и устойчивости современных пород свиней. Работа также затрагивает вопросы



визуализации и структурного анализа геномных особенностей, что имеет важное значение для управления генетическими ресурсами и поддержания генетического разнообразия популяций.

### **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 позволяет 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 г.; 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.; 71<sup>st</sup> Annual Meeting of European Federation of Animal Science (EAAP) Porto, Portugal, 2020 г., 72<sup>nd</sup> 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 Российской академии наук, 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).

### **Объем и структура работы**

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 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 института синтетической биологии ФГБУ «ЦСП» ФМБА России по науке, 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., научному консультанту, доктору биол. наук Усатову А.В.

### **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, внутрипородной стратификации и интенсивного отбора по селекционно-ценным признакам».

# 1. ОБЗОР ЛИТЕРАТУРЫ

## 1.1 Доместикация сельскохозяйственных животных

Породы сельскохозяйственных животных являются культурным достижением зоотехнической науки и представляют собой генетические ресурсы для экономики, а также уникальный материал для изучения генетических механизмов, лежащих в основе их адаптации к местному климату, условиям содержания и кормления, а также приоритетам людей относительно селекционно-значимых признаков.

Доместикация растений и животных – важнейшее событие за последние 13 тысяч лет истории человечества. Под domestiцированными (одомашненными) подразумевают виды, выведенные в неволе и тем самым модифицированные по сравнению со своими дикими предками, что делает их более полезными для людей, контролирующих их воспроизводство и пищу (Diamond, 2002). Таким образом, доместикация отличается от простого приручения диких животных. Африканские боевые слоны Ганнибала были, а современные азиатские рабочие слоны остаются просто прирученными дикими особями, а не особями генетически отличной популяции, рожденными и выращенными в неволе. Интересно отметить, что перспективными domestiцированными животными являются крупные млекопитающие (травоядные или всеядные), которых 148 видов (весом 45 кг и более). Однако только 14 из этих 148 видов были фактически domestiцированы. Аналогичная картина просматривается и у растений, где из 200 тысяч видов только 100 domestiцированы. Особенно удивительно то, что во многих случаях приручился только один из близкородственных видов. Например, лошади и ослы были domestiцированы, но ни один из четырех видов зебр не подвергся доместикации. Ключевой вопрос заключается в избирательности доместикации. Очень подробно этот вопрос рассмотрен в обзоре Diamond (2002), согласно которому доместикации могут быть подвергнуты только те виды, которые эволюционно к этому предрасположены. Например,

европейские коневоды, которые поселились в Южной Африке в 1600-х гг., и как африканские пастухи на протяжении предыдущих тысячелетий пытались приручить зебр. Через несколько столетий они сдались. Зебры злобны, а также у них периферийное зрение лучше, чем у лошадей, и они видят приближающуюся опасность, откидывают голову.

Diamond (2002) выделяет шесть основных причин, препятствующих domestikацию диких видов. Диета, которую нелегко предоставить в домашних условиях (отсутствие домашних муравьедов), медленный темп роста и длительный интервал между рожденьями (например, слоны и гориллы), неприятный нрав (гризли, медведи и носороги), нежелание размножаться в неволе (панды и гепарды), отсутствие иерархии доминирования следования за лидером (снежный баран и антилопа), а также склонность к панике в вольерах или при столкновении с хищниками (газели и олени, кроме северных оленей). У многих видов успешную domestikацию не удастся преодолеть именно из-за шестого препятствия.

Согласно имеющимся на сегодняшний день данным считается, что процесс domestikации животных берет начало с волка (*Canis lupus*), который начался не менее 15 тысяч лет назад (Freedman, Wayne, 2017; Larson et al., 2012; Frantz et al., 2016; Botigue et al., 2017). Это связывают со значительными экологическими и климатическими изменениями, сопровождавшими глобальный переход от пика последнего ледникового максимума, примерно 21 тысяч лет назад, к нынешнему голоценовому периоду (McHugo et al., 2019). Овцы (*Ovis aries*), козы (*Capra hircus*), безгорбый тауриновый скот (*Bos taurus*) и свиньи (*Sus scrofa*) были одними из первых животных, которые были domestikированы около 10–11 тысяч лет назад в районе Плодородного Полумесяца (условное название региона на Ближнем Востоке) (Larson et al., 2014; Zeder et al., 2011).

Род *Sus* возник в Юго-Восточной Азии во время видообразования в позднем миоцене или околограницы миоцена/плиоцена примерно 14–4 млн лет назад (Larson, 2014; Mona, 2007; Groenen, 2012). Предполагают, что

европейские кабаны отделились от *Sus scrofa* в Юго-Восточной Азии, в конце плейстоцена, между 0,5 и 0,9 млн лет назад (Larson, 2014; Scandura, 2008). Европейский кабан – теплолюбивое животное, распространён в южных и умеренных широтах. Начиная с позднего голоцена и до II века н.э. ареал европейского кабана практически не изменялся. В дальнейшем, изменения, связанные с распространением кабаны, были обусловлены климатическими факторами. Однако, начиная с XV–XVIII веков, большое негативное влияние на кабанов и их среду обитания оказывали антропогенные факторы. Европейский кабан подвергся массовому истреблению на всей территории Европы и России. Согласно данным, представленным в работе Данилкина (2020), в 1960–1661 гг. кабаны были уничтожены в Англии и Нидерландах, 1801–1803 гг. в Ютландии, Саксонии, Баварии и в других районах Германии, а также в Швейцарии. В это время небольшое количество кабанов сохранилось в Западной Европе. До 1915 г. кабаны были распространены в Белоруссии, но в дальнейшем их численность резко сократилась. В Киевской Руси резкое снижение численности кабанов произошло в XVIII веке и к XIX веку они практически исчезли. В начале XX века небольшое поголовье сохранялось в труднодоступных местах по берегам Днепра, в Прикарпатье и Карпатах, Грузии. В XX веке искусственное расселение кабанов дало начало их естественному восстановлению и расширению ареала. Так, на территорию СССР было выпущено около 9 тысяч кабанов, а непосредственно в России – около 8 тысяч. На начало 2019 г. общая численность кабанов в России составила около 286 тысяч, в США – 4 млн, в Европе – 4,5 млн особей.

Свиньи были одомашнированы около 10000–8000 лет назад. Однако относительно того, где и как это произошло, существуют различные мнения. Широкое распространение диких свиней на большей части территории Евразии усложняет классификацию археологических образцов как диких, так и домашних свиней и оставляет открытой возможность того, что свиньи были независимо одомашнированы в Европе и на Ближнем Востоке. В этом контексте интерес представляет работа Frantz с соавторами (Frantz et al., 2018),

результаты которой показали, что анатолийские кабаны, одомашнированные около 10500 лет назад, были предками домашних свиней, доставленных в Европу около 8500 лет назад. В дальнейшем доля генома ближневосточных домашних свиней в геноме европейских домашних свиных упала до 50% (поздний неолит, примерно 5000 лет назад) и постепенно снижалась, практически до полного исчезновения. Это означает, что европейские домашние свиньи возникли не в результате независимого процесса одомашнивания, а скорее в результате гибридизации между ближневосточными домашними свиньями и европейскими дикими кабанами.

Данные исследований по *S. scrofa* демонстрируют значительную интрогрессию и поток генов между популяциями кабанов и домашних свиней после одомашнивания, что указывает на совершенно иной процесс одомашнивания, т.е. не связанный с первоначальной одомашниванием ограниченного числа кабанов из дискретных местных популяций и в дальнейшем их в определенной степени генетической изоляции (Marshall et al., 2014).

Исторические и современные этнографические наблюдения за традиционным содержанием свиней, например, в Средиземноморье и Европе, указывают на обычную практику довольно свободного и обширного содержания домашних свиней, выпасающих на лесных пастбищах. Такое свиноводство было нормой в Европе и в таких обстоятельствах вполне вероятно, что скрещивание домашних свиней с диким кабаном было обычным явлением.

До XIX столетия большую часть Европы занимали местные свиньи. Различные экологические условия отдельных районов этого пространства оказывали свое влияние на животных (Wilkinson, 2013). С конца XVIII в. местных свиней начали улучшать путем скрещивания с романовскими, восточными, а в дальнейшем и китайскими свиньями. Большая селекционная работа привела к созданию ограниченного количества коммерческих пород.



В середине XX века большое количество племенных ассоциаций, действующих на региональном уровне, отвечали за свиноводство. Каждая из этих племенных ассоциаций и племенных компаний имела собственное племенное стадо, которое обычно основывалось на ограниченном количестве коммерческих пород, но часто происходило из национальных или региональных и, следовательно, уникальных популяций.

До XX века отрасль свиноводства в России развивалась в относительно небольших масштабах. Наибольшую активность свиноводство приобрело в конце XIX века, когда были созданы отечественные породы путем скрещивания нескольких популяций из разных пород, включая крупную белую, ландраса, среднюю белую, гемпшир, тамуорт, мангалица, местного дикого кабана и азиатских свиней. В 1980 г. в СССР были зарегистрированы 22 местные породы, которые играли важную роль в производстве свинины и вносили вклад в экономическое развитие страны. Несмотря на то, что эти породы основаны на импортном генетическом материале, целенаправленная селекционная работа, сфокусированная на решении определенных задач, диктуемых тем временем, привела к развитию пород с уникальными характеристиками. В связи с распадом СССР и последующими значительными социально-экономическими изменениями, отечественные породы стали практически неконкурентоспособными по сравнению с европейскими коммерческими породами. Это привело к тому, что многие из этих пород, хорошо адаптированных к местному климату, устойчивых к болезням и обладающих высокими качественными показателями мяса, были заменены международными коммерческими породами с более высокой производительностью.

Необходимо отметить, что значительное сокращение локальных пород свиней произошло во всем мире. За последние десятилетия в коммерческой племенной отрасли произошла значительная консолидация бизнеса за счет слияний и поглощений, в результате которых осталось ограниченное количество племенных компаний, работающих на международном уровне.

Племенные линии, принадлежащие этим компаниям, также испытали высокую степень консолидации. Селекционные линии, проигравшие в конкуренции с точки зрения производительности и генетического прироста, исчезли.

## **1.2 Подписи отбора**

Понимание того, как адаптация и отбор формируют закономерности генетической изменчивости, имеет большое значение, поскольку оно дает ценную информацию об изучении генетических механизмов, связанных с устойчивостью к болезням и признаками продуктивности. Домашние животные подвергаются сильному искусственному отбору для удовлетворения потребностей людей, таких как скорость роста, размер тела, состав мышц и размножение. Эти процессы отбора оставили изменения в областях генома, которые стали называть «подписями отбора» (selection signatures). Появление высокопроизводительных методов генотипирования и секвенирования способствовало обнаружению подписей отбора у домашних животных на уровне генома.

Выявление геномной основы, лежащей в основе местной адаптации, имеет первостепенное значение для эволюционной биологии и имеет множество применений в области природоохранной биологии, селекции сельскохозяйственных культур и животных, а также персонализированной медицины. Чтобы идентифицировать локусы, на которые в прошлом повлиял положительный отбор, генетики разработали методы, позволяющие сканировать геномы на наличие сигналов, характерных для этого процесса (Refoyo-Martínez et al., 2019). Эти сигналы могут быть основаны на паттернах гомозиготности гаплотипов (Voight et al., 2006; Sabeti et al., 2007), частотном спектре сайтов (Nielsen et al., 2005; Huber et al., 2016) или аллельной дифференциации между популяциями (Yi et al., 2010).

Методы, основанные на дифференциации населения, оказались особенно успешными в последние годы, поскольку им не требуется предположений о лежащем в основе демографическом процессе, который мог породить сигнал отбора, и, как правило, более надежны и масштабируемы для больших наборов данных по всему населению. Основаны они на вычислении попарных FST (Weir et al., 1984) или аналогичных мерах популяционной дифференциации между двумя популяционными панелями по SNP или окнам генома.

В результате географически ограниченного селективного давления положительный отбор может привести к повышению частоты определенного адаптивного аллеля до высокой частоты в популяции, где он приводит к благоприятному фенотипическому результату, но в других популяциях из отдаленных регионов он не дает селективного преимущества и остается нейтральным. Эта ситуация создает отличительный паттерн крайней популяционной дифференциации, т.е. локальной адаптации. Для поиска паттернов локальной адаптации хорошо зарекомендовал себя метод PBS (population branch statistic), основанный на сравнении различий частот аллелей, измеренных с помощью значений FST (Yi et al., 2010).

Гаплотипы в контексте популяционной геномики рассматриваются на основе предположения, что мутация, представленная альтернативным нуклеотидом (или аллелем) для варианта, может стать наследуемой, появившись у индивидуума, т.к. продукт этого гена (например, белок) изменен таким образом, что повышает либо репродуктивные шансы данного индивидуума, либо его выживание в окружающей среде (Abondio et al., 2022). Следовательно потомство этой особи с положительной мутацией будет иметь преимущество приспособленности в тех же условиях окружающей среды с точки зрения выживания и размножения, а со временем популяция будет обогащена субъектами, несущими один и тот же полезный аллель (Szpak et al., 2019). Поскольку генетический материал наследуется фрагментами во время гомологичной рекомбинации, то не только положительные мутации будут

увеличивать свою частоту в поколениях, но и окружающие их нейтральные варианты, которые относятся к одному и тому же сегменту (явление, называемое «автостопом») (Barton et al., 2000). Таким образом в геноме формируются регионы (гаплотипы) с высокой LD (неравновесие по сцеплению) и низкой генетической изменчивостью (Novembre et al., 2012). Поскольку рекомбинация разрывает связь между вариантами с течением времени, ожидается, что относительно недавние адаптивные события будут представлены сравнительно расширенными гаплотипами, в то время как более старые события отбора будут наблюдаться как меньшие гаплотипы в общей популяции (Stephan, 2019).

Ключевой характеристикой положительного отбора является то, что он вызывает необычно быстрое увеличение частоты аллеля, происходящее за достаточно короткое время, чтобы рекомбинация существенно не разрушала гаплотип, на котором происходит выбранная мутация. Таким образом, тест на положительный отбор включает в себя поиск основного гаплотипа с комбинацией высокой частоты и высокого ЕНН (гомозиготность по расширенному гаплотипу) по сравнению с другими основными гаплотипами в локусе (Sabeti et al., 2002). На основе гомозиготности по расширенному гаплотипу разработаны методы iHS (integrated haplotype score), nSI (number of segregating sites by length), iHH12 (integrated haplotype homozygosity pooled), XP-ЕНН (cross-population extended haplotype homozygosity), которые широко применяются в исследованиях, направленных на идентификацию сигналов положительного отбора у людей и сельскохозяйственных животных (Zorc et al., 2022; Zhong et al., 2023; Xia et al., 2023; Lukic et al., 2023; Salek et al., 2023; Ceccobelli S. et al., 2023; Wu et al., 2023).

Исследования подписей отбора у людей выявили геномные регионы, на которые повлияло эволюционное давление, связанное с выживанием, воспроизводством и адаптацией. Основные результаты включают локусы, связанные с иммунитетом (например, гены HLA), диетическими адаптациями (например, устойчивостью лактазы) и устойчивостью к окружающей среде

(например, пигментацией кожи, адаптацией к большой высоте). Эти исследования также раскрывают доказательства недавнего отбора, обусловленного культурными практиками, такими как диета и воздействие патогенов, и раскрывают генетические компромиссы, влияющие на здоровье и восприимчивость к болезням. Предполагают, что культурные и экологические факторы влияют на эволюционную траекторию формирования психоэмоциональных черт человека. В работе Вакоев с соавторами (Вакоев et al., 2023) определены у людей подписи отбора в генах *OXT*, *OXTR*, *AVP*, *AVPR1A* и *AVPR1B*, участвующих в регуляции психоэмоциональных реакций, таких как социальная связь, реакция на стресс и эмоциональное поведение. Варианты в генах *OXT* и *OXTR* были связаны с улучшенной социальной связью и эмоциональной регуляцией, в то время как полиморфизмы в гене *AVP* и его рецепторов (*AVPR1A* и *AVPR1B*) продемонстрировали доказательства отбора, связанного с реакцией на стресс и модуляцией агрессии. Результаты подчеркивают роль этих генетических вариантов в формировании психоэмоциональных черт человека и дают представление об их эволюционном и адаптивном значении.

Исследования генов циркадных ритмов *PER1*, *PER2* и *PER3* в различных человеческих популяциях показали сигналы положительного отбора, вероятно, связанные с адаптивной эволюцией в ответ на давление окружающей среды (Мишина и др., 2024). Результаты показали, что условия жизни, климат и другие внешние факторы напрямую влияют на генетическую структуру популяций, что приводит к вариациям в генах *PER* в разных этнических группах и географических точках. Примечательно, что многие из вариантов отбора в генах *PER1*, *PER2* и *PER3*, по-видимому, регулируют биологические процессы, связанные с основными современными заболеваниями, включая ожирение, рак, метаболический синдром, биполярное расстройство личности, депрессию, ревматоидный артрит, сахарный диабет, красную волчанку, инсульт и болезнь Альцгеймера, что делает их

чрезвычайно интересными мишенями для дальнейших исследований, направленных на выявление генетических причин заболеваний человека.

Исследования подписей отбора у сельскохозяйственных животных позволили определить геномные регионы, связанные с процессами domestikации, адаптации и продуктивности (ASAS). Такие исследования подчеркивают влияние как естественного, так и искусственного отбора, демонстрируя, как методы разведения формируют генетическую архитектуру сельскохозяйственных животных. У свиней гены, локализованные в областях подписей отбора, часто связаны с признаками, имеющими экономическое значение, такими как рост, воспроизводство, устойчивость к болезням.

Zhang с соавторами (Zhang et al., 2014) провели широкомасштабный анализ геномных данных (129 пород) для определения генов, отвечающих за адаптацию к различным условиям среды и за признаки, сформировавшиеся в процессе domestikации и селекции. Особое внимание уделено роли главного комплекса гистосовместимости (МНС) в поддержании генетического разнообразия и адаптации. В результате исследования были идентифицированы гены, связанные с адаптацией к высокогорью, температуре окружающей среды, domestikацией и искусственной селекцией, а также проанализированы особенности генетического разнообразия в регионе МНС.

В исследованиях свиней пород крупная белая, ландрас и дюрок датской селекции показано, что под селекционным давлением оказались гены, связанные с активностью обонятельных рецепторов и оксидоредуктазной активностью (Cai et al., 2020). Гены обонятельных рецепторов имеют решающее значение для обоняния, влияя на поведение, такое как поиск пищи и спаривание, которые жизненно важны для выживания и воспроизводства. Оксидоредуктазы — это ферменты, участвующие в окислительно-восстановительных реакциях, играющие ключевую роль в метаболических процессах и детоксикации активных форм кислорода, тем самым способствующие здоровью и функционированию клеток. Гены обонятельных

рецепторов определены у многих domesticiрованных видов, в том числе у крупного рогатого скота и свиней, что указывает на их важность в процессах domestикации и адаптации. Предполагают, что селекционное давление на эти гены могло быть значительным во время развития породы и было способно потенциально влиять на признаки, связанные с сенсорным восприятием и эффективностью метаболизма.

Atrian-Afiani с соавторами (Atrian-Afiani et al., 2023) сравнили австрийскую популяцию туропольских свиней (местная порода) с четырьмя коммерческими породами. В результате были определены гены, связанные с особенностями селекционных стратегий для разных пород свиней. Гены *ATP1A1*, *CASQ2*, *CD2*, *IGSF3*, *MAB21L3*, *NHLH2*, *SLC22A15* и *VANGL1* подверглись селекции у свиней крупной белой породы; гены *ARSB*, *BHMT*, *BHMT2*, *DMGDH* и *JMY* – у свиней породы дюрок. Для туропольских свиней под давлением селекции оказались ген *DPYD*, связанный с мышечным развитием, и ген *PTBP2*, предположительно связанный с характеристиками семени.

В работе Saini с соавторами (Saini et al., 2024) объектом исследования стали свиньи породы ландли (Landlly), выведенной в Индии на основе породы ландрас (75%) и одной из местных пород Индии гурра (Ghurrah) (25%). В областях подписей селекции определены гены, ответственные за развитие мышц (*IGF1* и *MYOD1*), воспроизводство (*ESR1*), устойчивость к болезням (*TLR4* и *CD163*) и стрессу (*HSP70*), связанные с поведением (*DRD2*) и цветом шерсти (*KIT*).

Анализ генетического разнообразия и подписей отбора у шести автохтонных пород свиней из Словении, Хорватии и Сербии (пятнистая банья, черная славянская, туропольская свинья, ласточкина мангалица, моравка и крскопольская свинья) показал, что каждая порода сохранила свою генетическую уникальность. С селекционным давлением связаны гены *MC4R*, *LEPR*, *RBP4*, *SCD*, *FASN*, *H-FABP*, *ESR1*, *GHR*, *IGF2*, *TLR2*, *TLR4* и др.,

участвующие в воспроизводстве, росте, метаболизме жирных кислот и устойчивости к болезням.

Исследование генетической истории кельтско-иберийской породы свиней показало несколько ключевых генов, находящихся под селекционным давлением, выделив такие черты, как качество мяса, адаптация к окружающей среде и domestикация (Arias et al., 2024). Такие гены, как *MC4R* и *PPARG*, связаны с отложением жира и мраморностью, улучшением вкуса и нежностью мяса. Адаптивные черты были связаны с такими генами, как *HSP70* (устойчивость к тепловому стрессу) и *PRKAG3* (энергетический метаболизм мышц), поддерживающими выживание в иберийском климате. Гены, связанные с domestикацией, такие как *DRD2*, вероятно, обеспечивают поведенческие черты, такие как послушание, в то время как ген *TLR4* связан с иммунитетом и устойчивостью к болезням.

Lee с соавторами (Lee et al., 2023) оценили генетическую дифференциацию между местными свиньями острова Чеджу (Корея) и свиньями пород беркшир и йоркшир. На основе этой работы было выдвинуто предположение о том, что гены *TJP1* (регулируют проницаемость эпителиальных и эндотелиальных клеточных барьеров), *ABAT* (имеют решающее значение для неврологического баланса и энергетического обмена) и *MMP25* (модулирует иммунных реакций) обеспечили уникальность местных свиней, их локальную адаптацию к окружающей среде и устойчивость к местным патогенам.

Масштабное геномное исследование свиней породы мейшань выявило уникальными фенотипическими характеристиками этой породы (Zhao et al., 2018). Были обнаружены области генома, подвергшиеся положительному отбору, включая гены, отвечающие за высокую плодовитость (*IGF1R*, *FoxO*) и морфологические особенности, такие как морщинистая кожа (*NFKB1*) и черная шерсть (*MC1R*). Исследование также показало генетический обмен между свиньями мейшань и европейскими породами, подтверждая их историческое влияние на формирование современных пород.



Wu с соавторами (Wu et al., 2023) провели исследования свиней породы дяньнань с маленькими ушами (DSE), китайская местная порода свиней южных районов провинции Юньнань. Эти свиньи демонстрируют наследственную адаптацию к высокой температуре и влажности. Результаты показали низкое генетическое разнообразие и высокую степень инбридинга, а также выявили гены, предположительно связанные с качеством мяса (*COL15A1*, *RPL3L* и *SLC9A3R2*), размером тела (*PALM2-AKAP2*, *NANS*, *TRAF7* и *PACSIN1*), адаптивностью (*CLDN9* и *E4F1*) и аппетитом (*GRM4*).

Южная Африка может похвастаться разнообразием популяций свиней, включая интенсивно выращиваемые коммерческие породы, а также местных и деревенских свиней, выращиваемых в условиях низкзатратных производственных систем. Hlongwane с соавторами (Hlongwane et al., 2024) изучили влияние отбора на геном южноафриканских свиней и выявили гены *NECAP1*, *KCNJ3*, *EPHB2*, *EPB41L3*, *METTL4*, *EPHA8*, *LYPLA2*, *FUCA1*, *PNRC2*, *SRSF10*, *MYOM3*, *SLC16A12*, *PANK1*, *PCGF5* и др., связанные с важными хозяйственными признаками, такими как качество мяса, репродукция, адаптация и устойчивость к болезням.

Интересным приложением исследований подписей отбора является использование этого метода как инструмента для непосредственного поиска генетических вариантов, связанных либо с признаками продуктивности, либо с какими-либо заболеваниями или дефектами. Так, Chen с соавторами (Chen et al., 2024) исследовали свиней китайской аборигенной породы лайу. На основе различий в фенотипах были выделены три группы с экстремальными значениями пяти показателей качества мяса (внутримышечный жир, потери при вытекании сока, водоудерживающая способность, pH через 45 минут и 24 часа после забоя). Анализируя данные генотипирования, авторы выявили гены, связанные с этими характеристиками, используя статистику FST и ХРЕНН. Ключевым результатом стало определение нескольких кандидатных генов, таких как *USF1*, *NDUFS2*, *PIGM*, *IGSF8* и *CASQ1*, влияющих на метаболизм липидов и развитие мышц.

Исследования Shen с соавторами (Shen et al., 2024) были направлены на поиск генов, связанных с качеством мяса. Выборку свиней крупной белой породы разделили на две группы на основе измерения фенотипов с низкими и высокими показателями. По результатам анализов определили перспективный геномный регион (SSC4 93544042–95179724), связанный с высоким содержанием внутримышечного жира. В этом регионе локализованы гены, играющие важную роль в липидном гомеостазе (*USF1*), липидном обмене и катаболизме (*PIGM*), энергетическом метаболизме (*NDUFS2*) и в развитии мышц (*IGSF8* и *CASQ1*).

Getmantseva с соавторами (Getmantseva et al., 2023) провели пилотные исследования, направленные на поиск геномных локусов и генов-кандидатов, связанных с дефектами конечностей у свиней. Для исследования были взяты свиньи породы ландрас и дюрок. Свиньи были разделены на две группы в зависимости от состояния их конечностей. В результате выявленные регионы перекрывались с QTL, связанными с признаками здоровья (параметры крови) и признаками мяса и туши (упитанность). Всего был идентифицирован 31 ген (17 генов у ландраса, 14 генов у дюрока). Три гена появились как в группе ландрас, так и в группе дюрок, включая *A2ML1* (SSC5), *ROBO2* (SSC13) и *MSI1* (SSC14).

Romanets с соавторами (Romanets et al., 2024) применили метод Fst для поиска вариантов, связанных с количеством и массой поросят при рождении, т.к. на сегодняшний день крайне значимой задачей является недопущение снижения массы поросят при рождении при сохранении высокой плодовитости свиноматок. По результатам анализа выделено 724 варианта, представляющих сигналы отбора по следующим признакам: количество поросят при рождении, многоплодие, масса поросят при рождении и масса одного поросенка при рождении. При этом выявлено 18 общих вариантов, которые являются потенциальными маркерами как для количества поросят при рождении, так и для массы, что крайне важно для селекционной работы по улучшению репродуктивных качеств свиноматок. Большинство из них

локализовано в генах, связанных с эмбриональным процессом, выживаемостью эмбрионов, а также различными патологиями, в дальнейшем связанными со снижением роста.

Таким образом, анализ генома свиньи показывает, что domestикация и селективное давление сформировали генетический профиль свиней. Сильные подписи отбора наблюдаются в генах, связанных с экономически и экологически значимыми признаками, такими как *MC4R* (рост и жировой обмен), *IGF1* (регуляция роста) и *LEP* (отложение жира и эффективность корма) (Groenen et al., 2012). Иммунозависимые гены, такие как *TLR4* и *CD163*, показывают отбор на устойчивость к болезням, в то время как гены, такие как *HSP70*, отражают адаптацию к тепловому стрессу в тропических регионах. Поведенческие гены, включая *DRD2*, были отобраны для улучшения темперамента в доместицированных условиях. Эти селективные давления подчеркивают двойное влияние естественной адаптации и направленной человеком селекции на такие признаки, как воспроизводство, иммунитет и толерантность к окружающей среде. Результаты подчеркивают важность понимания этих генетических адаптаций для улучшения программ разведения животных при сохранении генетического разнообразия.

### **1.3 Протяженные гомозиготные области**

В последнее время большой интерес приобретают исследования областей ROH (Runs of Homozygosity), т.к. они играют ключевую роль в оценке генетического разнообразия и демографической истории популяций. Эти протяженные области гомозиготности позволяют выявлять степень инбридинга, который часто связан с утратой генетического разнообразия, и отслеживать последствия исторических событий, таких как узкие места популяции или эффекты основателя. Кроме того, ROH дают ценную информацию о геномных участках, находящихся под действием направленного отбора, что особенно важно для понимания адаптивных

процессов и выявления селекционных подписей. Анализ ROH становится незаменимым инструментом в генетических исследованиях, включая изучение эволюции, устойчивости популяций и разработки селекционных программ.

В последнее время метод определения ROH находит все большее применение в поиске и идентификации участков генома, связанных с давлением отбора. Протяженные гомозиготные области (ROH, Runs of Homozygosity) – это непрерывные гомозиготные участки, которые присутствуют в генотипе животного из-за того, что оба родителя передали одинаковые гаплотипы своему потомству. Когда два одинаковых аллеля в локусе происходят от общего предка путем неслучайного спаривания (инбридинга), генотип считается аутозиготным или говорят об «идентичности по происхождению» (IBD, Identical by descent). Когда два одинаковых аллеля происходят из разных источников, генотип называется аллозиготным, или имеет место «идентичность по состоянию» (IBS, Identical by state). Методы для определения ROH обеспечивают более точную идентификацию аллелей в локусе IBD и широко используются в исследованиях людей и животных для точной оценки уровня аутозиготности (Kirin et al., 2010, Peripolli et al., 2017, Marchesi et al., 2018).

Количество ROH и распределение их по размеру отражают недавние и более ранние исторические события в популяции. Длинные участки ROH, как правило, расположены в областях с низкой рекомбинацией, т.е. в середине хромосомы, а самые маленькие ROH расположены в периферийных областях и наиболее часто встречаются в теломерных областях (Bosse et al., 2012). В соответствии с этим наличие длинных участков ROH указывает на недавнего общего предка у родительских форм особи. И наоборот, более короткие сегменты ROH происходят от далеких предков или могут включать в себя некоторые участки, не относящиеся к IBD (Howrigan et al., 2011). Идентификация и характеристика ROH позволяют получить представление о том, как структура популяции и демография изменялись с течением времени,

а также раскрыть генетические взаимоотношения между индивидуумами и определить уровень инбридинга в популяции.

Одной из сильных сторон анализа ROH служит то, что длинные гомозиготные сегменты могут быть надежно идентифицированы даже при относительно невысокой плотности маркеров (Ceballos et al., 2018). Это делает ROH привлекательным методом для изучения и понимания роли аутозиготности, а также в качестве возможного инструмента управления уровнем инбридинга для программ разведения в животноводстве. ROH – это прямая мера гомозиготности, которая по сравнению с другими методами геномной оценки менее подвержена воздействию отбора и ошибкам, вызванным вариациями частот аллелей (Gurgul et al., 2016). Следовательно, индекс  $F_{ROH}$  можно использовать для исправления ошибок родословных (Ferencakovic et al., 2017; Marras et al., 2015), и он рекомендован для определения IBD (Gurgul et al., 2016).

В животноводстве ROH впервые применил Sölkner J. (Solkner et al., 2010), а в более поздних исследованиях (Ferencakovic et al., 2013; Deirdre et al., 2012) обнаружили, что уровень инбридинга ( $F$ ), оцениваемый по ROH ( $F_{ROH}$ ), имеет ряд преимуществ по сравнению с методом FPED, предложенным Райтом–Кисловским.  $F_{ROH}$  может предсказать фактический процент аутозиготного генома более точно, чем FPED, также  $F_{ROH}$  может быть оценен в любом генотипированном индивидууме, даже при отсутствии родословной информации. К преимуществам ROH можно отнести также возможность изучения распределения аутозиготности по всему геному, поиск конкретных мест в геноме с более высокими уровнями аутозиготности и определение аутозиготности, возникающей от очень отдаленных общих предков (например, более 50 поколений назад).

#### **1.4 Основные методы определения ROH**

Методы, заложенные в программных обеспечениях, позволяющих проводить поиск участков ROH, можно классифицировать как

наблюдательные, модельно-ориентированные и методы для определения коротких последовательностей (Ceballos et al., 2018). Несмотря на условное разделение, большая часть этих программ способна решать общие задачи по поиску и идентификации ROH (Бакоев С.Ю., Гетманцева Л.В., 2019).

**Наблюдательные методы.** Алгоритмы, реализованные в PLINK (Purcell et al., 2007), сканируют каждую хромосому, при этом задается окно фиксированного размера, чтобы исключить ошибки и сбои генотипирования, а также редкие новые мутационные события. С целью поиска отрезков последовательных гомозиготных SNP это окно фиксированного размера перемещается вдоль длины хромосомы. На первом этапе ROH определяется вычислением пропорции полностью гомозиготных окон, которые охватывают SNP. Если эта пропорция выше определенного порога, SNP обозначается как находящийся в ROH. Далее ROH считается идентифицированным, если количество последовательных SNP в гомозиготном сегменте превышает определенный порог с точки зрения количества SNP и/или покрытой длины хромосомы. Как было отмечено выше, ROH – это непрерывные гомозиготные участки, которые присутствуют у животного из-за того, что родители передают идентичные гаплотипы своим потомкам. Чем длиннее сегменты ROH, тем более вероятно, что недавно в пределах родословной произошел инбридинг. Тем не менее необычно длительные гомозиготные участки могут сохраняться и у беспородных особей, возможно, из-за необычной мутации, неравновесного сцепления (LD, Linkage disequilibrium) и частоты рекомбинации в определенных местах генома (Danese et al., 2011). Простота подхода PLINK, если существует необходимость, позволяет перед вызовом ROH сократить для LD количество SNP.

Для идентификации ROH как особого случая IBD у индивидуума также можно использовать программу GERMLINE (Gusev et al., 2009), которая сопоставляет гаплотипы животных для вычисления идентичности по происхождению (IBD).

Платное программное обеспечение SNP & Variation Suite (SVS) Golden Helix не использует раздвижные окна, но рассматривает все SNP в гомозиготе в качестве возможной отправной точки для нового ROH. Каждый SNP классифицируется как «гомозиготный», «гетерозиготный» или «пропущенный вызов» и обеспечивает кластер гомозиготных областей с количеством SNPs в гомозиготном режиме, превышающим указанное для каждой хромосомы и отдельного индивидуума. Затем второй алгоритм группирует все идентифицированные области в кластеры и предоставляет список с минимальным количеством индивидуумов, которые имеют эти области в совокупности. Этот более современный и сложный метод позволяет пользователю определить группы параметров, таких как минимальный размер ROH в парах оснований, количество SNP, минимальная плотность, максимальный разрыв, максимальное количество гетерозигот и количество «пропущенных вызовов» (Curik et al., 2014).

В качестве альтернативы SVS в программной среде R можно использовать пакет detectRUNS (Marras et al., 2015), в котором реализованы методы определения ROH, при этом команда slidingRUNS использует метод, идентичный реализованному в PLINK, а команда consecutiveRuns – метод последовательного прогона по всей длине генома.

**Модельно-ориентированные методы.** Альтернативный наблюдательным реализован метод в программном обеспечении Beagle, которое использует скрытые модели Маркова (HMM) для учета фоновых уровней LD (Browning et al., 2010). Метод LROH, реализованный в программном обеспечении VCFtools (Danecek et al., 2009), также основан на использовании скрытой модели Маркова. HMM состоит из двух состояний для каждого SNP, которые представляют собой LROH. Вероятности выбросов в каждом SNP для каждого состояния зависят от вероятности наличия гетерозиготы, основанной на гетерозиготности SNP в популяции, и оценочной частоты ошибки генотипирования. Модель состоит из двух скрытых

состояний, а именно аутозиготные (A) и неаутозиготные ( $\neg A$ ). Если SNP  $i$  имеет генотипическое состояние  $X_i$  (равное 0, 1, 2, где 1 – гетерозиготное состояние) и скрытое состояние  $S_i$ , вероятности выбросов для двух состояний на каждом SNP определяются как

$$P(S_{i+1} = A | S_i = \neg A) = P(S_{i+1} = A(1 - e^{-2M(r_{i+1} - r_i)})), \quad (1)$$

где  $M$  – ожидаемое число мейозов со времени недавнего общего предка,  $r_i$  – местоположение SNP на генетической карте  $i$  в Морганах.

Также можно отметить программу Garlic (Szpiech et al., 2017), основанную на оценке плотности ядра Гаусса для балансовых шансов по всему геному. Распределение длин ROH моделируются как смесь трех гауссовых распределений, классифицирующих ROH по классам размеров: очень короткие ROH (от десятков до сотен килобайт), отражающие паттерны LD; промежуточный ROH (от сотен килобаз до 2 мегабаз), который появляется в результате фонового родства вследствие генетического дрейфа; и длинные ROH (свыше 1–2 Mb), происходящие из недавней родительской связанности (Pemberton et al., 2012).

Следует отметить подходы, в которых используется картирование гомозиготности для выявления генов редких рецессивных заболеваний (Pirrucci et al., 2014). Для картирования гомозиготности требуется оценка аутозиготной для каждого индивидуума доли генома, на основе которой вычисляется оценка LOD («логарифм шансов») для связи с указанным локусом. Точная оценка аутозиготности имеет решающее значение: недооценка приводит к завышенной оценке LOD и, таким образом, к ложному свидетельству сцепления (Magi et al., 2014); переоценка приводит к ложно-отрицательным результатам. Метод, заложенный в Garlic, вычисляет коэффициент лог-вероятности для  $k$ -го SNP  $i$ -го индивидуума наблюдаемого генотипа  $G_{ik}$  при гипотезах аутозиготности и неаутозиготности, включающего в себя предполагаемую частоту ошибок генотипа и популяционно-специфические частоты аллелей.



### **Методы, позволяющие определить короткие последовательности.**

Полногеномное секвенирование (WGS, Whole genome sequencing) позволяет получить данные высокого разрешения, на основе которых могут быть идентифицированы даже самые короткие ROH. Следует отметить, что частота ошибок при этом намного выше, чем для данных, полученных на основе панелей SNP. Методы HMM для данных последовательностей реализованы в программном обеспечении BCFtools/ROH (Narasimhan et al., 2016). Указанная программа обеспечивает низкую частоту появления ошибок и может использовать небольшие файлы (VCF, Variant Call Format), которые содержат только генотипы SNP и показатели качества. Посредством WGS можно наиболее точно обнаружить ROH и оценить вклад очень коротких ROH в депрессию инбридинга.

### **1.5 Варианты числа копий (CNV)**

Одним из самых сложных аспектов современных исследований генома является определение связи между генетической изменчивостью и фенотипами. В исследованиях генетической архитектуры значимых для разведения признаков сельскохозяйственных животных основной акцент делается на анализе SNP. Однако в последнее время стали вызывать интерес структурные варианты, в частности варианты числа копий (copy number variants, CNV). CNV определяется как область ДНК размером около 1000 п.н., состоящая из инверсий, сбалансированных транслокаций или геномных дисбалансов (вставок и делеций). По сравнению с SNP, CNV охватывают более широкие хромосомные области и потенциально могут отвечать за изменения в структуре генов, модификации в регуляции и приводить к значительным фенотипическим эффектам.

CNV могут влиять на экспрессию генов, а также на регуляцию, с потенциально большими фенотипическими эффектами. Несколько исследований CNV на людях показали связь с менделевскими заболеваниями

и сложными генетическими расстройствами, такими как шизофрения (The International Schizophrenia Consortium, 2008), рак (Shao et al., 2019; Shlien, Malkin et al., 2009) и различные врожденные дефекты (Hilger et al., 2020). Аналогичным образом в животноводстве все больше исследований демонстрируют, что CNVs оказывают причинно-следственные эффекты на фенотипическую изменчивость, например, CNVs в интроне 1 SOX5, вызывая фенотип горохового гребешка у кур (Wright et al., 2009), интронная дупликация 4,6 п.н. в STX17 связана с поседением волос и меланомой у лошадей (Sundström et al., 2012), дупликация FGF3, FGF4, FGF19 и ORAOV1 приводит к расчесыванию волос и предрасположенности к дермоидным синусным кистам у риджбеков (Salmon et al., 2007), а CNV и миссенс-мутации гена сигнального белка агутти (ASIP) приводят к различной окраске шерсти у коз (Fontanesi et al., 2009). Однако в настоящее время эта тема недостаточно изучена у свиней. KIT – первый ген свиньи, для которого было доказано, что дупликация гена и мутация сплайсинга, приводящая к пропуску экзона 17, отвечают за доминирующий белый фенотип и клетки периферической крови (количество и показатели эритроцитов, общее и дифференциальное количество лейкоцитов, уровни гематокрита и гемоглобина и компоненты сыворотки) (Sun et al., 2020; Chen et al., 2012; Johansson et al., 2005). CNV были связаны с несколькими фенотипами у свиней, такими как цвет шерсти (Rubin et al., 2012), толщина шпика (BF) (Fowler et al., 2013; Zappaterra et al., 2020) и качество мяса (Wang et al., 2015), что демонстрирует, что CNV можно считать перспективными маркерами экономически важных признаков.

Для обнаружения CNV использовали различные методы, начиная от надежных цитогенетических подходов, таких как кариотипирование и флуоресцентная гибридизация *in situ*, до прогнозирования CNV *in silico* по всему геному. В последнее время достигнуты значительные улучшения в точности и производительности идентификации CNV. Использование микрочипов и технологий секвенирования следующего поколения позволяет сравнивать CNV среди популяций в масштабе всего генома. Кроме того,

методы секвенирования открывают возможность идентификации CNV в сложных областях генома (Zhou et al., 2016). Для обнаружения сигнатур CNV у сельскохозяйственных животных, включая свиней, наиболее распространенным подходом является анализ на основе данных чипа SNP (Zhou et al., 2018). В результате были разработаны и опубликованы пакеты программного обеспечения для прогнозирования CNV, такие как PennCNV (Wang et al., 2007), QuantiSNP (Colella et al., 2007), GADA (Pique-Regi et al., 2010) и др.

У свиней, как и у многих других видов, все еще наблюдается недостаточный уровень знаний о свойствах CNV. Из всех тем, связанных с CNV, знания об их функциональных эффектах являются наиболее ограниченными. Несмотря на широкий диапазон числа и размера CNV, о которых сообщалось в предыдущих исследованиях, результаты показали, что большая часть CNV, вероятно, является общей для разных пород, а их функциональные эффекты недооцениваются у племенных свиней.

На сегодняшний день в области геномики домашнего скота большинство исследований были сосредоточены только на обнаружении CNV, и лишь немногие исследования были направлены на обнаружение возможных фенотипических CNV у домашнего скота с использованием исследований ассоциаций по всему геному (GWAS) (Zhou et al., 2018). Так, в исследовании Wang и др. (Wang et al., 2015) был проведен GWAS между CNV и признаками качества мяса у свиней. После коррекции частоты ложных обнаружений было идентифицировано в общей сложности восемь CNV на шести хромосомах, которые были значительно связаны по крайней мере с одним признаком качества мяса. Qiu и др. (Qiu et al., 2021) идентифицировали 35 CNVR со значительными ассоциациями с признаками роста и состояния тела с использованием GWAS на основе CNVR. Десять из этих CNVR были связаны как с признаками ADG, так и AGE у свиней породы дюрок, а четыре CNVR показали значительные ассоциации с ADG, AGE и BFT, что можно считать плеiotропной ролью CNVR в регуляции роста свиней и отложения

жира. Согласно результатам Ding и соавторов (Ding et al., 2022), анализ ассоциаций выявил 10 CNVR, один из которых был связан с площадью мышц поясницы, один с глубиной мышц поясницы и восемь с процентом постного мяса (LMP) у свиней породы дюрок.

### **1.6 Исследования ассоциаций по всему геному (GWAS)**

Наибольшую популярность для исследования генетической изменчивости в контексте генотип–фенотип на сегодняшний день имеют исследования ассоциаций по всему геному (GWAS). Они позволяют выявлять статистически значимые связи между генетическими вариантами и фенотипическими признаками, включая заболевания, продуктивность и адаптивные особенности. Эти исследования являются мощным инструментом для понимания наследственных механизмов и разработки новых методов диагностики и селекции.

Исследования ассоциаций по всему геному (GWAS) проверяют сотни тысяч генетических вариантов в геноме, чтобы определить статистически связанные варианты с определенным признаком или заболеванием (Uffelmann et al., 2021). Эта методология породила множество надежных ассоциаций для ряда признаков и заболеваний, и ожидается, что количество связанных вариантов будет неуклонно расти по мере увеличения размеров выборки GWAS. В общем виде модель для GWAS представляет линейную смешанную модель  $Q + K$  (Kennedy et al., 1992). Структура популяции контролируется за счет эффектов субпопуляции как фиксированные коварианты ( $Q$ ), а скрытое родство учитывается случайным членом с матрицей родства, определяющей генетическую ковариацию между особями ( $K$ ). Тестируемый эффект маркера моделируется как фиксированный параметр, а его значимость оценивается с помощью теста отношения правдоподобия (Lippert et al. 2011). Для уменьшения вычислительной нагрузки компоненты дисперсии оцениваются только один раз в «нулевой модели», без включения какого-либо эффекта

маркера, а затем они фиксируются на протяжении всего теста для всех маркеров.

Модели для полногеномных ассоциативных исследований должны обрабатывать многомерные данные и учитывать сложность генетических взаимодействий. Распространенные подходы включают линейные модели, такие как регрессия Лассо и регрессия Риджа для выбора и регуляризации признаков, а также передовые 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.

Регрессия Лассо, или регрессия с наименьшим абсолютным сжатием и оператором выбора, — это тип линейной регрессии, которая включает регуляризацию для улучшения производительности и интерпретируемости модели. Она добавляет штрафной член к функции потерь, в частности сумму абсолютных значений коэффициентов регрессии, известную как регуляризация L1. Этот метод штрафа помогает сжать некоторые коэффициенты до нуля, эффективно выполняя выбор признаков. Регрессия Лассо особенно полезна в наборах данных с признаками, обладающими мультиколлинеарностью, т.е. сильно коррелирующих. Сжимая нерелевантные коэффициенты признаков до нуля, она снижает сложность модели и повышает интерпретируемость. Сила регуляризации контролируется параметром настройки лямбда ( $\lambda$ ), который определяет степень штрафов. Большие значения лямбда приводят к обнулению большего количества коэффициентов, что приводит к более разреженной модели. Одним из главных преимуществ Лассо является его способность обрабатывать многомерные наборы данных и автоматически выбирать наиболее важные признаки. Однако регрессия Лассо чувствительна к масштабированию признаков, что требует надлежащей предварительной обработки данных.

Гребневая регрессия (RR), также известная как регуляризация Тихонова, является вариантом линейной регрессии, которая решает проблему

мультиколлинеарности и переобучения в наборах данных путем добавления штрафа за регуляризацию к модели. Использование гребневой регрессии в количественной генетике было впервые предложено Whittaker и др. (Whittaker et al., 1999). С тех пор гребневая регрессия, наряду с лассо-регрессией, завоевала большую популярность для реализации полногеномных исследований у людей и животных. В отличие от регрессии Лассо, регрессия гребня применяет регуляризацию L2, которая штрафует сумму квадратов коэффициентов. Этот штраф уменьшает величину коэффициентов регрессии до нуля, но никогда не заставляет их становиться равными нулю.

Модель гребневой регрессии минимизирует остаточную сумму квадратов (RSS) с помощью добавленного штрафного члена, пропорционального квадрату коэффициентов, контролируемого параметром настройки лямбда ( $\lambda$ ). Гребневая регрессия особенно эффективна, когда большинство предикторов вносят вклад в переменную отклика, поскольку она сохраняет все переменные в модели, одновременно уменьшая влияние менее важных. Это делает ее надежным выбором для обработки наборов данных с мультиколлинеарностью, где предикторы сильно коррелируют. Уменьшая дисперсию оценок, она улучшает предсказательную эффективность модели на невидимых данных. Однако гребневая регрессия не выполняет отбор признаков, поскольку все коэффициенты сокращаются, а не исключаются. Это делает ее менее подходящей, когда целью является определение небольшого подмножества важных предикторов.

### **1.7 Алгоритмы машинного обучения**

В последнее время применение моделей машинного обучения приобрело большой интерес из-за их огромной гибкости и способности улавливать закономерности в больших «зашумленных» наборах данных (Sigala et al., 2024). Примеры приложений машинного обучения в более широкой области омикс варьируются от идентификации последовательностей

ДНК (сайтов сплайсинга (Kevin et al., 2025), промоторов (Uddin et al., 2024), энхансеров (Smith et al., 2023), позиционирования нуклеосом (Hackett et al., 2024), таксономической аннотации (Mathieu et al., 2022), микробного энтеротипирования (Costea et al., 2018), обучения ошибкам последовательности (Callahan et al., 2016), классификации участков тела микробного хозяина и субъекта (Statnikov et al., 2013), прогнозирования вирусов (Hie et al., 2021), оценки трехмерной структуры белка (Ramakrishnan et al., 2023), вывода эволюционной популяционной генетики (Huang et al., 2023) и геномного отбора (Moeinizade et al., 2021).

Методы машинного обучения не требуют делать предположения о генетическом механизме, лежащем в основе рассматриваемого признака, например, об аддитивности эффектов, количестве и размере взаимодействий, глубине взаимодействий и т.д. В целом алгоритмы машинного обучения вполне эффективно могут справляться и с задачами прогнозирования, и с исследованиями генетической архитектуры признаков (David et al., 2022).

Основная цель алгоритмов машинного обучения – определить функцию, которая предсказывает неизвестный фенотип на основе выборочных данных генотипа (Mieth et al., 2016). Основной причиной принятия алгоритмов машинного обучения является то, что они хорошо подходят для разработки предсказательных моделей, когда количество признаков больше количества образцов. Так, наборы данных массива SNP генерируются для регулярного статистического тестирования в GWAS, и в результате только некоторые варианты могут пройти строгий уровень значимости по всему геному (например,  $p < 10^{-8}$ ).

Стандартной характеристикой результатов GWAS является то, что количество атрибутов ( $p$ ) значительно превосходит количество точек выборки ( $n$ ). Обычно это описывается как «проклятие размерности» или проблема большого  $p$  и малого  $n$ . В идеале это проблема для классической многомерной регрессии. Еще одно существенное различие между обычными фактическими стратегиями и методами машинного обучения заключается в том, что методы

машинного обучения не ожидают предположений о наследственных компонентах атрибута, на который ссылаются, например, аддитивность эффектов, количество и размер взаимодействий и область взаимодействий (Grinberg et al., 2020). По сравнению с анализом GWAS, модели машинного обучения предоставили лучшие средства изучения многолокусных генетических вариантов, а также их взаимодействий, которые предсказывают сложные признаки (Okser et al., 2014).

Алгоритмы машинного обучения, такие как Случайный лес (Random Forest), Машина градиентного спуска (Gradient Boosting Machine (GBM), Экстремальный градиентный спуск (Extremal Gradient Boosting (XGBoost)), имеют большую популярность в аспекте работы с геномными данными.

Случайный лес (Random Forest) – это метод ансамблевого обучения, используемый для задач классификации и регрессии. Он строит несколько деревьев решений во время обучения и объединяет их результаты для повышения точности прогнозирования и снижения переобучения. Алгоритм ансамблевого обучения работает путем построения нескольких деревьев решений во время фазы обучения и объединения их выходов для повышения точности и стабильности. Он вводит случайность на двух уровнях: путем выборки обучающих данных с заменой (бутстреп-выборка) и путем выбора случайного подмножества признаков для каждого разбиения в дереве. Эта случайность гарантирует, что деревья будут разнообразными, и уменьшает вероятность переобучения.

Для построения модели случайного леса необходимо определить ключевые параметры, к которым относятся количество деревьев ( $n$  оценщиков), определяющих размер ансамбля, и максимальная глубина каждого дерева (`max_depth`), контролирующая сложность отдельных деревьев. Кроме того, размер случайного подмножества признаков (`max_features`) используется для разбиения узлов, регулируя балансировку дисперсии и смещения. Другие гиперпараметры, такие как `min_samples_split` и `min_samples_leaf`, контролируют рост деревьев.



Окончательный вывод модели достигается путем усреднения предсказаний по всем деревьям для задач регрессии или использования голосования большинства для задач классификации. Такое агрегирование снижает переобучение, ослабляя влияние шума любого отдельного дерева. Случайный лес ценится за свою надежность, способность обрабатывать многомерные данные и встроенный рейтинг важности признаков, который помогает понять вклад отдельных предикторов. Алгоритм Случайный лес показал свою эффективность при обработке больших наборов данных со многими признаками, включая категориальные и непрерывные переменные, а также может фиксировать нелинейные связи. Однако тщательная настройка гиперпараметров с помощью таких методов, как перекрестная проверка, имеет важное значение для оптимизации его производительности. Модели Случайного леса могут быть вычислительно интенсивными и менее интерпретируемыми по сравнению с более простыми моделями, поскольку природа ансамбля затрудняет понимание отдельных прогнозов. Несмотря на это, данный метод широко используется в таких областях, как генетика и геномика.

Методология RF нашла применение для исследования генетической архитектуры различных заболеваний у людей, таких как рассеянный склероз, диабет, гипертония, рак молочной железы и др. (Kim et al., 2009; Goldstein et al., 2010; Chung, Chen, 2012; Sinoquet, 2018; Botta et al., 2014). Модели RF успешно применяют в исследовании генетической архитектуры сельскохозяйственных животных. В работе Alves с соавторами (Alves et al., 2022) представлена реализация GWAS с использованием подхода RF у крупного рогатого скота породы Неллор. В результате сканирования генома на основе RF и функционального анализа они определили геномные регионы, охватывающие гены-кандидаты с ключевыми функциями в фертильности, включая предимплантационную стадию и развитие эмбриона, жизнеспособность эмбриона, созревание мужских зародышевых клеток и распознавание феромонов. Кроме того, в работе обращено внимание на то, что

многие высокоранговые маркеры в GWAS на основе RF не демонстрируют сильного предельного линейного эффекта, но потенциально вовлечены в эпистатические «горячие точки» между геномными регионами в разных аутосомах.

Исследование ассоциаций по всему геному и прогнозирование фенотипов репродуктивных признаков у свиней крупной белой породы было проведено Zhang с соавторами (Zhang et al., 2024). В рамках этой работы протестированы модели машинного обучения, включая RF, деревья решений с градиентным усилением Gradient Boosted Decision Tree (GBDT) и легкая машина градиентного спуска (LightGBM). Несмотря на то, что RF показал хорошие результаты, по точности прогнозирования он уступил GBDT и LightGBM. Для исследования осуществимости и надежности реализации геномного предсказания в работе Xiang с соавторами (Xiang et al., 2023) были использованы различные алгоритмы машинного обучения, включая случайный лес (RF), опорную машину векторов (SVM), экстремальный градиентный спуск (XGBoost) и алгоритмы сверточной нейронной сети (CNN). Все анализы были обработаны на двух реальных наборах данных свиней: опубликованном наборе данных свиней PIC и наборе данных, содержащем данные из национального нуклеуса свиней в Чифэне, Северный Китай. В качестве изучаемых фенотипов были выбраны два признака: среднесуточный прирост и количество поросят при рождении. В результате они сделали вывод, что RF демонстрирует надежность в генетическом скрининге локусов и геномном отборе, но эффективность его ниже относительно других нелинейных моделей.

В целом RF – надежный базовый метод, он эффективен в обработке многомерных данных и сложных генетических взаимодействий и подходит для исследования генетической структуры. Однако его предсказательная точность часто зависит от конкретного анализируемого признака и структуры набора данных. В связи с этим для задач, требующих более высокой точности прогнозирования или более тонкой интерпретации модели, более

эффективными являются альтернативные модели машинного обучения, такие как машина градиентного усиления или экстремальное градиентное усиление.

Две широко используемые модели машинного обучения для анализа геномных данных в контексте генотип–фенотип – это машина градиентного усиления (GBM) и экстремальное градиентное усиление (XGBoost). Эти модели доказали свою высокую эффективность для задач прогнозирования, особенно при работе с большими и сложными наборами данных в исследованиях ассоциаций по всему геному.

Машина градиентного усиления (GBM) – это метод ансамблевого обучения, который строит модели последовательно. Он начинается с простой модели, обычно дерева решений, и итеративно улучшает ее, минимизируя ошибки прогнозирования. На каждом этапе GBM фокусируется на точках данных, которые были плохо предсказаны в предыдущих итерациях, эффективно «повышая» производительность модели. GBM использует неглубокие деревья решений в качестве базовых учеников. Каждое дерево фокусируется на изучении остатков (ошибок) предыдущей модели. Метод оптимизирует предопределенную функцию потерь (например, среднеквадратичную ошибку для регрессии) с помощью градиентного спуска. Каждое новое дерево обучается для минимизации градиента функции потерь относительно прогнозов модели. Прогнозы обновляются итеративно путем добавления выходных данных нового дерева, масштабированных по скорости обучения для управления вкладом каждого дерева. Модель легко настраивается, позволяя изменять такие параметры, как количество деревьев, скорость обучения и глубину дерева.

Экстремальный градиентный спуск (XGBoost) – это усовершенствованная реализация GBM, разработанная для скорости и эффективности. Модель на основе XGBoost включает в себя дополнительные методы, такие как регуляризация (для предотвращения переобучения) и параллельная обработка (для сокращения времени вычислений), что делает ее более подходящей для обработки крупномасштабных наборов данных и

данных высокой размерности. Улучшенная оптимизация в XGBoost достигается за счет использования производных второго порядка (информацию Гессе) функции потерь. Регуляризация включает L1 (лассо) и L2 (гребень) для уменьшения переобучения и улучшения обобщения.

Встроенные функции в XGBoost позволяют в некотором роде обрабатывать пропущенные данные, т.е. управлять пропущенными значениями, автоматически обучаясь тому, как с ними обращаться во время обучения. В отличие от GBM, который выращивает деревья до фиксированной глубины, XGBoost использует стратегию «максимальной глубины» в сочетании с обрезкой для обеспечения оптимального размера дерева. Кроме того, модель предлагает более точное управление гиперпараметрами, что позволяет исследователям лучше адаптировать модель к своему конкретному набору данных, и она предназначена для параллельной обработки деревьев, что делает ее значительно быстрее.

Модели машина градиентного спуска, экстремальный градиентный спуск и усовершенствованные варианты градиентного спуска эффективно применяются в исследованиях ассоциаций на уровне генома для выявления сложных генетических взаимодействий и повышения точности прогнозирования фенотипа (Bentéjac et al., 2021). В исследовании Guo с соавторами (Guo et al., 2021) был представлен GGInt-XGBoost – метод, использующий XGBoost для обнаружения взаимодействий генов в исследованиях случай–контроль. Подход был проверен с помощью моделирования и применен к данным ревматоидного артрита, успешно определив значимые взаимодействия генов. В работе Gill с соавторами (Gill et al., 2022) сравнили модели машинного обучения, включая XGBoost, с архитектурами глубокого обучения для предсказания фенотипов на основе генотипических данных в соевых бобах. Модели XGBoost и случайного леса превзошли модели глубокого обучения, продемонстрировав превосходную точность.

Эффективность трех ансамблевых методов RF, GBM и XGBoost для геномного прогнозирования племенной ценности крупного рогатого скота породы Брахман показало, что RF и GBM были эффективны в идентификации SNP, связанных с признаками роста (Li et al., 2018). Точность геномного прогнозирования с использованием 3000 лучших SNP, идентифицированных GBM, была сопоставима с использованием всей панели SNP (0,43), что подчеркивает потенциал GBM в геномной селекции.

Таким образом, GBM и XGBoost – мощные инструменты для прогнозирования сложных признаков и выявления закономерностей в геномных данных. В то время как GBM проще и эффективнее для небольших наборов данных, XGBoost предлагает повышенную производительность, масштабируемость и универсальность для более крупных и сложных наборов данных. Выбор правильной модели зависит от размера набора данных, вычислительных ресурсов и сложности решаемой задачи.

### **1.8 Искусственные нейронные сети**

Традиционные методы машинного обучения имеют ограничения, связанные с возможностями анализировать необработанные входящие данные, в связи с чем возникла необходимость разработать методы, позволяющие машине, получая необработанные данные, автоматически находить необходимые связи и несоответствия. Эти методы получили название обучение представлению.

Глубокое обучение – это подмножество методов обучения представлению с несколькими уровнями, полученные путем составления простых нелинейных модулей, преобразующих данные на одном уровне (начиная с необработанного ввода) и представляя их на более высоком уровне. При составлении достаточного количества таких преобразований модель обучается очень сложным функциям.

Во многих приложениях глубокого обучения используется архитектура нейронных сетей прямого распространения, обучение которых состоит из процесса отображения входных данных фиксированного размера в выходные данные фиксированного размера. Для перехода от одного слоя к другому рассчитывается взвешенная сумма входных данных из предыдущего слоя и передается результат через нелинейную функцию. В качестве нелинейной функции применяется линейный выпрямитель (ReLU), который обучается намного быстрее в сетях со многими слоями, что позволяет обучать глубокую контролируемую сеть без неконтролируемого предварительного обучения (Glorot et al., 2011).

Модель искусственной нейронной сети прямого распространения (ANN), также известная как глубокая нейронная сеть (DNN) или многослойный персептрон (MLP), является наиболее распространенным типом глубокой нейронной сети. Кроме того, все больший интерес привлекают другие типы глубоких нейронных сетей, такие как сверточные нейронные сети (CNN) и рекуррентные нейронные сети (RNN). В целом MLP хорошо работают с транзакционными (табличными) данными, при работе с изображениями наилучшим решением является применение CNN. При работе с последовательными данными лучшим выбором являются RNN.

Модели на основе глубокого обучения делают большие успехи в решении проблем, которые на протяжении многих лет нельзя было реализовать посредством искусственного интеллекта. Они применимы во многих областях науки, бизнеса и государственного управления. Так, модели на основе глубокого обучения превосходили другие методы машинного обучения в прогнозировании активности потенциальных молекул лекарств (Ma et al., 2015), анализе данных ускорителей частиц (Ciodaro et al., 2012; Kaggle, 2014), реконструкции мозговых цепей (Helmstaedter et al., 2014) и прогнозировании воздействия мутаций в некодирующей ДНК на экспрессию генов и заболевания (Leung et al., 2014, Xiong et al., 2015).

Кроме того, существуют различные успешные приложения DL для анализа геномных данных. Большой потенциал глубокое обучение показало для прогнозирования количественных признаков в программах разведения растений и животных. В исследовании Karansher с соавторами (Karansher et al., 2021) сравнили производительность двух моделей DL (многослойного персептрона и сверточной нейронной сети) с гребневой регрессией для прогнозирования пяти различных признаков яровой пшеницы с различной наследуемостью и генетической архитектурой, включая урожайность зерна, содержание белка в зерне, дату колошения, высоту растения и массу. Результаты показали, что модели на основе DL превосходили в прогнозировании по всем признакам. В работе Lozada с соавторами (Lozada et al., 2023) сравнивали точность прогнозирования для агрономических признаков перцев чили из Нью-Мексико. В результате они сделали вывод, что ни одна модель не была лучшей по всем признакам. Байесовская регрессия имела самую высокую среднюю точность для даты первого стручка и общего урожая с растения. Многослойный персептрон был самым лучшим для времени цветения и высоты растения, тогда как геномная модель BLUP имела самую высокую точность для ширины растения.

Модели DL особенно эффективны в многопризнаковых и многосредовых сценариях, где они могут использовать корреляции между признаками для повышения точности прогнозирования. Это позволяет повысить точность, особенно когда признаки умеренно или сильно коррелируют. В работе Montesinos-López с соавторами (Montesinos-López et al., 2018) изучали применение моделей многопризнакового, многосредового глубокого обучения (MT-ME DL) для геномного прогнозирования признаков растений. В результате модели MT-ME DL продемонстрировали превосходную точность прогнозирования по сравнению с моделями с одним признаком и традиционными методами геномного прогнозирования, особенно при включении взаимодействий с окружающей средой ( $G \times E$ ). Используя корреляции признаков и взаимодействия с окружающей средой, эти модели

обеспечивают практичный и эффективный подход и предлагают значительные преимущества в точности и масштабируемости.

Явными преимуществами при обработке данных, особенно изображений, обладают сверточные нейронные сети (CNN). Сверточные нейронные сети используют локальную связность и распределение веса, что значительно сокращает количество параметров по сравнению с полностью связанными архитектурами. Это делает их более эффективными при обнаружении пространственных закономерностей, что приводит к более высокой производительности в таких задачах, как анализ изображений, где локальные особенности (края, текстуры) имеют решающее значение. Особенностью вычислений в модели CNN являются фильтры (также называемые ядрами) к небольшим областям входных данных, выполняющие свертку для обнаружения локальных особенностей, таких как края или формы. Благодаря объединению слоев CNN уменьшают размерность промежуточных карт признаков, сосредотачиваясь на наиболее релевантных закономерностях, сохраняя при этом управляемые вычислительные требования. Их архитектура естественным образом имитирует работу изображений: соседние пиксели часто коррелируют, а фильтры изучают такие особенности, как края, углы и более сложные структуры по мере углубления. Эта локальная связность и инвариантность трансляции делают CNN пригодными для распознавания объектов или шаблонов на изображениях.

Эффективность систем машинного зрения на основе CNN побудила крупные технологические компании, включая Google, Facebook, Microsoft, IBM, Yahoo!, Twitter и Adobe и др., а также быстро растущее число стартапов инициировать научно-исследовательские и опытно-конструкторские проекты по распознаванию изображений, в том числе в здравоохранении и сельском хозяйстве.

Модели CNN позволяют решать задачи, связанные с клеточной классификацией. Разработана модель для классификации и дифференциации линий клеток мыши и человека, а также их радиорезистентных клонов на



основе данных микроскопических изображений линий (Toratani et al., 2018). Кроме того, CNN получили широкое распространение в качестве надежного инструмента для анализа изображений, связанных с заболеваниями, особенно для определения типов опухолей и их развития. Представлен метод для обнаружения рака легких на КТ-снимках, точность которого достигла 92,96% (Shankara et al., 2023, Rehman et al., 2023), опухолей головного мозга на основе анализа МРТ-сканов (Dewage et al., 2024, Xie et al., 2022), определения аномалий опухолей (Parageorgiou et al., 2024).

Поскольку любая геномная оценка базируется на данных фенотипа, большой интерес вызывают исследования, демонстрирующие применение CNN для анализа фенотипов. Vu с соавторами (Vu et al., 2024) использовали изображения КТ-сканов мышц живота. Они показали, что существуют значительные различия в плотности и объеме мышц живота в разных возрастных группах и между полами, подчеркивая закономерности дегенерации мышц и полового диморфизма.

В растениеводстве CNN применяются для точной оценки фенотипов, таких как признаки корней и побегов (Pound et al., 2017), подсчет листьев (Dobrescu et al., 2017; Giuffrida et al., 2018), классификация биотических и абиотических стрессов (Ghosal et al., 2018), подсчет семян в горшке (Uzal et al., 2018), обнаружение колосьев пшеницы (Hasan et al., 2018) и оценка морфологии растений и стадий развития (Wang et al., 2019).

В животноводстве исследования направлены на тестирование эффективности подходов на основе CNN для улучшения мониторинга благополучия и методов управления. Tian с соавторами (Tian et al., 2021) применили сверточные нейронные сети (CNN) для распознавания поведения крупного рогатого скота в различных условиях на ферме. Классификацию поведения проводили на основе наблюдений за кормлением, ходьбой и отдыхом. Их результаты показали высокую точность идентификации и надежное обнаружение поведения в различных условиях на ферме. Kashiha с соавторами (Kashiha et al., 2013) провели исследование по мониторингу свиней

в условиях группового содержания с использованием усовершенствованной системы визуализации и обработки видеоданных с камер, размещенных в свинарниках, что позволило автоматизировать наблюдение как за отдельными животными, так и за общей группой. Исследование установило эффективность систем мониторинга на основе CNN для улучшения управления животными в условиях группового содержания.

Одним из интересных и востребованных приложений CNN в животноводстве может быть анализ морфологии спермы, направленный на повышение точности и эффективности оценки качества. В работе Keller с соавторами (Keller et al., 2014) предложено в качестве фенотипа использовать высокоразрешающие микроскопические изображения спермы хряка для обучения сверточных нейронных сетей (CNN). Предложенная модель достигла высокой точности в идентификации и классификации морфологии спермы и продемонстрировала точную дифференциацию нормальных и аномальных форм спермы.

Сверточные нейронные сети (CNN) применяли для автоматизации определения признаков туши свиньи, используя фенотипические изображения и данные для точных прогнозов (Wei et al., 2024). Для анализа использовали фенотипические изображения, уделяя особое внимание ключевым признакам, таким как длина туши и толщина шпика. Предложенный подход продемонстрировал высокую точность в прогнозировании признаков туши со значительными улучшениями по сравнению с традиционными ручными методами. Nasirahmadi с соавторами (Nasirahmadi et al., 2019) сосредоточились на автоматическом обнаружении случаев кусания хвоста и оценке длины хвоста у свиней. Их результаты показали высокую точность определения поведения кусания хвоста и возможности измерения длины хвоста.

Преобразование генетических данных в искусственные изображения позволяет применить подходы машинного зрения и, в частности, сверточные нейронные сети к задачам генетики и геномики живых организмов. Этот подход, возможно, позволит преодолеть пространственные связи в

генетических данных, которые могли быть упущены из виду традиционными статистическими методами. Непосредственное использование представлений генома в виде изображения и дальнейший его анализ на основе машинного зрения относительно задач генотип–фенотип/заболевания встречается еще крайне редко. Как правило, в такого рода работах решаются задачи, связанные с преобразованием последовательностей геномных данных, анализа транскриптома и других «омных» данных.

Потенциал глубокого обучения для классификации геномных последовательностей, в частности экзон-интронного анализа и в смежных областях был продемонстрирован в работе Ben Nasr Barber с соавторами (Ben Nasr Barber, Elloumi Oueslati, 2024). Целью этой работы была классификация экзонов и интронов в ДНК человека с использованием передовых моделей, таких как ResNet-50, GoogleNet и 13-слойная CNN. Последовательности ДНК были закодированы как числовые данные или визуальные представления (например, прямое кодирование, бинарная матрица или преобразование сигнала). В результате модель ResNet-50 достигла высокой точности классификации; GoogleNet показала сопоставимые результаты; CNN достигла немного меньшей точности, но обеспечила лучшую интерпретируемость и была оптимизирована для данных, специфичных для домена.

Интересный подход анализа на основе изображений генетических данных представлен в работе Chen с соавторами (Chen et al., 2021). Целью их исследования была классификация шизофрении с использованием геномных данных, в частности SNV, выявленных с помощью GWAS и методов глубокого обучения. Для генерации изображений SNV были преобразованы в тепловые карты (наличие либо отсутствие), матрицы фиксированного размера и графовые изображения (визуализация генетических сетей или неравновесного сцепления). При необходимости дополнительно применяли синтетические преобразования искусственных изображений (вращение и переворачивание). В итоге искусственная визуализация изображений показала

более высокие результаты, чем базовое двоичное кодирование и интерпретируемость генетической изменчивости, связанной с шизофренией.

Анализ изображений с помощью моделей CNN позволяет идентифицировать кластеры инфекций и потенциальные источники вспышек (например, ВИЧ). В работе Kupperman с соавторами (Kupperman et al., 2022) предложено определять вспышки ВИЧ в реальном времени, используя генетические данные. Визуальные представления генетических данных, такие как филогенетические деревья и матрицы сходства последовательностей, использовали в качестве входных данных для повышения интерпретируемости модели.

Xu с соавторами (Xu et al., 2020) провели исследование нового подхода к классификации данных секвенирования РНК отдельных клеток. Они использовали в качестве входных данных профили экспрессии генов как изображения и применили к этим изображениям передовые методы извлечения признаков и классификации, т.к. CNN способна захватывать сложные шаблоны в формате 2D. Их результаты продемонстрировали улучшенную точность классификации и надежность и превзошли стандартные методы машинного обучения для данных секвенирования РНК отдельных клеток. Исследование установило эффективность преобразования профилей экспрессии генов в изображения и использования CNN. Эта методология является многообещающим направлением для анализа многомерных биологических данных.

Эффективность глубокого обучения в анализе сложных геномных данных подчеркивает огромный потенциал для научных открытий, в том числе в генетике и геномике сельскохозяйственных животных. Несмотря на это, не существует универсальной модели, «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 то, какая модель работает оптимально. Более того,

классические алгоритмы машинного обучения или даже линейные методы иногда могут превосходить глубокие архитектуры в определенных ситуациях из-за их простоты или лучшего соответствия структуре данных. В связи с этим необходимо проводить исследования, экспериментально сравнивать и определять наиболее подходящие решения для каждого уникального набора 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.

## 2 МАТЕРИАЛЫ И МЕТОДЫ

### 2.1 Материалы

Диссертационная работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования «Донской государственный аграрный университет» (ФГБОУ ВО Донской ГАУ) и на кафедре генетики Академии биологии и биотехнологии им. Д. И. Ивановского федерального государственного автономного образовательного учреждения высшего образования «Южный федеральный университет» (ЮФУ).

Работа проведена в период с 2013 по 2025 г. на базе лаборатории молекулярной диагностики и биотехнологии сельскохозяйственных животных в ФГБОУ ВО Донского ГАУ, кафедры генетики Академии биологии и биотехнологии им. Д. И. Ивановского федерального государственного автономного образовательного учреждения высшего образования «Южный федеральный университет», а также в качестве основного исполнителя грантов РНФ 19-76-10012 «Определение генетической архитектуры репродуктивных признаков чистопородных и гибридных свиноматок с использованием полногеномных ассоциативных исследований» и РНФ 23-76-10009 «Исследование генома свиней породы дюрок в аспекте породообразования, внутривидовой стратификации и интенсивного отбора по селекционно-ценным признакам».

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 ( $n = 73$ ) и 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 свиней крупной белой породы отечественной (КБ1) и импортной селекции (КБ2 и КБ3) были выбраны КБ1 ( $n = 99$ ), КБ2 ( $n = 207$ ) и КБ3 ( $n = 170$ ) соответственно. 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 синтезом данных их количество было увеличено до 5000 голов. Все животные содержались в одинаковых условиях кормления и содержания. 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 крупной белой породы без дефекта ( $n = 364$ ) и с дефектами ( $n = 204$ ), дефекты диагностировали визуально специалисты 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 рассчитывали 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 программе R-4.1.0 (R Core Team, 2019) с использованием функций-оболочек пакета R SambaR (<https://github.com/mennodejong1986/SambaR>). После фильтрации и прореживания 28915 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 рассчитывали с помощью программного обеспечения `Admixture-1.3` (Alexander, Lange, 2011). Неравновесие по сцеплению (LD) рассчитывали с использованием программы `plink`, оценку LD и его визуализацию осуществляли методом `LD decay`. Фазирование генотипов проводили с помощью программы `Beagle` (Browning et al., 2021). Аллель-специфическую расширенную гаплотипическую гомозиготность (ЕНН), интегрированную оценку гаплотипа (iHS), кросс-популяционную расширенную гаплотипическую гомозиготность (XP-ЕНН) и интегрированную гомозиготность гаплотипа (nSL) оценивали с помощью программы `selscan` (Szpiech, 2021).

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` основаны на скрытой марковской модели (HMM), тогда как `GADA` использует модель разреженного байесовского обучения (SBL). Для объединения отдельных CNV в регионы выполняли процедуру `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 были разделены на классы. Карты ROH были сформированы для каждого животного и сохранены в формате «JPEG» со следующими параметрами: ширина 1200 пикселей, высота 1200 пикселей, глубина цвета 24. Карты были сформированы для свиней пород дюрок и крупная белая. На основе созданных карт гомозиготности была составлена модель сверточных нейронных сетей (CNN).

## **2.2 Подготовка данных**

1. Из набора данных WB извлечены все SNP:



```
awk '{print$2}' WB.bim > WB_SNPs.txt
```

2. Из набора данных LLD извлечены варианты, присутствующие в наборе данных WB.

```
./plink1 --bfile LLD --extract WB_SNPs.txt --make-bed --out LLD1
```

3. Из набора данных LLD извлечены все SNP:

```
awk '{print$2}' LLD.bim > LLD_SNPs.txt
```

4. Из набора данных WB извлечены варианты, присутствующие в наборе данных LLD.

```
./plink1 --bfile WB --extract LLD_SNPs.txt --recode --make-bed --out WB1
```

5. Перед объединением данных LLD с данными WB необходимо убедиться, что файлы можно объединить, для этого выполнены три шага.

а) Проверено, что референсный геном похож в наборах данных WB и LLD.

```
awk '{print$2,$5}' LLD1.bim > 1kg_ref-list.txt
```

```
./plink1 --bfile WB1 --reference-allele 1kg_ref-list.txt --make-bed --out  
WB-adj
```

б) Устранены проблемы с цепями.

```
awk '{print$2,$5,$6}' LLD1.bim > LLD1_tmp
```

```
awk '{print$2,$5,$6}' WB-adj.bim > WB-adj_tmp
```

```
sort LLD1_tmp WB-adj_tmp |uniq -u > all_differences.txt
```

```
awk '{print$1}' all_differences.txt | sort -u > flip_list.txt
```

```
./plink1 --bfile WB-adj --flip flip_list.txt --reference-allele 1kg_ref-list.txt  
--make-bed --out corrected_WB
```

```
awk '{print$2,$5,$6}' corrected_WB.bim > corrected_WB_tmp
```

```
sort LLD1_tmp corrected_WB_tmp |uniq -u > uncorresponding_SNPs.txt
```

в) Удалены SNP, которые после предыдущих двух шагов все еще различаются между наборами данных.

```
awk '{print$1}' uncorresponding_SNPs.txt | sort -u >  
SNPs_for_exclusion.txt
```

```
./plink1 --bfile corrected_WB --exclude SNPs_for_exclusion.txt --make-bed --out WB2
```

```
./plink1 --bfile LLD1 --exclude SNPs_for_exclusion.txt --make-bed --out LLD2
```

6. Объединены наборы данных WB с данными LLD.

```
./plink1 --bfile WB2 --bmerge LLD2.bed LLD2.bim LLD2.fam --allow-no-sex --make-bed --out merge_LLDW
```

Плотность SNP для двух наборов данных рассчитывали с помощью пакета HandyCNV (Zhou et al., 2021) с установкой параметров: низкая плотность –  $\leq 20$  SNP/Mb, высокая плотность –  $\geq 50$  SNP/Mb.

Для преобразования данных SNP в форматы PED и MAP использовали PLINK (Purcell et al., 2007) и vcftools (Danecek et al., 2011). Флаги vcftools --depth и --site-depth применяли для расчета глубины прочтения для каждого животного и для каждого SNP.

Двоичные файлы (BED, RAW и BIM) были созданы из файлов PED и MAP с использованием PLINK (Purcell et al., 2007), т.е. с применением флагов --make-bed, --recode A, --chr -set 95 и allow-extra- chr).

## 2.3 Методы

Управление данными SNP и их анализ выполняли в программе R-4.1.0 (R Core Team, 2019) с использованием функций-оболочек пакета R SambaR (Menno et al., 2021).

Данные были импортированы в R и сохранены в объекте genlight с использованием функции read.PLINK пакета R adegenet-2.1.3 (Jombart T. 2008).

Данные были отфильтрованы с использованием функции «filterdata» пакета R SambaR с indmiss = 0,01, snpmiss = 0,01, min\_mac = 2, dohefilter = TRUE и min\_spacing = 500. Было оставлено 255 из 257 особей (44–73 на популяцию).

После фильтрации 29049 из 29320 SNP были сохранены. Этот отфильтрованный набор данных использовали для анализа выбора. Прореживание сократило набор данных еще больше – до 28915 SNP. Этот отфильтрованный и прореженный набор данных использовали для структурного анализа. Затем с помощью данной функции были отфильтрованы SNP на основе доли отсутствующих точек данных с учетом только сохраненных особей.

Обратный порядок – сначала фильтрация SNP, а затем фильтрация индивидуумов – может дать разные оценки недостающих данных и, как следствие, разное количество сохраненных SNP и индивидуумов.

Содержание GC в сохраненном наборе данных составило 0,51, а «трансверсия» соотношение «transitions\_vs\_transversions.filtered» составило 0,82.

### 2.3.1 Анализ структуры

Анализ соответствия (CA) был выполнен с использованием функции «dudi.coa» пакета R ade4-1.7.17 (Bougeard, Dray, 2018).

Анализ главных координат (PCoA) был выполнен с использованием функции «pcoa» пакета R ape-5.5 (Paradis, Schliep, 2018) на матрицах расстояний, содержащих три различные меры генетического расстояния:

- генетическое расстояние Нея, рассчитанное с помощью функции «stamppNeisD» пакета R StAMPP-1.6.2 (Pembleton et al., 2013);
- генетическое расстояние Хэмминга, вычисленное с помощью функции «bitwise.dist» пакета R poppr-2.9.2 (Kamvar et al., 2014);
- $\pi$  (попарное несходство последовательностей), вычисляемое с помощью функции «calcp $\pi$ » пакета R SambaR (Menno et al., 2021).

Анализ главных компонент (PCA) проводили с использованием функции «snpgdsPCA» пакета R SNPRelate-1.26.0 (Zheng et al., 2012).

Многомерное шкалирование (MDS) выполняли с использованием функции «cmdscale» (метрическая MDS) пакета R stats-4.1.0 (R Core Team, 2019) и функции «isoMDS» (неметрическая MDS) пакета R MASS-7.3.54 (Venables, Ripley, 2002) на евклидовой матрице расстояний, сгенерированной с помощью функции «dist» пакета R stats-4.1.0 (R Core Team, 2019).

Вероятности байесовского распределения популяций (BPA) были рассчитаны и построены с использованием функций «assign2pop» и «plotassign2pop» пакета R SambaR (Menno et al., 2021).

Коэффициенты смешивания рассчитывали с помощью функций «obj.snmf» пакета R LEA-3.4.0 (Frichot, Francois, 2014). Альфа была установлена на 10, допуск на 0,00001, количество итераций на 200.

Коэффициенты родословной рассчитывали с помощью программного обеспечения Admixture-1.3 (Alexander D.H., 2011), а графики построены с использованием функции «plotstructure» SambaR (Menno et al., 2021).

### 2.3.2 Анализ генетического разнообразия

Оценки неравновесия по сцеплению (LD) рассчитывали с использованием программы PLINK (-genome --r2 --ld -window-kb 1000000 --ld-window -r2 0).

Анализ HWE, (2D) свернутого спектра частот сайтов (SFS), анализ D Таджимы и анализ гетерозиготности по всему геному выполняли с использованием функции «calcdiversity» пакета R SambaR.

Более конкретно, векторы SFS, специфичные для популяции, были созданы с помощью функции «getfoldedsfs» пакета R SambaR, которая группирует SNP в классы на основе количества копий их минорного аллеля и которая впоследствии вычисляет размер каждого бина (т.е. количество SNP в каждом бине).

Геномный He (genomeHe) рассчитывали для каждого образца по формуле:

$$\text{genomeHe} = (\text{He\_seg} \cdot \text{N\_seg}) / (\text{N\_total}), \quad (2)$$

где  $N_{seg}$  – количество участков, разделяющихся внутри популяции, к которой принадлежит исследуемый индивидуум;  $He_{seg}$  – доля гетерозиготных участков в пределах исследуемой особи для сегрегирующих участков;  $N_{total}$  – общая длина всех генотипированных сайтов (как полиморфных, так и мономорфных), которые прошли настройки фильтра.

### 2.3.3 Анализ выбора

Анализы выбора проводили с использованием пакетов R *fsthet*-1.0.1 (Flanagan, Jones, 2018), *pcadapt*-4.3.3 (Privé et al., 2020) и функции «*gwdsfisher*» *SambaR* (Menno et al., 2021).

### 2.3.4 Демографический анализ

Исторические оценки  $Ne$  были выведены из сложенных векторов SFS с помощью программного обеспечения *stairway\_plot\_v2* (Liu, Fu, 2015).

Скорость миграции оценивали с использованием *BayesAss3-SNP* (Steven et al., 2019). Количество итераций было установлено равным 1000000, Burn-in – 100000, а значения дельты – 0,1. Выходные матрицы были преобразованы в графики потока генов с помощью функции «*plotmigration*» пакета R *SambaR* с использованием пакета R *circlize*-0.4.13.

### 2.3.5 Определение отбора с помощью статистики гаплотипа

Фазирование генотипов проводили с помощью программы *Beagle* (Browning et al., 2021), которая позволяет не вменять генотипы и не требует информации о скорости рекомбинации. Аллель-специфическую расширенную гаплотипическую гомозиготность (ЕНН), интегрированную оценку гаплотипа (iHS), кросс-популяционную расширенную гаплотипическую гомозиготность (XP-ЕНН), *nSL* и интегрированную гомозиготность гаплотипа оценивали с помощью программы *selscan* (Szpiech, Hernandez, 2014).

*Расширенная гомозиготность гаплотипа* (ЕНН) определяется как вероятность того, что две случайно выбранные хромосомы, несущие основной

аллель, являются гомозиготными в заданной окружающей хромосомной области (Sabeti et al., 2002). Она оценивается по образцу путем вычисления гомозиготности хромосомного фрагмента между фокальным маркером и другим маркером по формуле:

$$ENHS_{s,t} = \frac{1}{n_s(n_s - 1)} \sum_{k=1}^{K_{s,t}} n_k(n_k - 1), \quad (3)$$

где  $n_s$  представляет собой число хромосом, несущих основной аллель  $s$ ;  $K_{s,t}$  представляет собой число различных общих гаплотипов;  $n_k$  - число хромосом, относящихся  $k$ -тому общему гаплотипу.

*Интегрированный показатель гаплотипа (iHS)* (Voight et al., 2006) рассчитывали с использованием уравнения:

$$ENH_c(x_i) = \sum_{h \in C(x_i)} \frac{n_{h_2}}{n_{c_2}}, \quad (4)$$

где  $n_h$  - число наблюдаемых гаплотипов типа  $h_2$   $H_c(x_i)$ ;  $n_c$  - число наблюдаемых гаплотипов, несущих основной гаплотип ( $c \in C$ ) для отслеживания распада гомозиготности гаплотипа как для предковых, так и для производных гаплотипов от сайта запроса. Для расчета iHS на сайте мы сначала рассчитываем интегрированную гомозиготность гаплотипа (iHN) для предковых (0) и производных (1) гаплотипов ( $C := \{0, 1\}$ ) с помощью трапецевидной квадратуры:

$$iHN_c = \sum_{i=1}^{|\mathcal{D}|} \frac{1}{2} (ENH_c(x_{i-1}) + ENH_c(x_i)) g(x_{i-1}, x_i) + \sum_{i=1}^{|\mathcal{U}|} \frac{1}{2} (ENH_c(x_{i-1}) + ENH_c(x_i)) g(x_{i-1}, x_i), \quad (5)$$

где  $\mathcal{D}$  - набор маркеров ниже по течению от текущего локуса, так что  $x_i \in \mathcal{D}$  обозначает  $i$ -тый ближайший ниже по течению маркер от интересующего локуса ( $x_0$ ).  $\mathcal{U}$  и  $x_i \in \mathcal{U}$  определяются аналогично для вышестоящих маркеров.  $g(x_{i-1}, x_i)$  дает генетическое расстояние между двумя маркерами. Тогда, iHS вычисляется по формуле  $iHS = \ln \left( \frac{iHN_1}{iHN_0} \right)$ . В selscan iHS вычисляется командой `selscan --ihs --vcf pop1_chr.vcf.gz --map pop1_chr.map --pmap --threads 26 --out ihs_chr`, где `--vcf` принимает фазированный файл генотипов в формате `vcf.gz`, -

--map – файл карты генотипов из PLINK, --rmap – для использования физических расстояний вместо генетических для вычислений iHS и XP-EHH, --threads – количество ядер на компьютере и --out – выходной файл с расчетами iHS.

*Кросс-популяционная расширенная гаплотипическая гомозиготность* (XP-EHH) также была введена Sabeti и др. (Sabeti et al., 2007). Чтобы рассчитать XP-EHH между популяциями A и B на маркере x0, сначала нужно вычислить iHH для каждой популяции отдельно, интегрируя EHH всей выборки в популяции. В selscan для вычислений XP-EHH данные, предоставленные с флагами --hap, --tped или --vcf, соответствуют популяции A, а данные, предоставленные с флагами --ref, --tped-ref или --vcf-ref, соответствуют популяции B.

nSL – это статистика, связанная с iHS, введенная Ferrer-Admetlla и др. (Ferrer-Admetlla et al., 2014). nSL рассчитывается как логарифмическое отношение статистики SL, рассчитанной для предковых и производных пулов гаплотипов. Существенным отличием от iHS является удаление информации о генетической карте.

Поскольку мы не используем поляризованные данные (т.е. мы не знаем, какой аллель является предковым или производным). Если бы мы это делали, в selscan можно было бы применить веса к различным бинам частот аллелей, чтобы проверить, есть ли значительное отклонение в статистике iHS. Выходные файлы из трех функций (iHS, REHH и nSL) нормированы с помощью программы norm из пакета selscan, и все SNP, имеющие значения  $\text{crit} = 1$ , считаются вероятными генотипами, аллели которых подверглись отбору, т.е. прошли порог достоверности.

### 2.3.6 Вызов CNV

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

основаны на скрытой марковской модели (HMM), тогда как GADA использует модель разреженного байесовского обучения (SBL). В PennCNV вводится информация о маркерах SNP в текстовом формате. Файл отчета микрочипов Illumina был переформатирован с помощью `split_illumina_report.pl`. Ложные срабатывания были удалены с помощью `filter_cnv.pl`, что позволило выбрать CNV со стандартным отклонением  $LRR < 0,3$ . QuantiSNP требует исходный файл отчета Illumina в качестве входных файлов. Результаты с логарифмическим коэффициентом ниже 10 были удалены, как рекомендовали создатели программы. Пакет R GADA был взят в качестве третьей программы для вызова CNV. Первым шагом было нахождение наиболее вероятных точек-кандидатов для CNV с использованием функции модели разреженного байесовского обучения (SBL).

Для дальнейшего анализа CNV необходимо объединить отдельные CNV в регионы. Для этой цели выполняли процедуру Reciprocal Overlap (RO) с использованием пакета R CNVRanger v.1.18.0 (Silva et al., 2020), которая объединяет отдельные CNV в регионы с учетом указанного порога перекрытия `ro.thresh`. Для преобразования отдельных CNV в регионы выходные файлы трех методов PennCNV, QuantiSNP и GADA были объединены в один файл. Все CNV длиной более 5 Мб были удалены, т.к. CNV такой длины, скорее всего, будут ложноположительными. Другим критерием отбора CNV для включения в регионы является наличие перекрывающихся событий между двумя алгоритмами. Этот шаг был выполнен с использованием функции `subsetByOverlaps` пакета R GenomicRanges. Данные были преобразованы в файл, содержащий номер хромосомы, начало и конец CNV, название образца и состояние CN. Затем он был преобразован с помощью `GRangesList` в формат, подходящий для анализа данных CNV. Этот формат отображает список образцов, каждый из которых, в свою очередь, имеет свой собственный набор CNV. Затем была использована функция `populationRanges` с параметрами `mode = «RO»` и `ro.thresh = 0.51`, которые обозначают используемый метод Reciprocal Overlap и порог перекрытия между CNV в долях.



### 3 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

#### 3.1 Анализ генетического разнообразия у представителей рода *Sus*

На первом этапе рассмотрены генетические структуры свиней основных коммерческих пород (дюрок, ландрас и крупная белая) и европейских кабанов (Бакоев и др., 2023).

Для изучения генетической структуры популяций использовали анализ главных компонент (PCA). PCA 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 PCA можно визуализировать генетические различия между популяциями, выявляя кластеры, которые отражают историю миграций, изоляции или смешения популяций. По две главные компоненты (PC1 и PC2) и (PC1 и PC3) использовали для построения двумерных графиков, на которых можно наблюдать разделение популяций на основе их генетического сходства или различия, что позволяет изучать генетическую дистанцию между популяциями и реконструировать их эволюционную историю.

Кроме того, PCA помогает идентифицировать генетические маркеры, которые вносят наибольший вклад в различия между популяциями, что важно для понимания адаптации и отбора. Метод также устраняет избыточность данных, связанную с высокой корреляцией между SNP, и фильтрует шум, улучшая качество анализа. Таким образом, PCA является мощным инструментом в популяционной геномике 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 кабанами.

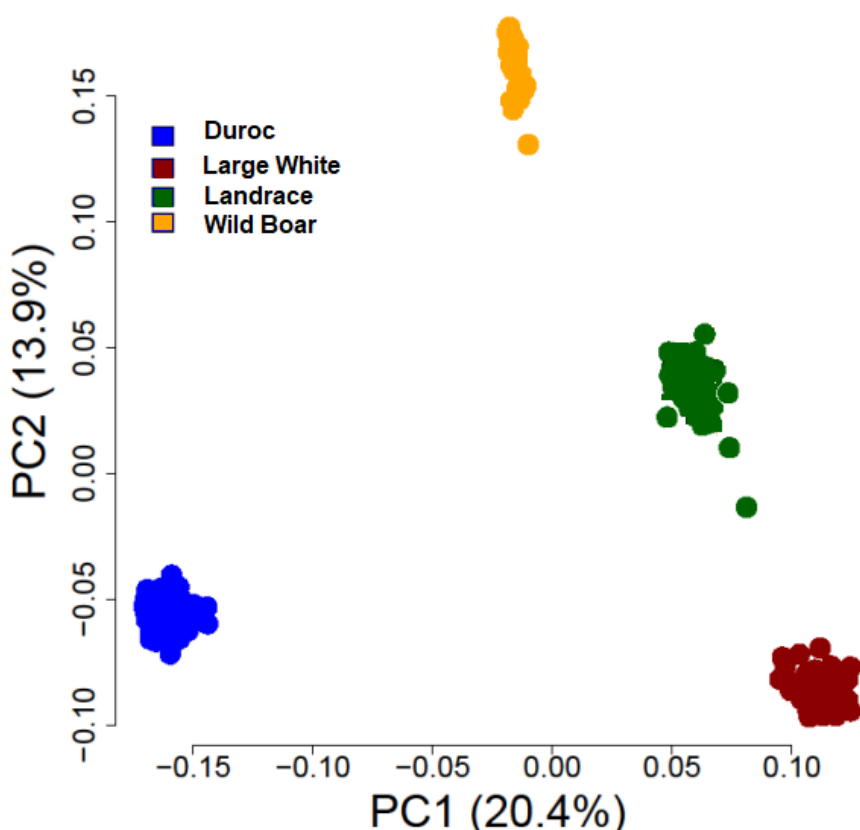


Рисунок 1 – Анализ главных компонент (PCA) для свиней коммерческих пород и диких кабанов. Представлены исследуемые популяции: дюрок (Duroc), крупная белая (Large White), ландрас (Landrace) и дикие кабаны (Wild Boar)

По первой компоненте наибольшая генетическая дифференциация определена между свиньями дюрок и свиньями пород крупная белая и ландрас. Между свиньями коммерческих пород и представителями их диких предков, кабанами, генетическая дифференциация прослеживается по второй компоненте.

Несмотря на явную дифференцированность пород, согласно анализу главных компонент, анализ примесей более детально демонстрирует смешанность пород на генетическом уровне (рисунок 2). На основе анализа с

применением указанных методов можно предположить, что у свиней крупной белой породы и кабанов прослеживается наличие общих примесей.

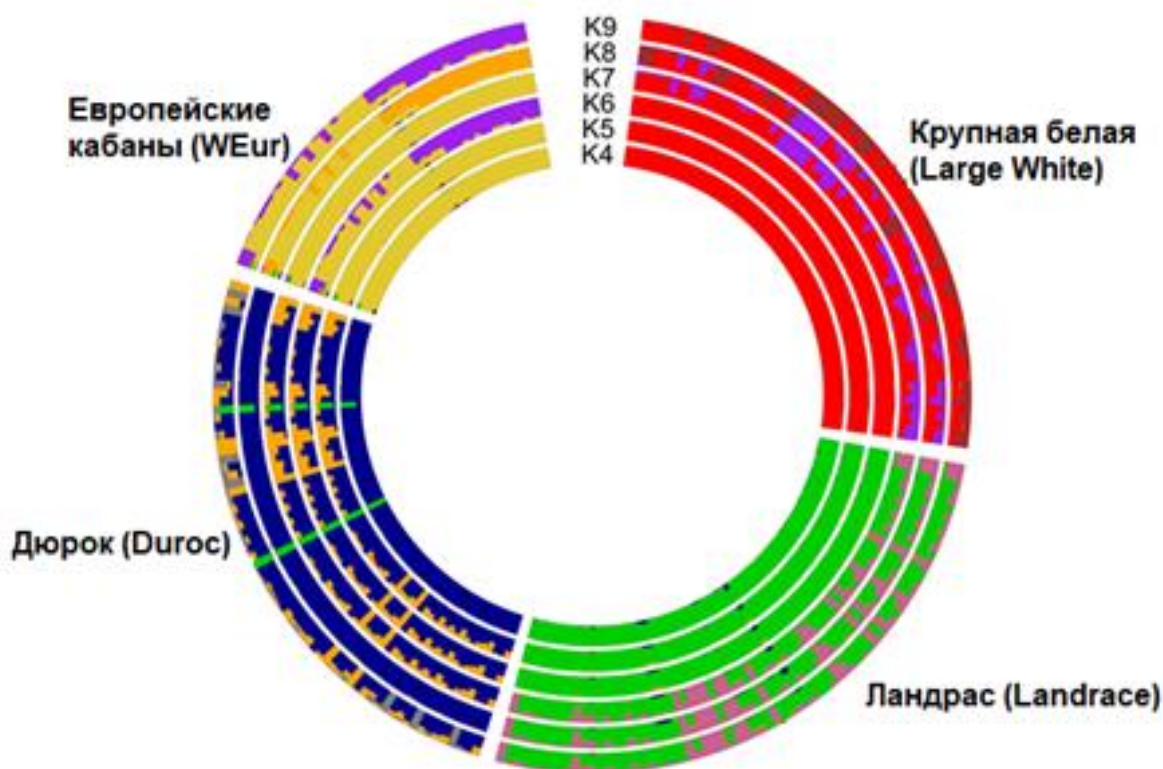


Рисунок 2 – График примесей (Admixture) для свиней коммерческих пород и диких кабанов при количестве кластеров от 4 до 9 ( $K = 4 \dots 9$ )

Следующий этап анализа генетического разнообразия направлен на определение филогенетического взаимоотношения между исследуемыми породами. Реализация этого этапа предполагает применение D-статистики Паттерсона, которая позволяет определить количественную оценку степени генетического обмена между дикими кабанов и современными коммерческими породами свиней, а также построить филогенетическое дерево. Первые два столбца таблицы 1, содержащие соответствующие породы, на основе D-статистики сравнивают, насколько они генетически близки с породой, расположенной в третьем столбце таблицы 1. Положительное значение D означает, что порода 2 (П2) ближе к породе 3 (П3), чем к породе 1 (П1). Отрицательное значение D означает, что П1 ближе к П3, чем П2.

Результаты D-статистики Паттерсона показали, что из трех коммерческих пород наибольшее генетическое сходство с европейскими кабанами имеют свиньи крупной белой породы.

Таблица 1 – Результаты D-статистики Паттерсона для свиней коммерческих пород и диких кабанов

| П1 | П2 | П3 | ВВАА | АВВА | ВАВА | D     | D_sd | nblocks | D_err | D_Z   | p-value  | sign | excess |
|----|----|----|------|------|------|-------|------|---------|-------|-------|----------|------|--------|
| Д  | Л  | КБ | 0,06 | 0,07 | 0,06 | 0,13  | 0,44 | 2452    | 0,01  | 14,55 | 5,87E-48 | **   | ABBA   |
| Д  | Л  | К  | 0,06 | 0,06 | 0,06 | 0,07  | 0,45 | 2452    | 0,01  | 7,18  | 6,96E-13 | **   | ABBA   |
| Д  | КБ | Л  | 0,06 | 0,07 | 0,06 | 0,11  | 0,49 | 2452    | 0,01  | 10,88 | 1,36E-27 | **   | ABBA   |
| Д  | КБ | К  | 0,06 | 0,06 | 0,06 | 0,05  | 0,50 | 2452    | 0,01  | 4,83  | 1,37E-06 | **   | ABBA   |
| Д  | К  | Л  | 0,06 | 0,06 | 0,06 | 0,04  | 0,56 | 2452    | 0,01  | 3,56  | 3,77E-04 | **   | ABBA   |
| Д  | К  | КБ | 0,06 | 0,06 | 0,06 | 0,05  | 0,54 | 2452    | 0,01  | 4,13  | 3,63E-05 | **   | ABBA   |
| Л  | КБ | Д  | 0,07 | 0,06 | 0,06 | -0,02 | 0,45 | 2452    | 0,01  | -2,34 | 1,91E-02 | *    | BABA   |
| Л  | КБ | К  | 0,07 | 0,06 | 0,06 | -0,02 | 0,41 | 2452    | 0,01  | -1,94 | 5,19E-02 | NS   | -      |
| Л  | К  | Д  | 0,06 | 0,06 | 0,06 | -0,02 | 0,55 | 2452    | 0,01  | -2,27 | 2,32E-02 | *    | BABA   |
| Л  | К  | КБ | 0,06 | 0,06 | 0,07 | -0,09 | 0,49 | 2452    | 0,01  | -8,85 | 8,44E-19 | **   | BABA   |
| КБ | К  | Д  | 0,06 | 0,06 | 0,06 | 0     | 0,52 | 2452    | 0,01  | -0,33 | 7,38E-01 | NS   | -      |
| КБ | К  | Л  | 0,06 | 0,06 | 0,07 | -0,07 | 0,49 | 2452    | 0,01  | -7,32 | 2,40E-13 | **   | BABA   |

Примечание: Д – дюрок, Л – ландрас, КБ – крупная белая, К – дикий кабан.

Для оценки нуклеотидного разнообразия рассчитали долю однонуклеотидных полиморфизмов (SNP) относительно общего количества проанализированных нуклеотидных позиций (рисунок 3). Домашние породы свиней демонстрируют относительно низкое генетическое разнообразие. У домашних свиней значения гетерозиготности варьируются от 30 до 35%, а нуклеотидное разнообразие находится в диапазоне 42–46%. 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.

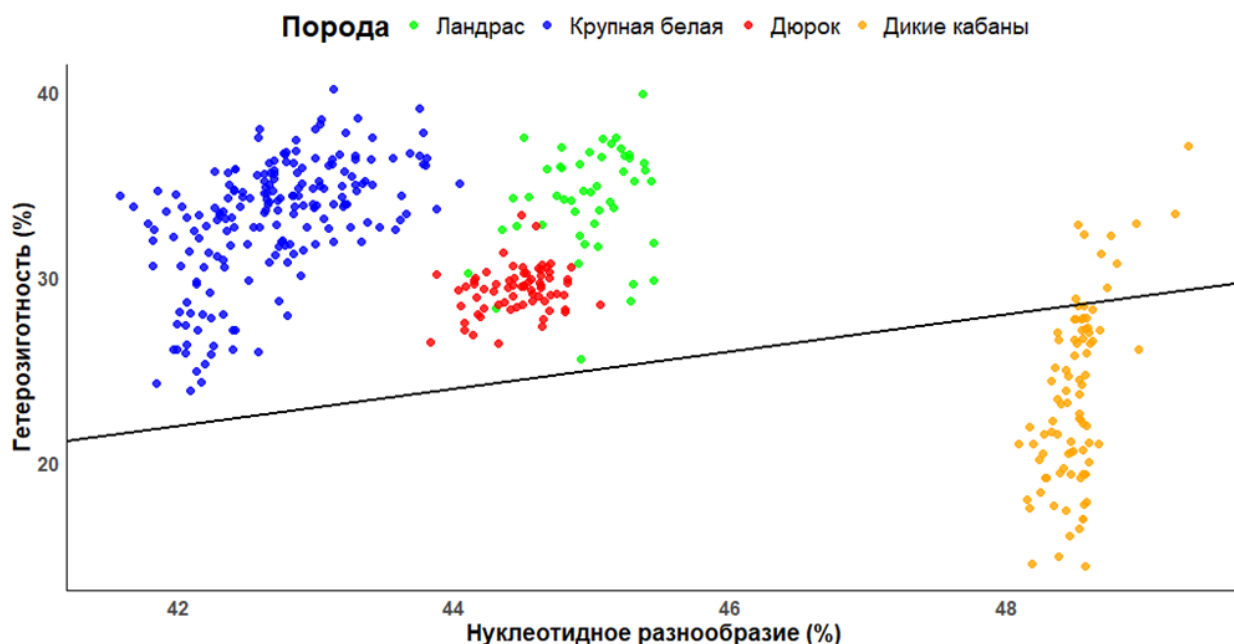


Рисунок 3 – Отношение нуклеотидного разнообразия и гетерозиготности в популяциях

Дикие кабаны, в отличие от домашних свиней, демонстрируют значительно более высокие показатели: гетерозиготность достигает 40%, а нуклеотидное разнообразие – 48%. Это связано с отсутствием искусственного отбора и большим генетическим разнообразием, характерным для естественных популяций. Высокие значения обоих показателей у диких кабанов подчеркивают их генетическую устойчивость и адаптивность к различным условиям среды.

**Плотность SNP.** Для изучения геномного распределения паттернов полиморфизмов ДНК в коммерческих породах и кабанов количество SNP было нанесено на график вдоль каждой хромосомы (рисунок 4). Были обнаружены неслучайные паттерны распределения SNP, при этом на всех хромосомах были обнаружены сильно различающиеся частоты SNP, что является свидетельством того, что полиморфизмы в геноме свиней распределены неслучайно.

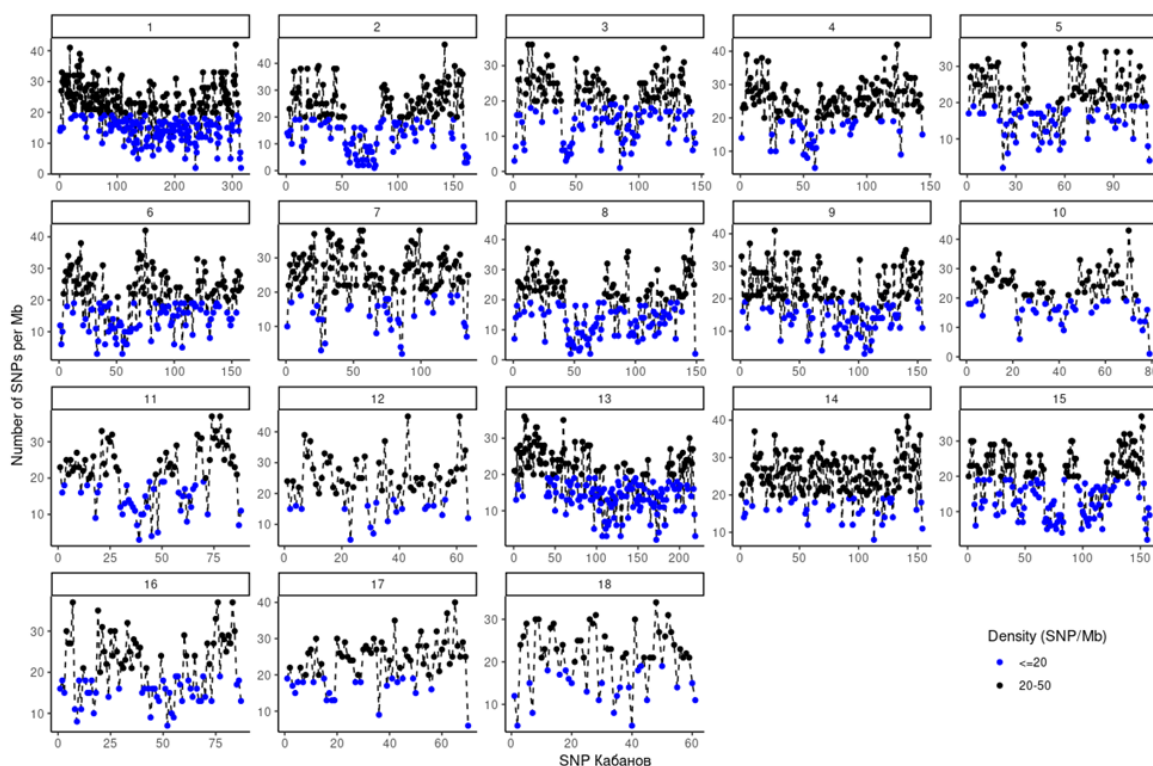


Рисунок 4 – Плотность SNP для диких кабанов

Для диких кабанов SNP распределены в диапазоне 20–50 SNP в 1 Mb 1336 регионов,  $\leq 20$ , соответственно, 1120 регионов со средним значением 20 SNP/Mb и стандартным отклонением  $+ SD = 28$  SNP/Mb и  $- SD = 12$  SNP/Mb. Для пород ландрас, крупная белая и дюрок SNP распределены в диапазоне 20–50 SNP в 1 Mb 1729 регионов,  $\leq 20$  662 регионов и  $\geq 50$ , соответственно, 56 регионов со средним значением 27 SNP/Mb и стандартным отклонением  $+ SD = 38$  SNP/Mb и  $- SD = 16$  SNP/Mb.

Относительно низкие полиморфизмы SNP наблюдались в регионах, выделенных синим цветом (рисунки 4 и 5). Возможные причины относительно низкой плотности SNP перечислены ниже.

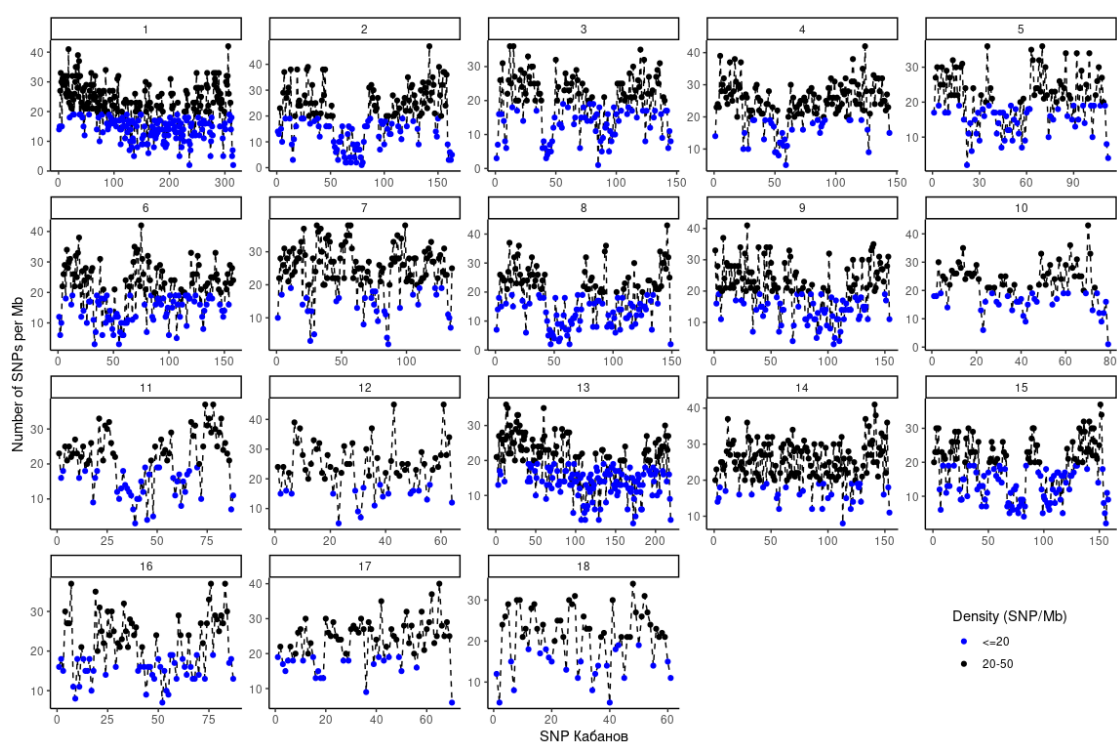


Рисунок 5 – Плотность SNP для свиней коммерческих пород

Низкая плотность SNP может указывать на генетически консервативные области, важные для жизнедеятельности организма, такие как гены или регуляторные элементы, где мутации могут быть фатальными. Это может также отражать эволюционную стабилизацию из-за сильного селективного давления. Технические ограничения, такие как низкое покрытие секвенирования или сложность анализа повторяющихся последовательностей, также могут влиять на выявление SNP.

Высокая плотность SNP, наблюдаемая у отечественных пород, может свидетельствовать о высоком генетическом разнообразии, важном для эволюции и адаптации. Такие области могут включать регуляторные элементы или локусы, связанные с предрасположенностью к заболеваниям. Это также может отражать популяционную структуру или технические аспекты анализа.

Таким образом, распределение SNP отражает как биологические процессы, так и методологические особенности, что важно для понимания генетической архитектуры и эволюционной динамики популяций.

**Неравновесие по сцеплению.** Не менее важным аспектом исследования генетического разнообразия в популяциях является оценка неравновесия по сцеплению. Linkage Disequilibrium (LD) отражает неслучайное наследование аллелей на хромосоме. Анализ LD позволяет оценить генетическую структуру, уровень разнообразия и эволюционные процессы, что важно для понимания адаптации и управления генетическими ресурсами.

Для оценки LD и его визуализации использовали метод LD decay. LD decay (распад неравновесия сцепления) – это мощный инструмент для изучения генетического разнообразия в популяциях. С помощью анализа распада LD можно оценить степень генетической изоляции, миграции, историю популяции, а также понять, как генетическая рекомбинация влияет на структуру генома.

Каждая из этих популяций имеет свои уникальные генетические характеристики, которые можно интерпретировать на основе скорости распада LD. Популяции пород дюрок и ландрас демонстрируют медленный распад LD (рисунок 6), что указывает на сниженное генетическое разнообразие, вероятно, обусловленное инбридингом, эффектом основателя или ограниченным генетическим пулом. Низкая скорость рекомбинации и генетическая изоляция способствуют сохранению ассоциаций между аллелями на больших расстояниях. Популяция диких кабанов, напротив, характеризуется быстрым распадом LD, что свидетельствует о высоком генетическом разнообразии, интенсивном потоке генов и высокой скорости рекомбинации, типичных для крупных и менее изолированных популяций. Популяция породы крупная белая занимает промежуточное положение, с умеренным уровнем генетического разнообразия и рекомбинации, что отражает баланс между генетическим обменом и изоляцией. Эти различия подчеркивают влияние эволюционной истории, размера популяции и уровня изоляции на генетическую структуру.



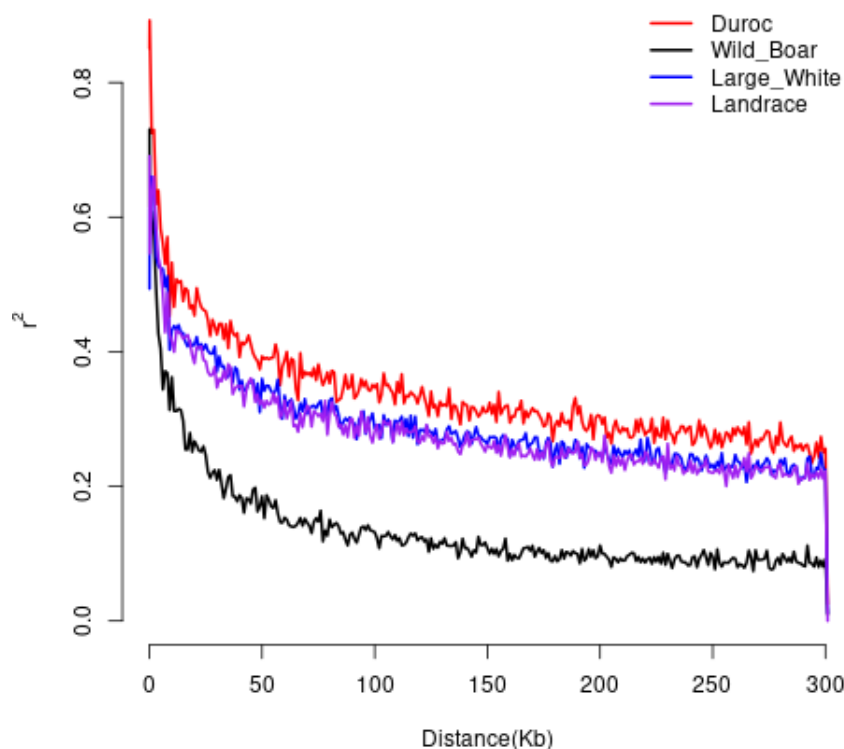


Рисунок 6 – График распада неравновесия по сцеплению (LD decay). Представлены исследуемые популяции: свиньи коммерческих пород дюрок (Duroc), крупная белая (Large White) и ландрас (Landrace), а также дикие кабаны (Wild Boar)

Таким образом, каждая из представленных популяций имеет свои уникальные генетические характеристики, которые отражают их эволюционную историю, уровень генетического разнообразия, скорость рекомбинации и степень изоляции. Эти факторы важны для понимания адаптивного потенциала и генетической структуры популяций.

Не менее важными мерами LD являются взаимная информация (MI) и сниженная взаимная информация (RMI), которые используются в исследованиях LD для количественной оценки взаимосвязи между генетическими вариантами и того, как эти варианты встречаются или изменяются вместе в популяции. В контексте генетики взаимная информация (MI) количественно оценивает количество информации, разделяющее два генетических варианта (SNP). Другими словами, MI измеряет, насколько меньше неопределенности о одном варианте SNP можно получить, если

известен другой вариант SNP. Сниженная взаимная информация (RMI) представляет собой нормализованную версию MI, которая устраняет влияние таких факторов, как размер популяции или различия в частотах аллелей.

Ниже представлен рисунок 7, полученный в результате расчета сниженной взаимной информации (RMI) как мера неравновесия по сцеплению (LD) в популяциях коммерческих пород, который показывает уровень генетической связи между различными локусами в геноме. Значения RMI, отфильтрованные порогом больше 0.45, указывают на сильную связь между аллелями в этих локусах, что может быть связано с ограниченной рекомбинацией или историческими событиями, такими как эффект основателя или инбридинг.

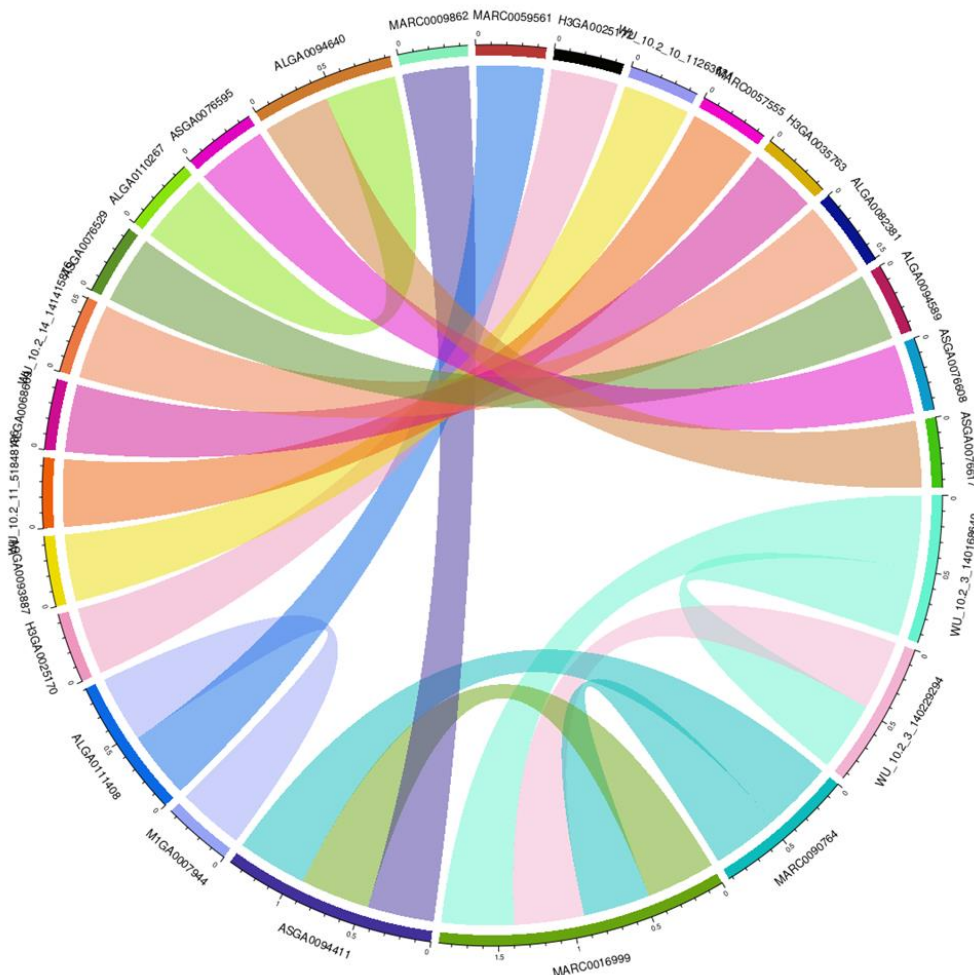


Рисунок 7 – Результаты расчета сниженной взаимной информации (RMI) как меры неравновесия по сцеплению (LD), показывающий уровень генетической связи между различными локусами в геноме

Высокие значения RMI ( $> 0.45$ ) могут свидетельствовать о регионах генома, где аллели наследуются вместе чаще, чем ожидается случайно. Это может быть полезно для идентификации генетических маркеров, связанных с важными хозяйственными признаками, или для понимания генетической структуры и истории популяций.

Таким образом, данные с отфильтрованными значениями RMI могут выделить ключевые области генома, где LD особенно силен, что может быть важно для дальнейшего генетического анализа и селекции.

Для оценки генетического разнообразия в контексте аутозиготности и распределения аутозиготных по происхождению сегментов в геномах кабанов и свиней основных коммерческих пород использовали модели, предложенные Druet с соавторами (Druet et al., 2017). Эти модели позволяют не только определять сегменты аутозиготности (HBD, Homozygous-by-Descent), но и сопоставлять с возрастом предков, от которых были унаследованы эти сегменты.

Идентификация областей HBD позволила определить коэффициенты геномного инбридинга у кабана и свиней. Средние значения коэффициента геномного инбридинга были практически на одном уровне во всех изучаемых группах и составили 0,23–0,29 (таблица 2). Наибольший максимум геномного инбридинга, равный 0,6, зафиксирован у кабанов. При этом максимальное значение инбридинга у свиней коммерческих пород составило 0,3.

Таблица 2 – Коэффициенты геномного инбридинга у диких кабанов и свиней коммерческих пород

| Название                   | Значение коэффициента геномного инбридинга |         |              |
|----------------------------|--------------------------------------------|---------|--------------|
|                            | 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       |

На рисунке 8 представлено распределение коэффициента инбридинга для свиней породы крупная белая в виде гистограммы. Столбцы выше в диапазоне 0,15–0,25 показывают преобладание умеренного инбридинга, характерного для плановой селекции. Наличие значений в области 0,30–0,35, даже с меньшей частотой, указывает на отдельные случаи сильного генетического сходства, требующие контроля и оценки структуры популяции и предоставляет возможность выявить потенциальные «зоны риска» для генетического разнообразия.

Преобладание низких и средних значений (0,20–0,22) может отражать успешное управление инбридингом. Однако «хвост» распределения в сторону 0,35 сигнализирует о необходимости корректировки стратегий разведения, чтобы избежать накопления вредных рецессивных аллелей и снижения устойчивости породы.

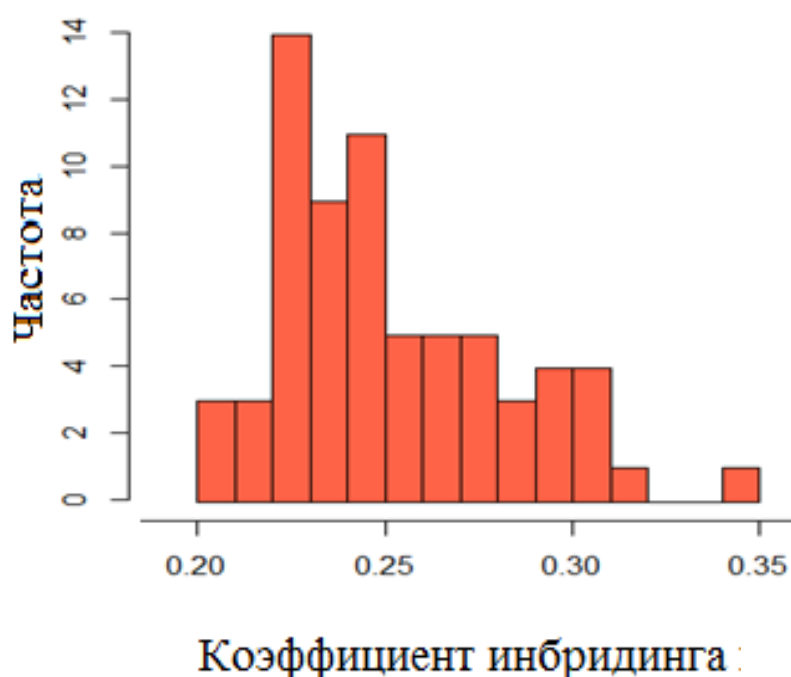


Рисунок 8 – Распределение коэффициента инбридинга свиней породы крупная белая

Распределение значений коэффициента инбридинга среди особей свиней породы ландрас (рисунок 9) показало, что наибольшие столбцы сосредоточены в зоне 0,22–0,26, это говорит о том, что большинство животных

подвергаются умеренному инбридингу, что типично для целенаправленной селекции. Для свиней породы ландрас, известных своей продуктивностью, поддержание баланса между селекцией и генетическим разнообразием критически важно. Если основная масса значений приходится на 0,22–0,26, это может отражать успешное внедрение мер, таких как ротация племенных животных или использование ДНК-тестов. Однако смещение части данных к 0,27–0,38 требует введения «свежей крови», строгого отслеживания родственных связей или кроссбридинга. Это поможет избежать накопления вредных мутаций, снижения иммунитета или плодовитости. Визуализация данных служит отправной точкой для принятия решений, направленных на сохранение устойчивости и конкурентоспособности породы в долгосрочной перспективе.

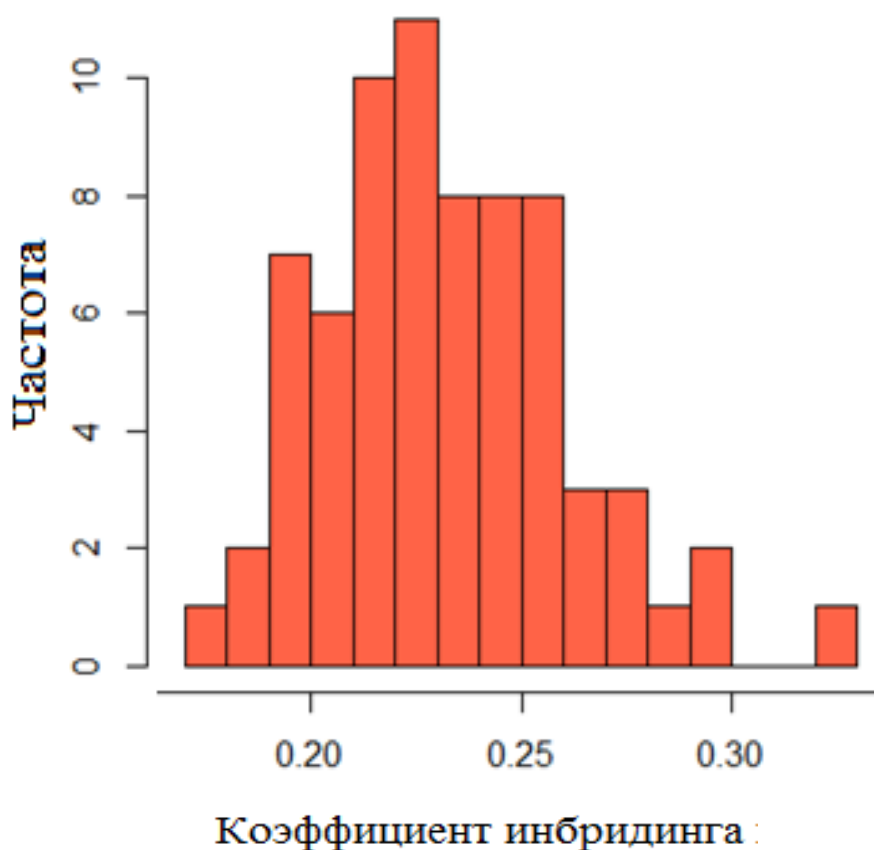


Рисунок 9 – Распределение коэффициента имбридинга свиней породы ландрас

На рисунке 10 представлено распределение коэффициента инбридинга для свиней породы дюрок, где основная масса данных сосредоточена в диапазоне 0,25–0,30, что указывает на умеренно высокий уровень генетического родства, характерный для интенсивной селекции. Наличие значений около 0,35, даже при низкой частоте, свидетельствует о случаях значительного инбридинга, которые могут негативно влиять на здоровье и продуктивность животных.

Для свиней породы дюрок, известной мясными качествами, поддержание баланса между закреплением желаемых признаков и сохранением генетической устойчивости особенно важно. Приближение значений к 0,35 сигнализирует о необходимости корректировки стратегии, а именно строгого контроля за родословными или применения ауткроссов, что поможет минимизировать риски инбредной депрессии, снижения иммунного ответа и повысит долгосрочную жизнеспособность породы.

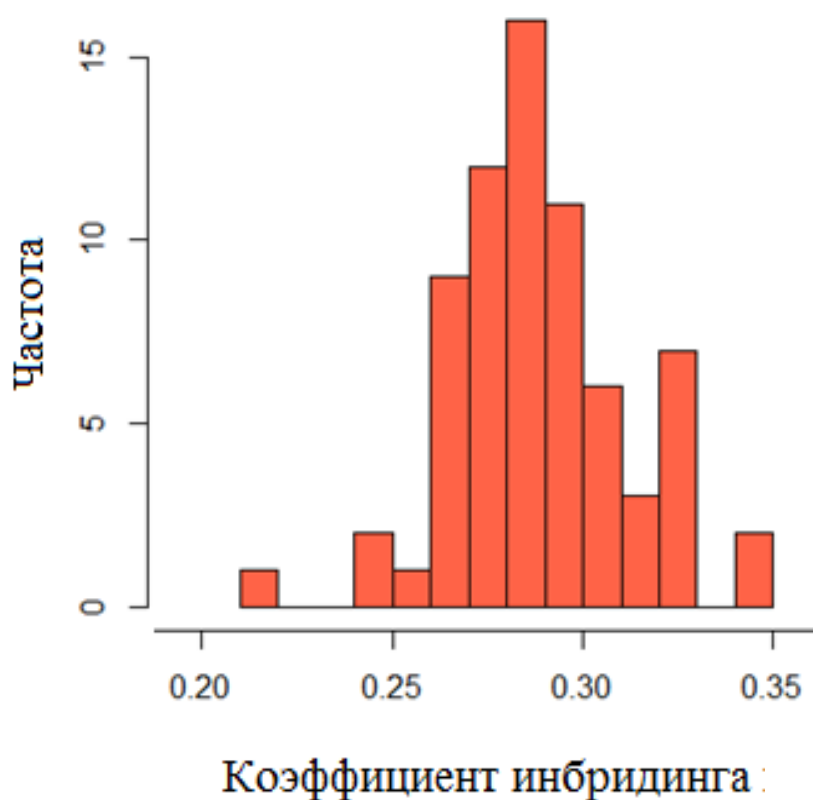
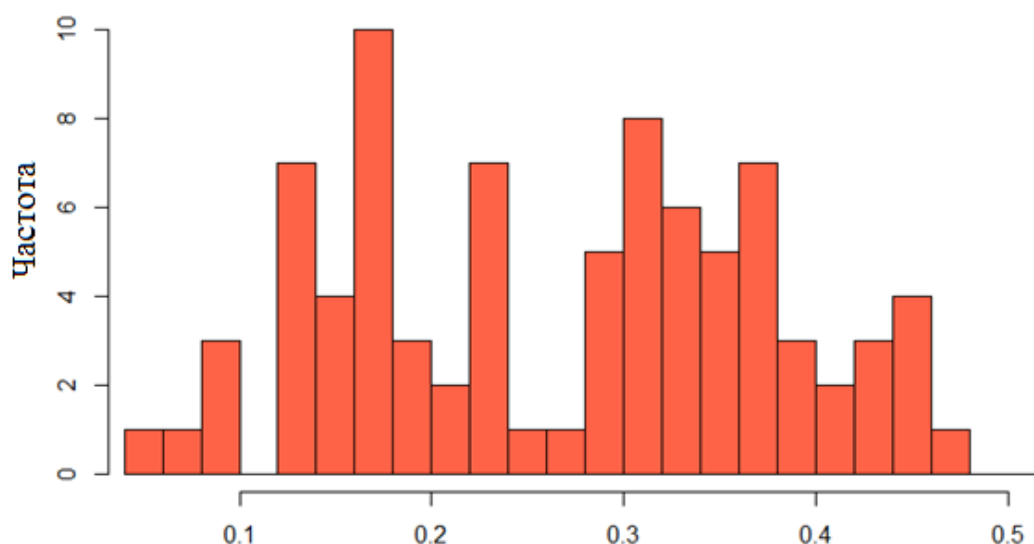


Рисунок 10 – Распределение коэффициента инбридинга свиней породы дюрок

На рисунке 11 отображено распределение коэффициента инбридинга для популяции диких кабанов свидетельствует о естественном уровне инбридинга, характерном для изолированных групп или популяций с ограниченным генетическим обменом. Наличие значений около 0,5, даже при низкой частоте, указывает на крайне высоком генетическом сходстве, которое в дикой природе часто связано с малым размером популяции, фрагментацией ареала или антропогенными факторами. Такое распределение помогает оценить риски снижения адаптивности и роста уязвимости к заболеваниям.

Для диких кабанов, в отличие от домашних пород, инбридинг часто возникает из-за естественных ограничений, таких как изоляция групп или сокращение численности. Преобладание коэффициентов в диапазоне 0,2–0,4 может отражать типичные для дикой природы процессы.



**Коэффициент инбридинга :**

Рисунок 11 – Распределение коэффициента инбридинга у диких кабанов

При оценках 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). Для породы крупная белая (рисунок 12) распределение классов  $R_k$  в геноме отражает ключевые этапы селекции, направленной на закрепление продуктивных признаков.

Классы  $R_k$ , соответствующие определённым временным периодам (например,  $R_k=16$  — 8 лет назад,  $R_k=32$  — 16 лет назад), могут указывать на периоды интенсивного отбора или введения новых генетических линий. Если, например, доминирует класс  $R_k=16$ , это говорит о недавних усилиях по улучшению таких характеристик, как конверсия корма или устойчивость к стрессовым условиям содержания. Одновременно наличие класса  $R_k=64$  (32 года назад) могло бы свидетельствовать о более ранних этапах формирования породы, когда закладывались её основные мясо-сальные качества.

Анализ классов  $R_k$  позволяет выделить «узкие места» в генетической стратегии. Концентрация в зоне  $R_k=16$ – $32$  (8–16 лет назад) указывает на циклы селекции, связанные с адаптацией к промышленным технологиям. Однако отсутствие классов с меньшими значениями  $R_k$  (например,  $R_k = 8, 4$  года назад) может свидетельствовать о недостаточном внедрении новых линий в последние годы.

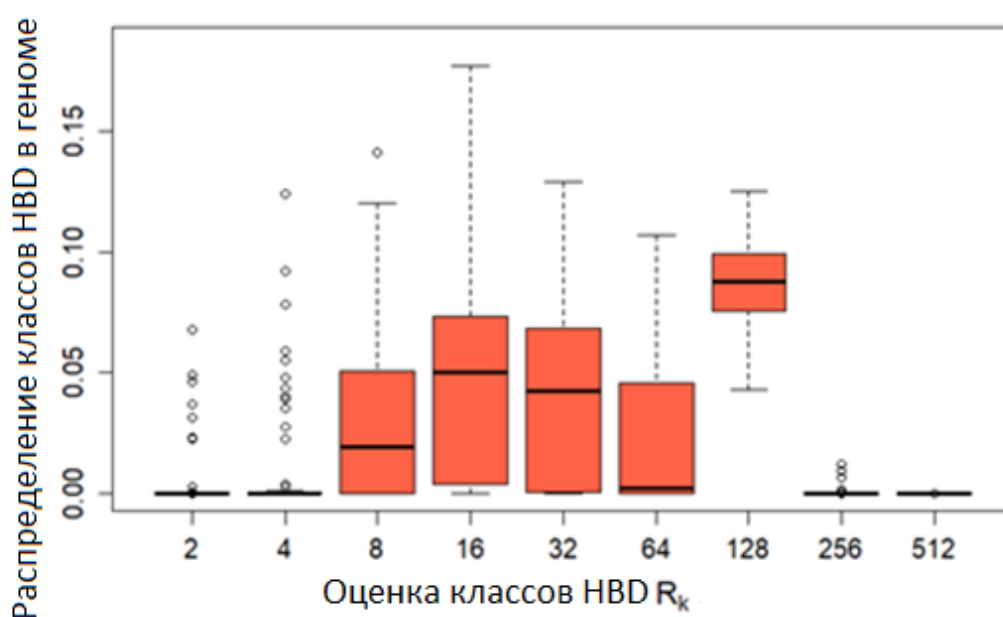


Рисунок 12 – Оценка аутозиготности в соответствии с возрастом для породы крупная белая

Для свиней породы ландрас (рисунок 13) распределение классов HBD в геноме демонстрирует особенности наследования идентичных сегментов ДНК, связанных с общими предками. Высокая частота класса  $R_k=16$  (8 лет



назад) в сочетании с долей генома 0,10–0,15 может указывать на этап целенаправленного отбора для улучшения таких признаков, как скорость роста или адаптация к условиям содержания. Наличие классов  $R_k=32-64$  (16–32 года назад) с аналогичными значениями подчеркивает более ранние циклы селекции, направленные на закрепление продуктивных характеристик.

Преобладание современных классов  $R_k = 2-8$  (1–4 года назад) с низкой долей HBD-сегментов (0,00–0,05) отражает использование стратегий, направленных на ротацию генетического материала или внедрение новых линий, для сохранения разнообразия. Однако концентрация в историческом классе  $R_k=128$  требует внимания к наследию длительного наследования идентичных сегментов, которое может влиять на устойчивость породы. Анализ таких распределений позволяет выделить периоды интенсивного отбора и скорректировать программы разведения, например, за счет оптимизации геномного анализа или диверсификации генетических ресурсов. Это способствует сохранению продуктивности и долгосрочной жизнеспособности ландрасов.

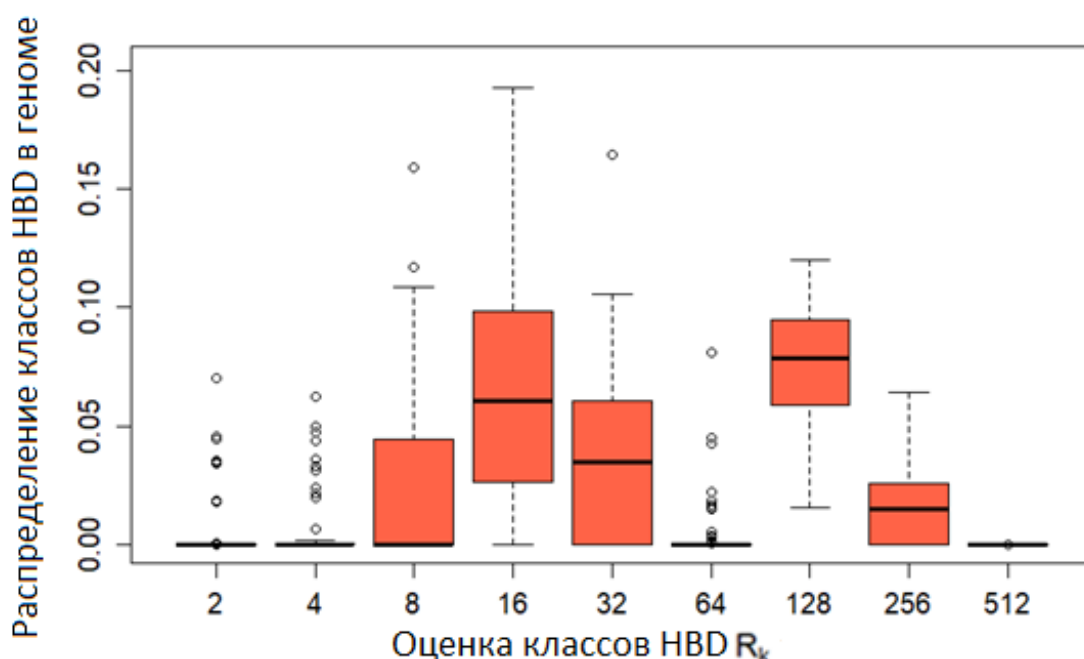


Рисунок 13 – Оценка аутозиготности в соответствии с возрастом для породы ландрас

Для породы дюрок (рисунок 14) класс  $R_k = 16$  (8 лет назад) с долей 0,15–0,20 может указывать на этап интенсивной селекции, направленной на улучшение мясных характеристик или адаптацию к промышленным условиям содержания. Наличие классов  $R_k = 32–64$  (16–32 года назад) с аналогичными значениями подчеркивает исторические циклы отбора, закрепившие базовые продуктивные качества породы.

Преобладание современных классов  $R_k = 2–8$  (1–4 года назад) с низкой долей HBD-сегментов (0,00–0,05) свидетельствует о внедрении мер по снижению генетического сходства. Однако высокая концентрация в классе  $R_k=128$  (64 лет назад) с долей 0,20–0,25 требует внимания к наследию длительного наследования идентичных сегментов, которое может способствовать накоплению скрытых рецессивных аллелей. Общая сумма HBD-сегментов позволяет оценить совокупную генетическую нагрузку, а именно если она превышает 0,20, это сигнализирует о необходимости диверсификации генетического материала. Такие меры помогут сохранить мясную продуктивность дюроков, минимизируя риски снижения адаптивности и устойчивости породы в долгосрочной перспективе.

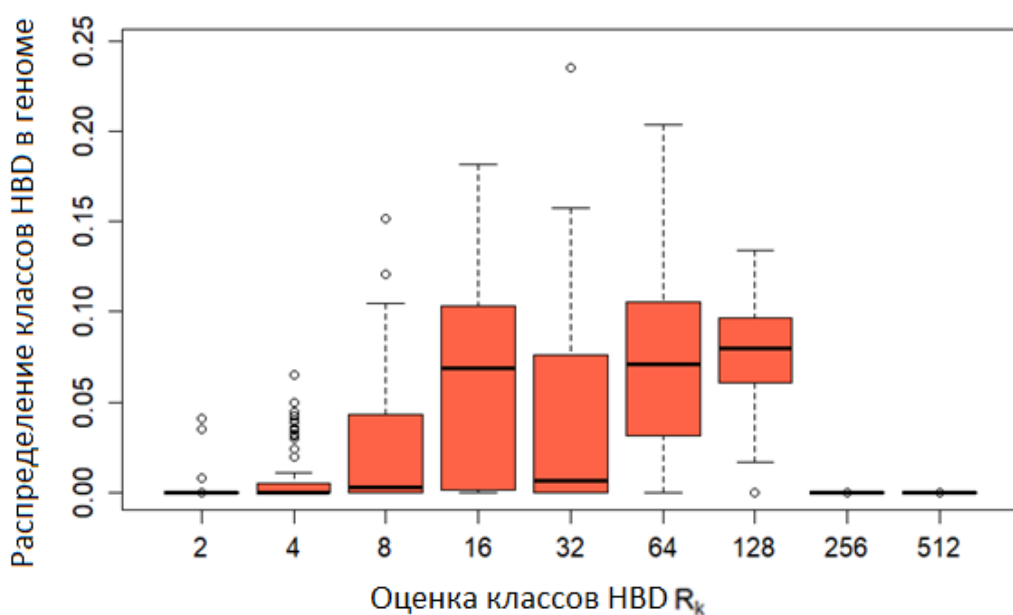


Рисунок 14 – Оценка аутозиготности в соответствии с возрастом для породы дюрок

Для диких кабанов (рисунок 15) высокая доля сегментов в классе  $R_k = 64$  (32 года назад) с уровнем 0,15–0,20 может указывать на период изоляции или сокращения численности популяции, вызванного изменениями среды обитания или антропогенным давлением. Наличие классов  $R_k = 128–512$  (64+ лет назад) с долей 0,25–0,30 подчеркивает длительное наследование идентичных участков ДНК, что характерно для популяций с ограниченным генетическим обменом на протяжении поколений.

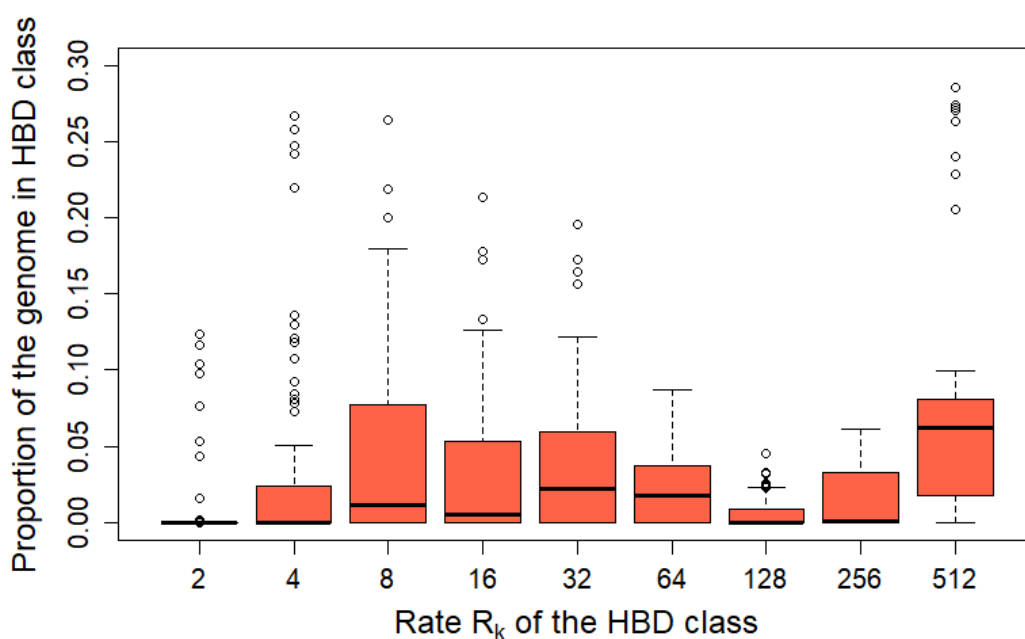


Рисунок 15 – Оценка аутозиготности в соответствии с возрастом для диких кабанов

Современные классы  $R_k = 2–16$  (1–8 лет назад) с низкой долей HBD-сегментов (0,00–0,10) могут отражать восстановление генетического разнообразия благодаря миграции между группами или природоохранным мерам. Однако высокая суммарная доля HBD-сегментов (например, 0,25–0,30) в исторических классах сигнализирует о рисках снижения адаптивности из-за накопления рецессивных аллелей.

В отличие от коммерческих пород, у диких кабанов высокая доля аутозиготности (около 0,08) связана с естественными процессами. Исторические значения  $R_k = 128–512$  с долей 0,25–0,30 отражают длительную

изоляция или сокращение популяции, например, из-за фрагментации ареалов. Низкая аутозиготность в современных классах ( $R_k = 2-16$ ) у кабанов может объясняться восстановлением миграционных потоков, тогда как у ландрасов и дюроков она связана с внедрением новых генетических линий. Это подчеркивает принципиальное различие: коммерческие породы управляют генетическим разнообразием через селекцию, а дикие популяции зависят от естественных механизмов, которые нарушаются антропогенными факторами. Сводная статистика реализованной аутозиготности в каждом классе (в пропорции от генома) для свиней и кабанов представлена в таблице 3.

Таблица 3 – Сводная статистика реализованной аутозиготности в каждом классе HBD (в пропорции от генома) у диких кабанов и свиней коммерческих пород

| Значение      | R_4      | R_8      | R_16     | R_32     | R_64     | R_128    | R_256    | R_512    |
|---------------|----------|----------|----------|----------|----------|----------|----------|----------|
| Крупная белая |          |          |          |          |          |          |          |          |
| mean          | 0,01033  | 0,03145  | 0,04991  | 0,04313  | 0,02537  | 0,08700  | 0,00044  | 6,24E-08 |
| se            | 0,00296  | 0,00441  | 0,00528  | 0,00454  | 0,00396  | 0,00203  | 0,00024  | 3,77E-08 |
| min           | 1,96E-20 | 1,1E-11  | 1,1E-19  | 9,73E-09 | 6E-10    | 0,04283  | 2,02E-22 | 3,88E-36 |
| max           | 0,12407  | 0,14137  | 0,17696  | 0,12901  | 0,10695  | 0,12548  | 0,01193  | 2,34E-06 |
| Ландрас       |          |          |          |          |          |          |          |          |
| mean          | 0,00611  | 0,02588  | 0,06276  | 0,03617  | 0,00414  | 0,07445  | 0,01699  | 2,95E-08 |
| se            | 0,00173  | 0,00436  | 0,00544  | 0,00427  | 0,00150  | 0,00294  | 0,00178  | 1,67E-08 |
| min           | 1,11E-11 | 7,25E-12 | 1,50E-11 | 1,09E-13 | 2,03E-18 | 0,015683 | 5,14E-16 | 3,62E-33 |
| max           | 0,06240  | 0,15884  | 0,19229  | 0,16412  | 0,08102  | 0,11963  | 0,06383  | 1,12E-06 |
| Дюрок         |          |          |          |          |          |          |          |          |
| mean          | 0,00810  | 0,02641  | 0,06215  | 0,04272  | 0,07026  | 0,07788  | 1,29E-06 | 1,81E-10 |
| se            | 0,00188  | 0,00444  | 0,00593  | 0,00641  | 0,00609  | 0,00303  | 1,22E-06 | 1,05E-10 |
| min           | 6,21E-14 | 8,74E-09 | 5,59E-11 | 3,35E-13 | 1,21E-13 | 7,56E-08 | 2,59E-43 | 3,43E-58 |
| max           | 0,06484  | 0,15194  | 0,18172  | 0,23490  | 0,20339  | 0,13380  | 8,57E-05 | 6,49E-09 |
| Дикие кабаны  |          |          |          |          |          |          |          |          |
| mean          | 0,03167  | 0,04906  | 0,03302  | 0,03802  | 0,02127  | 0,00593  | 0,01575  | 0,07625  |
| se            | 0,00691  | 0,00713  | 0,00505  | 0,00487  | 0,00247  | 0,00108  | 0,00206  | 0,00850  |
| min           | 3,79E-27 | 5,75E-23 | 1,1E-12  | 1,79E-11 | 2,79E-16 | 3,73E-41 | 2,45E-43 | 1,02E-17 |
| max           | 0,26631  | 0,26387  | 0,21361  | 0,19529  | 0,08660  | 0,04511  | 0,06136  | 0,28505  |

Каждая особь обладает индивидуальной аутозиготностью и у свиней коммерческих пород в целом все животные 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  $R_k = 512$ .

На кривой кумулятивной аутозиготности для исследуемых особей свиней породы крупная белая (рисунок 16) отображены кумулятивные пропорции аутозиготности генома в зависимости от пороговых значений  $R_k$  (2–512), соответствующих временным периодам в поколениях. Красная линия обозначает среднее значение аутозиготности в популяции, демонстрируя общий тренд. Если линия стабильно возрастает, это говорит о равномерном вкладе разных временных периодов. Резкие скачки на определенных  $R_k$  (например,  $R_k=16$  или 64) могут указывать на ключевые этапы селекции, такие как введение новых линий или усиление отбора. Плато при  $R_k = 128$  (примерно 64 года назад) указывает на завершение значимого влияния древних генетических событий на текущую структуру генома. Это означает, что сегменты ДНК, унаследованные от предков старше 64 года, вносят минимальный вклад в современную аутозиготность. Такое явление может быть связано с естественным «размытием» генетического наследования за счёт рекомбинаций или с тем, что основная селекционная работа проводилась в более поздние периоды. Выравнивание также подчеркивает однородность генетической истории особей в рамках этих временных рамок, что свидетельствует о стабильности базовых характеристик породы.

Именно во временных промежутках  $R_k = 2–128$  (1–64 года назад) сосредоточены ключевые этапы селекции, влияющие на продуктивность и адаптивность породы и важно контролировать недавние циклы разведения, чтобы избежать накопления вредных аллелей. Это позволяет сохранить баланс между поддержанием хозяйственно-полезных признаков и долгосрочной устойчивостью породы.

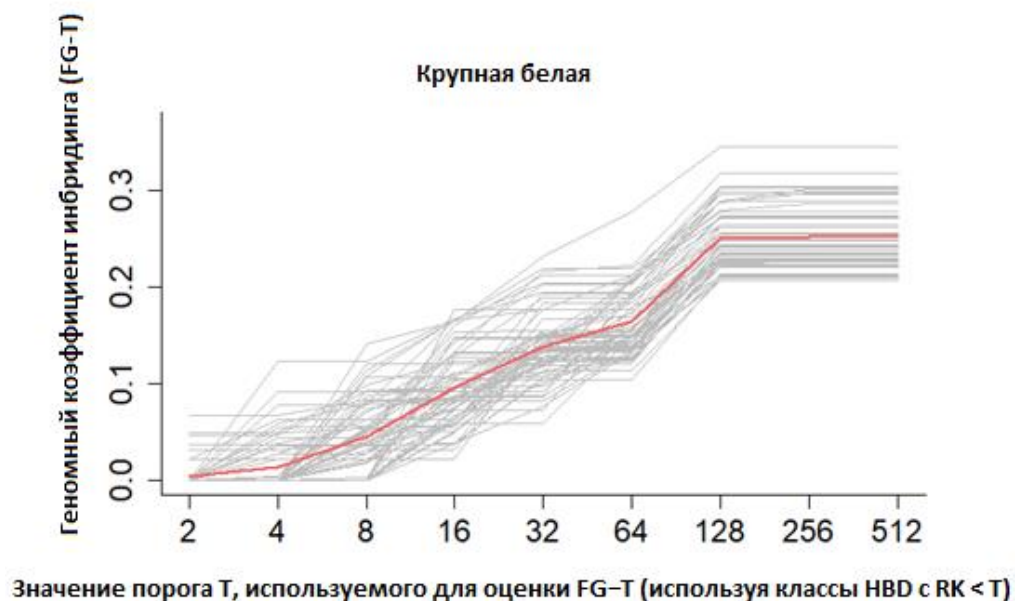


Рисунок 16 – Индивидуальная оценка аутозиготности для породы крупная белая. Красной линией обозначен общий тренд накопления аутозиготности

Линии на графике индивидуальной аутозиготности для свиней породы ландрас (рисунок 17) значительно варьируются. Особи с резким ростом пропорции на малых  $R_k$  ( $R_k = 8-32$ ) демонстрируют недавние генетические события. Линии, расположенные выше средней, указывают на повышенную долю HBD-сегментов, что может быть следствием естественного накопления гомозиготности в условиях ограниченного генетического разнообразия. Такие особи подвержены рискам проявления рецессивных аллелей, способных снижать адаптивность или вызывать заболевания.

Распределение аутозиготности также отражает влияние исторических процессов, которые закрепили определённые признаки. Выравнивание кривых при  $R_k = 128$  (примерно 64 года назад) свидетельствует, что древние генетические события минимально влияют на текущую структуру генома.

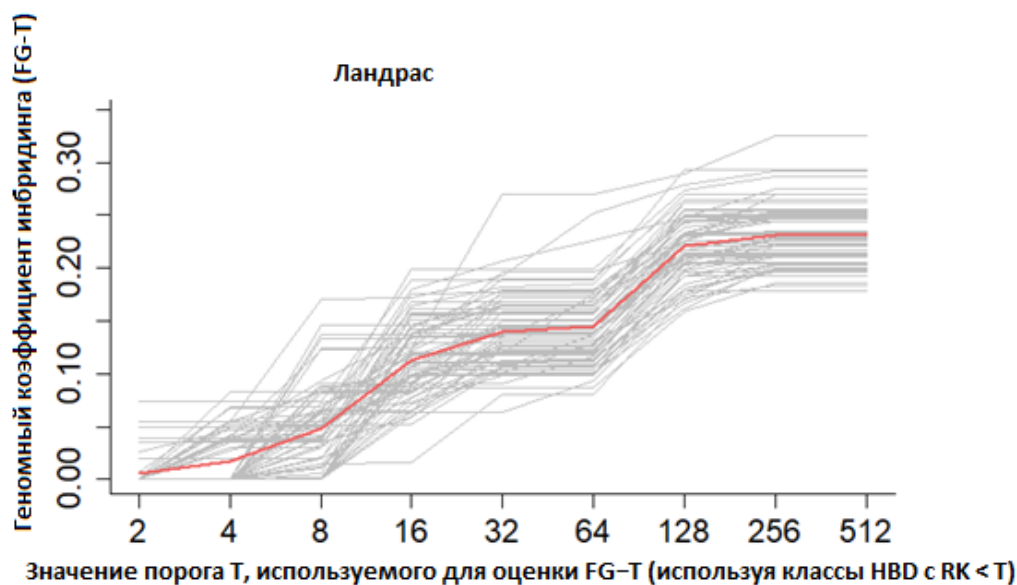


Рисунок 17 – Индивидуальная оценка аутозиготности для породы ландрас. Красной линией обозначен общий тренд накопления аутозиготности

Схожая с породами крупная белая и ландрас картина наблюдается индивидуальная аутозиготность и у породы дюрок (рисунок 18). Если резкие скачки на произошли на определенных  $R_k$  ( $R_k = 64$  и  $128$ ) для породы крупная белая и скачки на  $R_k$  ( $R_k = 8, 16, 64$  и  $128$ ) для породы ландрас которые указывают на ключевые этапы селекции, то для породы дюрок такие ключевые события произошли примерно 8, 16 и 64 года назад ( $R_k = 16, 32$  и  $128$ ).

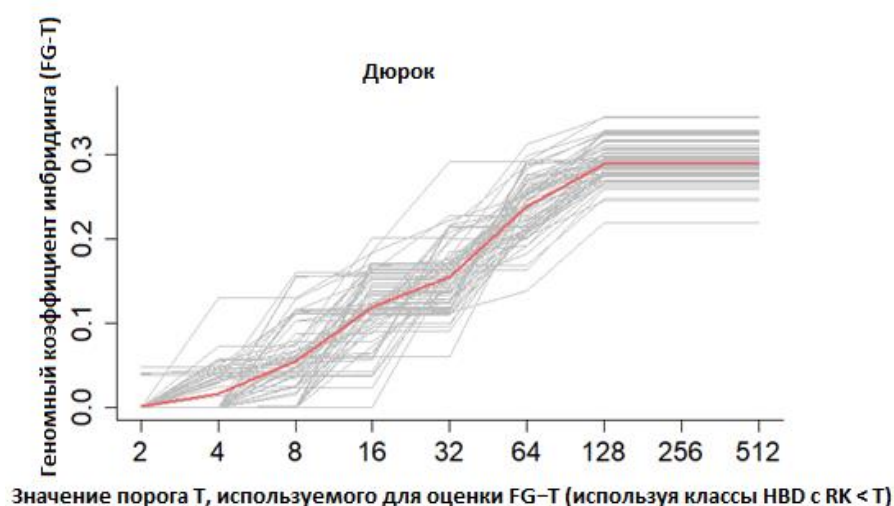


Рисунок 18 – Индивидуальная оценка аутозиготности для породы дюрок. Красной линией обозначен общий тренд накопления аутозиготности

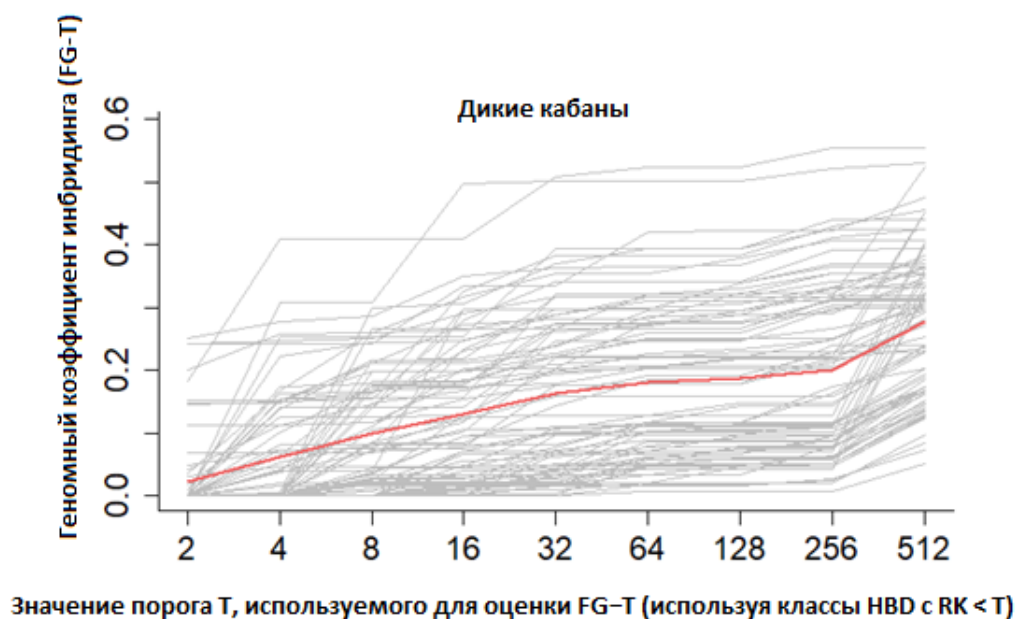


Рисунок 19 – Индивидуальная оценка аутозиготности для диких кабанов. Красной линией обозначен общий тренд накопления аутозиготности

Совершенно другое поведение аутозиготности демонстрируют дикие кабаны (рисунок 19). Резкие скачки на уровне  $Rk = 64$  и  $256$  свидетельствуют о ключевых изменениях в популяции диких кабанов.

С другой стороны, по сравнению с коммерческими породами в геноме диких кабанов присутствуют значения  $Rk = 512$  которые охватывают многовековые периоды, выявляя долгосрочные эффекты, такие как генетический дрейф, эффект основателя или исторические сокращения численности. Линии выше среднего значения указывают на особей с повышенной аутозиготностью, что связано с наследованием идентичных сегментов ДНК от ограниченного круга предков. Это может быть следствием естественной изоляции (географические барьеры) или антропогенных факторов (разрушение миграционных путей).

В целом у свиней коммерческих пород определено 30 416 сегментов HBD, в среднем 145 на одного животного; у кабанов – 11 598 сегментов, в среднем 133 на одного животного (таблица 4). Самый длинный сегмент размером 183,39 Mb (3403 SNP) установлен в группе кабанов в первой хромосоме. Среди свиней коммерческих пород наибольшую длину имел



сегмент размером 168,99 Mb (3 428 SNP) у свиней крупной белой породы, и он также был расположен в первой хромосоме (рисунок 20).

Таблица 4 – Сводная статистика сегментов HBD у свиней коммерческих пород и диких кабанов

| Статистика | Крупная белая ( $n = 68$ ) |                     | Ландрас ( $n = 71$ ) |                     | Дюрок ( $n = 70$ ) |                     | Дикие кабаны ( $n = 87$ ) |                     |
|------------|----------------------------|---------------------|----------------------|---------------------|--------------------|---------------------|---------------------------|---------------------|
|            | кол-во SNP                 | длина сегмента (Mb) | кол-во SNP           | длина сегмента (Mb) | кол-во SNP         | длина сегмента (Mb) | кол-во SNP                | длина сегмента (Mb) |
| $n$        | 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              |

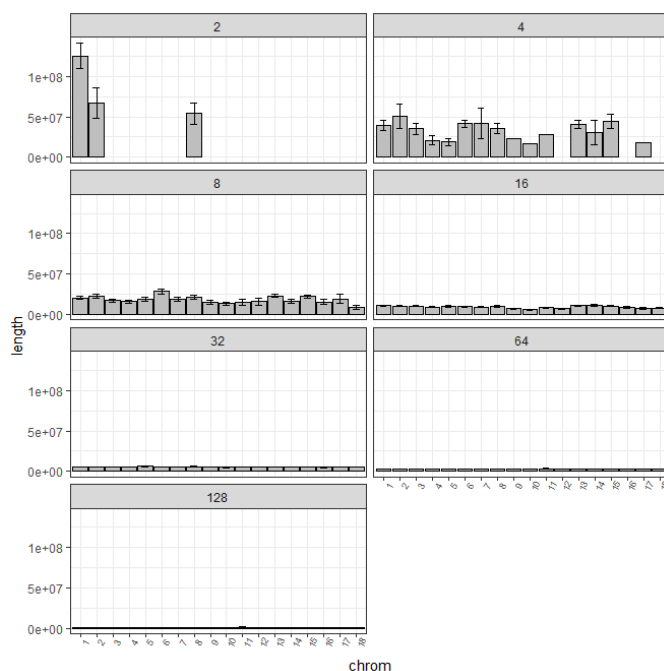


Рисунок 20 – Распределение сегментов HBD по хромосомам у свиней породы крупная белая

Распределение сегментов HBD по хромосомам с учетом различных классов для свиней породы ландрас представлено на рисунке 21. Длинные сегменты, принадлежащие классу  $R_k = 2$  определены в SSC1, SSC2, SSC4,

SSC5, SSC9, SSC13, SSC14 и SSC15; сегменты класса  $R_k = 4$  представлены в хромосомах SSC1, SSC2, SSC5, SSC7, SSC15 и SSC14.

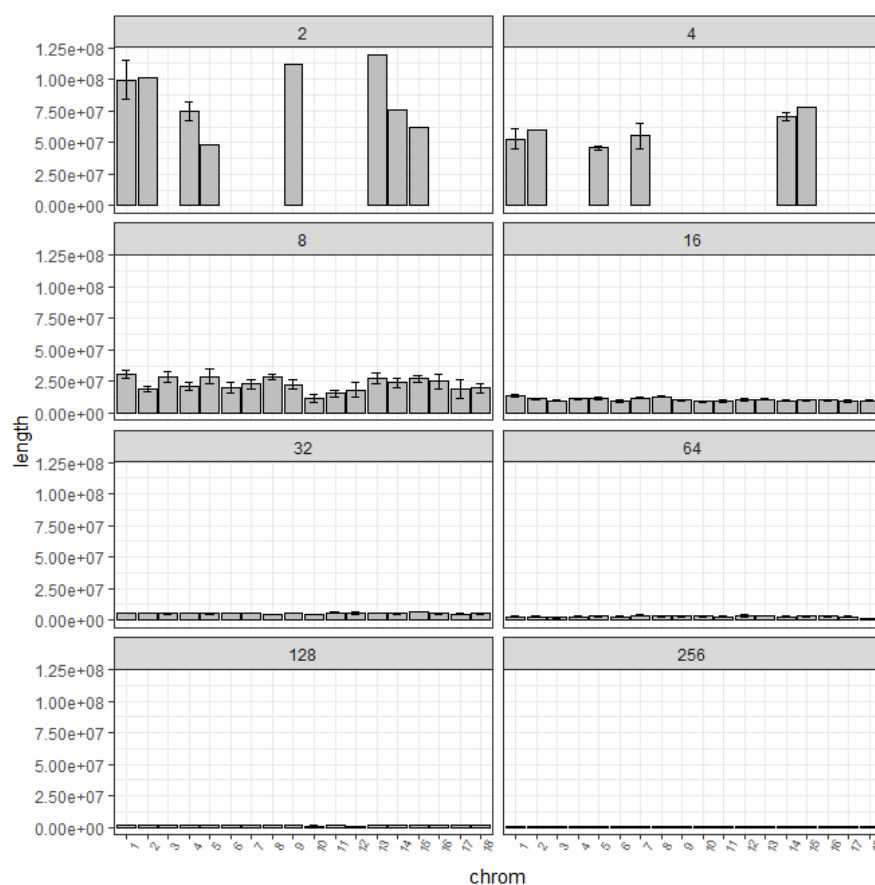


Рисунок 21 – Распределение сегментов HBD по хромосомам у свиней породы ландрас

У свиней породы дюрок длинные HBD-сегменты класса  $R_k = 2$  обнаружены исключительно на хромосоме SSC1 (рисунок 22), что указывает на недавние генетические события, связанные с этой хромосомой. Возможно, это результат целенаправленной селекции или случайного генетического дрейфа, затронувшего конкретный локус. Сегменты класса  $R_k = 4$  распределены более широко: они присутствуют на хромосомах SSC1, SSC3–SSC6, SSC8, SSC9, SSC13–SSC15 и SSC17. Такая локализация может отражать исторические этапы селекции, когда отбор по продуктивным признакам затронул гены, расположенные в этих регионах.

Для остальных классов HBD ( $R_k > 4$ ) сегменты различной длины выявлены практически во всех хромосомах. Это говорит о том, что геном дюроков сохраняет следы длительного наследования идентичных участков ДНК, характерного для замкнутых популяций с ограниченным генетическим разнообразием. Широкое распределение сегментов по хромосомам может быть связано как с историческим инбридингом, так и с неравномерным воздействием селекции на разные участки генома.

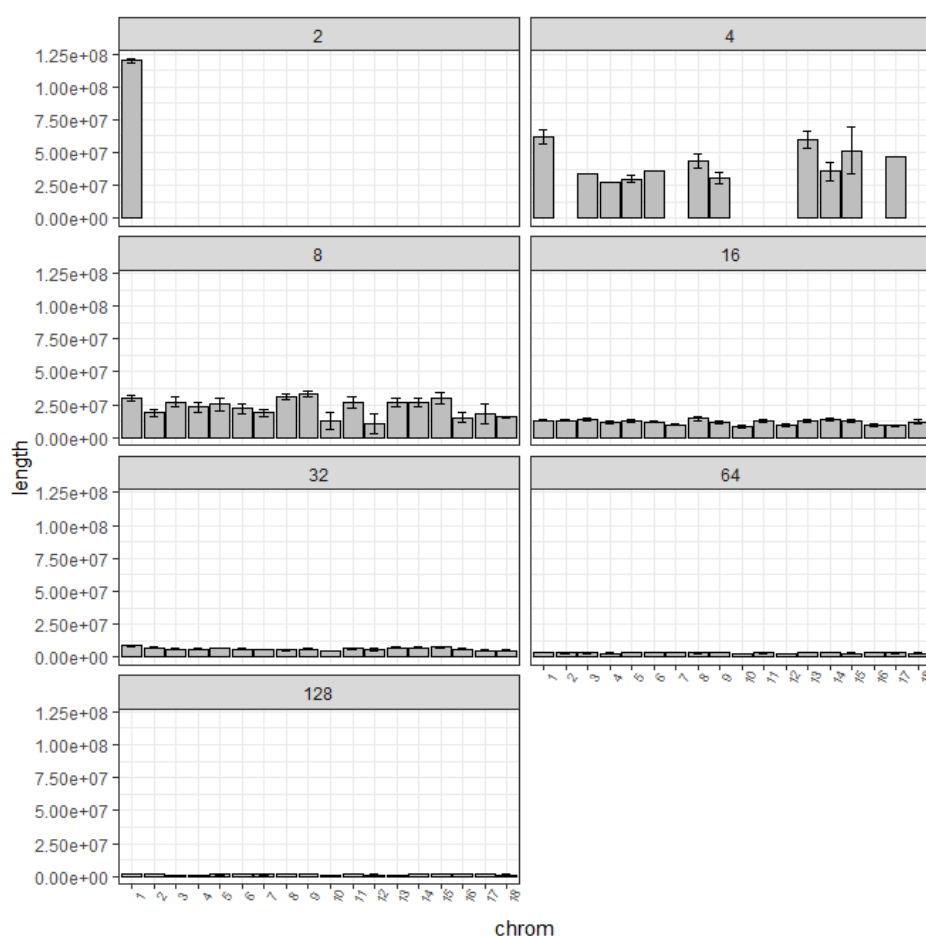


Рисунок 22 – Распределение сегментов HDB по хромосомам у свиней породы дюрок

У диких кабанов (рисунок 23), в отличие от коммерческих пород, длинные сегменты класса  $R_k = 2$  более представлены и локализованы во всех хромосомах за исключением SSC6, SSC7, SSC10, SSC16 и SSC17; при этом  $R_k = 4$  покрыты все хромосомы. Сегменты классов  $R_k = 8 - R_k = 128$  равномерно распределены по всем хромосомам во всех исследуемых группах. У

коммерческих пород короткие сегменты  $R_k = 256$  обнаружены только в породе ландрас. У кабанов, помимо сегментов  $R_k = 256$ , также присутствуют самые короткие сегменты класса  $R_k = 512$ .

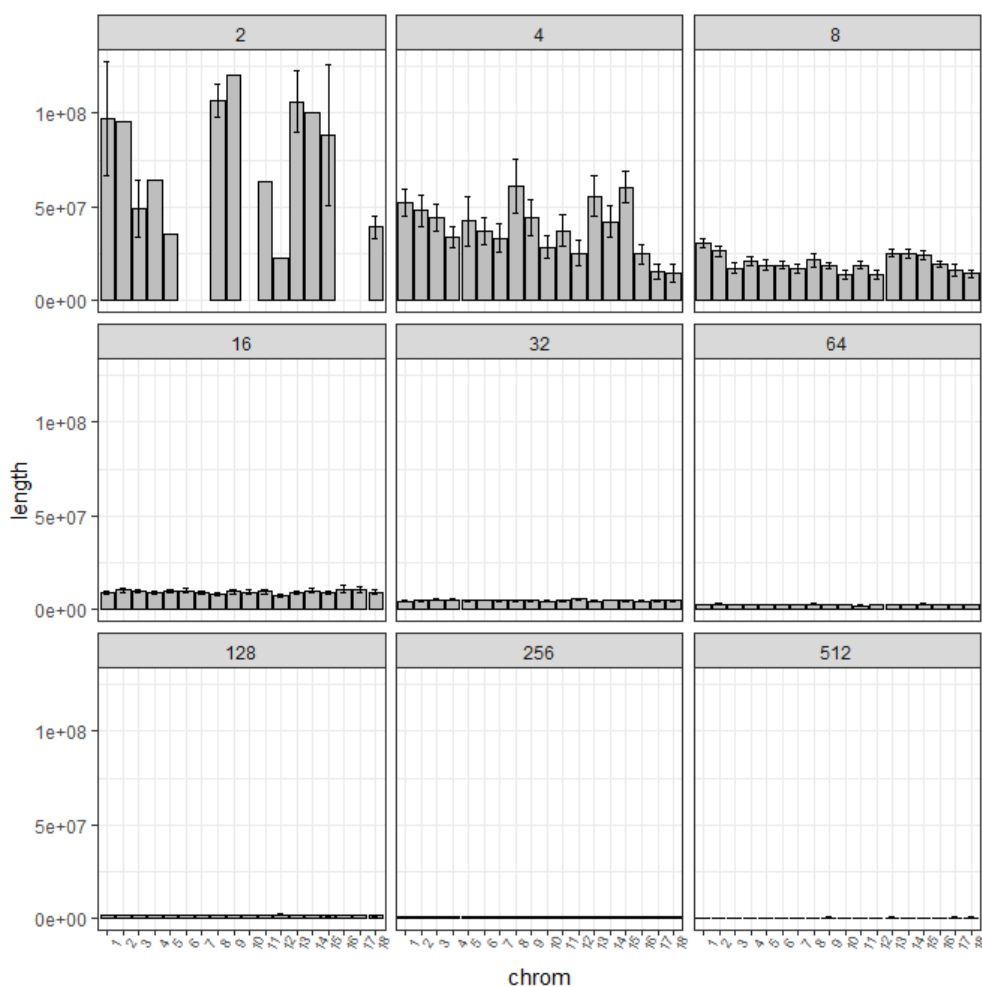


Рисунок 23 – Распределение сегментов HBD по хромосомам у диких кабанов

Частоты SNP (%) в обнаруженных HBD были оценены для каждой группы. У свиней породы крупная белая наиболее важные SNP-маркеры по недавней аутозиготности (рисунок 24) находятся в тех участках генома, где встречаются длинные, насыщенные гомозиготностью регионы. Особенно много таких SNP сосредоточено на хромосомах 1, 6, 9 и 13, каждая из которых содержит более 19 тысяч маркеров в рамках HBD-сегментов. Это говорит о том, что именно в этих хромосомах происходило частое наследование идентичных по происхождению участков ДНК – вероятно, в результате

использования близкородственных животных или селекции по определённым признакам.

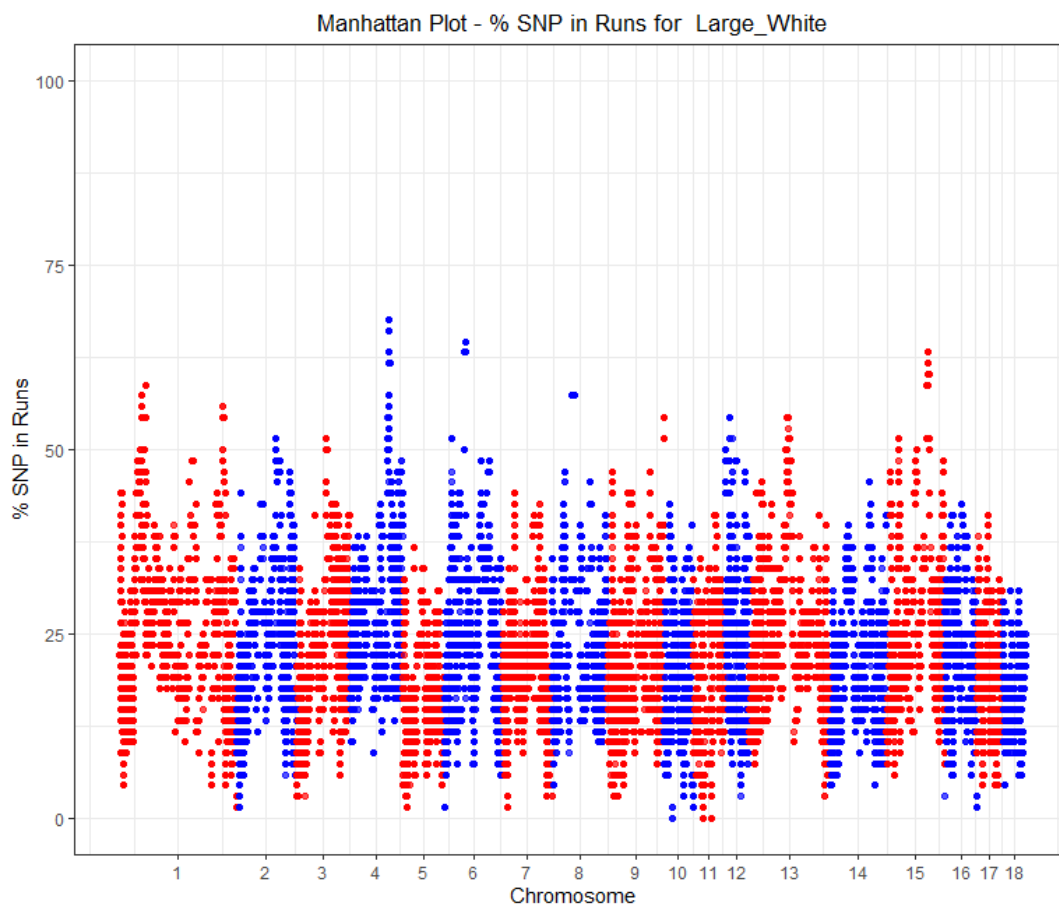


Рисунок 24 – Частоты SNP (%) в обнаруженных HBD у свиней породы крупная белая

На уровне конкретных участков особый интерес вызывают хромосома 7: от 7,4 до 11,3 млн п.о. (175 SNP), хромосома 10: 24,5–28,7 млн п.о. (173 SNP), хромосома 14: 139,8–142,6 млн п.о. (162 SNP), хромосома 5: 3,3–6,3 млн п.о. (160 SNP), хромосома 13: два перекрывающихся сегмента 8,8–12,0 млн п.о. (157–159 SNP), хромосома 12: 15,8–19,3 млн п.о. (156 SNP), хромосома 6: 139,5–142,4 млн п.о. (151 SNP).

Такие регионы генома могут быть признаками селекционного давления и требуют отдельного внимания при анализе генофонда. Их можно использовать для поиска функциональных генов, маркерной селекции или оценки степени инбридинга в популяции.

Для породы ландрас (рисунок 25) наиболее значимые SNP, связанные с недавней аутозиготностью, располагаются на хромосомах 1, 6, 2, 7 и 4, где наблюдаются максимальные суммы SNP в HBD-сегментах — от 19 000 до 25 000 маркеров. Это свидетельствует о наличии протяжённых регионов гомозиготности, характерных для сильного направленного отбора или узких популяционных «бутылочных горлышек». Средняя длина HBD-сегментов на этих хромосомах превышает 1.3 млн п.о., а в некоторых случаях достигает 3–4 млн п.о., что указывает на их недавнее происхождение.

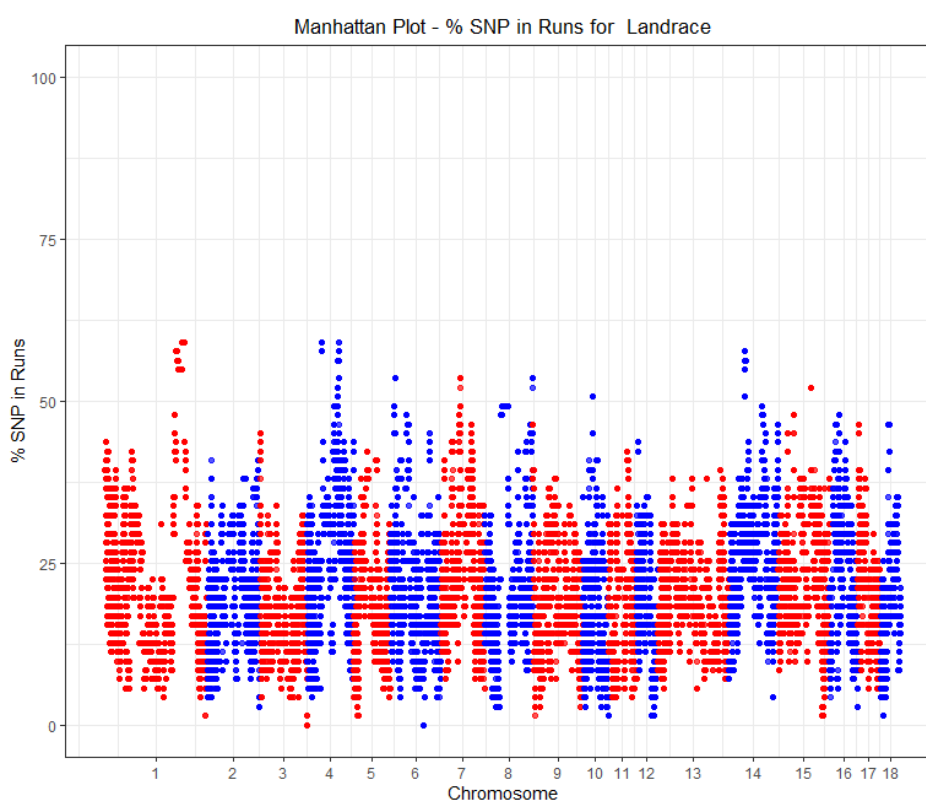


Рисунок 25 – Частоты SNP (%) в обнаруженных HBD у свиней породы ландрас

Наиболее насыщенные SNP-регионы включают хромосома 6: 139.4–143.3 Mb (201 SNP, 3.85 Mb), хромосома 15: 136.5–140.5 Mb (192 SNP, 4.03 Mb), хромосома 4: 3.1–6.6 Mb (184 SNP, 3.49 Mb), хромосома 1: 10.1–14.3 Mb (177 SNP, 4.17 Mb), хромосома 8: 143.9–148.1 Mb (174 SNP, 4.16 Mb), хромосома 12: 52.5–56.6 Mb (172 SNP, 4.15 Mb).

Наиболее значимые SNP в контексте аутозиготности для породы дюрок (рисунок 26) сосредоточены в длинных сегментах. В частности, наибольшее общее количество SNP в таких сегментах наблюдается на хромосомах 2, 14, 12 и 4 – от 15 до 17 тысяч SNP на каждую хромосому. Это указывает на высокую плотность участков недавней гомозиготности, что может быть связано с интенсивной селекцией или недавним инбридингом.

Среди конкретных регионов, выделяются хромосома 4: регион 3.4–7.8 Mb (237 SNP, 4.36 Mb), хромосома 14: 139.5–142.4 Mb (173 SNP, 2.96 Mb), хромосома 2: 135.5–139.9 Mb (171 SNP, 4.35 Mb), хромосома 11: 74.0–77.7 Mb (167 SNP, 3.69 Mb), хромосома 8: 133.5–137.5 Mb (160 SNP, 4.02 Mb), хромосома 9: 12.0–15.5 Mb (160 SNP, 3.46 Mb). Детальное изучение этих регионов позволяет выявить ключевые локусы, подвергшиеся селекции, и оценить риски, связанные с потерей генетического разнообразия.

Для каждой группы выбраны топ HBD при условии, что частота HBD составляет не менее 60% и не менее 10 SNP. У свиней породы ландрас условиям топ HDB соответствовали 9 сегментов, локализованные в SSC1, SSC4, SSC6–SSC8, SSC10, SSC14 и SSC15.

Манхэттенский сюжет для диких кабанов (рисунок 27), отображающий процентное содержание SNP в участках гомозиготности (ROH) по хромосомам 1-18, демонстрирует гетерогенное распределение. Отчетливые пики, указывающие на высокую долю SNP, локализованных в ROH (достигающую и превышающую 50-60%), наблюдаются на хромосомах 1, 4, 7, 15 и 18. Особо выделяются такие регионы на хромосоме 1 и один из наиболее высоких пиков (около 65%) на хромосоме 15. Подобное распределение свидетельствует о том, что недавние генетические процессы привели к формированию протяженных гомозиготных гаплотипов в этих специфических регионах генома дикого кабана, которые охватили значительную долю локальных полиморфных сайтов.

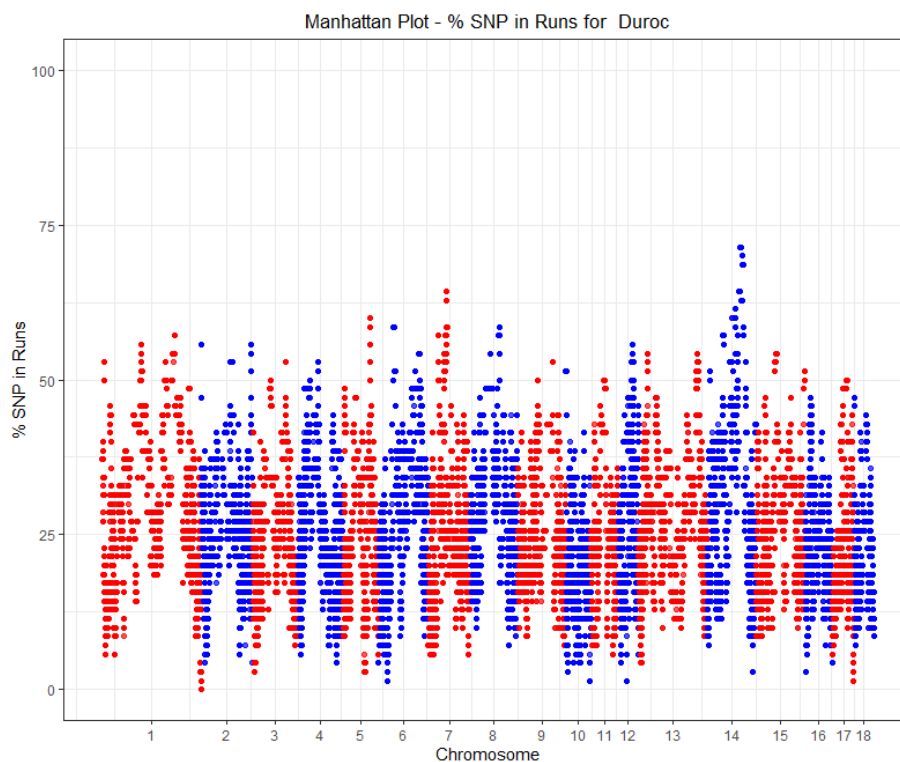


Рисунок 26 – Частоты SNP (%) в обнаруженных HBD у свиней породы дюрок

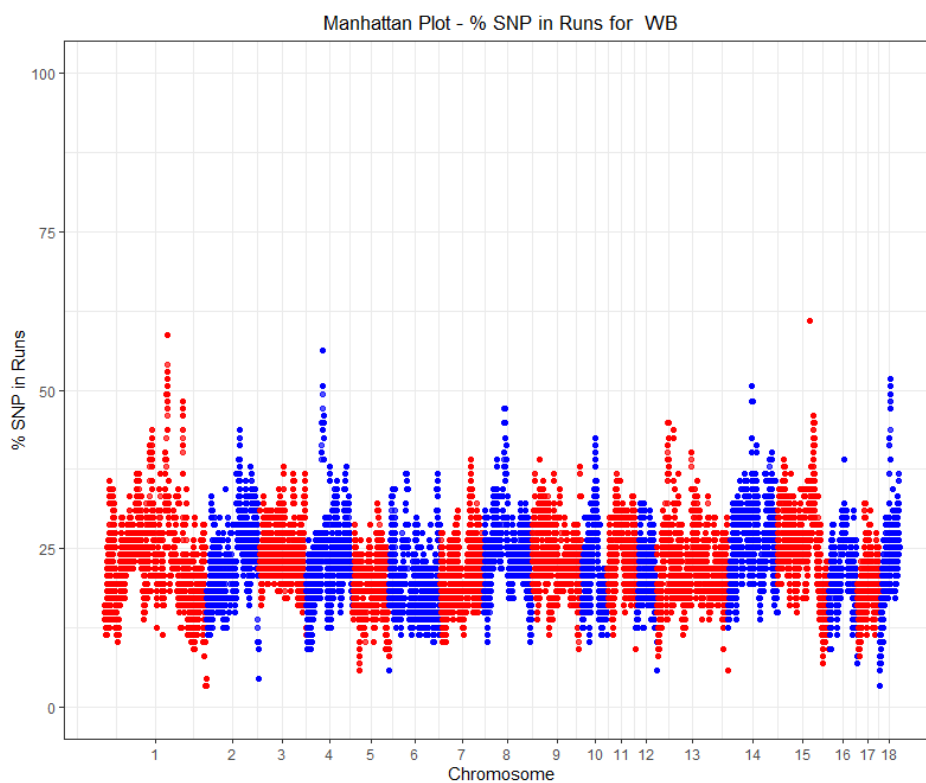


Рисунок 27 – Частоты SNP (%) в обнаруженных HBD диких кабанов



Длина топ HBD варьировала от 0,30 Mb (SSC7) до 27,05 Mb (SSC1) (таблица 5). У свиней породы дюрок определено наибольшее количество топ HBD, 35 сегментов, расположенных во всех хромосомах, за исключением SSC3, SSC11, SSC16-SSC18. Минимальная длина этих сегментов составила 0,23 Mb (SSC10), а максимальная – 15,32 Mb (SSC14). У диких кабанов определено только 3 топ HBD в SSC1 (1,38 Mb), SSC4 (2,11 Mb) и SSC14 (0,31 Mb).

Таблица 5 – Ключевые HBD у диких кабанов и свиней коммерческих пород

| Название SNP         |                     | Хромосома | nSNP | Позиция нуклеотидов |           | Длина,<br>п.н. |
|----------------------|---------------------|-----------|------|---------------------|-----------|----------------|
| Первая SNP           | Последняя SNP       |           |      | от                  | до        |                |
| Крупная белая        |                     |           |      |                     |           |                |
| ALGA0003892          | ALGA0003896         | 1         | 13   | 69968051            | 70512838  | 544787         |
| ALGA0008983          | H3GA0004187         | 1         | 22   | 278271067           | 278822935 | 551868         |
| ASGA0006558          | WU_10.2_1_280235304 | 1         | 31   | 279374670           | 280235304 | 860634         |
| H3GA0001733          | ALGA0003521         | 1         | 39   | 58990256            | 60455614  | 1465358        |
| ALGA0014711          | ALGA0106518         | 2         | 15   | 107102726           | 107512256 | 409530         |
| ASGA0015187          | DRGA0004042         | 3         | 22   | 80731857            | 81597192  | 865335         |
| ASGA0020965          | H3GA0013517         | 4         | 27   | 100967134           | 101855077 | 887943         |
| ALGA0026937          | M1GA0006152         | 4         | 27   | 103490318           | 104129681 | 639363         |
| MARC0027501          | ASGA0083623         | 4         | 112  | 104518531           | 108045490 | 3526959        |
| WU_10.2_6_20635220   | ASGA0027829         | 6         | 26   | 20635220            | 21310322  | 675102         |
| MARC0082369          | ASGA0082593         | 6         | 53   | 55261304            | 58859012  | 3597708        |
| ALGA0047870          | ASGA0082278         | 8         | 79   | 51464536            | 60241906  | 8777370        |
| ASGA0096216          | WU_10.2_9_153656647 | 9         | 55   | 152204065           | 153656647 | 1452582        |
| WU_10.2_12_10754934  | ASGA0053110         | 12        | 16   | 10754934            | 11073939  | 319005         |
| H3GA0037141          | DRGA0012788         | 13        | 25   | 107299413           | 108535538 | 1236125        |
| CASI0010149          | MARC0032201         | 13        | 26   | 104296385           | 106051998 | 1755613        |
| H3GA0044073          | ALGA0084620         | 15        | 27   | 32168704            | 33013541  | 844837         |
| WU_10.2_15_105914299 | DRGA0017581         | 15        | 139  | 105914299           | 113466125 | 7551826        |
| Ландрас              |                     |           |      |                     |           |                |
| ALGA0007502          | ASGA0005757         | 1         | 429  | 218007243           | 245059321 | 27052078       |
| ASGA0019551          | MARC0071111         | 4         | 33   | 44916881            | 46453617  | 1536736        |
| ALGA0026478          | 4_96337082          | 4         | 46   | 95052349            | 96337082  | 1284733        |
| WU_10.2_6_14710455   | ASGA0027643         | 6         | 13   | 14710455            | 15000888  | 290433         |
| H3GA0021824          | ASGA0034157         | 7         | 10   | 57707210            | 58003125  | 295915         |
| ASGA0040607          | ALGA0122175         | 8         | 13   | 147636271           | 148079360 | 443089         |
| WU_10.2_10_31038967  | WU_10.2_10_31551278 | 10        | 15   | 31038967            | 31551278  | 512311         |
| INRA0043983          | INRA0044020         | 14        | 47   | 50203210            | 52003103  | 1799893        |
| INRA0060572          | MARC0042652         | 15        | 19   | 100370184           | 101514834 | 1144650        |

Продолжение таблицы 5

| Название SNP         |                      | Хромосома | pSNP | Позиция нуклеотидов |           | Длина,<br>п.н. |
|----------------------|----------------------|-----------|------|---------------------|-----------|----------------|
| Первая SNP           | Последняя SNP        |           |      | от                  | до        |                |
| Дюрок                |                      |           |      |                     |           |                |
| INRA0000110          | ASGA0000414          | 1         | 21   | 4339088             | 4850701   | 511613         |
| ALGA0005533          | WU_10.2_1_126104360  | 1         | 57   | 122843794           | 126104360 | 3260566        |
| MARC0040499          | ALGA0007180          | 1         | 64   | 199842169           | 202387914 | 2545745        |
| DRGA0001874          | ALGA0007720          | 1         | 133  | 223911877           | 230761167 | 6849290        |
| WU_10.2_2_1111       | ASGA0097367          | 2         | 14   | 1111                | 365309    | 364198         |
| WU_10.2_2_156302473  | WU_10.2_2_157431842  | 2         | 51   | 156302473           | 157431842 | 1129369        |
| ASGA0010890          | ALGA0014472          | 2         | 79   | 92273830            | 96429165  | 4155335        |
| ASGA0096719          | DIAS0001827          | 4         | 31   | 62027221            | 62973622  | 946401         |
| ASGA0026488          | M1GA0008064          | 5         | 37   | 84166238            | 85112008  | 945770         |
| MARC0036138          | WU_10.2_6_129197330  | 6         | 42   | 127482521           | 129197330 | 1714809        |
| DIAS0004785          | ALGA0035333          | 6         | 50   | 47573330            | 51911721  | 4338391        |
| ASGA0033427          | DIAS0002292          | 7         | 16   | 49667222            | 50231835  | 564613         |
| ALGA0041169          | INRA0025403          | 7         | 17   | 49071381            | 49471090  | 399709         |
| ASGA0032187          | ASGA0032282          | 7         | 57   | 31107219            | 32728356  | 1621137        |
| WU_10.2_7_51175455   | DRGA0007600          | 7         | 59   | 51175455            | 53018942  | 1843487        |
| ALGA0041911          | MARC0049470          | 7         | 112  | 55622419            | 59080443  | 3458024        |
| DBWU0000670          | ALGA0047974          | 8         | 21   | 60770473            | 63475654  | 2705181        |
| DRGA0008674          | ALGA0048491          | 8         | 64   | 88497379            | 91443835  | 2946456        |
| WU_10.2_9_110799107  | ASGA0044085          | 9         | 23   | 110799107           | 112012220 | 1213113        |
| WU_10.2_10_2771      | MARC0026604          | 10        | 14   | 2771                | 241937    | 239166         |
| ASGA0054801          | H3GA0053469          | 12        | 20   | 46194492            | 46592669  | 398177         |
| WU_10.2_13_183393475 | ALGA0072857          | 13        | 10   | 183393475           | 183927662 | 534187         |
| DRGA0013280          | ALGA0110171          | 13        | 16   | 191245245           | 192020407 | 775162         |
| ASGA0059350          | WU_10.2_13_186219264 | 13        | 20   | 185586027           | 186219264 | 633237         |
| ASGA0090184          | MARC0068947          | 13        | 64   | 28561383            | 30329083  | 1767700        |
| WU_10.2_13_186567034 | ASGA0059468          | 13        | 95   | 186567034           | 190183426 | 3616392        |
| ASGA0066174          | ASGA0066192          | 14        | 10   | 123083682           | 123409339 | 325657         |
| ASGA0064195          | ASGA0064211          | 14        | 22   | 71748627            | 72469714  | 721087         |
| H3GA0038663          | H3GA0038713          | 14        | 28   | 8228429             | 8791250   | 562821         |
| ALGA0079001          | DIAS0002958          | 14        | 28   | 81831892            | 82760505  | 928613         |
| DRGA0014234          | ALGA0079760          | 14        | 28   | 96445285            | 97365662  | 920377         |
| MARC0008762          | ASGA0063279          | 14        | 45   | 49674562            | 51020901  | 1346339        |
| ASGA0064767          | DRGA0014194          | 14        | 80   | 88423846            | 91465078  | 3041232        |
| ALGA0079993          | ALGA0117921          | 14        | 371  | 100643377           | 115960354 | 15316977       |
| ALGA0085478          | H3GA0044463          | 15        | 110  | 60911648            | 66442766  | 5531118        |
| Дикие кабаны         |                      |           |      |                     |           |                |
| MARC0092489          | CASI0010228          | 1         | 20   | 194134592           | 195518093 | 1383501        |
| INRA0014030          | MARC0085408          | 4         | 32   | 49917991            | 52023825  | 2105834        |
| M1GA0018793          | ASGA0064288          | 14        | 11   | 75326862            | 75632971  | 306109         |

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) (рисунки 28–31). Однако представленность типов QTL отражает особенности селекционных стратегий для каждой из породы. Так, у свиней породы дюрок, относительно других пород и кабанов, QTL, связанные с мясной продуктивностью, наиболее представлены и составили более 80%. Наибольшую значимость показали признаки, ответственные за состав жирных кислот в мясе (содержание моновенасыщенных жирных кислот, стеариновой кислоты, насыщенных жирных кислот, пальмитолеиновой кислоты, пальмитиновой кислоты и олеиновой кислоты). Это согласуется с целенаправленной работой по совершенствованию породы дюрок, направленной на повышения признаков мясной продуктивности и качества мяса.

У свиней крупной белой породы, относительно других групп, значительно больше QTL (16,7%), связанных со здоровьем. Из признаков выделяется устойчивость/восприимчивость к актинобактериям плевропневмонии. Этот микроорганизм вызывает тяжелое высококонтагиозное заболевание свиней, часто отмечается внезапное начало болезни и короткое клиническое течение, высокая заболеваемость и смертность. Результаты Reiner с соавторами (Reiner et al., 2014) указывают на генетический фон устойчивости свиней к *Actinobacillus pleuropneumoniae*. Можно предположить, что исследуемые нами свиньи крупной белой породы имеют генетическую устойчивость к плевропневмонии. Помимо этого, у свиней крупной белой породы QTL имели значимое обогащение признаками, ответственными за количество сосков, а QTL, связанные с репродуктивными показателями, составили 5,8%. У свиней породы ландрас доля QTL, связанных с репродуктивными качествами, составила 5,4%. Селекционная работа над породами крупная белая и ландрас направлена, главным образом, на

повышение репродуктивных показателей. Доля репродуктивных показателей у свиней пород крупная белая и ландрас в областях топ HBD составила около 5,5–6,0%, что значительно выше относительно породы дюрок (1,8%), но при этом практически на одном уровне с кабанами (5,0%).

У диких кабанов значимое обогащение в областях топ HBD показали признаки, связанные с параметрами крови.

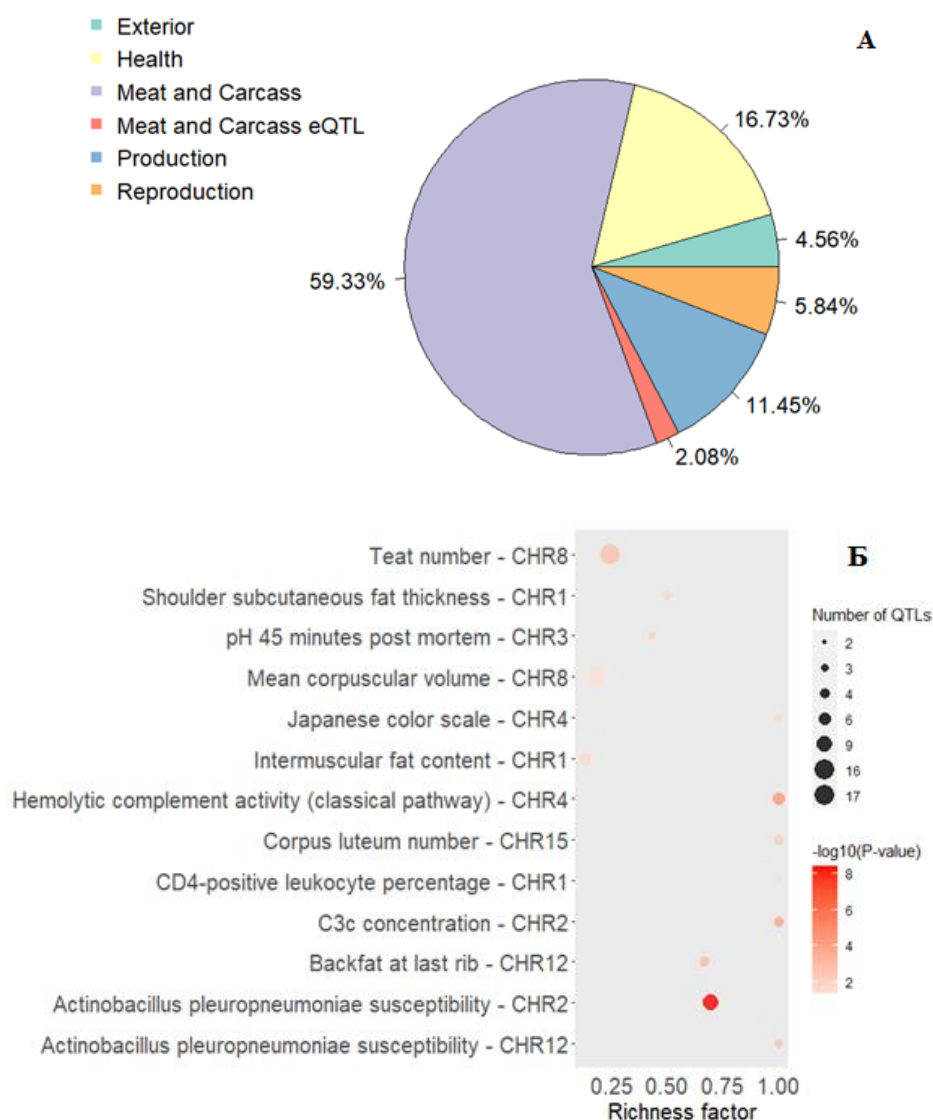


Рисунок 28 – Обогащение топ HBD локусами количественных признаков у свиней пород крупная белая: А – распределение QTL в топ HBD; Б – обогащение топ HBD

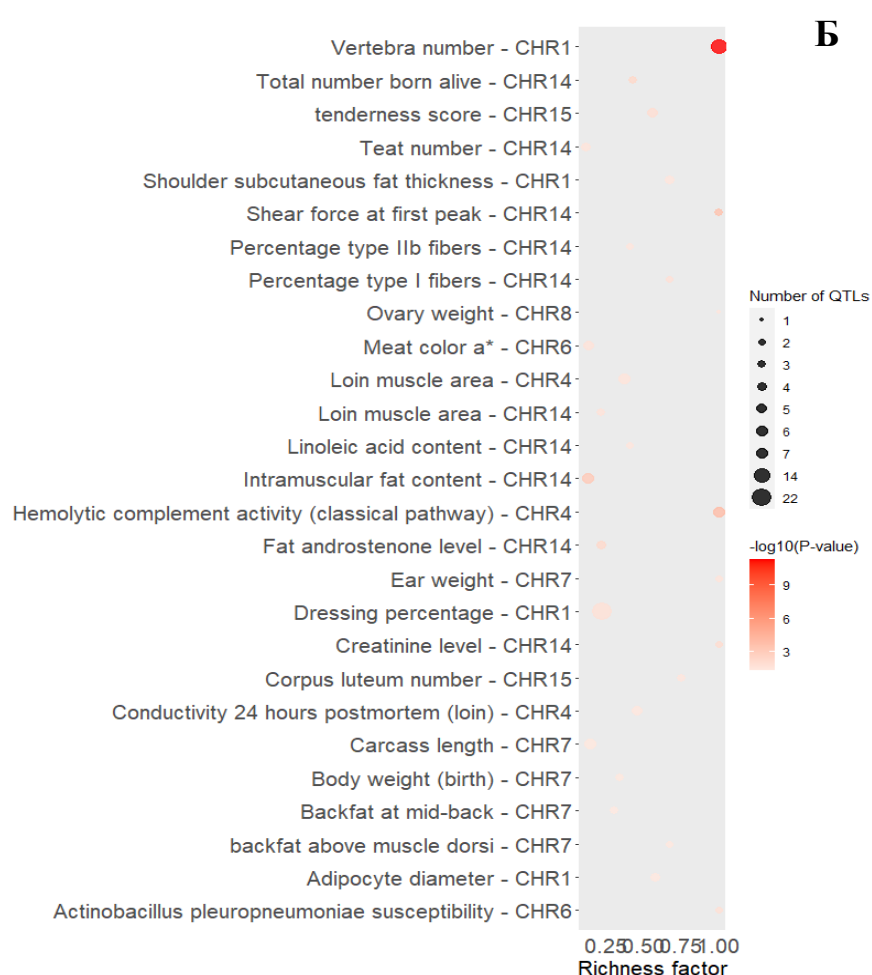
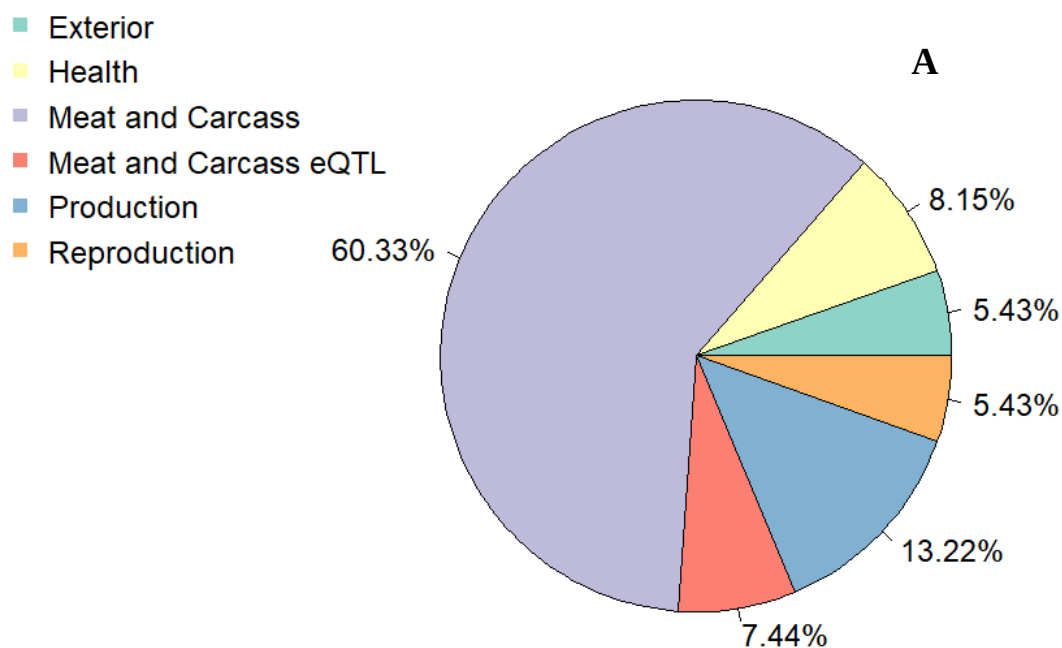


Рисунок 29 – Обогащение топ HBD локусами количественных признаков у свиней породы ландрас: А – распределение QTL в топ HBD; Б – обогащение топ HBD

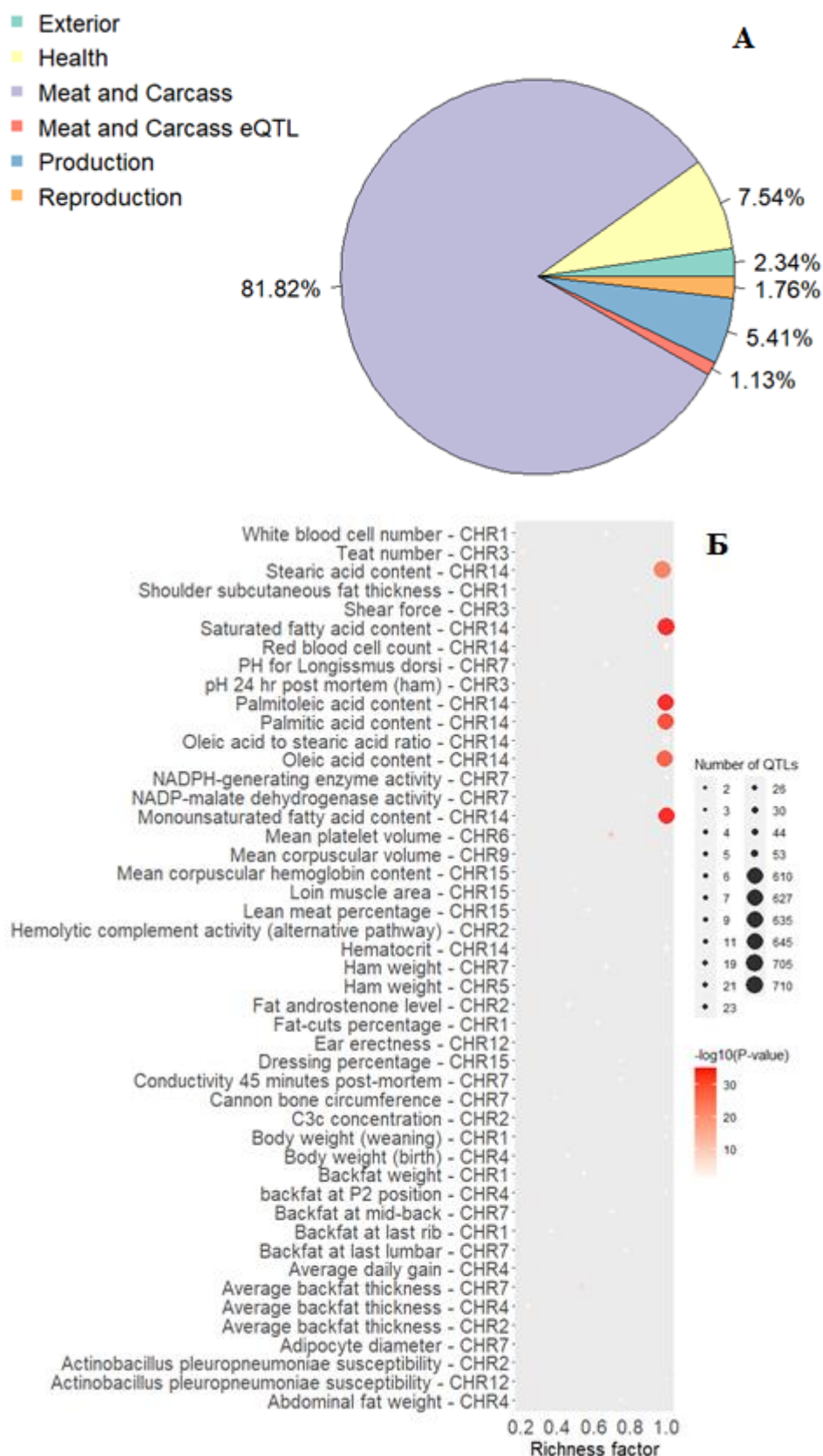


Рисунок 30 – Обогащение топ HBD локусами количественных признаков у свиней породы дюрок: А – распределение QTL в топ HBD, Б – обогащение топ HBD

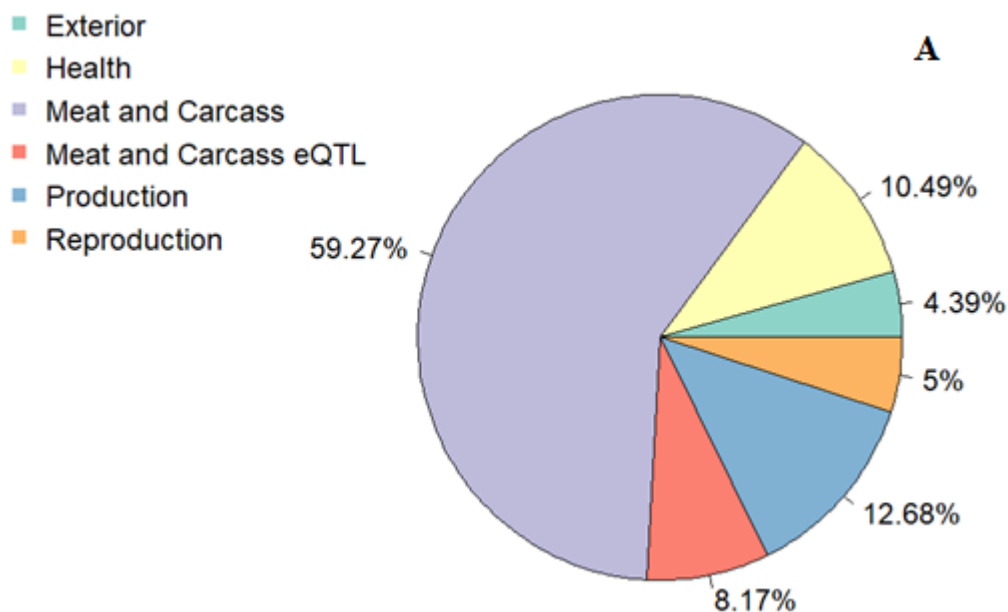


Рисунок 31 – Обогащение топ HBD локусами количественных признаков у диких кабанов: А – распределение QTL в топ HBD, Б – обогащение топ HBD у диких кабанов

Интересно отметить показатель активности аспартатаминотрансферазы (АСТ). АСТ – фермент, также известный как сывороточная глутаминовая щавелевоуксусная трансаминаза, присутствует в большей степени в клетках сердца и печени. Повышенный уровень АСТ в сыворотке рассматривают в качестве биомаркера мышечного стресса, тогда как высокий уровень АСТ связан с повреждением печени (Bovo et al., 2019). При этом исследования Schielke с соавторами (Schielke et al., 2019) показали, что дикие кабаны крайне подвержены поражению печени инфекционной природы (гепатиту Е), и вирус присутствует в популяции в течение длительного времени. Печень считается основным местом репликации вируса гепатита Е у свиней и важным органом в патогенезе заболевания (Risalde et al., 2017). Какую роль при этом играет гомозиготная область, связанная с активностью аспартатаминотрансферазы, на сегодняшний день неизвестно, но вполне допустимо, что области HBD вносят свой вклад в восприимчивость кабанов к вирусу гепатита Е.

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 кабанов покрыта сегментами HBD, предположительно происходящими от предков, живших около 206 лет назад. 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 кабанов. В областях топ HBD значимое обогащение показали признаки, связанные с параметрами крови, которые возможно интерпретировать в качестве предрасположенности 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 крупной белой породы, относительно других групп, значительно больше QTL (16,7%), связанных со здоровьем.

### **3.2 Анализ генетического разнообразия у свиней крупной белой породы из различных селекционных центров**

В течение длительного периода domestikации 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, а также обладающим высокой устойчивостью к заболеваниям. В дальнейшем процессы интенсификации захлестнули все мировое животноводство, в том числе Россию, и на смену низкопродуктивным свиньям пришли животные с высокой скоростью роста, высокими репродуктивными показателями, низким уровнем потреблением 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 крупной белой породы, являются всего 5–6 мировых генетических центров, при этом каждый из них имеет собственные целевые показатели и селекционные стандарты. Несмотря на то, что свиньи представляют одну

породу, они имеют внутripородную изменчивость, связанную с особенностями фенотипов селекционно-значимых признаков.

Для исследования внутripородной изменчивости свиней крупной белой породы были созданы три группы: КБ1 – образцы свиней крупной белой породы отечественной селекции; КБ2 и КБ3 – образцы свиней современной селекции из двух селекционных центров.

Результаты анализа главных компонент показали, что все три группы четко дифференцируются в отдельные классы (рисунок 32). Первые три компоненты PC1, PC2 и PC3 содержат, соответственно, 14,4; 5,8 и 1,5% дисперсии генетической дифференциации между свиньями крупной белой породы различных групп. По PC1 все группы различаются, при этом свиньи отечественной селекции занимают промежуточное положение, а в наибольшей степени различия проявляются между группами современной селекции. По PC2 свиньи современной селекции сбиваются в один кластер и дистанцируются от свиней отечественной селекции. По PC3 в большей степени прослеживается внутригрупповая изменчивость в группах КБ1 и КБ3.

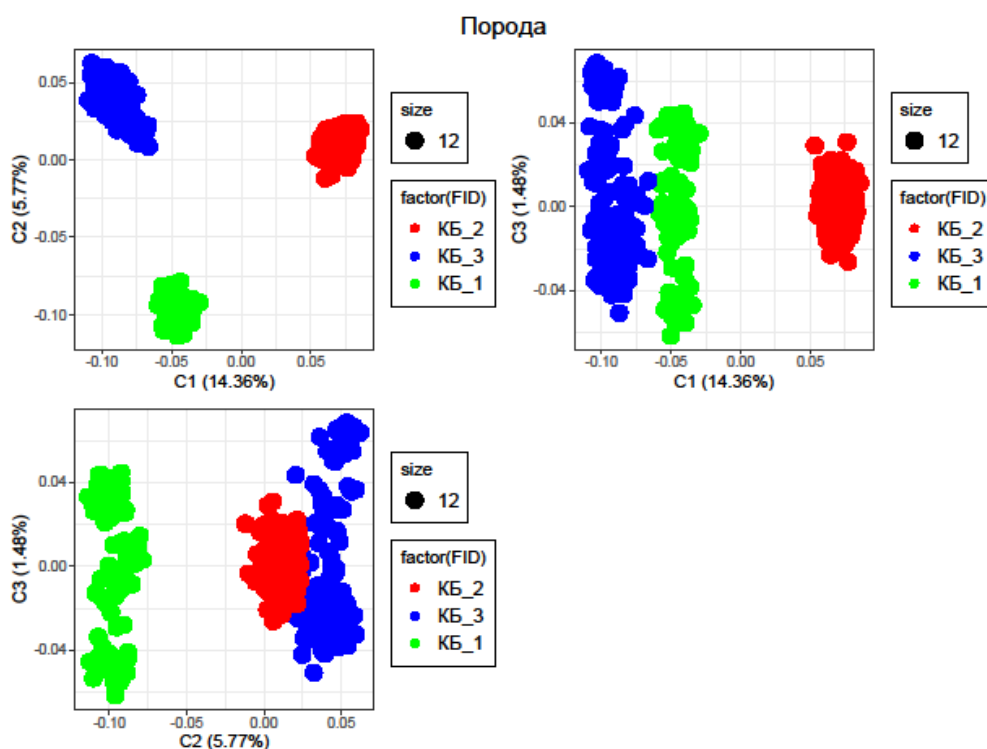


Рисунок 32 – Анализ главных компонент (РСА) для свиней крупной белой породы

Для проведения анализа примеси (Admixture) определили оптимальное число кластеров путем применения кросс-валидационного метода. Оптимальным числом кластеров считается точка абсциссы, в которой кривая имеет перегиб (рисунок 33). Расчеты показали, что оптимальное число кластеров (К) для изучаемых популяций равно  $K = 3$ . Результаты анализа примесей для  $K = 2 \dots 5$  представлен на рисунке 33.

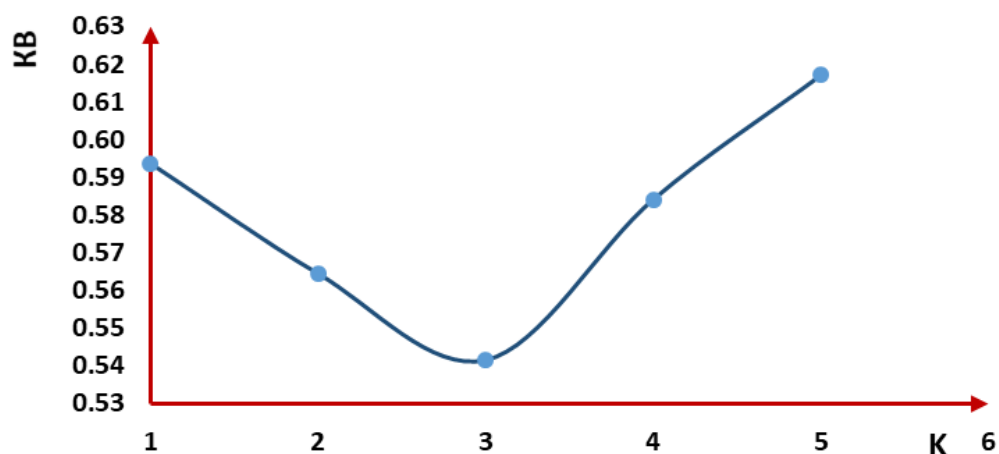


Рисунок 33 – График кросс-валидационной ошибки

График Admixture (рисунок 34) отражает степень смешанности (admixture) для трех линий свиней породы крупная белая (КБ1, КБ2 и КБ3) при варьировании предполагаемого числа предковых популяций (К) от 2 до 5. Каждая вертикальная линия соответствует отдельному генотипированному индивидууму, а сегменты разного цвета внутри линии отражают пропорциональное участие генома данного индивидуума от каждой из К гипотетических предковых популяций. Группировка индивидуумов по горизонтали соответствует их принадлежности к исходным трем линиям.

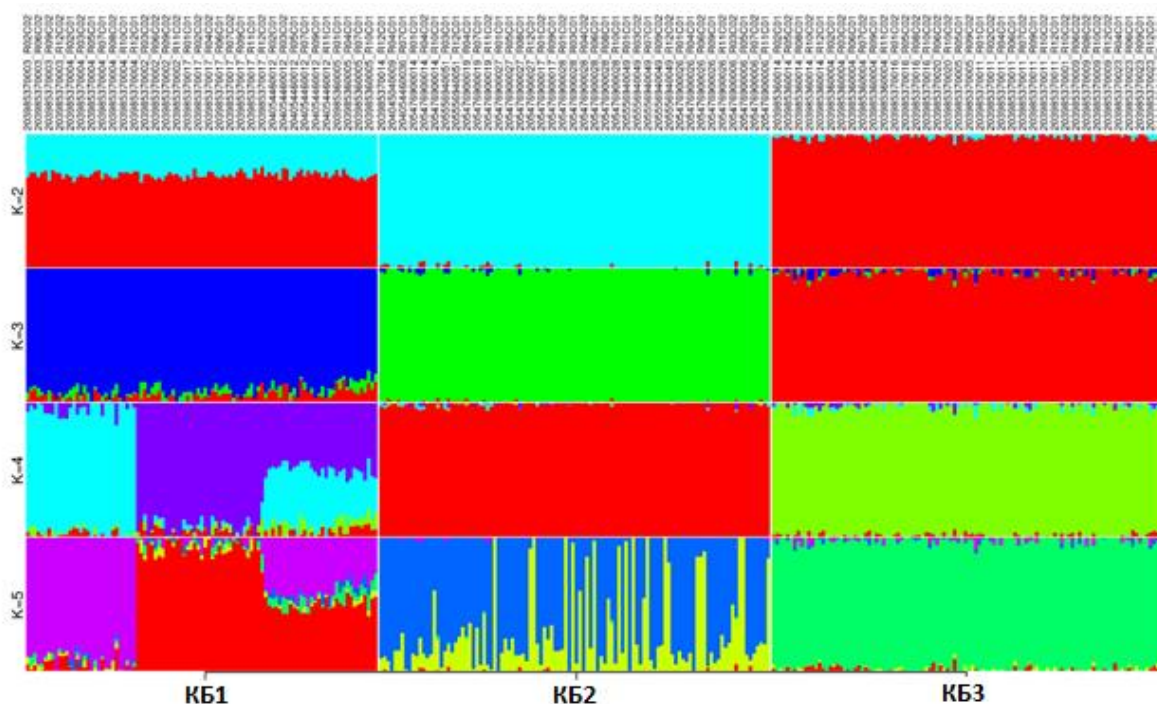


Рисунок 34 – График примесей (Admixture) для свиней крупной белой породы при количестве кластеров от 4 до 9 ( $K = 4 \dots 9$ )

При  $K=3$  наблюдается наиболее четкое разделение исследуемых групп, соответствующее априорной классификации на три линии. Линии KB1, KB2 и KB3 преимущественно ассоциированы с тремя различными, четко выраженными предковыми компонентами (условно синим, зеленым и красным соответственно), демонстрируя высокую генетическую однородность внутри каждой линии и минимальный уровень межпопуляционной примеси. Это свидетельствует о значительной генетической дифференциации между тремя исследуемыми линиями свиней на данном уровне иерархического кластерного анализа.

Увеличение значения  $K$  до 4 и 5 позволяет выявить более тонкую генетическую структуру и потенциальную субструктурированность внутри отдельных линий. При  $K=4$  линия KB1 начинает демонстрировать признаки гетерогенности, расщепляясь на два основных компонента с варьирующими пропорциями у разных особей. При  $K=5$  выраженная субструктура проявляется и в линии KB2, где особи характеризуются значительной

вариабельностью долей двух преобладающих предковых компонентов. В то же время, линия КБ3 сохраняет относительную генетическую гомогенность даже при более высоких значениях  $K$ , указывая на ее большую консолидированность по сравнению с двумя другими линиями.

Идентификация областей HBD позволила определить коэффициенты геномного инбридинга у свиней крупной белой породы. Средние значения коэффициента геномного инбридинга варьировали от 0,24 у свиней КБ1 до 0,29 у КБ2 (таблица 6).

Таблица 6 – Коэффициенты геномного инбридинга у свиней крупной белой породы

| Название | Значение коэффициента геномного инбридинга |         |              |
|----------|--------------------------------------------|---------|--------------|
|          | минимальное                                | среднее | максимальное |
| КБ1      | 0,1595                                     | 0,2394  | 0,3233       |
| КБ2      | 0,2475                                     | 0,2942  | 0,3802       |
| КБ3      | 0,1984                                     | 0,2535  | 0,3339       |

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 (рисунки 35–37).

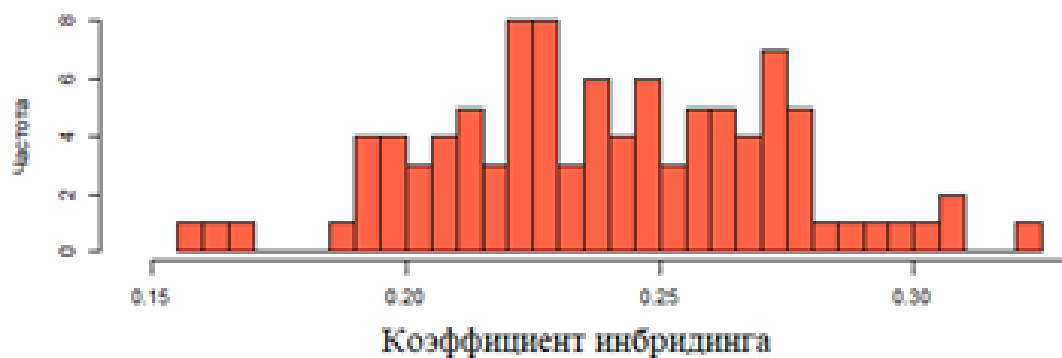


Рисунок 35 – Распределение коэффициента инбридинга у свиней крупной белой пород в группе КБ1

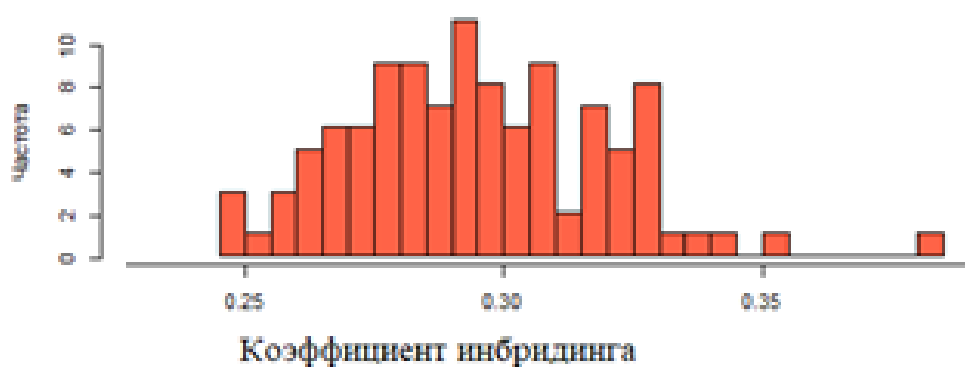


Рисунок 36 – Распределение коэффициента инбридинга у свиней крупной белой пород в группе КБ2; С – в группе КБ3

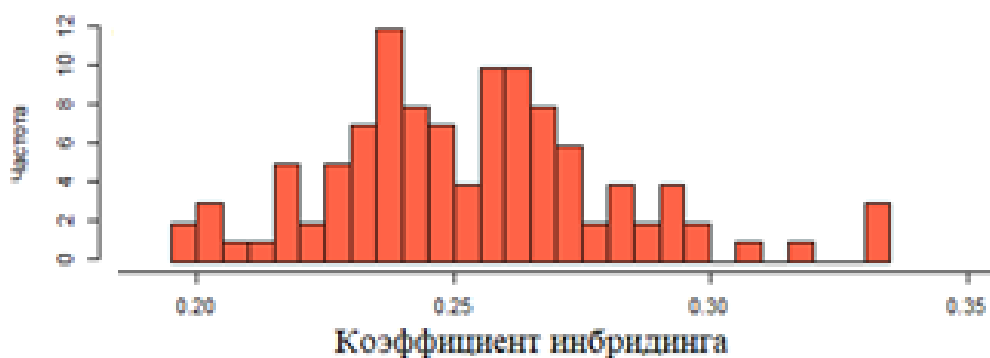


Рисунок 37 – Распределение коэффициента инбридинга у свиней крупной белой пород в группе КБ3

На представленном рисунке 38 визуализировано распределение гомозиготных по происхождению (HBD) сегментов в геномах свиней породы крупная белая группы КБ1. Анализ графика показывает, что основной вклад в общую аутозиготность генома современных свиней крупной белой породы вносят HBD-сегменты, унаследованные от предков, живших в промежутке от 8 (класс 16) до 128 (класс 256) поколений назад. Медианные значения доли генома и размеры межквартильных размахов для этих классов наиболее высоки. Пик медианной доли HBD приходится на класс 256, что соответствует общему происхождению от предков, живших примерно 128 поколений назад. В то же время, вклад недавнего инбридинга (классы 2 и 4, соответствующие 1 и 2 поколениям назад) и очень древнего общего происхождения (класс 512, 256 поколений назад) в наблюдаемую гомозиготность минимален, что отражается в низких медианных значениях и узких распределениях для этих классов.

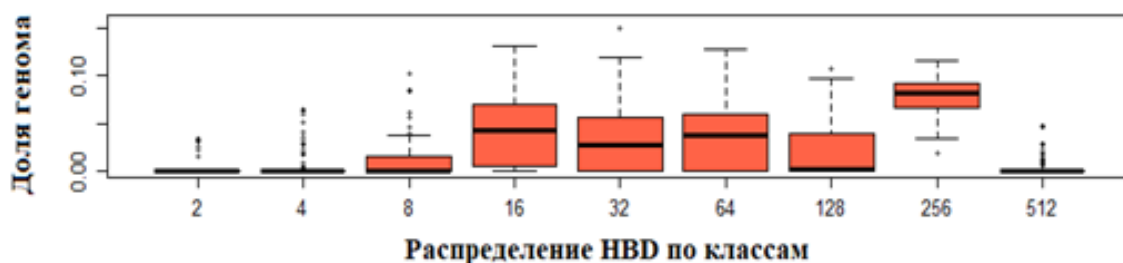


Рисунок 38 – Аутозиготность в соответствии с возрастом предков свиней коммерческих пород в группе КБ1

Значительная высота ящиков и длина усов для классов от 16 до 256, особенно для класса 256, указывает на существенную вариабельность в уровне HBD между отдельными животными для этих периодов происхождения. Индивидуальные истории разведения и родословные свиней в популяции различаются, и вклад предков, живших от 8 до 128 поколений назад, неодинаков для разных особей. Картина распределения HBD по классам отражает демографическую историю и практики разведения данной группы,

показывая, что основная часть накопленного инбридинга связана с предками из "среднего" прошлого, а не с самыми недавними или очень отдаленными временами.

Анализ распределения сегментов HBD для группы KB2 (рисунок 39) показывает иную картину по сравнению с группой KB1. Основной вклад в аутозиготность здесь вносят предки, жившие от 8 (класс 16) до 64 (класс 128) поколений назад. Наивысшая медианная доля HBD наблюдается для класса 128, что соответствует общим предкам, жившим примерно 64 поколения назад. Классы 16 и 32 также показывают существенный вклад, сравнимый между собой, но несколько ниже пикового значения. Интересно, что класс 64 (32 поколения) показывает меньшую медианную долю HBD, чем классы 16 и 32. Вклад как самых недавних предков (классы 2 и 4), так и относительно далеких (классы 256 и 512, соответствующие 128 и 256 поколениям) минимален, с медианами, близкими к нулю.

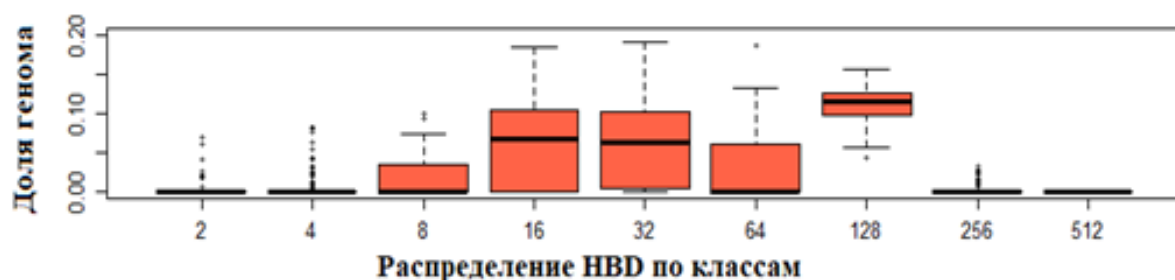


Рисунок 39 – Аутозиготность в соответствии с возрастом предков свиней коммерческих пород в группе KB2

Наблюдается значительная вариабельность в доле HBD между особями, особенно для классов 16, 32 и 128, что видно по высоте ящиков и длине усов. Однако ключевое отличие от группы KB1 заключается в смещении пика HBD с более давних предков (128 поколений назад для группы KB1) на более недавних (64 поколения назад). Это указывает на другую демографическую историю или специфику селекции в этой популяции, где наибольшее влияние на накопленный инбридинг оказали события или предки, относящиеся к периоду около 64 поколений назад. Резкое снижение вклада предков, живших



128 и 256 поколений назад (классы 256 и 512), также контрастирует с группой КБ1.

Анализ группы КБ3 (рисунок 40) выявляет сложную картину вклада предков разной давности в аутозиготность. Заметный вклад вносят предки, жившие 4 (класс 8), 8 (класс 16), 16 (класс 32), 64 (класс 128) и 128 (класс 256) поколений назад, что видно по поднятым оранжевым бокс-плотам. Наибольшая медианная доля HBD наблюдается для класса 128 (64 поколения назад). Классы 16 и 32 также демонстрируют высокие медианные значения, указывая на значительное влияние предков, живших 8 и 16 поколений назад. Интересной особенностью является относительно низкая медианная доля HBD для класса 64 (32 поколения назад) по сравнению с окружающими его классами (32 и 128). Как и для групп КБ1 и КБ2, вклад самых недавних (классы 2, 4) и самых древних (класс 512) предков минимален.

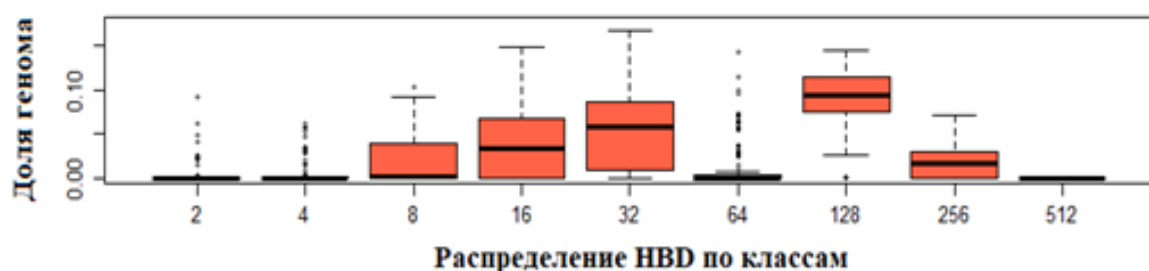


Рисунок 40 – Аутозиготность в соответствии с возрастом предков свиней коммерческих пород в группе КБ3

Значительная высота бокс-плотов и наличие усов/выбросов для классов 8, 16, 32, 128 и 256 свидетельствуют о высокой вариабельности уровня HBD между особями в группе КБ3, особенно для сегментов, унаследованных от предков, живших от 4 до 128 поколений назад. Пик HBD у предков 64-поколенной давности (класс 128) указывает на ключевой период или группу основателей в истории этой линии. Неожиданно низкий вклад предков 32-поколенной давности (класс 64) может отражать специфические события в разведении, такие как ауткроссинг или изменения в структуре популяции в тот период. В целом, распределение HBD в группе КБ3 отражает уникальную

комбинацию исторических событий, повлиявших на генетическую структуру этой конкретной группы свиней.

В целом уровень аутозиготности у свиней крупной белой породы 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%) (таблица 7).

Таблица 7 – Сводная статистика реализованной аутозиготности в каждом классе HBD (в процентах от генома)

| Классы/<br>статистика | R_4      | R_8      | R_16     | R_32     | R_64     | R_128    | R_256    | R_512    |
|-----------------------|----------|----------|----------|----------|----------|----------|----------|----------|
| Отечественная         |          |          |          |          |          |          |          |          |
| mean                  | 0,00604  | 0,01128  | 0,04483  | 0,03356  | 0,03560  | 0,02225  | 0,07993  | 0,00402  |
| se                    | 0,01459  | 0,02086  | 0,03710  | 0,03509  | 0,03411  | 0,02959  | 0,02004  | 0,01006  |
| min                   | 1,44E-27 | 9,55E-28 | 7,42E-23 | 8,88E-12 | 3,38E-21 | 1,81E-22 | 0,01872  | 8,13E-61 |
| max                   | 0,06444  | 0,10185  | 0,13037  | 0,14957  | 0,12683  | 0,10797  | 0,11643  | 0,04799  |
| Импортная 1           |          |          |          |          |          |          |          |          |
| mean                  | 0,00701  | 0,01730  | 0,06489  | 0,06173  | 0,02813  | 0,10966  | 0,00294  | 9,61E-08 |
| se                    | 0,01704  | 0,02479  | 0,05296  | 0,04957  | 0,04227  | 0,02406  | 0,00686  | 6,05E-07 |
| min                   | 5,23E-17 | 5,47E-19 | 1,08E-13 | 2,37E-25 | 6,29E-26 | 0,04317  | 5,45E-36 | 3,53E-59 |
| max                   | 0,08172  | 0,09991  | 0,18417  | 0,19123  | 0,18533  | 0,15613  | 0,03357  | 5,69E-06 |
| Импортная 2           |          |          |          |          |          |          |          |          |
| mean                  | 0,00649  | 0,02131  | 0,04197  | 0,05442  | 0,01430  | 0,09135  | 0,01878  | 2,37E-07 |
| se                    | 0,01466  | 0,02849  | 0,04127  | 0,04269  | 0,02980  | 0,03015  | 0,01643  | 1,08E-06 |
| min                   | 1,84E-19 | 6,62E-17 | 6,68E-13 | 4,36E-14 | 2,40E-32 | 7,75E-08 | 3,09E-16 | 1,65E-81 |
| max                   | 0,06245  | 0,10337  | 0,14771  | 0,16669  | 0,14361  | 0,14521  | 0,07107  | 6,86E-06 |

Учет эффективного размера популяции ( $N_e$ ) при анализе участков гомозиготности (ROH) является критически важным в связи с его фундаментальным влиянием на генетическую архитектуру популяций. Эффективный размер популяции определяет скорость накопления инбридинга и интенсивность генетического дрейфа, что напрямую отражается на распределении и характеристиках ROH (Wright, 1931; Charlesworth, 2009). В малых популяциях ( $N_e < 100$ ) усиление генетического дрейфа приводит к ускоренной фиксации аллелей, увеличивая вероятность возникновения длинных ROH, которые ассоциированы с недавними событиями инбридинга (Keller et al., 2011). Напротив, в популяциях с большим  $N_e$  преобладают короткие ROH, формирующиеся за счет древних общих предков, что отражает длительную историю рекомбинаций и стабильный генетический поток (Ceballos et al., 2018).

Исторические демографические события, такие как бутылочные горлышки или эффекты основателя, существенно модифицируют распределение ROH. Например, резкое сокращение  $N_e$  приводит к накоплению протяженных гомозиготных участков, которые могут быть ошибочно интерпретированы как следствие недавнего инбридинга, хотя на деле отражают последствия дрейфа (Pemberton et al., 2012). Это подчеркивает необходимость интеграции данных о  $N_e$  для дифференциации временных рамок инбридинга и точной реконструкции демографической истории (McQuillan et al., 2008).

Методологически, игнорирование  $N_e$  искажает калибровку параметров при выявлении ROH. Программные инструменты, такие как PLINK (Purcell et al., 2007) и BCFtools (Danese et al., 2021), используют пороговые значения длины и плотности ROH, которые зависят от ожидаемого уровня гетерозиготности, определяемого  $N_e$ . Без учета этого параметра возникает риск недооценки или переоценки генетического груза, связанного с инбредной депрессией (Kardos et al., 2016). В популяциях домашних животных, где  $N_e$  часто искусственно занижен из-за селекции, ROH служат индикатором не

только инбридинга, но и истории управляемого скрещивания (Bertolini et al., 2018).

Кроме того,  $N_e$  играет ключевую роль в прогнозировании устойчивости популяций к генетическим угрозам. В природоохранной биологии анализ RОН с учетом  $N_e$  позволяет оценить риск вымирания малых популяций, где накопление вредных аллелей в гомозиготных участках может снижать приспособленность (Frankham et al., 2017). Например, у видов, переживших бутылочное горлышко, высокая частота длинных RОН коррелирует с повышенной уязвимостью к заболеваниям (Robinson et al., 2019).

Интеграция  $N_e$  в анализ RОН обеспечивает точность интерпретации генетических данных, позволяя различить эффекты исторического дрейфа, инбридинга и демографических событий. Это особенно актуально в прикладных контекстах, таких как селекция сельскохозяйственных видов, где управление  $N_e$  направлено на минимизацию генетических рисков (Doekes et al., 2021), и в охране биоразнообразия, где оценка  $N_e$  через RОН помогает разрабатывать стратегии восстановления популяций (Hoffman et al., 2014). Пренебрежение этим параметром ведет к некорректным выводам о генетическом здоровье и адаптивном потенциале популяций. Для калибровки найденных значений HBD мы рассчитали значения  $N_e$  популяций отечественной и импортной селекции с помощью программы SNeP (Barbato et al., 2015).

Согласно значениям эффективного размера популяции, соответствующие определенным поколениям, определили, что группы КБ1, КБ2 и КБ3 представляют собой генетически разнообразные популяции, которые демонстрируют различия в изменении эффективного размера популяции ( $N_e$ ) и неравновесии по сцеплению (LD) с течением времени (таблица 8). В группе КБ1  $N_e$  снижается относительно медленно, что может свидетельствовать о стабильности генетического состава и меньшей уязвимости к генетическим узким горлышкам. В группах КБ2 и КБ3 наблюдаются более выраженные колебания  $N_e$ , что может указывать на более

высокую восприимчивость к генетическим дрейфам и потерям генетического разнообразия, особенно в контексте изменения генетической структуры популяции.

Таблица 8 – Сводная статистика динамики эффективного размера популяции в группах КБ1, КБ2 и КБ3

| Параметр/группа                   | КБ1     | КБ2     | КБ3     |
|-----------------------------------|---------|---------|---------|
| Текущий $N_e$ (GenAgo = 13)       | 145     | 78      | 75      |
| LD (GenAgo = 13)                  | 0,0448  | 0,0803  | 0,0833  |
| Items (GenAgo = 13)               | 581,666 | 490,106 | 527,561 |
| Исторический $N_e$ (GenAgo = 984) | 1084    | 973     | 990     |
| Сокращение $N_e$ (раз)            | 7.5×    | 12.5×   | 13.2×   |
| LD (GenAgo = 959)                 | 0,2891  | 0,3065  | 0,3174  |

Данные о неравновесии по сцеплению, характеризующие степень генетической связи между соседними маркерами, также показали различия между породами. В группе КБ1 наблюдается относительно медленное снижение LD, что может указывать на сохранение стабильных генетических связей между маркерами на протяжении времени. В то же время в группах КБ2 и КБ3 LD снижается быстрее, что может быть признаком более интенсивного генетического дрейфа, а также возможного воздействия внутривидового скрещивания, приводящего к нарушению сцепления между маркерами.

Трендовые модели, такие как сплайны и экспоненциальные кривые, подтверждают различия в динамике изменения  $N_e$  между группами (рисунок 41). Группы КБ1 и КБ2 показывают более плавные изменения, в то время как в группе КБ3 наблюдается более выраженная нестабильность, что может свидетельствовать о генетическом сужении и потере разнообразия в популяции. Эти различия в динамике  $N_e$  подчеркивают необходимость учета специфических генетических процессов в каждой породе для эффективного управления генетическим ресурсом и сохранения долгосрочного разнообразия.

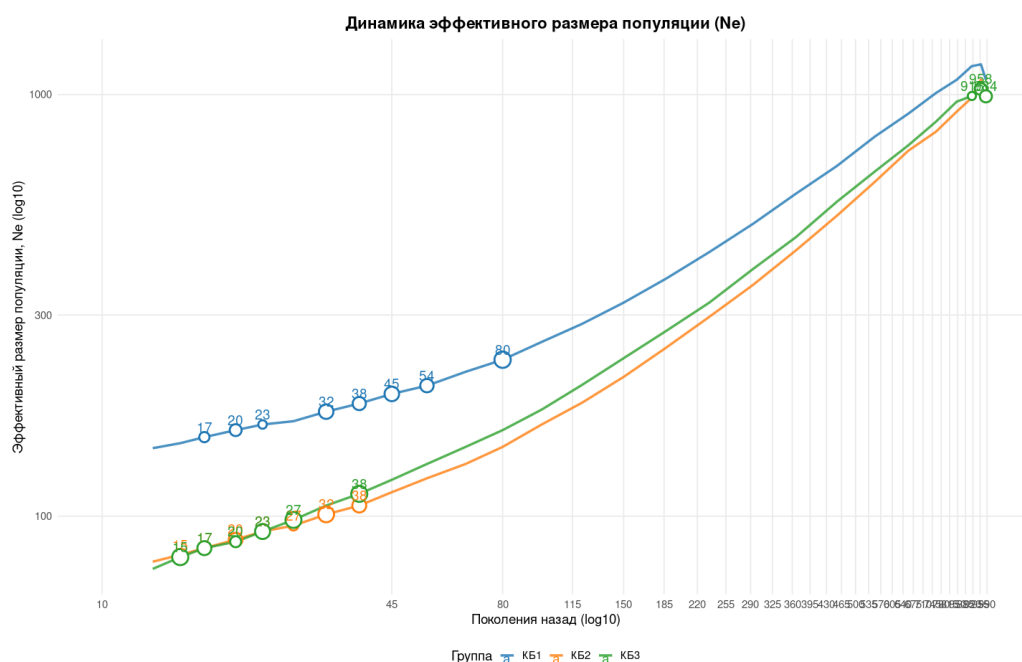


Рисунок 41 – Изменение эффективного размера популяции в группах. Окружности показывают значительные изменения в популяциях, радиус которых зависит от масштаба этих изменений

При оценке индивидуальной аутозиготности в группах можно отметить, что в целом особи придерживаются общей тенденции. Более компактно относительно центральной линии группируются свиньи в группе KB2 (рисунок 42), и после класса  $R_k = 256$  рост уровня аутозиготности прекращается у всех животных в группе. В группах KB1 и KB3 есть отдельные особи, которые незначительно отклоняются от общей траектории.

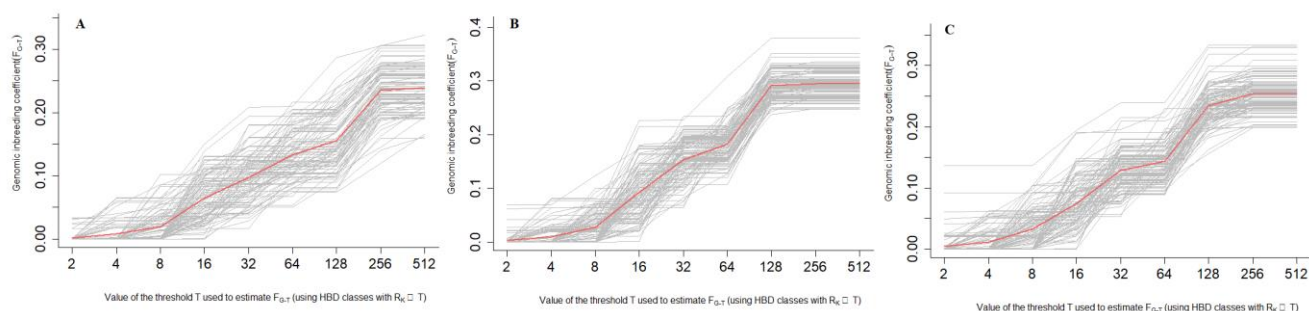
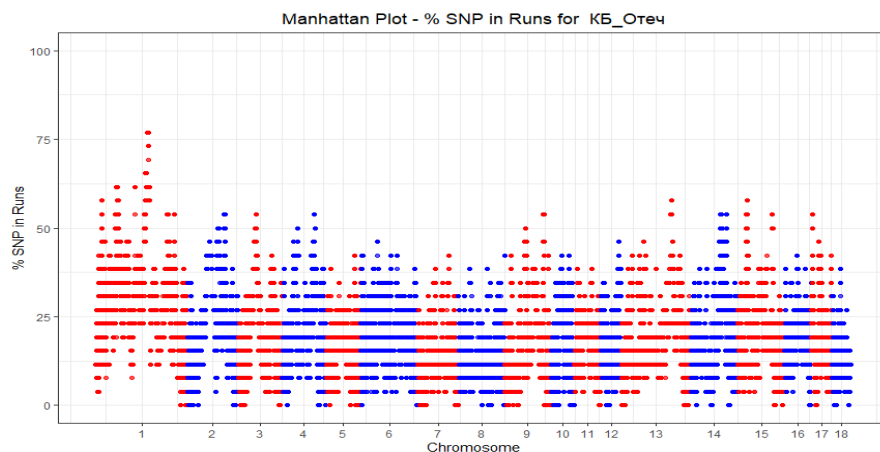


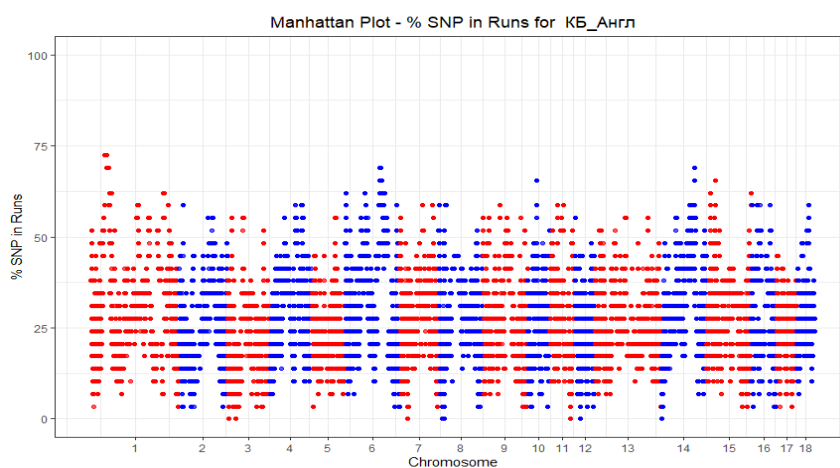
Рисунок 42 – Доля аутозиготности в соответствии с возрастом предков на индивидуальном уровне: А – в группе KB1; В – в группе KB2; С – в группе KB3

Частоты SNP (%) в обнаруженных HBD были оценены для каждой группы свиней КБ и нанесены на график в зависимости от положения SNP в аутосомах (рисунок 43).

А



В



С

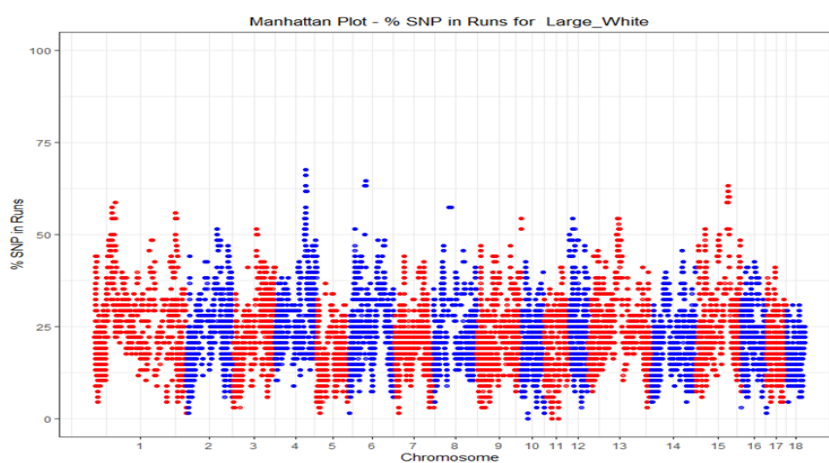


Рисунок 43 – Частоты SNP (%) в обнаруженных HBD у свиней КБ1 (А), КБ2 (В) и КБ3 (С)

Для каждой группы выбраны топ HBD при условии, что частота HBD составляет не менее 60% и входят не менее 10 SNP. В результате у КБ-1 определены 4 области, расположенные в SSC1 (таблица 9). У свиной группы КБ-2 обнаружены 10 областей, из которых 5 в SSC1, 2 в SSC6 и по одной в SSC10, SSC14 и SSC15. У свиной КБ-3 данным условиям соответствовало 7 сегментов, локализованных в хромосомах SSC4, SSC6 и SSC15.

Области топ HBD перекрывают QTL, среди которых наиболее представлены локусы, связанные с характеристиками мяса и туши, т.е. определяющие качество и пригодность свинины для потребления. У свиной КБ1 и КБ2 этот тип QTL с учетом генетических вариантов, влияющих на экспрессию гена или уровень активности гена в клетках (eQTL) составляет более 70% (рисунки 44, 45).

У свиной КБ3 (рисунок 46), относительно других групп, значительно больше QTL (16,7%), связанных со здоровьем. Помимо этого, у КБ3 QTL имеют значимое обогащение признаками, ответственными за количество сосков, а QTL, связанные с репродуктивными показателями, составляют 5,8%.



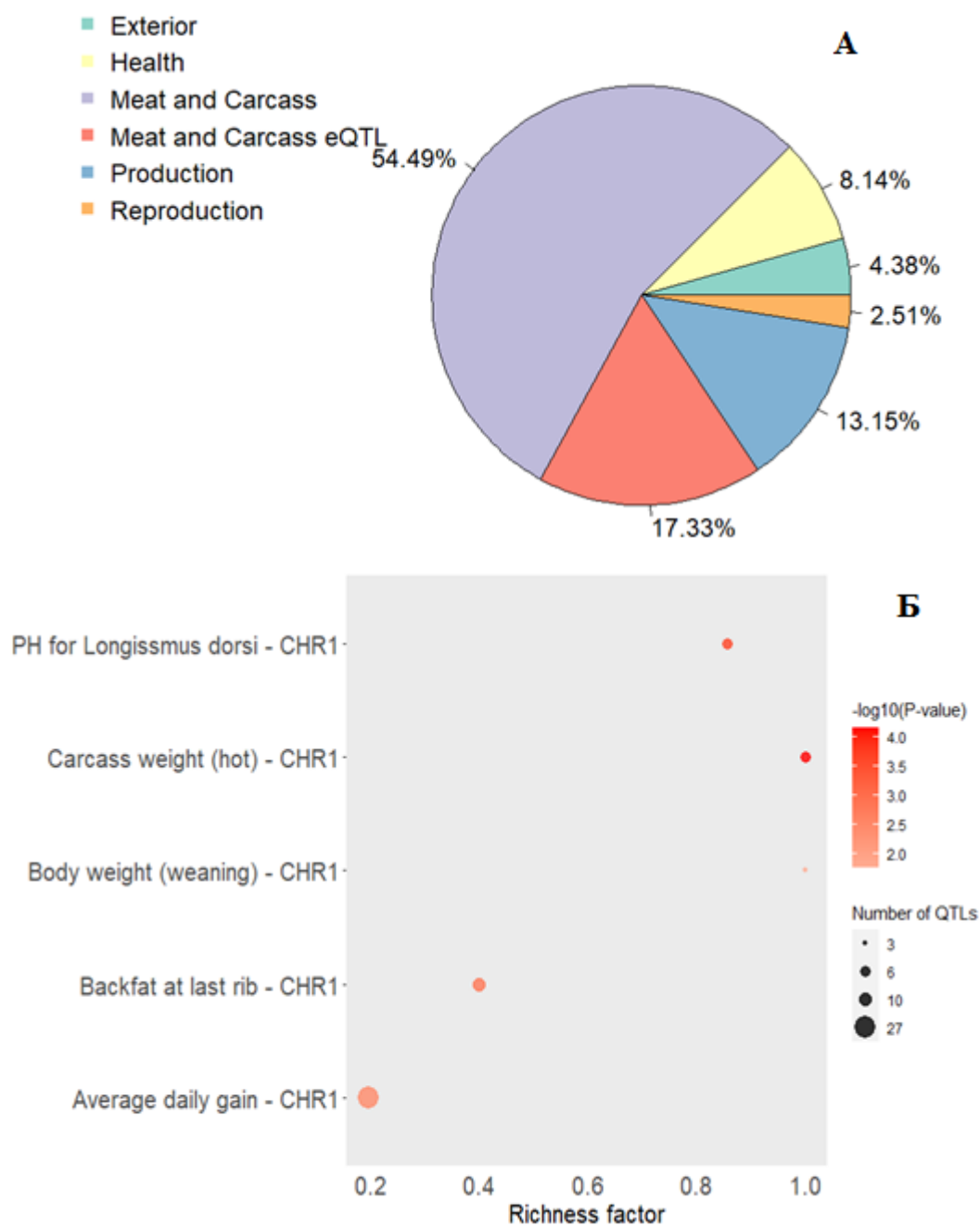


Рисунок 44 – Обогащение топ HBD локусами количественных признаков у свиней крупной белой породы: А – распределение QTL в топ HBD у КБ1, Б – обогащение топ HBD у КБ1

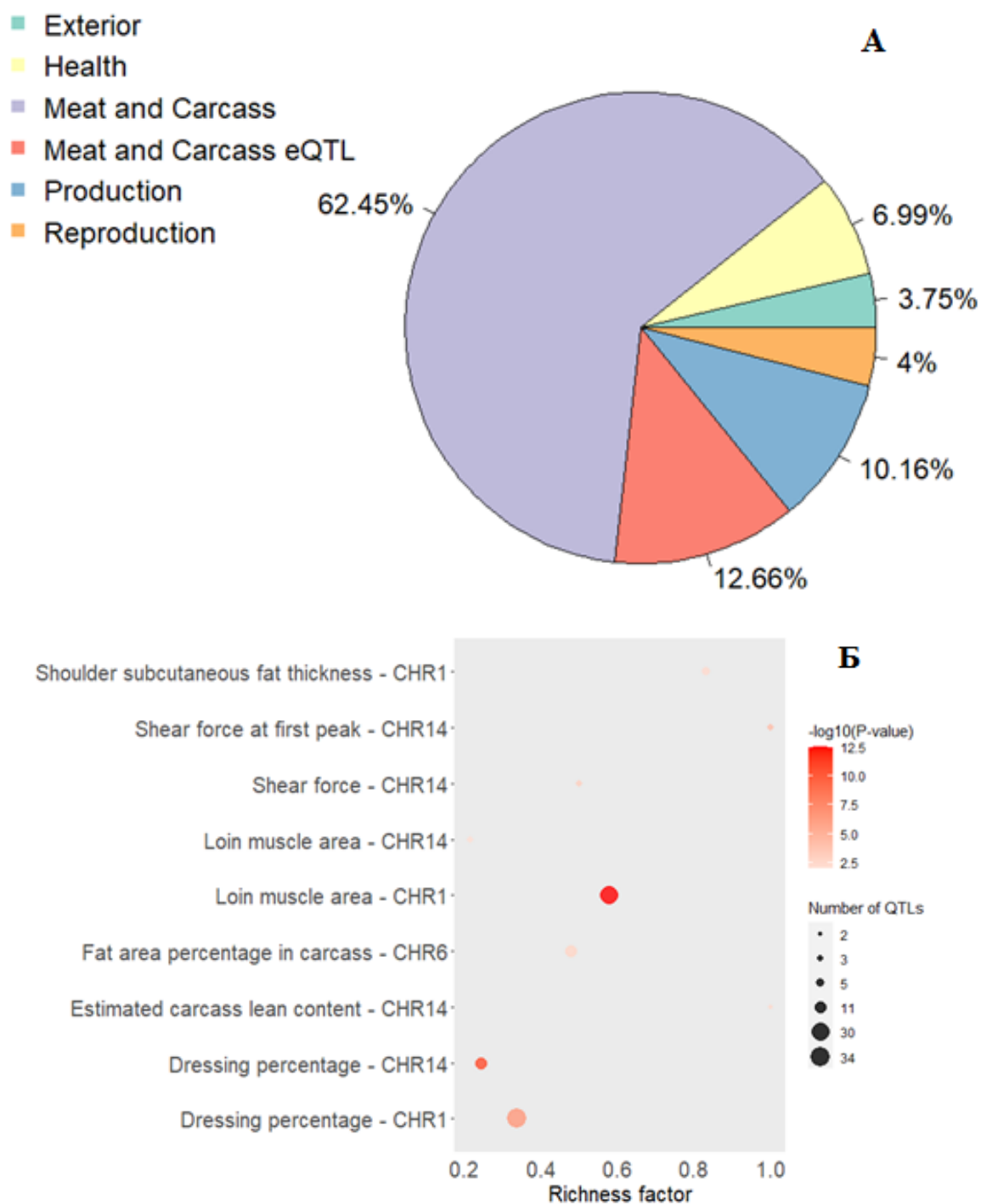


Рисунок 45 – Обогащение топ HBD локусами количественных признаков у свиней крупной белой породы: А – распределение QTL в топ HBD у KB2, Б – обогащение топ HBD у KB2

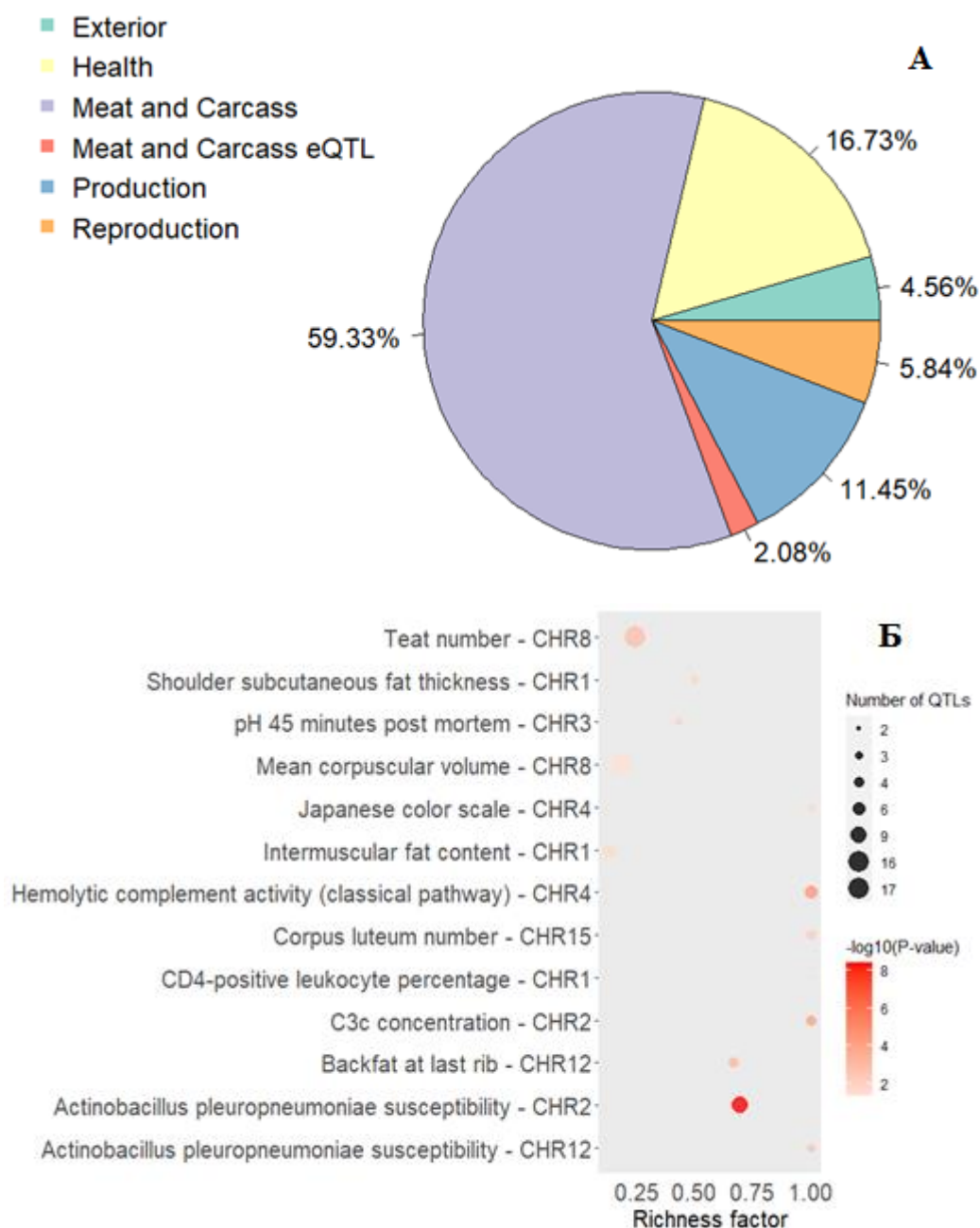


Рисунок 46 – Обогащение топ HBD локусами количественных признаков у свиней крупной белой породы: А – распределение QTL в топ HBD у КБЗ, Б – обогащение топ HBD у КБЗ

Для функциональной аннотации сегментов HBD были определены гены, локализованные в топ HBD у свиней крупной породы (таблица 9).

Таблица 9 – Гены, локализованные в топ HBD у свиней породы крупная белая

| Группа | Хромо-<br>сома | Кол-во<br>SNP | Первая<br>позиция | Последняя<br>позиция | Гены                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|--------|----------------|---------------|-------------------|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| КБ1    | 1              | 13            | 63704733          | 64301394             | <i>FHL5, GPR63, NDUFAF4, KLHL32</i>                                                                                                                                                                                                                                                                                                                                                                                                            |
|        | 1              | 14            | 116300444         | 117138985            | <i>PYGO1, DNAAF4, C15orf65, CCPG1, PIGB, RAB27, RSL24D1</i>                                                                                                                                                                                                                                                                                                                                                                                    |
|        | 1              | 17            | 149261646         | 150160033            | <i>ZNF407, CNDP1, CNDP2, DIPK1C, C18orf63, CYB5A, FBX015, TIMM21</i>                                                                                                                                                                                                                                                                                                                                                                           |
|        | 1              | 46            | 154796451         | 158940596            | <i>DSEL, CDH19, CDH7, SERPINB10, SERPINB7, SERPINB12, SERPINB8, SERPINB5, SERPINB2, SERPINB13, SERPINB11, VDS4B, PHLPP1, KDSR</i>                                                                                                                                                                                                                                                                                                              |
| КБ2    | 1              | 245           | 41940391          | 53513159             | <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              | 10            | 56563956          | 57069421             | <i>RNGTT</i>                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|        | 1              | 11            | 61986232          | 62371709             | –                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|        | 1              | 14            | 225017887         | 225614369            | <i>CEMIP2, ABHD17B, C9orf85, GDA</i>                                                                                                                                                                                                                                                                                                                                                                                                           |
|        | 1              | 18            | 227781084         | 229301982            | <i>TRPM6, C9orf40, NMRK1, CARNMT1, OSTF1, PCSK5</i>                                                                                                                                                                                                                                                                                                                                                                                            |
|        | 6              | 20            | 61261671          | 63135016             | <i>PEG3, AURKC, ZNF304, ZNF772, ZNF773, ZNF550, ZNF606, ZNF135, ZNF329, ZNF274, ZNF8, RPS5, ZNF584, ZNF446, SLC27A5, ZBTB45, TRIM28, CHMP2A, UBE2M, MZF1</i>                                                                                                                                                                                                                                                                                   |
|        | 6              | 84            | 107551091         | 113309905            | <i>RBBP8, CABLES1, TMEM241, RIOK3, RMC1, NPC1, ANKRD29, LAMA3, TTC39C, CABYR, OSBP1A, HRH4, ZNF521, SS18, PSMA8, TAF4B, KCTD1, AQP4, CHST9, CDH2</i>                                                                                                                                                                                                                                                                                           |
|        | 4              | 27            | 100967134         | 101855077            | <i>FCRL5, ETV3, ARHGEF11, LRRC71, PEAR1, NTRK1, INSRR</i>                                                                                                                                                                                                                                                                                                                                                                                      |
| КБ3    | 4              | 27            | 103490318         | 104129681            | <i>ZBTB7B, FLAD1, CKS1B, DCST2, PBXIP1, PMVK, KCNN3, ADAR, CHRNA2, UBE2Q1, TDRD10, SHE,</i>                                                                                                                                                                                                                                                                                                                                                    |
|        | 4              | 112           | 104518531         | 108045490            | <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, ZNF687, VPS72, PSMD4, PIP5K1A, TMOD4, SCN1, LYSD1, SEMA6C, GABPB2, CDC42SE1, BNIPL, PRUNE, ANXA9, CERS2, SETDB1, CTSS, HORMAD1, GOLPH3L, ECM1, TARS2, ADAMTSL4, RPRD2, PRPF3</i> |
|        | 6              | 26            | 20635220          | 21310322             | <i>Constrained elements for 39 eutherian mammals EPO_LOW_COVERAGE</i>                                                                                                                                                                                                                                                                                                                                                                          |
|        |                |               |                   |                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                |

Продолжение таблицы 9

| Группа | Хромо-сома | Кол-во SNP | Первая позиция | Последняя позиция | Гены                                                                                                                                                                     |
|--------|------------|------------|----------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| КБЗ    | 6          | 53         | 55261304       | 58859012          | <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         | 27         | 32168704       | 33013541          | <i>Constrained elements for 39 eutherian mammals EPO_LOW_COVERAGE</i>                                                                                                    |
|        | 15         | 139        | 105914299      | 113466125         | <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 КБ1 функциональный анализ генов, локализованных в топ HBD, показал, что кодируемые ими белки в большей степени представляют модуляторы активности белков и ингибиторы протеазы. Выполняемые ими биологические процессы и молекулярные функции сходятся на активности/ингибирование эндопептидаз, в том числе серинового типа (таблица 10).

Таблица 10 – Функции генов, локализованные в топ HBD, у свиней КБ1

| Название в соответствии с терминами GO PANTHER db   | Ожидаемое значение | Значение обогащения | Необработанное значение P | Значение P после FDR |
|-----------------------------------------------------|--------------------|---------------------|---------------------------|----------------------|
| Класс белка PANTHER                                 |                    |                     |                           |                      |
| Модулятор белок связывающей активности              | 0,96               | 8,36                | 3,12E-06                  | 3,07E-04             |
| Ингибитор протеазы                                  | 0,19               | 41,51               | 1,63E-11                  | 3,22E-09             |
| PANTHER GO-Slim биологический процесс               |                    |                     |                           |                      |
| Негативная регуляция активности эндопептидазы       | 0,07               | 69,33               | 1,32E-08                  | 2,91E-05             |
| Негативная регуляция активности пептидаз            | 0,07               | 67,13               | 1,54E-08                  | 1,69E-05             |
| Негативная регуляция протеолиза                     | 0,08               | 60,42               | 2,53E-08                  | 1,86E-05             |
| Регуляция протеолиза                                | 0,18               | 27,64               | 1,06E-06                  | 2,58E-04             |
| Регуляция активности пептидаз                       | 0,12               | 41,46               | 1,52E-07                  | 5,57E-05             |
| Регуляция активности эндопептидазы                  | 0,11               | 43,60               | 1,19E-07                  | 5,26E-05             |
| PANTHER GO-Slim молекулярная функция                |                    |                     |                           |                      |
| Активность ингибитора эндопептидазы серинового типа | 0,08               | 65,07               | 1,78E-08                  | 1,06E-05             |
| Активность ингибитора эндопептидазы                 | 0,11               | 45,97               | 9,28E-08                  | 2,77E-05             |
| Активность ингибитора пептидазы                     | 0,11               | 45,97               | 9,28E-08                  | 1,84E-05             |
| Активность регулятора пептидазы                     | 0,12               | 40,28               | 1,74E-07                  | 2,08E-05             |
| Активность регулятора ферментов                     | 0,64               | 7,79                | 4,05E-04                  | 3,02E-02             |
| Активность ингибитора ферментов                     | 0,18               | 28,38               | 9,30E-07                  | 9,24E-05             |
| Активность ингибитора молекулярной функции          | 0,21               | 23,76               | 2,18E-06                  | 1,85E-04             |
| Активность регулятора эндопептидазы                 | 0,12               | 42,72               | 1,32E-07                  | 1,96E-05             |

Сериновые эндопептидазы (протеазы) составляют почти треть всех протеаз и служат неизбежными компонентами, катализирующими гидролитические реакции, участвуют в различных физиологических процессах, таких как переваривание пищи, развитие эмбрионов и иммунная защита (Kanost et al., 2015). Благодаря своей каталитической активности некоторые сериновые протеазы обладают мощными противомикробными свойствами. Несколько исследований продемонстрировали эффективность некоторых протеаз в снижении вирулентности респираторно-синцитиального вируса (Bruno et al., 2022), а также вируса гриппа А (Sakai et al., 2014) у человека. Этот процесс опосредуется протеазами клетки-хозяина *in vivo*. Белки фрагментируются или инактивируются на поверхности, что нарушает проникновение вируса и снижает его патогенность. Кроме того, протеолитическое расщепление вирусных белков клеточными или вирусными протеазами описано на разных стадиях жизненного цикла вируса герпеса (Jovasevic et al., 2008).

Свиньи крупной белой породы адаптируются к различным условиям обитания и проявляют определенную устойчивость к заболеваниям. Однако отечественное поголовье, относительно современного коммерческого, отличалось более высокой устойчивостью к различным заболеваниям. Это связывают с тем, что отбор на высокий уровень производства мяса связан с повышенной восприимчивостью животных к стрессовым факторам окружающей среды и снижением иммунитета к инфекционным патогенам. Вероятно, 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 HBD, наиболее представленных у свиней из группы КБ1.

У свиней КБ2 гены, локализованные в топ HBD, связаны с процессом двухцепочного разрыва ДНК (таблица 11). Двухцепочечные разрывы в ДНК образуются в результате воздействия экзогенных агентов, таких как радиация и некоторые химические вещества, а также в результате эндогенных

процессов, включая репликацию и репарацию ДНК (Cannan et al., 2016). В той или иной мере все организмы страдают от двухцепочечных разрывов в ДНК в результате воздействия ионизирующей радиации. Основным источником естественной радиоактивности земного происхождения являются продукты распада радона – радиоактивного газа, образующегося в почвах и горных породах и накапливающегося в закрытых помещениях (дома, школы, рабочие места). Дополнительным источником естественного излучения могут быть продукты или остатки некоторых промышленных секторов, где используются материалы, содержащие радионуклиды природного происхождения. Кроме того, искусственная радиоактивность возникает в результате антропогенной деятельности, связанной с использованием радиоактивных источников в медико-диагностических, промышленных и научных исследованиях.

В настоящее время в связи с переходом на интенсивные технологии производства в свиноводстве преобладает закрытая система содержания, используются различные диагностические приборы, концентрированные кормовые смеси, что вполне может выступать в качестве повышенной нагрузки ионизирующей радиации. Возможно, обогащение сегментов HBD генами, связанными с процессом двухцепочного разрыва ДНК, можно рассматривать как процесс адаптации к воздействиям современных интенсивных технологий на организм свиней.

Кроме того, у свиней KB2 в сегментах HBD представлены гены, биологические функции которых связаны 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). Ядерно-плазматический транспорт является важным процессом у эукариот и отвечает за процессы обмена белками между ядром и цитоплазмой. Различные белки, такие как факторы транскрипции, гистоны и регуляторы клеточного цикла, должны быть транспортированы в ядро, что делает крайне необходимым наличие сигнала ядерной локализации. Ядерно-

цитоплазматический транспорт функционально и механически разнообразен и служит не только для обеспечения работы механизмов базальной репликации, транскрипции и обработки, но также для регулирования клеточного цикла, активации и репрессии транскрипции, циркадных ритмов и множества других процессов (Lu et al., 2021).

Таблица 11 – Функции генов, локализованные в топ HBD, у свиней КБ2

| Название в соответствии с терминами GO PANTHER db                                      | Ожидаемое значение | Значение обогащения | Необработанное значение P | Значение P после FDR |
|----------------------------------------------------------------------------------------|--------------------|---------------------|---------------------------|----------------------|
| Класс белка PANTHER                                                                    |                    |                     |                           |                      |
| Фактор транскрипции цинковых пальцев C2H2                                              | 1,86               | 7,54                | 5,62E-09                  | 1,11E-06             |
| Фактор транскрипции цинковых пальцев                                                   | 2,20               | 6,36                | 4,51E-08                  | 4,44E-06             |
| ДНК-связывающий фактор транскрипции                                                    | 4,87               | 3,28                | 2,45E-05                  | 1,21E-03             |
| Ген-специфический регулятор транскрипции                                               | 5,20               | 3,66                | 7,47E-07                  | 4,91E-05             |
| PANTHER GO-Slim биологический процесс                                                  |                    |                     |                           |                      |
| Процессинг двухцепочечного разрыва ДНК                                                 | 0,01               | 69,63               | 1,78E-02                  |                      |
| Импорт белка, несущего NLS, в ядро                                                     | 0,04               | 27,85               | 3,87E-02                  |                      |
| Путь ERAD                                                                              | 0,22               | 9,13                | 2,14E-02                  |                      |
| PANTHER GO-Slim молекулярная функция                                                   |                    |                     |                           |                      |
| Связывание стерола                                                                     | 0,11               | 26,95               | 2,44E-04                  | 7,26E-02             |
| Связывание стероидов                                                                   | 0,13               | 23,21               | 3,68E-04                  | 3,13E-02             |
| Активность ДНК-связывающего репрессора транскрипции, специфичная для РНК-полимеразы II | 0,31               | 12,95               | 3,08E-04                  | 4,59E-02             |
| Активность ДНК-связывающего фактора транскрипции, специфичная для РНК-полимеразы II    | 3,74               | 3,20                | 3,59E-04                  | 3,57E-02             |

Эндоплазматический ретикулум является основным местом биогенеза мембранных и секретируемых белков, биосинтеза липидов и хранения внутриклеточного кальция. Кроме того, обширная структура эндоплазматического ретикулума влияет на общую архитектуру и организацию эукариотических клеток. Система расщепления белков с



помощью эндоплазматического ретикулума осуществляет посттрансляционный контроль над липидным гомеостазом, организацией ядерной оболочки, контактами между органеллами (например, митохондриями, эндосомами), потоком кальция и даже передачей врожденных иммунных сигналов. Зависимый от аутофагии и независимый от аутофагии клиренс через эндолизосомы или прямая экстракция неправильно локализованных мембранных белков являются альтернативными стратегиями регуляции эндоплазматического ретикулума.

Можно предположить, что концентрация в сегментах HBD генов, связанных с функционированием ядерно-цитоплазматического транспорта в ядре и пути ERAD, у представителей современного поголовья свиней крупной белой породы 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. Раннее предполагали, что гормон роста и рецептор гормона роста (комплекс GH/GHR) осуществляют свои функции только на плазматической мембране. Однако недавние открытия показали, что комплекс GH/GHR может перемещаться в ядра клеток, возможно, по пути ERAD, где он выполняют важные физиологические функции (Hainan et al., 2018).

Интересно отметить, что если роль ядерно-цитоплазматического транспорта в ядре и пути ERAD могли оказать положительное влияние на признаки продуктивности свиней, то с другой стороны предполагают, что эти процессы также могут быть связаны с репликацией цирковируса свиней типа 2 (Shuai et al., 2011), а также с репродуктивно-респираторным синдромом свиней (Lee et al., 2005). 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 2-го типа (1994 г.), эпидемическая диарея свиней (2013 г.), африканская чума свиней в Европе и Азии (2007 г.).

Среди локусов, расположенных в топ HBD, у современных свиней крупной белой породы, как в группе КБ2, так и в группе КБ3, определены гены, относящиеся к классу транскрипционных факторов, содержащих кластер цинковых пальцев C2H2-типа (таблица 12). Анализ генов полицинковых пальцев показал, что они возникли из небольшой предковой группы эукариотических факторов транскрипции и в процессе эволюции сильно подвергались процессам дупликации (Emerson et al., 2009). В дополнение к точечным мутациям, посредством которых реализуется адаптивная эволюция, дупликация или делеция целых доменов может привести к быстрой функциональной дивергенции среди недавно дублированных паралогов. В тех случаях, когда кластеры цинковых пальцев C2H2-типа связываются с целевыми сайтами ДНК, дупликация или делеция домена может радикально изменить мишень связывания белка. Дупликация этих генов произошла на ранних стадиях эволюции, что может свидетельствовать об их роли в фундаментальных и/или консервативных функциях организмов, например, таких как эмбриональное развитие и общая регуляция транскрипции в организме (Maskeh et al., 2018). Добавление доменов типа C2H2, возможно, было необходимо для функций, специфичных для высших организмов, таких как комплексная регуляция транскрипции генов, обеспечивающая развитие, дифференцировку и/или функционирование сложных органов и тканей.

В целом скоординированная эволюция селекционно-значимых признаков современных свиней крупной белой породы произошла путем изменения регуляторных генов, контролирующих наборы эффекторных генов, в чем, вероятно, сыграли роль и транскрипционные факторы, содержащие кластер цинковых пальцев C2H2-типа.

Таблица 12 – Функции генов, локализованные в топ HBD, у свиней КБЗ

| Название в соответствии с терминами GO PANTHER db                                      | Ожидаемое значение | Значение обогащения | Необработанное значение P |
|----------------------------------------------------------------------------------------|--------------------|---------------------|---------------------------|
| Класс белка PANTHER                                                                    |                    |                     |                           |
| Мембраносвязанная сигнальная молекула                                                  | 0,18               | 10,88               | 1,60E-02                  |
| Кальцийсвязывающий белок                                                               | 0,59               | 8,50                | 3,77E-04                  |
| Фосфолипаза                                                                            | 0,32               | 6,24                | 4,30E-02                  |
| Киназа                                                                                 | 0,78               | 5,14                | 8,38E-03                  |
| Фактор транскрипции цинковых пальцев C2H2                                              | 2,72               | 3,31                | 1,79E-03                  |
| PANTHER GO-Slim биологический процесс                                                  |                    |                     |                           |
| Процесс биосинтеза холестерина                                                         | 0,02               | 47,58               | 2,59E-02                  |
| Сигнальный путь рецептора инсулиноподобного фактора роста                              | 0,03               | 38,06               | 3,10E-02                  |
| Высвобождение секвестрированных ионов кальция в цитозоль                               | 0,14               | 14,64               | 9,37E-03                  |
| Метаболический процесс ацил-КоА                                                        | 0,19               | 10,57               | 1,68E-02                  |
| Процесс биосинтеза фосфатидилинозитолфосфата                                           | 0,20               | 9,76                | 1,95E-02                  |
| Позитивная регуляция активности ДНК-связывающего фактора транскрипции                  | 0,21               | 9,52                | 2,04E-02                  |
| PANTHER GO-Slim молекулярная функция                                                   |                    |                     |                           |
| Экзополифосфатазная активность                                                         | 0,03               | 31,72               | 3,60E-02                  |
| Активируемая кальцием активность калиевых каналов                                      | 0,07               | 29,28               | 2,72E-03                  |
| тРНК-специфическая активность аденозиндезаминазы                                       | 0,04               | 23,79               | 4,61E-02                  |
| Активность фосфолипазы C                                                               | 0,35               | 5,77                | 3,92E-03                  |
| Активность трансмембранного рецептора протеинтирозинкиназы                             | 0,29               | 6,92                | 3,59E-02                  |
| Активность ДНК-связывающего репрессора транскрипции, специфичная для РНК-полимеразы II | 0,45               | 6,64                | 1,14E-02                  |

У свиней группы КБЗ, в отличие от свиней КБ2, в сегментах HBD присутствуют гены, кодирующие белки, принадлежащие к классам

фосфолипаз, киназ, мембраносвязанных сигнальных молекул и принимающие участие в процессах биосинтеза холестерина, сигнального пути рецептора инсулиноподобного фактора роста, метаболический процесс ацил-КоА и др., так или иначе участвующих в процессах, связанных с транспортом, активацией и синтезом жирных кислот. Так, ацил-КоА являются значимыми регуляторными молекулами и промежуточными продуктами метаболизма. Начальным этапом их синтеза является активация жирных кислот одной из 13 изоформ длинноцепочечной ацил-КоА-синтетазы. Продукты ацил-КоА регулируют метаболические ферменты и сигнальные пути, участвуя в процессах синтеза сложных липидов, таких как триацилглицерин, фосфолипиды и эфиры холестерина (Grevengoe et al., 2014).

Жировая ткань в организме может быть локализована в виде висцеральных, подкожных, межмышечных или внутримышечных депо с уменьшением интенсивности нарастания в процессе развития. Каждое жировое депо проявляет специфические метаболические свойства, которые влияют на метаболизм всего организма путем секреции гормонов, адипоцитокинов и регуляторных белков (Poklucar et al., 2020). Гормоны, полученные из жировой ткани, влияют на потребление пищи, воспалительные реакции или чувствительность к инсулину. У свиней жировая ткань является основным местом липогенеза (превращение глюкозы в триглицериды) и обеспечивает синтез более 80% жирных кислот. Биосинтез жирных кислот зависит от активности ключевых ферментов, физиологического состояния животного и генетики. Несмотря на то, что в целом считается, что программы селекции привели к снижению упитанности и содержанию внутримышечного жира у современного поголовья свиней, в группе КБЗ можно предположить, что одним из акцентов селекционной работы являлось повышение качества мяса. Здесь также следует отметить, что свиньи, представляющие группу КБЗ, действительно отличаются высоким качеством мяса и уровнем внутримышечного жира.

Подводя итоги анализа HBD у свиней крупной белой породы из различных селекционных групп, можно выделить следующие особенности. По результатам наших исследований самые короткий HBD были унаследованы в группе КБ1 от предков около 128 лет (класс  $R_k = 256$ ), что в целом как раз является периодом формирования крупной белой породы. В дальнейшем крупная белая порода участвовала в создании и совершенствовании большинства современных европейских пород, а также локальных пород, созданных на территории СССР. Наши исследования показали, что общая аутозиготность у современных свиней КБ2 и КБ3 составила около 0,25–0,29 доли генома, при этом около 0,1 доли генома покрыта сегментами аутозиготности, относящимися к классу  $R_k = 128$ . На основании этого можно предположить, что большой вклад в аутозиготность этих групп внесли предки около 64 лет назад. Это период накладывается на период роста интенсификационных процессов в свиноводстве и формирования коммерческого поголовья, отличающегося высокими производственными показателями.

Общие сегменты аутозиготности в популяции идентифицируют точки отбора. Исследования, направленные на изучение сегментов HBD свиней крупной белой породы, разводимых до внедрения в производство процесса интенсификации и после, позволили выявить различия между популяциями. Кроме того, даже у свиней, относящихся к современному поголовью, но содержащихся в различных генетических центрах, определены отличительные особенности, связанные с различными целями и направлением отбора, что делает такие участки наиболее перспективные для поиска потенциальных генов кандидатов, связанных с уровнем продуктивности и устойчивостью к заболеваниям.

### 3.3 Идентификация сигналов положительного отбора в геномах свиней крупной белой породы из различных селекционных центров

Для достижения цели по обнаружению признаков отбора в популяции свиней крупной белой породы были использованы различные подходы к набору данных, включающему ~60 тысяч SNP-маркеров (после этапов фильтрации для анализа оставили 32 тысяч SNP). На первом этапе сигналы отбора были установлены в каждой исследуемой популяции: свиней отечественной селекции (группа КБ1), импортной селекции двух центров (группа КБ2 и группа КБ3), используя три метода расчетов на основе паттернов гомозиготности гаплотипов iHS, nSI и iH12. В качестве значимых сигналов рассматривали значения, показавшие критерий равный 1 ( $\text{crit} = 1$ ).

По результатам трех методов у свиней КБ1 определено 2739 варианта (iHS – 2021, nSI – 1068 и iH12 – 431) (рисунок 47).

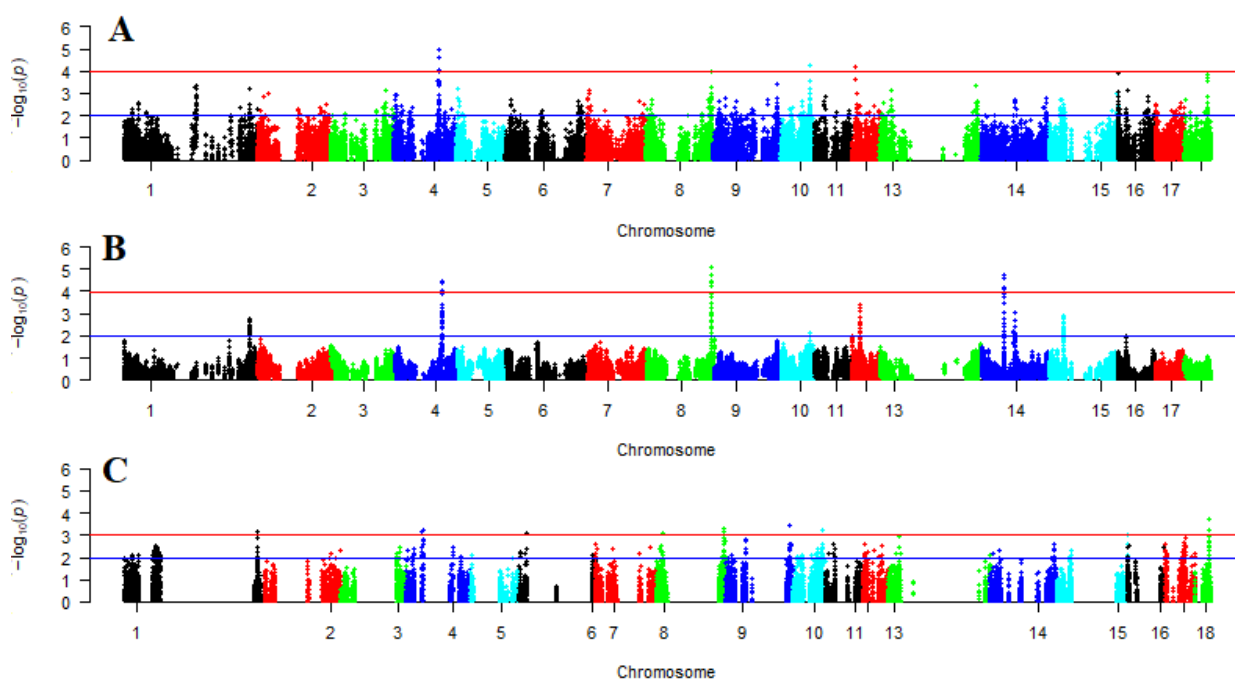


Рисунок 47 – Полногеномное распределение сигналов отбора у свиней КБ1:  
А – iHS; В – iH12; С – nSI

В общей сложности для КБ1 было обнаружено 808 генов, находящихся под давлением отбора. По данным базы PigQTLdb, гены под действием отборов в группе КБ1 в большей степени пересекаются с локусами,

ответственными за мясо и тушу свиней. Круговая диаграмма (А) (рисунок 48) иллюстрирует распределение выявленных генетических маркеров по основным группам хозяйственно-полезных признаков. Подавляющее большинство маркеров (58,28%) относится к категории «Мясо и туша», подчеркивая генетическую важность или фокус исследования на качестве мясной продукции и параметрах туши. Значительные доли маркеров также связаны с признаками «Здоровья» (15,93%) и общей «Продуктивности» (11,07%). Меньший вклад вносят маркеры, ассоциированные с «Репродуктивностью» (5,77%), «Экстерьером» (4,96%) и экспрессией генов, влияющих на мясные качества («Мясо и туша eQTL», 3,99%).

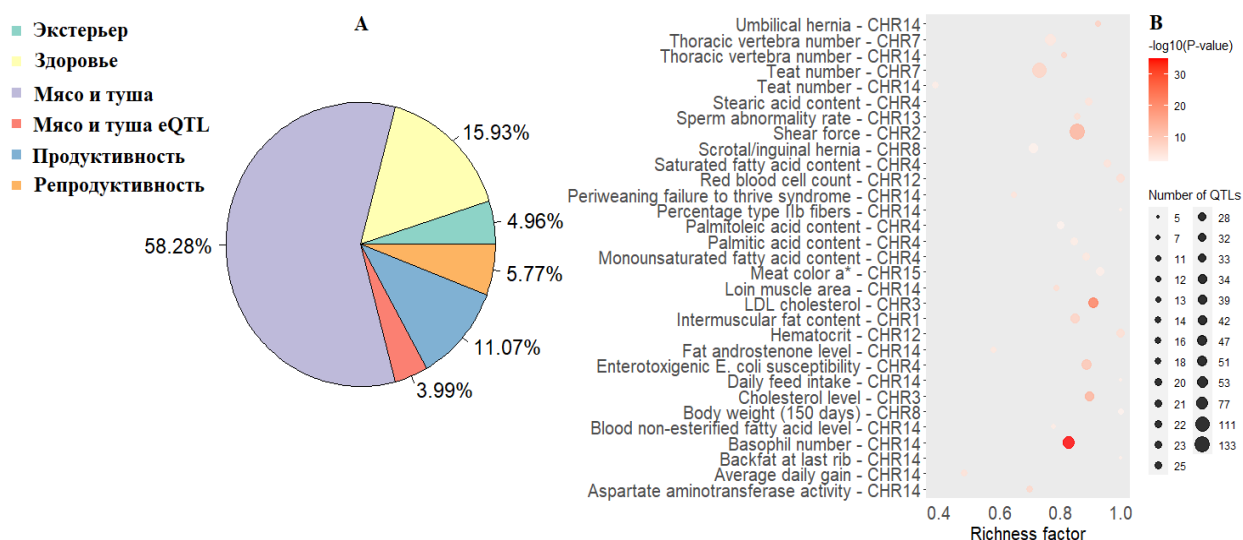


Рисунок 48 – Обогащение вариантов отбора локусами количественных признаков у свиней КБ1

График (В) детализирует информацию по отдельным признакам, выявляя наиболее значимые ассоциации. Особо выделяется признак «Количество базофилов» (Basophil number), локализованный на 14-й хромосоме, который демонстрирует исключительно высокую статистическую значимость и связан с большим числом генетических маркеров. Также сильные генетические сигналы обнаружены для таких показателей, как «Усилие среза» мяса (Shear force, хромосома 2), уровень «Холестерина

ЛПНП» (LDL cholesterol, хромосома 3) и «Вес тела в 150 дней» (Body weight (150 days), хромосома 8). Кроме того, наблюдается концентрация важных маркеров на 14-й хромосоме, связанных не только с базофилами, но и с другими признаками, включая «Толщину шпика на последнем ребре» (Backfat at last rib) и «Среднесуточный привес» (Average daily gain), указывая на плейотропное действие или сцепление генов в этом регионе.

У свиней КБ2 выявлено 966 вариантов сигналов отбора (iHS – 863, nSI – 316 и iH12 – 26) (рисунок 49).

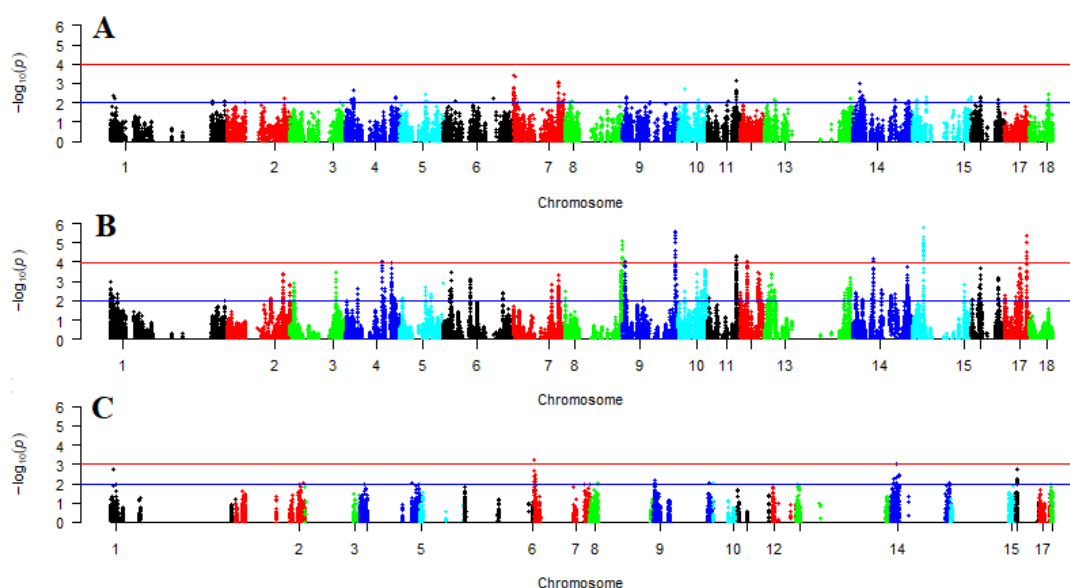


Рисунок 49 – Полногеномное распределение сигналов отбора у свиней КБ2:  
А – iHS; В – iH12; С – nSI

В группе КБ2 гены под действием отборов в большей степени также пересекаются с локусами, ответственными 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 (рисунок 50). Как и в группе КБ1, доминирующей категорией является «Мясо и туша», на которую приходится 61.4% всех выявленных ассоциаций. Категория «Здоровье» занимает второе место с 14.78%, за ней следует «Продуктивность» с 9.82%. Признаки «Экстерьера» (6.22%) и «Репродуктивности» (6.06%) представлены примерно в равной, но меньшей степени. Наименьшая доля



маркеров (1.71%) приходится на категорию «Мясо и туша eQTL», связанную с экспрессией генов в контексте мясных качеств.

График обогащения (В) предоставляет детализацию по конкретным признакам. Высокую статистическую значимость демонстрируют маркеры, связанные с «Активностью гемолитического компонента (классический путь)» на хромосоме 4 и «Уровнем жирового андростенона» на хромосоме 14. Также значимые сигналы наблюдаются для «Уровня холестерина ЛПНП» (LDL cholesterol, хромосома 3) и «Средней толщины шпика» (Average backfat thickness, хромосома 7). Признак «Потери влаги при хранении» (Drip loss, хромосома 1) выделяется большим количеством ассоциированных с ним маркеров (QTLs), хотя и с умеренной статистической значимостью. Эти результаты указывают на важные геномные регионы, влияющие на иммунный ответ, качество мяса, метаболизм жиров и влагоудерживающую способность мяса.

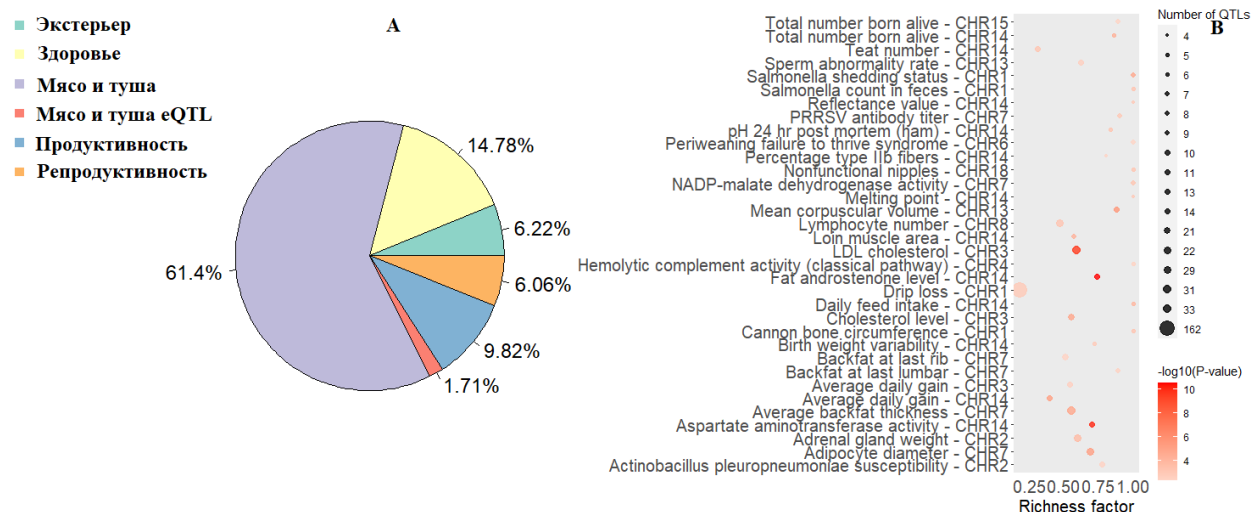


Рисунок 50 – Обогащение вариантов отбора локусами количественных признаков у свиней KB2

У свиней KB3 установлено 2778 вариантов сигналов отбора (iHS – 1124, pSI – 563 и iH12 – 1623) (рисунок 51).

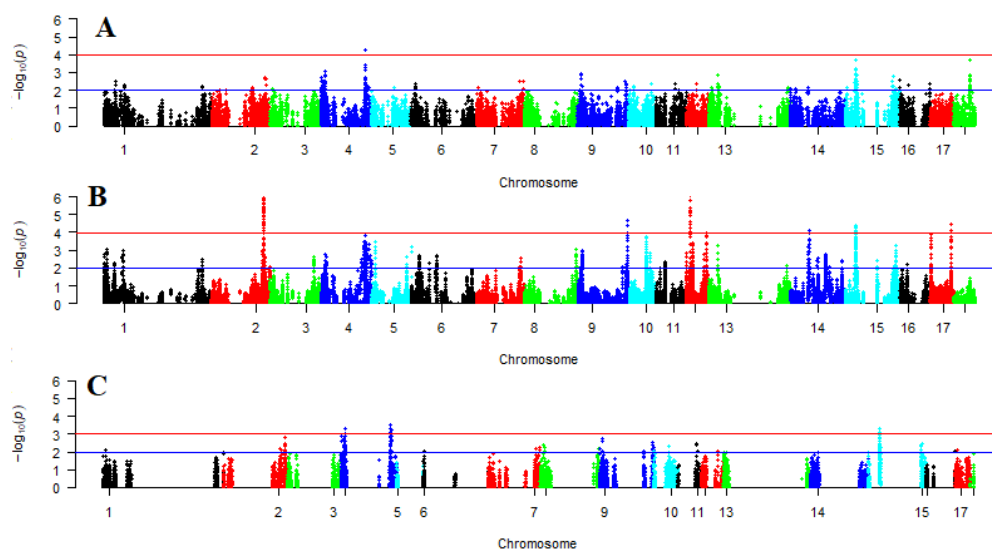


Рисунок 51 – Полногеномное распределение сигналов отбора у свиней KB3:  
А – iHS; В – iH12; С – nSI

В группе KB3 в целом наблюдалась картина, аналогичная группе KB2. Гены под действием отбора пересекаются с локусами, ответственными за мясо и тушу свиней, здоровье и продуктивность (рисунок 52). Категория «Мясо и туша» вновь является преобладающей, охватывая 59.89% выявленных маркеров. Существенная доля маркеров (16.62%) относится к признакам «Здоровья», а категория «Продуктивность» составляет 9.83%. Меньшие доли приходятся на «Экстерьер» (5.71%) и «Репродуктивность» (5.47%). Категория «Мясо и туша eQTL» представлена наименьшей долей (2.48%). Это распределение указывает на сильный акцент исследования или генетический контроль признаков, связанных с мясными качествами и здоровьем.

График обогащения (В) выделяет несколько конкретных признаков с заметными генетическими ассоциациями. Наиболее статистически значимые сигналы (насыщенный красный цвет) обнаружены для «Количества красных кровяных телец» (Red blood cell count, хромосома 12) и «Гематокрита» (Hematocrit, хромосома 12), причем оба признака также связаны с большим количеством QTL (крупные точки). Кроме того, значимые ассоциации выявлены для «Восприимчивости к энтеротоксигенной *E. coli*» (Enterotoxigenic *E. coli* susceptibility, хромосома 4) и «Средней толщины

шпики» (Average backfat thickness, хромосома 2). Эти результаты подчеркивают важность геномных регионов на хромосомах 12, 4 и 2 для контроля показателей крови, устойчивости к заболеваниям и жиротложения.

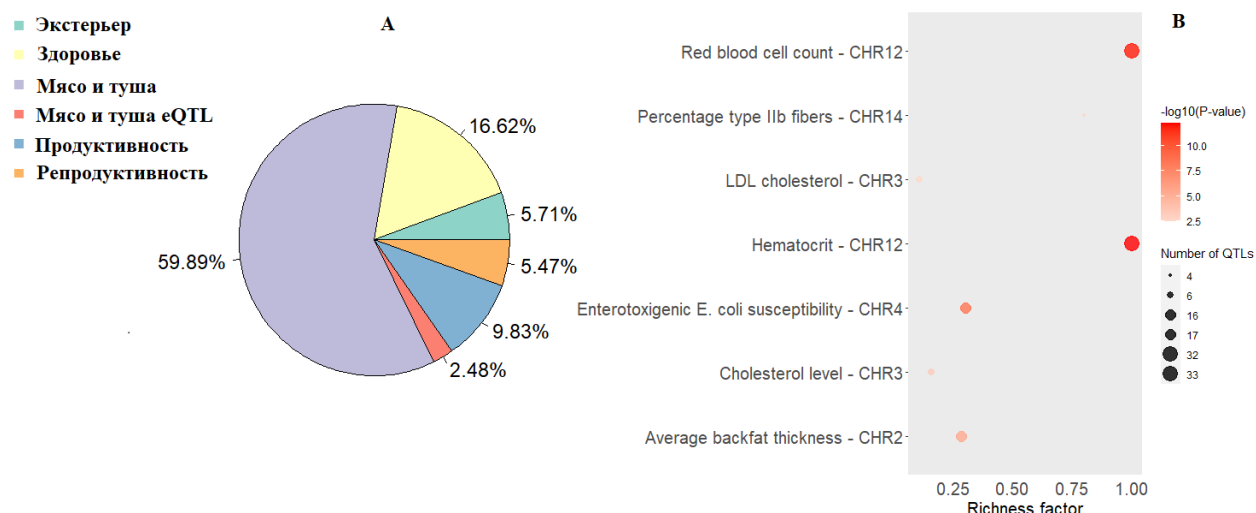


Рисунок 52 – Обогащение вариантов отбора локусами количественных признаков у свиней КБЗ

Общие варианты отбора для всех анализируемых групп свиней крупной белой породы определены в SSC4, 7, 8 и 14. Кроме того, 6 общих вариантов локализованы в SSC9, 2 из которых – в интронах гена *TENM4* (таблица 13).

Таблица 13 – Общие сигналы отбора у свиней трех групп

| SSC | Позиция   | Номер rs     | Вариант в геноме                   | Ген           |
|-----|-----------|--------------|------------------------------------|---------------|
| 14  | 25989860  | rs80973471   | intergenic variant                 |               |
| 4   | 16486634  | rs80809199   | intron variant                     | <i>ZHX2</i>   |
| 7   | 106025579 | rs81295294   | intergenic variant                 |               |
| 8   | 18271949  | rs81406415   | non coding transcript exon variant | <i>lncRNA</i> |
| 9   | 13492371  | rs81257618   | intron variant                     | <i>TENM4</i>  |
| 9   | 13580920  | rs330708382  | intron variant                     | <i>TENM4</i>  |
| 9   | 13846118  | rs81477175   | non coding transcript exon variant | <i>lncRNA</i> |
| 9   | 13871291  | rs81420099   | intergenic variant                 |               |
| 9   | 13961806  | rs81420072   | upstream gene variant              | <i>lncRNA</i> |
| 9   | 14014071  | rs3472119487 | downstream gene variant            | <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. Геномы позвоночных включают четыре гена *TENM* (*TENM1–TENM4*), при этом у членистоногих имеются только два гена тенеурина 1, а у других беспозвоночных – только один ген (Wides et al., 2019).

У людей тенерины связаны с различными генетическими нарушениями. Относительно *TENM4*, его роль установлена для гаструляции и индукции мезодермы во время эмбрионального развития и для развития олигодендроцитов типа I/II в головном мозге (Zhang et al., 2022).

У свиней *TENM4* является критически важным онтогенетическим геном, т.к. он участвует в развитии сомитов и, следовательно, скелета. В работе Rohrer с соавторами (Rohrer et al., 2015) выдвинуто предположение о наличии ассоциативной связи между генетическими вариантами гена *TENM4* и количеством груднопоясничных позвонков у свиней. Роль вариантов гена *TENM4*, приводящих к слиянию грудных и/или поясничных позвонков у мышей представлена в работе Lossie и соавторов (Lossie et al., 2005). 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.

Развитие позвоночного столба является основополагающим элементом для выживания млекопитающих, в связи с чем он в высшей степени консервативен и строго регулируется. Количество позвонков у большинства видов млекопитающих довольно консервативно. Немногие виды демонстрируют фенотипическую изменчивость этих признаков и, вероятно, не обладают генетической изменчивостью. Так, количество шейных позвонков у большинства видов млекопитающих составляет семь, но количество хвостовых позвонков у млекопитающих колеблется от 2 до 37 (Owen, 1853). Существуют различия в количестве грудных и поясничных позвонков у разных видов, но сумма грудных и поясничных позвонков (грудопоясничных) в большей степени консервативна, т.к. у большинства видов составляет 19.

Одним из видов, у которых существует вариация количества позвонков, является домашняя свинья. При этом у диких кабанов и свиней породы мейшан (место происхождения Китай) количество грудопоясничных позвонков равно 19, а у коммерческих свиней европейского происхождения варьирует от 21 до 23 (King, Roberts, 1960).

Сигналы отбора, локализованные в гене *TENM4*, вполне вероятно связаны и с его ролью в формировании и других значимых признаков посредством сложных генных сетей и метаболических процессов. На уровне фенотипа это подтверждается корреляционными связями между количеством грудопоясничных позвонков и селекционными признаками. Так, отмечают связь между количеством позвонков у свиней и длиной туши, постным мясом, скоростью роста (Borchers et al., 2004). Однако с большим количеством позвонков также связывают повышение мертворожденных поросят (Rees, 1954), дефекты искривления спины (Holl et al., 2008), хромоту и проблемы с движением (Duckworth et al., 1968). В целом это является общим правилом, при улучшении признаков продуктивности в популяции повышаются дефекты и различные аномалии. В связи с этим ген *TENM4* может выступать в качестве перспективного гена-кандидата как для значимых признаков продуктивности, так и анализа отрицательных корреляций между продуктивностью и здоровьем.

Один из сигналов отбора локализован в гене цинковых пальцев и гомеобоксов 2 *ZHX2*, который принадлежит к семейству генов *ZHX* (*ZHX1–ZHX3*), обнаруженных только у позвоночных. Все три гена *ZHX* имеют одинаковую уникальную структуру кодирующей области в третьем экзоне.

Несколько исследований на людях выявили причастность *ZHX2* к сердечно-сосудистым заболеваниям, на мышах получили результаты, свидетельствующие о влиянии *ZHX2* на уровень липидов в сыворотке (Nail et al., 2020). *ZHX2* играет значимую роль в регуляции  $\alpha$ -фетопротейна (АФП) в печени.

У свиней концентрация альфа-фетопротеина (наряду с альфа-1-антитрипсином и трансферрином) в сыворотке увеличивается с возрастом до 40–50 дней беременности, а затем прогрессивно снижается при рождении (Ahn et al., 2014). Основными местами синтез альфа-фетопротеина (АФП) в онтогенезе являются печень и желточный мешок плодов.

Можно предположить, что сигналы отбора в гене *ZHX2* у свиней связаны с регуляцией синтез альфа-фетопротеина, т.к. главная особенность альфа-фетопротеина – это высокое сродство к полиненасыщенным жирным кислотам. На уровне фенотипа это может проявляться в развитии мышечной и внутримышечной жировой ткани. Так в работе Óvilo (Óvilo et al., 2014) на основе полногеномных ассоциативных исследований отмечена связь гена *ZHX2* с различиями в профиле мышечного транскриптома у чистокровных иберийских свиней и гибридного поголовья, полученного при скрещивании с хряками породы дюрок.

По результатам поиска определено 53 варианта, показавших сигналы отбора только у свиней импортной селекции, из них 21 вариант локализован в межгенных областях, 4 – в длинных некодирующих РНК (LncRNA) и 31 вариант в 26 генах (таблица 14).

Таблица 14 – Сигналы отбора, идентифицированные только у свиней импортной селекции

| SSC | Позиция   | Номер rs    | Вариант в геноме                   | Ген            |
|-----|-----------|-------------|------------------------------------|----------------|
| 1   | 248654503 | rs80942179  | intergenic variant                 | -              |
| 1   | 248711404 | rs336098849 | intergenic variant                 | -              |
| 1   | 248794352 | rs81314465  | intergenic variant                 | -              |
| 1   | 6732845   | rs80856514  | non coding transcript exon variant | <i>lncRNA</i>  |
| 3   | 110968085 | rs81375763  | intron variant                     | <i>BABAM2</i>  |
| 3   | 11298970  | rs339350272 | intergenic variant                 | -              |
| 3   | 122158369 | rs81377655  | intron variant                     | <i>NBAS</i>    |
| 4   | 114016613 | rs80974420  | intergenic variant                 | -              |
| 4   | 4505948   | rs344384537 | intron variant                     | <i>FAM135B</i> |
| 4   | 5517026   | rs325576014 | intergenic variant                 | -              |
| 5   | 4713576   | rs81382770  | intergenic variant                 | -              |
| 5   | 5383332   | rs80937862  | intron variant                     | <i>lncRNA</i>  |
| 5   | 89362545  | rs81386282  | intron variant                     | <i>CRADD</i>   |

Продолжение таблицы 14

| SSC | Позиция   | Номер rs               | Вариант в геноме        | Ген             |
|-----|-----------|------------------------|-------------------------|-----------------|
| 6   | 20650131  | rs81392250             | intergenic variant      | -               |
| 7   | 106923304 | rs80947173             | intergenic variant      | -               |
| 7   | 107951062 | rs338619524            | intergenic variant      | -               |
| 7   | 113568204 | rs80985247             | synonymous variant      | <i>TRIP11</i>   |
| 7   | 116379097 | rs81396316             | intron variant          | <i>DICER1</i>   |
| 8   | 12888246  | rs325979027            | intron variant          | <i>LCORL</i>    |
| 8   | 13713034  | rs330623309            | intergenic variant      | -               |
| 9   | 124790147 | rs81417044             | intron variant          | <i>NCF2</i>     |
| 9   | 124913515 | rs81234791             | intron variant          | <i>APOBEC4</i>  |
| 9   | 125628612 | rs81417300             | intron variant          | <i>LncRNA</i>   |
| 9   | 126968369 | rs81262046             | intron variant          | <i>HMCN1</i>    |
| 9   | 126973412 | rs705432331            | intron variant          | <i>HMCN1</i>    |
| 9   | 13775191  | rs787818425            | intron variant          | <i>TENM4</i>    |
| 9   | 13857010  | rs81420094             | intergenic variant      | -               |
| 9   | 13925619  | rs321131687            | intergenic variant      | -               |
| 10  | 46476233  | rs81424925             | intron variant          | <i>FAM171A1</i> |
| 10  | 46643939  | rs324335825            | upstream gene variant   | <i>NMT2</i>     |
| 11  | 61442298  | rs344297345            | intron variant          | <i>GPC5</i>     |
| 11  | 63668462  | rs1108031663           | intron variant          | <i>DCT</i>      |
| 12  | 7668938   | rs81437773             | upstream gene variant   | <i>LncRNA</i>   |
| 13  | 201799863 | rs81442794             | intergenic variant      | -               |
| 13  | 25861930  | rs81320832             | intron variant          | <i>TRAK1</i>    |
| 13  | 27151296  | rs81444244             | intergenic variant      | -               |
| 13  | 47743525  | rs81445390             | upstream gene variant   | <i>SLC25A26</i> |
| 14  | 129544197 | rs340340080            | intergenic variant      | -               |
| 14  | 129554596 | rs80793133             | intergenic variant      | -               |
| 14  | 129565608 | rs333023007            | intergenic variant      | -               |
| 14  | 129586806 | rs319669226            | intron variant          | <i>INPP5F</i>   |
| 14  | 129611113 | rs80911564             | intron variant          | <i>INPP5F</i>   |
| 14  | 130408520 | rs80786021             | intron variant          | <i>PLPP4</i>    |
| 14  | 25374812  | rs339777110            | intron variant          | <i>TMEM132D</i> |
| 14  | 25408487  | rs341078558            | intron variant          | <i>TMEM132D</i> |
| 14  | 25446354  | rs80924625             | intron variant          | <i>TMEM132D</i> |
| 14  | 25517226  | rs81300965             | intron variant          | <i>TMEM132D</i> |
| 14  | 44463226  | rs80836442             | intergenic variant      | -               |
| 14  | 44986228  |                        |                         | <i>MN1</i>      |
| 14  | 8411858   | rs80866597             | intron variant          | <i>ADAM28</i>   |
| 15  | 122185946 | rs332967326            | intergenic variant      | -               |
| 16  | 7850860   | rs81345229             | intergenic variant      | -               |
| 17  | 54601392  | rs789089933/rs80877089 | downstream gene variant | <i>ZNF217</i>   |

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 (Vázquez-Gómez et al., 2020) был определен как вариант материнского феномена искажения передачи информации (Transmission ratio distortion, TRD) для иберийских свиней. Феномен искажения передачи информации возникает, когда один из аллелей в гетерозиготном локусе не передаются в равной степени потомству (Villena et al., 2000). Этот феномен встречается у людей (Meyer et al., 2012), других млекопитающих (Wu et al., 2005), насекомых (Solignac et al., 2004), растений (Koide et al., 2008). Различные биологические механизмы могут вызывать искажение передачи информации до или после оплодотворения (Fishman et al., 2019). 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 третьей хромосоме из трех вариантов отбора два локализованы в генах *BABAM2* и *NBAS*. Ген *BABAM2* кодирует высококонсервативный белок у млекопитающих. Он не имеет идентифицируемых функциональных доменов, но обладает двумя убиквитин-конъюгирующими ферментами, подобными семейству (UEV-подобных) областей (Li et al., 1995). Экспрессия гена *BABAM2* зафиксирована на двухклеточной стадии у преимплантационных эмбрионов, показывая его роль в раннем эмбриональном развитии (Chen et al., 2013). Текущее понимание роли *BABAM2* ограничено и в целом сводится к тому, что он регулирует клеточный цикл и предотвращает клеточное старение (Chung et al., 2020).

Продукты гена *NBAS* играют существенную роль в контроле экспрессии генов, участвуя в нонсенс-опосредованном распаде. В основе нонсенс-опосредованного распада происходит расщепление мРНК, содержащих стоп-кодона в открытых рамках считывания, что предохраняет клетку от дефектных белков. Предполагают, что нонсенс-опосредованный распад



сыграл особую роль в эволюционном процессе у эукариот, способствуя появлению новых генов (Aoki et al., 2009). Уникальные функции *NBAS* могут являться причиной широкой фенотипической изменчивости. У людей генетические варианты гена *NBAS* связаны с широким спектром заболеваний, таких как поражения печени, частые инфекции, скелетные и иммунологические аномалии (Thong et al., 2021).

У свиней вариант rs81377655 в гене *NBAS* определен как кандидат для среднесуточного прироста и массы при рождении в полногеномных исследованиях, проведенных Qiao с соавторами (Qiao et al., 2015). Кроме того, роль *NBAS* отмечена в работе, направленной на поиск вариантов, связанных с убийством свиноматкой своего новорожденного потомства в течение 24 часов после рождения, загрызает их до смерти (каннибализм свиноматок) (Quilter et al., 2012). Это расстройство поведения встречается у свиноматок при первом опоросе, связывают его с условиями содержания и кормления, но также оно имеет сильный генетический компонент, наследуемость которого составляет 0,4–0,9 (Chen et al., 2007; Кнар, Merks, 1987). Следует отметить, что при создании коммерческого поголовья (и по настоящее время) любое агрессивное поведение свиней рассматривается как сильнейший дефект, и такие особи сразу подлежат отбраковке (в том числе все их ближайшие родственники). Вполне вероятно, что целенаправленный отбор в этом направлении отразился в качестве сигнала отбора в гене *NBAS*. Наряду с геном *NBAS* в работе Quilter с соавторами (Quilter et al., 2015) с канибализмом свиноматок детектирован ген *FAM135B*, в котором также определен сигнал отбора, идентифицированный только у свиней импортной селекции. В других работах *FAM135B* представлен как ген-кандидат для признаков роста у мясного скота (Seabury et al., 2017) и воспроизводительных качеств у свиней породы дюрок (Zhang et al., 2019).

В пятой хромосоме сигналы отбора были определены в межгенной области, *LncRNA* (длинная некодирующая РНК) и гене *CRADD*. Вариант rs81382770, локализованный в межгенной области, ранее в работе Rongrong

с соавторами (Rongrong et al., 2019) был представлен как генетический предиктор для внутримышечного жира у свиней породы дюрок. Ген *CRADD* в исследованиях Ramos с соавторами (Ramos et al., 2009) отмечен как ген-кандидат для роста и отложения жира у гибридов схемы Беркшир × Йоркшир. Связь гена *CRADD* с развитием мышц и качеством мяса также продемонстрирована в работе Wang с соавторами (Wang et al., 2022). Кроме того, *CRADD* рассматривается как ген-кандидат, ассоциированный с признаками мясной продуктивности у мясомолочных коз (Селионова и др., 2022). В целом при создании коммерческих пород свиней приоритетным направлением селекционной работы являлись признаки, связанные с ростом и отложением жира, с чем, вероятно, и связаны сигналы отбора, идентифицированные у свиней крупной белой породы в пятой хромосоме.

В седьмой хромосоме сигналы отбора обнаружены в генах *TRIP11* и *DICER1*. Ген *TRIP11* кодирует белок, который взаимодействует с рецептором гормона щитовидной железы  $\beta$  и участвует в регуляции метаболизма липидов (Zhu et al., 2010). Мыши, лишённые всех изоформ рецепторов тиреоидных гормонов, демонстрируют пониженную температуру тела, задержку роста и повышенное количество жировой ткани (Golozoubova et al., 2004). У свиней ген *TRIP11* показал связь с внутримышечным жиром у гибридов мейшан × дюрок (Sato et al., 2005) и толщиной шпика у свиней породы дюрок (Xue et al., 2022).

Белковый продукт гена *DICER1* играет роль в контроле трансляции белка и крайне важен для продукции зрелых микроРНК (Murchison, Hannon, 2004). На сегодняшний день микроРНК считается системой посттранскрипционного контроля, имеющей решающее значение почти для всех биологических процессов, включая развитие и функционирование тканей, а также репродуктивных функций (Carletti, Christenson, 2009). В работах с мышами было продемонстрировано, что Dicer1-зависимый путь miRNA играет важную роль в развитии фолликулов, оогенезе и сперматогенезе (Lei et al., 2009). У свиней также продемонстрирована роль

*DICER1* в регуляции репродуктивной системы в исследовании Knapczyk-Stwora с соавторами (Knapczyk-Stwora et al., 2015), которые впервые представили данные об экспрессии и клеточной локализации *DICER1* в семенниках и яичниках эмбрионов свиней.

В восьмой хромосоме сигналы отбора определены в области гена *LCORL*. По литературным данным, *LCORL* ассоциируется с ростом у людей (Weedon et al., 2008), массой тела овец (Al-Mamun et al., 2015), с размером тела у собак, крупного рогатого скота, лошадей и свиней (Rubin et al., 2012). Результаты Rubin (Rubin et al., 2012) продемонстрировали сильные сигналы отбора по локусу гена *LCORL* в коммерческих линиях европейских домашних свиней, подвергнутых интенсивному отбору для повышения откормочных и мясных качеств. При этом в популяциях свиней, не подвергшихся интенсификации, в частности иберийских свиней, в области гена *LCORL* сигналов отбора не обнаружено. Кроме того, сигналы отбора по гену *LCORL* определены у овец швейцарских пород (Signer-Hasler et al., 2019), которые считаются достаточно крупными животными. Например, черноносые овцы Вале в двухлетнем возрасте достигают в холке 72–78 см и имеют массу 70–90 кг, а бараны – 75–83 см и 80–125 кг.

На основании вышеперечисленного, сигнал отбора в гене *LCORL*, вероятно, связан с генетической архитектурой признаков роста свиней, таких как среднесуточный прирост и длина туловища.

В девятой хромосоме установлено 8 сигналов отбора, два из которых расположены в межгенной области, 1 в *LncRNA* и 5 в генах *NCF2*, *APOBEC4*, *HMCN1* (2 сигнала отбора) и *TENM4*. Ген *NCF2* кодирует никотинамидадениндинуклеотидфосфатоксидазу (НАДФН), принимающую активное участие в клеточной противомикробной защитной системе. У людей варианты гена *NCF2* связывают с развитием аутоиммунных заболеваний и нарушением иммунологической толерантности, что приводит к повреждению тканей и органов в результате продукции аутоантител и хронического воспаления (Bakutenko et al., 2022). У свиней роль *NCF2* отмечена в

устойчивости свиней к сальмонеллезной инфекции (Uthe et al., 2011). Эта инфекция в большей степени приводит к локализованному энтероколиту, который наносит значительный экономический ущерб свиноводству. Кроме того, существуют бессимптомные носители, которые составляют долгосрочную зоонозную угрозу и служат резервуаром для заражения соседних свиней, боен и продуктов из свинины. Профилактика и борьба с сальмонеллезом у свиней приобретают огромное значение не только для благополучия животных и повышения эффективности свиноводства, но и для сведения к минимуму рисков для здоровья населения. Все это является импульсом для борьбы с инфекцией, в том числе за счет генетического улучшения поголовья. Вполне вероятно, это могло оставить отпечаток на геномном уровне, в том числе затронув последовательность гена *NCF2* у современного поголовья свиней крупной белой породы.

Белковый продукт гена *APOBEC4* является членом семейства цитидиндеаминаз AID/APOBEC. Несмотря на быстрый прогресс в исследованиях AID/APOBEC, остается много вопросов, в том числе очень мало известно об APOBEC4. Белки AID/APOBEC имеют общую способность вносить мутации в ДНК и РНК путем дезаминирования цитидина до уридина. Возникновение семейства генов AID/APOBEC, по-видимому, совпало с появлением линии позвоночных и эволюцией адаптивного иммунитета (Conticello et al., 2008). AID/APOBECs являются мощными инструментами для улучшения иммунного ответа, однако их уникальная активность – вставка мутаций в нуклеиновые кислоты – поднимает много вопросов относительно их участия в других биологических путях (Marino et al., 2016). Редактирование РНК путем дезаминирования цитидина до уридина считается важным регуляторным механизмом развития видов, а также адаптации сельскохозяйственных животных к различным условиям (Li et al., 2023). На основании имеющейся информации можно предположить, что сигналы отбора в гене *APOBEC4* связаны с изменением изоформ мРНК и возникающим в

результате этого физиологическим изменениям, способствующим адаптации к новым условиям содержания, либо устойчивостью к новым патогенам.

Ген *HMCN1* кодирует гемицентины, которые являются членами семейства фибулинов и определяют правильную морфологию и функциональное поведение специализированных типов клеток преимущественно в эластических тканях (Welcker et al., 2021). Сравнительно мало данных о тканевой локализации и функции гемицентинов у млекопитающих. В работе Lin с соавторами (Lin et al., 2020) было показано, что *HMCN1* образует тонкие внеклеточные дорожки вдоль базальных мембран некоторых структур эластической ткани, однако остались вопросы относительно функции этих структур, т.к. генетическая абляция *HMCN1* и *HMCN2*, а также двойной нокаут с помощью технологии CRISPR/Cas9 не приводили к явным фенотипическим изменениям у мышей.

При исследовании свиней китайских пород и кабанов было отмечено, что ген *HMCN1* находится под сильным искусственным отбором (Yaling et al., 2017). Аналогичные результаты для свиней крупной белой породы получены и в нашей работе. Можно предположить, что это связано с его функциями в формировании структуры эластической ткани и участием в подвижности и морфологии сперматозоидов. Так, при исследовании транскриптома свиней пород ландрас и дюрок van Son с авторами (van Son et al., 2020) определили связь между геном *HMCN1* и гиперактивной подвижностью сперматозоидов. Кроме того, функции гена *HMCN1* связывают с устойчивостью иммунной системы, в частности к казеозному лимфадениту у овец (Kyselová et al., 2023). Это заразное хроническое бактериальное заболевание также известно, как псевдотуберкулез, оно поражает мелких жвачных животных, лошадей, крупный рогатый скот, свиней, диких животных и даже людей.

Интересно отметить, что все гены в 9-й хромосоме, которые затронуты сигналами отбора у современных свиней крупной белой породы, в той или иной степени связаны с иммунной системой, что может быть крайне интересным материалом для дальнейших исследований, связанных с оценкой

генетической архитектуры устойчивости свиней к патогенам различного происхождения.

В 10 хромосоме сигналы отбора обнаружены в интроне гена *FAM171A1* и перед началом кодирующей последовательности (upstream gene variant) гена *NMT2*. Открытию гена *FAM171A1* предшествовало открытие высокоэкспрессированного белка в астроцитах головного мозга, в связи с чем его называли астропринцином (APCN). Далее по результатам клонирования и секвенирования 10-й хромосомы человека последовательность была аннотирована как C10orf38 и впоследствии названа FAM171A1. Экспрессия гена *FAM171A1* зафиксирована в центральной нервной системе, различных тканях, включая трофобласты плаценты, скелетные и сердечные мышцы, почки и поджелудочную железу (Karpio et al., 2021). Основной функцией APCN является регуляция клеточной сигнализации и метаболизма липидов. У свиней обнаружена связь между вариантами гена *FAM171A1* и содержанием жира в туше (Wang et al., 2022). В соответствии с его функцией и экстенсивной экспрессией в различных клетках *FAM171A1* можно считать потенциальным геном-кандидатом положительного отбора на мясные признаки (содержание жира и выход постного мяса) у свиней.

Ген *NMT2* кодирует N-миристоилтрансферазу – фермент, обладающий лизинмиристоилтрансферазной активностью. Этот фермент играет важную роль в регуляции различных сигнальных каскадов в клетках, включая ростовые факторы и ионные проникновения (Kosciuk et al., 2020). В исследованиях выявлена связь между геном *NMT2* и заболеваниями у людей, такими как рак, диабет и некоторые неврологические заболевания, хотя механизмы этой связи еще не до конца изучены. Связь между геном *NMT2* и какими-либо признаками свиней не была обнаружена в научной литературе, однако *NMT2* может быть важным фактором в регуляции липидного обмена и, следовательно, в фенотипе мясных качеств свиней.

В 11-й хромосоме сигналы отбора обнаружены в интроне генов *GPC5* и *DCT*. *GPC5* является членом семейства глипикановых генов. Глипиканы

представляют собой семейство гепарансульфатных протеогликанов, которые связаны с экзоцитоплазматической поверхностью плазматической мембраны с помощью гликозил-фосфатидилинозитолового якоря. Эти белки могут играть роль в контроле клеточного деления и регуляции роста (Yang et al., 2013). Мутации с потерей функции в гене *GPC6* вызывают аутосомно-рецессивную омодисплазию у мышей, которая характеризуется низким ростом, укороченными конечностями и лицевым дисморфизмом (Campos-Xavier et al., 2009). По результатам полногеномных ассоциативных исследований выявлена связь *GPC5* с массой тела у домашних кур (Gu et al., 2012; Ming-Shan et al., 2016). В работе Wang с соавторами (Wang et al., 2021) были определены области в геноме свиней Лайу, подвигнутые интрогрессии от современных европейских свиней. В одной из этих областей локализован ген *GPC5*. Кроме того, в этой работе локально протестировали генетические варианты гена *GPC5*, и ассоциативный анализ показал значимую связь с массой тела и окружностью грудной клетки. Все это может свидетельствовать в пользу положительного отбора вариантов в гене *GPC5* на массу тела и скорость роста у свиней.

В 13-й хромосоме сигналы отбора определены в генах *TRAK1* и *SLC25A26*, которые задействованы в митохондриальном транспорте. Митохондрии генерируют большую часть АТФ, необходимую для жизнедеятельности организма. Митохондриальный протеом составляют примерно 1500 белков и РНК, из которых 13 белков, 2 рибосомные РНК и 22 транспортные РНК кодируются митохондриальной ДНК, все остальные – ядерной ДНК. Митохондрии имеют антероградный и ретроградный транспорт, чередующийся с нечастыми паузами и переключением направлений в аксонах и дендритах (Saxton&Hollenbeck, 2012). Сложными транспортными свойствами митохондрий управляют Miro1, адаптеры TRAK, координирующие противоположные моторы, и другие факторы (Cantu et al., 2023). TRAK1 играет роль в регуляции транспорта митохондрий по аксонам нейронов. Белок, кодируемый геном *TRAK1*, клеточный аксонный

транспортный белок 1, связывается с митохондриальными моторными белками и микротрубулами в аксонах, обеспечивая эффективный и направленный транспорт митохондрий к местам, где они могут выполнять свои функции. TRAK1 также участвует в регуляции митохондриальной физиологии и митохондриальной последовательности событий в ответ на различные сигналы и стрессовые состояния. Белок, кодируемый геном *SLC25A26*, является одним из членов семейства переносчиков ядерных мембран SLC25, которые участвуют в переносе различных метаболитов через митохондриальные мембраны. Это самое большое семейство транспортеров, представленное 53 членами, от *SLC25A1* до *SLC25A53*, и широко распространенное у эукариот (Palmieri, 2013). *SLC25A26* специфично транспортирует аденин-нуклеотиды через внутреннюю митохондриальную мембрану. Этот транспортер играет важную роль в обеспечении митохондрий энергией, влияет на митохондриальную дыхательную цепь и процессы окислительного фосфорилирования.

Значимая роль *TRAK1* и *SLC25A26* в метаболизме энергии, нуклеотидов, жирных кислот и аминокислот, а также в обеспечении кофакторов для ферментов и предшественников для репликации, транскрипции и трансляции в митохондриях позволяет рассматривать положительный отбор в этих генах как необходимость в обеспечении организма ресурсами, необходимыми для интенсивного роста и развития современного поголовья свиней крупной белой породы.

Наибольшее количество сигналов отбора определено в 14-й хромосоме. Они распределились следующим образом: 4 SNP – в межгенной области, 2 SNP в гене *INPP5F*, 4 SNP в гене *TMEM132D* и по одному SNP в генах *PLPP4*, *MN1* и *ADAM28*.

Белковый продукт гена *INPP5F* принадлежит к семейству фосфатаз внутриклеточных сигнальных путей и играет роль в регуляции фосфорилирования инозитолфосфатов в клетках. Инозитолполифосфат-5-фосфатазы участвуют во многих клеточных событиях, включая передачу



сигналов инсулина, эндоцитоз, пролиферацию и активацию гемопоэтических клеток и полимеризацию актина (Quinlan, 2006). *INPP5F* участвует в регуляции клеточной мембраны и сигнальных молекул, контролируя процессы клеточной пролиферации, выживания и миграции. Некоторые варианты гена связывают с различными патологическими состояниями у людей, включая рак, гипертрофию сердца и нейродегенеративные заболевания (Zhao et al., 2020). Zhu с соавторами (Zhu et al., 2009) обнаружили, что у мышей *INPP5F* является важным эндогенным регулятором размера кардиомиоцитов и реакции сердца на стресс. Относительно свиней имеются данные о связи полиморфизма гена *INPP5F* со среднесуточным приростом (Zhou et al., 2009). При этом существует предположение, что целенаправленная селекция на увеличение массы тела и мускулатуры непреднамеренно привела к изменениям в сердечно-сосудистой системе свиней, включая снижение объема крови и уровня гемоглобина, а также к более низкому относительному весу сердца и сердечного выброса (Müller et al., 2000). Essen с соавторами (van Essen et al., 2018) пришли к выводу, что у современных свиней с массой тела более 150 кг непропорционально низкие значения ударного объема и сердечного выброса. Это обусловлено непропорционально малым конечным диастолическим объемом левого желудочка и связано с заметной повышением артериального давления в аорте. На сегодняшний день в свиноводстве растет обеспокоенность относительно состояния сердечно-сосудистой системы у современных свиней и снижения адаптивности к стрессу, в связи с чем сигналы отбора в гене *INPP5F*, возможно, связаны с функциональными особенностями сердечно-сосудистой системы современных свиней и требуют дальнейшего исследования в аспекте продуктивности и здоровья животных.

Ген *PLPP4* кодирует магний-независимую фосфолипидфосфатазу с широкой субстратной специфичностью, которая выполняет разнообразные функции, катализируя дефосфорилирование ряда биоактивных липидных медиаторов для ослабления активации клеток, а также играют важную роль для высвобождения и притока внутриклеточного  $\text{Ca}^{2+}$ . По результатам

исследований можно отметить роль *PLPP4* в развитии и прогрессировании рака у людей. Есть сведения о связи *PLPP4* с яйценоскостью и значительной дифференциальной экспрессии в фолликулах яичников на разных стадиях развития (Chen et al., 2021). Кроме того, в работе Samaha (Samaha, 2020) отмечена роль *PLPP4* в наследственных заболеваниях, поражающих крупных кошек. У свиней установлена связь гена *PLPP4* с биохимическими показателями крови (билирубином) (Bovo et al., 2019). Все это свидетельствует о вероятности того, что ген *PLPP4* связан с устойчивостью/предрасположенностью свиней к заболеваниям, которые на сегодня еще недостаточно изучены и, соответственно, информация об этих генах не представлены в базах данных и открытых источниках.

Ген *TMEM132D* кодирует белок, точная функция которого неизвестна и функционально не охарактеризована ни у одного организма. Экспрессия генов семейства *TMEM132* сильно распространена в нервной системе различных животных и человека (Wang et al., 2021). *TMEM132D* является геном-кандидатом, генотипы риска которого связаны с тревогой у мышей, паническими расстройствами, различными неврологическими и психиатрическими заболеваниями у людей (Erhardt et al., 2011). Существует связь между влиянием окружающей среды и генетическими факторами восприимчивости, связанными с тревогой (Naik et al., 2018). Белки семейства *TMEM132* регулируют структуру и функцию нейронов, тем самым формируя изменения в мозге, лежащие в основе межиндивидуальной предрасположенности к тревоге. Так, у крупного рогатого скота ген *TMEM132D* подвергся давлению отбора, что связывают с процессами domestikация (Qanbari et al., 2014). Относительно свиней можно предположить, что сигналы отбора в гене *TMEM132D* связаны с адаптацией к стрессу в современном технологическом производстве. В работе Grohmann (Grohmann et al., 2023) с соавторами также отметили, что ген *TMEM132D* находится под действием искусственного отбора у свиней для признаков со сложной архитектурой. Кроме этого, по данным Zhang с соавторами (Zhang et

al., 2019), ген *TMEM132D* потенциально связан с репродуктивными признаками свиней. Вопрос относительно того, каким образом коррелируют восприимчивость к тревоге, стресс и продуктивные качества (в том числе репродуктивные) у свиней на сегодняшний день остается открытым, но очевидно, что сигналы отбора в гене *TMEM132* функционально обоснованы и вносят вклад в развитие нервной системы и, возможно, признаков продуктивности свиней.

Ген *MN1* является орфанным геном, т.е. у него нет гомологов в других линиях, в отличие от генов, эволюция которых обусловлена механизмами дупликации (Tautz et al., 2011). Орфанные гены составляют до трети генов во всех геномах, включая бактерии, археи и фаги, и считаются особенно важными для таксон-специфической адаптации развития и взаимодействия с окружающей средой (Long et al., 2003). Участие гена *MN1* было определено в фенотипах миелоидного лейкоза у людей, а также в черепно-лицевых дефектах, затрагивающих перепончатые кости черепа (Tian et al., 2021). В результате нокаута *Mn1* предположили, что лейкозный фенотип этого гена является лишь побочным продуктом, а основная функция *Mn1* заключается в регуляции развития перепончатых костей черепа (Meester-Smoor et al., 2005). Кроме того, в работе Pallares с соавторами показано, что ген *Mn1* может выступать как ключевой ген в генетической архитектуре размера и формы черепа.

Тот факт, что форма и размер черепа являются высоконаследственными полигенными признаками (Geiger et al., 2018), а кроме того, они интегрированы с мозгом и сенсорными системами, с дыхательной и пищеварительной системами, позволяет предположить, что вариант в гене *MN1* оказался под давлением отбора у коммерческих свиней крупной белой породы в результате некоторых морфологических и физиологических преобразований, возможно, связанных с формой черепа и значимыми признаками продуктивности, например, со скороспелостью и толщиной шпика. В качестве подтверждения этого предположения могут выступать

большие различия в форме черепа между дикими кабанами и домашними свиньями (Joseph et al., 2014). Также непосредственно у свиней крупной породы, от ее создания до настоящего времени, морфология черепа несколько раз менялась. Взаимосвязь между формой черепа и толщиной шпика была исследована Whelehan и Gregory (Whelehan et al., 1985), которые показали, что свиньи с коротким широким черепом, как правило, были худее, чем свиньи с длинным узким черепом. Прямые доказательства о связи гена *MN1* с фенотипами качества мяса у свиней получены Óvilo с соавторами (Óvilo et al., 2014) на основе исследования мышечного транскриптома. Кроме того, этот ген оказался под давлением отбора при оценке свиней ливенской породы (местная порода свиней, выведенная в СССР, мясо-сального направления) и свиней пород крупная белая и ландрас (Chernukha et al., 2024).

В целом в результате генетического отбора на повышение скорости роста и снижение толщины шпика произошли изменения в гене *MN1*, что позволяет его рассматривать в качестве гена-кандидата откормочных и мясных признаков свиней. Кроме того, в дальнейшем интересны исследования в аспекте использования формы черепа при оценке состава туши и других селекционных показателей.

Ген *ADAM28* принадлежит к семейству дезинтегринов и металлопротеиназ, группы мембранных белков, участвующих во множестве биологических процессов, включающих межклеточные взаимодействия и взаимодействия клеток с матриксом, включая гаметогенез и оплодотворение (Wolfsberg et al., 1995). Предполагают, что репродуктивные гены, в том числе участвующие в оплодотворении и сперматогенезе, развиваются более быстрыми темпами по сравнению с генами, связанными с другими тканями, и более подвержены положительному отбору (Civetta et al., 1998; Singh et al., 2000; Barb et al., 2004). В частности, *ADAM28* преимущественно экспрессируется в придатках яичка, что указывает на его непосредственную роль в созревании сперматозоидов (Howard et al., 2000), и, согласно нашим результатам, он подвергся положительному отбору у коммерческого

поголовья свиней крупной белой породы. Связь между геном *ADAM28* и репродуктивными признаками крупного рогатого скота представлена в работе Nicolini с соавторами (Nicolini et al., 2018), которые провели сканирование генома быков-производителей голштинской породы и оценку их фертильности. Данные о связи гена *ADAM28* с репродуктивными качествами свиней на сегодняшний день отсутствуют в открытых источниках. Имеются данные о связи *ADAM28* с уровнем среднего корпускулярного гемоглобина у свиней, представленные Lindholm-Perry с соавторами (Lindholm-Perry et al., 2017). Гематологические и клинико-биохимические параметры крови отражают физиологический статус животных и вполне могут оказаться под давлением в результате целенаправленного отбора по селекционно-значимыми признакам. Все это свидетельствует о том, что вариант гена *ADAM28* подвергся давлению отбора, т.к. он прямо или косвенно связан с физиологическими процессами, регулирующими продуктивность свиней коммерческого поголовья.

В 17-й хромосоме вариант отбора определен в гене *ZNF217*. *ZNF217*, член семейства Krüppel-подобных генов, напрямую контролирует экспрессию целевого гена и является двухсторонним фактором транскрипции (Fahmé et al., 2022). Функциональный анализ показал, что он связан с взаимодействием между системами свертывания крови и комплемента, передачи сигналов АФК (активные формы кислорода), нейрогенезом и регуляцией процессов связывания белков и катаболических процессов (Oczkiewicz et al., 2019). По результатам работы Yibin Qiu с соавторами (Yibin et al., 2023) этот ген был отмечен как молекулярный маркер для генетического улучшения массы свиных туш. Кроме того, экспрессия в жировой ткани гена *ZNF217* зависит от источника пищевых жиров (Oczkiewicz et al., 2019). Все это свидетельствует о том, что селекционные предпочтения людей и изменение рационов питания способствовали положительному отбору в гене *ZNF217* у современных свиней крупной белой породы.

Таким образом, микроэволюционные изменения на уровне популяции могут стать предметом положительного отбора, а их совокупный эффект может вызвать большие изменения в фенотипе. По результатам поиска вариантов положительного отбора у современного поголовья свиней крупной белой породы можно отметить, что изменения генома могут быть связаны с фенотипическими характеристиками селекционно-значимых признаков. В первой хромосоме варианты отбора связаны с материнским феноменом искажения передачи информации и, вероятно, оказывают влияние на процессы эмбриогенеза у свиней. В раннем эмбриональном развитии также участвуют варианты положительного отбора, локализованные в 3-ей хромосоме. 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 крупной белой породы в той или иной степени связаны с иммунной системой, что может быть крайне интересным материалом для дальнейших исследований, связанных с оценкой 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, которые обусловлены интенсивными технологиями, реализуемыми на сегодняшний день в ведущих селекционно-генетических центрах. На основании этого варианты положительного отбора могут выступать в качестве кандидатов селекционно значимых признаков, с устойчивостью к стрессам и различным заболеваниям.

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

Среди селекционно-значимых признаков репродуктивные показатели, в частности количество поросят при рождении (TNB), имеют первостепенное значение как экономический признак, играющий решающую роль в эффективности производства. Количество поросят – это полигенетический признак, определяемый множеством регуляторных генных путей. Исследование молекулярных механизмов, контролирующих репродуктивные процессы, должно привести к лучшему пониманию генетической регуляции воспроизводства. Для поиска значимых генетических предикторов, ассоциированных с этим признаком, используют полногеномные панели SNP в качестве маркеров и статистические методы, которые одновременно фиксируют эффекты большого количества SNP.

На сегодняшний день все больший интерес вызывает поиск значимых предикторов для оценки племенной ценности по селекционно-значимым признакам, в частности количеству поросят при рождении, на основе полногеномного ассоциативного анализа (GWAS), реализуемого с помощью различных линейных моделей и алгоритмов машинного обучения.

Процесс применения GWAS можно разделить на 5 этапов:

- вменение (импутация) отсутствующих и кодировка данных, а также анализ распределения изучаемого целевого признака (фенотипа);

- фильтрация данных – поиск отсутствующих данных, анализ неравновесия по сцеплению (LD), уменьшение размерности данных;
- разделение данных на обучающие и тестовые наборы;
- выбор и обучение моделей;
- оценка моделей путем использования тестовых данных.

Анализ отсутствующих данных (NA) является важным этапом при разработке модели, так как подавляющее большинство алгоритмов машинного обучения не работают при наличии NA. На рисунке 53 визуализированы отсутствующие генотипы в исследуемой выборке.

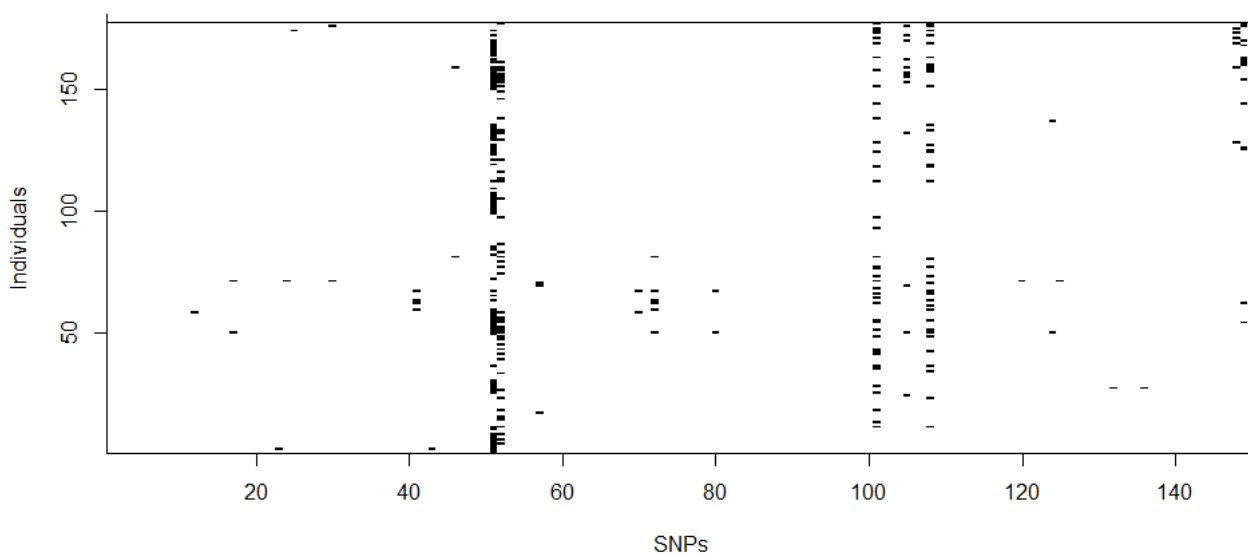


Рисунок 53 – Визуализация отсутствующих генотипов в данных генотипирования

В связи с тем, что имеются отсутствующие данные, можно применить следующие действия:

- удалить наблюдения, которые являются неполными (строки);
- удалить переменные, которые содержат пропущенные значения (SNP);
- восстановить недостающие значения, используя имеющуюся информацию (вменение).

Удаление строк можно применять только в том случае, если их много и процент неполных записей очень низок, что неприменимо в нашем случае.



Третья стратегия позволяет восстановить недостающие данные. Перед применением были удалены столбцы, у которых количество NA превышало 50%. Таким образом, для дальнейшей обработки остались 52241 SNP в качестве предикторов.

Полногеномные ассоциативные исследования (GWAS) являются мощным методом для определения генетических вариантов, связанных с различными селекционными признаками. Однако для идентификации воспроизводимых значимых ассоциаций по всему геному GWAS требуются очень большие размеры выборки. Недостатки в дизайне исследования и ошибки в определении генотипа могут привести к систематическим ошибкам в исследованиях генетических ассоциаций, что приведет к увеличению числа ложноположительных и ложноотрицательных ассоциаций. Применение инструментов искусственного интеллекта, включая генерацию синтетических данных с помощью генеративно-состязательных сетей (GAN), позволяет преодолевать ограничения, связанные с малыми объемами исходных данных. Это открывает возможности для исследований взаимосвязей «генотип–фенотип» даже при недостаточном размере выборки, так как алгоритмы ИИ способны компенсировать дефицит информации за счет моделирования реалистичных данных. Тип распределения переменной имеет фундаментальное значение практически во всех исследованиях, связанных с данными, от проверки гипотез до создания моделей. Кроме того, для синтезируемых данных первостепенное значение имеет точное определение типа распределения исследуемых данных. В качестве анализируемого признака выбрано количество просят при рождении у свиноматок крупной белой породы (рисунок 54).

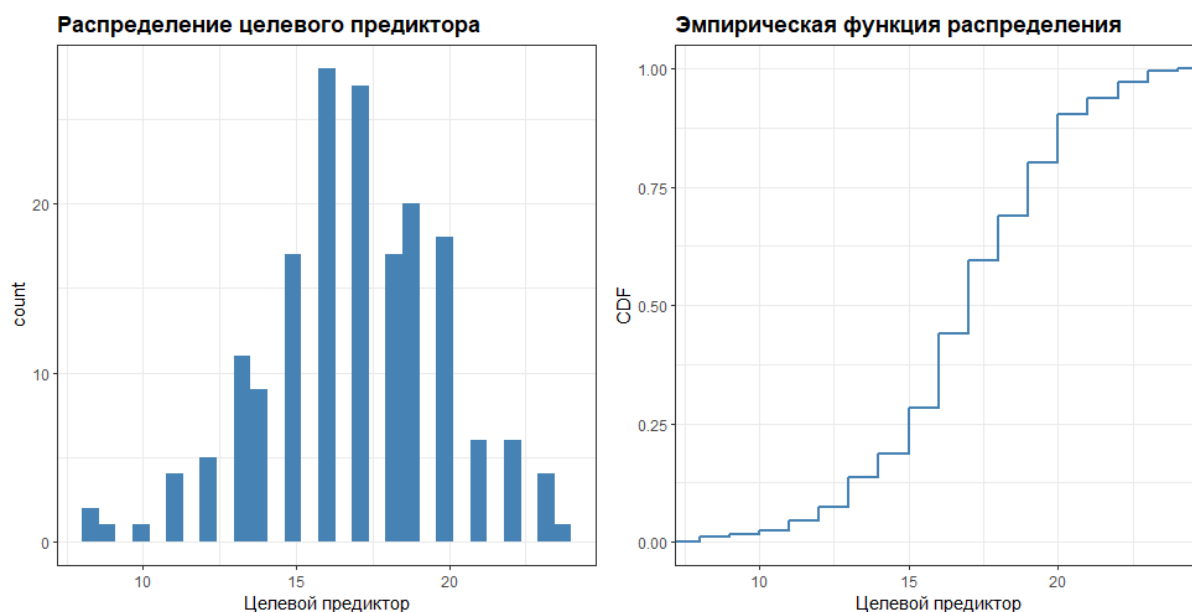


Рисунок 54 – Распределение данных по количеству поросят при рождении

Для точного определения типа распределения изучаемых данных использовали метод максимального правдоподобия, реализованный в пакете UnivariateML. Это позволило выполнить следующие действия:

- подогнать модель (распределение) к данным;
- оценить параметры распределения по максимальному правдоподобию;
- оценить неопределенность и доверительные интервалы параметров с помощью начальной загрузки;
- сравнить различные подходящие модели с использованием показателей AIC и BIC;
- создать кривые QQ, PP и плотности для диагностики;
- учитывая заданное распределение, оценить вероятности, квантили и смоделировать новые данные.

В таблице 15 приведены значения AIC для 10 типов распределения, которые расположены в убывающем порядке. В верхней части таблицы приведены наиболее приемлемые типы распределения для исследуемого признака.

Таблица 85 – Значение AIC для анализируемых типов распределения

| №  | Распределение          | df | AIC       |
|----|------------------------|----|-----------|
| 1  | mlweibull(dff\$TNB)    | 2  | 893,9000  |
| 2  | mlgamma(dff\$TNB)      | 2  | 906,8896  |
| 3  | mllnorm(dff\$TNB)      | 2  | 916,4869  |
| 4  | mlinvgauss(dff\$TNB)   | 2  | 917,6553  |
| 5  | mlbetapr(dff\$TNB)     | 2  | 927,6022  |
| 6  | mlinvgamma(dff\$TNB)   | 2  | 929,4536  |
| 7  | mlinvweibull(dff\$TNB) | 2  | 1001,5434 |
| 8  | mlrayleigh(dff\$TNB)   | 1  | 1128,4711 |
| 9  | mlexp(dff\$TNB)        | 1  | 1356,7151 |
| 10 | mllgamma(dff\$TNB)     | 2  | -         |

Наиболее точным типом распределения для количества поросят при рождении является распределение Вейбулла, которое принадлежит к асимптотическому распределению третьего типа крайних членов вариационного ряда. Доверительные интервалы  $C(0,05; 0,95)$  ограничены значениями (17,74407; 18,481728). Полная картина эмпирического распределения признака, а также графики QQ и PP представлены на рисунке 55.

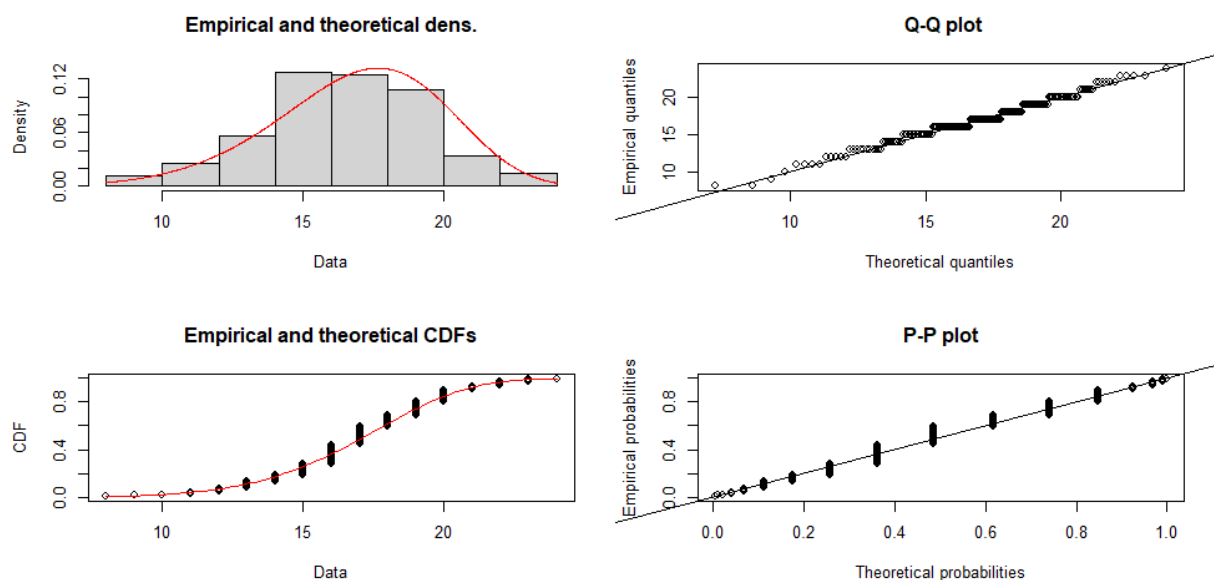


Рисунок 55 – Эмпирическое распределение исследуемого признака

В результате применения комбинированного метода Copula и GAN получена выборка из 5000 образцов, качество (идентичность) которых

составила 94% ( $\text{evaluate}(\text{sample}, \text{data}) = 0.9455388$ ), которые были использованы для ассоциативного анализа алгоритмами машинного обучения.

Для успешной реализации прогнозного или ассоциативного моделирования на основе алгоритмов машинного обучения требуется разделить данные на обучающие (тренировочный набор) и тестовый набор. Основные характеристики обучающего набора в контексте использования генотипированных данных:

- набор объединяет молекулярные (независимые переменные) и фенотипические (зависимые переменные) данные;
- набор содержит достаточно наблюдений (животных в данном случае) и предикторов (молекулярные данные), чтобы иметь возможность обучать модель.

Схема разделения данных на тренировочный и тестовый наборы представлена на рисунке 56. В зависимости от поставленной задачи разделение набора данных на обучающий и тестовый наборы проводится с применением различных стратегий.

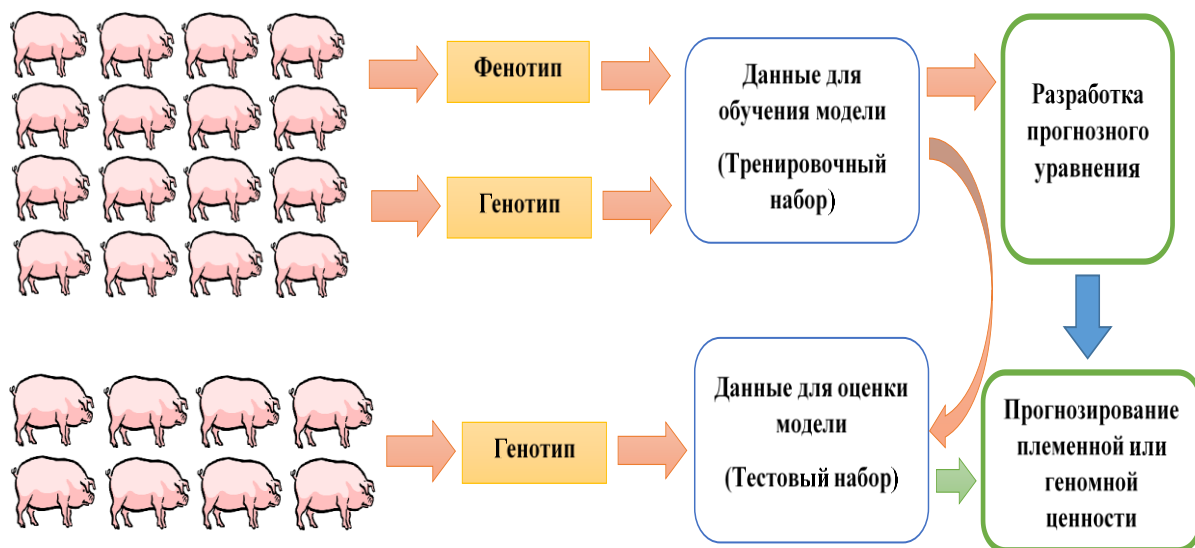


Рисунок 56 – Схематическое изображение тренировочного и тестового наборов, необходимых для реализации ассоциативного анализа

Одна из стратегий состоит в том, что разделение данных проводится в различных пропорциях, начиная с пропорций 80 : 10 : 10 до 20 : 10 : 80, где

первое число – обучающий набор, второе – валидационный набор и третье число – тестовый набор. При этом пропорции меняются последовательно, для каждой пропорции определяется точность модели, а затем вычисляется ее среднее значение. Другая же стратегия предполагает разделение данных в пропорциях 70 : 20 : 10 или 60 : 20 : 20, а точность модели улучшается путем подбора ее различных параметров. Для поиска значимых ассоциаций между генетическими вариантами и количеством поросят при рождении акцент сделан на подборе параметров модели и данные разделили в пропорциях 70 : 20 : 10.

Не менее важным процессом подготовки наборов данных для моделирования является кодирование генотипов. Генотип двухаллельного SNP по своей природе является категориальным и может быть либо гомозиготным, либо гетерозиготным. Например, в упрощенном виде он может быть представлен парой аллелей AA и BB для гомозиготных генотипов и AB для гетерозиготного генотипа. Распространенным способом кодирования генотипа является либо кодировка  $\{-1, 0, 1\}$ , либо  $\{0, 1, 2\}$ . При этом первое и третье числа кодируют гомозиготы, а второе число – гетерозиготу. Кодировки генотипа  $\{0, 1, 2\}$  и  $\{-1, 0, 1\}$  соответствуют измерению аддитивного эффекта при SNP. Закодированный генотип  $\{0, 1, 2\}$  представляет количество минорных аллелей в генотипе, а генотип  $\{-1, 0, 1\}$  связан с уровнем доминирования (Kim et al., 2020). Также для анализа влияния эффекта доминирования его кодируют как 0 для гомозиготного старшего AA, а гетерозиготный AB и гомозиготный минорный аллель BB кодируют как 1. Следует отметить, что много исследований посвящены проблеме кодирования генотипов для проведения ассоциативного анализа, и наилучшей политикой считается поиск различных методов ассоциации, результаты которого будут устойчивыми при различных кодировках генотипов.

При проведении ассоциативного анализа для поиска значимых предикторов количества поросят при рождении приняли кодировку  $\{0, 1, 2\}$ .

При вычислении уравнений модели закодированные генотипы учитывали как непрерывные числа с учетом аддитивного эффекта.

Для достижения наибольшей точности при поиске значимых генетических вариантов для изучаемого признака использовали 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)) и 2 линейные модели с регуляризаторами Лассо-регрессия (Lasso-regression (LR)) и Гребневая регрессия (Ridge-regression (RR)).

На следующем этапе вычислили мощность статистического теста для двух моделей, показавших наиболее значимые результаты. Вычисление мощности статистического теста является фундаментальным этапом в генетических исследованиях, так как оно позволяет оценить вероятность того, что используемый метод анализа сможет обнаружить реальные генетические эффекты, если они существуют. Для этого мы выполнили симуляцию генетических данных и оценку мощности для анализа ассоциации между генотипами и фенотипом, включая размер выборки  $n = 548$ , число однонуклеотидных полиморфизмов  $SNP = 32000$ , размер эффекта  $= 0.1$ , уровень значимости  $\alpha = 0.05$  и количество симуляций  $b = 1000$ . Для каждой симуляции проверяли значимость  $SNP$  на основе их важности, определенной моделями. Результаты анализа были использованы для расчета мощности, которая представляет собой долю симуляций, в которых были обнаружены значимые  $SNP$ . Актуальность вычисления мощности обусловлено тем, что данные могут быть подвержены различным источникам вариабельности, таким как генетическая изменчивость или ошибки измерений.

Для оценки качества модели рассчитывали показатели MAE, MSE, RMSE, R2 и RMSLE (таблица 16).

Таблица 16 – Метрики моделей

| Метрики                                                                | Тренировочный набор |                      | Тестовый набор   |
|------------------------------------------------------------------------|---------------------|----------------------|------------------|
|                                                                        | среднее значение    | стандарт. отклонение | среднее значение |
| Глубокое обучение (Deep Learning (DL))                                 |                     |                      |                  |
| mae                                                                    | <b>0,590020</b>     | 0,071773             | <b>0,4750844</b> |
| mean_residual_deviance                                                 | 0,559832            | 0,128388             | 0,3502782        |
| mse                                                                    | 0,559832            | 0,128388             | 0,3502782        |
| r <sup>2</sup>                                                         | <b>0,939155</b>     | 0,011273             | <b>0,9618232</b> |
| residual_deviance                                                      | 0,559832            | 0,128388             | 0,3502782        |
| rmse                                                                   | 0,745025            | 0,084581             | 0,5918431        |
| rmsle                                                                  | 0,043911            | 0,005592             | 0,03451773       |
| Экстремальный градиентный спуск (Extremal Gradient Boosting (XGBoost)) |                     |                      |                  |
| mae                                                                    | <b>1,510068</b>     | 0,071681             | <b>1,871898</b>  |
| mean_residual_deviance                                                 | 3,617699            | 0,413383             | 5,514274         |
| mse                                                                    | 3,617699            | 0,413383             | 5,514274         |
| r <sup>2</sup>                                                         | <b>0,604665</b>     | 0,031571             | <b>0,3989991</b> |
| residual_deviance                                                      | 3,617699            | 0,413383             | 5,514274         |
| rmse                                                                   | 1,899250            | 0,108266             | 2,348249         |
| rmsle                                                                  | 0,111984            | 0,008008             | 0,1389622        |
| Машина градиентного спуска (Gradient Boosting Machine (GBM))           |                     |                      |                  |
| mae                                                                    | <b>1,884282</b>     | 0,075253             | <b>1,863851</b>  |
| mean_residual_deviance                                                 | 5,517185            | 0,427668             | 5,470074         |
| mse                                                                    | 5,517185            | 0,427668             | 5,470074         |
| r <sup>2</sup>                                                         | <b>0,393505</b>     | 0,050699             | <b>0,4038164</b> |
| residual_deviance                                                      | 5,517185            | 0,427668             | 5,470074         |
| rmse                                                                   | 2,347314            | 0,090063             | 2,338819         |
| rmsle                                                                  | 0,137453            | 0,005693             | 0,1384256        |
| Случайный лес (Random Forest (RF))                                     |                     |                      |                  |
| mae                                                                    | <b>1,983754</b>     | 0,106916             | <b>0,7669743</b> |
| mean_residual_deviance                                                 | 5,979893            | 0,606600             | 0,9717731        |
| mse                                                                    | 5,979893            | 0,606600             | 0,9717731        |
| r <sup>2</sup>                                                         | <b>0,343705</b>     | 0,041655             | <b>0,3938947</b> |
| residual_deviance                                                      | 5,979893            | 0,606600             | 0,9717731        |
| rmse                                                                   | 2,442547            | 0,124092             | 0,9857855        |
| rmsle                                                                  | 0,143224            | 0,008028             | 0,05970351       |
| Лассо-регрессия (Lasso-regression (LR))                                |                     |                      |                  |
| mae                                                                    | <b>0,392218</b>     | 0,018287             | <b>0,7721181</b> |
| mean_residual_deviance                                                 | 0,256187            | 0,040168             | 0,9856217        |
| mse                                                                    | 0,256187            | 0,040168             | 0,9856217        |
| null_deviance                                                          | 2152,24316          | 251,82135            | 0,8122487        |
| r <sup>2</sup>                                                         | 0,935614            | 0,008178             | <b>0,9015521</b> |
| residual_deviance                                                      | 90,214560           | 18,565221            | 0,99255541       |
| rmse                                                                   | 0,5612258           | 0,0568789            | 0,06122583       |
| Гребневая регрессия (Ridge-regression (RR))                            |                     |                      |                  |
| mae                                                                    | <b>0,383172</b>     | 0,017936             | <b>0,7669743</b> |
| mean_residual_deviance                                                 | 0,233283            | 0,031785             | 0,9717731        |
| mse                                                                    | 0,233283            | 0,031785             | 0,9717731        |
| null_deviance                                                          | 2756,310800         | 249,617690           | 0,8938947        |
| r <sup>2</sup>                                                         | <b>0,974526</b>     | 0,002510             | <b>0,9717731</b> |
| residual_deviance                                                      | 70,136820           | 10,107722            | 0,9857855        |
| rmse                                                                   | 0,482020            | 0,032305             | 0,05970351       |

Вывод о качестве модели можно сделать путем сравнения значений указанных показателей между результатами модели на обучающем и тестовом наборах данных.

Глубокое обучение (Deep Learning) и гребневая регрессия (Ridge-regression (RR)) превзошли другие модели по всем критериям оценки, использованным в настоящем исследовании. Коэффициенты детерминации для глубокого обучения и гребневой регрессии составили 0,939 и 0,975 соответственно.

Мощность статистического теста, вычисленная для моделей глубокое обучение и гребневой регрессии равна 1. Это означает, что указанные модели обладают максимальной способностью обнаруживать реальные генетические эффекты при заданных условиях симуляции, т.е. модели успешно идентифицировали значимые однонуклеотидные полиморфизмы (SNP) во всех проведенных симуляциях, также свидетельствует о том, что аналитический подход является чрезвычайно чувствительным к наличию генетических эффектов, и вероятность пропуска истинно положительных результатов (ложноотрицательных ошибок) минимальна. Это положительный результат, который указывает на высокую надежность и эффективность используемых моделей.

Генетические варианты, установленные в моделях DL и RR, были ранжированы по значениям важности, и для дальнейшего анализа выбраны те, которые превышали квантиль 0,99. Результаты анализа обогащения генов, в которых локализованы значимые генетические варианты, определенные по модели DL, представлены на рисунке 57. В качестве основных путей определены деградация лизина, сигнальные пути TGF- $\beta$  и сигнальные пути, регулирующие плюрипотентность клеток. Этот анализ выявил три ключевых биологических процесса, которые статистически значимо ( $p=0.04$ ) связаны с изучаемым набором генов или белков. Этими процессами являются: деградация аминокислоты лизина, сигнальные пути TGF-бета и сигнальные пути, контролирующие плюрипотентность (способность дифференцироваться



во многие типы клеток) стволовых клеток. Обнаружение этих путей указывает на их вероятную важную роль в биологическом контексте, из которого были получены исходные данные.

Среди выявленных путей наибольшую степень обогащения демонстрирует деградация лизина (примерно в 75 раз выше ожидаемого), что подчеркивает возможное значение метаболических изменений в исследуемой системе. Также очень существенно представлены сигнальные пути TGF-бета (обогащение ~55 раз) и регуляция плюрипотентности стволовых клеток (обогащение ~40 раз). Это указывает на высокую активность этих сигнальных каскадов, которые часто взаимосвязаны и играют критическую роль в развитии, дифференцировке клеток и поддержании свойств стволовых клеток.

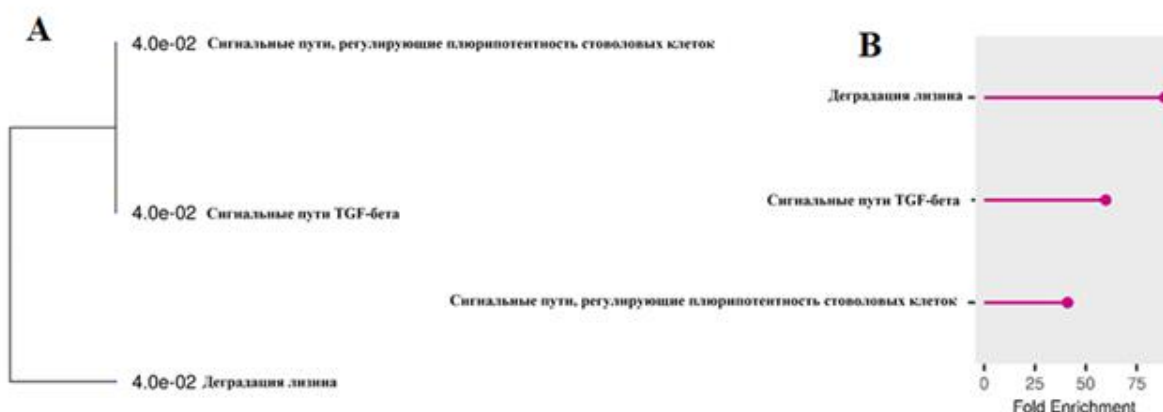


Рисунок 57 – Результаты обогащения генов, на основе алгоритма DL:  
А – дерево путей, В – распределение путей по степени обогащенности

Результаты анализа обогащения генов, в которых локализованы значимые генетические варианты, определенные по модели DL, представлены на рисунке 58.

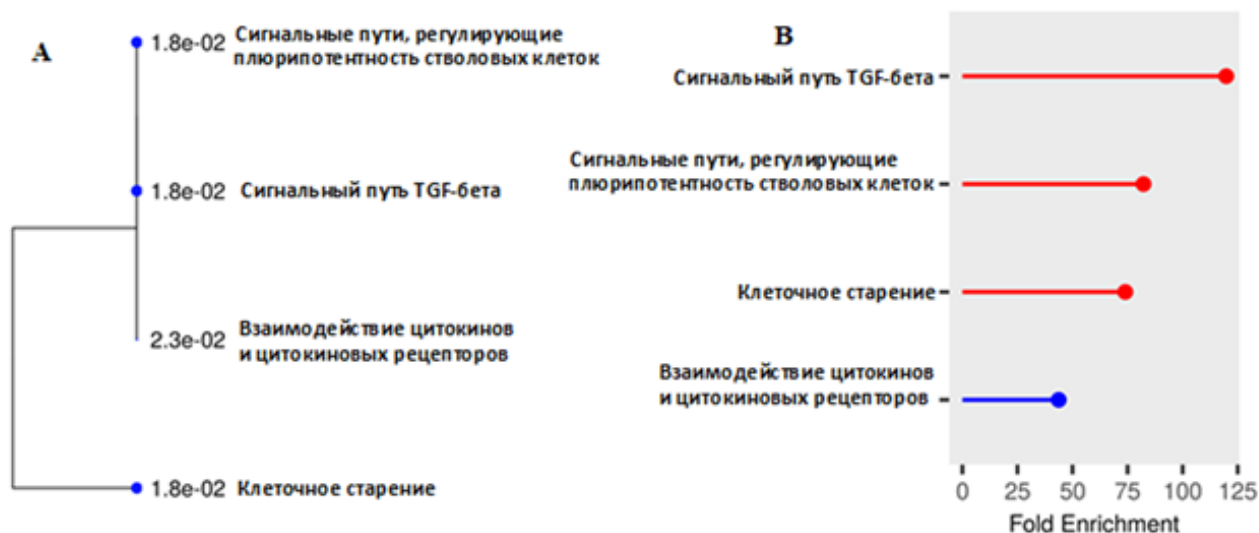


Рисунок 58 – Результаты обогащения генов, на основе алгоритма RR:  
А – дерево путей, В – распределение путей по степени обогащенности

Анализ выявляет четыре биологических процесса, статистически значимо связанных с исследуемым набором данных. К ним относятся сигнальные пути, регулирующие плюрипотентность стволовых клеток, сигнальный путь TGF-бета, клеточное старение (все с p-value 0.018) и взаимодействие цитокинов с их рецепторами (p-value 0.023). Это указывает на то, что процессы, связанные с поддержанием свойств стволовых клеток, клеточной коммуникацией через TGF-бета и цитокины, а также процессы клеточного старения, играют важную роль в изучаемом биологическом контексте.

Наибольшее обогащение наблюдается для сигнального пути TGF-бета (более чем в 115 раз выше ожидаемого), что подчеркивает его центральное значение. Пути, связанные с регуляцией плюрипотентности стволовых клеток (обогащение ~80 раз) и клеточным старением (обогащение ~75 раз), также демонстрируют очень высокую степень обогащения. Взаимодействие цитокинов и их рецепторов также является значимым (обогащение ~45 раз), хотя и в меньшей степени по сравнению с тремя другими путями. Различие в цвете (красный для трех путей, синий для цитокинового взаимодействия)

может указывать на разную направленность изменений (например, активация против подавления) или принадлежность к разным группам анализируемых генов/белков. В качестве основных путей определены клеточное старение, взаимодействие цитокинов и цитокиновых рецепторов, а также, аналогично модели DL, сигнальные пути TGF- $\beta$  и сигнальные пути, регулирующие плюрипотентность клеток.

Следует отметить, что более половины значимых генетических вариантов по моделям DL и RR были общими. Кроме того, были определены общие варианты по моделям DL и RR и подписям отбора, установленные ранее методами iHS, pSL и HH12 в этой популяции (таблица 17). Эти варианты локализованы в генах *NBAS*, *LRRC8C*, *LRFN3*, *BCL11B*, *CC2D2A*, *KCNK2*, *ESRRG*, *ENSSSCG00000039658*, *NEBL*, *CCDC3*, *TNFRSF19*, *FAM117A*, *ULK4*, *SGCZ*, *TUSC3*, *GHRHR*, *MINDY4*, *INMT*. Все эти гены в той или иной степени связаны со сложными процессами, сопровождающими супоросность (беременность) свиноматок, и обеспечивают изменения, непрерывно происходящие как в желтых телах, так и в эндометрии в зависимости от эмбриональной, предимплантационной или внутриутробной стадий.

Функциональной единицей яичника млекопитающих является фолликул. По мере развития фолликулов соматические клетки пролиферируют и дифференцируются, при этом гранулезные клетки приобретают способность реагировать на половые стероиды, включая эстрогены, андрогены и прогестагены (Vanderhyden et al., 1993). Фоликулогенез – это сложный процесс, требующий четко регулируемой экспрессии и взаимодействия генов (Yin et al., 2012). Решающую роль в развитии фолликулов яичников играют микроРНК, воздействуя на гены, участвующие в фоликулогенезе (Yao et al., 2010). Так, miR-383 может участвовать в пролиферации гранулезных клеток, развитии и созревании ооцитов (Shengchen et al., 2022). Транскрипция miR-383 управляется геном *SGCZ*, который ранее был определен по моделям DL и RR как маркер плодовитости свиней, а также подвергся положительному отбору в этой

популяции. miR-383 расположена в первом интроне гена *SGCZ*. Интронные микроРНК могут происходить из общего транскрипта с генами хозяина или независимо транскрибируются с помощью собственного промотора, подобного РНК-полимеразе II (Hidetaka et al., 2023). В работе Yin с соавторами (Yin et al., 2012) было представлено, что мРНК-383 экспрессируется с геном *SGCZ* в тканях мыши и в гранулезных клетках, обработанных TGF- $\beta$ 1, что указывает на то, что они котранскрибируются. На основании чего можно предположить, что ген *SGCZ* играет важную роль в регуляции роста фолликулов яичников и фертильности и у свиней.

Раннее эмбриональное развитие млекопитающих состоит из ряда высококонсервативных, регулируемых и предсказуемых событий клеточного деления, включая оплодотворение, дробление, активацию зиготического генома, уплотнение морулы и образование бластоцисты. На первых этапах эмбрионального развития контроль осуществляется материнскими транскриптами и белками. Далее инициируется активация зиготического генома, и материнские факторы постепенно исчезают. Активация зиготического генома является важнейшим процессом раннего эмбрионального развития (Schulz et al., 2019). Любые нарушения указанных выше процессов могут негативно влиять на эмбриональное развитие.

У свиней активация зиготического генома происходит с 2-клеточной на 4-клеточную стадию развития эмбриона (Zhang et al., 2022). На 4-й день после оплодотворения эмбрион подвергается процессу уплотнения, что связано с усилением межклеточной адгезии и приобретением полярности клеток, придавая ему вид шелковицы, называемой морула.

На 5-й день происходит образование бластоцисты. Этот процесс тесно связан с внутренней клеточной массой (ICM) и спецификацией трофобласта (TE). Одним из генов, подвергнувшихся положительному отбору и определенным по результатам моделей DL и RR, является ген *CCDC3*. Он кодирует белок, экспрессируемый в эндотелиальных клетках и жировой ткани, и положительно регулирует адипогенез и накопление

липидов. В работе Сао с соавторами (Сао et al., 2014) были проанализированы транскриптомы предимплантационных эмбрионов свиней и мышей. Экспрессия гена *CCDC3* отмечена и у свиней, и у мышей несмотря на вывод авторов о том, что дифференциация трофэктодермы регулируется разными сигнальными путями у эмбрионов свиней и мышей. Это позволяет выдвинуть предположение о значительной роли гена *CCDC3* в предимплантационном эмбриональном развитии, в том числе у свиней.

Таблица 97 – Варианты, определенные по моделям DL и RR и подписям отбора

| SSC | Позиция   | ih12 | iHS  | nSL  | RR   | DL   | Номер rs    | Тип варианта          | Название гена             |
|-----|-----------|------|------|------|------|------|-------------|-----------------------|---------------------------|
| 1   | 26068717  | 2,10 | –    | –    | 0,53 | –    | rs80949671  | upstream gene variant | <i>LncRNA</i>             |
| 1   | 26505944  | 2,16 | –    | –    | 0,51 | –    | rs81351637  | intergenic variant    | -                         |
| 1   | 26593023  | –    | 2,56 | 2,14 | 0,52 | –    | rs81351644  | intergenic variant    | -                         |
| 1   | 26625743  | –    | 2,47 | 2,04 | 0,61 | 0,67 | rs81351664  | intergenic variant    | -                         |
| 1   | 30564821  | –    | 2,46 | –    | 0,75 | –    | rs80857633  | intergenic variant    | -                         |
| 3   | 122158369 | –    | 2,05 | –    | 0,51 | –    | rs81377655  | intron variant        | <i>NBAS</i>               |
| 4   | 125690120 | –    | 2,22 | 2,30 | 0,50 | –    | rs81339593  | intergenic variant    | -                         |
| 4   | 126987143 | –    | –    | 2,01 | 0,58 | 0,67 | rs80958447  | intron variant        | <i>LRRC8C</i>             |
| 5   | 10852290  | –    | 2,17 | –    | –    | 0,69 | rs80842148  | intergenic variant    | -                         |
| 6   | 45342655  | 2,52 | –    | –    | 0,64 | –    | rs81395771  | upstream gene variant | <i>LRFN3</i>              |
| 7   | 14152430  | –    | 2,15 | 2,04 | 0,52 | –    | rs326183563 | intron variant        | <i>LncRNA</i>             |
| 7   | 108105220 | 2,18 | –    | –    | 0,50 | –    | rs80793422  | intergenic variant    | -                         |
| 7   | 120229737 | –    | 2,66 | –    | 0,50 | –    | rs80808822  | intron variant        | <i>BCL11B</i>             |
| 7   | 120610621 | –    | 2,96 | –    | 0,56 | 0,66 | rs80935654  | intergenic variant    | -                         |
| 8   | 10796231  | –    | –    | 2,31 | –    | 0,71 | rs334772761 | intron variant        | <i>CC2D2A</i>             |
| 9   | 128673870 | 2,18 | 2,28 | –    | 0,53 | 0,65 | rs335981516 | intergenic variant    | <i>KCNK2</i>              |
| 10  | 6950011   | –    | 2,14 | –    | –    | 0,67 | rs81428487  | intergenic variant    | <i>ESRRG</i>              |
| 10  | 47212377  | –    | 2,11 | –    | 0,58 | –    | rs81477650  | intron variant        | <i>CDNF</i>               |
| 10  | 53675235  | 2,43 | –    | –    | 0,61 | –    | rs339993261 | intron variant        | <i>NEBL</i>               |
| 10  | 53697286  | 2,28 | –    | –    | 0,51 | 0,63 | rs81314157  | intron variant        | <i>NEBL</i>               |
| 10  | 53738869  | 2,21 | –    | –    | 0,57 | 0,77 | rs81241974  | intron variant        | <i>NEBL</i>               |
| 10  | 59106264  | –    | 2,29 | 2,09 | 0,51 | –    | rs81426512  | intron variant        | <i>CCDC3</i>              |
| 11  | 2538185   | –    | 2,22 | –    | –    | 0,67 | rs331680910 | intergenic variant    | <i>TNFRSF19</i>           |
| 12  | 25912338  | –    | 2,86 | –    | 0,53 | –    | rs81239149  | intron variant        | <i>FAM117A</i>            |
| 13  | 25296621  | 2,45 | –    | –    | –    | 0,65 | rs81284602  | intron variant        | <i>ULK4</i>               |
| 14  | 123823825 | –    | 2,48 | –    | 0,59 | 0,71 | rs320763504 | intergenic variant    | -                         |
| 15  | 29638798  | 2,05 | 2,45 | 2,60 | 0,70 | 0,84 | rs81452296  | intergenic variant    | -                         |
| 15  | 121460992 | –    | –    | 2,08 | 0,60 | –    | rs341236717 | intron variant        | <i>ENSSSCG00000032522</i> |

Продолжение таблицы 17

|    |           |      |      |      |      |      |              |                                    |               |
|----|-----------|------|------|------|------|------|--------------|------------------------------------|---------------|
| 15 | 121927668 | –    | 3,16 | 2,92 | 0,51 | –    | rs81455316   | intergenic variant                 | -             |
| 15 | 121959926 | 2,07 | 2,38 | 2,26 | 0,62 | –    | rs81455304   | regulatory region variant          | -             |
| 15 | 122065360 | 2,37 | –    | –    | 0,59 | 0,69 | rs325379930  | upstream gene variant              | -             |
| 15 | 122153233 | 2,19 | 2,77 | 2,48 | 0,59 | 0,67 | rs337997805  | intergenic variant                 | -             |
| 15 | 122263611 | 2,39 | –    | –    | 0,56 | –    | rs80856455   | intergenic variant                 | -             |
| 15 | 122294134 | 2,27 | –    | –    | 0,69 | 0,72 | rs80965619   | intergenic variant                 | -             |
| 15 | 122374213 | 2,06 | 2,34 | 2,33 | 0,59 | –    | rs318986640  | intergenic variant                 | -             |
| 16 | 7850860   | –    | 2,07 | –    | 0,56 | –    | rs81345229   | intergenic variant                 | -             |
| 17 | 2890084   | 2,22 | –    | –    | 0,50 | –    | rs81465731   | intron variant                     | -             |
| 17 | 3013297   | 2,12 | –    | –    | 0,55 | –    | rs81465735   | intron variant                     | <i>SGCZ</i>   |
| 17 | 3316742   | 2,61 | –    | –    | –    | 0,69 | rs81466107   | intron variant                     | <i>TUSC3</i>  |
| 17 | 3448507   | 2,00 | –    | –    | –    | 0,69 | rs3471300045 | intron variant                     | <i>TUSC3</i>  |
| 17 | 40763707  | –    | 2,04 | –    | 0,52 | –    | rs318485305  | intergenic variant                 | -             |
| 17 | 53760514  | 3,76 | –    | –    | 0,55 | –    | rs80892682   | intergenic variant                 | -             |
| 17 | 53970256  | 2,39 | –    | –    | 0,56 | –    | -            |                                    | -             |
| 18 | 41692890  | 2,04 | 2,25 | –    | 0,51 | 0,73 | rs319528527  | intergenic variant                 | -             |
| 18 | 41992583  | –    | 2,61 | –    | 0,52 | –    | rs81339141   | regulatory region variant          | -             |
| 18 | 42038739  | –    | 3,14 | –    | 0,54 | –    | rs337936594  | intron variant                     | <i>GHRHR</i>  |
| 18 | 42078871  | –    | 3,20 | –    | 0,51 | –    | -            |                                    | -             |
| 18 | 42126037  | –    | 3,70 | –    | 0,54 | –    | rs80945004   | intron variant                     | <i>MINDY4</i> |
| 18 | 42222435  | –    | 2,93 | –    | 0,63 | 0,69 | rs324752471  | downstream gene variant            | <i>INMT</i>   |
| 18 | 42547236  | –    | 2,25 | –    | 0,67 | 0,90 | rs341575491  | non coding transcript exon variant | <i>LncRNA</i> |

В работе Aikawa с соавторами (Aikawa et al., 2022) представлены результаты по изучению экспрессии генов на ранних сроках беременности у мышей. Они обнаружили, что уровни эпителиальных генов динамически меняются с течением времени, и это связано с различными механизмами транскрипции в каждый момент времени. Так, предполагают, что на 5-й день в эпителии повышается экспрессия генов (в том числе *LRRC8C* и *LRFN3*), активируются ангиогенез и цитокиновые пути, увеличивается проницаемость кровеносных сосудов, окружающих места имплантации. Ген *LRRC8C* кодирует одну из субъединиц гетеромерного белка, регулирующих объем анионных каналов. Домены LRR представляют собой эволюционно консервативные белковые структуры, и предполагается, что они способствуют врожденному иммунному ответу. Роль *LRRC8C* отмечена во многих иммунных процессах (Manolia et al., 2022). В работе de Castro (de Castro et al., 2024) *LRRC8C* рассматривается как один из генов, участвующих во взаимодействии эмбриона и эндотелия у кобыл. Эндометрий имеет решающее значение для имплантации бластоцисты. Любые сбои в этом процессе ставят под угрозу последующие исходы беременности.

Различные гены цитокинов связаны с липидным обменом у свиней. Ген *TNFRSF19* является членом суперсемейства рецепторов фактора некроза опухоли и участвует во взаимодействии цитокинов и цитокиновых рецепторов (Wu et al., 2020). *TNFRSF19* у свиней может играть решающую роль в накоплении липидов и иметь отношение к метаболизму жирных кислот. В работе Mekchay с соавторами (Mekchay et al., 2022) показано, что варианты гена *TNFRSF19* связаны с содержанием внутримышечного жира и уровнем арахидоновой кислоты у свиней. Кроме того, высокий уровень экспрессии *TNFRSF19* отмечается в процессе развития эмбриона. Все это может свидетельствовать о значимой роли гена *TNFRSF19* в эмбриональном развитии и плодовитости свиней.

Ряд генов, определенных по анализам DL и RR и подписям отбора, связан с процессами, обеспечивающими рецептивность эндометрия и



имплантацию эмбриона. Рецептивность эндометрия осуществляется за счет существенных структурных изменений в цитоскелете и плазматической мембране эпителиальных клеток, которые облегчают адгезию эмбрионов. Так, ген *NEBL* кодирует белок, принадлежащий к семейству небулиновых актин-связывающих белков, которые важны в миофибрилlogenезе и сборке Z-линии. Предполагается, что варианты гена *NEBL* связаны с различными кардиомиопатиями (Perrot et al., 2016). Кроме того, роль *NEBL* определена в недостаточно высокой рецептивности эпителия и неудачной имплантации эмбриона. Так, Zhou с соавторами (Zhou et al., 2024) определили, что у женщин с рецидивирующей неудачей имплантации значительно повышался уровень *ACTN1* в секреторной фазе эндометрия, при этом уровни *NEBL* значительно снижались, и повышался уровень волокон F-актина, вызывая заметное ухудшение адгезии бластоцисты, а сверхэкспрессия *NEBL* заметно восстанавливала способность клеток к адгезии.

Наиболее значительная потеря эмбрионов у свиней происходит во время имплантации на ранних сроках беременности; смертность достигает 30% в период между 10-м и 30-м днями беременности (Kridli et al., 2016). Предполагается, что основной причиной является маточно-плацентарная недостаточность. Плацента необходима для развития плода и успешного исхода беременности, она служит органом питания, защиты и газообмена, регулирует рост и развитие плода. Важную роль в ранней плацентации, а также пролиферации и дифференцировки трофобласта играет белок, кодируемый геном *ESRRG*. Гомозиготные мутантные эмбрионы, полученные путем целенаправленного разрушения гена *ESRRG*, имеют серьезное нарушение формирования плаценты и погибают через 10,5 дней (Gao et al., 2019).

Любые изменения в структуре плаценты могут быть связаны с риском нарушений развития нервной системы, что может привести к задержке эмбрионального развития и различным аномалиям. Ряд генов,

идентифицированных по результатам анализа DL и RR и подписей селекции, может быть рассмотрен в данном контексте.

Ген *BCL11B* кодирует транскрипционный фактор, необходимый для постнатального развития гиппокампа, и его дефицит приводит к структурным дефектам головного мозга (Lessel et al., 2018). Кроме того, *BCL11B* участвует в детерминации и поддержании линии Т-клеток, а также играет значительную роль в развитии врожденных лимфоидных клеток группы 2 (ILC), продуцирующих эффекторные цитокины 2-го типа (Yu et al., 2015). У людей генетические варианты в гене *BCL11B* связаны с нарушением развития Т-клеток (Punwani et al., 2016), а недостаточная экспрессия *BCL11B* наблюдается в плацентах с задержкой внутриутробного развития и может приводить как к иммунному, так и к нейрональному дисбалансу (Majewska et al., 2019).

Ген *CDNF* (*Armetl1*) экспрессируется в нескольких тканях мыши и человека, включая эмбриональный и постнатальный мозг мыши. Нейротрофический фактор глиальной клеточной линии (GDNF) является членом суперсемейства факторов роста TGF- $\beta$ , обладающих нейротрофической активностью в отношении дофаминергических нейронов среднего мозга, а также развивающихся и зрелых мотонейронов ствола головного мозга и спинного мозга. GDNF-нулевые мыши, независимо от их целевой мутации, демонстрируют полную агенезию почек из-за отсутствия индукции зачатка мочеточника, являющегося ранней стадией развития почек. У этих мышей также нет кишечных нейронов, что, вероятно, объясняет наблюдаемый стеноз привратника и расширение двенадцатиперстной кишки (Sánchez et al., 1996).

Ген *ULK4* кодирует белок, выполняющий несколько функций в головном мозге, включая участие в пролиферации нейрональных клеток и регуляции клеточного цикла, тогда как функции за пределами центральной нервной системы в значительной степени неизвестны (Liu et al., 2017).

Ген *CC2D2A* кодирует белок, играющий решающую роль в формировании ресничек. Первичные реснички участвуют во множестве

различных сигнальных путей, однако молекулярные механизмы, лежащие в основе сенсорной функции первичных ресничек, недостаточно изучены. Дисфункция ресничек лежит в основе развития тяжелых заболеваний, обычно называемых цилиопатиями (Knockenhauer, Schwartz, 2015). Цилиопатии имеют плеiotропный фенотип и поражают несколько органов. В работе Tsyklauri с соавторами (Tsyklauri et al., 2021) представлена связь между цилиопатией и нарушением регуляции иммунной и кроветворной систем. Типичные симптомы цилиопатий могут возникать из-за дисфункции ресничек во время эмбрионального развития. Генетические варианты в *CC2D2A* связаны с целым спектром различных синдромов цилиопатии, включающим неврологические заболевания, такие как синдром Жубера, синдром Меккеля и летальная мультиорганная цилиопатия (Barroso-Gil et al., 2021). Как эти варианты приводят к такому широкому спектру различных клинических проявлений с частично перекрывающимися фенотипами, до конца не понятно. Однако, согласно полученным ранее результатам по моделям и подписям отбора, можно предположить о значительной роли гена *CC2D2A* в плодовитости свиноматок.

Точный контроль ионного потока (кальция, натрия и калия) способствует процессам внутриутробного развития, таким как пролиферация, миграция и дифференцировка нейронов. Дисфункция ионных каналов вызывает ряд патологий, называемых каналопатиями. Ген *KCNK2* (также известный как *TREK1*) кодирует белок, принадлежащий к семейству калиевых каналов с двухпоровым доменом. Высокий уровень экспрессии *KCNK2* отмечена в нескольких областях мозга, таких как обонятельная луковица, гиппокамп, мозжечок и кора, в связи с чем этот ген стал рассматриваться как мишень для ряда патологий, поражающих центральную нервную систему (Djillani et al., 2019). Дисфункцию ионных каналов в основном изучали в контексте постнатальных дифференцированных нейронов, однако теперь стало известно о роли ионных каналов в пренатальных нарушениях головного мозга, приводящих к различным порокам развития коры головного мозга и

аномалиям на ранних и средних стадиях эмбрионального развития (Smith, Walsh, 2020). Роль гена *KCNK2* в эмбриональном развитии и плодовитости свиней на сегодняшний день неизвестна, однако широкое распространение, разнообразные стимулы и регуляции *KCNK2* в человеческом организме, плейотропные роли, которые он может играть в физиологии и физиопатологии, а также представленность в моделях по поиску ассоциации с плодовитостью и подписями отбора делают его интересным объектом для более дальнейших исследований в контексте репродуктивности свиней.

Ген *TUSC3* необходим для усвоения клеточного магния и является ключевой молекулой в эмбриональном развитии (Yu et al., 2017). *TUSC3* имеет ограниченный характер экспрессии, с наибольшей представленностью в яичниках, плаценте, предстательной железе, семенниках, жировой ткани и легких. Исследование полногеномной экспрессии мРНК показало, что он широко распространен в тканях на всех стадиях эмбрионального развития (Jin et al., 2008). Zhou с соавторами (Zhou et al., 2009) показали, что гены *MagT1* и *TUSC3* лежат в основе ключевого компонента клеточного транспорта  $Mg^{2+}$  и необходимы для эмбрионального развития позвоночных. *MagT1* и *TUSC3* играют центральную роль в эмбриональном развитии позвоночных, которая не может быть компенсирована другими предполагаемыми переносчиками  $Mg^{2+}$ . В целом у свиней магний играет значительную роль в проявлении фенотипов репродуктивных признаков свиноматок, таких как интервал между отъемом и течкой, количеством рожденных, живых и отъемных поросят, влияет на переваримость сырой клетчатки и сырого протеина на поздних сроках беременности и в период лактации.

Таким образом, для повышения точности поиска предикторов, ассоциированных с количеством поросят при рождении, объем выборки увеличили за счет синтезируемых данных на основе алгоритма генеративно-состязательных сетей (GAN). В результате объем выборки составил 5000 образцов. Для ассоциативного анализа использовали 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)) и 2 линейные модели с регуляризаторами – Лассо-регрессия (Lasso-regression (LR)) и Гребневая регрессия (Ridge-regression (RR)). На основании метрик (MAE, MSE, RMSE, R2, RMSLE) модели Глубокое обучение (Deep Learning) и Гребневая регрессия (Ridge-regression (RR)) превзошли другие модели по всем критериям оценки. Генетические варианты, установленные моделями DL и RR, участвуют в путях, связанных со взаимодействием цитокинов и цитокиновых рецепторов, деградацией лизина, сигнальными путями TGF- $\beta$ , плюрипотентностью клеток и клеточным старением. Для поиска причинных вариантов были выбраны общие SNP, определенные по моделям DL и RR и подписям отбора (методами iHS, nSL и HH12) в этой популяции. Варианты локализованы в генах *NBAS*, *LRRC8C*, *LRFN3*, *BCL11B*, *CC2D2A*, *KCNK2*, *ESRRG*, *CDNF*, *NEBL*, *CCDC3*, *TNFRSF19*, *FAM117A*, *ULK4*, *SGCZ*, *TUSC3*, *GHRHR*, *MINDY4*, *INMT*. Все эти гены играют значительную роль в период супоросности 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*). 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*).

Эти результаты позволяют предположить важные биологические пути, связанные с плодовитостью свиноматок, а также данные о ключевых генетических вариантах, которые могут выступать как потенциальные маркеры фертильности и включаться в будущие программы 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 крупной белой породы определили 3098 CNV: 429, 810 и 1859 CNV соответственно. CNV размером менее 50 Kb исключали, а для анализа отбирались 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. Максимально количество CNV на одно животное в исследуемой выборке составило 52, минимальное 2 CNV (рисунок 59).

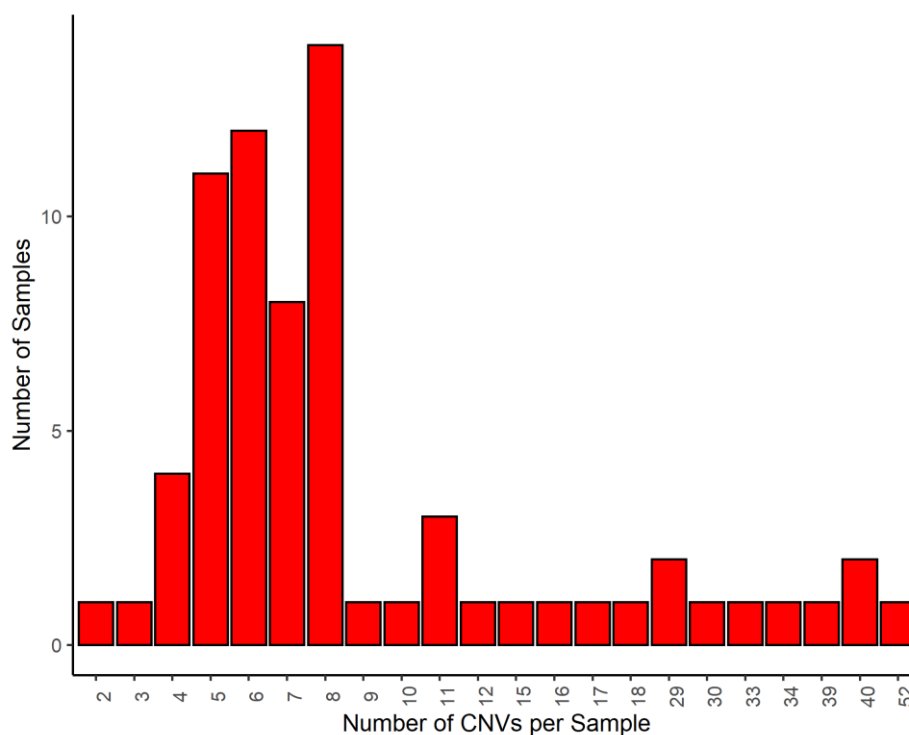


Рисунок 59 – Распределение количества CNV на образец

На рисунке 60 показано распределение различных CNV по хромосомам. Как видно, дубликации равномерно распределены по хромосомам, а делеции в основном локализованы на хромосомах 3, 4, 7 и 8.

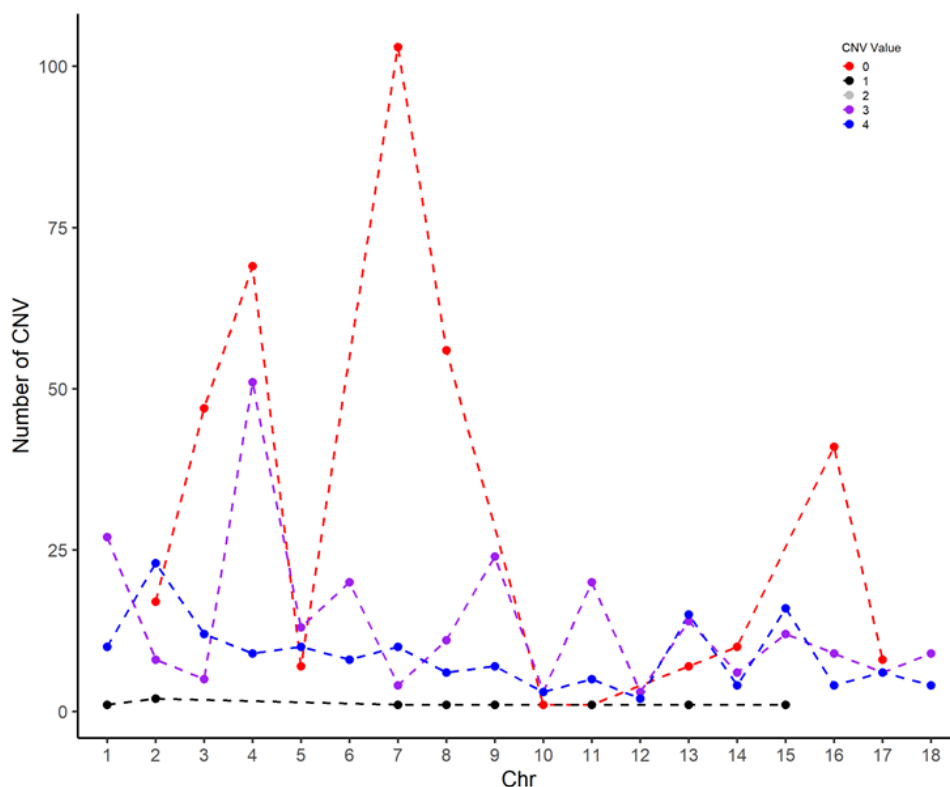


Рисунок 60 – Распределение CNV по хромосомам. Указано значение CNV value: 0 – полная делеция; 1 – однокопийная делеция; 2 – нормальная (гетерозигота) и нормальная (гомозигота); 3 – одиночная дупликация; 4 – двойная дупликация)

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 Mb. Потери копий (Loss) содержат 12 регионов, 91 регион содержит приросты копий (Gain), а 9 регионов включают оба типа, потери и приросты копий (Mixed) (рисунок 61).

Карта визуализирует участки хромосом, где наблюдаются отличия в количестве копий ДНК по сравнению с референсным геномом или средней нормой для популяции. Белым цветом показаны участки с нормальным числом копий, красным - участки с увеличением числа копий (Gain), а зеленым - участки с уменьшением числа копий (Loss). Черным цветом отмечены

"смешанные" регионы (Mixed), где могут присутствовать и потери, и приобретения в разных образцах или сложные вариации.

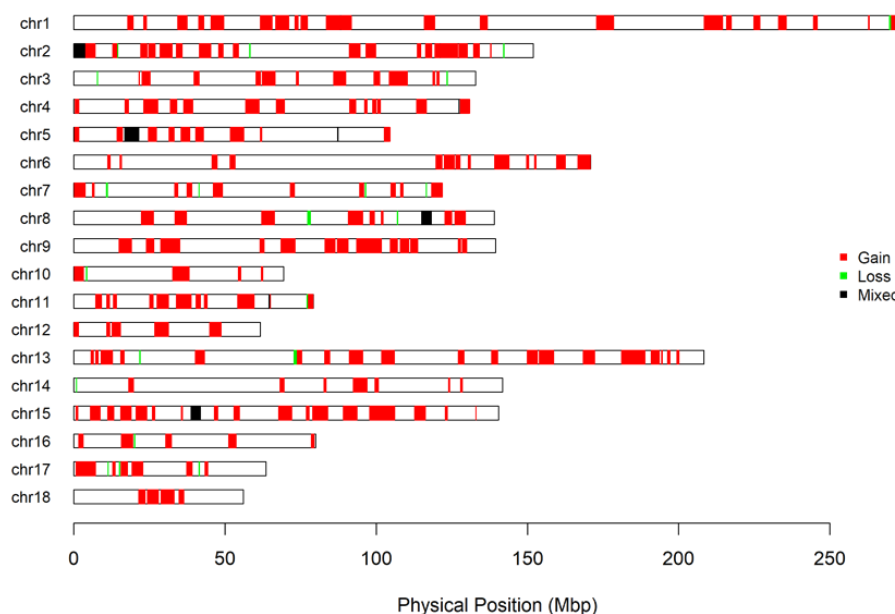


Рисунок 61 – Полногеномная карта CNVR у свиней крупной белой породы

Анализ карты показывает, что CNV распределены по всем исследованным хромосомам крупной белой свиньи, однако их плотность и тип варьируют. Наиболее заметной особенностью является явное преобладание регионов с увеличением числа копий (красные сегменты) над регионами с потерями (зеленые сегменты). Увеличения числа копий встречаются на всех хромосомах и часто образуют кластеры или занимают протяженные участки, например, на хромосомах 1, 6, 9, 13, 15. Это свидетельствует о том, что дупликации сегментов ДНК являются распространенным типом структурной изменчивости в геноме данной породы.

Потери генетического материала (зеленые сегменты) встречаются значительно реже и обычно представлены более короткими участками, разбросанными по разным хромосомам (например, chr2, chr3, chr7, chr8, chr10, chr13, 16, 17). Наличие как потерь, так и приобретений указывает на значительную структурную пластичность генома крупной белой породы. Эти вариации числа копий могут затрагивать гены, влияющие на важные



хозяйственно-полезные признаки, такие как рост, качество мяса, репродуктивные способности и устойчивость к заболеваниям, и могут быть результатом как естественных процессов мутагенеза, так и искусственного отбора в ходе селекции породы.

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; с размножением – 5,88% и внешним видом – 5,42% (рисунок 62).

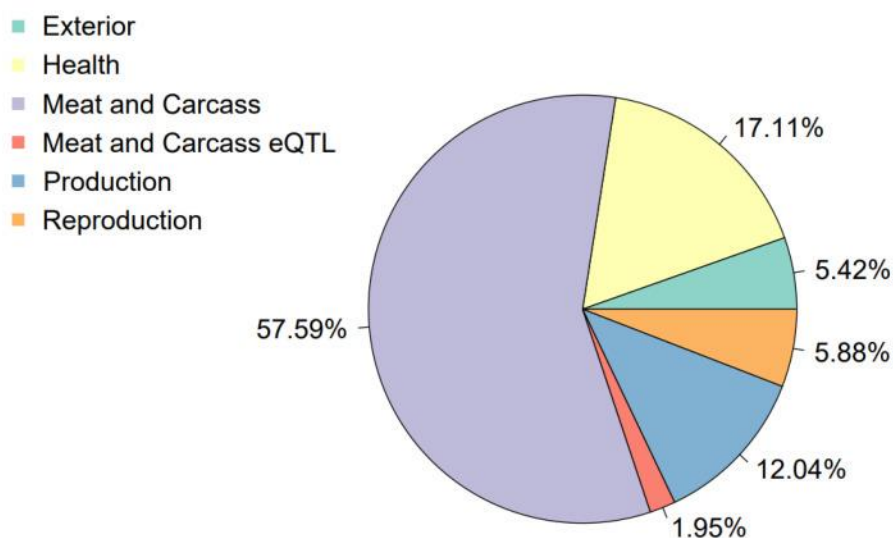


Рисунок 62 – Процент QTL, затронутых CNVR

Для проведения полногеномного ассоциативного анализа были выбраны признаки, представляющую наибольшую селекционную ценность в аспекте мясной и откормочной продуктивности свиней: скороспелость (возраст достижения 100 кг) (дн.), 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), представляющий источник значительных экономических потерь.

По результатам CNV-GWAS определили 29 CNV, ассоциированных с дефектами конечностей, уровень значимости которых был меньше или равен 0,05 без поправки на множественное тестирование (рисунок 63). Варианты расчетов, где использовали поправки на множественное тестирование методом FDR, показали 3 CNV на уровне статистической значимости (рисунок 64). CNV локализованы в хромосомах 1 (1:41300464-42881624), 3 (3:123404335-123460528) и 6 (6:142358005-143865882).

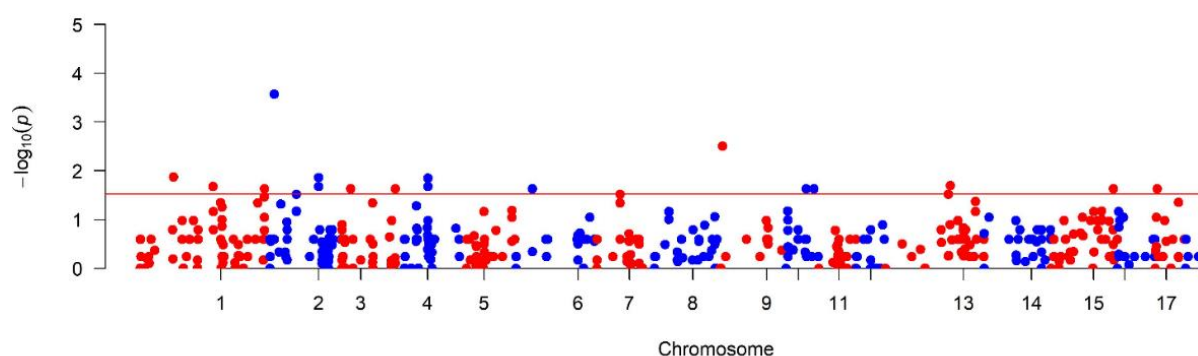


Рисунок 63 – Полногеномные ассоциации CNV (на уровне зонда) с ЗНШ (без поправки на FDR)

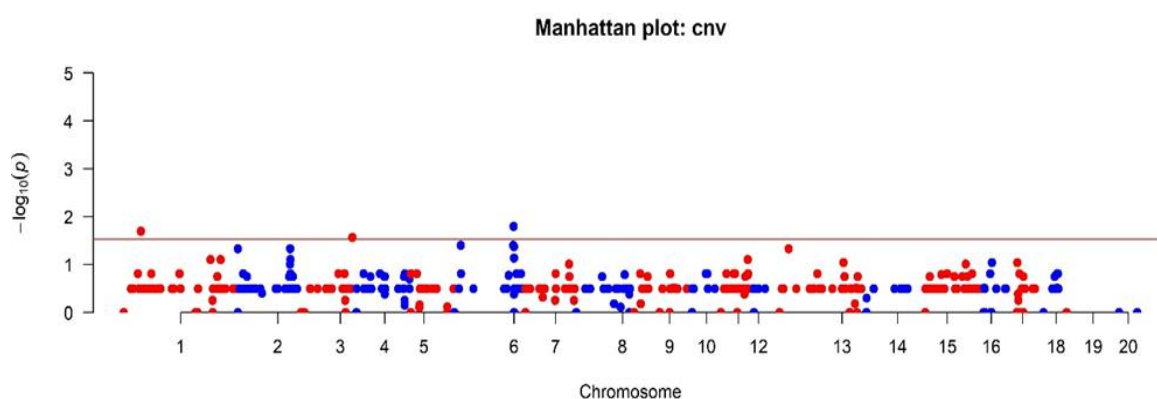


Рисунок 64 – Полногеномные ассоциации CNV (на уровне зонда) с ЗНШ (FDR)

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. В качестве основных функциональных путей для

*СТН* можно выделить метаболические пути, результатом которых является образование или изменение, необходимое для правильного функционирования биологической системы за счет регулирования серосодержащих аминокислот, таких как метионин и цистеин. Метионин метаболизируется в печени и тонком кишечнике путем трансметилирования, реметилирования и транссульфирования. Цистатионин-γ-лиаза, кодируемая геном *СТН*, катализирует последнюю стадию транссульфирования от метионина к цистеину (Kahindi et al., 2018). Помимо синтеза белка, метионин и цистеин влияют на производство цитокинов, глутатиона и белков-предшественников амилазы. Недостаточное количество этих кислот приводит к снижению иммунного ответа, выработке антител и пролиферацию Т-клеток у свиней, а также развитие лимфоидных органов. Все это может быть причинами возникновения дефектов на конечностях свиней. Связь между соотношением серосодержащих аминокислот и экспрессией гена *СТН* дает основание предположить, что CNV в этом гене играет роль в метаболизме серосодержащих аминокислот, тем самым провоцирует процессы, связанные с образованием шишек/наростов у свиней.

Ген *SRSF11* кодирует фактор сплайсинга пре-мРНК и является богатым аргинином-серином белком семейства SR (Sliškoć et al., 2022). 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 (Zappaterra et al., 2020).

Кроме того, CNV 6:142358005-143865882 перекрывается с QTL по признакам мяса и туши, отвечающим за содержание внутримышечного жира. Гипертрофия адипоцитов происходит в результате сдвига метаболизма в сторону липогенеза (Hood, Allen, 1977). Метаболические процессы, происходящие в адипоцитах внутримышечного жира, связаны с риском сердечно-сосудистых заболеваний, резистентности к инсулину, а также с

развитием мышечной ткани (Zappaterra et al., 2020). Связь маркеров упитанности с дефектами конечностей у свиней также описана в исследовании Fan с соавторами (Fan et al., 2013). Авторы проанализировали три породы, чтобы определить генетические маркеры, влияющие на хромоту и слабость ног. Среди шести наиболее значимых SNP, обнаруженных в породе ландрас, был вариант в гене *SRSF12*.

Следующий по значимости CNV 1:41300464-42881624 перекрывается с геном *MAN1A1*, который кодирует белок, участвующий в гликозилировании белка. На сегодняшний день мало что известно об ассоциациях *MAN1A1* с селекционными признаками свиней. Определено, что *MAN1A1* участвует в путях различных типов биосинтеза N-гликанов. Посредством гликозилирования модулируются клеточные функции, включая клеточную адгезию (Phoomak et al., 2018), и происходит модификация белков, играющих ключевую роль в гомеостатических функциях, таких как воспаление, морфогенез и иммунитет (Hamester et al., 2019).

Дефекты N-гликозилирования белков являются наиболее распространенными причинами генетически обусловленных заболеваний, при которых нарушаются синтез гликанов, их присоединение к гликопротеинам и гликолипидам, а также синтез гликозилфосфатидилинозитола (Иванов и др., 2018). Поражение печени является признаком почти всех нарушений гликозилирования. Болезнь может проявляться гепатомегалией, холестазом или печеночной недостаточностью. Нередки тяжелые психомоторные нарушения, гипотония, судороги, черепно-лицевой дисморфизм и желудочно-кишечные расстройства. В случае с дефектами на ногах у свиней можно предложить, что шишки и наросты на конечностях имеют какую-либо связь с заболеванием печени, а CNV в гене *MAN1A1* играют роль в этом процессе, и их можно рассматривать в качестве маркеров предрасположенности.

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 варианты,

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. Это  $Mg^{2+}$ -зависимые ферменты фосфатазы фосфатидной кислоты, которые выполняют ключевую реакцию в биосинтезе глицеролипидов (Csaki et al., 2010). *LPIN1* был идентифицирован во время позиционного клонирования мутантного гена, лежащего в основе липодистрофии у мышей с жировой дистрофией печени (fld) (Shlien, Malkin, 2009). *LPIN1* экспрессируется в скелетных мышцах и играет ключевую роль в функционировании мышц (Donkor et al., 2007).

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. Липин выступает в роли транскрипционного фактора, изменяя экспрессию генов, участвующих в окислении жирных кислот и митохондриальном метаболизме. Такая двойственная функция делает липин уникальным белком, без корректной работы которого невозможно функционирование адипоцитов и гепатоцитов. Исследования показали, что *LPIN1* задействован в синтезе молочного жира у коз (Shi et al., 2013), крупного рогатого скота (Bionaz et al., 2008) и буйволов (Zhou et al., 2022). Нарушение экспрессии гена вызывает задержки в развитии у насекомых и повышение риска возникновения хронических заболеваний у млекопитающих (Reue, 2009). Дефицит липина у человека может лежать в основе таких болезней, как атеросклероз, колит, раковые опухоли и заболевание печени. В мышечной печени дефицит *LPIN1* препятствует нормальному накоплению липидов и индуцирует экспрессию ключевых адипогенных генов. Данных, подтверждающих связь *LPIN1* с какими-либо признаками или заболеваниями у свиней, в открытых источниках нет. Однако его роль в метаболизме глицеролипидов и глицерофосфолипидов, синтезе

жиров и сигнальном пути фосфолипазы D дает все основания предположить, что LPIN1 может быть связан с дефектами конечностей у свиней.

По результатам ассоциативного анализа между CNV и процентом мясности определены 10 CNV (рисунок 65), уровень значимости которых был меньше или равен 0,05 без поправки на множественное тестирование. При использовании FDR остался только один CNV (SSC10:3605230), соответствующий уровню статической значимости (рисунок 66).

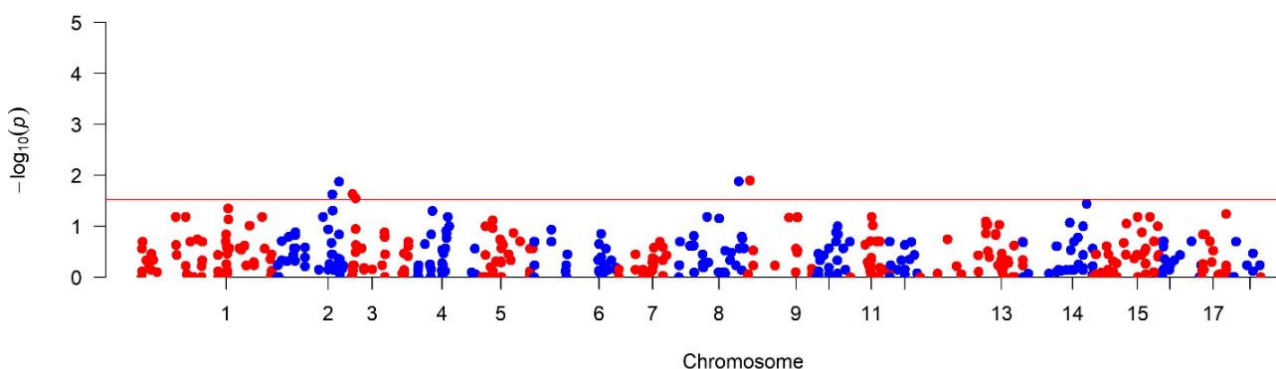


Рисунок 65 – Полногеномные ассоциации CNV с процентом мясности (без поправки на FDR)

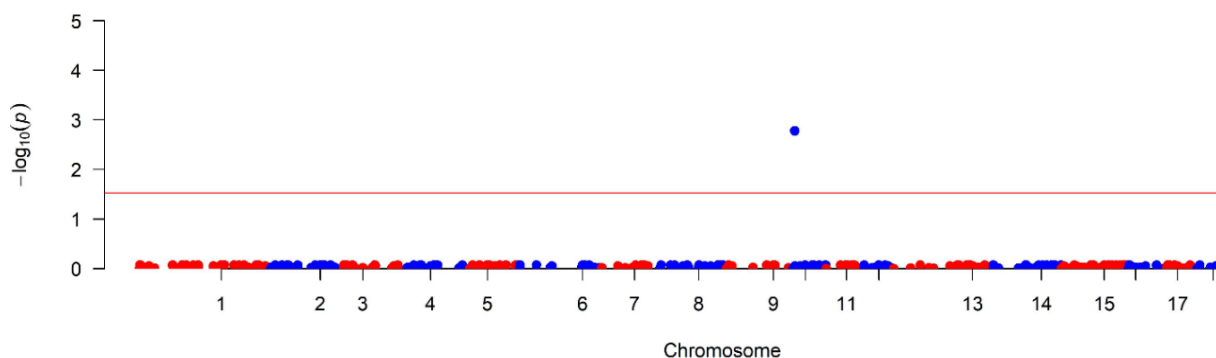


Рисунок 66 – Полногеномные ассоциации CNV с процентом мясности (FDR)

По результатам анализа толщины шпика определены 32 CNV, уровень значимости которых был меньше или равен 0,05 без поправки на множественное тестирование (рисунок 67). Варианты расчетов, где использовались поправки на множественное тестирование методом FDR, не показали значимых CNV, ассоциированных с толщиной шпика (рисунок 68).

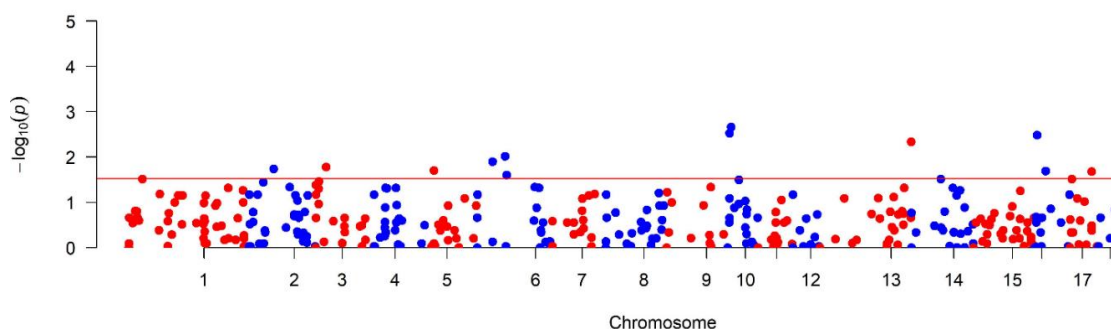


Рисунок 67 – Полногеномные ассоциации CNV с толщиной шпика (без поправки на FDR)

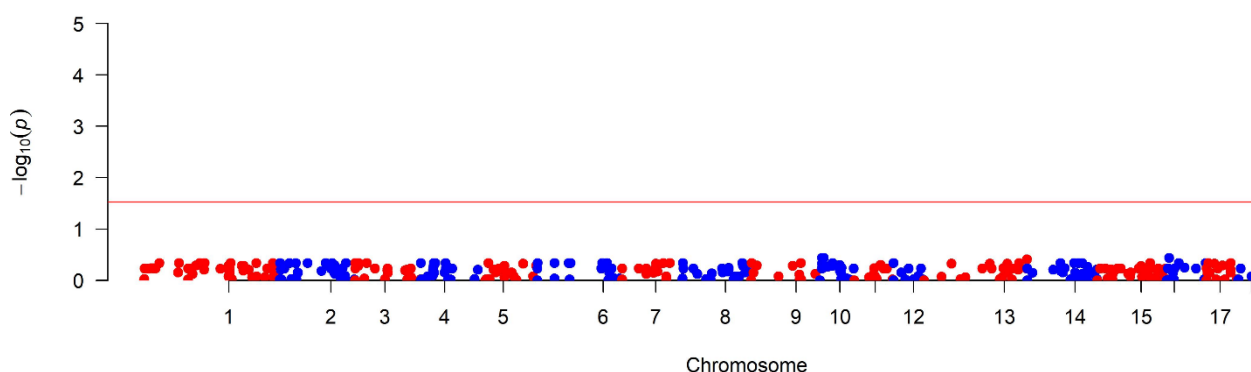


Рисунок 68 – Полногеномные ассоциации CNV с толщиной шпика (FDR)

По результатам анализа между CNV и скороспелостью определены 11 CNV при уровне значимости меньше или равно 0,05 без поправки на FDR (рисунок 69). Варианты расчетов, где использовались поправки на множественное тестирование, показали только один CNV (SSC10:3605230), также ранее показавший связь и с процентом мясности (рисунок 70).

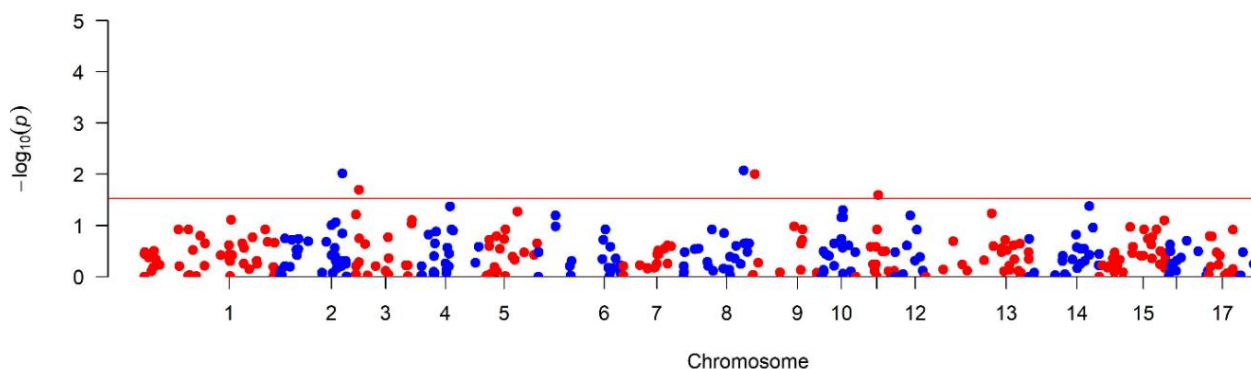


Рисунок 69 – Полногеномные ассоциации CNV со скороспелостью (без поправки на FDR)

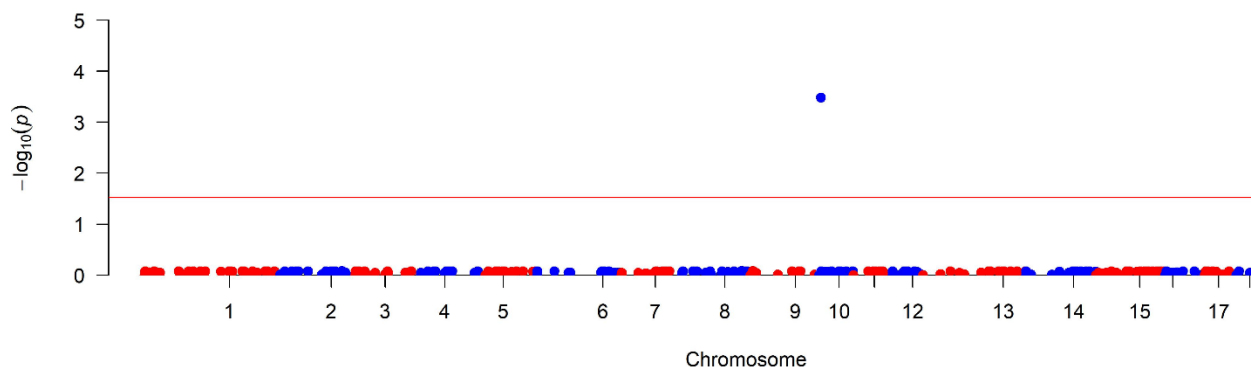


Рисунок 70 – Полногеномные ассоциации CNV со скороспелостью (FDR)

CNV (SSC10:3605230) перекрывается с геном *BRINP3*, кодирующим костный морфогенетический белок/ретиноевой кислоты. *BRINP3* связан с несколькими клеточными функциями, такими как пролиферация, миграция и запрограммированная гибель клеток. По результатам полногеномных ассоциативных исследований была отмечена связь *BRINP3* с качеством мяса в популяции коммерческих свиней, полученных от трехпородных скрещиваний (Gao et al., 2021), толщину подкожного жира, процент мышечной массы и скорость роста у хряков-производителей линии Maxgro (Ирландия) (Reyer et al., 2017). Кроме того, в ряде работ отмечено, что в ген *BRINP3* был подвержен положительному отбору у свиней и лошадей (Asadollahpour et al., 2020), вероятно, в результате целенаправленной селекционной работы, в том числе связанной с интенсификацией процессов роста и развития сельскохозяйственных животных. Все это свидетельствует о том, что ген *BRINP3* и CNV, перекрывающий этот ген, требуют дальнейших исследований, т.к. могут быть перспективными генетическими маркерами для продуктивных признаков свиней.

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 *LPIN1*) могут выступать как



генетические маркеры предрасположенности к дефектам конечностей свиней. CNV локализованы (либо близлежащие от них) в генах, ответственных за метаболизм аминокислот, жирных кислот, глицеролипидов и глицерофосфолипидов, тем самым связанные с иммунным ответом, функциями печени, содержанием внутримышечного жира и упитанностью животных. Полученные результаты согласуются с ранее полученными результатами (Вакоев et al., 2021), согласно которым предрасположенность к образованию шишек и наростов в области скакательного сустава конечностей может реализовываться посредством генетических вариантов, ассоциированных с функциями печени, почек, гематологическими признаками, восприимчивостью к инфекциям, а также быть связанной с 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 (ROH)**

Анализ гомозиготных участков (ROH) в геноме является ключевым элементом в понимании генетической структуры и эволюции популяций. Различные истории популяции приводят к формированию уникального ландшафта длинных и коротких ROH в геноме. Это предполагает, что геномная мозаика ROH может отражать межвидовую и межпородную

изменчивость. Кроме того, появляется все больше аргументов, подтверждающих связь между ROH и сложными фенотипами, в том числе различными заболеваниями у людей и экстерьерными дефектами у животных. Однако при оценке связи между ROH и сложными фенотипами в качестве переменных ROH используют длину (долю) областей гомозиготности, покрывающую геном (в том числе с учетом классов ROH, характеризующих время наследования данных сегментов от предков), и общее количество сегментов. Существуют подходы на основе ассоциативных анализов, но в этом случае достаточно сложно выбрать анализируемые сегменты, а также необходимы достаточно большие выборки, что создает на сегодня серьезные сложности при исследовании связей между ROH и фенотипами.

Одним из возможных подходов для исследования ROH являются создание карт ROH (визуализация геномной мозаики ROH) 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 ROH 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.

Таким образом, опираясь на уникальность мозаики ROH в геноме и инструментарий глубокого обучения в области анализа изображений, разработан алгоритм для создания карт областей гомозиготности, позволяющий обнаружить и оценить уникальную архитектуру ROH как для отдельной особи, так и для определенной популяции. Предлагаемая стратегия состоит из трех основных этапов: обнаружение ROH, их визуализация с

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

Для обнаружения ROH 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 ROH, 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 ROH. Идентификация ROH посредством plink является одним из наиболее широко используемых инструментов для геномного анализа, т.к. быстро и достаточно эффективно позволяет обнаруживать ROH и обрабатывает 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 этим для исследования демографической истории и генетического разнообразия лучше использовать RZooROH.

В случае обнаружения ROH с целью построения карты гомозиготности более предпочтителен plink, в силу своей простоты использования и скорости. В программе предусмотрены пороговые значения, настройки которых крайне важны. Для поиска ROH у свиней использовали данные генома, полученные на основе генотипирования с помощью чипов. Эти данные прошли базовый контроль качества. Вопрос относительно сокращения данных полногеномного

генотипирования по LD решали в процессе тестирования параметров идентификации ROH.

Далее уже в самой программе использовали следующие параметры настройки:

- 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 kb, 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 kb) между последовательными гомозиготными SNP в пределах потенциального ROH. 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.

Тестирование влияния сокращения геномных данных по LD (--indep-pairwise) и количества допустимых гетерозиготных вызовов (--homozyg-het и --homozyg-window-het) на поиск ROH провели на основе данных полногеномного генотипирования (65287 SNP) 70 свиней крупной белой породы.

Для оценки влияния ограничений количества гетерозиготных вызовов в гомозиготном участке выбрали значение параметра --homozyg-window-het 5. Это максимально допустимое смягчение данного фильтра по литературным данным. Поскольку используется большое скользящее окно для анализа, несмотря на мягкий фильтр, при поиске будет отсекается большое количество сегментов. Мы также рассмотрели, насколько снижение порога влияет на количество обнаруженных ROH (рисунок 71).

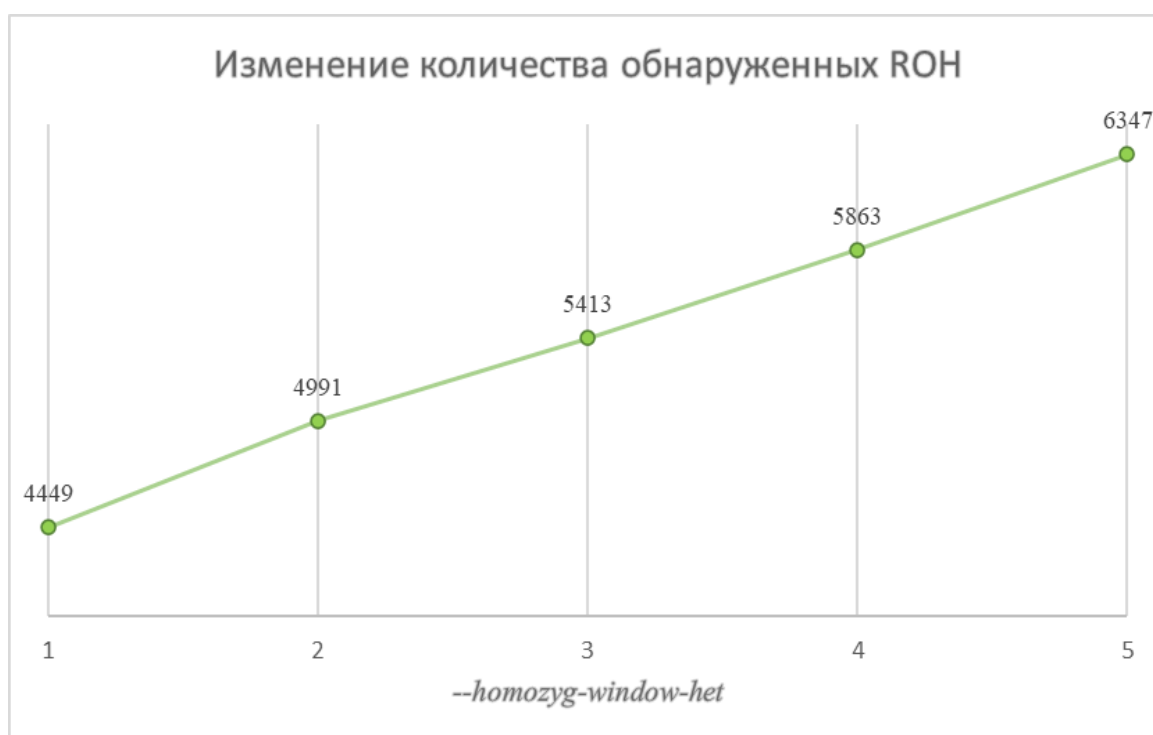


Рисунок 71 – Изменение количества обнаруженных ROH в зависимости от заданного значения параметра `--homozyg-window-het`

С каждым разом количество обнаруженных участков увеличивалось на ~500 значений. Далее данные были протестированы 4 раза, с учетом различных параметров `indep-pairwise`, `homozyg-het` и `homozyg-window-het` 5 (таблица 18).

Таблица 108 – Условия тестирования параметров для поиска сегментов ROH

| Группа | Параметр                                                                                                      | Описание                                                                                                                               |
|--------|---------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| ROH_1  | <code>--homozyg-het 1</code><br><code>--homozyg-window-het 1</code>                                           | Используем строгий порог по количеству возможных гетерозиготных участков в результате                                                  |
| ROH_2  | <code>--homozyg-window-het 5</code>                                                                           | Убираем все ограничения по количеству гетерозигот в ROH и увеличиваем порог допущения таких фрагментов во время сканирования окна до 5 |
| ROH_3  | <code>--indep-pairwise 50 5 0.8</code><br><code>--homozyg-het 1</code><br><code>--homozyg-window-het 1</code> | Обрезаем LD и оставляем строгий фильтр по количеству гетерозиготных вызовов                                                            |

|       |                                                                     |                                                                   |
|-------|---------------------------------------------------------------------|-------------------------------------------------------------------|
|       |                                                                     |                                                                   |
| ROH_4 | -- <i>indep-pairwise</i> 50 5 0.8<br>-- <i>homozyg-window-het</i> 5 | Оставляем LD обрезку, убрав ограничения по количеству гетерозигот |

Влияние выбранных параметров на итоговое количество обнаруженных ROH представлено на рисунке 72.

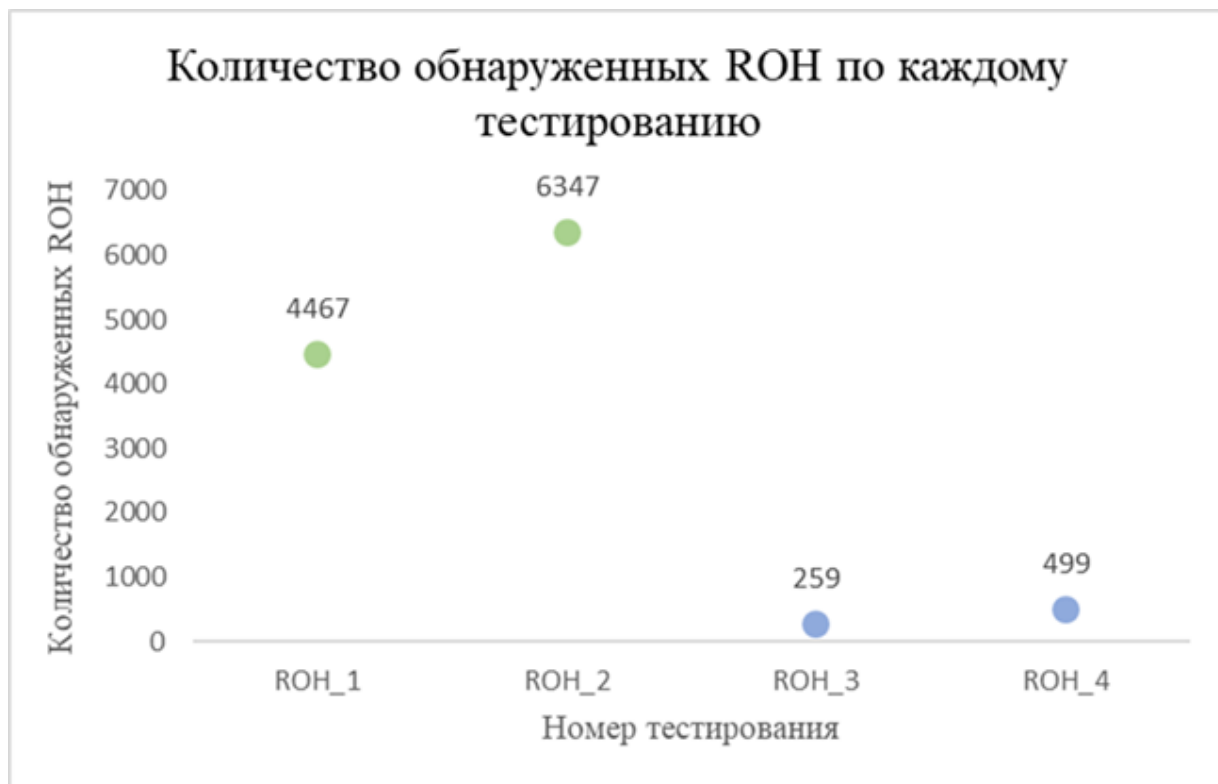


Рисунок 72 – Результаты поиска ROH для четырех тестирований (зеленым цветом обозначены тестирования без обрезки LD, голубым – с обрезкой LD)

Применение параметра по сокращению LD существенно уменьшает количество найденных участков, при сопоставлении соответствующих тестирований ROH\_1-ROH\_3 и ROH\_2-ROH\_4 количество сегментов сокращается примерно на 92–94%. Повышение количества допустимых гетерозиготных вызовов увеличивает итоговый результат в ~1,5–2 раза.

Средние значения длины сегментов гомозиготности между группами не имели сильных различий, тем не менее можно отметить, что максимальные и

минимальные значения варьируются в зависимости от изменений параметров. Исходя из данных таблицы 19, при обрезке LD отсекаются максимально длинные и короткие участки ROH. Также стоит отметить, что в случае отсутствия LD обрезки, ослабление порога гетерозиготных вызовов ведет к определению более длинных участков ROH.

Таблица 119 – Описательная статистика длины сегментов ROH (Mb), определенных в различных группах

| Группа | Кол-во сегментов | Mean    | Std     | Min     | 25%     | 50%     | 75%     | Max      |
|--------|------------------|---------|---------|---------|---------|---------|---------|----------|
| ROH_1  | 4467             | 4448,49 | 2536,73 | 1227,89 | 2794,71 | 3848,49 | 5283,67 | 23913,72 |
| ROH_2  | 6347             | 4612,58 | 2645,56 | 1240,99 | 2874,08 | 3964,14 | 5475,29 | 28253,70 |
| ROH_3  | 259              | 4512,76 | 1943,73 | 1756,64 | 3517,04 | 4217,33 | 4850,56 | 16736,98 |
| ROH_4  | 499              | 4535,35 | 1716,37 | 1554,69 | 3583,70 | 4217,33 | 5024,44 | 16736,98 |

Распределение длин ROH в каждой хромосоме показало аналогичные результаты, у групп ROH\_1 и ROH\_2 длина сегментов была гораздо больше в группах ROH\_1 и ROH\_2 по сравнению с группами ROH\_3 и ROH\_4, в особенности на хромосоме 2, 3, 6, 14 (рисунок 73).

Оценка коэффициента инбридинга в группах ROH\_1, ROH\_2, ROH\_3 и ROH\_4 составила 0,12; 0,18; 0,007 и 0,014 соответственно, что еще раз подтверждает, что сокращение по LD очень сильно влияет на результаты оценки гомозиготности и сильно занижает этот показатель.



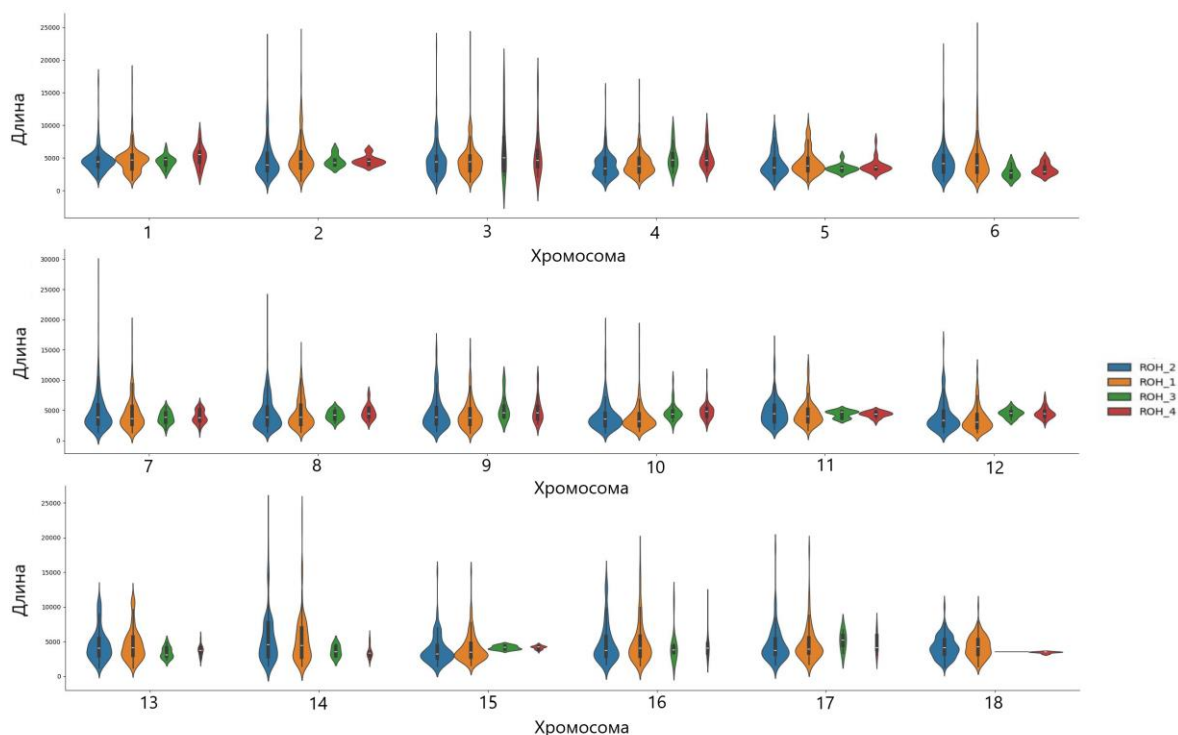


Рисунок 73 – Распределение длин ROH по хромосомам

В целом 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. Основываясь на полученных результатах и обзора литературных данных по вопросам оптимизации параметров при оценке гомозиготности был сделан вывод, что несмотря на то, что обрезка LD снижает вероятность появления случайных (неаутозиготных) сегментов, в случае использования данных полногеномного генотипирования в ней нет необходимости. Целесообразнее применять строгие фильтры при вызове команды `--homozyg`, что будет достаточно для снижения вероятности обнаружения ложноположительных результатов. Соответственно, для идентификации сегментов ROH на основе данных полногеномного генотипирования рекомендуем использовать следующие параметры: «`plink --bfile data --homozyg --homozyg-window-snp 50 --homozyg-snp 50 --homozyg-window-missing 1 --homozyg-kb 1000 --homozyg-density 100 --homozyg-gap 500 --homozyg-window-het 1 --homozyg-het 1 --allow-extra-chr`».

На рисунке 74 показана карта RОН для одного животного. Каждая строка на рисунке соответствует хромосоме, они располагаются сверху вниз, от 1-й до 18-й хромосомы. Сегменты RОН разделены на классы в соответствии с их длиной в Mb и представлены отдельными цветами. Так сегменты менее 2 Mb показаны зеленым, 2–4 Mb – красным, 4–6 Mb – синим, 6–8 Mb – бирюзовым, 8–16 Mb – оранжево-желтым и более 16 Mb – серым.

Карты RОН были сформированы для каждого животного и сохранены в формате «JPEG» со следующими параметрами: разрешение 1200 × 1200, ширина 1200 пикселей, высота 1200 пикселей, горизонтальное разрешение 96 точек на дюйм, разрешений по вертикали 96 точек на дюйм, глубина цвета 24.

Карты были сформированы для свиней породы дюрок (DR,  $n = 500$ ) и свиней крупной белой породы (LW,  $n = 568$ ). Кроме того, для оценки связи между ландшафтом сегментов RОН и наличием опухолей в области скакательного сустава на задних конечностях свиней (далее шишки на задних конечностях) создали две группы свиней без шишек (LW1,  $n = 364$ ) и с шишками (LW2,  $n = 204$ ).

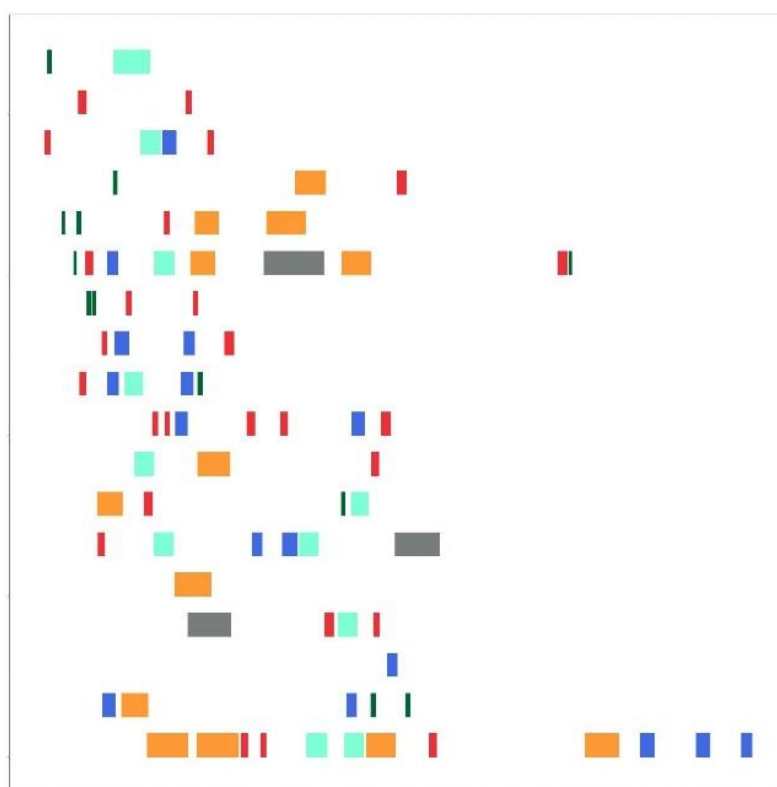


Рисунок 74 – Карта RОН для одного животного

### 3.6.1 Модель межпородной дифференциации

CNN получили широкое распространение для решения задач компьютерного зрения, где в качестве входных данных используют изображения. CNN представляет специализированную форму глубоких нейронных сетей для анализа входных данных, содержащих некоторую форму пространственной структуры. В связи с этим для реализации поставленных задач на основе визуализированных карт ROH мы ставили цель разработать модель для определения межпородной дифференциации свиней коммерческих пород крупная белая и дюрок, а также оценить внутрипородную дифференциацию относительно бинарного фенотипа, применили модель на основе сверточных нейронных сетей (CNN). Основные компоненты этой модели:

- 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 (тренировка модели) и валидации (оценка точности модели);
- развитие модели глубокого обучения – формируются различные слои, такие как сверточные слои и предвыборки, глобально усредняющий слой и несколько плотных слоев. Происходит компиляция модели с использованием функции потерь, оптимизатора с заданным темпом обучения и метрик 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 75;

- 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.

После идентификации сегментов RОН и распределением их на классы в соответствии с их длиной были построены карты RОН каждого животного. Сегменты RОН разделены на классы и, в соответствии с их длиной в Mb, представлены отдельными цветами. Так, сегменты менее 2 Mb даны зеленым, 2–4 Mb – красным, 4–6 Mb – синим, 6–8 Mb – бирюзовым, 8–16 Mb – оранжево-желтым и более 16 Mb – серым.

Далее на основе визуализированных карт RОН была разработана модель для определения межпородной дифференциации свиней пород крупная белая и дюрок.

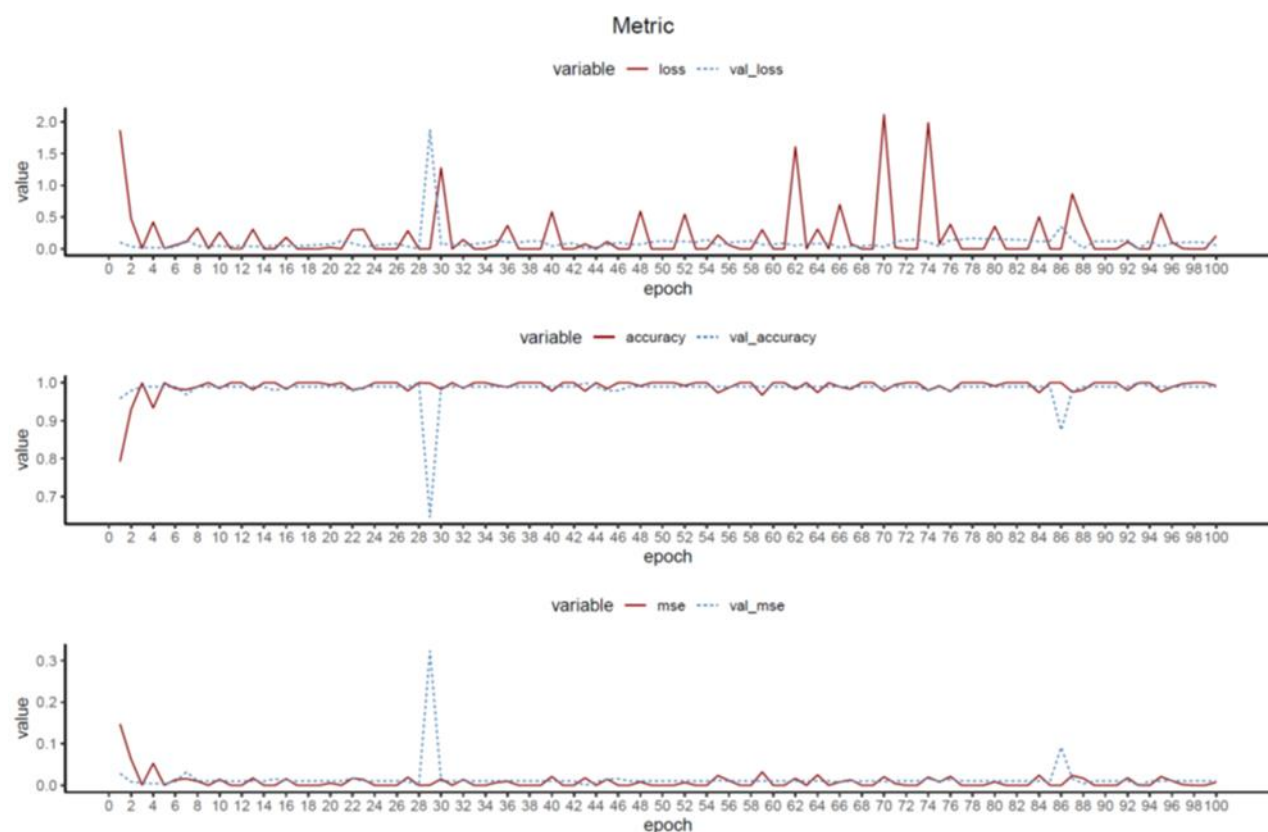


Рисунок 75 – Процесс обучения модели и валидации модели на 100 эпохах

Модель имела все метрики, такие как точность, чувствительность, специфичность, положительная и отрицательная прогностическая ценность,

равные 1,0. Это свидетельствует о безошибочной работе модели и 100%-ному определению породы свиней по геномному ландшафту сегментов ROH (таблица 20).

Таблица 20 – Матрица путанности и метрики модели межпородной дифференциации

| Прогноз      | Наблюдаемые:<br>ROH_of_Duroc | Наблюдаемые:<br>ROH_of_LW |
|--------------|------------------------------|---------------------------|
| ROH_of_Duroc | 10                           | 0                         |
| ROH_of_LW    | 0                            | 10                        |

#### Performance Metrics:

- **Accuracy:** 1 (100%)  
95% Confidence Interval: (0.8316, 1)
- **No Information Rate:** 0.5
- **p-Value [Acc > NIR]:** 9.537e-07
- **Kappa:** 1
- **Mcnemar's Test P-Value:** NA

#### Class-Specific Metrics:

- **Sensitivity (True Positive Rate):** 1.0
- **Specificity (True Negative Rate):** 1.0
- **Positive Predictive Value (Precision):** 1.0
- **Negative Predictive Value:** 1.0
- **Prevalence:** 0.5
- **Detection Rate:** 0.5
- **Detection Prevalence:** 0.5
- **Balanced Accuracy:** 1.

Приведенные метрики свидетельствуют о совершенной работе модели на предоставленных данных. Значение Accuracy (1.0) указывает на то, что все объекты выборки были классифицированы верно. Доверительный интервал (0.8316, 1) демонстрирует, что даже при учете статистической погрешности точность модели остается высокой (нижняя граница >83%). Низкий p-value (9.537e-07) подтверждает, что результат значительно превосходит случайный уровень (No Information Rate = 0.5), характерный для сбалансированных классов (Prevalence = 0.5). Коэффициент Каппа (1) отражает абсолютное

согласие между предсказаниями и истинными метками, что исключает влияние случайного угадывания. Все класс-специфичные метрики (Sensitivity, Specificity, Precision, NPV) равны 1.0, что указывает на отсутствие ложноположительных и ложноотрицательных случаев. Сбалансированная точность (Balanced Accuracy = 1) подчеркивает равномерное качество классификации для обоих классов, что критично при равной распространенности классов (Detection Rate = 0.5).

Кривая ROC (Receiver Operating Characteristic), которая визуализирует компромисс между истинно положительной (TPR) и ложноположительной (FPR) долями классификатора при варьировании порога классификации, также показывает обнадеживающие результаты (рисунок 76). AUC (Area Under Curve), количественно оценивающее способность модели разделять классы: значение 1.0 соответствует идеальному разделению, 0.5 – случайному угадыванию, равно 0,997. Показатель ROC-AUC не зависит от порога классификации и устойчива к дисбалансу классов, что делает ее ключевой метрикой для оценки общей дискриминативной способности модели в бинарных задачах.

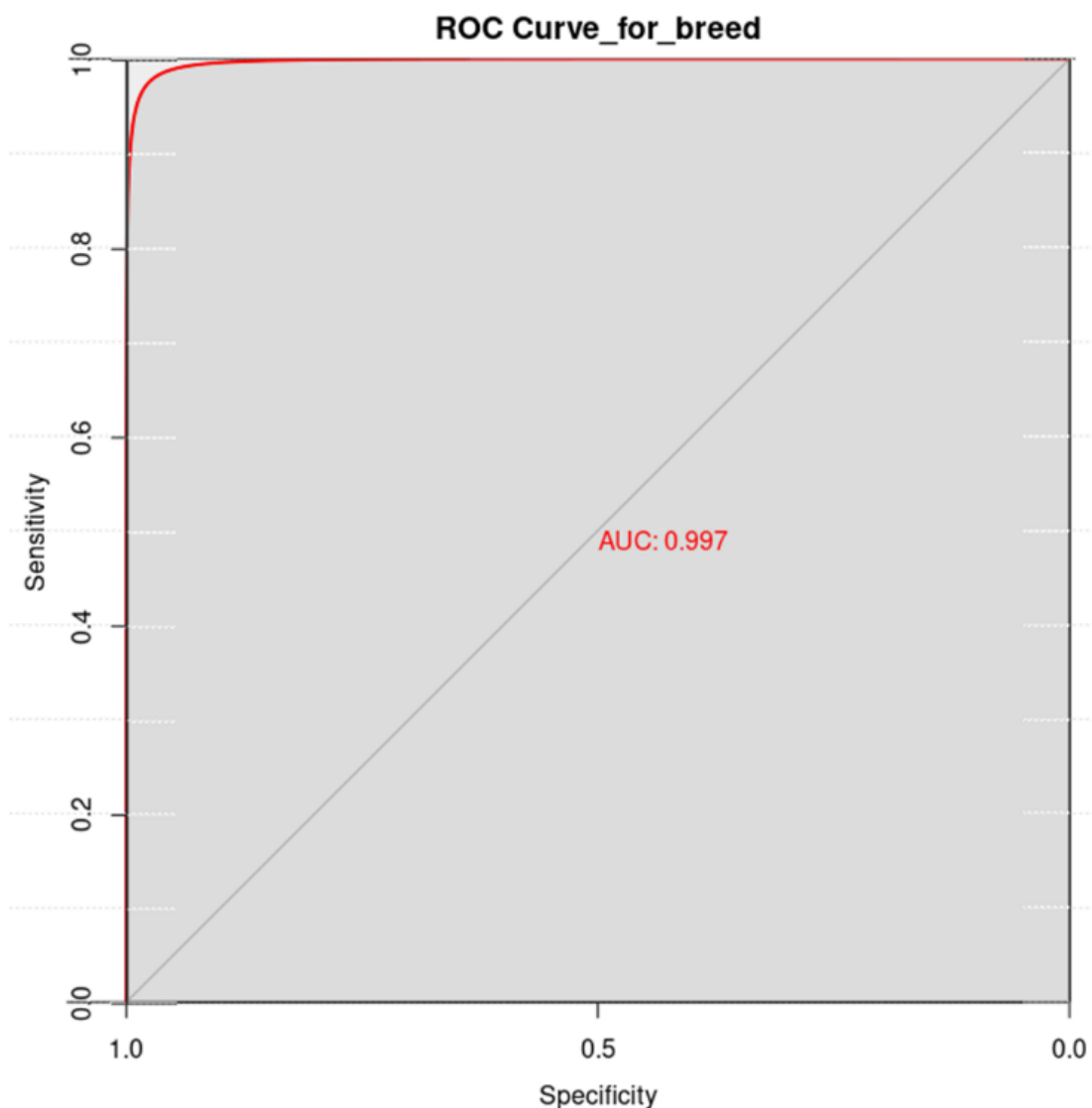


Рисунок 76 – Кривая ROC-AUC для оценки эффективности классификации пород. Площадь под кривой (AUC) составляет 0,997

В результате модель для классификации карт ROH относительно межпородной дифференциации свиней крупной белой породы и дюрок имела все метрики, такие как точность, чувствительность, специфичность, положительная и отрицательная прогностическая ценность, равными 1,0. Это свидетельствует о безошибочной работе модели и 100%-ному определению породы свиней по геномному ландшафту сегментов ROH (таблица 20).

### 3.6.2 Модель дифференциации шишек на ногах свиней

Для дифференциации шишек на ногах свиней породы крупная белая на первом этапе была применена модель дифференциации пород, т.к. входные данные сгенерированы по схожему принципу, но результат оказался неудовлетворительным. Шишки на ногах свиней – это комплексное явление, зависящее от метаболических, иммунных и воспалительных процессов, и эти факторы достаточно тесно взаимосвязаны. На наш взгляд, этот факт является основной причиной такого неудовлетворительного результата. По этой причине мы привлекли инструмент топологического анализа данных. Топологический анализ данных (TDA) представляет собой междисциплинарный подход, основанный на методах алгебраической топологии, направленный на выявление устойчивых геометрических и структурных закономерностей в данных. В отличие от классических методов, фокусирующихся на метрических или статистических свойствах, таких как PCA, MDS и т.д., TDA исследует данные как топологические пространства, выделяя инварианты (такие как компоненты связности, циклы или многомерные полости), которые сохраняются при непрерывных деформациях. Этот подход особенно эффективен для анализа высокомерных и нелинейно организованных данных, где традиционные методы снижения размерности (например, PCA) могут терять существенную информацию. Ключевым инструментом TDA является персистентная гомология, позволяющая количественно оценивать «устойчивость» топологических признаков при изменении масштаба анализа, что обеспечивает многоуровневое представление структуры данных.

Применение TDA в машинном обучении открывает возможности для генерации новых признаков, отражающих глобальную геометрию данных. Построение симплициальных комплексов на основе точечных облаков позволяет выявить кластеры, петли или границы, которые могут быть интерпретированы как дополнительные признаки для классификации или регрессии. Важным преимуществом TDA является его устойчивость к шуму:



кратковременные топологические артефакты, возникающие из-за выбросов, автоматически фильтруются в ходе анализа. Это делает топологические признаки ценным дополнением к традиционным признакам, особенно в задачах, где ключевую роль играет форма данных.

Персистентная гомология – это метод в рамках TDA, систематически фиксирующий рождение и исчезновение топологических признаков (например, циклов, полостей) при постепенном «расширении» пространства данных, задаваемом параметром фильтрации (например, радиусом окрестности точек). Результатом анализа являются персистентные диаграммы или штрих-коды, где каждый элемент соответствует топологическому признаку с координатами (рождение, смерть), отражающими его значимость на разных масштабах. Эти диаграммы кодируют информацию об иерархической структуре данных, что позволяет преобразовывать их в векторные представления (например, персистентные образы или статистики устойчивости) для интеграции в алгоритмы машинного обучения.

Включение персистентных признаков в модели машинного обучения способно повысить их интерпретируемость и точность за счет учета топологической информации. В задачах классификации изображений признаки, описывающие наличие и устойчивость циклов или изолированных компонент, могут дополнить стандартные сверточные признаки, улучшая распознавание форм. На сегодняшний день интеграция таких методов демонстрирует потенциал в областях, где геометрия данных играет критическую роль, – от биоинформатики до анализа финансовых временных рядов.

Использование персистентной гомологии в анализе изображений карт ROH позволяет выявлять скрытые топологические паттерны, которые могут быть связаны с генетической предрасположенностью к образованию шишек на ногах свиней. Участки ROH, представляющие собой длинные гомозиготные сегменты генома, могут содержать гены, ответственные за воспалительные процессы, иммунный ответ или метаболические нарушения,

способствующие развитию патологии. Сверточные нейронные сети фокусируются на локальных текстурах и интенсивностях пикселей, но могут упускать глобальные структурные особенности, такие как кластеризация ROH или их пространственное распределение. Топологический анализ преобразует карты ROH в симплициальные комплексы, выделяя инварианты (например, число связных компонент или циклов), которые отражают организацию генетически значимых регионов. Это особенно важно для выявления сложных взаимосвязей между гомозиготностью и фенотипическими проявлениями.

Персистентную гомологию изображений карт ROH рассчитывали с помощью пакета `ripserr`, который обеспечивает фильтрацию случайных вариаций и выделение устойчивых топологических признаков, сохраняющиеся на разных масштабах анализа. Карты ROH могут варьировать по длине, количеству и распределению гомозиготных участков из-за генетической гетерогенности популяции. Методы машинного обучения, опирающиеся только на интенсивность или форму отдельных участков, могут переобучаться на случайные флуктуации. Топологические дескрипторы, напротив, кодируют информацию об иерархии структур: короткоживущие циклы на персистентной диаграмме соответствуют шуму, а длительные – устойчивым кластерам ROH, которые могут быть ассоциированы с генетическими локусами, ответственными за развитие шишек. Это позволяет модели CNN фокусироваться на биологически значимых паттернах, таких как плотные скопления гомозиготных участков или их специфическое пространственное распределение.

Добавление топологических признаков в CNN обогатило нашу модель информацией о глобальной структуре карт ROH, что важно для анализа сложных генетических данных. Персистентные диаграммы могут выявить наличие кластеров ROH, которые коррелируют с повышенной вероятностью развития шишек, или специфические топологические конфигурации, связанные с наследственными заболеваниями. Такие признаки дополняют стандартные сверточные карты, обеспечивая более полное описание

генетической архитектуры. Кроме того, визуализация персистентных диаграмм позволяет интерпретировать, какие топологические свойства (например, устойчивые циклы или кластеры) ассоциированы с фенотипом, что важно для валидации генетических гипотез и повышения доверия к прогнозам модели. Это играет важную роль в решении нашей задачи, где ключевое значение имеет выявление сложных взаимосвязей между геномом и фенотипом, таких как анализ предрасположенности к появлению шишек у свиней.

Далее на основе визуализированных карт ROH была разработана модель на основе сверточных нейронных сетей для определения наличия/отсутствия дефектов у свиней. Модель продемонстрировала умеренную точность (~76%), что свидетельствует о ее способности различать животных с дефектами и без них на основе карт ROH. Результаты в виде матрицы путанности и метрики модели дифференциации шишек на ногах свиней приведены в таблице 21.

Модель демонстрирует точность (Accuracy) 78.57%, что превышает базовый уровень случайного угадывания (No Information Rate = 64.29%), подтвержденное значимым  $p$ -value (0.03466). Доверительный интервал (63.19–89.7%) указывает на относительно высокую вариабельность оценки, что может быть связано с ограниченным объемом выборки. Коэффициент Каппа (0.54) отражает умеренное согласие между предсказаниями и истинными метками, учитывая дисбаланс классов (Prevalence = 35.71% для положительного класса). Сбалансированная точность (Balanced Accuracy = 77.41%) подчеркивает, что модель сохраняет разумный баланс между Sensitivity (73.33%) и Specificity (81.48%), хотя несколько лучше идентифицирует отрицательные классы.

Таблица 21 – Матрица путанности и метрики модели дифференциации шишек на ногах свиней

| Прогноз         | Наблюдаемые:<br>0 (нет дефекта) | Наблюдаемые:<br>Reference: 1 (с дефектом) |
|-----------------|---------------------------------|-------------------------------------------|
| 0 (Нет дефекта) | 22                              | 4                                         |
| 1 (С дефектом)  | 5                               | 11                                        |

#### Performance Metrics:

- **Accuracy:** 0.7857 (78.57%)  
95% Confidence Interval: (0.6319, 0.897)
- **No Information Rate:** 0.6429
- **P-Value [Acc > NIR]:** 0.03466
- **Kappa:** 0.5401

#### Class-Specific Metrics:

- **Sensitivity (True Positive Rate):** 0.7333 (73.33%)
- **Specificity (True Negative Rate):** 0.8148 (81.48%)
- **Positive Predictive Value (Precision):** 0.6875 (68.75%)
- **Negative Predictive Value:** 0.8462 (84.62%)
- **Prevalence:** 0.3571 (35.71%)
- **Detection Rate:** 0.2619 (26.19%)
- **Detection Prevalence:** 0.3810 (38.10%)
- **Balanced Accuracy:** 0.7741 (77.41%)

Чувствительность и специфичность модели (рисунок 77) 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 (50%) указывает на наличие дополнительных факторов, влияющих на фенотип. При этом высокая отрицательная прогностическая ценность (89%) означает, что модель достаточно уверенно идентифицирует животных без дефектов. Это делает ее полезной для исключения особей, не имеющих предрасположенности к заболеванию. В

целом результаты модели показали, что анализ карт ROH на основе CNN позволяет определять генетическую предрасположенность к появлению шишек/наростов у свиней.

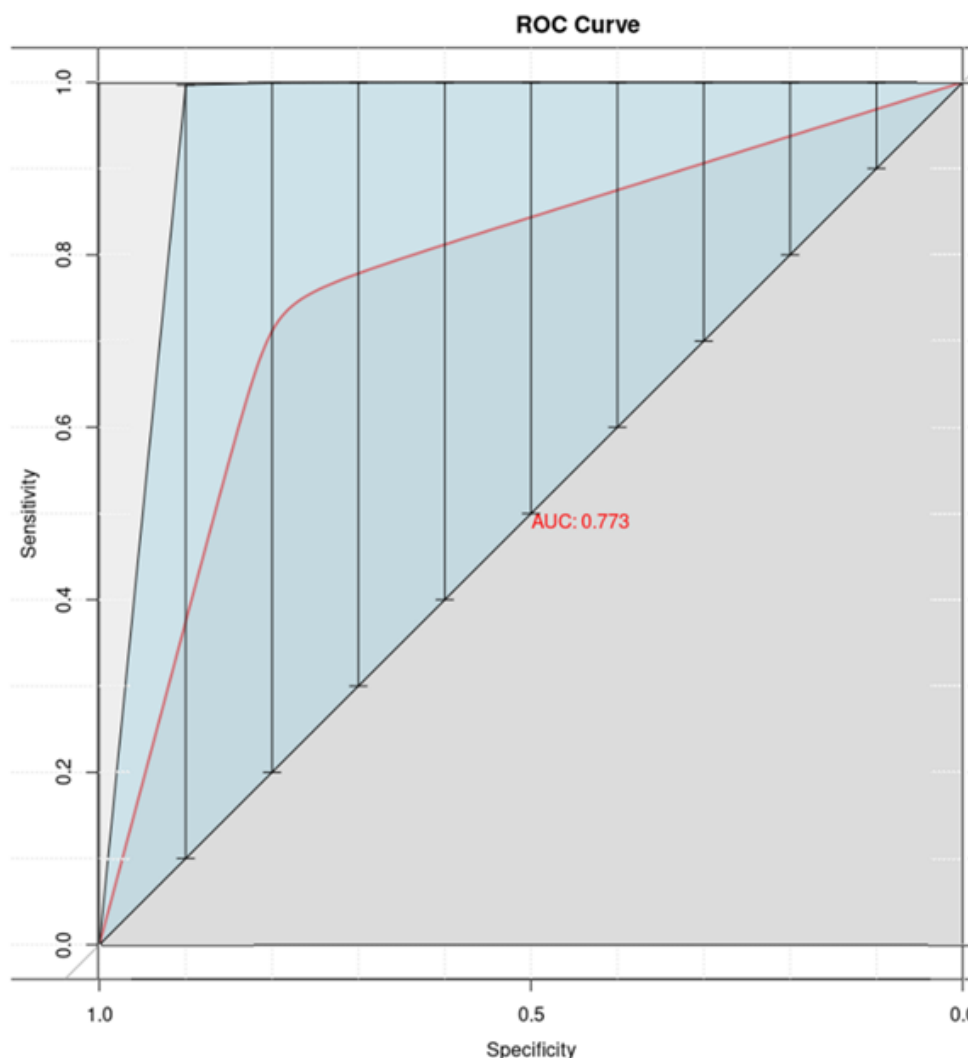


Рисунок 77 – Кривая ROC-AUC для оценки эффективности классификации шишек на ногах свиней. Площадь под кривой (AUC) составляет 0,773

На следующем этапе анализа мы по аналогии с моделями XGBoost и RandomForest определили области важности, т.е. те области, на основе которых модель CNN способна дифференцировать животных с шишками и без шишек.

На рисунке 78 изображена в качестве примера область важности для случайно взятых животных двух классов.

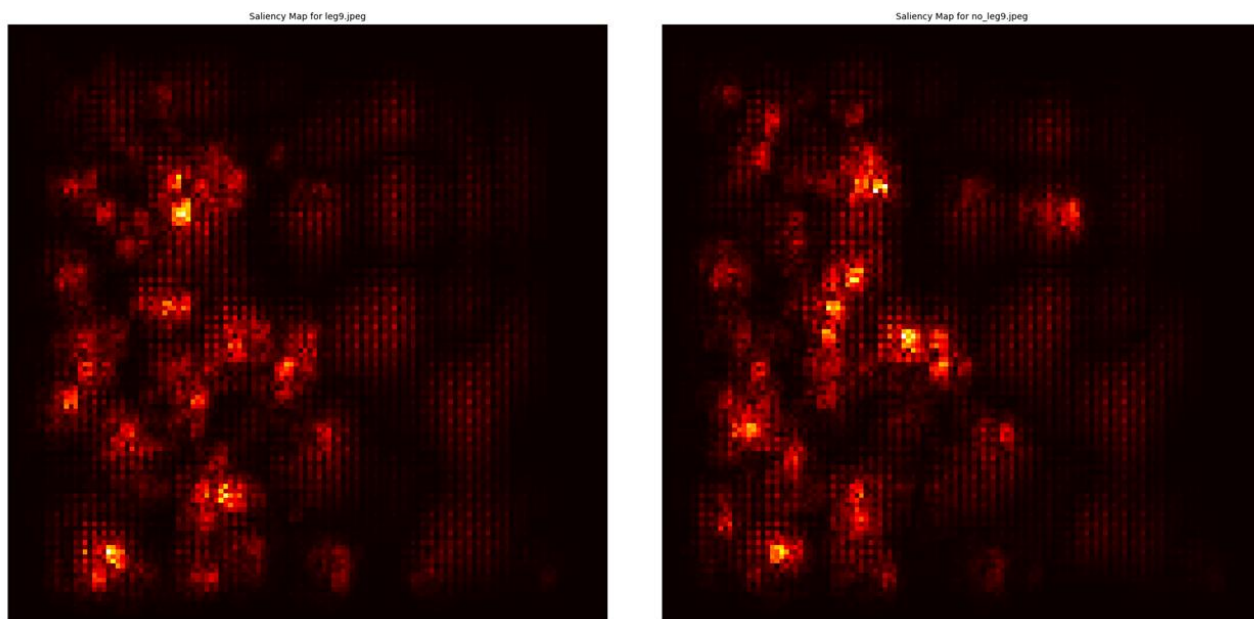


Рисунок 78 – Области важности для двух классов карт ROH классификации шишек на ногах свиней. Области важности выделены более ярким цветом

Далее по всем картам рассчитали усредненные области и для наглядности и более четких контуров перевели в черно-белый формат (рисунок 79).

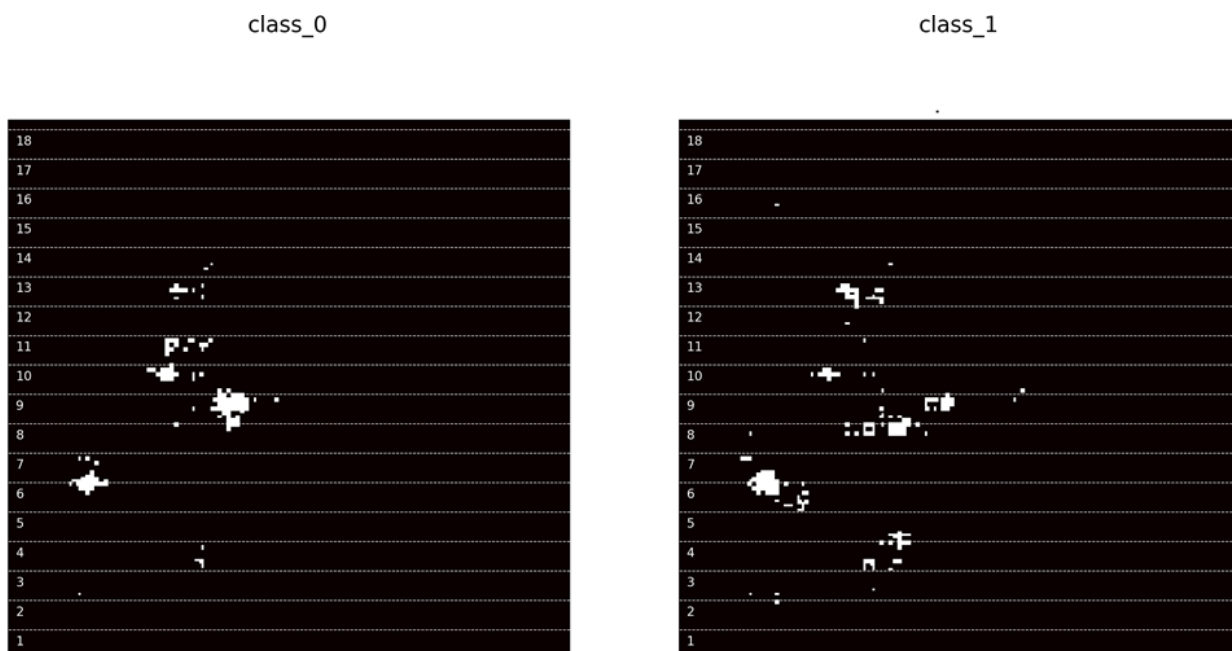


Рисунок 79 – Усредненные области важности для двух классов карт ROH классификации шишек на ногах свиней

На рисунке 79 четко прослеживается разница между двумя классами, причем некоторые области совпадают, а для класса с больными животными выявляется существенно больше, чем для класса с отсутствием заболевания. Данный подход позволяет более глубоко проанализировать разницу в областях гомозиготности между больными и здоровыми животными, а также выявлять гены, ответственные за заболевание.

Таким образом, различные истории популяций приводят к популяционно-специфическим распределениям длинных и коротких RОН. Происхождение этих островов является предметом споров. Предполагают, что они связаны с положительным отбором, однако другое потенциальное биологическое объяснение заключается в том, что острова RОН включают небольшие инверсии, которые подавляют рекомбинацию (Curtis et al., 2008). Тем не менее на сегодняшний день в целом был сделан вывод, что геномная мозаика RОН уникальна между видами и популяциями. Это нашло подтверждение и при использовании классификации карт RОН относительно межпородной дифференциации свиней (породы крупная белая и дюрок).

При анализе RОН важно правильно настроить параметры поиска, такие как количество SNP в гомозиготных участках, минимальная длина сегментов и плотность SNP. Снижение порога для гетерозиготных вызовов увеличивает количество найденных сегментов, однако это может приводить к включению менее точных сегментов в анализ. Использование строгих фильтров, таких как сокращение LD, уменьшает количество ложных сегментов и улучшает результаты, но при этом важно учитывать, что для полногеномного генотипирования такие меры могут быть излишними.

Разработанная модель на основе карт RОН и сверточных нейронных сетей (CNN) продемонстрировала отличные результаты. Все метрики, включая точность, чувствительность, специфичность и прогностическую ценность (положительную и отрицательную), были равны 1,0, что указывает на безошибочную работу модели и на 100% точное определение породы свиней по геномному ландшафту сегментов RОН. Это подтверждает

эффективность использования карты ROH и алгоритма на основе сверточных нейронных сетей (CNN) для успешной межпородной дифференциации.

Для модели, оценивающей генетическую изменчивость, связанную с дефектами конечностей (шишками), метрики показали достаточно хорошие результаты. Точность, чувствительность, специфичность и прогностическая ценность составили  $\sim 0,8$ , что указывает на то, что области гомозиготности могут быть связаны с формированием фенотипа дефектов конечностей. Полученные результаты демонстрируют возможность использования карт ROH для исследования связей между геномом и фенотипами сложных признаков животных, однако модели для этих задач требует дальнейшей настройки и оптимизации. Создание карт ROH с использованием данных полногеномного генотипирования и их дальнейший анализ с помощью глубоких нейронных сетей (CNN) может 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 людей.



## ЗАКЛЮЧЕНИЕ

1. Проведен анализ генетической изменчивости и уровня аутозиготности свиней трех коммерческих пород: крупная белая, ландрас и дюрок, а также диких кабанов. Коэффициенты геномного инбридинга у 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 селекций 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 относилась 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 превзошли другие методы по всем критериям оценки (MAE, MSE, RMSE,  $R^2$ , 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*). В результате проведенных исследований выявлены 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 крупной белой породы. 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 (RON) 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 0,78. Эти результаты демонстрируют эффективность данного подхода и подчеркивают потенциал использования карт RON и глубоких нейронных сетей для анализа селекционно-ценных признаков свиней.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Abondio, P., Cilli, E., Luiselli, D. Inferring Signatures of Positive Selection in Whole-Genome Sequencing Data: An Overview of Haplotype-Based Methods // *Genes (Basel)*. – 2022. – V. 13, № 5. – P. 926. DOI: 10.3390/genes13050926
2. Ahn, J.Y., Kim, I.Y., Oh, S.J., Hwang, H.S., Yi, S.S., Kim, Y.N., Shin, J.H., Yoon, Y.S., Seong, J.K. Proteomic analysis of domestic pig pancreas during development using two-dimensional electrophoresis and matrix-assisted laser desorption/ionization-time of flight mass spectrometry // *Lab Anim Res.* – 2014. – V. 30, № 2. – P. 45–53. DOI: 10.5626/larc.2014.30.2.45
3. Aikawa, S., Hirota, Y., Fukui, Y., Ishizawa, C., Iida, R., Kaku, T., Hirata, T., Akaeda, S., Hiraoka, T., Matsuo, M., Osuga, Y. A gene network of uterine luminal epithelium organizes mouse blastocyst implantation // *Reprod Med Biol.* – 2022. – V. 21, № 1. – P. e12435. DOI: 10.1002/rmb2.12435
4. Alexander, D.H., Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation // *BMC Bioinformatics.* – 2011. – V. 12. – P. 246. DOI: 10.1186/1471-2105-12-246
5. Alexandra, N. Nail, Jeramiah, J. Smith, Martha, L. Peterson, Brett, T. Spear. Evolutionary Analysis of the Zinc Finger and Homeoboxes Family of Proteins Identifies Multiple Conserved Domains and a Common Early Chordate Ancestor // *Genome Biology and Evolution.* – 2020. – V. 12, № 3. – P. 174–184. DOI: 10.1093/gbe/evaa034
6. Al-Mamun, H.A., Kwan, P., Clark, S.A., Ferdosi, M.H., Tellam, R., Gondro, C. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight // *Genet Sel Evol.* – 2015. – V. 47, № 1. – P. 66. DOI: 10.1186/s12711-015-0102-9
7. Alves, A.A.C., da Costa, R.M., Fonseca, L.F.S., Carneiro, R., Ventura, R.V., Rosa, G.J.M., Albuquerque, L.G. A random forest-based genome-wide scan reveals fertility-related candidate genes and potential inter-chromosomal epistatic

regions associated with age at first calving in Nellore cattle // *Frontiers in Genetics*. – 2022. – V. 13. – P. 834724. DOI: 10.3389/fgene.2022.834724

8. Amanda, K. Lindholm-Perry, Virginia, M. Cederberg, Jeremy, R. Miles, Andrew, P. Foote. Expression of cytokine genes and receptors in white blood cells associated with divergent body weight gain in beef steers // *Agri Gene*. – 2017. – V. 6. – P. 37–39. DOI: 10.1016/j.agrigene.2017.05.002

9. Amjad, Rehman, Majid, Harouni, Farzaneh, Zogh, Tanzila, Saba, Mohsen, Karimi, Faten, S. Alamri, Gwanggil, Jeon. Detection of lungs tumors in CT scan images using convolutional neural networks // *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. – 2023. – V. 21, № 4. – P. 769–777. DOI: 10.1109/TCBB.2023.3266695

10. Andolfatto, P. Adaptive evolution of non-coding DNA in *Drosophila* // *Nature*. – 2005. – V. 437. – P. 1149–1152. DOI: 10.1038/nature04228

11. Aoki, T., Ichimura, S., Itoh, A., Kuramoto, M., Shinkawa, T., Isobe, T., et al. Identification of the neuroblastoma-amplified gene product as a component of the syntaxin 18 complex implicated in Golgi-to-endoplasmic reticulum retrograde transport // *Mol Biol Cell*. – 2009. – V. 20, № 11. – P. 2639–2649. DOI: 10.1091/mbc.E09-05-0437

12. Arias, K.D., Lee, H., Bozzi, R., Álvarez, I., Gutiérrez, J.P., Fernández, A., Menéndez, J., Beja-Pereira, A., Goyache, F. Ascertaining the genetic background of the Celtic-Iberian pig strain: A signatures of selection approach // *J. Anim. Breed. Genet.* – 2024. – V. 141, № 1. – P. 96–112. DOI: 10.1111/jbg.12729

13. Asadollahpour, Nanaei, H., Ayatollahpardi, A., Esmailizadeh, A. Whole-genome sequence analysis reveals candidate genomic footprints and genes associated with reproductive traits in Thoroughbred horse // *Reprod Dom Anim*. – 2020. – T. 55. – C. 200–208. DOI: 10.1111/rda.13656

14. Atrian-Afiani, F., Berger, B., Draxl, C., Solkner, J., Meszaros, G. Selective Sweeps in the Austrian Turopolje and Other Commercial Pig Populations // *Animals*. – 2023. – T. 13, № 24. – C. 3749. DOI: 10.3390/ani13243749

15. Auton, A., Bryc, K., Boyko, A.R., et al. Global distribution of genomic diversity underscores rich complex history of continental human populations // *Genome Res.* – 2009. – V. 19. – P. 795–803. DOI: 10.1101/gr.085257.108
16. Bakoev, S., Kolosov, A., Bakoev, F., Romanets, T., Getmantseva, L., Kostyunina, O., Bakoev, N., Koshkina, O. Analysis of homozygous-by-descent (HBD) segments for purebred and crossbred pigs in Russia // *Life.* – 2021. – V. 11, № 8. – P. 861. DOI: 10.3390/life11080861
17. Bakoev, S.Y., Korobeinikova, A.V., Mishina, A.I., Kabieva, S.S., Mitrofanov, S.I., Akinshina, A.A., Snigir, A.I., Yudin, S.M., Yudin, V.S. Genomic signatures of positive selection in human populations of the OXT, OXTR, AVP, AVPR1A and AVPR1B gene variants related to the regulation of psychosocial response // *Genes.* – 2023. – V. 14, № 11. – P. 2053. DOI: 10.3390/genes14112053
18. Bakutenko, I.Y., Haurylchyk, I.D., Nikitchenko, N.V., Sechko, E.V., Kozyro, I.A., Tchitchko, A.M., Batyan, G.M., Sukalo, A.V., Ryabokon, N.I. Neutrophil cytosolic factor 2 (NCF2) gene polymorphism is associated with juvenile-onset systemic lupus erythematosus, but probably not with other autoimmune rheumatic diseases in children // *Molecular Genetics & Genomic Medicine.* – 2022. – V. 10. – P. e1859. DOI: 10.1002/mgg3.1859
19. Barbato M, Orozco-terWengel P, Tapio M and Bruford MW // *Front. Genet.* 2015. – V. 6. – P. 109. DOI: 10.3389/fgene.2015.00109
20. Barroso-Gil, M., Olinger, E., Ramsbottom, S.A., Molinari, E., Miles, C.G., Sayer, J.A. Update of genetic variants in CEP120 and CC2D2A—With an emphasis on genotype-phenotype correlations, tissue specific transcripts and exploring mutation specific exon skipping therapies // *Mol Genet Genomic Med.* – 2021. – V. 9. – P. e1603. DOI: 10.1002/mgg3.1603
21. Barton, N.H. Genetic Hitchhiking // *Phil. Trans. R. Soc. Lond. B.* – 2000. – V. 355. – P. 1553–1562. DOI: 10.1098/rstb.2000.0637
22. Behravan, H., Hartikainen, J.M., Tengstrom, M. et al. Machine learning identifies interacting genetic variants contributing to breast cancer risk: A case study

in Finnish cases and controls // *Sci Rep.* – 2018. – V. 8. – P. 13149. DOI: 10.1038/s41598-018-31243-8

23. Ben Nasr, Barber, F., Elloumi, Oueslati, A. Human exons and introns classification using pre-trained Resnet-50 and GoogLeNet models and 13-layers CNN model // *J Genet Eng Biotechnol.* – 2024. – V. 22, № 1. – P. 100359. DOI: 10.1016/j.jgeb.2024.100359

24. Bentéjac, C., Csörgő, A., Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms // *Artif Intell Rev.* – 2021. – V. 54. – P. 1937–1967. DOI: 10.1007/s10462-020-09884-7

25. Bertolini, F., Servin, B., Talenti, A., Rochat, E., Kim, E.S., Oget, C., Palhière, I., Crisà, A., Catillo, G., Steri, R., Amills, M., Colli, L., Marras, G., Milanesi, M., Nicolazzi, E., Rosen, B.D., Van Tassell, C.P., Guldbrandtsen, B., Sonstegard, T.S., Tosser-Klopp, G., Stella, A., Rothschild, M.F., Joost, S., Crepaldi, P., AdaptMap consortium. Signatures of selection and environmental adaptation across the goat genome post-domestication // *Genet Sel Evol.* 2018. – V. 19, № 1. – C. 57. DOI: 10.1186/s12711-018-0421-y

26. Bionaz, M., Loor, J.J. ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation // *J. Nutr.* – 2008. – V. 138. – P. 1019–1024. DOI: 10.1093/jn/138.6.1019

27. Bjelland, D., Weigel, K., Vukasinovic, N., et al. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding // *J Dairy Science.* – 2013. – V. 96. – P. 4697–4706. DOI: 10.3168/jds.2012-6502

28. Borchers, N., Reinsch, N., Kalm, E. The number of ribs and vertebrae in a Pietrain cross: variation, heritability and effects on performance traits // *J Anim Breed Genet.* – 2004. – V. 121. – P. 392–403. DOI: 10.1111/j.1439-0388.2004.00461.x

29. Bosse, M., Megens, H.J., Madsen, O., et al. Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape

// PLoS Genetics. – 2012. – V. 8. – P. e1003100. DOI: 10.1371/journal.pgen.1003100

30. Botigue, L.R., Song, S., Scheu, A., Gopalan, S., Pendleton, A.L., Oetjens, M., Taravella, A.M., Seregely, T., Zeeb-Lanz, A., Arbogast, R.-M., et al. Ancient European dog genomes reveal continuity since the Early Neolithic // Nat. Commun. – 2017. – V. 8. – P. 16082. DOI: 10.1038/ncomms16082

31. Botta, V., Louppe, G., Geurts, P., Wehenkel, L. Exploiting SNP correlations within random forest for genome-wide association studies // PLoS ONE. – 2014. – V. 9, № 4. – P. e93379. DOI: 10.1371/journal.pone.0093379

32. Bougeard, S., Dray, S. Supervised Multiblock Analysis in R with the ade4 Package // Journal of Statistical Software. – 2018. – V. 86, № 1. – P. 1–17. DOI: 10.18637/jss.v086.i01

33. Bovo, S., Mazzoni, G., Bertolini, F., Schiavo, G., Galimberti, G., Gallo, M., Dall'Olio, S., Fontanesi, L. Genome-wide association studies for 30 haematological and blood clinical-biochemical traits in Large White pigs reveal genomic regions affecting intermediate phenotypes // Sci Rep. – 2019. – V. 9, № 1. – P. 7003. DOI: 10.1038/s41598-019-43491-x

34. Broman, K., Webber, J. Long homozygous chromosomal segments in reference families from the Centre d'Etude du Polymorphisme Humain // American Journal of Human Genetics. – 1999. – V. 65. – P. 1493–1500. DOI: 10.1086/302656

35. Browning, B.L., Tian, X., Zhou, Y., Browning, S.R. Fast two-stage phasing of large-scale sequence data // Am J Hum Genet. – 2021. – V. 108, № 10. – P. 1880–1890. DOI: 10.1016/j.ajhg.2021.08.003

36. Browning, S.R., Browning, B.L. High-resolution detection of identity by descent in unrelated individuals // Am. J. Hum. Genetics. – 2010. – V. 86. – P. 526–539. DOI: 10.1016/j.ajhg.2010.02.013

37. Bruno, Rafael, Pereira, Lopes, Gabriel, Soares, da Silva, Gabriela, de Lima, Menezes, Bruno, Rafael, Pereira, Lopes, et al. Serine proteases in neutrophil extracellular traps exhibit anti-Respiratory Syncytial Virus activity // International



Immunopharmacology. – 2022. – V. 106. – P. 108573. DOI: 10.1016/j.intimp.2022.108573

38. Burrow, H.M. The effects of inbreeding in beef cattle // *Animal Breed Abstract.* – 1993. – V. 61. – P. 737–751.

39. Cai, Z., Sarup, S., Østersen, T., Nielsen, B., Fredholm, M., Sørensen, P., Guldbrandtsen, J., Lund, M.S., Ernst, C.W. Genomic diversity revealed by whole-genome sequencing in three Danish commercial pig breeds // *J. Anim. Sci.* – 2020. – V. 98, № 7. – P. skaa229. DOI: 10.1093/jas/skaa229

40. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., Holmes, S.P. DADA2: High-resolution sample inference from Illumina amplicon data // *Nat. Methods.* – 2016. – V. 13. – P. 581–583. DOI: 10.1038/nmeth.3869

41. Campos-Xavier, A.B., Martinet, D., Bateman, J., Belluoccio, D., Rowley, L., Tan, T.Y., Baxova, A., Gustavson, K.H., Borochowitz, Z.U., Innes, A.M., Unger, S., Beckmann, J.S., Mittaz, L., Ballhausen, D., Superti-Furga, A., Savarirayan, R., Bonafe, L. Mutations in the heparan-sulfate proteoglycan glypican 6 (GPC6) impair endochondral ossification and cause recessive omodysplasia // *Am. J. Hum. Genet.* – 2009. – V. 84, № 6. – P. 760–770. DOI: 10.1016/j.ajhg.2009.04.007

42. Cannan, W.J., Pederson, D.S. Mechanisms and Consequences of Double-Strand DNA Break Formation in Chromatin // *J Cell Physiol.* – 2016. – V. 231, № 1. – P. 3–14. DOI: 10.1002/jcp.25237

43. Canty, J.T., Hensley, A., Aslan, M. et al. TRAK adaptors regulate the recruitment and activation of dynein and kinesin in mitochondrial transport. *Nat Commun* 14, 1376 (2023). <https://doi.org/10.1038/s41467-023-36945-8>

44. Cao, S., Han, J., Wu, J., Li, Q., Liu, S., Zhang, W., Pei, Y., Ruan, X., Liu, Z., Wang, X., Lim, B., Li, N. Specific gene-regulation networks during the pre-implantation development of the pig embryo as revealed by deep sequencing // *BMC Genomics.* – 2014. – V. 15, № 1. – P. 4. DOI: 10.1186/1471-2164-15-4

45. Carletti, M.Z., Christenson, L.K. MicroRNA in the ovary and female reproductive tract // *J Anim Sci.* – 2009. – V. 87, № 14 Suppl. – P. E29–E38. DOI: 10.2527/jas.2008-1692

46. Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M., Wilson, J.F. Runs of homozygosity: windows into population history and trait architecture // *Nature Reviews Genetics*. – 2018. – V. 19. – P. 220. DOI: 10.1038/nrg.2017.164
47. Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M., Wilson, J.F. Runs of homozygosity: windows into population history and trait architecture // *Nature Reviews Genetics*. – 2018. – V. 19, № 4. – P. 220–234. DOI: 10.1038/nrg.2017.164
48. Ceballos, F.C., Kramer, B., Robles, M., The feasibility of picture-based insurance (PBI): Smartphone pictures for affordable crop insurance // *Development Engineering*. – 2019. – V. 4. – P. 100042. DOI: 10.1016/j.deveng.2019.100042
49. Ceccobelli, S. et al. A comprehensive analysis of the genetic diversity and environmental adaptability in worldwide Merino and Merino-derived sheep breeds // *Genet Sel Evol*. – 2023. – V. 55, № 1. – P. 24. DOI: 10.1186/s12711-023-00741-4
50. Charlesworth, B. Effective population size and patterns of molecular evolution and variation // *Nat. Rev. Genet.* – 2009. – V. 10. – P. 195–205. DOI:10.1038/nrg2526
51. Chen, C., Gilbert, C. L., Yang, G., Guo, Y., Segonds-Pichon, A., Ma, J., Evans, G., Brenig, B., Sargent, C., Affara, N., Huang, L. Maternal infanticide in sows: Incidence and behavioural comparisons between savaging and non-savaging sows at parturition // *Applied Animal Behaviour Science*. – 2008. – V. 109. – P. 238–248. DOI: 10.1016/j.applanim.2007.04.006
52. Chen, C., Qiao, R., Wei, R., Guo, Y., Ai, H., Ma, J., Ren, J., Huang, L. A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits // *BMC Genomics*. – 2012. – V. 13. – P. 733. DOI: 10.1186/1471-2164-13-733
53. Chen, E., Tang, M.K., Yao, Y., Yau, W.W., Lo, L.M., Yang, X., Chui, Y.L., Chan, J., Lee, K.K. Silencing BRE expression in human umbilical cord perivascular (HUCPV) progenitor cells accelerates osteogenic and chondrogenic

differentiation // PLoS ONE. – 2013. – V. 8. – P. e67896. DOI: 10.1371/journal.pone.0067896

54. Chen, G., Kobek-Kjeldager, C., Jensen, L.D., Kristensen, J.K., Kaiser, M., Thodberg, K., Zhang, G., Rong, L. Experimental study on temperature difference between the interior and exterior of the vehicle transporting weaner pigs // Biosystems Engineering. – 2024. – V. 247. – P. 119–131. DOI: 10.1016/j.biosystemseng.2023.11.013

55. Chen, M., Wang, J., Wang, Y., Fu, J., Liu, J-F. Genome-wide detection of selection signatures in Chinese indigenous Laiwu pigs revealed candidate genes regulating fat deposition in muscle // BMC Genet. – 2018. – V. 19, № 1. – P. 31. DOI: 10.1186/s12863-018-0609-9

56. Chen, X., Chen, D.G., Zhao, Z., Zhan, J., Ji, C., Chen, J. Artificial image objects for classification of schizophrenia with GWAS-selected SNVs and convolutional neural network // Patterns (N Y). – 2021. – V. 2, № 8. – P. 100303. DOI: 10.1016/j.patter.2021.100303

57. Chen, X., Sun, X., Chimbaka, I.M., Qin, N., Xu, X., Liswaniso, S., Xu, R., Gonzalez, J.M. Transcriptome Analysis of Ovarian Follicles Reveals Potential Pivotal Genes Associated With Increased and Decreased Rates of Chicken Egg Production // Front Genet. – 2021. – V. 12. – P. 622751. DOI: 10.3389/fgene.2021.622751

58. Chernukha, I.M., Kotenkova, E.A., Fedulova, L.V. SNP-based genetic signatures revealed breeding effects in indigenous Livni compared with Landrace and Large White breeds // Foods and Raw Materials. – 2024. – V. 12, № 2. – P. 283–307. DOI: 10.30720/frm.2024.12.2.283

59. Chowdhury, M.Z.I., Leung, A.A., Walker, R.L. et al. A comparison of machine learning algorithms and traditional regression-based statistical modeling for predicting hypertension incidence in a Canadian population // Sci. Rep. – 2023. – V. 13. – P. 13. DOI: 10.1038/s41598-022-27264-x

60. Chung, C.Y.T., Lo, P.H.Y., Lee, K.K.H. Babam2 Regulates Cell Cycle Progression and Pluripotency in Mouse Embryonic Stem Cells as Revealed by

Induced DNA Damage // *Biomedicines*. – 2020. – V. 8, № 10. – P. 397. DOI: 10.3390/biomedicines8100397

61. Chung, R-H., Chen, Y-E. A two-stage random forest-based pathway analysis method // *PLoS ONE*. – 2012. – V. 7, № 5. – P. e36662. DOI: 10.1371/journal.pone.0036662

62. Ciodaro, T., De Stefano, D., de Seixas, J., Damazio, D. Online particle detection with neural networks based on topological calorimetry information // *J. Phys. Conf. Series*. – 2012. – V. 368. – P. 012030. DOI: 10.1088/1742-6596/368/1/012030

63. Civetta, A., Singh, R.S. Sex-related genes, directional sexual selection, and speciation // *Mol. Biol. Evol.* – 1998. – V. 15. – P. 901–909. DOI: 10.1093/oxfordjournals.molbev.a026048

64. Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., Ragoussis, J. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data // *Nucleic Acids Res.* – 2007. – V. 35, № 6. – P. 2013–2025. DOI: 10.1093/nar/gkm077

65. Conticello, S.G. The AID/APOBEC family of nucleic acid mutators // *Genome Biol.* – 2008. – T. 9, № 6. – P. 229. DOI: 10.1186/gb-2008-9-6-229

66. Costa, P.I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M.J., Bushman, F.D., de Vos, W.M., Ehrlich, S.D., Fraser, C.M., Hattori, M., et al. Enterotypes in the landscape of gut microbial community composition // *Nat. Microbiol.* – 2018. – V. 3. – P. 8–16. DOI: 10.1038/s41564-017-0070-7

67. Csaki, L.S., Reue, K. Multifunctional lipid metabolism proteins // *Annu. Rev. Nutr.* – 2010. – V. 30. – P. 257–272. DOI: 10.1146/annurev.nutr.012809.104729

68. Curik, I., Ferencakovic, M., Solkner, J. Inbreeding and runs of homozygosity: A possible solution to an old problem // *Livestock Science*. – 2014. – V. 166. – P. 26–34. DOI: 10.1016/j.livsci.2014.04.008

69. Danecek, P., Auton, A., Abecasis, G. The variant call format and VCFtools // Bioinformatics. – 2011. – V. 27. – P. 2156–2158. DOI: 10.1093/bioinformatics/btr330
70. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R. M., Li, H. Twelve years of SAMtools and BCFtools // GigaScience. 2021. – T. 10, № 2. DOI: 10.1093/gigascience/giab008
71. Daniel, C. Hackett, Joshua, Isaacson, Shirley, Weishi, Li, Karla, Tame-Narvaez, Michael, L. Wagman. Machine Learning Neutrino-Nucleus Cross Sections // FERMILAB-PUB-24-0948-T.
72. David, O. Enoma, Janet, Bishung, Theresa, Abiodun, Olubanke, Ogunlana, Victor, Chukwudi, Osamor. Machine learning approaches to genome-wide association studies // Journal of King Saud University - Science. – 2022. – V. 34, № 4. – P. 101847. DOI: 10.1016/j.jksus.2022.101847
73. de Castro, T., van Heule, M., Domingues, R.R., Jacob, J.C.F., Daels, P.F., Meyers, S.A., Conley, A.J., Dini, P. Embryo-endometrial interaction associated with the location of the embryo during the mobility phase in mares // Sci. Rep. – 2024. – V. 14, № 1. – P. 3151. DOI: 10.1038/s41598-024-48614-4
74. Deirdre, C.P., Donagh, P., Sinead, M., et al. Runs of homozygosity and population history in cattle // BMS Genetics. – 2012. – V. 13. – P. 70. DOI: 10.1186/1471-2156-13-70
75. Dewage, K.A.K.W., Hasan, R., Rehman, B., Mahmood, S. Enhancing brain tumor detection through custom convolutional neural networks and interpretability-driven analysis // Information. – 2024. – V. 15, № 10. – P. 653. DOI: 10.3390/info15100653
76. Ding, R., Zhuang, Z., Qiu, Y. et al. A composite strategy of genome-wide association study and copy number variation analysis for carcass traits in a Duroc pig population // BMC Genomics. – 2022. – V. 23. – P. 590. DOI: 10.1186/s12864-022-08804-1

77. Djillani, A., Mazella, J., Heurteaux, C., Borsotto, M. Role of TREK-1 in Health and Disease, Focus on the Central Nervous System // *Front. Pharmacol.* – 2019. – V. 10. – P. 379. DOI: 10.3389/fphar.2019.00379
78. Dobrescu, A., Giuffrida, M.V., Tsiftaris, S.A. Leveraging multiple datasets for deep leaf counting // In: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). – IEEE. – P. 2072–2079. DOI: 10.1109/ICCVW.2017.243
79. Doekes, H.P., Bijma, P., Windig, J.J. How Depressing Is Inbreeding? A Meta-Analysis of 30 Years of Research on the Effects of Inbreeding in Livestock // *Genes (Basel)*. 2021. – V. 12, № 6. – P. 926. DOI: 10.3390/genes12060926
80. Donkor, J., Sariahmetoglu, M., Dewald, J., Brindley, D.N., Reue, K. Three mammalian lipins act as phosphatidate phosphatases with distinct tissue expression patterns // *J. Biol. Chem.* – 2007. – V. 282. – P. 3450–3457. DOI: 10.1074/jbc.M610745200
81. Druet, T., Gautier, M. A model-based approach to characterize individual inbreeding at both global and local genomic scales // *Molecular Ecology*. – 2017. – V. 26. – P. 5820–5841. DOI: 10.1111/mec.14368
82. Duckworth, J.E., Holmes, W. Selection for carcass length in Large White pigs // *Anim Prod.* – 1968. – V. 10. – P. 359–372. DOI: 10.1017/S0003356100028967
83. Emerson, R.O., Thomas, J.H. Adaptive evolution in zinc finger transcription factors // *PLoS Genet.* – 2009. – V. 5, № 1. – P. e1000325. DOI: 10.1371/journal.pgen.1000325
84. Erhardt, A. et al. TMEM132D, a new candidate for anxiety phenotypes: evidence from human and mouse studies // *Mol. Psychiatry*. – 2011. – V. 16. – P. 647–663. DOI: 10.1038/mp.2010.109
85. Eusebi, P.G., Sevane, N., Pizarro, M., Dunner, S. Gene expression profiles underlying aggressive behavior in the prefrontal cortex of cattle // *Biorxiv*. – 2020. – DOI: 10.1101/2020.07.09.194704

86. Fahme, P., Ramadan, F., Le, D.T., Thi, K.-O.N., Ghayad, S.E., Hussein, N., Diaz, C., Croset, M., Clezardin, P., Cohen, P.A. The Intricate Interplay between the ZNF217 Oncogene and Epigenetic Processes Shapes Tumor Progression // *Cancers*. – 2022. – V. 14, № 24. – P. 6043. DOI: 10.3390/cancers14246043

87. Fan, S., Hansen, M.E., Lo, Y., et al. Going global by adapting local: A review of recent human adaptation // *Science*. – 2016. – V. 354. – P. 54–59. DOI: 10.1126/science.aaf1080

88. Ferencakovic, M., Hamzic, E., Gredler, B., et al. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations // *J. Animal Breeding Genetics*. – 2013. – V. 130. – P. 286–293. DOI: 10.1111/j.1439-0388.2012.00949.x

89. Ferencakovic, M., Solkner, J., Kaps, M., et al. Genome-wide mapping and estimation of inbreeding depression of semen quality traits in a cattle population // *J Dairy Sci*. – 2017. – V. 100, № 6. – P. 4721–4730. DOI: 10.3168/jds.2016-12534

90. Ferrer-Admetlla, A., Liang, M., Korneliussen, T., Nielsen, R. On detecting incomplete soft or hard selective sweeps using haplotype structure // *Molecular Biology and Evolution*. – 2014. – V. 31. – P. 1275–1291. DOI: 10.1093/molbev/msu117

91. Fishman, L., McIntosh, M. Standard Deviations: The biological bases of transmission ratio distortion // *Annu. Rev. Genet.* – 2019. – V. 53. – P. 347–372. DOI: 10.1146/annurev-genet-120418-025250

92. Flanagan, S. P., Jones, A. G. fsthet: Fst-Heterozygosity Smoothed Quantiles // *CRAN.R-project.org/package=fsthet*. – 2018.

93. Fontanesi, L., Beretti, F., Riggio, V., Gómez, González, E., Dall'Olio, S., Davoli, R., Russo, V., Portolano, B. Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colors // *Cytogenet Genome Res.* – 2009. – V. 126, № 4. – P. 333–347. DOI: 10.1159/000268089

94. Fowler, K.E., Pong-Wong, R., Bauer, J., Clemente, E.J., Reitter, C.P., Affara, N.A., Waite, S., Walling, G.A., Griffin, D.K. Genome wide analysis reveals

single nucleotide polymorphisms associated with fatness and putative novel copy number variants in three pig breeds // BMC Genomics. – 2013. – V. 14. – P. 784. DOI: 10.1186/1471-2164-14-784

95. Frankham, Ballou, R., Ralls, J., Eldridge, K., Dudash, M., Fenster, M., Lacy, C., Sunnucks, R., Paul. Genetic Management of Fragmented Animal and Plant Populations//Oxford. online edn., Oxford Academic. 2017. DOI: 10.1093/oso/9780198783411.001.0001

96. Frantz, L.A., Mullin, V.E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., Linderholm, A., Mattiangeli, V., Teasdale, M.D., Dimopoulos, E.A., et al. Genomic and archaeological evidence suggest a dual origin of domestic dogs // Science. – 2016. – V. 352. – P. 1228–1231. DOI: 10.1126/science.aaf3161

97. Freedman, A.H., Gronau, I., Schweizer, R.M., Ortega-Del Vecchyo, D., Han, E., Silva, P.M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., et al. Genome sequencing highlights the dynamic early history of dogs // PLoS Genet. – 2014. – V. 10. – P. e1004016. DOI: 10.1371/journal.pgen.1004016

98. Frichot, E., Francois, O. LEA: an R package for Landscape and Ecological Association studies // Under Review. – 2014. – URL <http://membres-timc.imag.fr/Olivier.Francois/lea.html>

99. Fuller, Z.L., Nino, E.L., Patch, H.M., et al. Genome-wide analysis of signatures of selection in populations of African honey bees (*Apis mellifera*) using new web-based tools // BMC Genomics. – 2015. – V. 16. – P. 518. DOI: 10.1186/s12864-015-1742-6

100. Gao, G., Gao, N., Li, S., Kuang, W., Zhu, L., Jiang, W., Yu, W., Guo, J., Li, Z., Yang, C., Zhao, Y. Genome-Wide Association Study of Meat Quality Traits in a Three-Way Crossbred Commercial Pig Population // Front. Genet. – 2021. – V. 12. – P. 614087. DOI: 10.3389/fgene.2021.614087

101. Geiger, M., Sanchez-Villagra, M.R. Similar rates of morphological evolution in domesticated and wild pigs and dogs // Front. Zool. – 2018. – V. 15. – P. 23. DOI: 10.1186/s12983-018-0261-9



102. Getmantseva, L., Kolosova, M., Fede, K., Korobeinikova, A., Kolosov, A., Romanets, E., Bakoev, F., Romanets, T., Yudin, V., Keskinov, A., Bakoev, S. Finding predictors of leg defects in pigs using CNV-GWAS // *Genes*. – 2023. – V. 14, № 11. – P. 2054. DOI: 10.3390/genes14112054
103. Ghosal, S., Blystone, D., Singh, A.K. et al. An explainable deep machine vision framework for plant stress phenotyping // *Proc. Natl. Acad. Sci. USA*. – 2018. – V. 115, №. 18. – P. 4613. DOI: 10.1073/pnas.1715975115
104. Gibson, J., Morton, N. E., Collins, A. Extended tracts of homozygosity in outbred human populations // *Human Molecular Genetics*. – 2006. – V. 15, № 5. – P. 789–795. DOI: 10.1093/hmg/ddi493
105. Gill, M., Anderson, R., Hu, H. et al. Machine learning models outperform deep learning models, provide interpretation and facilitate feature selection for soybean trait prediction // *BMC Plant Biol*. – 2022. – V. 22. – P. 180. DOI: 10.1186/s12870-022-03575-5
106. Giuffrida, M.V., Minervini, M., Tsafaris, S.A. Learning to count leaves in rosette plants // In: CVPPP workshop - BMVC (British Machine Vision Association). – The British Machine Vision Association (BMVA). – 2015. – V. 13. – DOI: 10.5244/C.29.CVPPP.1
107. Glassey, B., Civetta, A. Positive Selection at Reproductive ADAM Genes with Potential Intercellular Binding Activity // *Molecular Biology and Evolution*. – 2004. – V. 21, № 5. – P. 851–859. DOI: 10.1093/molbev/msh082
108. Glorot, X., Bordes, A., Bengio, Y. Deep sparse rectifier neural networks // In *Proc. 14th International Conference on Artificial Intelligence and Statistics*. – 2011. – P. 315–323.
109. Goldstein, B.A., Hubbard, A.E., Cutler, A. et al. An application of Random Forests to a genome-wide association dataset: Methodological considerations & new findings // *BMC Genet*. – 2010. – V. 11. – P. 49. DOI: 10.1186/1471-2156-11-49
110. Golozubova, V., Gullberg, H., Matthias, A., et al. Depressed thermogenesis but competent brown adipose tissue recruitment in mice devoid of all

hormone-binding thyroid hormone receptors // *Mol. Endocrinol.* – 2004. – T. 18. – P. 384–401. DOI: 10.1210/mend.2003-0292

111. Grevengoed, T.J., Klett, E.L., Coleman, R.A. Acyl-CoA metabolism and partitioning // *Annu Rev Nutr.* – 2014. – V. 34. – P. 1–30. DOI: 10.1146/annurev-nutr-071813-105652

112. Grinberg, N.F., Orhobor, O.I., King, R.D. An evaluation of machine-learning for predicting phenotype: studies in yeast, rice, and wheat // *Mach. Learn.* – 2020. – V. 109. – P. 251–277. DOI: 10.1007/s10994-019-05848-5

113. Groenen, M.A., Archibald, A.L., Uenishi, H. Analyses of pig genomes provide insight into porcine demography and evolution // *Nature.* – 2012. – V. 491, – № 7424. – P. 393–398. DOI: 10.1038/nature11622

114. Grohmann, C.J., Shull, C.M., Crum, T.E. et al. Analysis of polygenic selection in purebred and crossbred pig genomes using generation proxy selection mapping // *Genet Sel Evol.* – 2023. – V. 55. – P. 62. DOI: 10.1186/s12711-023-00754-2

115. Gu, X., Feng, C., Ma, L., et al. Genome-wide association study of body weight in chicken F2 resource population // *PLoS One.* – 2011. – V. 6. – P. e21872. DOI: 10.1371/journal.pone.0021872

116. Guo, Y., Wu, C., Yuan, Z. et al. Gene-based testing of interactions using XGBoost in genome-wide association studies // *Front. Cell. Dev. Biol.* – 2021. – V. 9. – P. 801113. DOI: 10.3389/fcell.2021.801113

117. Gurgul, A., Szmatoła, T., Topolski, P., et al. The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle // *Journal of Applied Genetics.* – 2016. – V. 57. – P. 527–530.

118. Gusev, A., Lowe, J.K., Stoffel, M., et al. Whole population, genome-wide mapping of hidden relatedness // *Genom. Res.* – 2009. – V. 19. – P. 318–326. DOI: 10.1101/gr.085028.108

119. Haibo, Gao, Rui, Gao, Linfeng, Zhang, Wenchao, Xiu, Ruge, Zang, Hong, Wang, Yong, Zhang, Hong, Wang, Jianyu, Chen, Yong, Zhang, Jiayu, Chen, Yawei, Gao, Shaorong, Gao. Esrrb plays important roles in maintaining self-renewal

of trophoblast stem cells (TSCs) and reprogramming somatic cells to induced TSCs // *Journal of Molecular Cell Biology*. – 2019. – V. 11, № 6. – P. 463–473. DOI: 10.1093/jmcb/mjz024

120. Hamester, F., Legler, K., Wichert, B., Kelle, N., Eylmann, K., Rossberg, M., Ding, Y., Kürti, S., Schmalfeldt, B., Milde-Langosch, K., et al. Prognostic relevance of the Golgi mannosidase MAN1A1 in ovarian cancer: Impact of N-glycosylation on tumor cell aggregation // *Br. J. Cancer*. – 2019. – V. 121. – P. 944–953. DOI: 10.1038/s41416-019-0607-2

121. Hasan, M.M., Chopin, J.P., Laga, H. et al. Detection and analysis of wheat spikes using convolutional neural networks // *Plant Methods*. – 2018. – V. 14, № 1. – P. 100. DOI: 10.1186/s13007-018-0364-5

122. Helmstaedter, M., Briggman, K.L., Turaga, S.C., Jain, V., Seung, H.S., Denk, W. Connectomic reconstruction of the inner plexiform layer in the mouse retina // *Nature*. – 2013. – V. 500, № 7461. – P. 168–174. DOI: 10.1038/nature12346

123. Hidetaka, Tokuno, Tatsuya, Itoga, Jumpei, Kasuga, Kazumi, Hasuko, Hiroaki, Masuyama, Yoshimi, Ben– №. Method for estimating disease risk from microbiome data using structural equation modeling // *Front. Microbiol.* – 2023. – V. 14. – P. 1035002. DOI: 10.3389/fmicb.2023.1035002

124. Hie, B., Zhong, E.D., Berger, B., Bryson, B. Learning the language of viral evolution and escape // *Science*. – 2021. – V. 371. – P. 284–288. DOI: 10.1126/science.abe6487

125. Hilger, A.C., Dworschak, G.C., Reutter, H.M. Lessons Learned from CNV Analysis of Major Birth Defects // *Int. J. Mol. Sci.* – 2020. – V. 21, № 21. – P. 8247. DOI: 10.3390/ijms21218247

126. Hlongwane, N.L., Dzomba, E.F., Hadebe, K., van der Nest, M.A., Pierneef, R., Muchadeyi, F.C. Identification of Signatures of Positive Selection That Have Shaped the Genomic Landscape of South African Pig Populations // *Animals*. – 2024. – V. 14, № 2. – P. 236. DOI: 10.3390/ani14020236

127. Hoffmann, T.C., Glasziou, P.P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D.G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S.E.,

Dixon-Woods, M., McCulloch, P., Wyatt, J.C., Chan, A.W., Michie, S. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide //BMJ. 2014. – V. 348. – P. 1687. DOI: 10.1136/bmj.g1687

128. Holl, J.W., Rohrer, G.A., Shackelford, S.D., Wheeler, T.L., Koohmaraie, M. Estimates of genetic parameters for kyphosis in two crossbred swine populations // J Anim Sci. – 2008. – V. 86. – P. 1765–1769. DOI: 10.2527/jas.2007-0525

129. Hood, R.L., Allen, C.E. Cellularity of porcine adipose tissue: Effects of growth and adiposity // J. Lipid Res. – 1977. – V. 18. – P. 275–284. DOI: 10.1016/S0022-2275(20)41676-1

130. Hood, S.E., Kofler, V.X., Chen, Q., Scott, J., Ortega, J., Lehmann, M. Nuclear translocation ability of lipin differentially affects gene expression and survival in fed and fasting *Drosophila* // J. Lipid Res. – 2020. – V. 61. – P. 1720–1732. DOI: 10.1194/jlr.RA120001051

131. Howard, L., Maciewicz, R.A., Blobel, C.P. Cloning and characterization of ADAM28: evidence for autocatalytic pro-domain removal and for cell surface localization of mature ADAM28 // The Biochemical Journal. – 2000. – V. 348 Pt 1. – P. 21–27. DOI: 10.1042/0264-6021:3480021

132. Howrigan, D., Simonson, M., Keller, M. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms // BMC Genomics. – 2011. – V. 12. – P. 460. DOI: 10.1186/1471-2164-12-460

133. Huang, X., Rymbekova, A., Dolgova, O., Lao, O., Kuhlwilm, M. Harnessing deep learning for population genetic inference // Nat. Rev. Genet. – 2023. – V. 25. – P. 61–78. DOI: 10.1038/s41576-023-00641-3

134. Huber, C.D., DeGiorgio, M., Hellmann, I., Nielsen, R. Detecting recent selective sweeps while controlling for mutation rate and background selection // Mol Ecol. – 2016. – V. 25. – P. 142–156. DOI: 10.1111/mec.13571

135. Hui, Y., Xiong, B., Lee, L.J., Frey, B.J. The human splicing code reveals new insights into the genetic determinants of disease // Science. – 2015. – V. 347. – P. 6218. DOI: 10.1126/science.aaa6852

136. Ivanov, D.O., Novikova, V.P., Pohlebkina, A.A. Congenital disorders of glycosylation // *Pediatr.* – 2018. – V. 9. – P. 5–15. DOI: 10.17816/PED935-15
137. J. C. Whittaker, R. Thompson, and M. C. Denham. Marker-assisted selection using ridge regression // *Ann. Hum. Genet.* – 1999. – V. 63, № 4. – P. 366.
138. Jin, J., Desai, B.N., Navarro, B., Donovan, A., Andrews, N.C., Clapham, D.E. Deletion of *Trpm7* disrupts embryonic development and thymopoiesis without altering  $Mg^{2+}$  homeostasis // *Science.* – 2008. – V. 322, № 5902. – P. 756–760. DOI: 10.1126/science.1163564
139. Johansson, A., Pielberg, G., Andersson, L., Edfors-Lilja, I. Polymorphism at the porcine Dominant white/KIT locus influence coat colour and peripheral blood cell measures // *Animal Genetics.* – 2005. – V. 36. – P. 288–296. DOI: 10.1111/j.1365-2052.2005.01320.x
140. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers // *Bioinformatics.* – 2008. – V. 24. – P. 1403–1405. DOI: 10.1093/bioinformatics/btn312
141. Joseph, Owen, Keith, Dobney, Allowen, Evin, Thomas, Cucchi, Greger, Larson, Una, Strand, Vidarsdottir. The zooarchaeological application of quantifying cranial shape differences in wild boar and domestic pigs (*Sus scrofa*) using 3D geometric morphometrics // *Journal of Archaeological Science.* – 2014. – V. 43. – P. 159–167. DOI: 10.1016/j.jas.2013.12.003
142. Jovasevic, V., Liang, L., Roizman, B. Proteolytic cleavage of VP1-2 is required for release of herpes simplex virus 1 DNA into the nucleus // *J. Virol.* – 2008. – V. 82. – P. 3311–3319. DOI: 10.1128/JVI.01805-07
143. Kaggle. Higgs boson machine learning challenge // Kaggle. – 2014. – <https://www.kaggle.com/c/higgs-boson>
144. Kahindi, R.K., Regassa, A., Htoo, J.K., Nyachoti, C.M. Growth performance and expression of genes encoding enzymes involved in methionine and cysteine metabolism in piglets fed increasing sulphur amino acid to lysine ratio during enterotoxigenic *Escherichia coli* challenge // *Can. J. Anim. Sci.* – 2018. – V. 98. – P. 333–340. DOI: 10.1139/cjas-2017-0027

145. Kamvar, Z.N., Tabima, J.F., Grunwald, N.J. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction // *PeerJ*. – 2014. – V. 2. – P. e281. DOI: 10.7717/peerj.281
146. Kanost, M.R., Jiang, H.B. Clip-domain serine proteases as immune factors in insect hemolymph // *Curr. Opin. in Insect Sci.* – 2015. – V. 11. – P. 47–55. DOI: 10.1016/j.cois.2015.07.001
147. Kaprio, T., Lindstrom, A.M., Rasila, T. et al. Elevated tumor expression of Astroprincin (FAM171A1) is an independent marker of poor prognosis in colon cancer // *BMC Gastroenterol.* – 2021. – V. 21. – P. 341. DOI: 10.1186/s12906-021-03703-8
148. Kashiha, M., C. Bahr, S.A. Haredasht, S. Ott, C.P.H. Moons, T.A. Niewold, F.O. Ödberg, D. Berckmans. Automatic monitoring of pigs water use by cameras // *Comput. Electron. Agric.* – 2013. – V. 90. – P. 164–169. DOI: 10.1016/j.compag.2012.09.003
149. Keller, A., Maus, M., Keller, E., Kerns, K. Deep learning classification method for boar sperm morphology analysis // *Andrology*. – 2024. – V. 12, № 9. – P. 1738. DOI: 10.1111/andr.13855
150. Keller, L.F., Waller, D.M. Inbreeding effects in wild populations // *Trends in Ecology & Evolution*. – 2002. – V. 17. – P. 230–241. DOI: 10.1016/S0169-5347(02)02589-X
151. Kennedy, B.W., Quinton, M., van Arendonk, J.A.M. Estimation of effects of single genes on quantitative traits // *J. Anim. Sci.* – 1992. – V. 70. – P. 2000-2012.
152. Kevin, Yang, Nathaniel, Islas, San, Jewell, Di, Wu, Anupama, Jha, Caleb, M. Radens, Jeffrey, A. Pleiss, Kristen, W. Lynch, Yoseph, Barash, Peter, S. Choi. Machine learning-optimized targeted detection of alternative splicing // *Nucleic Acids Research*. – 2025. – V. 53, № 3. – P. gkae1260. DOI: 10.1093/nar/gkae1260

153. Kim, E.-S., Sonstegard, T.S., Rothschild, M.F. Recent artificial selection in U.S. Jersey cattle impacts autozygosity levels of specific genomic regions // BMC Genomics. – 2015. – V. 16. – P. 302. DOI: 10.1186/s12864-015-1625-8
154. Kim, S., Lee, C. A consistent approach to the genotype encoding problem in a genome-wide association study of continuous phenotypes // PLoS One. – 2020. – V. 15, № 7. – P. e0236139. DOI: 10.1371/journal.pone.0236139
155. Kim, Y., Wojciechowski, R., Sung, H. et al. Evaluation of random forests performance for genome-wide association studies in the presence of interaction effects // BMC Proc. – 2009. – V. 3 Suppl 7. – P. S64. DOI: 10.1186/1753-6561-3-S7-S64
156. King, J.W.B., Roberts, R.C. Carcass length in the bacon pig; its association with vertebrae numbers and prediction from radiographs of the young pig // Anim Prod. – 1960. – V. 2. – P. 59–65. DOI: 10.1017/S0003356100043578
157. Kirin, M., McQuillan, R., Franklin, C.S., et al. Genomic runs of homozygosity record population history and consanguinity // PLoS ONE. – 2010. – V. 5. – P. e13996. DOI: 10.1371/journal.pone.0013996
158. Knap, P.W., Merks, J.W.M. A note on the genetics of aggressiveness of primiparous sows towards their piglets // Livestock Production Science. – 1987. – V. 17. – P. 161–167. DOI: 10.1016/0301-6226(87)90017-9
159. Knapczyk-Stwora, K., Belej, A., Grzesiak, M., Slomczynska, M. Effect of gestational antiandrogen treatment on Dicer1 expression in the porcine fetal gonads // Acta Histochemica. – 2015. – V. 117, № 8. – P. 725–731. DOI: 10.1016/j.acthis.2015.02.003
160. Knockenhauer, K.E., Schwartz, T.U. Structural Characterization of Bardet-Biedl Syndrome 9 Protein (BBS9) // Journal of biological chemistry. – 2015. – V. 290, № 32. – P. 19569–19583. DOI: 10.1074/jbc.M115.651206
161. Koide, Y., Onishi, K., Nishimoto, D., Baruah, A.R., Kanazawa, A., Sano, Y. Sex-independent transmission ratio distortion system responsible for reproductive barriers between Asian and African rice species // New Phytol. – 2008. – V. 179. – P. 888–900. DOI: 10.1111/j.1469-8137.2008.02541.x

162. Kosciuk, T., Price, I.R., Zhang, X. et al. NMT1 and NMT2 are lysine myristoyltransferases regulating the ARF6 GTPase cycle // *Nat Commun.* – 2020. – V. 11. – P. 1067. DOI: 10.1038/s41467-020-14432-8
163. Kostyunina, O., Traspov, A., Bakoev, N., Prytkov, Y., Bardukov, N., Karpushkina, T., Economov, A., Domsy, I., Seryodkin, I., Senchik, A. Genetic diversity, admixture and analysis of homozygous-by-descent (HBD) segments of Russian wild boar // *Biology.* – 2022. – V. 11, № 2. – P. 283–307. DOI: 10.3390/biology11020283
164. Kouji, Sakai, Yasushi, Ami, Maino, Tahara, et al. The Host Protease TMPRSS2 Plays a Major Role in In Vivo Replication of Emerging H7N9 and Seasonal Influenza Viruses // *ASM Journals Journal of Virology.* – 2013. – V. 88, № 10. – P. DOI: 10.1128/jvi.03677-13
165. Kridli, R.T., Khalaj, K., Bidarimath, M., Tayade, C. Placentation, maternal-fetal interface, and conceptus loss in swine // *Theriogenology.* – 2016. – V. 85. – P. 135–144. DOI: 10.1016/j.theriogenology.2015.10.024
166. Kyselova, J., Tichy, L., Sztankoova, Z., Markova, J., Kavanova, K., Beinhauerova, M., Muskova, M. Comparative Characterization of Immune Response in Sheep with Caseous Lymphadenitis through Analysis of the Whole Blood Transcriptome // *Animals (Basel).* – 2023. – V. 13, № 13. – P. 2144. DOI: 10.3390/ani13132144
167. Lan, Hainan, Liu, Huilin, Muhammad, Akram, Khan, Zheng, Xin, Yang, YuJiang, Zhang, Hui, Yao, Naiquan. The basic route of the nuclear translocation porcine growth hormone (GH)-growth hormone receptor (GHR) complex (pGH/GHR) in porcine hepatocytes // *General and Comparative Endocrinology.* – 2018. – V. 266. – P. 101–109. DOI: 10.1016/j.ygcen.2018.07.005
168. Lander, E. S., Botstein, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps // *Genetics.* – 1989. – V. 121, № 1. – P. 185–199.



169. Larson, G. D.Q. Fuller. The evolution of animal domestication // *Annu. Rev. Ecol. Evol. Syst.* – 2014. – V. 45. – P. 115–136. DOI: 10.1146/annurev-ecolsys-110512-135833
170. Lee, C., Hodgins, D., Calvert, J.G., Welch, S.K., Jolie, R., Yoo, D. Mutations within the nuclear localization signal of the porcine reproductive and respiratory syndrome virus nucleocapsid protein attenuate virus replication // *Virology.* – 2006. – V. 346, № 1. – P. 238–250. DOI: 10.1016/j.virol.2005.10.024
171. Lee, E., Jang, J.C., OH, S.H. The current status of Korean native pig production // *J Anim Sci Technol.* – 2023. – V. 65, № 6. – P. 1169–1179. DOI: 10.5187/JAST.2023.65.6.1169
172. Lei, L., Jin, S., Gonzalez, G., Behringer, R.R., Woodruff, T.K. The regulatory role of Dicer in folliculogenesis in mice // *Molecular and Cellular Endocrinology.* – 2010. – V. 315, № 1–2. – P. 63–73. DOI: 10.1016/j.mce.2009.10.007
173. Lencz, T., Lambert, C., Derosse, P., et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia // *Proceedings of the National Academy of Sciences of the United States of America.* – 2007. – V. 104. – P. 19942–19947. DOI: 10.1073/pnas.0704196104
174. Leroy, G. Inbreeding depression in livestock species: review and meta analysis // *Animal Genetics.* – 2014. – V. 45, № 5. – P. 618–628. DOI: 10.1111/age.12192
175. Lessel, D., Gehbauer, C., Bramswig, N.C., Schluth-Bolard, C., Venkataramanappa, S., van Gassen, K.L.I., Hempel, M., Haack, T.B., Baresic, A., et al. BCL11B mutations in patients affected by a neurodevelopmental disorder with reduced type 2 innate lymphoid cells // *Brain.* – 2018. – V. 141. – P. 2299–2311. DOI: 10.1093/brain/awy154
176. Leung, M.K., Xiong, H.Y., Lee, L.J., Frey, B.J. Deep learning of the tissue-regulated splicing code // *Bioinformatics.* – 2014. – V. 30. – P. i121–i129. DOI: 10.1093/bioinformatics/btu378

177. Li, B., Zhang, N., Wang, Y.-G., George, Andrew, W., Reverter, A., Li, Y. Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods // *Front. Genet.* – 2018. – V. 9. – P. 237. DOI: 10.3389/fgene.2018.00237
178. Li, L., Xu, X., Xiao, M., Huang, C., Cao, J., Zhan, S., Guo, J., Zhong, T., Wang, L., Yang, L., et al. The Profiles and Functions of RNA Editing Sites Associated with High-Altitude Adaptation in Goats // *International Journal of Molecular Sciences.* – 2023. – V. 24, № 4. – P. 3115. DOI: 10.3390/ijms24043115
179. Li, L., Yoo, H., Becker, F.F., Ali-Osman, F., Chan, J.Y. Identification of a brain- and reproductive-organs-specific gene responsive to DNA damage and retinoic acid // *Biochem. Biophys. Res. Commun.* – 1995. – V. 206. – P. 764–774. DOI: 10.1006/bbrc.1995.1118
180. Lin, M.-H., Pope, B.D., Sasaki, T., Keeley, D.P., Sherwood, D.R., Miner, J.H. Mammalian hemicentin 1 is assembled into tracks in the extracellular matrix of multiple tissues // *Developmental Dynamics.* – 2020. – V. 249. – P. 775–788. DOI: 10.1002/dvdy.218
181. Lippert, C., Listgarten, J., Liu, Y. et al. FaST linear mixed models for genome-wide association studies // *Nat Methods.* – 2011. – V. 8. – P. 833–835. DOI: 10.1038/nmeth.1681
182. Liu, M., Xu, P., O'brien, T., Shen, S. Multiple roles of Ulk4 in neurogenesis and brain function // *Neurogenesis.* – 2017. – V. 4. – P. 10. DOI: 10.1080/23262133.2017.1367310
183. Liu, P., Liang, Y., Li, L., Lv, X., He, Z., Gu, Y. Identification of selection signatures and candidate genes related to environmental adaptation and economic traits in Tibetan pigs // *Animals.* – 2024. – V. 14, № 4. – P. 654. DOI: 10.3390/ani14040654
184. Liu, X., Fu, Y.X. Exploring population size changes using SNP frequency spectra // *Nature Genetics.* – 2015. – V. 47, № 5. – P. 555–559. DOI: 10.1038/ng.3218

185. Lohmueller, K.E., Albrechtsen, A., Li, Y., et al. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome // *PLoS Genet.* – 2011. – V. 7, № 10. – P. e1002326. DOI: 10.1371/journal.pgen.1002326
186. Long, M., Betran, E., Thornton, K., Wang, W. The origin of new genes: glimpses from the young and old // *Nature Rev. Genet.* – 2003. – V. 4. – P. 865–875. DOI: 10.1038/nrg1187
187. Lossie, A.C., Nakamura, H., Thomas, S.E., Justice, M.J. Mutation of l7Rn3 shows that Odz4 is required for mouse gastrulation // *Genetics.* – 2005. – V. 169. – P. 285–299. DOI: 10.1534/genetics.104.035590
188. Lozada, D.N., Sandhu, K.S., Bhatta, M. Ridge regression and deep learning models for genome-wide selection of complex traits in New Mexican Chile peppers // *BMC Genom Data.* – 2023. – V. 24. – P. 80. DOI: 10.1186/s12863-023-01179-6
189. Lu, J., Wu, T., Zhang, B. et al. Types of nuclear localization signals and mechanisms of protein import into the nucleus // *Cell Commun Signal.* – 2021. – V. 19. – P. 60. DOI: 10.1186/s12964-021-00702-9
190. Ma, J., Sheridan, R.P., Liaw, A., Dahl, G.E., Svetnik, V. Deep neural nets as a method for quantitative structure-activity relationships // *J. Chem. Inf. Model.* – 2015. – V. 55. – P. 263–274. DOI: 10.1021/ci5006468
191. Mackeh, R., Marr, A.K., Fadda, A., Kino, T. C2H2-Type Zinc Finger Proteins: Evolutionarily Old and New Partners of the Nuclear Hormone Receptors // *Nuclear Receptor Signaling.* – 2018. – V. 15. – P. 10. DOI: 10.1016/j.nursig.2018.02.001
192. Magi, A., Tattini, L., Palombo, F., et al. H3M2: detection of runs of homozygosity from whole-exome sequencing data // *Bioinformatics.* – 2014. – V. 30. – P. 2852–2859. DOI: 10.1093/bioinformatics/btu432
193. Majewska, M., Lipka, A., Pauksztó, L., Jastrzebski, J.P., Szeszko, K., Gowkielewicz, M., Lepiarczyk, E., Jozwik, M., Majewski, M.K. Placenta

Transcriptome Profiling in Intrauterine Growth Restriction (IUGR) // *Int J Mol Sci.* – 2019. – V. 20, № 6. – P. 1510. DOI: 10.3390/ijms20061510

194. Manolia, R., Fiacco, Todd, A., Binder, Devin, K. Structure-function relationships of the LRRC8 subunits and subdomains of the volume-regulated anion channel (VRAC) // *Frontiers in Cellular Neuroscience.* – 2022. – V. 16. – P. 962714. DOI: 10.3389/fncel.2022.962714

195. Marchesi, J.A.P., Buzanskas, M.E., Cantao, M.E., et al. Relationship of runs of homozygosity with adaptive and production traits in a paternal broiler line // *Animal.* – 2018. – V. 12, № 6. – P. 1126–1134. DOI: 10.1017/S1751731117002653

196. Marino, D., Perkovic, M., Hain, A., Jaguva, Vasudevan, A.A., Hofmann, H., Hanschmann, K.-M., et al. APOBEC4 Enhances the Replication of HIV-1 // *PLoS ONE.* – 2016. – V. 11, № 6. – P. e0155422. DOI: 10.1371/journal.pone.0155422

197. Marras, G., Gaspa, G., Sorbolini, S., et al. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy // *Animal Genetics.* – 2015. – V. 46, № 2. – P. 110–121. DOI: 10.1111/age.12255

198. Marshall, F. B., Dobney, K., Denham, T., and Capriles, J. M. Evaluating the roles of directed breeding and gene flow in animal domestication // *Proc. Natl. Acad. Sci. U.S.A.* – 2014. – V. 111. – P. 6153–6158. DOI: 10.1073/pnas.1320863111

199. Mathieu, A., Leclercq, M., Sanabria, M., Perin, O., Droit, A. Machine Learning and Deep Learning Applications in Metagenomic Taxonomy and Functional Annotation // *Front. Microbiol.* – 2022. – V. 13. – P. 811495. DOI: 10.3389/fmicb.2022.811495

200. McHugo, M., Talati, P., Armstrong, K., Vandekar, S.N., Blackford, J.U., Woodward, N.D., Heckers, S. Hyperactivity and Reduced Activation of Anterior Hippocampus in Early Psychosis // *Am J Psychiatry.* – 2019. – V. 176, № 12. – P. 1030–1038. DOI: 10.1176/appi.ajp.2019.18091025

201. McQuillan, R., Leutenegger, A., Abdel-Rahman, R., et al. Runs of homozygosity in European populations // *American Journal of Human Genetics*. – 2008. – V. 83. – P. 359–372. DOI: 10.1016/j.ajhg.2008.08.015
202. McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., et al. Runs of homozygosity in European populations // *The American Journal of Human Genetics*. – 2008. – V. 83. – P. 359–372. DOI: 10.1016/j.ajhg.2008.08.015
203. Meester-Smoor, M.A., Vermeij, M., van Helmond, M.J., Molijn, A.C., van Wely, K.H., et al. Targeted disruption of the Mn1 oncogene results in severe defects in development of membranous bones of the cranial skeleton // *Mol. Cell Biol.* – 2005. – V. 25. – P. 4229–4236. DOI: 10.1128/MCB.25.11.4229-4236.2005
204. Mekchay, S., Pothakam, N., Norseeda, W., Supakankul, P., Teltathum, T., Liu, G., Naraballobh, W., Khamlor, T., Sringarm, K., Krutmuang, P. Association of IFNA16 and TNFRSF19 Polymorphisms with Intramuscular Fat Content and Fatty Acid Composition in Pigs // *Biology (Basel)*. – 2022. – V. 11, № 1. – P. 109. DOI: 10.3390/biology11010109
205. Menno, J. de Jong, Joost, F. de Jong, A. Rus, Hoelzel, Axel, Janke, 2021, SambaR: an R package for fast, easy and reproducible population-genetic analyses of biallelic SNP datasets // *Molecular Ecology Resources*. – DOI: 10.1111/1755-0998.13339
206. Meyer, W.K., Arbeithuber, B., Ober, C., Ebner, T., Tiemann-Boege, I., Hudson, R.R., Przeworski, M. Evaluating the evidence for transmission distortion in human pedigrees // *Genetics*. – 2012. – V. 191. – P. 215–232. DOI: 10.1534/genetics.112.140482
207. Mianmian, Yin, Mingrong, Lu, Guidong, Yao, Hui, Tian, Jie, Lian, Lin, Liu, Meng, Liang, Yong, Wang, Fei, Sun. Transactivation of microRNA-224 is involved in transforming growth factor- $\beta$ -mediated mouse granulosa cell proliferation and granulosa cell function by targeting Smad4 // *Mol. Endocrinol.* – 2010. – V. 24. – P. 540–551. DOI: 10.1210/me.2009-0347
208. Mieth, B., Kloft, M., Rodríguez, J.A., Sonnenburg, S., Vobruba, R., Morcillo-Suárez, C., Farré, X., Marigorta, U.M., Fehr, E., Dickhaus, T., Blanchard,

G., Schunk, D., Navarro, A., Müller, K.-R. Combining multiple hypothesis testing with machine learning increases the statistical power of genome-wide association studies // *Sci. Rep.* – 2016. – V. 6. – P. 36671. DOI: 10.1038/srep36671

209. Moeinizada, S., Hu, G., Wang, L. A reinforcement learning approach to resource allocation in genomic selection // *Intell. Syst. Appl.* – 2021. – V. 14. – P. 100076. DOI: 10.1016/j.isystems.2021.100076

210. Montesinos-López, O., Montesinos-López, A., Crossa, J., Gianola, D., Hernández-Suárez, C., Martín-Vallejo, J. Multi-trait, multi-environment deep learning model for genomic-enabled prediction of plant traits // *G3: Genes|Genomes|Genetics.* – 2018. – V. 8. – P. 3829–3840. DOI: 10.1534/g3.118.200728

211. Moon, S., Kim, T.H.Y., Lee, K.T., et al. A genome-wide scan for signatures of directional selection in domesticated pigs // *BMC Genomics.* – 2015. – V. 16. – P. 130. DOI: 10.1186/s12864-015-1392-5

212. Muller, E., Moser, G., Bartenslager, H., Gelderman, H. Trait values of growth, carcass and meat quality in Wild Boar, Meishan and Pietrain pigs as well as their crossbred generations // *J. Anim. Breed. Genet.* – 2000. – V. 117. – P. 189–202. DOI: 10.1046/j.1439-0388.2000.00269.x

213. Murchison, E. P., Hannon, G. J. miRNAs on the move: miRNA biogenesis and the RNAi machinery // *Current Opinion in Cell Biology.* – 2004. – V. 16, № 3. – P. 223–229. DOI: 10.1016/j.ceb.2004.03.003

214. Naik, R.R., Sotnikov, S.V., Diepold, R.P. et al. Polymorphism in Tmem132d regulates expression and anxiety-related behavior through binding of RNA polymerase II complex // *Transl Psychiatry.* – 2018. – V. 8. – P. 1. DOI: 10.1038/s41398-018-0192-9

215. Narasimhan, V., Danecek, P., Scally, A., et al. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from nextgeneration sequencing data // *Bioinformatics.* – 2016. – V. 32. – P. 1749–1751. DOI: 10.1093/bioinformatics/btw158

216. Nasirahmadi, A., Sturm, B., Edwards, S., Jeppsson, K.H., Olsson, A.C., Müller, S., Hensel, O. Deep Learning and Machine Vision Approaches for Posture Detection of Individual Pigs // *Sensors (Basel)*. – 2019. – V. 19, № 17. – P. 3738. DOI: 10.3390/s19173738
217. Nicolini, P., Amorin, R., Han, Y. et al. Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle // *BMC Genet.* – 2018. – V. 19. – P. 14. DOI: 10.1186/s12863-018-0601-0
218. Novembre, J., Han, E. Human Population Structure and the Adaptive Response to Pathogen-Induced Selection Pressures // *Philos. Trans. R Soc. Lond. B Biol. Sci.* – 2012. – V. 367. – P. 878–886. DOI: 10.1098/rstb.2011.0247
219. O'Brien, A.M.P., Utsunomiya, Y.T., Meszaros, G., et al. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle // *Genetics Selection Evolution.* – 2014. – V. 46. – P. 19. DOI: 10.1186/1297-9686-46-19
220. Oczkiewicz, M., Szmatoła, T., Swiatkiewicz, M. Source of Dietary Fat in Pig Diet Affects Adipose Expression of Genes Related to Cancer, Cardiovascular, and Neurodegenerative Diseases // *Genes.* – 2019. – V. 10. – P. 948. DOI: 10.3390/genes10120948
221. Oczkiewicz, M., Szmatoła, T., Swiatkiewicz, M. Source of Dietary Fat in Pig Diet Affects Adipose Expression of Genes Related to Cancer, Cardiovascular, and Neurodegenerative Diseases // *Genes.* – 2019. – V. 10, № 12. – P. 948. DOI: 10.3390/genes10120948
222. Okser, S., Pahikkala, T., Aittokallio, T. Genetic variants and their interactions in disease risk prediction - Machine learning and network perspectives // *BioData Min.* – 2013. – V. 6. – P. 10. DOI: 10.1186/1756-0381-6-10
223. Ovilo, C., Benitez, R., Fernandez, A. et al. Longissimus dorsi transcriptome analysis of purebred and crossbred Iberian pigs differing in muscle characteristics // *BMC Genomics.* – 2014. – V. 15. – P. 413. DOI: 10.1186/1471-2164-15-413

224. Owen, R. Descriptive catalogue of the osteological series contained in the museum of the Royal College of Surgeons of England // London: Royal College of Surgeons. – 1853.

225. P. Zhao, Y. Yu, W. Feng, H. Du, J. Yu, H. Kang, X. Zheng, Z. Wang, G.E. Liu, C.W. Ernst, R. Ran, J. Xu, J. Liu. Evidence of evolutionary history and selective sweeps in the genome of Meishan pig reveals its genetic and phenotypic characterization // *Gigascience*. – 2018. – V. 7, № 5. – P. giy058. DOI: 10.1093/gigascience/giy058

226. Padmanabhan, G. An convolutional neural network of low complexity for tumor anomaly detection // In: Yang, X.S., Sherratt, R.S., Dey, N., Joshi, A. (eds) *Proceedings of Eighth International Congress on Information and Communication Technology*. – ICICT 2023. – Lecture Notes in Networks and Systems. – V. 696. – Springer, Singapore.

227. Palmieri F. The mitochondrial transporter family SLC25: identification, properties and physiopathology. *Mol Aspects Med*. 2013 Apr-Jun;34(2-3):465-84. doi: 10.1016/j.mam.2012.05.005. Epub 2012 Dec 23. PMID: 23266187.

228. Paradis, E., Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R // *Bioinformatics*. – 2018. – V. 35. – P. 526–528. DOI: 10.1093/bioinformatics/bty882

229. Pemberton, T.J., Absher, D., Feldman, M.W., et al. Genomic patterns of homozygosity in worldwide human populations // *Am. J. Hum. Genetics*. – 2012. – V. 91. – P. 275–292. DOI: 10.1016/j.ajhg.2012.07.019

230. Pembleton, L. W., Cogan, N. O. I., Forster, J. W. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations // *Molecular Ecology Resources*. – 2013. – V. 13. – P. 946–952. DOI: 10.1111/1755-0998.12119

231. Peripolli, E., Munari, D.P., Silva, M., et al. Runs of homozygosity: current knowledge and applications in livestock // *Animal Genetics*. – 2017. – V. 48, № 3. – P. 255–271. DOI: 10.1111/age.12533



232. Perrot, A., Tomasov, P., Villard, E., Faludi, R., Melacini, P., Lossie, J., Lohmann, N., Richard, P., De Bortoli, M., Angelini, A., Varga-Szemes, A., Sperling, S.R., Simor, T., Veselka, J., Ozcelik, C., Charron, P. Mutations in NEBL encoding the cardiac Z-disk protein nebulin are associated with various cardiomyopathies // Arch Med Sci. – 2016. – V. 12, № 2. – P. 263–278. DOI: 10.5114/aoms.2016.62279
233. Pertile, S., Silva, F., Salvian, M., et al. Selecao e associacao genomica ampla para o melhoramento genetico animal com uso do metodo ssGBLUP // Pesquisa Agropecuaria Brasileira. – 2016. – V. 51. – P. 1729–1736. DOI: 10.1590/s0100-204x2016001000004
234. Phoomak, C., Silsirivanit, A., Park, D., Sawanyawisuth, K., Vaeteewoottacharn, K., Wongkham, C., Lam, E.W., Pairojkul, C., Lebrilla, C.B., Wongkham, S. O-GlcNAcylation mediates metastasis of cholangiocarcinoma through FOXO3 and MAN1A1 // Oncogene. – 2018. – V. 37. – P. 5648–5665. DOI: 10.1038/s41388-018-0366-1
235. Pippucci, T., Magi, A., Gialluisi, A., et al. Detection of runs of homozygosity from whole exome sequencing data: state of the art and perspectives for clinical, population and epidemiological studies // Hum. Hered. – 2014. – V. 77. – P. 63–72. DOI: 10.1159/000365279
236. Pique-Regi, R., Caceres, A., Gonzalez, J.R. R-Gada: a fast and flexible pipeline for copy number analysis in association studies // BMC Bioinformatics. – 2010. – V. 11. – P. 380. DOI: 10.1186/1471-2105-11-380
237. Poklucar, K., Candek-Potokar, M., Batorek, Lukac, N., Tomazin, U., Skrlep, M. Lipid Deposition and Metabolism in Local and Modern Pig Breeds: A Review // Animals (Basel). – 2020. – V. 10, № 3. – P. 424. DOI: 10.3390/ani10030424
238. Pound, M.P., Atkinson, J.A., Wells, D.M., Pridmore, T.P., French, A.P. Deep learning for multi-task plant phenotyping // bioRxiv. – 2017. – <https://www.biorxiv.org/content/early/2017/10/17/204552>.
239. Prive, F., Luu, K., Vilhjalmsen, B.J., Blum, M.G. Performing highly efficient genome scans for local adaptation with R package pcadapt version 4 //

Molecular Biology and Evolution. – 2020. – V. 37, № 7. – P. 2153–2154. DOI: 10.1093/molbev/msaa076

240. Punwani, D., Zhang, Y., Yu, J., Cowan, M.J., Rana, S., Kwan, A., Adhikari, A.N., Lizama, C.O., Mendelsohn, B.A., Fahl, S.P., et al. Multisystem Anomalies in Severe Combined Immunodeficiency with Mutant BCL11B // N. Engl. J. Med. – 2016. – V. 375. – P. 2165–2176. DOI: 10.1056/NEJMoa1604747

241. Purcell, S., Neale, B., Todd-Brown, K., et al. PLINK: a tool set for whole-genome association and population-based linkage analysis // American Journal of Human Genetics. – 2007. – V. 81. – P. 559-575. DOI: 10.1086/519795

242. Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T.M., Fries, R., Nielsen, R., Simianer, H. Classic selective sweeps revealed by massive sequencing in cattle // PLoS Genet. – 2014. – V. 10, № 2. – P. e1004148. DOI: 10.1371/journal.pgen.1004148

243. Qiao, R., Gao, J., Zhang, Z. et al. Genome-wide association analyses reveal significant loci and strong candidate genes for growth and fatness traits in two pig populations // Genet Sel Evol. – 2015. – V. 47. – P. 17. DOI: 10.1186/s12711-015-0104-9

244. Qiu, Y., Ding, R., Zhuang, Z. et al. Genome-wide detection of CNV regions and their potential association with growth and fatness traits in Duroc pigs // BMC Genomics. – 2021. – V. 22. – P. 332. DOI: 10.1186/s12864-021-07654-7

245. Quilter, C.R., Sargent, C.A., Bauer, J., Bagga, M.R., Reitter, C.P., Hutchinson, E.L., Southwood, O.I., Evans, G., Mileham, A., Griffin, D.K., Affara, N.A. An association and haplotype analysis of porcine maternal infanticide: a model for human puerperal psychosis? // Am J Med Genet B Neuropsychiatr Genet. – 2012. – V. 159B, № 8. – P. 908–927. DOI: 10.1002/ajmg.b.32057

246. Quinlan, L.R. Phosphoinositides, inositol phosphates, and phospholipase C in embryonic stem cells // Methods Mol Biol. – 2006. – V. 329. – P. 127–149. DOI: 10.1385/1-59745-034-X:127

247. R Core Team. R: A language and environment for statistical computing // R Foundation for Statistical Computing, Vienna, Austria. – 2020. – URL <https://www.R-project.org/>.
248. Raffaella, E., Sigala, V., Vasiliki, Lagou, Aleksey, Shmeliov, Samaneh, Kouchaki, Muhammad, Awais, Inga, Prokopenko, Adam, Mahdi, Ayse, Demirkan. Machine Learning to Advance Human Genome-Wide Association Studies // Genes. – 2024. – V. 15, № 1. – P. 34. DOI: 10.3390/genes15010034
249. Ramakrishnan, G., Baakman, C., Heijl, S., Vroling, B., van Horck, R., Hiraki, J., Xue, L.C., Huynen, M.A. Understanding structure-guided variant effect predictions using 3D convolutional neural networks // Front. Mol. Biosci. – 2023. – V. 10. – P. 1204157. DOI: 10.3389/fmolb.2023.1204157
250. Ramey, H.R., Decker, J.E., Mckay, S.D., et al. Detection of selective sweeps in cattle using genome-wide SNP data // BMC Genomics. – 2013. – V. 14. – P. 382. DOI: 10.1186/1471-2164-14-382
251. Ramos, A.M., Pita, R.H., Malek, M., Lopes, P.S., Guimaraes, S.E.F., Rothschild, M.F. Analysis of the mouse High-growth region in pigs // J. Anim. Breed. Genet. – 2009. – V. 126. – P. 404–412. DOI: 10.1111/j.1439-0388.2008.00735.x
252. Rees, Evans, E.T. Investigations on the vertebral column of the Welsh pig // Proc. Br. Soc. Anim. Prod. – 1954. – C. 65–73.
253. Refoyo-Martinez, A., da Fonseca, R.R., Halldorsdottir, K., Arnason, E., Mailund, T., Racimo, F. Identifying loci under positive selection in complex population histories // Genome Res. – 2019. – V. 29, № 9. – P. 1506–1520. DOI: 10.1101/gr.244673.118
254. Reiner, G., Bertsch, N., Hoeltig, D., Selke, M., Willems, H., Gerlach, G.F., Tuemmler, B., Probst, I., Herwig, R., Drungowski, M., Waldmann, K.H. Identification of QTL affecting resistance/susceptibility to acute *Actinobacillus pleuropneumoniae* infection in swine // Mamm Genome. – 2014. – V. 25, № 3–4. – P. 180–191. DOI: 10.1007/s00335-014-9507-x

255. Reue, K. The lipin family: Mutations and metabolism // *Curr. Opin. Lipidol.* – 2009. – V. 20. – P. 165–170. DOI: 10.1097/MOL.0b013e32832adee5
256. Reyer, H., Varley, P.F., Murani, E., Ponsuksili, S., Wimmers, K. Genetics of body fat mass and related traits in a pig population selected for leanness // *Sci Rep.* – 2017. – V. 7, № 1. – P. 9118. DOI: 10.1038/s41598-017-09491-7
257. Risalde, M.A., Rivero-Juarez, A., Romero-Palomo, F., Frias, M., Lopez-Lopez, P., Cano-Terriza, D., et al. Persistence of hepatitis E virus in the liver of non-viremic naturally infected wild boar // *PLoS ONE.* – 2017. – V. 12, № 11. – P. e0186858. DOI: 10.1371/journal.pone.0186858
258. Robinson, P., Turk, D., Jilka, S., Cella, M. Measuring attitudes towards mental health using social media: investigating stigma and trivialisation // *Soc Psychiatry Psychiatr Epidemiol.* 2019. – V. 54, № 1. – P. 51–58. DOI: 10.1007/s00127-018-1571-5
259. Rohrer, G.A., Nonneman, D.J., Wiedmann, R.T. et al. A study of vertebra number in pigs confirms the association of vertnin and reveals additional QTL // *BMC Genet.* – 2015. – V. 16. – P. 129. DOI: 10.1186/s12863-015-0268-1
260. Romanets, E., Bakoev, S., Romanets, T., Kolosova, M., Kolosov, A., Bakoev, F., Tretiakova, O., Usatov, A., Getmantseva, L. Evaluation of genetic differentiation and search for candidate genes for reproductive traits in pigs // *Animal Bioscience.* – 2024. – V. 37, № 5. – P. 832–838. DOI: 10.1080/25290731.2023.2235582
261. Rongrong, D., Ming, Y., Jianping, Q., Shaoyun, L., Zhanwei, Z., Shenping, Z., Enqin, Z., Linjun, H., Zicong, L., Gengyuan, C., Wen, H., Zhenfang, W., Jie, Y. Single-Locus and Multi-Locus Genome-Wide Association Studies for Intramuscular Fat in Duroc Pigs // *Frontiers in Genetics.* – 2019. – V. 10. – P. 619. DOI: 10.3389/fgene.2019.00619
262. Rubin, C.J., Megens, H.J., Martinez, Barrio, A., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, O., Jern, P., Jorgensen, C.B., Archibald, A.L., Fredholm, M., Groenen, M.A., Andersson, L. Strong signatures of selection in

the domestic pig genome // Proceedings of the National Academy of Sciences of the USA. – 2012. – V. 109. – P. 19529–19536. DOI: 10.1073/pnas.1203884109

263. Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E. S. Detecting recent positive selection in the human genome from haplotype structure // Nature. – 2002. – V. 419. – P. 832–837. DOI: 10.1038/nature01140

264. Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S. & The International HapMap Consortium. Genome-wide detection and characterization of positive selection in human populations // Nature. – 2007. – V. 449, № 7164. – P. 913–918. DOI: 10.1038/nature06250

265. Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S. & The International HapMap Consortium. Genome-wide detection and characterization of positive selection in human populations // Nature. – 2007. – V. 449, № 7164. – P. 913–918. DOI: 10.1038/nature06250

266. Saini, T., Chauhan, A., Ahmad, S.F., Kumar, A., Vaishnav, S., Singh, S., Mehrotra, A., Bhushan, B., Gaur, G.K., Dutt, T., Elucidation of population stratifying markers and selective sweeps in crossbred Landlly pig population using genome-wide SNP data // Mamm Genome. – 2024. – V. 35, № 2. – P. 170–185. DOI: 10.1007/s00335-024-9926-3

267. Salek, Ardestani, S., Zandi, M.B., Vahedi, S.M., Janssens, S. Population structure and genomic footprints of selection in five major Iranian horse breeds // Anim Genet. – 2022. – V. 53, № 5. – P. 627–639. DOI: 10.1111/agi.14790

268. Salmon, Hillbertz, N.H., Isaksson, M., Karlsson, E.K., Hellmén, E., Pielberg, G.R., Savolainen, P., Wade, C.M., von Euler, H., Gustafson, U., Hedhammar, A., Nilsson, M., Lindblad-Toh, K., Andersson, L., Andersson, G. Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and

predisposition to dermoid sinus in Ridgeback dogs // *Nat. Genet.* – 2007. – V. 39, – № 11. – P. 1318-1320. DOI: 10.1038/ng.2007.4

269. Samaha, G. Advances in felid genetics and genomics // A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy. – 2021.

270. Sanchez, M., Silos-Santiago, I., Frisen, J. et al. Renal agenesis and the absence of enteric neurons in mice lacking GDNF // *Nature.* – 1996. – V. 382. – P. 70–73. DOI: 10.1038/382070a0

271. Sandhu, K.S., Lozada, D.N., Zhang, Z., Pumphrey, M.O., Carter, A.H. Deep Learning for Predicting Complex Traits in Spring Wheat Breeding Program // *Frontiers in Plant Science.* – 2021. – V. 11. DOI: 10.3389/fpls.2020.602117

272. Sato, S., Hasebe, H., Sato, S., Asahi, Y., Hayashi, T., Kobayashi, E., Sugimoto, Y. High-resolution physical mapping and construction of a porcine contig spanning the intramuscular fat content QTL // *Animal Genetics.* – 2006. – V. 37. – P. 113–120. DOI: 10.1111/j.1365-2052.2006.01469.x

273. Schielke, A., Sachs, K., Lierz, M., Appel, B., Jansen, A., Johne, R. Detection of hepatitis E virus in wild boars of rural and urban regions in Germany and whole genome characterization of an endemic strain // *Virology.* – 2009. – V. 6. – P. 58. DOI: 10.1186/1743-422X-6-58

274. Schulz, K.N., Harrison, M.M. Mechanisms regulating zygotic genome activation // *Nat Rev Genet.* – 2019. – V. 20. – P. 221–234. DOI: 10.1038/s41576-018-0073-9

275. Seabury, C.M., Oldeschulte, D.L., Saatchi, M., Beever, J.E., Decker, J.E., Halley, Y.A., Bhattarai, E.K., Molaei, M., Freetly, H.C., Hansen, S.L., et al. Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle // *BMC Genom.* – 2017. – V. 18. – P. 386. DOI: 10.1186/s12864-017-3780-2

276. Saxton W.M., Hollenbeck P.J. The axonal transport of mitochondria // *J. Cell Sci.* – 2012. – V. 125(Pt 9). – P. 2095–2104. DOI: 10.1242/jcs.053850.

277. Shankara, C., Hariprasad, S.A., Latha, D.U. Detection of lungs tumors in CT scan images using convolutional neural network // SN COMPUT. SCI. – 2023. – V. 4. – P. 225. DOI: 10.1007/s42979-023-01174-6

278. Shao, X., Lv, N., Liao, J., Long, J., Xue, R., Ai, N., Xu, D., Fan, X. Copy number variation is highly correlated with differential gene expression: a pan-cancer study // BMC Med Genet. – 2019. – V. 20. – P. 175. DOI: 10.1186/s12881-019-0909-5

279. Shen, Y., Chen, Y., Zhang, S., Wu, Z., Lu, X., Liu, W., Liu, B., Zhou, X. Smartphone-based digital phenotyping for genome-wide association study of intramuscular fat traits in longissimus dorsi muscle of pigs // Animal Genetics. – 2024. – Jan 30.

280. Shengchen, Gu, Xupeng, Zang, Lei, Jiang, Ting, Gu, Fanming, Meng, Sixiu, Huang, Gengyuan, Cai, Zicong, Li, Zhenfang, Wu, Jie, Enqin, Zheng. Differential MicroRNA Expression in Porcine Endometrium Related to Spontaneous Embryo Loss during Early Pregnancy // Int. J. Mol. Sci. – 2022. – V. 23, № 15. – P. 8157. DOI: 10.3390/ijms23158157

281. Shi, H., Luo, J., Zhu, J., Li, J., Sun, Y., Lin, X., Zhang, L., Yao, D., Shi, H. PPAR $\gamma$  regulates genes involved in triacylglycerol synthesis and secretion in mammary gland epithelial cells of dairy goats // PPAR Res. – 2013. – P. 310948. DOI: 10.1155/2013/310948

282. Shlien, A., Malkin, D. Copy number variations and cancer // Genome Med. – 2009. – V. 1, № 6. – P. 62. DOI: 10.1186/gm62

283. Shlien, A., Malkin, D. Copy number variations and cancer // Genome Med. – 2009. – V. 1, № 6. – P. 62. DOI: 10.1186/gm62

284. Shuai, J., Fu, L., Zhang, X. et al. Functional exchangeability of the nuclear localization signal (NLS) of capsid protein between PCV1 and PCV2 in vitro: Implications for the role of NLS in viral replication // Virol J. – 2011. – V. 8. – P. 341. DOI: 10.1186/1743-422X-8-341

285. Signer-Hasler, H., Burren, A., Ammann, P., Drogemüller, C., Flury, C. Runs of homozygosity and signatures of selection: a comparison among eight local

Swiss sheep breeds // *Anim Genet.* – 2019. – V. 50. – P. 512–525. DOI: 10.1111/agi.14023

286. Silva, J., Giachetto, P., Silva, L., et al. Genomic variants revealed by invariably missing genotypes in Nelore cattle // *PLoS ONE.* – 2015. – V. 10. – P. e0136035. DOI: 10.1371/journal.pone.0136035

287. Silva, V.H.d., Ramos, M., Groenen, M., Crooijmans, R., Johansson, A., Regitano, L., Coutinho, L., Zimmer, R., Waldron, L., Geistlinger, L. CNVRanger: association analysis of CNVs with gene expression and quantitative phenotypes // *Bioinformatics.* – 2020. – V. 36, № 3. – P. 972–973. DOI: 10.1093/bioinformatics/btz866

288. Singh, R.S., Kulathinal, R.J. Sex gene pool evolution and speciation: A new paradigm // *Genes Genet. Sys.* – 2000. – V. 75. – C. 119–130.

289. Sinoquet, C. A method combining a random forest-based technique with the modeling of linkage disequilibrium through latent variables, to run multilocus genome-wide association studies // *BMC Bioinformatics.* – 2018. – V. 19. – P. 106. DOI: 10.1186/s12859-018-2100-6

290. Slišković, I., Eich, H., Müller-McNicoll, M. Multifunctionality of selenoproteins // *Biochem. Soc. Trans.* – 2022. – V. 50. – P. 187–198. DOI: 10.1042/BST20210325

291. Smith, G.D., Ching, W.H., Cornejo-Paramo, P. et al. Decoding enhancer complexity with machine learning and high-throughput discovery // *Genome Biol.* – 2023. – V. 24. – P. 116. DOI: 10.1186/s13059-023-02907-y

292. Smith, R.S., Walsh, C.A. Ion Channel Functions in Early Brain Development // *Trends in Neurosciences.* – 2020. – V. 43, № 2. – P. 103–114. DOI: 10.1016/j.tins.2019.11.003

293. Solignac, M., Vautrin, D., Baudry, E., Mougél, F., Loiseau, A., Cornuet, J.-M. A Microsatellite-Based linkage map of the honeybee, *Apis mellifera* L // *Genetics.* – 2004. – V. 167. – P. 253–262. DOI: 10.1534/genetics.103.021401

294. Solkner, J., Ferencakovic, M., Gredler, B., et al. Genome-wide mapping and estimation of inbreeding depression of semen quality traits in a cattle population



// J Dairy Sci. – 2017. – V. 100, №. 6. – P. 4721-4730. DOI: 10.3168/jds.2016-12534

295. Somavilla, A.L., Sonstegard, T.S., Higa, R.H., et al. A genome-wide scan for selection signatures in Nellore cattle // *Animal Genetics*. – 2014. – V. 45. – P. 771–781. DOI: 10.1111/age.12139

296. Sorbolini, S., Marras, G., Gaspa, G., et al. Detection of selection signatures in Piemontese and Marchigiana cattle, two breeds with similar production aptitudes but different selection histories // *Genetics Selection Evolution*. – 2015. – V. 47. – P. 52. DOI: 10.1186/s12711-015-0107-0

297. Statnikov, A., Hanaff, M., Narendra, V., Konganti, K., Li, Z., Yang, L., Pei, Z., Blaser, M.J., Alferis, C.F., Alekseyenko, A.V. A comprehensive evaluation of multicategory classification methods for microbiome data // *Microbiome*. – 2013. – V. 1. – P. 11. DOI: 10.1186/2049-2618-1-11

298. Steven, M. Mussmann, Marlis, R. D., Tyler, K. C., Michael, E. D. BA3-SNPs: Contemporary migration reconfigured in BayesAss for next-generation sequence data // 2019.

299. Stoffel, M.A., Johnston, S.E., Pilkington, J.G. et al. Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal // *Nat Commun*. – 2021. – V. 12. – P. 2972. DOI: 10.1038/s41467-021-22972-5

300. Sun, G., Liang, X., Qin, K., Qin, Y., Shi, X., Cong, P., Mo, D., Liu, X., Chen, Y., He, Z. Functional Analysis of KIT Gene Structural Mutations Causing the Porcine Dominant White Phenotype Using Genome Edited Mouse Models // *Front Genet*. – 2020. – V. 11. – P. 138. DOI: 10.3389/fgene.2020.00138

301. Sundström, E., Imsland, F., Mikko, S., Wade, C., Sigurdsson, S., Pielberg, G.R., Golovko, A., Curik, I., Seltenhammer, M.H., Sölkner, J., Lindblad-Toh, K., Andersson, L., Andersson, G. Copy number expansion of the STX17 duplication in melanoma tissue from Grey horses // *BMC Genomics*. – 2012. – V. 13. – P. 365. DOI: 10.1186/1471-2164-13-365

302. Szpak, M., Xue, Y., Ayub, Q., Tyler-Smith, C. How Well Do We Understand the Basis of Classic Selective Sweeps in Humans? // *FEBS Lett.* – 2019. – V. 593, №. 14. – P. 1431–1448. DOI: 10.1002/1873-3468.13493
303. Szpiech, Z.A. selscan 2.0: scanning for sweeps in unphased data // *Bioinformatics.* – 2021. – V. 40. – P. btae006. DOI: 10.1093/bioinformatics/btae006
304. Szpiech, Z.A., Blant, A., Pemberton, T.J., et al. Genomic Autozygosity Regions Likelihood-based Inference and Classification // *Bioinformatics.* – 2017. – V. 33. – P. 2059-2062. DOI: 10.1093/bioinformatics/btx092
305. Tang, G.Q., Xue, J., Lian, M.J., et al. Inbreeding and genetic diversity in three imported Swine breeds in china using pedigree data // *Asian-Australasian journal of animal sciences.* – 2013. – V. 26, №. 6. – P. 755–765. DOI: 10.5713/ajas.2012.12496
306. Tautz, D., Domazet-Lošo, T. The evolutionary origin of orphan genes // *Nat Rev Genet.* – 2011. – V. 12. – P. 692–702. DOI: 10.1038/nrg3026
307. The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia // *Nature.* – 2008. – V. 455. – P. 237–241. DOI: 10.1038/nature07239
308. Thong, J.Y., Li, Z., Halim, A., Wang, X., Halim, M.Z., Zhai, X. WES Reveals Novel Heterozygous NBAS Gene Mutations Associated with Fanconi Syndrome in a Patient with SOPH Syndrome: Case Report // *Asploro Journal of Pediatrics and Child Health.* – 2021. – V. 3, – №. 1. – P. 1. DOI: 10.29333/asploro-21001
309. Tian, Q., Shu, L., Zhang, P., Zeng, T., Cao, Y., Xi, H., Peng, Y., Wang, Y., Mao, X., Wang, H. MN1 Neurodevelopmental Disease-Atypical Phenotype Due to a Novel Frameshift Variant in the MN1 Gene // *Front. Mol. Neurosci.* – 2021. – V. 14. – P. 789778. DOI: 10.3389/fnmol.2021.789778
310. Toratani, M., Konno, M., Asai, A., Koseki, J., Kawamoto, K., Tamari, K., Li, Z., Sakai, D., Kudo, T., Satoh, T., Sato, K., Motooka, D., Okuzaki, D., Doki, Y., Mori, M., Ogawa, K., Ishii, H. A convolutional neural network uses microscopic

images to differentiate between mouse and human cell lines and their radioresistant clones // *Cancer Res.* – 2018. – V. 78, №. 23. – P. 6703–6707. DOI: 10.1158/0008-5472.CAN-18-1427

311. Tsyklauri, O., Niederlova, V., Forsythe, E., Prasai, A., Drobek, A., Kasperek, P., Sparks, K., Trachtulec, Z., Prochazka, J., Sedlacek, R., Beales, P., Huranova, M., Stepanek, O. Bardet-Biedl Syndrome ciliopathy is linked to altered hematopoiesis and dysregulated self-tolerance // *EMBO Rep.* – 2021. – V. 22, №. 2. – P. e50785. DOI: 10.15252/embr.202051655

312. Uddin, M. M., Shiddike, J., Ahmed, A., Ahsan, T. Promoter prediction in DNA classification using machine learning algorithms // 2024 3rd International Conference on Sentiment Analysis and Deep Learning (ICSADL). – Bhimdatta, Nepal. – P. 254-260. DOI: 10.1109/ICSADL61749.2024.00047

313. Uffelmann, E., Huang, Q.Q., Munung, N.S. et al. Investigations of associations across the whole genome // *Nat Rev Methods Primers.* – 2021. – V. 1. – P. 59. DOI: 10.1038/s43586-021-00056-9

314. Uthe, J.J., Qu, L., Couture, O., Bearson, S.M.D., O'Connor, A.M., McKean, J.D., Torres, Y.R., Dekkers, J.C.M., Nettleton, D., Tuggle, C.K. Use of bioinformatic SNP predictions in differentially expressed genes to find SNPs associated with *Salmonella* colonization in swine // *J Anim Breed Genet.* – 2011. – V. 128. – P. 354–365. DOI: 10.1111/j.1439-0388.2010.00924.x

315. Uzal, L.C., Grinblat, G.L., Namías, R. et al. Seed-per-pod estimation for plant breeding using deep learning // *Comput Electron Agric.* – 2018. – V. 150. – P. 196–204. DOI: 10.1016/j.compag.2018.05.017

316. van Essen, G.J., Te Lintel Hekkert, M., Sorop, O., Heinonen, I., van der Velden, J., Merkus, D., Duncker, D.J. Cardiovascular Function of Modern Pigs Does not Comply with Allometric Scaling Laws // *Sci Rep.* – 2018. – V. 8, №. 1. – P. 792. DOI: 10.1038/s41598-017-18932-3

317. van Son, M., Tremoen, N.H., Gaustad, A.H. et al. Transcriptome profiling of porcine testis tissue reveals genes related to sperm hyperactive motility // *BMC Vet Res.* – 2020. – V. 16. – P. 161. DOI: 10.1186/s12917-020-02337-8

318. Vanderhyden, B.C., Cohen, J.N., Morley, P. Mouse oocytes regulate granulosa cell steroidogenesis // *Endocrinology*. – 1993. – V. 133. – P. 423–426. DOI: 10.1210/en.133.1.423
319. VanRaden, P. Efficient methods to compute genomic predictions // *Journal of Dairy Science*. – 2008. – V. 91. – P. 4414–4423. DOI: 10.3168/jds.2007-0980
320. Vazquez-Gomez, M., Hijas-Villalba, M.M., Varona, L., Ibanez-Escriche, N., Rosas, J.P., Negro, S., Noguera, J.L., Casellas, J. Maternal Transmission Ratio Distortion in Two Iberian Pig Varieties // *Genes (Basel)*. – 2020. – V. 11, №. 9. – P. 1050. DOI: 10.3390/genes11091050
321. Venables, W. N., Ripley, B. D. Modern Applied Statistics with S (Fourth) // Springer. – 2002. – URL <http://www.stats.ox.ac.uk/pub/MASS4>
322. Villena, F.P.-M.D., Casa-Esperon, E.D.L., Briscoe, T.L., Sapienza, C. A Genetic test to determine the origin of maternal transmission ratio distortion: Meiotic drive at the mouse om locus // *Genetics*. – 2000. – V. 154. – P. 333–342.
323. Voight, B. F., Kudaravalli, S., Wen, X., Pritchard, J. K. A map of recent positive selection in the human genome // *PLoS Biology*. – 2006. – V. 4. – P. e72. DOI: 10.1371/journal.pbio.0040072
324. Vu, P.T., Chahine, C., Chatterjee, N., MacLean, M.T., Swago, S., Bhattarai, A., Thompson, E.W., Ikhlas, A., Oteng, E., Davidson, L., Tran, R., Hazim, M., Raghupathy, P., Verma, A., Duda, J., Gerstein, H., Pang, J., Le, S., Wang, L. CT imaging-derived phenotypes for abdominal muscle and their association with age and sex in a medical biobank // *Sci Rep*. – 2024. – V. 14, №. 1. – P. 14807. DOI: 10.1038/s41598-024-23804-8
325. W.L. Fan, C.S. Ng, C.F. Chen, M.Y.J. Lu, Y.H. Chen, C.J. Liu, S.K. Chen, C.K. Chen, J.J. Chen, C.T. Mao, Y.T. Lai, W.S. Lo, W.H. Chang, W.H. Li. Genome-wide patterns of genetic variation in two domestic chickens // *Genome Biol Evol*. – 2013. – V. 5. – P. 1376–1392. DOI: 10.1093/gbe/evt064

326. Wadhwa, RR, Piekenbrock, M, Scott, JG. ripserr: Calculate Persistent Homology with Ripser-based Engines; version 0.1.0. <https://github.com/rrrlw/ripserr>.

327. Waldman, Y.Y., Biddanda, A., Dubrovsky, M., et al. The genetic history of Cochin Jews from India // *Hum. Genet.* – 2016. – V. 135. – P. 1127–1143. DOI: 10.1007/s00439-016-1708-9

328. Wang, H., Wu, C., Yuan, Z. et al. Genome-wide association study identifying genetic variants associated with carcass backfat thickness, lean percentage and fat percentage in a four-way crossbred pig population using SLAF-seq technology // *BMC Genomics.* – 2022. – V. 23. – P. 594. DOI: 10.1186/s12864-022-08675-w

329. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S., Hakonarson, H., Bucan, M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data // *Genome Research.* – 2007. – V. 17. – P. 1665–1674. DOI: 10.1101/gr.6853807

330. Wang, K.W., Wang, P. Jiang, S., Chen, Y., Jiang, D., Xiao, S., Jiang, W., Zeng, Y., Xu, Y., Qiu, X., Li, X., Li, M., Tang, X.G. Epigenome-wide DNA methylation analysis reveals differentially methylation patterns in skeletal muscle between Chinese Chenghua and Qingyu pigs // *Journal of Integrative Agriculture.* – 2022. – V. 21, №. 6. – P. 1731-1739. DOI: 10.1163/j.ijas.2022.00039

331. Wang, L., Xu, L., Liu, X., Zhang, T., Li, N., Hayel, H., Zhang, Y., Yan, H., Zhao, K., Liu, G. Copy number variation-based genome wide association study reveals additional variants contributing to meat quality in Swine // *Sci Rep.* – 2015. – V. 5. – P. 12535. DOI: 10.1038/srep12535

332. Wang, X., Hu, X., Yan, D., Shrestha, S., Pless, R., Poland, J. High-throughput phenotyping with deep learning gives insight into the genetic architecture of flowering time in wheat // *GigaScience.* – 2019. – V. 8. – P. 1–11. DOI: 10.1093/gigascience/giz050

333. Wang, X., Zhang, H., Huang, M., et al. Whole-genome SNP markers reveal conservation status, signatures of selection, and introgression in Chinese Laiwu pigs // *Evol Appl.* – 2021. – V. 14. – P. 383–398. DOI: 10.1111/eva.13289
334. Weedon, M.N., Lango, H., Lindgren, C.M., Wallace, C., Evans, D.M., Mangino, M., et al. Genome-wide association analysis identifies 20 loci that influence adult height // *Nat Genet.* – 2008. – V. 40. – P. 575–583. DOI: 10.1038/ng.177
335. Wei, J., Wu, Y., Tang, X., Liu, J., Huang, Y., Wu, Z., Li, X., Zhang, Z. Deep Learning-Based Automated Approach for Determination of Pig Carcass Traits // *Animals.* – 2024. – V. 14, №. 16. – P. 2421. DOI: 10.3390/ani14162421
336. Welcker, D., Stein, C., Feitosa, N.M. et al. Hemicentin-1 is an essential extracellular matrix component of the dermal–epidermal and myotendinous junctions // *Sci Rep.* – 2021. – V. 11. – P. 17926. DOI: 10.1038/s41598-021-96754-9
337. Whelehan, O.P., Gregory, N.G. Further studies on skull shape in relation to carcass fatness in pigs // *J. Sci. Food Agric.* – 1985. – V. 36. – P. 963-969. DOI: 10.1002/jsfa.2740360813
338. Wides, R. The natural history of Teneurins: A billion years of evolution in three key steps // *Front. Neurosci.* – 2019. – V. 13. – P. 109. DOI: 10.3389/fnins.2019.00109
339. Wilkinson, S., Zen, H. Lu., Megens, H.J, Archibald, A.L., Haley, C., Jackson, I.J., Groenen, M.A.M., Crooijmans, R.P.M.A., Ogden, R, Wiener, P. Signatures of Diversifying Selection in European Pig Breeds // *PLoS Genetics.* – 2013. – April 25.
340. Wolfsberg, T.G., Straight, P.D., Gerena, R.L., Huovila, A.-P.J., Primakoff, P., Myles, D.G., White, J.M. ADAM, a widely distributed and developmentally regulated gene family encoding membrane proteins with a disintegrin and metalloprotease domain // *Dev Biol.* – 1995. – V. 169, №. 1. – P. 378–83. DOI: 10.1006/dbio.1995.1161

341. Wright, D., Boije, H., Meadows, J.R., Bed'hom, B., Gourichon, D., Vieaud, A., Tixier-Boichard, M., Rubin, C.J., Imsland, F., Hallböök, F., Andersson, L. Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens // *PLoS Genet.* – 2009. – V. 5, №. 6. – P. e1000512. DOI: 10.1371/journal.pgen.1000512
342. Wright, S. Coefficients of Inbreeding and Relationship // *The American Naturalist.* – 1922. – V. 56. – P. 330-338. DOI: 10.1086/279852
343. Wright, S. Evolution in mendelian populations // *Genetics.* 1931. – V. 16, №. 2. – C. 97–159. DOI: /10.1093/genetics/16.2.97
344. Wu, G., Hao, L., Han, Z., Gao, S., Latham, K.E., de Villena, F.P.-M., Sapienza, C. Maternal transmission ratio distortion at the mouse om locus results from meiotic drive at the second meiotic division // *Genetics.* – 2005. – V. 170. – P. 327–334. DOI: 10.1534/genetics.104.035590
345. Wu, L., Zhou, R., Diao, J., Chen, X., Huang, J., Xu, K., Ling, L., Xia, W., Li, X. Differentially expressed circular RNAs in orbital adipose/connective tissue from patients with thyroid-associated ophthalmopathy // *Exp. Eye Res.* – 2020. – V. 196. – P. 108036. DOI: 10.1016/j.exer.2020.108036
346. Wu, X., Zhang, H., Long, H., Zhang, D., Yang, L., Ji, C., Huang, J., Zhao, Y., Han, R., Che, D. Genome-Wide Selection Signal Analysis to Investigate Wide Genomic Heredity Divergence between Eurasian Wild Boar and Domestic Pig // *Animals (Basel).* – 2023. – V. 13, №. 13. – P. 2158. DOI: 10.3390/ani13132158
347. Wucher, V., Sodaei, R., Amador, R., Irimia, M., Guigo, R. Day-night and seasonal variation of human gene expression across tissues // *PLoS Biol.* – 2023. – V. 21, №. 2. – P. e3001986. DOI: 10.1371/journal.pbio.3001986
348. Xia, T., Nakayama, K. Signatures of adaptation in myopia-related genes on the sunlight exposure hypothesis // *J Physiol Anthropol.* – 2023. – V. 42, №. 1. – P. 25. DOI: 10.1007/s42521-023-00124-0
349. Xie, Y., Zaccagna, F., Rundo, L., Testa, C., Agati, R., Lodi, R., Manners, D.N., Tonon, C. Convolutional neural network techniques for brain tumor

classification (from 2015 to 2022): Review, challenges, and future perspectives // *Diagnostics*. – 2022. – V. 12, №. 8. – P. 1850. DOI: 10.3390/diagnostics12081850

350. Xue, Y., Liu, S., Li, W., Mao, R., Zhuo, Y., Xing, W., Liu, J., Wang, C., Zhou, L., Lei, M., et al. Genome-Wide Association Study Reveals Additive and Non-Additive Effects on Growth Traits in Duroc Pigs // *Genes*. – 2022. – V. 13. – P. 1454. DOI: 10.3390/genes13081454

351. Y. Xu, Z. Zhang, L. You, J. Liu, Z. Fan, X. Zheng, Z. Wang, G.E. Liu, C.W. Ernst, R. Ran, J. Xu, J. Liu. Evidence of evolutionary history and selective sweeps in the genome of Meishan pig reveals its genetic and phenotypic characterization // *Gigascience*. – 2018. – V. 7, №. 5. – P. giy058. DOI: 10.1093/gigascience/giy058

352. Yang, J., Lee, S., Goddard, M., et al. GCTA: a tool for genome-wide complex trait analysis // *American Journal of Human Genetics*. – 2011. – V. 88. – P. 76-82. DOI: 10.1016/j.ajhg.2010.11.011

353. Yang, X., Zhang, Z., Qiu, M., et al. Glypican-5 is a novel metastasis suppressor gene in non-small cell lung cancer // *Cancer Lett*. – 2013. – V. 341. – P. 265–273. DOI: 10.1016/j.canlet.2013.07.018

354. Yao, G., Yin, M., Lian, J., Tian, H., Liu, L., Li, X., Sun, F. Transactivation of microRNA-224 is involved in transforming growth factor- $\beta$ -mediated mouse granulosa cell proliferation and granulosa cell function by targeting Smad4 // *Mol Endocrinol*. – 2010. – V. 24. – P. 540–551. DOI: 10.1210/me.2009-0347

355. Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., et al. Sequencing of 50 human exomes reveals adaptation to high altitude // *Science*. – 2010. – V. 329. – P. 75–78. DOI: 10.1126/science.1190373

356. Yibin, Qiu, Zhanwei, Zhu, Fanming, Meng, Donglin, Ruan, Longlong, Peng, Rongrong, Ding, Gengyuan, Cai, Ming, Yang, Jie, Enqin, Zheng. Identification of candidate genes associated with carcass component weights in



commercial crossbred pigs through a combined GWAS approach // *Journal of Animal Science*. – 2023. – V. 101. – P. skad121. DOI: 10.1093/jas/skad121

357. Yu, X., Zhai, C., Fan, Y., Zhang, J., Liang, N., Liu, F., Cao, L., Wang, J., Du, J., TUSC3: a novel tumour suppressor gene and its functional implications // *J Cell Mol Med*. – 2017. – V. 21, №. 9. – P. 1711–1718. DOI: 10.1111/jcmm.13312

358. Yu, Y., Wang, C., Clare, S., Wang, J., Lee, S.C., Brandt, C., Burke, S., Lu, L., He, D., Jenkins, N.A., Copeland, N.G., Dougan, G., Liu, P. The transcription factor Bcl11b is specifically expressed in group 2 innate lymphoid cells and is essential for their development // *J Exp Med*. – 2015. – V. 212, №. 6. – P. 865–874. DOI: 10.1084/jem.20141609

359. Zappaterra, M., Gioiosa, S., Chillemi, G., Zambonelli, P., Davoli, R. Muscle transcriptome analysis identifies genes involved in ciliogenesis and the molecular cascade associated with intramuscular fat content in Large White heavy pigs // *PLoS One*. – 2020. – V. 15, №. 5. – P. e0233372. DOI: 10.1371/journal.pone.0233372

360. Zavarez, L.B., Utsunomiya, Y.T., Carmo, A.S., et al. Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes // *Frontiers in Genetics*. – 2015. – V. 6. – P. 1-8. DOI: 10.3389/fgene.2015.00261

361. Zhang, H., Bao, S., Zhao, X., Bai, Y., Lv, Y., Gao, P., Li, F., Zhang, W. Genome-wide association study and phenotype prediction of reproductive traits in Large White pigs // *Animals*. – 2024. – V. 14, №. 23. – P. 3348. DOI: 10.3390/ani14233348

362. Zhang, Q., Guldbrandtsen, B., Bosse, M., et al. Runs of homozygosity and distribution of functional variants in the cattle genome // *BMC Genomics*. – 2015. – V. 16. – P. 542. DOI: 10.1186/s12864-015-1742-6

363. Zhang, R., Lahens, N.F., Ballance, H.I., Hughes, M.E., Hogenesch, J.B. A circadian gene expression atlas in mammals: implications for biology and medicine // *Proc Natl Acad Sci USA*. – 2014. – V. 111. – P. 168–174. DOI: 10.1073/pnas.1418968111

364. Zhang, T., Zheng, Y., Kuang, T., Yang, L., Jiang, H., Wang, H., Zhao, Y., Han, R., Che, D. Arginine Regulates Zygotic Genome Activation in Porcine Embryos Under Nutrition Restriction // *Front. Vet. Sci.* – 2022. – V. 9. – P. 921406. DOI: 10.3389/fvets.2022.921406

365. Zhang, X., Lin, P.Y., Liakath-Ali, K., Sudhof, T.C. Teneurins assemble into presynaptic nanoclusters that promote synapse formation via postsynaptic non-teneurin ligands // *Nat Commun.* – 2022. – V. 13, №. 1. – P. 2297. DOI: 10.1038/s41467-022-29545-4

366. Zhang, Z., Chen, Z., Ye, S., He, Y., Huang, S., Yuan, X., Chen, Z., Zhang, H., Li, J. Genome-Wide Association Study for Reproductive Traits in a Duroc Pig Population // *Animals.* – 2019. – V. 9. – P. 732. DOI: 10.3390/ani9070732

367. Zhang, Z., Chen, Z., Ye, S., He, Y., Huang, S., Yuan, X., Chen, Z., Zhang, H., Li, J. Genome-Wide Association Study for Reproductive Traits in a Duroc Pig Population // *Animals.* – 2019. – V. 9. – P. 732. DOI: 10.3390/ani9070732

368. Zhao, X., Chang, S., Liu, X., Wang, S., Zhang, Y., Lu, X., Zhang, T., Zhang, H., Wang, L. Imprinting aberrations of SNRPN, ZAC1 and INPP5F genes involved in the pathogenesis of congenital heart disease with extracardiac malformations // *J Cell Mol Med.* – 2020. – V. 24, №. 17. – P. 9898-9907. DOI: 10.1111/jcmm.15303

369. Zheng, X., Levine, D., Shen, J., Gogarten, S., Laurie, C., Weir, B. A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data // *Bioinformatics.* – 2012. – V. 28, №. 24. – P. 3326-3328. DOI: 10.1093/bioinformatics/bts642

370. Zhong, Z.Q., Li, R., Wang, Z., Tian, S.S., Xie, X.F., Wang, Z.Y., Na, W., Wang, Q.S., Pan, Y.C., Xiao, Q. Genome-wide scans for selection signatures in indigenous pigs revealed candidate genes relating to heat tolerance // *Animal.* – 2023. – V. 17, №. 7. – P. 100882. DOI: 10.1016/j.animal.2023.100882

371. Zhou, F., Fan, X., Miao, Y. LPIN1 promotes triglycerides synthesis and is transcriptionally regulated by PPAR $\gamma$  in buffalo mammary epithelial cells // *Sci. Rep.* – 2022. – V. 12. – P. 2390. DOI: 10.1038/s41598-022-06114-w
372. Zhou, H., Clapham, D.E. Mammalian MagT1 and TUSC3 are required for cellular magnesium uptake and vertebrate embryonic development // *Proc Natl Acad Sci U S A.* – 2009. – V. 106. – P. 15750–15755. DOI: 10.1073/pnas.0904809106
373. Zhou, J., Liu, L., Lopdell, T.J., Garrick, D.J., Shi, Y. HandyCNV: Standardized Summary, Annotation, Comparison, and Visualization of Copy Number Variant, Copy Number Variation Region, and Runs of Homozygosity // *Front Genet.* – 2021. – V. 12. – P. 731355. DOI: 10.3389/fgene.2021.731355
374. Zhou, Q.Y., Huang, J.N., Zhu, M.J., Zhao, S.H. Molecular characterization and association analysis with production traits of the porcine INPP5F gene // *Mol Biol Rep.* – 2009. – V. 36. – P. 1095–1098. DOI: 10.1007/s11033-008-9346-1
375. Zhou, X., Duan, J., Zhou, W., Zhang, A., Chen, Q. Upregulated  $\beta$ -actinin-1 impairs endometrial epithelial cell adhesion by downregulating NEBL in recurrent implantation failure // *iScience.* – 2024. – V. 27, №. 3. – P. 3. DOI: 10.1016/j.isci.2024.104174
376. Zhou, Y., Connor, E.E., Wiggans, G.R. et al. Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle // *BMC Genomics.* – 2018. – V. 19. – P. 314. DOI: 10.1186/s12864-018-4699-5
377. Zhou, Y., Utsunomiya, Y.T., Xu, L., Hay, E.H.A., Bickhart, D.M., Sonstegard, T.S., et al. Comparative analyses across cattle genders and breeds reveal the pitfalls caused by false positive and lineage-differential copy number variations // *Sci Rep.* – 2016. – V. 6. – P. 29219. DOI: 10.1038/srep29219
378. Zhu, B., Zhang, Q., Pan, Y., Mace, E.M., Yang, E. et al. A cell-autonomous mammalian 12 hr clock coordinates metabolic and stress rhythms // *Cell Metab.* – 2017. – V. 25. – P. 1305–1319. DOI: 10.1016/j.cmet.2017.03.008

379. Zhu, W., Trivedi, C.M., Zhou, D., Yuan, L., Lu, M.M., Epstein, J.A. Inpp5f is a polyphosphoinositide phosphatase that regulates cardiac hypertrophic responsiveness // *Circ Res.* – 2009. – V. 105. – P. 1240–1247.

380. Zhu, X., Cheng, S.Y. New insights into regulation of lipid metabolism by thyroid hormone // *Curr Opin Endocrinol Diabetes Obes.* – 2010. – V. 17(5). – P. 408–413.

381. Zorc, M., Skorput, D., Gvozdanovic, K., Margeta, P., Karolyi, D., Lukovic, Z., Salajpal, K., Savic, R., Munoz, M., Bovo, S., Djurkin Kusec, I., Radovic, C., Kusec, G., Candek Potokar, M., Dovc, P. Genetic diversity and population structure of six autochthonous pig breeds from Croatia, Serbia, and Slovenia // *Genet Sel Evol.* – 2022. – V. 54(1). – №. 30.

382. Бакоев С. Ю., Гетманцева Л. В. Методы оценки инбридинга и подписей селекции сельскохозяйственных животных на основе протяженных гомозиготных областей // *Достижения науки и техники АПК.* – 2019. – Т. 33. DOI: 10.24411/0235-2451-2019-10000.

383. Бакоев С.Ю., Романец Т.С., Коробейникова А.В., Мишина А.И., Колосов А.Ю., Гетманцева Л.В. Генетическое разнообразие и оценка аутозиготности в популяциях дикого европейского кабана и домашних свиней // *Экологическая генетика.* – 2023. – Т. 21, №. 4. – С. 640–649.

384. Мишина А. И., Бакоев С. Ю., Ооржак А. Ю., Кескинов А. А., Кабиева Ш. Ш., Коробейникова А. В., Юдин В. С., Боброва М. М., Шестаков Д. А., Макаров В. В., Гетманцева Л. В. Поиск сигналов положительного отбора генов циркадных ритмов PER1, PER2, PER3 в различных популяциях людей // *Вавиловский журнал генетики и селекции.* – 2024. – Т. 28, №. 6. – С. 640-649.

385. Селионова М.И., Айбазов А.М., Сермягин А.А., Белоус А.А., Зиновьева Н.А. Полногеномный анализ в поиске генов-кандидатов, связанных с признаками мясной продуктивности у мясо-молочных коз // *Теория и практика мясопереработки.* – 2022. – Т. 7(4). – С. 288–295.

## Приложения

### Приложение 1. Скрипт для расчета и визуализации nsl, ihs и ihh12

```
system("plink1.exe --bfile doct --keep-allele-order --keep doca1.txt --make-bed -
-out doct1")#
#MDS
system("plink1.exe --allow-extra-chr --noweb --bfile doct1 --geno 0.1 --genome --
out pca_doct")
system("plink1.exe --allow-extra-chr --noweb --bfile doct1 --geno 0.1 --read-genome
pca_doct.genome --cluster --mds-plot 4 --out plot_run_LW_LR")

## % in PCA #
system('plink1.exe --allow-extra-chr --bfile doct1 --distance-matrix --out glob_pig
breeds')
dist_pig<-read.table("glob_pigbreeds.mdist",header=F)
fam <- data.frame(famids=read.table("glob_pigbreeds.mdist.id")[,1])
mds_pig <- cmdscale(dist_pig,eig=T,3)
eigenvec_pig <- cbind(fam,mds_pig$points)
(eigen_values <- round(((mds_pig$eig)/sum(mds_pig$eig))*100,2))[1:3]

#####
#####
#####PCA ggplot#####
test <- read.table('plot_run_LW.mds',header=T)
df1 <- test[,c(4,5,6,3,1)]
df1[,4] <- ifelse(df1$FID=="КБ_1", "1", as.vector(df1[,4]))
df1[,4] <- ifelse(df1$FID=="КБ_3", "2", as.vector(df1[,4]))
df1[,4] <- ifelse(df1$FID=="КБ_2", "3", as.vector(df1[,4]))
df1 <- df1[order(df1$SOL,decreasing = F),]
df1
#C1-C2
library(ggplot2)
library(ggpubr)
p1 <- ggplot(df1) +
  geom_point(aes(x=C1, y=C2, color=factor(FID), shape=factor(FID),size=12)) +
  scale_color_manual(values=c('red',
                              'blue',
                              'green',
                              "goldenrod",
                              "brown",
```

```

        "orange",
        'yellow',
        "darkolivegreen1",
        "pink",
        'red',
        'blue',
        'green',
        'goldenrod',
        'gray',
        'darkolivegreen1',
        'black',
        'khaki'

    )) +
  scale_shape_manual(values=c(
    16,
    16,
    16,
    16,
    16,
    16,
    16,
    16,
    16,
    16,
    16,16,16,16,16,16,16,8,16,0,13,1,2,3,4,5,1,2,3,4,5,6,0,8)) +
  #stat_ellipse(aes(x=C1,y=C2,color=factor(FID)),
  #type = "norm", size=1, linetype = 1, alpha=0.8) +
  theme_bw() +
  theme(legend.position = 'right',
        legend.background = element_rect(fill='white', size=0.5,linetype='solid',color='black'),
        legend.text = element_text(colour="black", size = 12))+
  guides(color = guide_legend(override.aes = list(size=2)))+
  #guides(color=guide_legend(""),fill=guide_legend("")) +
  #cleanup+
  labs(x = "C1 (14.36%)",
       y = "C2 (5.77%)")
p2 <- ggplot(df1) +
  geom_point(aes(x=C1, y=C3, color=factor(FID), shape=factor(FID),size=12)) +
  scale_color_manual(values=c('red',
                              'blue',
                              'green',

```

```

        "goldenrod",
        "brown",
        "orange",
        'yellow',
        "darkolivegreen1",
        "pink",
        'red',
        'blue',
        'green',
        'goldenrod',
        'gray',
        'darkolivegreen1',
        'black',
        'khaki'

    )) +
  scale_shape_manual(values=c(
    16,
    16,
    16,
    16,
    16,
    16,
    16,
    16,
    16,
    16,16,16,16,16,16,16,16,8,16,0,13,1,2,3,4,5,1,2,3,4,5,6,0,8)) +
  #stat_ellipse(aes(x=C1,y=C2,color=factor(FID)),
  #type = "norm", size=1, linetype = 1, alpha=0.8) +
  theme_bw() +
  theme(legend.position = 'right',
        legend.background = element_rect(fill='white', size=0.5,linetype='solid',color='black'),
        legend.text = element_text(colour="black", size = 12))+
  guides(color = guide_legend(override.aes = list(size=2)))+
  #guides(color=guide_legend(""),fill=guide_legend("")) +
  #cleanup+
  labs(x = "C1 (14.36%)",
       y = "C3 (1.48%)")
p3 <- ggplot(df1) +
  geom_point(aes(x=C2, y=C3, color=factor(FID), shape=factor(FID),size=12)) +
  scale_color_manual(values=c('red',

```

```

        'blue',
        'green',
        "goldenrod",
        "brown",
        "orange",
        'yellow',
        "darkolivegreen1",
        "pink",
        'red',
        'blue',
        'green',
        'goldenrod',
        'gray',
        'darkolivegreen1',
        'black',
        'khaki'

    )) +
    scale_shape_manual(values=c(
      16,
      16,
      16,
      16,
      16,
      16,
      16,
      16,
      16,
      16,16,16,16,16,16,16,8,16,0,13,1,2,3,4,5,1,2,3,4,5,6,0,8)) +
    #stat_ellipse(aes(x=C1,y=C2,color=factor(FID)),
    #type = "norm", size=1, linetype = 1, alpha=0.8) +
    theme_bw() +
    theme(legend.position = 'right',
          legend.background = element_rect(fill='white', size=0.5,linetype='solid',color='black'),
          legend.text = element_text(colour="black", size = 12))+
    guides(color = guide_legend(override.aes = list(size=2)))+
    #guides(color=guide_legend(""),fill=guide_legend("")) +
    #cleanup+
    labs(x = "C2 (5.77%",
         y = "C3 (1.48%)")

```



```

final_plot <- ggarrange(p1, p2, p3)
final_plot <- annotate_figure(final_plot, top = text_grob("Порода", size = 15))
final_plot

##

library(rehh)
library(rehh.data)
library(dplyr)
library(readr)
library(data.table)
##iht_dom
d0 <- list.files(pattern = "100bins.norm") %>%
  lapply(read.table, stringsAsFactors=T, sep = "\t") %>%
  bind_rows
colnames(d0) <- c("SNP", "Pos", "P", "IHS1", "IHS2", "IHS", "Norm", "Criteria")
d0$pvalue <- 2*pnorm(-abs(d0$Norm))
d0$CHROM <- sub(":.*", "", d0$SNP)
d0$CHROM <- as.numeric(d0$CHROM)
d0 <- d0[order(d0$pvalue),]
fwrite(d0, "domestic.txt", sep = " ")
jpeg("dom_iht.jpeg", width = 30, height = 10, units = "cm", res = 400)
qqman::manhattan(d0, chr="CHROM", bp="Pos", p="pvalue", snp="SNP", ylim = c(
0,6), cex = 0.5, col = c("black", "red", "green", "blue", "cyan"), genomewideline=-log
10(1.01E-04), suggestiveline = -log10(1.01E-02))
dev.off()
##iht_imp1
d0 <- list.files(pattern = "100bins.norm") %>%
  lapply(read.table, stringsAsFactors=T, sep = "\t") %>%
  bind_rows
colnames(d0) <- c("SNP", "Pos", "P", "IHS1", "IHS2", "IHS", "Norm", "Criteria")
d0$pvalue <- 2*pnorm(-abs(d0$Norm))
d0$CHROM <- sub(":.*", "", d0$SNP)
d0$CHROM <- as.numeric(d0$CHROM)
d0 <- d0[order(d0$pvalue),]
fwrite(d0, "imp_1.txt", sep = " ")
jpeg("imp1_iht.jpeg", width = 30, height = 10, units = "cm", res = 400)
qqman::manhattan(d0, chr="CHROM", bp="Pos", p="pvalue", snp="SNP", ylim = c(
0,6), cex = 0.5, col = c("black", "red", "green", "blue", "cyan"), genomewideline=-log
10(1.01E-04), suggestiveline = -log10(1.01E-02))
dev.off()

```

```

##ihs_imp2
d0 <- list.files(pattern = "100bins.norm") %>%
  lapply(read.table, stringsAsFactors=T, sep = "\t") %>%
  bind_rows
colnames(d0) <- c("SNP", "Pos", "P", "IHS1", "IHS2", "IHS", "Norm", "Criteria")
d0$pvalue <- 2*pnorm(-abs(d0$Norm))
d0$CHROM <- sub(":.*", "", d0$SNP)
d0$CHROM <- as.numeric(d0$CHROM)
d0 <- d0[order(d0$pvalue),]
fwrite(d0, "imp_2.txt", sep = " ")
jpeg("imp2_ihs.jpeg", width = 30, height = 10, units = "cm", res = 400)
qqman::manhattan(d0, chr="CHROM", bp="Pos", p="pvalue", snp="SNP", ylim = c(
0,6), cex = 0.5, col = c("black", "red", "green", "blue", "cyan"), genomewideline=-log
10(1.01E-04), suggestiveline = -log10(1.01E-02))
dev.off()

##xpehh_dom_imp1
d0 <- list.files(pattern = "out.norm") %>%
  lapply(read.table, stringsAsFactors=T, sep = "\t", header = T) %>%
  bind_rows
d0$CHROM <- sub ( ":.*", "", d0$id )
d0$pvalue <- 2*pnorm(-abs(d0$normxpehh))
d0 <- d0[,c(1,11,2,4,5,7,10,12)]
colnames(d0) <- c("SNP", "CHR", "POSITION", "freq_A", "iHH_A", "iHH_D", "crit", "pv
alue")
d0 <- d0[order(d0$pvalue),]
d0$CHR <- as.numeric(d0$CHR)
fwrite(d0, "dom_imp_1.txt", sep = " ")
jpeg("dom_imp1.jpeg", width = 30, height = 10, units = "cm", res = 400)
qqman::manhattan(d0, chr="CHR", bp="POSITION", p="pvalue", snp="SNP", main =
"Манхэттенский сюжет:ihs-статистика для КБ_1", ylim = c(0,6), cex = 0.5, col = c
("black", "red", "green", "blue", "cyan"), genomewideline=-log10(1.01E-04), suggest
iveline = -log10(1.01E-02))
dev.off()

##xpehh_dom_imp2
d0 <- list.files(pattern = "out.norm") %>%
  lapply(read.table, stringsAsFactors=T, sep = "\t", header = T) %>%
  bind_rows
d0$CHROM <- sub ( ":.*", "", d0$id )
d0$pvalue <- 2*pnorm(-abs(d0$norm))

```

```

d0 <- d0[,c(1,11,2,4,5,7,10,12)]
colnames(d0) <- c("SNP","CHR","POSITION","freq_A","iHH_A","iHH_D", "crit","pvalue")
d0 <- d0[order(d0$pvalue),]
d0$CHR <- as.numeric(d0$CHR)
fwrite(d0, "dom_imp_2.txt",sep = " ")
jpeg("dom_imp2.jpeg",width = 30, height = 10, units = "cm", res = 400)
qqman::manhattan(d0,chr="CHR",bp="POSITION",p="pvalue",snp="SNP", main =
"Манхэттенский сюжет:ihs-статистика для КБ_1", ylim = c(0,6), cex = 0.5,col = c
("black","red","green","blue","cyan"),genomewideline=-log10(1.01E-04),suggest
iveline = -log10(1.01E-02))
dev.off()

##ihh12_dom
d1 <- fread("dom_ihh12_chr_1.ihh12.out.norm")
d2 <- fread("dom_ihh12_chr_2.ihh12.out.norm")
d3 <- fread("dom_ihh12_chr_3.ihh12.out.norm")
d4 <- fread("dom_ihh12_chr_4.ihh12.out.norm")
d5 <- fread("dom_ihh12_chr_5.ihh12.out.norm")
d6 <- fread("dom_ihh12_chr_6.ihh12.out.norm")
d7 <- fread("dom_ihh12_chr_7.ihh12.out.norm")
d8 <- fread("dom_ihh12_chr_8.ihh12.out.norm")
d9 <- fread("dom_ihh12_chr_9.ihh12.out.norm")
d10 <- fread("dom_ihh12_chr_10.ihh12.out.norm")
d11 <- fread("dom_ihh12_chr_11.ihh12.out.norm")
d12 <- fread("dom_ihh12_chr_12.ihh12.out.norm")
d13 <- fread("dom_ihh12_chr_13.ihh12.out.norm")
d14 <- fread("dom_ihh12_chr_14.ihh12.out.norm")
d15 <- fread("dom_ihh12_chr_15.ihh12.out.norm")
d16 <- fread("dom_ihh12_chr_16.ihh12.out.norm")
d17 <- fread("dom_ihh12_chr_17.ihh12.out.norm")
d18 <- fread("dom_ihh12_chr_18.ihh12.out.norm")
d0 <- rbind(d1,d2,d3,d4,d5,d6,d7,d8,d9,d10,d11,d12,d13,d14,d15,d16,d17,d18 )
rm(d1,d2,d3,d4,d5,d6,d7,d8,d9,d10,d11,d12,d13,d14,d15,d16,d17,d18 )
d0$CHROM <- sub ( ".*", "", d0$id )
d0 <- d0[,c(1,7,2,3,4,5,6)]
colnames(d0) <- c("SNP","CHR","POSITION","freq_A","iHH12","norm","crit")
d0$pvalue <- 2*pnorm(-abs(d0$norm))
fwrite(d0, "dom_h12.txt", sep = " ")
d0$CHR <- as.numeric(d0$CHR)
jpeg("dom_h12.jpeg",width = 30, height = 10, units = "cm", res = 400)

```

```
qqman::manhattan(d0,chr="CHR",bp="POSITION",p="pvalue",snp="SNP", main =
"Манхэттенский сюжет:Н12-статистика для отечественной селекции",ylim = c
(0,6),cex = 0.5,col = c("black","red","green","blue","cyan"),genomewideline = -lo
g10(1.01E-04),suggestiveline = -log10(1.01E-02))#,genomewideline=-log10(0.05/n
row(gwasp)),suggestiveline = F
dev.off()
```

```
##ihh12_imp1
```

```
d0 <- list.files(pattern = "out.norm") %>%
  lapply(read.table, stringsAsFactors=T, sep = "\t", header = T) %>%
  bind_rows
d0$CHROM <- sub ( ".*" , "" , d0$id )
d0 <- d0[,c(1,7,2,3,4,5,6)]
d0$CHROM <- as.numeric(d0$CHROM)
colnames(d0) <- c("SNP","CHR","POSITION","freq_A","iHH12","norm", "crit" )
d0$pvalue <- 2*pnorm(-abs(d0$norm))
fwrite(d0, "imp_h12.txt", sep = " ")
jpeg("h12_imp1.jpeg",width = 30, height = 10, units = "cm", res = 400)
qqman::manhattan(d0,chr="CHR",bp="POSITION",p="pvalue",snp="SNP", main =
"Манхэттенский сюжет:Н12-статистика для импортной селекции1",ylim = c(0,
6), cex = 0.5,col = c("black","red","green","blue","cyan"),genomewideline=-log10
(1.01E-04),suggestiveline = -log10(1.01E-02))#,genomewideline=-log10(0.05/ nro
w(gwasp)),suggestiveline = F dev.off()#ylim = c(0,50) #d32 <- d31[order(d31$pval
ue),]
dev.off()
```

```
##ihh12_imp2
```

```
d0 <- list.files(pattern = "out.norm") %>%
  lapply(read.table, stringsAsFactors=T, sep = "\t", header = T) %>%
  bind_rows
d0$CHROM <- sub ( ".*" , "" , d0$id )
d0 <- d0[,c(1,7,2,3,4,5,6)]
d0$CHROM <- as.numeric(d0$CHROM)
colnames(d0) <- c("SNP","CHR","POSITION","freq_A","iHH12","norm", "crit")
d0$pvalue <- 2*pnorm(-abs(d0$norm))
fwrite(d0, "imp2_h12.txt", sep = " ")
jpeg("h12_imp1.jpeg",width = 30, height = 10, units = "cm", res = 400)
qqman::manhattan(d0,chr="CHR",bp="POSITION",p="pvalue",snp="SNP", main =
"Манхэттенский сюжет:Н12-статистика для импортной селекции_2",ylim = c(
0,6), cex = 0.5,col = c("black","red","green","blue","cyan"),genomewideline=-log
10(1.01E-04),suggestiveline = -log10(1.01E-02))#,genomewideline=-log10(0.05/ n
```

```

row(gwasp)),suggestiveline = F dev.off())#ylim = c(0,50) #d32 <- d31[order(d31$p
value),]
dev.off()

##nsl
##domestic
d0 <- list.files(pattern = ".norm") %>%
  lapply(read.table, stringsAsFactors=T, sep = "\t") %>%
  bind_rows
colnames(d0) <- c("SNP","POSITION","freq_A","sl1", "sl10","nsl", "norm","crit" )
d0$CHROM <- sub ( ".*" , "" , d0$SNP)
d0$CHROM <- as.numeric(d0$CHROM)
d0 <- d0[,c(1,9,2,3,6,7,8)]
colnames(d0) <- c("SNP","CHR","POSITION","freq_A","nsl","norm", "crit")
d0$pvalue <- 2*pnorm(-abs(d0$norm))
fwrite(d0, "dom_nsl.txt", sep = " ")
jpeg("nsl_dom.jpeg",width = 30, height = 10, units = "cm", res = 400)
qqman::manhattan(d0,chr="CHR",bp="POSITION",p="pvalue",snp="SNP", main =
"Манхэттенский сюжет:nsl-статистика для местной селекции",ylim = c(0,6), se
x = 0.5,col = c("black","red","green","blue","cyan"),genomewideline=-log10(1.01
E-03),suggestiveline = -log10(1.01E-02))#,genomewideline=-log10(0.05/ nrow(gw
asp)),suggestiveline = F dev.off())#ylim = c(0,50) #d32 <- d31[order(d31$pvalue),]
dev.off()

##imp1
d0 <- list.files(pattern = ".norm") %>%
  lapply(read.table, stringsAsFactors=T, sep = "\t") %>%
  bind_rows
colnames(d0) <- c("SNP","POSITION","freq_A","sl1", "sl10","nsl", "norm","crit" )
d0$CHROM <- sub ( ".*" , "" , d0$SNP)
d0$CHROM <- as.numeric(d0$CHROM)
d0 <- d0[,c(1,9,2,3,6,7,8)]
colnames(d0) <- c("SNP","CHR","POSITION","freq_A","nsl","norm", "crit")
d0$pvalue <- 2*pnorm(-abs(d0$norm))
fwrite(d0, "dom_nsl.txt", sep = " ")
jpeg("nsl_dom.jpeg",width = 30, height = 10, units = "cm", res = 400)
qqman::manhattan(d0,chr="CHR",bp="POSITION",p="pvalue",snp="SNP", main =
"Манхэттенский сюжет:nsl-статистика для импортной селекции_1",ylim = c(0,
6), cex = 0.5,col = c("black","red","green","blue","cyan"),genomewideline=-log10
(1.01E-03),suggestiveline = -log10(1.01E-02))#,genomewideline=-log10(0.05/ nro

```

```

w(gwas)),suggestiveline = F dev.off())#ylim = c(0,50) #d32 <- d31[order(d31$pval
ue),]
dev.off()

##imp2
d0 <- list.files(pattern = ".norm") %>%
  lapply(read.table, stringsAsFactors=T, sep = "\t") %>%
  bind_rows
colnames(d0) <- c("SNP","POSITION","freq_A","sl1", "sl10","nsI", "norm","crit" )
d0$CHROM <- sub ( ".*" , "" , d0$SNP)
d0$CHROM <- as.numeric(d0$CHROM)
d0 <- d0[,c(1,9,2,3,6,7,8)]
colnames(d0) <- c("SNP","CHR","POSITION","freq_A","nsI","norm", "crit")
d0$pvalue <- 2*pnorm(-abs(d0$norm))
fwrite(d0, "dom_nsl.txt", sep = " ")
jpeg("nsl_dom.jpeg",width = 30, height = 10, units = "cm", res = 400)
qqman::manhattan(d0,chr="CHR",bp="POSITION",p="pvalue",snp="SNP", main =
"Манхэттенский сюжет:nsl-статистика для импортной селекции_2",ylim = c(0,
6), cex = 0.5,col = c("black","red","green","blue","cyan"),genomewideline=-log10
(1.01E-03),suggestiveline = -log10(1.01E-02))#,genomewideline=-log10(0.05/ nro
w(gwas)),suggestiveline = F dev.off())#ylim = c(0,50) #d32 <- d31[order(d31$pval
ue),]
dev.off()

```

## Приложение 2. Скрипт для расчета и визуализации CNV

```
library(CNVRanger)
library(AnnotationHub)
library(regionerR)
#library(BSgenome.Btaurus.UCSC.bosTau6.masked)
library(SummarizedExperiment)
library(curatedTCGAData)
library(TCGAutils)
library(Gviz)
library(dplyr)
library(data.table)
#BiocManager::install("CNVRanger",force = TRUE)

data.dir <- "/Desktop/статья_cnv/LW"
cnv.loc <- file.path(data.dir, " LW _goodcnv_segment.txt")
cnv.calls <- read.delim(cnv.loc, sep = "\t",as.is=TRUE)
cnv.calls <- GenomicRanges::makeGRangesListFromDataFrame(cnv.calls,
                                                           split.field="ID", keep.extra.columns=TRUE)

cnv.calls <- sort(cnv.calls)
cnv.calls
ra <- RaggedExperiment::RaggedExperiment(cnv.calls)
ra <- assay(ra)
ra <- as.data.frame(ra)
glimpse(ra)
raa <- as.data.frame(t(ra))
fwrite(raa, '/Desktop/статья_cnv/LW/cnv_code.txt', sep = '\t', row.names = T, col.names = T)
ro.cnvrs <- populationRanges(cnv.calls, mode="RO", ro.thresh=0.51)
ro.cnvrs <- assay(ro.cnvrs)
raf <- sort(ro.cnvrs)
ro <- as.data.frame(ro.cnvrs)
fwrite(ro, '/Desktop/статья_cnv/LW/cnvrs.txt', sep = '\t', row.names = T, col.names = T)

ah <- AnnotationHub::AnnotationHub()
ahDb <- AnnotationHub::query(ah, pattern = c("Sus csrofa", "EnsDb"))
ahDb
ahEdb <- ahDb[["AH109389"]]
## loading from cache
## require("ensemblDb")
```

```

bt.genes <- ensemblDb::genes(ahEdb)
GenomeInfoDb::seqlevelsStyle(bt.genes) <- "UCSC"
bt.genes
sel.genes <- subset(bt.genes, seqnames %in% paste0("chr", 1:26))
sel.genes <- subset(sel.genes, gene_biotype == "protein_coding")
sel.cnvrs <- subset(cnvrs, seqnames %in% paste0("chr", 1:26))
cnvrs <- populationRanges(cnv.calls, density=0.1, est.recur=TRUE)
subset(cnvrs, pvalue < 0.05)
cnvrs

olaps <- GenomicRanges::findOverlaps(sel.genes, sel.cnvrs, ignore.strand=FALSE)
qh <- S4Vectors::queryHits(olaps)
sh <- S4Vectors::subjectHits(olaps)
cgenes <- sel.genes[qh]
cgenes$type <- sel.cnvrs$type[sh]
subset(cgenes, select = "type")
cnvOncoPrint(cnv.calls, cgenes, top.features = 35, top.samples = 100)

###

remotes::install_github(repo = "JH-Zhou/HandyCNV")
library(HandyCNV)
cnv_penn_ars <- ("C:/Users/Siroj/Desktop/статья_cnv/Salsk/Salsk2.goodcnv")
cnv_penn_ars <- cnv_clean(penncnv = cnv_penn_ars, #replace input file when running your own data
                        penn_id_sep = "cnv/", #take the unique pattern as a separator to extract the sample ID
                        drop_length = 5, # the maximum CNV length threshold (in Mb); CNVs larger than this value will be deleted
                        folder = " /Desktop/статья_cnv/LW/LW_Filter/") #create a new folder to save results
sheep <- get_refgene(gene_version = "Sheep_Oar_v3.1_UCSC")
#sheep <- read.table("SSC.refGene.txt")
cnv_summary_plot(clean_cnv = cnv_penn_ars, #standard file was generated by 'cnv_clean' function in section 1
                length_group = c(0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 1, 2, 5), # optional, set group of vectors to divide CNV length, unit is Mb. such as vector of 'c(0.05, 0.3, 1)', means divide the CNV length into four group: '<0.05Mb', '0.05 - 0.3Mb', '0.3-1Mb' and '>1Mb', maximum can accept 11 valueses
                col_0 = "red", #optional, customize color for 0 copy CNV

```



```

col_1 = "black", #optional, customize color for o copy CNV
col_3 = "purple", #optional, customize color for o copy CNV
col_4 = "blue", #optional, customize color for o copy CNV
plot_sum_1 = TRUE, #optional, make sum combination plot 1
height_sum1 = 26, #optional, adjust the size of sum plot 1
width_sum1 = 20, #optional, adjust the size of sum plot 1
plot_sum_2 = TRUE, #optional, make sum combination plot 2
height_sum2 = 20, #optional, adjust the size of sum plot 2
width_sum2 = 27, #optional, adjust the size of sum plot 2
folder = "C:/Users/Siroj/Desktop/статья_cnv/Salsk/")
#> Deviding the length group...
#> Plotting...
#> The CNV distribution plot has been saved in the 'cnv_summary_plot_penn' directory.
#> The CNV length group plot was saved in the 'cnv_summary_plot_penn' directory.
#> The box plot of CNV lengths was saved in the 'cnv_summary_plot_penn' directory.
#> Distribution of CNV type on chromosome plot was saved in working directory.
#> The number of CNV in Individual plot was saved in working directory.
#> cnv_summary_plot_1 was saved in working directory.
#> cnv_summaray_polt_2 was saved in working directory.
#> Task done.

cnvr_penn_ars <- call_cnvr(clean_cnv = cnv_penn_ars, #standard file was generated by 'cnv_clean' function in section 1
                           chr_set = 26, #Set the maximum number of chromosomes (29 autosomes for cattle)
                           folder = "C:/Users/Siroj/Desktop/статья_cnv/Salsk/")
call_gene(refgene = sheep, #standard file was generated by 'get_refgene' function in section 5.1
          interval = cnvr_penn_ars, #optional, standard file was generated by 'call_cnvr' function in section 4
          clean_cnv = cnv_penn_ars, #optional, standard file was generated by 'cnv_clean' function in section 1
          folder = "/Desktop/статья_cnv/LW/Geness/")
cnvr_plot(cnvr = " /Desktop/статья_cnv/LW/cnvr.txt", #standard file was generated by 'call_cnvr' function in section 4
          assembly = "SSC_11.0_UCSC",
          loss_col = "green2", #optional, adjust the color of Loss type of CNVR
          gain_col = "red", #optional, adjust the color of Gain type of CNVR

```

```

mixed_col = "black", #optional, adjust the color of Mixed type of CNVR
legend_x = 300, #optional, adjust the horizontal position of legend
legend_y = 30, #optional, adjust the vertical position of legend
#height_1 = 14, #optional, adjust the height of CNVR plot
#width_1 = 10, #optional, adjust the width of CNVR plot
folder = "C:/Users/Siroj/Desktop/статья_cnv/Salsk/")

```

```

library(data.table)
setwd("/Desktop/статья_cnv/LW/LW_Filter/")
mydata <- read.table("LW_filter.goodcnv")
data1<-strsplit(as.character(unique(mydata[,1])),":")
data2<-do.call(rbind,data1)
data3<-strsplit(data2[,2],"-")
data4<-do.call(rbind,data3)
data_5 <- cbind(data2[,1],data4)
write.table(data_5,"cnv.bed",sep="\t",quote=FALSE,row.names=FALSE,col.names=FALSE)
library(GenomicRanges)
CNV <- data_5
str(CNV)
start<-as.numeric(CNV[,2])
end<-as.numeric(CNV[,3])
CNV_GRanges<-GRanges(seqnames=CNV[,1],IRanges(start,end))
CNV_segment<-disjoin(CNV_GRanges)
#Get the raw CNV segments
raw_CNV_segment<-data.frame(seqnames=seqnames(CNV_segment),start=start(CNV_segment),end=end(CNV_segment))
#Filter the CNV segments smaller than 50 bp (see Note 2)
good_CNV_segment <- raw_CNV_segment[which(raw_CNV_segment[,3]-raw_CNV_segment[,2] >= 50),]
write.table(good_CNV_segment," good_CNV_segment.bed",sep="\t",quote=FALSE,row.names=FALSE,col.names=FALSE)

```

### Приложение 3. Скрипт для расчета ROH и визуализации для модели

```
library(tidyverse)
library(data.table)
library(snpStats)
library(ggplot2)
# check SNP call rates -----
# plink name
landr_plink_name <- "/Desktop/rich/LR "
# read merged plink data
landr_bed <- paste0(landr_plink_name, ".bed")
landr_bim <- paste0(landr_plink_name, ".bim")
landr_fam <- paste0(landr_plink_name, ".fam")
full_sample <- read.plink(landr_bed, landr_bim, landr_fam)

# snp genotyping summary
snp_sum <- col.summary(full_sample$genotypes) %>%
  as_tibble(rownames = "snp.name") %>%
  left_join(full_sample$map, ., by = "snp.name")
# filter call rate and monomorphic snps
snp_sum %>%
  as_tibble() %>%
  filter((Call.rate < 0.95) | (MAF == 0)) %>%
  .$snp.name %>%
  write_lines("/Desktop/rich/imp_low_call_snps095.txt")

# filter low call rate SNPs
system(paste0("plink1.exe --bfile LR --exclude imp_low_call_snps095.txt ",
  "--make-bed --out LR_1"))
system(paste0("plink1.exe --bfile LR_Voron1 --out output/ROH/roh ",
  # "--keep output/ROH/ids_surv.txt",
  "--homozyg --homozyg-window-snp 50 --homozyg-snp 50 --homozyg-kb 1
200 ",
  "--homozyg-gap 300 --homozyg-density 200 --homozyg-window-missing 2
",
  "--homozyg-het 1 ",
  "--homozyg-window-het 1 "))
library("GenomicRanges")
rich <- fread("output/ROH/roh.hom", stringsAsFactors = T)
rich1 <- GRanges(rich)
BiocManager::install("biocLite", force = TRUE)
```

```

library(HandyCNV)
roh_visual(clean_roh = "output/ROH/roh.hom",
            individual_id = "203985370027_R04C02",
            #height_1 = 14, #optional
            #width_1 = 22, #optional
            col_short = "green", #optional
            col_mid = "yellow", #optional
            col_long = "red")

####
library(dplyr)
chr_data <- fread("sscr11.txt", stringsAsFactors = T)
colnames(chr_data) <- c("CHR", "size_BP", "Scaffolds")
chr_data <- chr_data %>%
  mutate(size_KB = size_BP / 1000)

autosomal_genome_size <- chr_data %>%
  .[2:19, ] %>%
  summarise(sum_KB = sum(size_KB)) %>%
  as.numeric()
file_path <- "output/ROH/roh.hom"
roh_lengths <- fread(file_path)

# max ROH length
roh_lengths[which.max(roh_lengths$KB), ]

# ROH overview -----
# descriptive ROH statistics
num_roh_per_ind <- roh_lengths %>% group_by(IID) %>% tally()
summary(num_roh_per_ind$n)
sd(num_roh_per_ind$n)

# inbreeding coefficients
froh <- roh_lengths %>%
  dplyr::group_by(IID) %>%
  dplyr::summarise(KBAVG = mean(KB), KBSUM = sum(KB)) %>%
  mutate(FROH = KBSUM/autosomal_genome_size) %>%
  mutate(FROH_cent = FROH - mean(FROH))
mean(froh$FROH)
range(froh$FROH)

```

```

# longest ROH
roh_lengths %>% arrange(-KB)

# longest ROH proportional to chr size
chr_sizes <- chr_data %>% .[-1, ] %>%
  mutate(CHR = str_replace(CHR, "Chromosome ", "")) %>%
  mutate(CHR = as.integer(CHR))

roh_lengths %>%
  arrange(desc(KB)) %>%
  left_join(chr_sizes, by = "CHR") %>%
  mutate(prop_chr = KB / size_KB) %>%
  arrange(desc(prop_chr))
#plot(froh$KBSUM, num_roh_per_ind$n)

# ROH length and abundance in the 1% least and most inbred individuals
num_roh_per_ind %>%
  left_join(froh) %>%
  # top_frac(-0.01, FROH) %>% # top 1% least inbred individuals
  top_n(-7, FROH) %>% # top 1% most inbred individuals
  summarise(mean(n), mean(KBAVG))

# Supplementary Figure FROH / ROH across individuals -----
#~~ ROH for some individuals -----
# Figure 1A, ROH for most and least inbred individuals

all_roh <- roh_lengths %>%
  group_by(IID) %>%
  summarise(sum_roh = sum(KB)) %>%
  ungroup() %>%
  arrange(desc(sum_roh))

longest_roh <- all_roh %>%
  top_n(7)
shortest_roh <- all_roh %>%
  top_n(-7)
num_ind <- 14

extreme_roh <- rbind(longest_roh, shortest_roh)

df <- roh_lengths %>%

```

```

mutate(POS1 = POS1 / 1e+6,
       POS2 = POS2 / 1e+6,
       MB = KB / 1000)

df <- df %>% filter(IID %in% extreme_roh$IID) %>%
mutate(IID = factor(IID, levels = extreme_roh$IID))

yax <- data.frame(IID = fct_inorder(levels(df$IID))) %>%
mutate(yax = seq(from = 2,
                 to = 2*length(unique(df$IID)),
                 by = 2))

df <- left_join(df, yax, by = "IID")

shade <- df %>%
group_by(CHR) %>%
summarise(min = min(POS1), max = max(POS2)) %>%
mutate(min = case_when(CHR == 2 | CHR == 4 | CHR == 6 | CHR == 8 | CHR == 1
0 |
                     CHR == 12 | CHR == 14 | CHR == 16 | CHR == 18 ~ 0,
                     TRUE ~ min)) %>%
mutate(max = case_when(CHR == 2 | CHR == 4 | CHR == 6 | CHR == 8 | CHR ==
10 |
                     CHR == 12 | CHR == 14 | CHR == 16 | CHR == 18 ~ 0,
                     TRUE ~ max))

col <- c("#4c566a", "#d8dee9")
col <- c("#1E3231", "#9aadbfb")
chr_names <- as.character(1:18)
names(chr_names) <- as.character(1:18)
chr_names[c(3,5, 7, 9, 11, 13, 15, 17)] <- ""
source("theme_simple.R")
df %>%
filter(MB > 5) %>%
filter(CHR %in% 1:26) %>%
ggplot() +
geom_rect(data=shade, aes(xmin=min, xmax=max, ymin=0, ymax=num_ind*2 +
1),
         alpha=0.5, fill = "#e4e4e4") + # "#f7f7f7" "#e4e4e4"
geom_hline(data = yax, aes(yintercept = yax), color = "#d8dee9", size = 0.4) +
geom_rect(aes(xmin = POS1, xmax = POS2, ymin = yax - 0.5, ymax = yax + 0.9,

```

```

        fill = as.factor(CHR)), col = "grey", size = 0, alpha = 1) +
scale_fill_manual(values = rep(col, 18)) +
scale_color_manual(values = rep(col, 18)) +
scale_y_reverse(expand = c(0, 0)) +
theme_simple(axis_lines = TRUE, grid_lines = FALSE, base_size = 13, base_famil
y = "Helvetica") +
facet_grid(~CHR,scales = 'free_x', space = 'free_x', switch = 'x',
          labeller = as_labeller(chr_names)) +
theme(#strip.placement = 'outside',
      axis.text.x = element_blank(),
      axis.ticks.x = element_blank(),
      axis.ticks.y = element_blank(),
      panel.spacing = unit(0, "lines"),
      plot.margin = margin(r = 0.5, l = 0.1, b = 0.1, t = 0.1, unit = "cm"),
      axis.line.x = element_blank(),
      legend.position="none",
      axis.title.x = element_text(margin=margin(t=0)),
      axis.title.y = element_text(margin=margin(r=0)),
      axis.text.y = element_text(colour = "white"),
      axis.line.y = element_blank()) +
coord_cartesian(clip = 'off') +
xlab("Chromosome") +
ylab("Individuals") -> ROH_per_ind
ROH_per_ind
#ggsave("figs/fig1b_roh_per_ind_5Mb.jpg", ROH_per_ind, width = 7, height = 3)

# needs to be a grob to display axis labels correctly
pg <- ggplotGrob(ROH_per_ind)

for(i in which(grepl("strip-b", pg$layout$name))){
  pg$grobs[[i]]$layout$clip <- "off"
}

ROH_per_ind_grob <- as.ggplot(pg)
ggsave("figs/Fig1A.pdf", ROH_per_ind_grob, width = 6, height = 2.5)
#ggsave("figs/fig1b_roh_per_ind_5Mb.jpg", ROH_per_ind_grob, width = 6, heigh
t = 3.5)

#p1 / p_roh_classes + plot_layout(heights = c(1, 0.8))

# ROH classes -----

```

```

# expected ROH length using cM/Mb from Johnston et al (2016)
length_dist <- data.frame(g = c(1, 2,2^2, 2^3, 2^4,2^5,2^6,2^7,2^8,2^9,2^10,2^11,2^12,2^13)) %>%
  mutate(ROH_length_cM = 100 / (2*g)) %>%
  mutate(ROH_length_Mb = ROH_length_cM * 0.7816743)

prop_IBD_df <- roh_lengths %>%
  mutate(length_Mb = KB/1000) %>%
  mutate(class = case_when(#length_Mb >= 39.083715000 ~ 1,
    # length_Mb < 39.083715000 & length_Mb >= 19.541857500 ~ 2,
    length_Mb >= 19.541857500 ~ 2,
    length_Mb < 19.541857500 & length_Mb >= 9.770928750 ~ 4,
    #length_Mb < 9.770928750 & length_Mb >= 6.513952500 ~ 6,
    length_Mb < 9.770928750 & length_Mb >= 4.885464375 ~ 8,
    # length_Mb < 4.885464375 & length_Mb >= 3.908371500 ~ 10,
    length_Mb < 4.885464375 & length_Mb >= 2.442732188 ~ 16,
    length_Mb < 2.442732188 & length_Mb >= 1.2 ~ 32)) %>% # 0.610683047 1.22
1366094
  mutate(length_class = case_when(
    #class == 1 ~ ">39 (1G)",
    #class == 2 ~ ">19.5-39 (2g)",
    class == 2 ~ "2g",
    class == 4 ~ "2-4g",
    #class == 6 ~ "6.5-9.7 (6G)",
    class == 8 ~ "4-8g",
    # class == 10 ~ "3.9-4.9 (10G)",
    class == 16 ~ "8-16g",
    class == 32 ~ "16-32g"
    # class == 128 ~ "0.6-0.3 (128G)"
  ))

fwrite(roh,"LR.hom", sep = "\t")
# add IBD of non-ROH snps if wanted
# bind_rows(homs)

prop_IBD_df_with_0 <- prop_IBD_df %>%
  # add missing length classes as 0
  ungroup() %>%
  tidyr::complete(length_class, nesting(IID)) %>%
  mutate(class = ifelse(is.na(class), length_class, class)) %>%

```



```

mutate(prop_IBD = ifelse(is.na(prop_IBD), 0, prop_IBD))

prop_IBD_df_with_0 %>%
  group_by(length_class) %>%
  summarise(mean(prop_IBD))

prop_IBD_df_with_0 %>%
  group_by(length_class) %>%
  #summarise(sum(prop_IBD > 0)/ 5925)
  filter(prop_IBD > 0) %>%
  summarise(mean(prop_IBD))

library(viridis)
library(ggghalves)
col_pal <- plasma(6)
col_pal <- paste0("#", (c("21295c","204683","1763a1","0a96d6","65bee2","BAE2
F2")))
col_pal <- paste0("#", (c("01161e","124559","598392","84a3a1","aec3b0","eff6e
0")))
#col_pal <- paste0("#", c("432371","714674","9f6976","cc8b79","e39d7a","faae7
b"))
p_roh_length <- prop_IBD_df_with_0 %>%
  mutate(prop_IBD = prop_IBD * 100) %>%
  ggplot(aes(length_class, prop_IBD, fill = length_class)) +
  geom_half_point(side = "l", shape = 21, alpha = 0.3, stroke = 0.1, size = 2, color =
"#4c566a",
               transformation_params = list(height = 0, width = 1.3, seed = 1)) +
  geom_half_boxplot(side = "r", outlier.color = NA,
                   width = 0.6, lwd = 0.3, color = "black",
                   alpha = 0.8) +
  theme_simple(axis_lines = TRUE, grid_lines = FALSE, base_size = 13,
               base_family = "Helvetica") +
  ylab("% genome") +
  scale_x_discrete(guide = guide_axis(n.dodge = 2)) +
  scale_fill_manual(values = col_pal, name = "ROH class (Mb)") +
  theme(legend.position = "none",
        plot.margin = margin(r = 0.5, l = 0.1, b = 0.1, t = 0.1, unit = "cm"),
        #axis.ticks.x = element_blank(),
        axis.title=element_text(size = rel(1.1)),
        axis.text = element_text(color = "black")) +
  xlab("ROH length class in Mb (~ generations to MRCA)")

```

p\_roh\_length

```
roh <- prop_IBD_df
library(tidyverse)
plink_roh_names <- c("FID", "IID", "PHE", "CHR", "SNP1", "SNP2", "POS1", "POS2",
",
", "KB", "NSNP", "DENSITY", "PHOM", "PHET", "length_Mb", "class", "length_class")
if(length(colnames(roh)) == length(plink_roh_names)){
  cat("ROH with input file in PLINK format was detected, converting to HandyCNV standard formats\n")
  colnames(roh) <- c("FID", "Sample_ID", "PHE", "Chr", "Start_SNP", "End_SNP", "Start", "End", "Length", "Num_SNP", "DENSITY", "PHOM", "PHET", "length_Mb", "class", "length_class")
  handycnv_name <- c("Sample_ID", "Chr", "Start", "End", "Num_SNP", "Length", "Start_SNP", "End_SNP", "length_class")
  roh <- roh %>% select(handycnv_name)
  roh$Chr <- as.numeric(roh$Chr)
  roh <- dplyr::filter(roh, roh$Chr %in% c(1:max(roh$Chr)))
}
individual_id <- "202179440023_R01C02"
cnv_indiv <- roh[which(roh$Sample_ID == individual_id),]
chr_coord <- data.frame("Chr" = seq(1,max(as.integer(roh$Chr),1)))
chr_coord$x <- max(roh$End) #adjust x for geom_text
chr_coord$y <- (chr_coord$Chr-1)*5 + 1 #adjust y for geom_text
cnv_indiv_coord <- merge(cnv_indiv, chr_coord, all.x = TRUE, sort = FALSE)

#4.
folder = "/Desktop/rich/roh_visual/"
width_1 = 26
height_1 = 10
indiv_name <- paste0(folder, "/ROH_of_", individual_id, ".png")
indiv_title <- paste0("ROH Distribution of Individual ", individual_id)
png(res = 350, filename = indiv_name, width = width_1, height = height_1, units = "cm")
#fruit <- unique(roh$Sample_ID)
#for ( i in fruit){
#individual_id=i
#cnv_indiv <- roh[which(roh$Sample_ID == individual_id),]
#chr_coord <- data.frame("Chr" = seq(1,max(roh$Chr),1))
```

```

#chr_coord$x <- max(roh$End) #adjust x for geom_text
#chr_coord$y <- (chr_coord$Chr-1)*5 + 1 #adjust y for geom_text
#cnv_indiv_coord <- merge(cnv_indiv, chr_coord, all.x = TRUE, sort = FALSE)
#}
indiv_plot <- ggplot(cnv_indiv_coord) +
  geom_rect(aes(xmin = Start/1000000, xmax = End/1000000, ymin = (Chr-1)*5, y
max = (Chr-1)*5 + 3, fill = length_class)) +
  #scale_fill_gradient(low = col_short, high = col_long) +
  scale_fill_manual(values = c("#EF3038", "#006633", "#4169E1", "#7FFFD4", "#FE
FE22")) +
  geom_text(aes(x = x/1000000, y, label = Chr), size = 4) +
  theme_bw() +
  scale_y_continuous(breaks = seq(0, max(roh$End)/1000000, by = 20), labels = N
ULL) +
  scale_x_continuous(breaks = seq(0, max(roh$End)/1000000, by = 10)) +
  labs(x = "Physical Position (Mb)", y = "Autosome", title = indiv_title, fill = "Length
")
indiv_plot
dev.off()
cat("Task done, plot was stored in working directory.\n")
roh <- read.table("LR_Voron_roh.hom", sep = "\t", header = T)
fwrite(roh, "LR_Voron.hom", sep = "\t", col.names = T)

```

```

roh <- read.table("D7_HBD.hom.txt", sep = "\t", header = T)
cnv_visual <- function(individual_id) {
  cnv_indiv <- roh[which(roh$Sample_ID == individual_id),]
  chr_coord <- data.frame("Chr" = seq(1, max(as.integer(roh$Chr), 1)))
  chr_coord$x <- max(roh$End) #adjust x for geom_text
  chr_coord$y <- (chr_coord$Chr-1)*5 + 1 #adjust y for geom_text
  cnv_indiv_coord <- merge(cnv_indiv, chr_coord, all.x = TRUE, sort = FALSE)

```

#4.

```

folder <- "/Desktop/rich/roh_visual/"
indiv_name <- paste0(individual_id, ".jpeg")
indiv_title <- paste0("ROH Distribution of Individual ", individual_id)
jpeg(filename = paste0(folder, "/", indiv_name), width = 500, height = 500, unit
s = "px")
col_0 = "#006633"

```

```

col_1 = "#EF3038"
col_2 = "#4169E1"
col_3 = "#7FFFD4"
col_4 = "#FF7F24"
col_5 = "#838B8B"
color_copy <- c("R1" = col_0,
               "R2" = col_1,
               "R3" = col_2,
               "R4" = col_3,
               "R5" = col_4,
               "R6" = col_5)
indiv_plot <- ggplot(cnv_indiv_coord) +
  geom_rect(aes(xmin = Start/1000000, xmax = End/1000000, ymin = (Chr-1)*5,
               ymax = (Chr-1)*5 + 3, fill = length_class)) +
  #scale_fill_gradient(low = col_short, high = col_long) +
  scale_fill_manual(values = color_copy) +
  geom_text(aes(x = x/1000000, y, label = Chr), size = 4) +
  theme_bw() +
  scale_y_continuous(breaks = seq(0, max(roh$End)/1000000, by = 20), labels =
NULL) +
  scale_x_continuous(breaks = seq(0, max(roh$End)/1000000, by = 10)) +
  theme(legend.position="none") +
  labs( title = NULL)
print(indiv_plot)
dev.off()

}

cnv_visual("DD52077")

table(roh$length_class)
a <- roh[roh$length_class == "R6",]
a <- roh[roh$Sample_ID == "DD52077",]
table(a$length_class)
folder <- "C:/Users/Siroj/Desktop/rich/roh_visual/"
indiv_name <- paste0("LR", individual_id, ".jpeg")
indiv_title <- paste0("ROH Distribution of Individual ", individual_id)
jpeg(filename = paste0(folder, "/", indiv_name), width = 500, height = 500, units
= "px")
col_0 = "#006633"
col_1 = "#EF3038"

```

```

col_2 = "#4169E1"
col_3 = "#7FFFD4"
col_4= "#FEFE22"
color_copy <- c("2g" = col_0,
               "2-4g" = col_1,
               "4-8g" = col_2,
               "8-16g" = col_3,
               "16-32g" = col_4)
indiv_plot <- ggplot(cnv_indiv_coord) +
  geom_rect(aes(xmin = Start/1000000, xmax = End/1000000, ymin = (Chr-1)*5, y
max = (Chr-1)*5 + 3, fill = length_class)) +
  #scale_fill_gradient(low = col_short, high = col_long) +
  scale_fill_manual(values = color_copy) +
  #geom_text(aes(x = x/1000000, y, label = Chr), size = 4) +
  theme_bw() +
  scale_y_continuous(breaks = seq(0, max(roh$End)/1000000, by = 20), labels = N
ULL) +
  scale_x_continuous(breaks = seq(0, max(roh$End)/1000000, by = 10)) +
  theme(legend.position="none") +
  labs( title = NULL)
indiv_plot
dev.off()
print(indiv_plot)

```

#### Приложение 4. Скрипт для расчета HBD и для графиков

```
library(RZooRoH)
library(data.table)
library(psych)# библиотека для описательной статистике
library(dplyr)
#В рабочую папку загружаются файлы LLDW (2"FAM", "MAP", "PED") и plink1
# на основе папки LLDW.MAP создаем Duroc.txt, Landrace.txt, Large_White.txt
, WB.txt

system("plink1.exe --bfile Dur_leg_69 --keep dom_ID.txt --make-bed --out DOM"
)#
system("plink1.exe --bfile doct1 --keep imp1_ID.txt --make-bed --out IMP1")
system("plink1.exe --bfile doct1 --keep imp2_ID.txt --make-bed --out IMP2")

system("plink1.exe --bfile Dur_leg_692 --recode oxford --autosome --out Dur_leg
_69")
system("plink1.exe --bfile IMP1 --recode oxford --autosome --out IMP1")
system("plink1.exe --bfile IMP2 --recode oxford --autosome --out IMP2")

file1 <- ("Dur_leg_69.gen")
file2 <- ("IMP1.gen")
file3 <- ("IMP2.gen")

BBB_GP1 <- zoodata(genofile = file1, zformat = "gp")
BBB_GP2 <- zoodata(genofile = file2, zformat = "gp")
BBB_GP3 <- zoodata(genofile = file3, zformat = "gp")

LRR=zoomodel()
LRR

bbb_LW_1 <- zoorun(LRR, BBB_GP1,nT = 5)
bbb_LW_2 <- zoorun(LRR, BBB_GP2, nT=5)
bbb_LW_3 <- zoorun(LRR, BBB_GP1, nT=5)

save(bbb_LW_1,file="LW_1_model.RData")# сохранить в папке результаты, по
считанные на основе модели bbb_LW <- zoorun(LRR, BBB_GP1)
save(bbb_LW_B,file="LW_B_model.RData")# сохранить в папке результаты, по
считанные на основе модели bbb_LR <- zoorun(LRR, BBB_GP2)

load("LW_imp1_model.RData") # загрузить
```

```

load("LW_imp2_model.RData") # загрузить

load("RU_EU_model.RData")
load("WB_AS_model.RData")
load("WB_EU_model.RData")
par(mar=c(1.1,4.1,1.1,2.1))
par(mfcol = c(3, 1))
zooplot_prophbd(list(dom= bbb_LW_1), cols = 'tomato',style = 'boxplot')#предст
авляет долю генома, связанного с различными классами HBD на уровне поп
уляции
zooplot_prophbd(list(imp1= bbb_LW_2), cols = 'tomato',style = 'boxplot')
zooplot_prophbd(list(imp2= bbb_LW_3), cols = 'tomato',style = 'boxplot')

#Доля генома, связанная с различными классами HBD для четырех популяци
й
zooplot_prophbd(list(Dom=bbb_LW_1,IMP1= bbb_LW_2,IMP2= bbb_LW_3, style
='lines'))

par(mar=c(1.1,4.1,1.1,2.1))
par(mfcol = c(3, 1))
zooplot_prophbd(list(Домашняя=bbb_LW_1),style='lines', cumulative = TRUE)

popDOM <- list(Отечественная= bbb_LW_1)
popIMP1<-list(Импортная1=bbb_LW_2)
popIMP2 <- list(Импортная2= bbb_LW_3)

par(mfcol = c(3, 1))
zooplot_individuals(popDOM, cumulative = TRUE)# те же результаты, что и фун
кция zooplot_prophbd, но на индивидуальном уровне. Каждая особь предста
влен линией (без диаграмм или диаграмм).
zooplot_individuals(popIMP1, cumulative = TRUE)
zooplot_individuals(popIMP2, cumulative = TRUE)

pop <- list(Домашняя= bbb_LW_1,Импортная1=bbb_LW_2, Импортная2= bbb_
LW_3)
zooplot_partitioning(pop, randomids = TRUE, nrandom = c(20,20,20), plotids = F
ALSE,
                ylim=c(0,0.5), nonhbd = FALSE)
#представляет для каждой особи долю генома в каждом классе HBD. Каждда
я особь представлена в виде столбчатой диаграммы. Общая высота предста

```

```

вляет собой общий уровень аутозиготности. Вклад каждого класса HBD (в% о
т генома) представлен в виде полосы разного цвета
par(mar=c(4.1,4.1,1.1,1.1))
par(mfcol = c(3, 1))
x <- 1-bbb_LW_1@realized[,10]# рассчитываем доля генома HBD
hist(x,nc=30,main="",ylab= "Частота",xlim=c(0.15,0.4),col='tomato')# строим гис
тограмму для инбридинга
summary(a)
y <- 1-bbb_LW_2@realized[,10]
hist(y,nc=30,main="",ylab= "Частота",xlim=c(0.15,0.4),col='tomato')

a <- 1-bbb_LW_3@realized[,10]
hist(a,nc=30,main="",ylab= "Частота",xlim=c(0.15,0.4),col='tomato')

# Таблица 1. Сводная статистика реализованной аутозиготности в каждом кл
асс HBD (в процентах от генома)
bbb_LW_1@realized # Значения [i, j] - это вклады класса j (усредненные по вс
ем позициям) в геном индивидуума i по всему геному.
df <-realized(bbb_LW_1)
realized(bbb_LW_1)#извлечь всю таблицу результатов
write.table(df,"dom_model.txt",append = F,quote = F,sep = " ", col.names = T, ro
w.names = T)# записать все результаты в таблицу
df_D7<-read.table('dom_model.txt', sep = ' ', header = T)# записали как таблицу
для работы
df_D7
table1 <- describe(x = df[,c(2:10)], fast = T)# описательная (фаст) по всей выбор
ке
write.table(table1,"dom_AYT_STAT.txt",append = F,quote = F,sep = " ", col.name
s = T, row.names = T)# результаты по сводной статистике аписали в таблицу L
W_AYT_STAT.txt
table1

df <-realized(bbb_LW_2)
realized(bbb_RU_EU)#извлечь всю таблицу результатов
write.table(df,"RU_EU_model.txt",append = F,quote = F,sep = " ", col.names = T,
row.names = T)# записать все результаты в таблицу
df_RU_EU<-read.table('RU_EU_model.txt', sep = ' ', header = T)# записали как т
аблицу для работы
df_RU_EU
table2 <- describe(x = df[,c(2:10)], fast = T)# описательная (фаст) по всей выбор
ке

```



```
write.table(table2,"LW1_AYT_STAT.txt",append = F,quote = F,sep = " ", col.names = T, row.names = T)# результаты по сводной статистике аписали в таблицу LR_AYT_STAT.txt
```

```
bb_WB_AS@realized # Значения [i, j] - это вклады класса j (усредненные по всем позициям) в геном индивидуума i по всему геному.
df <- realized(bbb_LW_3)
realized(bbb_WB_AS)#извлечь всю таблицу результатов
write.table(df,"WB_AS_model.txt",append = F,quote = F,sep = " ", col.names = T, row.names = T)# записать все результаты в таблицу
df_WB_AS<-read.table('WB_AS_model.txt', sep = ' ', header = T)# записали как таблицу для работы
df_WB_AS
table3 <- describe(x = df[,c(2:10)], fast = T)# описательная (фаст) по всей выборке
write.table(table3,"LW2_AYT_STAT.txt",append = F,quote = F,sep = " ", col.names = T, row.names = T)# результаты по сводной статистике аписали в таблицу LR_AYT_STAT.txt
```

```
bb_WB_EU@realized # Значения [i, j] - это вклады класса j (усредненные по всем позициям) в геном индивидуума i по всему геному.
df <- realized(bbb_WB_EU)
realized(bbb_WB_EU)#извлечь всю таблицу результатов
write.table(df,"WB_EU_model.txt",append = F,quote = F,sep = " ", col.names = T, row.names = T)# записать все результаты в таблицу
df_WB_EU<-read.table('WB_EU_model.txt', sep = ' ', header = T)# записали как таблицу для работы
df_WB_EU
table4 <- describe(x = df_WB_EU[,c(2:10)], fast = T)# описательная (фаст) по всей выборке
write.table(table4,"WB_EU_AYT_STAT.txt",append = F,quote = F,sep = " ", col.names = T, row.names = T)# результаты по сводной статистике аписали в таблицу LR_AYT_STAT.txt
```

```
#Сегменты HBD, идентифицированные с помощью алгоритма Витерби,
dim(bbb_D7@hbdseg)[1]
table5 <- bbb_D7@hbdseg
write.table(table5,"D7_длины_HBD.txt",append = F,quote = F,sep = " ", col.names = T, row.names = T)
summary(bbb_D7@hbdseg$length)
summary(bbb_D7@hbdseg$number_snp)
```

```
dim(bbb_RU_EU@hbdseg)[1]
table6 <- bbb_RU_EU@hbdseg
write.table(table6,"RU_EU_длины_HBD.txt",append = F,quote = F,sep = " ", col.names = T, row.names = T)
summary(bbb_RU_EU@hbdseg$length)
summary(bbb_RU_EU@hbdseg$number_snp)
```

## Приложение 5. Скрипт для модели с наличием и отсутствием шишек на ногах свиней

```
# Загрузка пакетов
library(ripserr)
library(keras)
library(abind)
library(tensorflow)
library(magick)
library(parallel)
library(caret)
library(randomForest)
library(ggplot2)
library(rsample)

# Пути к папкам с изображениями
folder_0 <- /Leg/Picture_ROH/train/no_leg/" # Папка с изображениями класса 0
folder_1 <- /Leg/Picture_ROH/train/leg/" # Папка с изображениями класса 1

# Функция для загрузки и предобработки цветных изображений
preprocess_image <- function(image_path, target_size = c(128, 128)) {
  img <- image_read(image_path)
  img <- image_resize(img, paste0(target_size[1], "x", target_size[2], "!")) # Изменяем размер
  img <- image_data(img, channels = "rgb") # Получаем данные изображения в RGB
  img <- as.numeric(img) / 255 # Нормализация
  return(img)
}

# Загрузка изображений и создание меток
image_paths_0 <- list.files(folder_0, pattern = "\\\\.jpeg$", full.names = TRUE)
image_paths_1 <- list.files(folder_1, pattern = "\\\\.jpeg$", full.names = TRUE)
images_0 <- lapply(image_paths_0, preprocess_image)
images_1 <- lapply(image_paths_1, preprocess_image)

# Объединение изображений и создание меток
images <- c(images_0, images_1)
labels <- c(rep(0, length(images_0)), rep(1, length(images_1)))

# Преобразование списка изображений в массив
images <- abind(images, along = 0)
print(dim(images)) # Проверяем размерность массива

# Преобразование меток в one-hot encoding
labels <- to_categorical(labels, 2)
```

```

# Функция для случайного сэмплирования точек
sample_points <- function(image, num_samples = 460) {
  height <- dim(image)[1]
  width <- dim(image)[2]
  points <- which(image > 0, arr.ind = TRUE)
  sampled_indices <- sample(1:nrow(points), num_samples)
  return(points[sampled_indices, ])
}

# Функция для вычисления персистентных гомологий с использованием rips
err
compute_persistence <- function(image, num_samples = 460) {
  # Сэмплируем точки из изображения
  points <- sample_points(image, num_samples)
  # Вычисляем персистентные гомологии с помощью ripserr
  diagram <- vietoris_rips(points, dim = 1) # Вычисляем гомологии до 1-го изме
рения

  return(diagram)
}

# Функция для извлечения топологических фич
topologicalFeatures <- function(diagram) {
  numH0 <- length(which(diagram[, 1] == 0)) # Количество компонент связности
(H0)
  numH1 <- length(which(diagram[, 1] == 1)) # Количество петель (H1)
  if (numH1 != 0) {
    persistence <- diagram[diagram[, 1] == 1, 3] - diagram[diagram[, 1] == 1, 2] # Пр
одолжительность жизни петель
    medianPerst <- median(persistence) # Медиана продолжительности жизни п
етель
    maxPerst <- max(persistence) # Максимум продолжительности жизни пете
ль
  } else {
    medianPerst <- 0
    maxPerst <- 0
  }

  return(c(numH0, numH1, medianPerst, maxPerst))
}

# Параллельное вычисление персистентных гомологий с использованием m
clapply

```

```

num_cores <- detectCores() - 1 # Оставляем одно ядро свободным
print(paste("Используется", num_cores, "ядер")) # Проверяем количество ядер

# Параллельное вычисление с использованием mclapply
persistence <- mclapply(1:dim(images)[1], function(i) {
  compute_persistence(images[i, , ])
}, mc.cores = num_cores)
# Инициализация data frame для хранения топологических фич
features_df <- data.frame()
# Вычисление топологических фич для каждого изображения
for (i in 1:dim(images)[1]) {
  diagram <- persistence[[i]] # Персистентная диаграмма для i-го изображения
  features <- topologicalFeatures(diagram) # Извлекаем фичи
  features_df <- rbind(features_df, features) # Добавляем в data frame
}
# Задаем имена столбцов
colnames(features_df) <- c("H0", "H1", "H1MD", "H1ML")
# Нормализация фич
features_df <- scale(features_df)
print(head(features_df))
features_df <- features_df[,-c(1)]
# Разделение данных на тренировочную, валидационную и тестовую выборки
# Создаем data.frame с индексами и метками
data_df <- data.frame(
  images = 1:dim(images)[1], # Индексы изображений
  labels = apply(labels, 1, which.max) - 1 # Метки (0 или 1)
)
data_df <- data_df[sample(1:nrow(data_df)), ]
# Разделение данных на тренировочную, валидационную и тестовую выборки с использованием rsample
set.seed(123)
data_split <- initial_split(data_df, prop = 0.9, strata = "labels") # Стратификация по меткам
train_indices <- training(data_split)$images
temp_indices <- testing(data_split)$images

val_test_split <- initial_split(data_df[temp_indices, ], prop = 0.7, strata = "labels")

```

```

val_indices <- training(val_test_split)$images
test_indices <- testing(val_test_split)$images
# Тренировочная выборка
x_train <- images[train_indices, , ]
features_train <- features_df[train_indices, ]
y_train <- labels[train_indices, ]
# Валидационная выборка
x_val <- images[val_indices, , ]
features_val <- features_df[val_indices, ]
y_val <- labels[val_indices, ]
# Тестовая выборка
x_test <- images[test_indices, , ]
features_test <- features_df[test_indices, ]
y_test <- labels[test_indices, ]
# Проверка баланса классов
print(table(apply(y_train, 1, which.max) - 1)) # Тренировочная выборка
print(table(apply(y_val, 1, which.max) - 1)) # Валидационная выборка
print(table(apply(y_test, 1, which.max) - 1)) # Тестовая выборка
# Проверка корректности топологических фич
labels_numeric <- apply(y_train, 1, which.max) - 1 # Преобразуем метки в числовой формат
# 1. Визуализация персистентных диаграмм
par(mfrow = c(2, 2)) # Разделяем графическое окно на 2x2
for (i in 1:4) {
  plot(persistence[[i]], main = paste("Диаграмма для изображения", i))
}

# 2. Корреляция топологических фич с метками
correlation <- cor(features_train, labels_numeric)
print("Корреляция фич с метками:")
print(correlation)
# 3. Кластеризация топологических фич
set.seed(123)
kmeans_result <- kmeans(features_train, centers = 2) # 2 кластера для двух классов
print("Результаты кластеризации:")
print(table(kmeans_result$cluster, labels_numeric))
# 4. Обучение простой модели на топологических фичах
labels_factor <- as.factor(labels_numeric) # Преобразуем метки в фактор
model_lr <- train(
  x = features_train,

```

```

y = labels_factor,
method = "glm", # Логистическая регрессия
trControl = trainControl(method = "cv", number = 10) # Кросс-валидация
)
print("Результаты логистической регрессии:")
print(model_lr)
# 5. Сравнение распределений фич для разных классов
features_df_plot <- as.data.frame(features_train)
features_df_plot$class <- as.factor(labels_numeric)
ggplot(features_df_plot, aes(x = class, y = H1)) + geom_boxplot() + ggtitle("Boxplot для фичи H1")
ggplot(features_df_plot, aes(x = class, y = H1MD)) + geom_boxplot() + ggtitle("Boxplot для фичи H1MD")
ggplot(features_df_plot, aes(x = class, y = H1ML)) + geom_boxplot() + ggtitle("Boxplot для фичи H1ML")

# 6. Проверка значимости фич с помощью тестов
t_test_result_H1 <- t.test(features_train[labels_numeric == 0, "H1"], features_train[labels_numeric == 1, "H1"])
print("Результаты t-теста для фичи H1:")
print(t_test_result_H1)
t_test_result_H1MD <- t.test(features_train[labels_numeric == 0, "H1MD"], features_train[labels_numeric == 1, "H1MD"])
print("Результаты t-теста для фичи H1MD:")
print(t_test_result_H1MD)
t_test_result_H1ML <- t.test(features_train[labels_numeric == 0, "H1ML"], features_train[labels_numeric == 1, "H1ML"])
print("Результаты t-теста для фичи H1ML:")
print(t_test_result_H1ML)
# 7. Анализ важности фич в модели
model_rf <- randomForest(x = features_train, y = labels_factor, importance = TRUE)
print("Важность фич в случайном лесе:")
print(importance(model_rf))
varImpPlot(model_rf)
# Создание модели CNN с аугментацией данных
input_image <- layer_input(shape = c(128, 128, 3), name = "input_image") # 3 канала (RGB), размер 300x300
input_features <- layer_input(shape = c(3), name = "input_features") # 4 фичи: H0, H1, H1MD, H1ML
# Аугментация данных

```

```

data_augmentation <- keras_model_sequential() # %>%
  #layer_random_rotation(factor = 0.1) %>%
  #layer_random_zoom(height_factor = 0.1, width_factor = 0.1) %>%
  #layer_random_flip(mode = "horizontal")
cnn_output <- input_image %>%
  #data_augmentation() %>% # Применяем аугментацию
  layer_conv_2d(filters = 32, kernel_size = c(3, 3), activation = 'relu', kernel_regul
arizer = regularizer_l2(l = 0.0001)) %>%
  #layer_batch_normalization() %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_conv_2d(filters = 64, kernel_size = c(3, 3), activation = "relu", kernel_regu
larizer = regularizer_l2(l = 0.0001)) %>%
  layer_batch_normalization() %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_dropout(rate = 0.25) %>%
  layer_conv_2d(filters = 128, kernel_size = c(3, 3), activation = "relu", kernel_reg
ularizer = regularizer_l2(l = 0.0001)) %>%
  #layer_batch_normalization() %>%
  layer_conv_2d(kernel_size = c(3, 3), filters = 256,
    activation = "relu", padding = "valid") %>%
  layer_max_pooling_2d(pool_size = c(2, 2)) %>%
  layer_flatten() %>%
  layer_dense(units = 512, activation = "relu", kernel_regularizer = regularizer_l2(
l = 0.0001)) %>%
  layer_dropout(rate = 0.3) # %>%
#layer_dense(units = 100, activation = "relu", kernel_regularizer = regularizer_l2(
l = 0.0001)) %>%
#layer_dropout(rate = 0.6) # Увеличиваем Dropout
# Объединение CNN и топологических фич
combined <- layer_concatenate(list(cnn_output, input_features))

# Выходной слой
output <- combined %>%
  layer_dense(units = 2, activation = "sigmoid")
  #layer_dense(units = 2, activation = "softmax") # 2 класса: 0 и 1

# Создание модели
model <- keras_model(inputs = list(input_image, input_features), outputs = outp
ut)
binary_focal_crossentropy <- function(gamma = 2.0, alpha = 0.25) {
  function(y_true, y_pred) {

```



```

# Преобразование меток и предсказаний в нужный формат
y_true <- k_cast(y_true, 'float32')
y_pred <- k_cast(y_pred, 'float32')

# Вычисление бинарной кросс-энтропии
bce <- k_binary_crossentropy(y_true, y_pred)

# Вычисление весов для фокальной потери
p_t <- k_exp(-bce)
focal_loss <- k_pow(1 - p_t, gamma) * bce

# Взвешивание классов
alpha_factor <- k_ones_like(y_true) * alpha
alpha_t <- y_true * alpha_factor + (1 - y_true) * (1 - alpha_factor)

# Итоговая фокальная потеря
focal_loss <- alpha_t * focal_loss

return(k_mean(focal_loss, axis = -1))
}
}
focal_loss <- function(gamma = 2.0, alpha = 0.25) {
  function(y_true, y_pred) {
    y_true <- k_cast(y_true, 'float32')
    y_pred <- k_cast(y_pred, 'float32')

    bce <- k_binary_crossentropy(y_true, y_pred)
    p_t <- k_exp(-bce)
    focal_loss <- k_pow(1 - p_t, gamma) * bce
    alpha_factor <- k_ones_like(y_true) * alpha
    alpha_t <- y_true * alpha_factor + (1 - y_true) * (1 - alpha_factor)
    focal_loss <- alpha_t * focal_loss
    return(k_mean(focal_loss, axis = -1))
  }
}

# Компиляция модели
model %>% compile(
  optimizer = optimizer_adamax(learning_rate = 0.0001), # Используем Adam
  loss = binary_focal_crossentropy(gamma=2),#"binary_crossentropy", #binary_focal_crossentropy(gamma=2), focal_loss(gamma = 2.0, alpha = 0.25),

```

```

metrics = c("accuracy"))

# Ранняя остановка
early_stopping <- callback_early_stopping(
  monitor = "val_loss",
  patience = 5,
  restore_best_weights = TRUE
)

lr_schedule <- function(epoch, lr) {
  initial_lr <- 0.001 # Начальная скорость обучения
  drop_factor <- 0.5 # Снижение скорости
  epochs_drop <- 5 # Каждые 5 эпох будет уменьшаться скорость
  factor=0.1
  patience=5
  min_lr=1e-6

  # Расчет новой скорости обучения
  lr <- initial_lr * drop_factor ^ floor(epoch / epochs_drop)
  return(lr)
}

callback <- callback_learning_rate_scheduler(schedule = lr_schedule)
lr_schedule_exponential <- function(epoch, lr) {
  initial_lr <- 0.001 # Начальная скорость обучения
  decay_rate <- 0.01 # Коэффициент экспоненциального уменьшения
  # Расчет новой скорости обучения
  lr <- initial_lr * exp(-decay_rate * epoch)
  return(lr)
}

# Создаем callback для экспоненциального изменения скорости обучения
callback_exponential <- callback_learning_rate_scheduler(schedule = lr_schedule_exponential)

# Вычисление весов классов
class_counts <- table(apply(y_train, 1, which.max) - 1)
class_weights <- as.list(1 / (class_counts / sum(class_counts)))
print("Веса классов:")
print(class_weights)
train_samples <- length(train_indices)
valid_samples <- length(val_indices)
# Обучение модели с использованием валидационной выборки и class_weights
#rm(history)

```

```

set.seed(123)
history <- model %>% fit(
  x = list(x_train, features_train),
  y = y_train,
  epochs = 100,
  steps_per_epoch = as.integer(train_samples/10),
  #batch_size = 32,
  validation_data = list(list(x_val, features_val), y_val),
  validation_steps = as.integer(valid_samples/10),
  callbacks = callback)#,
  #callbacks = list(callback_exponential),
  #callbacks = list(early_stopping),
  #class_weight = class_weights) # Передаем веса классов
history
# Оценка модели на тестовой выборке
model %>% evaluate(list(x_test, features_test), y_test)
#save_model_hdf5(model, ' /Leg/model_leg.h5')
#save_model_tf(model, ' /Leg/model_leg_0.8.keras')
# Визуализация истории обучения
plot(history)
test_results <- model %>% evaluate(list(x_val, features_val), y_val)
print(paste("Test Loss:", test_results$loss))
print(paste("Test Accuracy:", test_results$accuracy))
# Предсказание классов на тестовых данных
y_pred <- model %>% predict(list(x_test, features_test))
# Преобразование предсказаний в метки классов
y_pred_classes <- apply(y_pred, 1, which.max) - 1 # Преобразуем one-hot encoding в метки (0 или 1)
y_true_classes <- apply(y_test, 1, which.max) - 1 # Преобразуем true labels в метки (0 или 1)
# Создание confusion matrix
conf_matrix <- table(Actual = y_true_classes, Predicted = y_pred_classes)
print("Confusion Matrix:")
print(conf_matrix)
# Визуализация confusion matrix с помощью ggplot2
conf_matrix_df <- as.data.frame(conf_matrix)
ggplot(conf_matrix_df, aes(x = Actual, y = Predicted, fill = Freq)) +
  geom_tile() +
  geom_text(aes(label = Freq), color = "white", size = 6) +
  scale_fill_gradient(low = "blue", high = "red") +
  labs(title = "Confusion Matrix", x = "Actual", y = "Predicted") +

```

```

theme_minimal()
# Использование confusionMatrix из пакета caret для более детального анализа
conf_matrix_caret <- confusionMatrix(as.factor(y_true_classes), as.factor(y_pred_classes), positive = "1")
print("Detailed Confusion Matrix (caret):")
print(conf_matrix_caret)
# Предсказания
predictions <- ensemble_model %>% predict(x_test) # x_test - тестовые изображения
# Для финальных предсказаний можно усреднять результаты
final_predictions <- apply(predictions, 1, function(x) which.max(x))
# Оценка модели
score <- ensemble_model %>% evaluate(x_test, y_test)
cat('Test loss:', score$loss, '\n')
cat('Test accuracy:', score$accuracy, '\n')
predictions <- model %>% predict(list(x_val, features_val))
predicted_classes <- apply(predictions, 1, which.max) - 1
true_classes <- apply(y_val, 1, which.max) - 1
confusion_matrix <- table(predicted_classes, true_classes)
print("Конфузионная матрица:")
print(confusion_matrix)

custom_objects <- list(
  'binary_focal_crossentropy_gamma___2_' = binary_focal_crossentropy(gamma = 2.0, alpha = 0.25)
)
# Load the model with the custom objects
model <- load_model_tf('/Leg/model_leg_0.8.keras', custom_objects = custom_objects)
model %>% evaluate(list(x_val, features_val), y_val)
#save_model_hdf5(model, '/Leg/model_leg.h5')
#save_model_tf(model, '/Leg/model_leg_0.8.keras')
# Визуализация истории обучения
plot(history)
test_results <- model %>% evaluate(list(x_test, features_test), y_test)
print(paste("Test Loss:", test_results$loss))
print(paste("Test Accuracy:", test_results$accuracy))
# Предсказание классов на тестовых данных
y_pred <- model %>% predict(list(x_test, features_test))
# Преобразование предсказаний в метки классов

```

```

y_pred_classes <- apply(y_pred, 1, which.max) - 1 # Преобразуем one-hot encoding в метки (0 или 1)
y_true_classes <- apply(y_test, 1, which.max) - 1 # Преобразуем true labels в метки (0 или 1)
# Создание confusion matrix
conf_matrix <- table(Actual = y_true_classes, Predicted = y_pred_classes)
print("Confusion Matrix:")
print(conf_matrix)

# Визуализация confusion matrix с помощью ggplot2
conf_matrix_df <- as.data.frame(conf_matrix)
ggplot(conf_matrix_df, aes(x = Actual, y = Predicted, fill = Freq)) +
  geom_tile() +
  geom_text(aes(label = Freq), color = "white", size = 6) +
  scale_fill_gradient(low = "blue", high = "red") +
  labs(title = "Confusion Matrix", x = "Actual", y = "Predicted") +
  theme_minimal()
# Использование confusionMatrix из пакета caret для более детального анализа
conf_matrix_caret <- confusionMatrix(as.factor(y_true_classes), as.factor(y_pred_classes), positive = "1")
print("Detailed Confusion Matrix (caret):")
print(conf_matrix_caret)
library(pROC)
roc_curve <- roc(y_true_classes, y_pred_classes)
# Визуализация ROC-кривой
plot(roc_curve, main="ROC Curve", col="blue", lwd=2)
# Добавление AUC на график
auc_value <- auc(roc_curve)
legend("bottomright", legend=paste("AUC =", round(auc_value, 2)), col="blue", lwd=2)
# Вывод AUC в консоль
print(paste("AUC:", auc_value))
library(PRROC)
PRROC_obj <- roc.curve(scores.class0 = y_pred_classes, weights.class0=y_true_classes, curve=TRUE)
plot(PRROC_obj)
library(ROCit)
## Warning: package 'ROCit' was built under R version 3.5.2
ROCit_obj <- rocit(score=y_pred_classes, class=y_true_classes, negref=1)
plot(ROCit_obj)

```

```

ksplot(ROCit_obj)
library(pROC)
pROC_obj <- roc(y_true_classes, y_pred_classes)
# Сглаживание ROC-кривой с использованием большего количества точек
smoothed_roc <- smooth(pROC_obj, method = "density", n = 1000) # Увеличива
ем количество точек для сглаживания
# Построение ROC-кривой с настройками
plot(smoothed_roc, col = "red", lwd = 2, main = "ROC Curve",
     auc.polygon = TRUE, max.auc.polygon = TRUE, grid = TRUE,
     print.auc = TRUE, show.thres = TRUE, xlim = c(1, 0), mar = c(4, 4, 2, 2))
# Добавление доверительного интервала для чувствительности
sens.ci <- ci.se(smoothed_roc)
# Рисуем доверительный интервал, не перезаписывая цвет линии ROC
plot(sens.ci, type = "shape", col = rgb(173, 216, 230, alpha = 127, maxColorValue
= 255),
     add = TRUE, no.roc = TRUE)
##
pROC_obj <- roc(y_true_classes, y_pred_classes,
               smoothed = TRUE,
               # arguments for ci
               ci=TRUE, ci.alpha=0.9, stratified=FALSE,
               # arguments for plot
               plot=TRUE, auc.polygon=TRUE, max.auc.polygon=TRUE, grid=FALSE,
               print.auc=TRUE, show.thres=TRUE)
sens.ci <- ci.se(pROC_obj)
plot(sens.ci, type="shape", col="lightblue")
plot(sens.ci, type="bars")

```

**Приложение 6. Патент на изобретение «Способ оценки плодовитости свиней пород ландрас и крупная белая»**





**Приложение 7. Патент на изобретение «Способ диагностики  
предрасположенности свиней породы дюрок к бурситам на основе вариантов  
SNP rs81327419 (SSC13) в гене ROBO2»**





**Приложение 8. Свидетельство о государственной регистрации программ для ЭВМ «База данных аутосомных ДНК-маркеров свиней»**

**РОССИЙСКАЯ ФЕДЕРАЦИЯ**



**СВИДЕТЕЛЬСТВО**  
о государственной регистрации базы данных

**№ 2015621623**

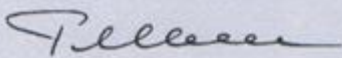
**БАЗА ДАННЫХ АУТОСОМНЫХ ДНК-МАРКЕРОВ  
СВИНЕЙ**

Правообладатель: *Федеральное государственное бюджетное образовательное учреждение высшего образования «Донской государственный аграрный университет» (RU)*

Авторы: *Гетманцева Любовь Владимировна (RU), Колосов Анатолий Юрьевич (RU), Леонова Мария Анатольевна (RU), Бакоев Сирождин Юсуфович (RU), Бакоев Некруз Фарходович (RU), Радюк Анастасия Владимировна (RU)*

Заявка № **2015621146**  
Дата поступления **09 сентября 2015 г.**  
Дата государственной регистрации  
в Реестре баз данных **02 ноября 2015 г.**

Руководитель Федеральной службы  
по интеллектуальной собственности

 **Г.П. Ивлиев**





**Приложение 9. Свидетельство о государственной регистрации программ для ЭВМ «Оценка информативности ДНК-маркеров. Метод Шеннона»**

**РОССИЙСКАЯ ФЕДЕРАЦИЯ**



**СВИДЕТЕЛЬСТВО**

о государственной регистрации программы для ЭВМ

**№ 2016661414**

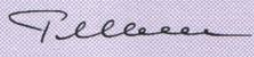
**Оценка информативности ДНК-маркеров. Метод Шеннона**

Правообладатель: *Федеральное государственное бюджетное образовательное учреждение высшего образования «Донской государственный аграрный университет» (RU)*

Авторы: *Колосов Анатолий Юрьевич (RU), Гетманцева Любовь Владимировна (RU), Леонова Мария Анатольевна (RU), Бакоев Сирождин Юсуфович (RU), Бакоев Некруз Фарходович (TJ)*

Заявка № **2016618776**  
Дата поступления **12 августа 2016 г.**  
Дата государственной регистрации  
в Реестре программ для ЭВМ **07 октября 2016 г.**

Руководитель Федеральной службы  
по интеллектуальной собственности

  
 **Г.П. Ивлиев**



**Приложение 10. Свидетельство о государственной регистрации программ для ЭВМ «Оценка моделей прогнозирования племенной ценности свиней»**

**РОССИЙСКАЯ ФЕДЕРАЦИЯ**



**СВИДЕТЕЛЬСТВО**  
о государственной регистрации программы для ЭВМ

**№ 2017660746**

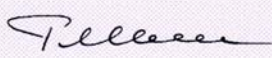
**Оценка моделей прогнозирования племенной ценности свиней**

Правообладатель: **Федеральное государственное бюджетное образовательное учреждение высшего образования «Донской государственный аграрный университет» (RU)**

Авторы: **Колосов Анатолий Юрьевич (RU), Гетманцева Любовь Владимировна (RU), Леонова Мария Анатольевна (RU), Бакоев Сироджин Юсуфович (RU)**

Заявка № **2017617918**  
Дата поступления **03 августа 2017 г.**  
Дата государственной регистрации  
в Реестре программ для ЭВМ **25 сентября 2017 г.**

Руководитель Федеральной службы  
по интеллектуальной собственности

 **Г.П. Ивлиев**

