

**ПОИСК ГЕНОМНЫХ ВАРИАНТОВ, АССОЦИИРОВАННЫХ С
ПРОДУКТИВНЫМИ КАЧЕСТВАМИ У ЛОШАДЕЙ, НА ОСНОВЕ
АНАЛИЗА SNP-ГЕНОТИПОВ**

Бейшова Индира Салтановна, директор испытательного центра, НАО «Западно-Казахстанский аграрно-технический университет имени Жангир хана»

Гриценко Диляра Александровна, заведующая лабораторией молекулярной биологии, РГП на ПХВ «Институт биологии и биотехнологии растений»

Шамекова Малика Хабидулаевна, заведующая лабораторией селекции и биотехнологии, РГП на ПХВ «Институт биологии и биотехнологии растений»

Пожарский Александр Сергеевич, научный сотрудник лаборатории молекулярной биологии, РГП на ПХВ «Институт биологии и биотехнологии растений»

Ульянова Татьяна Владимировна, старший научный сотрудник лаборатории биотехнологии и диагностики инфекционных болезней, НАО «Западно-Казахстанский аграрно-технический университет имени Жангир хана»

Ковальчук Александр Михайлович, заведующий лабораторией биотехнологии и диагностики инфекционных болезней, НАО «Западно-Казахстанский аграрно-технический университет имени Жангир хана»

Салимова Динара Калилуловна, докторант Института ветеринарной медицины и животноводства НАО «Западно-Казахстанский аграрно-технический университет имени Жангир хана»

Аннотация. В настоящей работе представлены результаты полногеномного поиска ассоциаций однонуклеотидных полиморфизмов (SNP) с продуктивными качествами у отечественных пород лошадей, полученные на основании данных SNP-генотипирования животных, проведенного с помощью набора реагентов Equine 80 kHTS («IlluminaInc.», США).

Ключевые слова: полногеномный поиск ассоциаций, лошади, SNP, тип джабе, адайский тип, найманский тип, костанайская порода, кушумская порода, мугалжарская порода.

В 2007 году полная последовательность генома лошади стала общедоступной, благодаря этому исследованию совершен значительный прорыв в области геномики лошадей [1]. Результаты этой работы послужили ключевым источником для создания коммерческих массивов SNP, что позволило проводить высокопроизводительное генотипирование лошадей. В 2011 году появились массивы ДНК-генотипирования первого и второго поколения, включающие 54 602 и 74 500 SNP-маркеров соответственно [2]. Как следствие, стало возможным проведение полногеномно-ассоциативных

исследований (GWAS) у лошадей. Среди наиболее впечатляющих результатов GWAS, полученных к настоящему времени, стоит отметить обнаружение SNP на 18 хромосоме, влияющих на экспрессию гена миостатина (*MSTN*), который связан со спортивными качествами у чистокровных лошадей [3].

Достигнут значительный прогресс в расшифровке менделевских признаков, включая выявление ряда аллельных вариантов, ответственных за разные окрасы шерсти [4], а также ассоциированных с генетическими дефектами [5].

В настоящее время имеется ограниченное количество информации о генетике полигенных количественных признаков у лошадей, включая их продуктивные качества. Исследование количественных признаков осложняется отсутствием информации о данных и фенотипах животных, а также сложной генетическими механизмами, которые служат основой для этих признаков. Необходимо отметить, что подобные исследования для казахских местных лошадей ранее не проводились, но являются бесспорно актуальными, так как коневодство в Казахстане является одной из ведущих отраслей животноводства, а изучение генетических особенностей формирования важных признаков уотечественных пород лошадей является важным для сохранения и совершенствования популяции.

Целью настоящей работы было исследование полногеномных ассоциаций SNP с продуктивными качествами в популяциях отечественных пород лошадей.

Исследовали биоматериал (волосыные луковицы) казахской породы типов джабе ($n = 631$), адай ($n = 303$) и найман ($n = 158$), мугалжарской ($n = 584$) кушумской ($n = 226$) и костанайской ($n = 116$) пород, отобранных с хозяйств различных регионов Казахстана.

Сбор промеров у лошадей проводили до отбора проб, при этом измеряли следующие показатели: высота в холке, косая длина туловища, обхват груди, обхват пясти и живую массу.

ДНК выделяли с помощью коммерческого набора ДНК-Экстран-2 (ООО «Синтол», Россия) в соответствии с рекомендациями производителя. Перед генотипированием на SNP-чипах полученные препараты проходили контроль качества: измеряли концентрацию двуцепочечной ДНК на флуориметре Qubit 4.0, «Invitrogen, Life Technologies», США), и проверяли её качество с помощью гель-электрофореза. Генотипирование проводили на чипе Equine 80kHTS («Illumina Inc.», США).

Контроль качества генотипирования проводили с помощью программы PLINK 1.9 [6]. По его результатам было отобрано 60 987 SNP.

Полногеномный ассоциативный анализ проводили с использованием PLINK 1.9. Результаты анализа генетической структуры, полученные ранее [7], были приняты во внимание при отборе образцов для GWAS. Анализ связи был проведен по показателям живой массы и промеров лошадей. Животные в возрасте до трех лет и образцы-выбросы были исключены из анализа. Корреляционный тест Пирсона был использован для обеспечения независимости фенотипических переменных от возраста. Переменная промеров была определена с использованием измерений высоты в холке, косой длины

туловища, обхвата груди, обхвата пясти. Эти параметры были нормализованы путем расчета среднего значения и деления на стандартное отклонение, затем был проведен анализ главных компонент (PCA), и первый компонент был выбран в качестве новой переменной промеров животных.

Для выполнения полногеномного поиска ассоциаций применялся тест линейной регрессии с адаптивным методом перестановок Монте-Карло, с коррекцией р-значения для учета множественных сравнений (команда PLINK '-linearperm'). SNP, находящиеся в состоянии сильного неравновесия связей, были исключены из анализа ($r^2 > 0,7$).

Маркеры с р-значением ниже установленного порога в 0,001 были аннотированы с использованием инструмента VEP (varianteffectpredictor) [8] и веб-сервера DAVID [9]. В качестве ссылки для аннотации использовалась сборка генома лошади EquCab3.0 (GCA_002863925.1).

Полногеномный анализ связи SNP-маркеров с промерами и живой массой лошадей был проведен для всех животных с доступными фенотипическими данными, за исключением лошадей в возрасте до 3 лет и отдаленных генотипов, выявленных с помощью анализа генетической структуры; всего было исследовано 1533 особи. Для проверки отсутствия влияния возраста в анализируемой выборке, мы провели корреляционный анализ Пирсона для всех фенотипических переменных. Показатели высоты в холке и косой длины туловища показали отсутствие значимой корреляции при пороговом уровне значимости $p=0,05$ ($p = 0,8388$ и $0,4211$ соответственно). Слабые корреляции были выявлены для обхвата груди ($0,0841$, $p = 0,000507$), обхвата пясти ($0,1011$, $p = 2,9229 \cdot 10^{-5}$) и живой массы ($0,1121$, $p = 3,496 \cdot 10^{-6}$). Показатели промеров были объединены с использованием PCA; первый основной компонент, описывающий 81% общей вариации, был выбран в качестве новой переменной.

GWAS-анализ проводился с использованием алгоритма линейной регрессии, реализованного в программном обеспечении PLINK, и включал адаптивную коррекцию р-значений на основе теста перестановок Монте-Карло. С целью визуализации распределения статистически значимых полиморфных сайтов и их распределения по хромосомам для каждого анализируемого параметра строили Манхэттенские графики (Manhattanplot). По оси Y откладывали значения отрицательного логарифма уровня значимости для каждого полиморфного сайта ($-\log p$). По оси X откладывали хромосомы, в которых локализованы полиморфные сайты. Чем меньше значение р, тем больший уровень значимости имеет показатель и тем выше он расположен относительно начала координат по оси Y.

При выбранном уровне значимости $p = 0,001$ было обнаружено, что 81 и 84 SNP имеют статистически значимые ассоциации с промерами и живой массой, соответственно. Среди выявленных полиморфизмов, 60 SNP были связаны с известными генами лошади с помощью сервера VEP Ensembl и с соответствующими биологическими процессами с помощью DAVID.

Между двумя наборами маркеров, связанных с соответствующими признаками (промеры и живая масса), почти не было перекрытия. Только два полиморфизма, BIEC2_117960 и BIEC2-187196, показали значительную

ассоциацию с обоими признаками. Первый маркер был связан с геном *OR4C269P*, для которого не было доступной генной онтологической аннотации, в то время как последний был связан с геном экто-5'-нуклеотидазы (*NT5E*), который участвует в метаболизме аденозинфосфатов. Идентифицированные гены играют регуляторную или сигнальную роль в разнообразных биологических процессах, охватывая широкий спектр уровней, начиная с клеточного и заканчивая организменным. Гены *BMP6*, *DDR3* и *CREB3L1* участвуют в развитии и метаболизме соединительных тканей, включая костную. Эти гены содержат SNP, которые, согласно нашим данным, ассоциированы с живой массой у отечественных лошадей. Ген *BMP6* содержал три полиморфизма, ассоциированных с продуктивными качествами, что было самым высоким числом полиморфизмов среди всех генов. Ряд генов, включая *DPF1*, *GNAT3*, *NEGR1* и т.д., были аннотированы как участвующие в развитии нервной системы. Так, ген *NEGR1* связан с регуляцией пищевого и двигательного поведения, в то время как ген *GNAT3* влияет на восприятие вкуса. Гены *BMP6*, *RELA1*, *AIM2*, *PDE4D* и *IGF1R* помимо своих других функций, участвуют в регуляции иммунной системы. Ген *EIF2AK4* связан с клеточной реакцией на холодостресс и дефицит белков.

Среди всех функционально аннотированных генов можно отметить некоторые определенные аспекты биологических процессов, потенциально связанных с интересующими признаками. Во-первых, развитие соединительных тканей и костной системы, которые имеют решающее значение для поддержания животным своего веса и размеров. Во-вторых, развитие нервной системы: более специфическое влияние генов *GNAT3* и *NEGR1* на предпочтения лошадей в пище и, следовательно, косвенное воздействие на их рост может представлять интересную тему для будущих исследований. В-третьих, регуляция иммунных процессов, которые могут оказывать влияние на рост путем воздействия на общее состояние здоровья.

Проведенное полногеномное ассоциативное исследование позволило выявить следующее. Обнаружено, что 60 SNP ассоциированы с одним из двух исследуемых признаков (живая масса и промеры) и связаны с функционально аннотированными генами лошадей. Среди идентифицированных генов были гены, участвующие в различных биологических процессах в качестве регуляторных и сигнальных факторов. Необходимо отметить, что почти все значимые полиморфизмы были независимо связаны с промерами или живой массой несмотря на то, что между этими признаками существует очевидная корреляция. Среди всех функционально аннотированных генов можно отметить некоторые определенные аспекты биологических процессов, потенциально связанных с интересующими признаками. Однако следует иметь в виду, что аннотации генов, выполненные с помощью «Geneontology», основаны главным образом на данных о человеке и модельных животных. В результате истинная физиологическая роль идентифицированных генов у лошадей может несколько отличаться. Кроме того, возможные ассоциации вариантов, которые пока не были идентифицированы, требуют

дополнительного исследования с учетом обновленных аннотационных данных для геномов лошадей.

Работа выполнена в рамках научного проекта грантового финансирования МНВО РК на 2022-2024 гг. ИРН № AP14870614 «Генетическое маркирование продуктивных качеств казахской лошади типа джабе на основе SNP-генотипирования с широким покрытием генома», а также научно-технической программы ПЦФ МСХ РК на 2021-2023 гг. ИРН № BR10764999 «Разработка технологий эффективного управления селекционным процессом и сохранения генофонда в коневодстве».

Библиографический список

1. Wade C.M. Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse/C.M. Wade, E. Giulotto, S. Sigurdsson, et al. // *Science*. – 2009. – V. 326. – P. 865-867.
2. Finno C.J. Applied equine genetics/C.J. Finno, D.L. Bannasch // *Equine Veterinary Journal*. - 2014. – V. 46. – P. 538-544.
3. Hill E.W. A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses/E.W. Hill, B.A. McGivney, J. Gu, et al. // *BMC Genomics*. – 2010. – V. 11. – P. 552-1-552-10.
4. Rieder S. Molecular tests for coat colours in horses/S. Rieder // *J Anim Breed Genet*. – 2009. – V. 126. – P. 415-424.
5. Brosnahan M.M. Equine clinical genomics: A clinician's primer/M.M. Brosnahan, S.A. Brooks, D.F. Antczak // *Equine Vet*. – 2010. – V. 42. – P. 658-670.
6. Chang C.C. Second-generation PLINK: rising to the challenge of larger and richer datasets / C.C. Chang, C.C. Chow, L.C. Tellier, et al. // *GigaScience*. - 2015. – V. 4. – P. 7-1-7-16.
7. Бейшова И.С. Изучение генетического разнообразия отечественных пород лошадей с использованием полногеномного анализа SNP / И.С. Бейшова, Д.А. Гриценко, М.Х. Шамекова, А.С. Пожарский, Т.В. Ульянова, А.М. Ковальчук // *Izdenister Natigeler*. – 2023. - № 3 (99). - С. 48-58.
8. McLaren W. The Ensembl Variant Effect Predictor/ W. McLaren, L. Gil, S.E. Hunt, et al. // *Genome Biology*. – 2016. – V. 17. – P. 122-1-122-14.
9. Sherman B.T. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)/ B.T. Sherman, M. Hao, J. Qiu, et al. // *Nucleic Acids Research*. - 2022. – V. 50. – P. 216-221.