

Методы генотипирования крупного рогатого скота для post-GWAS аннотирования SNPs

М. В. Бытов¹, О. В. Соколова¹, Н. А. Безбородова¹, А. С. Красноперов¹, А. Г. Исаева¹✉

¹Уральский федеральный аграрный научно-исследовательский центр Уральского отделения Российской академии наук, Екатеринбург, Россия

✉ E-mail: isaeva.05@bk.ru

Аннотация. Полногеномный анализ ассоциаций является одним из ключевых инструментов для выяснения генетической этиологии различных фенотипов, в том числе заболеваний и степени предрасположенности к ним. В последующем для статистически значимых генетических маркеров необходимо проведение валидационных исследований на независимых выборках. Такие валидационные post-GWAS-исследования позволяют тестировать генетические маркеры, высокая степень ассоциации которых с фенотипом подтверждается независимо от размера выборки, позволяя опровергнуть ложноположительные результаты изначального анализа ассоциаций. При выборе методики генотипирования для валидационных post-GWAS-исследований необходимо учитывать размер выборки и количество генетических маркеров для изучения, поскольку методики различаются по производительности и необходимым финансовым ресурсам. **Цель** данной работы – проведение генотипирования крупного рогатого скота по SNPs rs137396952 и rs134055603, для которых была показана высокая степень ассоциации с развитием кетоза в предыдущих GWAS-исследованиях, и описание современных методов генотипирования в зависимости от их производительности. Используемые **методы** генотипирования включают в себя технологии TaqMan и High-Resolution Melt Analysis. Анализ генотипов проведен с помощью веб-инструмента SNPStats. По полученным **результатам** генотипирования с помощью указанных технологий продемонстрированы особенности аллельной дискриминации методов. Проведено тестирование результатов генотипирования, в результате которого показано, что rs134055603 не подчиняется равновесию Харди – Вайнберга в исследованной выборке животных. **Научная новизна.** Полученные результаты по генотипированию могут быть использованы в дальнейших ассоциативных тестах с физиологически ценными параметрами, в том числе устойчивости к заболеваниям крупного рогатого скота молочного направления.

Ключевые слова: GWAS, крупный рогатый скот, генотипирование, SNP, ДНК, ПЦР, электрофорез.

Для цитирования: Бытов М. В., Соколова О. В., Безбородова Н. А., Красноперов А. С., Исаева А. Г. Методы генотипирования крупного рогатого скота для post-GWAS аннотирования SNPs // Аграрный вестник Урала. 2023. № 06 (235). С. 67–75. DOI: 10.32417/1997-4868-2023-235-06-67-75.

Дата поступления статьи: 14.03.2023, **дата рецензирования:** 10.04.2023, **дата принятия:** 20.04.2023.

Постановка проблемы (Introduction)

Полногеномные ассоциативные исследования (Genome-Wide Association Study, GWAS) определяют генетические области, которые могут играть важную роль в развитии фенотипов. Аннотирование этих статистически значимых пиков ассоциаций – локусов (Quantitative Trait Loci, QTL) – в виде генов и их функциональных частей, в которых находятся значимые SNPs, является важным шагом в процессе GWAS для проведения связи с тестируемым фенотипом [1]. Однако некоторые из этих пиков могут быть ложноположительными [2]. Для достоверно значимых генетических маркеров необходимо проведение валидационных исследований

на независимых выборках. Такие валидационные исследования позволяют определить генетические маркеры, высокая степень ассоциации которых с фенотипом подтверждается независимо от размера выборки или ложноположительных результатов изначального анализа ассоциаций. При выборе методики генотипирования необходимо учитывать размер выборки и количество генетических маркеров для изучения.

На сегодняшний день для генотипирования организмов исследователями используется ряд методов в различных модификациях, однако их можно объединить в три основные группы: полимеразная цепная реакция (ПЦР), секвенирование, ДНК-

чипы [3]. В свою очередь, каждый из них можно разделить на подгруппы.

Так, генотипирование с помощью ПЦР можно разделить на ПЦР по конечной точке и в режиме реального времени. ПЦР-методы включают ПЦР с использованием аллель-специфичных праймеров (как с детекцией по конечной точке – KASP, так и в режиме реального времени – rhAmp), гибридационных зондов (технология TaqMan и другие), а также ферментов рестрикции (ПЦР-ПДРФ). Для использования отдельных методов требуется соответствующее оборудование (например, амплификатор с детекцией в реальном времени для технологии TaqMan [4]). Перечисленные технологии имеют примерно одинаковую производительность, которая ограничена количеством каналов детекции в амплификаторе, а также стоимостью. Для проведения функциональной аннотации отдельных SNPs в post-GWAS-исследованиях на большом количестве особей данную группу методов можно считать оптимальной по соотношению цены и производительности.

Секвенирование исторически принято подразделять на секвенирование по Сэнгеру; секвенирование второго поколения, которое является наиболее распространенной технологией и реализовано компаниями Illumina Inc., Thermo Fisher Scientific, GeneMind Biosciences; секвенирование третьего поколения, которое реализовано компаниями MGI Tech, Nanopore Inc., Pacific Biosciences of California Inc. [5]. Для выполнения секвенирования отдельных участков генома на небольшом количестве особей оптимальным можно считать NGS второго поколения.

Использование ДНК-чипов является ключевой технологией в GWAS для поиска новых QTL и SNPs, ассоциированных с фенотипом [6]. Наиболее распространенным вариантом метода можно считать технологию от компании Illumina Inc. Метод позволяет генотипировать особей по SNPs, распределенных по всему геному. Для валидационных исследований необходим выбор конкретных SNPs, для которых была показана высокая степень ассоциации с фенотипом, поэтому для таких целей требуется подробный анализ данных предыдущих GWAS для создания кастомного чипа. Использование ДНК-чипов является оптимальным при генотипировании особей по очень большому количеству полиморфизмов (ДНК-чипы высокой плотности от компании Illumina Inc. способны детектировать до 777 тысяч SNPs).

Стоит отдельно отметить метод генотипирования с помощью MALDI-TOF MS. Эта технология реализована компаниями Agena Bioscience Inc. (ранее Sequenom Inc.) и Hexin Instrument Co. [7]. Данный вид масс-спектрометрии в качестве downstream-анализа после мини-секвенирования

(SBE, single-base extension) оптимально подходит для генотипирования до 30 SNPs одновременно для относительно большого количества особей (до 500).

В 2019 г. группой канадских ученых был проведен полногеномный анализ ассоциаций с уровнем β-гидроксibuтирата в качестве маркера субклинического кетоза крупного рогатого скота молочного направления продуктивности во время первой и последующих лактаций [8]. В результате исследования был выявлен ряд QTL: на 6, 14 и 20 аутосомах. SNPs rs137396952 и rs134055603 в межгенном регионе GC-NPFFR2 показали высокую степень ассоциации с обозначенными фенотипами.

Таким образом, цель нашего исследования – проведение генотипирования крупного рогатого скота по SNPs rs137396952 и rs134055603, описание современных методов генотипирования в зависимости от их производительности.

Методология и методы исследования (Methods)

Экстракция ДНК проведена из цельной крови с помощью Diatom DNA Prep 400 (ООО «Лаборатория Изоген», Россия) в соответствии с указаниями производителя. Концентрацию выделенной ДНК измеряли с помощью набора Qubit 1X dsDNA HS (Thermo Fisher Scientific, США). В дальнейшем для работы использовали только образцы с концентрацией ДНК выше 5 нг/мкл. Было исследовано 240 образцов цельной крови.

Для дизайна олигонуклеотидов было использовано программное обеспечение FastPCR 6.8.12 [9; 10]. ПЦР в реальном времени по технологии TaqMan проводили с помощью наборов с Taq-полимеразой Hot Start (ООО «Диаэм», Россия). Конечный объем реакционной смеси составил 10 мкл, объем ДНК – 1 мкл. Конечная концентрация $MgCl_2$ – 1,16 мМ, каждого dNTP – 0,2 мМ, Taq-полимеразы – 5 е. а.; 0,35 мкмоль каждого праймера и 0,9 мкмоль каждого зонда. Последовательность олигонуклеотидов (ООО «ДНК-Синтез», Россия) представлена в таблице 1. ПЦР в реальном времени с дальнейшим анализом кривой плавления (High-resolution Melt analysis, HRM) была проведена с 1 мкл продукта амплификации после TaqMan в конечном объеме 10 мкл. Концентрация в конечном объеме $MgCl_2$ – 1,25 мМ, каждого dNTP – 0,22 мМ, Taq-полимеразы – 6,5 е. а.; Eva488 20x (аналог EvaGreen; производство Lumiprobe RUS Ltd, Россия) – 0,225 мкл; 0,45 мкмоль каждого праймера (последовательности в таблице 1). Длина целевого фрагмента была проверена с помощью гель-электрофореза на агарозной пластине, окрашенной бромистым этидием. Для определения оптимальной температуры отжига праймеров предварительно была выполнена амплификация в температурном градиенте с дальнейшим выбором оптимального протокола. Для проведения HRM-анализа после этапа амплификации был использован следующий

температурный протокол: 95 °C – 5 мин., 50 °C – 2 мин., плавление от 70 °C до 85 °C с инкрементом 0,1 °C в течение 2 с.

Для аллельной дискриминации по технологии TaqMan было использовано ПО Manager Bio-Rad (США). Для анализа HRM было использовано ПО с открытым кодом – hrmR, написанное на языке R [11].

Статистическая проверка результатов генотипирования на соответствие равновесию Харди – Вайнберга проведена с помощью веб-инструмента SNPStats [12].

Результаты (Results)

Технология TaqMan позволяет проводить генотипирование особей по интересующим SNPs по двум аллельным вариантам сразу. Так, в ходе генотипирования по rs134055603 с помощью аллельной дискриминации в ПО Manager Bio-Rad была определена аллельная частота в исследованной выборке: A* – 51 %, G* – 49 %. Генотип A/A составлял 28 %, A/G – 47 %, G/G – 25 %, соответствуя равновесию Харди – Вайнберга по точному критерию Фишера (p -value = 0,37).

Анализ кривой плавления для rs134055603 показал идентичные результаты с генотипированием по технологии TaqMan. Так, с помощью ПО hrmR было определено, что оптимальное количество кластеров кривых плавления – 3 (рис. 1, а) и соответствует количеству генотипов, обнаруженных по TaqMan. Программа hrmR позволяет работать с изначальными данными по результатам плавления (рис. 1, б), а также проводит нормализацию по одному из образцов для более эффективной визуализации результатов (рис. 1, с). В качестве алгоритма кластеризации выбран метод k -ближайших соседей.

В ходе генотипирования по rs137396952 с помощью аллельной дискриминации в ПО Manager Bio-Rad была определена аллельная частота в исследованной выборке: T* – 64 %, C – 36 %. Генотип T/C от общей выборки составлял 71 %, а T/T – 29 %.

Анализ кривых плавления для rs137396952 показал, что оптимальное количество кластеров кривых плавления – 2 (рис. 2, а). По этим данным также проведена нормализация (рис. 2, б).

Таблица 1

Использованные для генотипирования олигонуклеотиды

Название	Последовательность	Длина ампликона, п. н.
rs137396952_F	ATGCAGCAGAAACAAGGGTTAAA	225
rs137396952_R	GTACAGCCACTGTGCAACAAC	
rs137396952_pT	HEX-GA+TT+CAGCATG+G+T+G+TCAG-BHQ2	
rs137396952_pC	Cy5-GATT+CAGCATG+G+C+G+TCAG-BHQ3	
rs137396952_HRM_F	GCAAGAGGAATAGATTCAGCA	71
rs137396952_HRM_R	ATCTTCATCCTCATGTTTCATAGT	
rs134055603_F	GACAAGGCTTTTGATAGGTGAAA	316
rs134055603_R	CAAAGCAACCACACAATGTTG	
rs134055603_pA	HEX-CAT+TT+TCT+T+A+GA+CT+T+CTG-BHQ1	
rs134055603_pG	Cy5-CATTTTCT+T+G+GA+CT+T+CTG-BHQ3	
rs134055603_HRM_F	ATTCTCAAATAATTTTGGGTTCAA	94
rs134055603_HRM_R	CACAGACTTTGATACTATAAACAC	

Примечание. «+» используется для обозначения LNA модификации следующего нуклеотида.

Table 1
Oligonucleotides utilized for genotyping

Name	Sequence	Amplicon length, b. p.
rs137396952_F	ATGCAGCAGAAACAAGGGTTAAA	225
rs137396952_R	GTACAGCCACTGTGCAACAAC	
rs137396952_pT	HEX-GA+TT+CAGCATG+G+T+G+TCAG-BHQ2	
rs137396952_pC	Cy5-GATT+CAGCATG+G+C+G+TCAG-BHQ3	
rs137396952_HRM_F	GCAAGAGGAATAGATTCAGCA	71
rs137396952_HRM_R	ATCTTCATCCTCATGTTTCATAGT	
rs134055603_F	GACAAGGCTTTTGATAGGTGAAA	316
rs134055603_R	CAAAGCAACCACACAATGTTG	
rs134055603_pA	HEX-CAT+TT+TCT+T+A+GA+CT+T+CTG-BHQ1	
rs134055603_pG	Cy5-CATTTTCT+T+G+GA+CT+T+CTG-BHQ3	
rs134055603_HRM_F	ATTCTCAAATAATTTTGGGTTCAA	94
rs134055603_HRM_R	CACAGACTTTGATACTATAAACAC	

Note. “+” is used to label LNA modified following nucleotide.

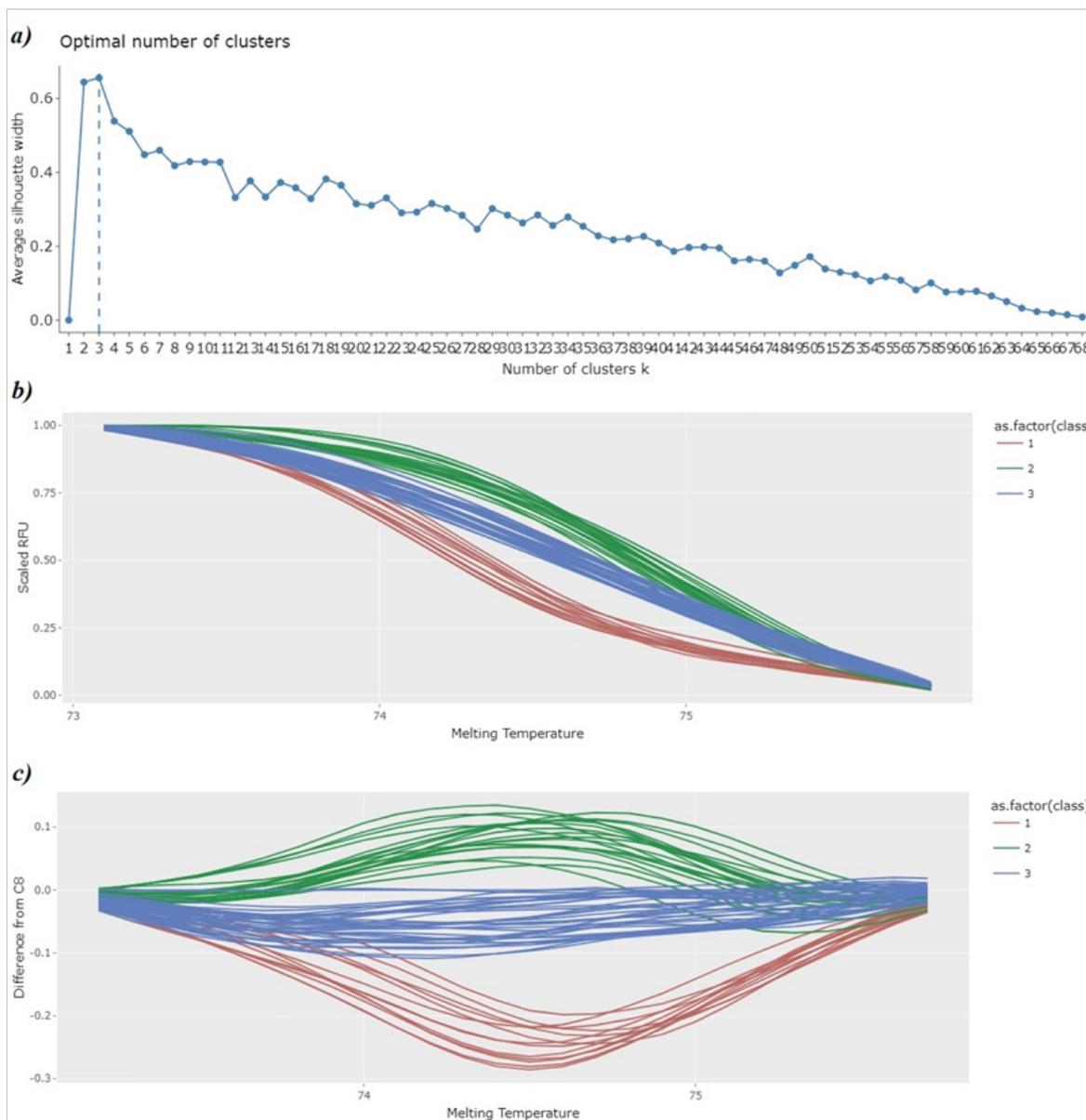


Рис. 1. Анализ кривых плавления для rs134055603:

a) оптимальное количество кластеров, определяемое по расстоянию между членами кластера и соседним кластером; б) исходные данные после кластеризации; в) визуализация нормализованных данных

Fig. 1. rs134055603 melt curve analysis:

a) optimal number of clusters identified by distance between cluster members and neighboring cluster; b) raw data after clusterizations; c) normalized data visualization

Таким образом, результаты анализа HRM подтверждают, что для rs137396952 в исследованной выборке животных гомозигот по аллелю С* обнаружено не было.

Для rs137396952 обнаружено статистически значимое ($p < 0,0001$) несоответствие равновесию Харди – Вайнберга, поскольку в исследованной выборке не обнаружено гомозигот по одной из аллелей.

Обсуждение и выводы (Discussion and Conclusion)

В ходе работы рассмотрены основные технологии генотипирования организмов: ПЦР-методы, секвенирование, ДНК-чипы.

Достигнуты существенные экспериментальные и методологические успехи, которые позволяют изучать гораздо более распространенные геноварианты в больших выборках. Дизайн чипов высокой плотности, использование импутации генотипов и увеличение размеров референтных панелей для импутации позволяют проводить подробный полногеномный анализ полиморфизмов. Разработка инструментов для метаанализа данных упрощает агрегирование результатов GWAS и способствует увеличению размеров выборки генетических исследований. При использовании ДНК-чипов в GWAS-

исследованиях десятилетие назад был общепризнан порог статистической значимости $p = 5 \times 10^{-8}$, однако в современных исследованиях используют поправку Бонферрони на множественную проверку гипотез, чтобы избежать ошибки первого рода – ложноположительных ассоциаций [13].

В обзорной статье немецкого исследователя S. Qanbari показано, что неравновесное сцепление у крупного рогатого скота протягивается на промежуток чуть меньше 40 тысяч пар оснований, что немногим больше, чем у человека [14]. Сравнения уровней проводятся не только для межвидовых, но и для межпородных [15] и гибридных [16] особенностей неравновесного сцепления. Успех GWAS с использованием SNP в качестве генетических маркеров частично зависит от неравновесного сцепления конкретных маркеров в популяции. Неравновесное сцепление – это свойство SNP в геномной последовательности, которое обозначает их физическую или адаптивно выгодную при наследовании сцепленность [17].

GWAS позволяют выявить тысячи ассоциаций SNP с признаками по всему геному, связывая генетические вариации с сотнями, если не тысячами сложных заболеваний и признаков. Однако лишь

небольшая часть этих статистических ассоциаций была тщательно исследована, чтобы определить:

- какой вариант или варианты являются причинными (в том числе определение ложноположительных ассоциаций);
- каковы молекулярные функции причинных вариантов (в том числе нарушение кодирующей функции генов);
- какие гены затронуты причинными вариантами (аннотация полиморфизмов);
- как изменения в функции или регуляции причинных генов приводят к изменению риска заболевания [18].

Учитывая все отмеченные ограничения GWAS, крайне важно выполнить анализ GWAS downstream, цель которого состоит в том, чтобы использовать результат ассоциации для выполнения задач:

- возможность экстраполяции предыдущего результата на другие экспериментальные выборки;
- выявление новых значимых функциональных вариантов;
- идентификация новых генов восприимчивости к болезням, ассоциаций «генотип – фенотип» и сети путей передачи биологических сигналов;
- проведение оценки полигенного риска.

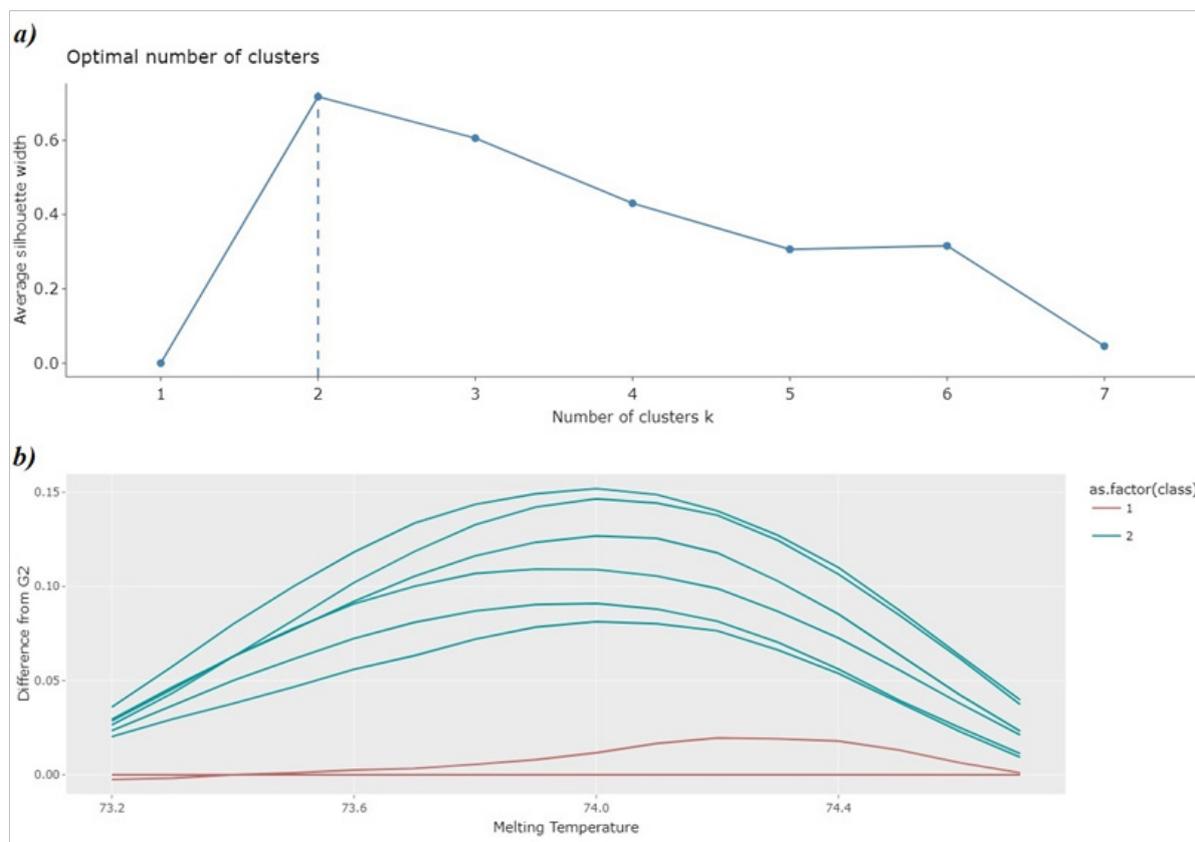


Рис. 2. Анализ кривых плавления для rs137396952:

a) оптимальное количество кластеров, определяемое по расстоянию между членами кластера и соседним кластером; b) визуализация нормализованных данных

Fig. 2. rs137396952 melt curve analysis:

a) optimal number of clusters identified by distance between cluster members and neighboring cluster; b) normalized data visualization

На сегодняшний день наиболее распространенные подходы к выполнению анализа GWAS downstream включают: проведение исследования по отдельным SNP, использование полиморфизма отдельных генов, изучение полиморфизма генов в рамках целых клеточных путей [19].

За последние несколько лет произошел резкий подъем в количестве исследований организации генома в ядре, и в настоящее время существует множество доказательств того, что хромосомы могут изгибаться и образовывать петли в масштабах тысяч и миллионов оснований и что эти петли играют важную роль в регуляции транскрипции. В то время как транскрипция гена запускается на промоторе, энхансеры и другие дистальные регуляторные элементы также влияют на транскрипцию гена, физически взаимодействуя со своими промоторами-мишенями, а часто и друг с другом посредством петлевых пересечений хроматина. Таким образом, физический контакт между дистальными регуляторными элементами и промоторами может рассматриваться как свидетельство регуляторной функции этих элементов. Сочетание методов фиксации конформации хромосом (3C) с высокопроизводительным секвенированием позволило исследовать все дальние контакты в геноме (Hi-C, подход «все против всех»), а также с превосходной глубиной и разрешением все дальние контакты, затрагивающие любую интересующую область, например, промотор гена (4C, подходы «один против всех») [20].

Хотя определение молекулярного механизма, с помощью которого генетический вариант, связанный с фенотипом, влияет на экспрессию гена, важно с точки зрения генетической регуляции, более практический вопрос заключается в том, как небольшие изменения в уровнях экспрессии гена влияют на фенотипы клеток и организма. Несмотря на то что во многих исследованиях сообщалось о генетических вариантах, которые изменяют цис-регуляторную функцию (изменяют участки некодирующей ДНК, которые регулируют транскрипцию соседних генов), механизмы, с помощью которых результирующие изменения в экспрессии генов влияют на риск развития заболевания, часто не исследованы или неизвестны. Некоторые исследования функционально связывают уровни экспрессии причинного гена с соответствующими заболеваниями фенотипами, но многие из этих исследований основываются на неточно контролируемой сверхэкспрессии, нокдауне или нокауте. Повторить различия в экспрессии генов, относящиеся к локусу риска заболевания, сложно по крайней мере по двум причинам: во-первых, сила эффекта eQTL (локусы количественных признаков экспрессии) с точки зрения кратности изменения экспрессии обычно неизвестна, не сообщается или мала; во-вторых, технически сложно количественно точно проконтролировать

сверхэкспрессию или нокдаун гена. Чтобы решить эту проблему, в некоторых исследованиях проводят поиск корреляции между уровнями экспрессии интересующего гена и фенотипами. То есть при обнаружении казуального аллеля заболевания проводят, например, нокдаун ассоциированного с вариантом гена. Таким образом, роль ассоциированного с фенотипом гена доказывается или опровергается. При подтверждении можно предполагать конкретный механизм воздействия аллеля на развитие фенотипа [18].

Основное ограничение рассмотренных выше подходов заключается в том, что они являются корреляционными. Таким образом, для определения фенотипических эффектов аллель-специфических изменений в экспрессии генов редактирование генома может быть лучшим подходом. Путем мутации причинного локуса с одного аллеля на другой результирующие изменения экспрессии генов и клеточных фенотипов:

1) с большей вероятностью будут физиологически значимыми, чем наблюдаемые в экспериментах со сверхэкспрессией или нокдауном;

2) могут быть напрямую функционально связаны с генетическим полиморфизмом.

Идентификация типов клеток, связанных с заболеванием, предполагает, что значимые SNP представлены в специфически активных патогенных типах клеток. Методы обогащения SNP объединяют результаты GWAS с различными геномными аннотациями и отдают приоритет типам клеток, в которых связанные варианты перекрывают аннотации чаще, чем ожидалось бы при случайном совпадении. Например, специфическая для типа клеток активность геномной области (например, локуса GWAS) может быть определена уровнями экспрессии генов в этой области. Существует подход, который определяет, насколько специфичны для определенного типа клеток те гены, которые имеют высокую экспрессию в отдельных типах клеток по сравнению со всеми другими клетками. Если для данного признака аллельные варианты SNP чрезмерно представлены (обогащены) для генов, специфически экспрессируемых в данном типе клеток, то этот тип клеток имеет приоритет для дальнейшего изучения [21].

Таким образом, мы предполагаем, что повышенное внимание к downstream функциональному анализу уже идентифицированных локусов GWAS, а не к поиску еще большего количества локусов GWAS, скорее всего, принесет больше пользы в понимании физиологических процессов. На данный момент мультиплексная ПЦР или аналоги, а также кастомные ДНК-чипы представляются наиболее оптимальными технологиями для анализа по результатам ранее проведенных GWAS-исследований.

В результате нашей работы отобраны два SNPs, для которых по данным предыдущих GWAS-исследований показана высокая степень ассоциации с развитием субклинического кетоза крупного рогатого скота молочного направления продуктивности. Для данных полиморфизмов разработаны кастомные ПЦР-системы для генотипирования животных. Использовано open-source программное обеспечение для анализа кривой плавления высокого разрешения с целью валидации результатов генотипирования по комплементарной ей технологии – TaqMan. Проведена оценка аллельных частот по каждому локусу на соответствие равновесию

Харди – Вайнберга. Показано соответствие равновесию для rs134055603, а для rs137396952 – несоответствие. Данные будут использованы в ассоциативных тестах с фенотипом «восприимчивость к кетозу» у крупного рогатого скота молочного направления при помощи веб-инструмента SNPStats.

Благодарности (Acknowledgements)

Работа выполнена в рамках проекта РНФ № 22-16-00021 «Изучение ассоциаций молекулярно-генетических маркеров с ценными физиологическими признаками сельскохозяйственных животных с целью направленной селекции для повышения адаптационного потенциала и долголетия».

Библиографический список (References)

1. Uffelmann E., Huang Q. Q., Munung N. S., de Vries J., Okada Y., Martin A. R., Martin H. C., Lappalainen T., Posthuma D. Genome-wide association studies // *Nature Reviews Methods Primers*. 2021. Vol. 1. No. 1. Article number 59. DOI: 10.1038/s43586-021-00056-9.
2. Oetting W. S., Jacobson P. A., Israni A. K. Validation Is Critical for Genome-Wide Association Study-Based Associations // *American Journal of Transplantation*. 2017. Vol. 17. No. 2. Pp. 318–319. DOI: 10.1111/ajt.14051.
3. Goodwin S., McPherson J. D., McCombie W. R. Coming of age: ten years of next-generation sequencing technologies // *Nature Reviews Genetics*. 2016. Vol. 17. No. 6. Pp. 333–351. DOI: 10.1038/nrg.2016.49.
4. Kovalchuk S. N., Arkhipova A. L. Development of TaqMan PCR assay for genotyping SNP rs211250281 of the bovine *agpat6* gene // *Animal Biotechnology*. 2022. DOI: 10.1080/10495398.2022.2077742.
5. Heather J. M., Chain B. The sequence of sequencers: The history of sequencing DNA // *Genomics*. 2016. Vol. 107. No. 1. DOI: 10.1016/j.ygeno.2015.11.003.
6. Jiang J., Ma L., Prakapenka D., VanRaden P. M., Cole J. B., Da Y. A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle // *Frontiers in Genetics*. 2019. Vol. 10. Article number 412. DOI: 10.3389/fgene.2019.00412.
7. Liu Y., Jiao Y., Li P., Zan L. MALDI-TOF-MS-based high throughput genotyping of mutations associated with body measurement traits in cattle // *Mammalian Genome*. 2020. Vol. 31. No. 7-8. Pp. 228–239. DOI: 10.1007/s00335-020-09840-6.
8. Nayeri S., Schenkel F., Fleming A., Kroezen V., Sargolzaei M., Baes C., Canovas A., Squires J., Miglior F. Genome-wide association analysis for beta-hydroxybutyrate concentration in Milk in Holstein dairy cattle // *BMC Genetics*. 2019. Vol. 20. No. 1. Article number 58. DOI: 10.1186/s12863-019-0761-9.
9. Kalendar R., Khassenov B., Ramankulov Y., Samuilova O., Ivanov K. I. FastPCR: An in silico tool for fast primer and probe design and advanced sequence analysis // *Genomics*. 2017. Vol. 109. No. 3-4. Pp. 312–319. DOI: 10.1016/j.ygeno.2017.05.005.
10. Kalendar R., Lee D., Schulman A. H. Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis // *Genomics*. 2011. Vol. 98. No. 2. Pp. 137–144. DOI: 10.1016/j.ygeno.2011.04.009.
11. Pavlo H. pavlohrab/hrmR: Pre-release version (v0.1-alpha) // *Zenodo*. 2021. DOI: 10.5281/zenodo.4491296.
12. Sole X., Guino E., Valls J., Iñiesta R., Moreno V. SNPStats: a web tool for the analysis of association studies // *Bioinformatics*. 2006. Vol. 22. No. 15. Pp. 1928–1929. DOI: 10.1093/bioinformatics/btl268.
13. Chen Z., Boehnke M., Wen X., Mukherjee B. Revisiting the genome-wide significance threshold for common variant GWAS // *G3 (Bethesda)*. 2021. Vol. 11. No. 2. Article number jkaa056. DOI: 10.1093/g3journal/jkaa056.
14. Qanbari S. On the Extent of Linkage Disequilibrium in the Genome of Farm Animals // *Frontiers in Genetics*. 2019. Vol. 10. Article number 1304. DOI: 10.3389/fgene.2019.01304.
15. Fabbri M. C., Dadousis C., Bozzi R. Estimation of Linkage Disequilibrium and Effective Population Size in Three Italian Autochthonous Beef Breeds // *Animals (Basel)*. 2020. Vol. 10. No. 6. Article number 1034. DOI: 10.3390/ani10061034.
16. Singh A., Kumar A., Mehrotra A., Pandey A. K., Mishra B. P., Dutt T. Estimation of linkage disequilibrium levels and allele frequency distribution in crossbred Vrindavani cattle using 50K SNP data // *PLoS One*. 2021. Vol. 16. No. 11. Article number e0259572. DOI: 10.1371/journal.pone.0259572.
17. Joiret M., Mahachie John J. M., Gusareva E. S., Van Steen K. Confounding of linkage disequilibrium patterns in large scale DNA based gene-gene interaction studies // *BioData Mining*. 2019. Vol. 12. Article number 11. DOI: 10.1186/s13040-019-0199-7.

18. Gallagher M. D., Chen-Plotkin A. S. The Post-GWAS Era: From Association to Function // American Journal of Human Genetics. 2018. Vol. 102. No. 5. Pp. 717–730. DOI: 10.1016/j.ajhg.2018.04.002.
19. Adam Y., Samtal C., Brandenburg J. T., Falola O., Adebisi E. Performing post-genome-wide association study analysis: overview, challenges and recommendations // F1000Research. 2021. Vol. 10. Article number 1002. DOI: 10.12688/f1000research.53962.1.
20. Kempfer R., Pombo A. Methods for mapping 3D chromosome architecture // Nature Reviews Genetics. 2020. Vol. 21. No. 4. Pp. 207–226. DOI: 10.1038/s41576-019-0195-2.
21. Cano-Gamez E., Trynka G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases // Frontiers in Genetics. 2020. Vol. 11. Article number 424. DOI: 10.3389/fgene.2020.00424.

Об авторах:

Максим Владимирович Бытов¹, младший научный сотрудник отдела геномных исследований и селекции животных, ORCID 0000-0002-3622-3770, AuthorID 1121043; +7 904 380-41-51, bytovmaks@mail.ru

Ольга Васильевна Соколова¹, доктор ветеринарных наук, ведущий научный сотрудник отдела геномных исследований и селекции животных, ORCID 0000-0002-1169-4090, AuthorID 648613; +7 908 920-84-36, nauka_sokolova@mail.ru

Наталья Александровна Безбородова¹, кандидат ветеринарных наук, старший научный сотрудник отдела геномных исследований и селекции животных, ORCID 0000-0003-2793-5001, AuthorID 665979; +7 904 981-72-14, n-bezborodova@mail.ru

Александр Сергеевич Красноперов¹, кандидат ветеринарных наук, старший научный сотрудник отдела экологии и незаразной патологии животных, ORCID 0000-0002-7838-4126, AuthorID 655970; +7 903 083-31-32, marafon.86@list.ru

Альбина Геннадьевна Исаева¹, доктор биологических наук, ведущий научный сотрудник лаборатории биологических технологий отдела ветеринарно-лабораторной диагностики с испытательной лабораторией, ORCID 0000-0001-8395-1247, AuthorID 665717; +7 902 872-89-10, isaeva.05@bk.ru

¹ Уральский федеральный аграрный научно-исследовательский центр Уральского отделения Российской академии наук, Екатеринбург, Россия

Cattle genotyping methods for post-GWAS annotation of SNPs

M. V. Bytov¹, O. V. Sokolova¹, N. A. Bezborodova¹, A. S. Krasnoperov¹, A. G. Isaeva¹✉

¹ Ural Federal Agrarian Scientific Research Centre of the Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russia

✉ E-mail: isaeva.05@bk.ru

Abstract. Genome-wide association analysis is one of the key tools for elucidating the genetic etiology of various phenotypes, including diseases and the degree of predisposition to them. Subsequently, for statistically significant genetic markers, it is necessary to conduct validation studies on independent cohorts. These post-GWAS validation studies test genetic markers that are strongly associated with the phenotype, regardless of sample size, allowing to identify false-positive results from the initial association analysis. When choosing a genotyping technique for post-GWAS validation studies, consideration should be given to the sample size and the number of genetic markers planned to be studied, since genotyping techniques differ in throughput and cost. **The aim** of this paper is to describe modern methods of genotyping depending on their performance and to carry out genotyping of cattle for the SNPs rs137396952 and rs134055603, for which a high degree of association with the development of ketosis was shown in previous GWAS studies. Utilized genotyping **methods** include TaqMan and High-Resolution Melt Analysis; genotype analysis was performed using the SNPStats web tool. When comparing the **results** of genotyping using these technologies, the specificity of allelic discrimination carried out using these methods was demonstrated. Testing of the genotyping results had shown that rs134055603 does not obey the Hardy-Weinberg equilibrium in the studied cohort of animals. **Scientific novelty.** Obtained genotyping results will be used in further association tests with physiologically valuable parameters of dairy cattle, including resistance to diseases.

Keywords: GWAS, cattle, genotyping, SNP, DNA, PCR, electrophoresis.

For citation: Bytov M. V., Sokolova O. V., Bezborodova N. A., Krasnoperov A. S., Isaeva A. G. Metody genotipirovaniya krupnogo rogatogo skota dlya post-GWAS annotirovaniya SNPs [Cattle genotyping methods for post-GWAS annotation of SNPs] // Agrarian Bulletin of the Urals. 2023. No. 06 (235). Pp. 67–75. DOI: 10.32417/1997-4868-2023-235-06-67-75. (In Russian.)

Date of paper submission: 14.03.2023, **date of review:** 10.04.2023, **date of acceptance:** 20.04.2023.

Authors' information:

Maksim V. Bytov¹, junior researcher of the department of animal genomics and selection, ORCID 0000-0002-3622-3770, AuthorID 1121043; +7 904 380-41-51, bytovmaks@mail.ru

Olga V. Sokolova¹, doctor of veterinary sciences, leading researcher of the department of animal genomics and selection, ORCID 0000-0002-1169-4090, AuthorID 648613; +7 908 920-84-36, nauka_sokolova@mail.ru

Natalya A. Bezborodova¹, candidate of veterinary sciences, senior researcher of the department of animal genomics and selection, ORCID 0000-0003-2793-5001, AuthorID 665979; +7 904 981-72-14, n-bezborodova@mail.ru

Aleksandr S. Krasnoperov¹, candidate of veterinary sciences, senior researcher of the department of ecology and animals' noncontagious pathology, ORCID 0000-0002-7838-4126, AuthorID 655970; +7 903 083-31-32, marafon.86@list.ru

Albina G. Isaeva¹, doctor of biological sciences, leading researcher of the laboratory of biological technologies of the department of veterinary laboratory diagnostics with a testing laboratory, ORCID 0000-0001-8395-1247, AuthorID 665717; +7 902 872-89-10, isaeva.05@bk.ru

¹ Ural Federal Agrarian Scientific Research Centre of the Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russia