IRSTI 34.23.37 https://doi.org/10.26577/bb106120266

A. Imanbay[1*], A. Amirgaliyeva[2], K. Ergali[2], M. Begmanova[2], G. Abylkassymova[2], A. Zhaxylykova[2], Zh. Kumarbekov[2], Zh. Seisenova[2], A. Zagitov[1]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]Institute of Genetics and Physiology SC MSHE RK, Almaty, Kazakhstan
*e-mail: phd.imanbay@gmail.com

GENOME-WIDE ASSOCIATION STUDY OF ENDURANCE GENES IN PROFESSIONAL WRESTLERS

The study aimed to identify genetic factors associated with endurance in professional wrestlers from the Republic of Kazakhstan using a genome-wide association study (GWAS). The dataset included 58 elite combat-sport athletes and 60 non-athletic controls. Genotyping was conducted on the iScan platform (Illumina) with the Infinium Global Screening Array-24. Quality control procedures in PLINK v1.9 and subsequent analyses in R (qqman) retained 296,772 SNPs after filtering for call rate, Hardy–Weinberg equilibrium, minor allele frequency, and exclusion of sex-chromosomal and mitochondrial variants. Principal component analysis showed no substantial population stratification, and the genomic inflation factor ($\lambda = 1.02$) indicated minimal confounding. At the suggestive significance level ($p < 1\times10^{-5}$), several loci showed differences between athletes and controls, including rs12916133, rs7765401, GSA-rs1682809, and rs4777639. These variants are located near PPEF2, SV2B, CDC37L1, and the non-coding region LOC105370982, which are genes and regulatory elements involved in calcium-dependent signaling, synaptic regulation, proteostasis, and muscular adaptation. Although none reached genome-wide significance ($p < 5\times10^{-8}$), the detected trends point to a contribution of regulatory and neurometabolic pathways to endurance-related traits in wrestlers.

The study is limited by the moderate sample size and the need for replication in independent Kazakhstani and international cohorts. Nevertheless, the findings form a basis for personalized training strategies and future integrative multi-omics research in sports genomics.

Keywords: GWAS, single nucleotide polymorphism (SNP), endurance, sports genetics, wrestlers, genetic polymorphism.

А. Иманбай[1*], А. Амиргалиева[2], К. Ергали[2], М. Бегманова[2], Г. Абылкасымова[2], А. Жаксылыкова[2], Ж. Құмарбеков[2], Ж. Сейсенова[2], А. Загитов[1]

[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
[2]ҚР ҒЖБМ ҒК Генетика және физиология институты, Алматы, Қазақстан
*e-mail: phd.imanbay@gmail.com

Кәсіби борысшылардағы шыдамдылықпен байланысты кең ауқымды геномдық ассоциативті зерттеу (GWAS)

Зерттеудің мақсаты – Қазақстан Республикасының кәсіби балуандарында төзімділік фенотипімен байланысатын генетикалық факторларды кең ауқымды геномдық ассоциативті талдау (GWAS) әдісі арқылы анықтау. Іріктемеге 58 жоғары білікті жекпе-жек спортшылары және бақылау тобына спорттық тәжірибесі жоқ 60 қатысушы алынды. Генотиптеу Illumina iScan платформасында Infinium Global Screening Array-24 жинағы арқылы жүргізілді. PLINK v1.9 бағдарламасында сапалық сүзгіден өткізу және R тіліндегі (qqman пакеті) статистикалық талдаудан кейін генотиптеу сапасы, Харди–Вайнберг теңдігі, минорлы аллель жиілігі және X/Y хромосомалары мен митохондриялық нұсқаны алып тастау критерийлеріне сәйкес 296 772 SNP талданды. Негізгі компоненттер талдауы нақты популяциялық стратификацияның жоқтығын, ал геномдық инфляция коэффициенті ($\lambda = 1.02$) шатастырушы факторлардың төмен деңгейін көрсетті. Болжамды маңыздылық шегінде ($p < 1\times10^{-5}$) спортшылар мен бақылау тобы арасында айырмашылығы бар бірнеше локус анықталды: rs12916133, rs7765401, GSA-rs1682809 және rs4777639. Бұл варианттар кальций сигнализациясы, синаптикалық беріліс, протеостаз және қаңқа бұлшықетінің бейімделу процестеріне қатысатын PPEF2, SV2B, CDC37L1 гендеріне және LOC105370982 кодталмайтын аймағына жақын орналасқан. Жалпы геномдық маңыздылық деңгейіне ($p < 5\times10^{-8}$) жетпегеніне қарамастан, алынған нәтижелер төзімділікке әсер ететін реттеуші және нейрометаболикалық жолдардың маңызын көрсетеді.

Зерттеудің шектеулері – іріктеменің салыстырмалы түрде шағын көлемі мен нәтижелерді Қазақстанның және халықаралық тәуелсіз когорталарында қайталап тексеру қажеттілігі болады. Соған қарамастан, бұл деректер спорттық геномика саласында жеке бағытталған жаттығу стратегияларын әзірлеуге және мультиомдық зерттеулерді дамытуға негіз бола алады.

Түйін сөздер: GWAS, SNP, төзімділік, спорт генетикасы, балуандар, генетикалық полиморфизм.

А. Иманбай[1*], А. Амиргалиева[2], К. Ергали[2], М. Бегманова[2], Г. Абылкасымова[2], А. Жаксылыкова[2], Ж. Кумарбеков[2], Ж. Сейсенова[2], А. Загитов[1]

[1]Казахский национальный университет имени аль-Фараби, Алматы, Казахстан
[2]Институт генетики и физиологии Комитета науки Министерства науки и высшего образования Республики Казахстан, Алматы, Казахстан
*e-mail: phd.imanbay@gmail.com

Геном-широкое ассоциативное исследование генов выносливости у профессиональных борцов

Целью работы было выявление генетических детерминант, связанных с развитием выносливости у профессиональных борцов Республики Казахстан, с использованием метода широкогеномного ассоциативного анализа (GWAS). В исследование были включены 58 высококвалифицированных спортсменов единоборств и 60 участников контрольной группы, не занимающихся спортом. Генотипирование выполнено на платформе iScan (Illumina) с использованием набора Infinium Global Screening Array-24. После процедур качества в PLINK v1.9 и последующей статистической обработки в R (пакет qqman) для анализа было оставлено 296 772 SNP-варианта, прошедших фильтры по частоте генотипирования, равновесию Харди–Вайнберга, минорной аллельной частоте и исключению вариантов X-, Y-хромосом и митохондриальной ДНК. Анализ главных компонент не выявил выраженной стратификации популяции, а коэффициент геномной инфляции ($\lambda = 1.02$) подтвердил отсутствие существенных смешивающих факторов. На уровне предположительной значимости ($p < 1\times10^{-5}$) были идентифицированы локусы, различающиеся между спортсменами и контролем: rs12916133, rs7765401, GSA-rs1682809 и rs4777639. Эти варианты располагаются вблизи генов PPEF2, SV2B, CDC37L1 и некодирующей области LOC105370982, участвующих в кальций-зависимой сигнализации, регуляции синаптической передачи, поддержании протеостаза и адаптации скелетной мышцы. Хотя ни один SNP не достиг общегеномного уровня значимости ($p < 5\times10^{-8}$), выявленные тенденции свидетельствуют о роли регуляторных и нейрометаболических механизмов в развитии выносливости у борцов.

Ограничениями исследования являются сравнительно небольшой размер выборки и необходимость подтверждения результатов в независимых казахстанских и международных когортных выборках. Тем не менее полученные данные формируют основу для разработки персонализированных тренировочных подходов и дальнейших мультиомных исследований в спортивной геномике.

Ключевые слова: GWAS, однонуклеотидный полиморфизм (SNP), выносливость, спортивная генетика, борьба, генетическая вариабельность.

Introduction

Sports performance and success in elite athletic disciplines reflect multifactorial phenomena shaped by the interplay of genetic determinants, phenotypic adaptations, and environmental influences. Among the key components of athletic capacity, endurance – typically assessed through VO_2max/VO_2peak, cardiorespiratory fitness (CRF), and indices of aerobic metabolic efficiency which have a distinct polygenic architecture, as demonstrated by large-scale genomic studies conducted in recent years (Williams et al., 2017; Zhao et al., 2020).

According to Ahmetov et al. (2023), a total of 251 genetic markers have been reported in association with diverse athletic performance traits. Of these, only 128 markers (~51%) have been replicated in at least two independent studies, including 41 associated with endurance, 45 with power/strength, and 42 with mixed phenotypes (Semenova et al., 2023).

A recent narrative review by Barkın Bıçakçı et al. (2024) highlighted the genes ACTN3, ACE, PPARA, VEGFA, and ADRB2 as key contributors to endurance phenotypes. However, the authors emphasized the need for multi-omics approaches and larger, standardized cohorts to ensure robust inference (Bıçakçı et al., 2024).

The shift from the traditional *candidate gene* paradigm to genome-wide association study (GWAS) has enabled the identification of dozens to hundreds of independent loci of small individual effect, which collectively explain a meaningful proportion of the variability in CRF and endurance-related traits. Multicenter

GWAS have revealed the involvement of pathways related to mitochondrial biogenesis, calcium homeostasis, endothelial function, and neuromuscular regulation (Eynon et al., 2011; Zhao et al., 2020).

Despite significant progress in population-based cohorts, the translation of genomic findings to sport-specific disciplines remains limited as most existing studies have been conducted in cyclic endurance sports (e.g., running, cycling), whereas combat sports, including wrestling, are largely underrepresented. The competitive profile of wrestlers is characterized by high-intensity intermittent load patterns, substantial reliance on buffering capacity, and a pronounced demand for aerobic compensation, which may indicate a distinct genetic architecture underlying their performance traits (Youn et al., 2021).

Recent domain-specific research indicates that genes such as PPARGC1A, PPARA, AMPD1, NRF1/NRF2, and ACTN3 play roles in the regulation of energy metabolism, oxidative phosphorylation, and resistance to fatigue (Maksimycheva, 2023; Tharabenjasin et al., 2019). Nevertheless, large-scale meta-analyses highlight substantial heterogeneity of genetic effects across ethnic groups and among diverse athletic specializations (Psatha et al., 2024; Ramadan et al., 2025).

Kazakhstan has long-standing and successful traditions in a wide range of combat sports – including Kazakh kuresi, Greco-Roman and freestyle wrestling, judo, sambo, karate, taekwondo, and Brazilian jiu-jitsu – with national athletes consistently demonstrating strong performance at international competitions. Owing to this rich athletic heritage, Kazakhstani combat-sport athletes represent a highly valuable cohort for investigating the genetic foundations of elite performance. A deeper understanding of their genetic predispositions Kazakhstani athletes may contribute to advancing personalized sports medicine and strengthening the country's position in the rapidly developing field of sports genomics.

In the present study, we applied genome-wide association analysis among professional combat-sport athletes from Kazakhstan to identify single nucleotide polymorphisms (SNPs) associated with endurance-related athletic characteristics, evaluate their relevance for endurance-related physiological traits, and explore their potential application in the development of an athlete's genetic performance profile.

Materials and methods

Peripheral blood samples stabilized with EDTA were collected from 58 elite combat-sport athletes specializing in wrestling (Masters of Sport and Candidates for Master of Sport (Republic of Kazakhstan), all of whom were medalists at national and Asian championships. The control group consisted of 60 non-athlete volunteers with no history of systematic sports training. All participants provided written informed consent prior to enrollment. The study protocol was approved by the Local Ethics Committee of the Institute of Genetics and Physiology, Ministry of Science and Higher Education of the Republic of Kazakhstan (Protocol No. 5, dated July 25, 2022).

Genomic DNA was extracted from frozen (–20°C) stabilized peripheral blood samples treated with EDTA using the GeneJet Genomic DNA Purification Kit (Thermo Scientific, USA) and the ReliaPrep Blood gDNA Miniprep System (Promega, USA), according to the manufacturers' protocols. DNA samples were stored at –20°C until analysis.

The concentration and purity of the isolated DNA were evaluated using a NanoDrop One spectrophotometer (ThermoScientific, USA), a Quantus fluorometer (Promega, USA), and 1.0% agarose gel electrophoresis. Only DNA samples with an A260/A280 purity ratio of 1.75–1.80 were included in subsequent genotyping. Samples containing RNA contamination (a ratio of 1.8–2.0) or protein impurities (ratio 1.5–1.7) were additionally washed and ethanol-precipitated to achieve the required purity.

Genome-wide SNP genotyping was performed using the Infinium Global Screening Array-24 Kit (Illumina, USA), containing 654,027 genome-wide markers. Genomic DNA samples were required to meet quality criteria, including a minimum concentration of 50 ng/μL and an A260/A280 purity a ratio of 1.75–1.80. Library preparation and sample handling were performed on the Freedom Evo automated workstation (Tecan, USA). Genotyping was then conducted on the Illumina iScan System following the Infinium HTS Automated Workflow provided by the manufacturer.

Initial processing of the SNP genotyping data was performed using GenomeStudio v2010.3 (Illumina, USA), ensuring a minimum call rate threshold of 98%. The bioinformatic pipeline included: verification of cluster and genotype quality ($p < 0.05$), determination of chromosomal location of SNPs (autosomal, sex chromosomes, mitochondrial genome), assessment of genotype status (homozygous/heterozygous at polymorphic loci). SNPs located in mitochondrial DNA and sex chromosomes were excluded, leaving only autosomal variants for downstream analyses.

A case–control genome-wide association analysis was carried out in PLINK v1.9. The quality con-

trol workflow involved excluding individuals with more than 10% missing genotypes (--mind 0.1), SNPs with a call rate below 95% (--geno 0.05), and markers deviating from Hardy–Weinberg equilibrium (--hwe 0.001). Following the application of these filters, a total of 296,772 SNPs were retained for downstream association testing.

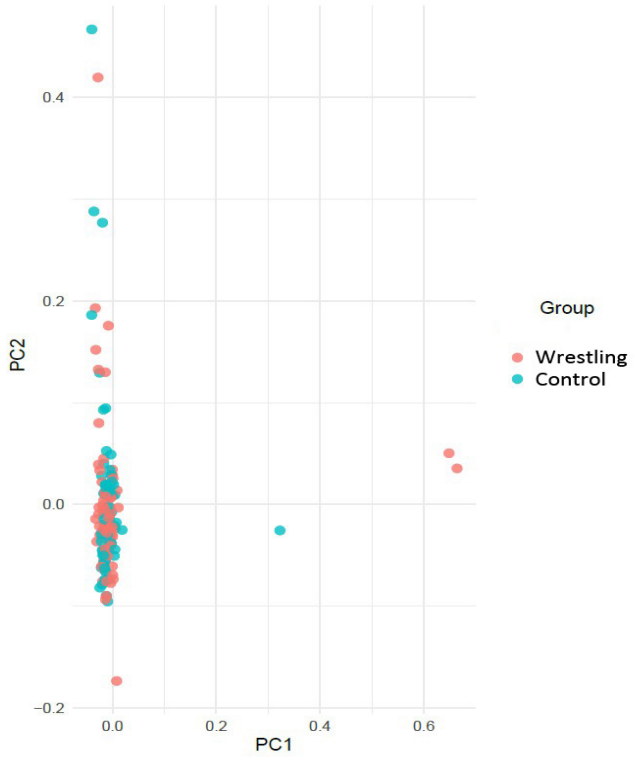
Data visualization was conducted in R v4.2.2 using the qqman package (Turner, 2018), and functional annotation of significant SNPs was performed using the BioMart package (Durinck et al., 2005, 2009).

Results and discussion

To ensure the validity of the genome-wide association analysis (GWAS), we first evaluated whether population stratification was present, as violations of genetic homogeneity can lead to spurious associations. For this purpose, a principal component analysis (PCA) was performed to visualize the extent of genetic similarity and differentiation among study participants. PCA reduces the multidimensional structure of genome-wide variation into a lower-dimensional space in which the distances between individuals reflect their relative genetic proximity.

Principal component analysis (PCA) demonstrated that the athlete cohort did not form a distinct genetic cluster and was not genetically differentiated from the control group. A substantial overlap was observed in the distribution of individuals from both groups, indicating the absence of significant population stratification (Fig. 1). This finding supported the validity of the cohort formation strategy and suggested that the observed associations were unlikely to result from underlying population structure, but rather reflected genuine genetic effects related to the endurance phenotype.

Figure 1
PCA on genetic similarity between athletes and control cohorts



A genome-wide association study (GWAS) was conducted to identify genetic factors associated with endurance in professional wrestlers. Several single nucleotide polymorphisms (SNPs) showed statistically significant associations with studied phenotype. As shown in Figure 2A, the distribution of p-values is presented in a circular Manhattan plot, where $-\log_{10}(p)$ values are plotted

against the chromosomal positions of the respective SNPs.

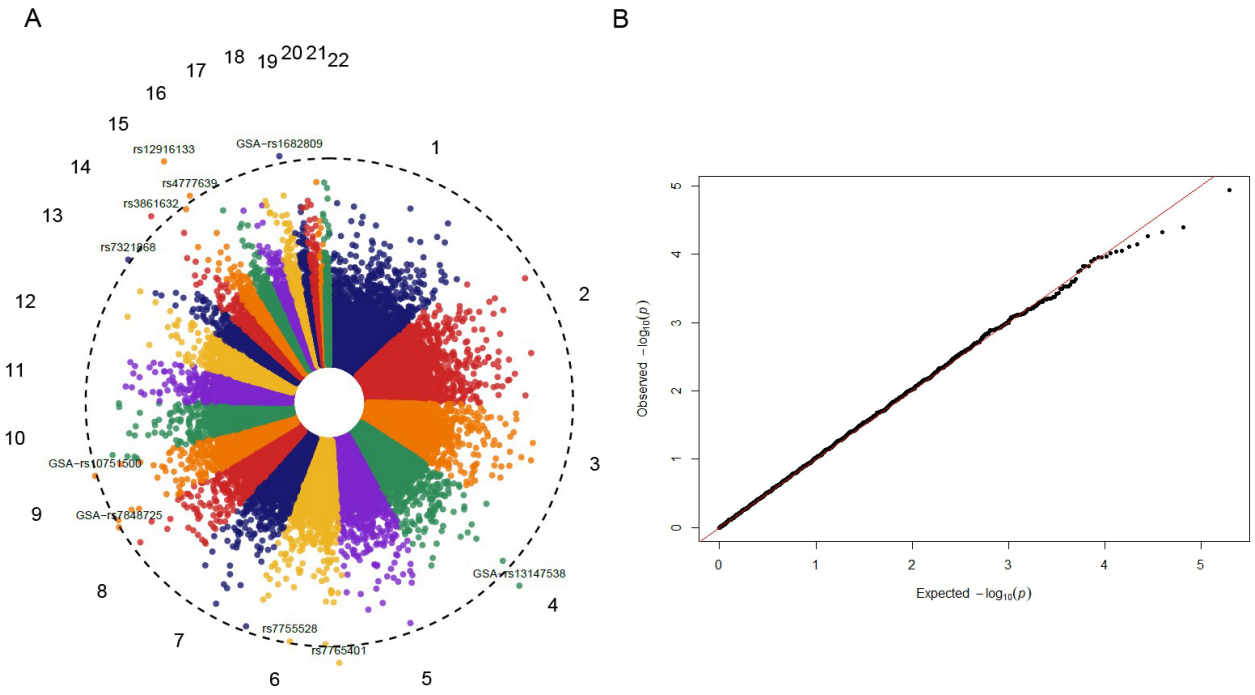
Among the variants exceeding the predefined suggestive significance threshold ($p < 1\times10^{-5}$), the loci rs7765401, rs12916133, rs3861632, and GSA-rs1682809 were identified, indicating plausible associations with endurance-related traits. These loci warrant further functional characterization due to their biological plausibility and statistical significance.

For example, the frequency of the A allele of rs12916133 in the general population was 0.421, whereas in the athlete group it was 0.3525, suggesting possible adaptive genetic differences between wrestlers and non-athletes. Comparable allele frequency differences were also observed for rs3861632, rs4777639, and rs7321868. These findings allowed to evaluate the contribution of the identified factors to the endurance phenotype formation and provided a foundation for ethnogenetic and functional analyses.

The circular Manhattan plot illustrated the distribution of SNP markers across all chromosomes (Fig. 2a). The radial axis represented the level of statistical significance ($-log_{10}(p)$). One SNP exhibited the strongest association signal and was highlighted in the outer ring of the plot, indicating its potential relevance to the endurance phenotype in professional wrestlers.

Figure 2

(a) Circular Manhattan plot of GWAS results; (b) Q–Q plot based on SNP-specific p-values, analysed independently of sport specialisation while accounting for the predisposition profile



To evaluate the distribution of *p*-values and assess potential statistical inflation, a quantile–quantile (Q–Q) plot was constructed (Fig. 2b). The observed $-log_{10}(p)$ values closely followed those expected under the null hypothesis, indicating the absence of systematic bias, population stratification, or technical artefacts that could have influenced the results. A slight deviation in the upper right tail of the distribution reflected the presence of true associations identified in the analysis and corresponded to the significant SNP markers detected.

Thus, the GWAS identified several SNPs significantly associated with the endurance phenotype. The results of quality control procedures, PCA, and Q–Q analysis supported the robustness of the model and confirmed the absence of systematic bias. These findings provided a basis for further functional interpretation of the identified loci and for in-depth investigation of the molecular mechanisms underlying endurance development in athletes.

Following this, odds ratios (ORs) were calculated for the identified variants, and the results are pre-

sented in Table 1. Notably, the markers rs12916133 ($p = 1.15\times10^{-5}$, OR = 0.3121), rs7765401 ($p = 4.77\times10^{-5}$, OR = 2.965), GSA-rs1682809 ($p = 7.24\times10^{-5}$, OR = 3.01), and rs4777639 ($p = 7.89\times10^{-5}$, OR = 0.2611) showed the highest level of association with endurance-related performance indicators. Several of the detected variants such as rs12916133 and rs7321868 – indicate a possible protective effect (OR < 1), while others, including rs7765401 and rs1682809, may be associated with physiological characteristics that contribute to enhanced endurance capacity.

Table 1
SNPs identified by GWAS and associated with endurance performance in professional wrestlers

CHR	SNP	BP	A1	F_A	F_U	A2	CHISQ	P	OR
4	GSA-rs13147538	76818367	A	0.02459	0.1864	G	16.84	4.07E-05	0.11
6	rs7765401	9104960	A	0.5656	0.3051	G	16.54	4.77E-05	2.965
9	GSA-rs10751500	136057106	A	0.582	0.3305	G	15.28	9.28E-05	2.82
9	GSA-rs7848725	4678754	G	0.5902	0.339	A	15.21	9.64E-05	2.808
13	rs7321868	64864846	A	0	0.1186	G	15.37	8.83E-05	0
14	rs3861632	78540046	T	0.2705	0.5254	C	16.3	5.40E-05	0.3349
15	rs12916133	91705749	A	0.3525	0.6356	G	19.24	1.15E-05	0.3121
15	rs4777639	93810409	A	0.1066	0.3136	G	15.58	7.89E-05	0.2611
19	GSA-rs1682809	3165895	A	0.4836	0.2373	G	15.75	7.24E-05	3.01

Several genetic loci associated with endurance performance in professional wrestlers were identified. Functional characterisation of these loci may aid the clarification of the biological pathways through which they influence endurance-related phenotypes. The *PPEF2* gene (*rs13147538*), located on chromosome 4 (*4q25*), encodes a serine/threonine phosphatase containing EF-hand-like calcium-binding domains. The SNP *rs13147538* was situated within the intronic region and the 5′-untranslated region (5′-UTR), suggesting a potential regulatory role in gene transcription. Although direct evidence linking this factor to endurance performance was lacking, phosphatases were known to regulate energy metabolism, calcium signalling, and cellular stress responses. These processes are central to physiological adaptation during prolonged physical exercise. Therefore, regulatory variation in the PPEF2 gene may affect molecular pathways that determine recovery dynamics and endurance. Further studies, including eQTL analyses and promoter activity experiments, are needed to clarify the functional role of this variant (ENCODE Project Consortium, 2020; NCBI, 2025; UniProt Consortium, 2024). The rs7765401 variant is located in an immunoregulatory region on chromosome 6 that includes the HLA-C and MICB genes. Although this SNP has not been directly characterized in previous GWAS studies, literature reports indicate that endurance athletes show increased expression of mitochondrial oxidative metabolism genes and reduced inflammatory activity in peripheral leukocytes. These data suggest that immuno-inflammatory regulation and recovery play an important role in endurance adaptation (Liu et al., 2017).

Overall, our results support a multilevel genetic architecture of endurance. This architecture encompasses several biological mechanisms, including fatty acid metabolism (RAB8A), immune regulation (HLA), and membrane glycolipid stability (GBGT1). The variant rs10751500, located near the GBGT1 gene on chromosome 9, also showed an association with endurance (OR ≈ 2.82). GBGT1 encodes a glycosyltransferase involved in the biosynthesis of glycolipids. These glycolipids contribute to the formation of membrane rafts, which in turn ensure the assembly of receptor complexes and membrane stability. Current studies suggest that membrane glycolipids play an important role in modulating stress-response signaling pathways activated during high-intensity exercise (National Center for Biotechnology Information, 2025; Furukawa et al., 2016). The rs10751500 variant may reflect genetic differences in membrane organization and signaling that influence muscle adaptation to exercise load.

The CDC37L1 gene (rs7848725), located in an intronic region on chromosome 9, influences gene

expression and supports the proper assembly and stability of proteins under stress conditions. The roles of molecular chaperones in maintaining proteostasis are well known. This locus is a promising candidate for sports genomics (Li et al., 2021; Peak et al., 2021). rs7321868 is located on chromosome 13. This variant was not found in any of the athletes, but was detected only in the control group (OR → 0). The closest functional gene is SPRY2, which is an inhibitor of the MAPK/ERK signaling pathway. This pathway is crucial for muscle regeneration, plasticity and hypertrophy. Increased SPRY2 activity has been shown to reduce MAPK-mediated adaptive responses and may negatively (Milillo et al., 2015). Satellite cells, the cell cycle, and skeletal muscle regeneration. affect muscle recovery and exercise tolerance (Forcina & Rudnicki, 2019; Dumont et al., 2015).

rs3861632, located near the DLGAP5 gene on chromosome 14, showed a clear protective effect (OR ≈ 0.33). DLGAP5 is involved in the regulation of the cell cycle. Therefore, rs3861632 may be a marker that reflects individual differences in regenerative potential among athletes (Chen et al., 2024)

The SV2B gene (rs12916133), located on chromosome 15, is involved in the regulation of synaptic neurotransmitter release. Intronic rs12916133 may affect gene expression and the efficacy of neuronal signal transmission (Morgans et al., 2009; Stout et al., 2019). Since neuromotor coordination and rapid reaction times are particularly important in combat sports, this locus, although not directly associated with endurance, may indirectly influence athletic performance. The noncoding locus LOC105370982 (rs4777639) on chromosome 15 is located in an intronic region and is likely to affect transcriptional regulation or alternative splicing. Given that a significant proportion of GWAS signals associated with athletic traits are located in noncoding regulatory elements, rs4777639 is a promising candidate for functional annotation, such as chromatin accessibility analysis or enhancer activity studies (ENCODE Consortium, 2020). Another variant of particular interest is GSA-rs1682809, located near the RAB8A gene on chromosome 19. According to the literature, RAB8A functions as a lipid droplet receptor involved in the delivery of long-chain fatty acids to mitochondria and facilitates β-oxidation (Ouyang et al., 2023). Since endurance performance is directly dependent on the ability of muscles to use fat as a primary energy source, the observed association (OR ≈ 3.0) suggests that this variant may enhance metabolic adaptation.

Conclusion

In this genome-wide association analysis, nine single nucleotide polymorphisms (GSA-rs13147538, rs7765401, GSA-rs10751500, GSA-rs7848725, rs7321868, rs3861632, rs12916133, rs4777639, and GSA-rs1682809) were identified as significantly associated with athletic performance traits in professional wrestlers. Of particular interest are the loci situated within genes that may serve as novel candidate markers influencing the development of aerobic capacity and endurance-related phenotypes. The functional characteristics of these genes – including their involvement in calcium signaling, synaptic transmission, and proteostatic regulation – are consistent with the physiological mechanisms underlying adaptation to prolonged and high-intensity physical exertion in combat-sport athletes.

The obtained results expand the existing catalog of genetic determinants contributing to interindividual variability in sports performance and further support the polygenic nature of endurance and recovery-related traits. Despite the limited sample size and the need for subsequent functional validation, this study provides a foundation for future replication efforts and multi-omics research focused on athletic populations from Central Asia. Integrating genomic insights into training practices may significantly enhance precision sports medicine and enable personalized athlete development in Kazakhstan.

Acknowledgements

The authors gratefully acknowledge the participation of all athletes involved in this study. The authors further express their deep appreciation to L.B. Djansugurova, N.K. Altynova, G.S. Zhunussova, N.V. Mit, A. Kasymbekova, A. Erzhan, and B.O. Bekmanov for their invaluable support throughout all stages of the research and for their substantial contributions to its successful execution.

Funding

The present research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan under Project No. BR1854139, “Development of an integrated system for training highly qualified athletes and promising Olympic reserve candidates in priority sports using physiological and genetic assessment”, for the period 2023–2024.

Author Contributions

A. Imanbay: Formal Analysis, Writing – Original Draft, Writing – Review & Editing; A. Amirgaliyeva: Investigation, Resources, Supervision; K. Ergali: Investigation, Validation, Formal Analysis; M. Begmanova: Investigation, Visualization; G. Abylkassymova: Investigation, Software; A. Zhaxylykova: Investigation, Data Curation; Zh. Kumarbekov: Investigation, Data Curation; Zh. Seisenova: Investigation, Methodology; A. Zagitov: Investigation, Methodology.

References

Bıçakçı, B., G. Öztürk, and M. Atalay. (2024). Genetic determinants of endurance: A narrative review. *International Journal of Molecular Sciences, 25*(23), 13041. https://doi.org/10.3390/ijms252313041

Durinck, S., et al. (2005). BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics, 21*(16), 3439–3440. https://doi.org/10.1093/bioinformatics/bti525

Durinck, S., P. T. Spellman, E. Birney, and W. Huber. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols, 4*(8), 1184–1191. https://doi.org/10.1038/nprot.2009.97

Dumont, N. A., Bentzinger, C. F., & Rudnicki, M. A. (2015). Satellite cells and skeletal muscle regeneration. *Comprehensive Physiology, 5*(3), 1027–1059. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10778731/

The ENCODE Project Consortium. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, *583*(7818), 699–710. https://doi.org/10.1038/s41586-020-2493-4

Eynon, N., J. R. Ruiz, Y. Meckel, M. Morán, and A. Lucia. (2011). Mitochondrial biogenesis related endurance genotype score and sports performance in athletes. *Mitochondrion, 11*(1), 64–69. https://doi.org/10.1016/j.mito.2010.07.004

Furukawa, K., W. Aixinjueluo, et al. (2016). Regulation of the human GBGT1 (Forssman Synthase) gene. https://pmc.ncbi.nlm.nih.gov/articles/PMC5164903/

Forcina, L., & Rudnicki, M. A. (2019). An overview about the biology of skeletal muscle satellite cells. *International Journal of Molecular Sciences, 20*(6), Article 1457. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6446479/

Liu, D., M. A. Sartor, and G. A. Nader. (2017). Exercise-induced gene expression and adaptations in endurance athletes. *BMC Genomics, 18*, 1–13. https://doi.org/10.1186/s12864-016-3388-5

Maksimycheva T. Yu., Kondratyeva E. I., Popova V. M. Molecular-genetic foundations of energy exchange and physical qualities of man. Research perspectives. Experimental and Clinical Gastroenterology. 2023;217(9): 222–230. (In Russ.) https://doi.org/10.31146/1682-8658-ecg-217-9-222-230

Milillo, A., La Carpia, F., Costanzi, S., D'Urbano, V., Martini, M., Lanuti, P., Marchisio, M., & Gurrieri, F. (2015). A SPRY2 mutation leading to MAPK/ERK pathway inhibition is associated with an autosomal dominant form of IgA nephropathy. *European Journal of Human Genetics, 23*, 1673–1678. https://doi.org/10.1038/ejhg.2015.52

Morgans, C. W., Kensel-Hammes, P., Hurley, J. B., Burton, K., Idzerda, R., McKnight, G. S., & Bajjalieh, S. M. (2009). Loss of the synaptic vesicle protein SV2B results in reduced neurotransmission and altered synaptic vesicle protein expression in the retina. *PLoS ONE, 4*(4), e5230. https://doi.org/10.1371/journal.pone.0005230

National Center for Biotechnology Information. (2025, November 25). *GBGT1 globoside alpha-1,3-N-acetylgalactosaminyltransferase 1 (FORS blood group) [Homo sapiens]*. Gene. https://www.ncbi.nlm.nih.gov/gene/26301

Li, L., Tao, X., Li, Y., Gao, Y., & Li, Q. (2021). CDC37L1 acts as a suppressor of migration and proliferation in gastric cancer by down-regulating CDK6. Journal of Cancer, 12(11), 3145–3153. https://doi.org/10.7150/jca.56097

NCBI. (2025). Gene Database. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/gene/

Ouyang, Q., Q. Chen, S. Ke, et al. (2023). RAB8A as a mitochondrial receptor for lipid droplets in skeletal muscle: Implications for fatty acid oxidation and endurance. *Developmental Cell, 58*(4), 289–305.e6. https://doi.org/10.1016/j.devcel.2023.01.007

Peak, S. L., Gracia, L., Lora, G., & Jinwal, U. K. (2021). Hsp90-interacting Co-chaperones and their Family Proteins in Tau Regulation: Introducing a Novel Role for Cdc37L1. *Neuroscience*, *453*, 312–323. https://doi.org/10.1016/j.neuroscience.2020.11.020

Psatha, A., et al. (2024). Genetic polymorphisms and athlete status: A systematic review and meta-analysis. *Human Genomics, 18*(1), 21. https://doi.org/10.1186/s40246-024-00621-9

Ramadan, W., et al. (2025). Effect of polymorphisms of genes related to sport performance in adolescent athletes. *Life, 15*(3), 477. https://doi.org/10.3390/life15030477

Semenova, E. A., E. C. R. Hall, and I. I. Ahmetov. (2023). Genes and athletic performance: The 2023 update. *Genes, 14*(6), 1235. https://doi.org/10.3390/genes14061235

Stout, K. A., Dunn, A. R., Hoffman, C., & Miller, G. W. (2019). The synaptic vesicle glycoprotein 2: Structure, function, and disease relevance. ACS Chemical Neuroscience, 10(9), 3927–3938. https://doi.org/10.1021/acschemneuro.9b00351

Tharabenjasin, P., N. Pabalan, and P. Suntarattiwong. (2019). Association of the PPARGC1A Gly482Ser polymorphism with athletic performance: A meta-analysis. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0200967

Turner, S. (2018). qqman: An R package for visualizing GWAS results using Q–Q and Manhattan plots. *Journal of Open Source Software, 3*(25), 731. https://doi.org/10.21105/joss.00731

UniProt Consortium. (2024). UniProt Knowledgebase. https://www.uniprot.org/

Chen, M., Zhang, S., Wang, F., He, J., Jiang, W., & Zhang, L. (2024). DLGAP5 promotes lung adenocarcinoma growth via upregulating PLK1 and serves as a therapeutic target. *Journal of translational medicine*, *22*(1), 209. https://doi.org/10.1186/s12967-024-04910-8

Williams, C. J., et al. (2017). Genes to predict VO_2max trainability: A systematic review. *BMC Genomics, 18*, 559. https://doi.org/10.1186/s12864-017-4192-6

Wen, X., Jiao, L., & Tan, H. (2022). MAPK/ERK Pathway as a Central Regulator in Vertebrate Organ Regeneration. *International journal of molecular sciences*, *23*(3), 1464. https://doi.org/10.3390/ijms23031464

Youn, B. Y., S. G. Ko, and J. Y. Kim. (2021). Genetic basis of elite combat sports athletes: A systematic review. *Biology of Sport, 38*(4), 667–675. https://doi.org/10.5114/biolsport.2022.102864

Zhao, Y., et al. (2020). Four loci are associated with cardiorespiratory fitness and endurance performance in young Chinese females. *Scientific Reports, 10*, 10117. https://doi.org/10.1038/s41598-020-67045-y

Information about the authors:

A. Imanbay (corresponding author) – PhD candidate of al-Farabi Kazakh National University, Senior Lecturer, Faculty of Medicine and Healthcare, al-Farabi Kazakh National University (Almaty, Kazakhstan; e-mail: phd.imanbay@gmail.com).

A. Amirgaliyeva – Senior Research Scientist at the Laboratory of Molecular Genetics, Institute of Genetics and Physiology SC MSHE RK (Almaty, Kazakhstan; e-mail: almira-71@mail.ru).

K. Ergali – Junior Research Scientist at the Laboratory of Molecular Genetics, РГП «Institute of Genetics and Physiology SC MSHE RK (Almaty, Kazakhstan; e-mail: ergali.0394@mail.ru).

M. Begmanova – Senior Research Scientist at the Laboratory of Molecular Genetics, Institute of Genetics and Physiology SC MSHE RK (Almaty, Kazakhstan; e-mail: bmamura@mail.ru).

G. Abylkassymova – Senior Research Scientist at the Laboratory of Molecular Genetics, Institute of Genetics and Physiology SC MSHE RK (Almaty, Kazakhstan; e-mail: gulyam05@mail.ru).

A. Zhaxylykova – Junior Research Scientist at the Laboratory of Molecular Genetics, Institute of Genetics and Physiology SC MSHE RK (Almaty, Kazakhstan; e-mail: asel_zhaksylykov@mail.ru).

Zh. Kumarbekov – Junior Research Scientist at the Laboratory of Molecular Genetics, Institute of Genetics and Physiology SC MSHE RK (Almaty, Kazakhstan; e-mail: zhantkm@gmail.com).

Zh. Seisenova – Senior Laboratory Assistant, Laboratory of Molecular Genetics, RSE REM "Institute ofGenetics and Physiology" SC MSHE RK (Almaty, Kazakhstan, e-mail: s_seisenova@mail.ru).

A. Zagitov – master's student of al-Farabi Kazakh National University (Almaty, Kazakhstan; e-mail: almazzagit0420@gmail.com).

Авторлар туралы мәлімет:

А. Иманбай (корреспондент-автор) – әл-Фараби атындағы Қазақ ұлттық университетінің докторанты, әл-Фараби атындағы Қазақ ұлттық университеті Медицина және денсаулық сақтау факультетінің аға оқытушысы (Алматы, Қазақстан; e-mail: phd.imanbay@gmail.com).

А. Әмірғалиева – ҚР ҒЖБМ ҒК «Генетика және физиология институты» молекулалық генетика зертханасының аға ғылыми қызметкері (Алматы, Қазақстан; e-mail: almira-71@mail.ru).

Қ. Ергали – ҚР ҒЖБМ ҒК «Генетика және физиология институты» РМК молекулалық генетика зертханасының кіші ғылыми қызметкері (Алматы, Қазақстан; e-mail: ergali.0394@mail.ru).

М. Бегманова – ҚР ҒЖБМ ҒК «Генетика және физиология институты» молекулалық генетика зертханасының аға ғылыми қызметкері (Алматы, Қазақстан; e-mail: bmamura@mail.ru).

Г. Әбілқасымова – ҚР ҒЖБМ ҒК «Генетика және физиология институты» молекулалық генетика зертханасының аға ғылыми қызметкері (Алматы, Қазақстан; e-mail: gulyam05@mail.ru).

А. Жақсылықова – ҚР ҒЖБМ ҒК «Генетика және физиология институты» молекулалық генетика зертханасының кіші ғылыми қызметкері (Алматы, Қазақстан; e-mail: asel_zhaksylykov@mail.ru).

Ж. Құмарбеков – ҚР ҒЖБМ ҒК «Генетика және физиология институты» молекулалық генетика зертханасының кіші ғылыми қызметкері (Алматы, Қазақстан; e-mail: zhantkm@gmail.com).

Ж. Сейсенова – ҚР ҒЖБМ ҒК ШЖҚ РМК «Генетика және физиология институты» молекулалық генетика зертханасының аға зертханашысы (Алматы, Қазақстан; e-mail: s_seisenova@mail.ru).

А. Загитов – әл-Фараби атындағы Қазақ ұлттық университетінің магистранты (Алматы, Қазақстан; e-mail: almazzagit0420@gmail.com).

Сведения об авторах:

А. Иманбай (автор-корреспондент) – докторант Казахского национального университета имени аль-Фараби, старший преподаватель факультета медицины и здравоохранения Казахского национального университета имени аль-Фараби (Алматы, Казахстан, e-mail: phd.imanbay@gmail.com).

А. Амиргалиева – старший научный сотрудник лаборатории молекулярной генетики Института генетики и физиологии КН МНВО РК (Алматы, Казахстан, e-mail: almira-71@mail.ru).

К. Ергали – младший научный сотрудник лаборатории молекулярной генетики, РГП «Институт генетики и физиологии» КН МНВО РК (Алматы, Казахстан, e-mail: ergali.0394@mail.ru).

М. Бегманова – старший научный сотрудник лаборатории молекулярной генетики Института генетики и физиологии КН МНВО РК (Алматы, Казахстан, e-mail: bmamura@mail.ru).

Г. Абылкасымова – старший научный сотрудник лаборатории молекулярной генетики Института генетики и физиологии КН МНВО РК (Алматы, Казахстан, e-mail: gulyam05@mail.ru).

А. Жаксылыкова – младший научный сотрудник лаборатории молекулярной генетики Института генетики и физиологии КН МНВО РК (Алматы, Казахстан, e-mail: asel_zhaksylykov@mail.ru).

Ж. Кумарбеков – младший научный сотрудник лаборатории молекулярной генетики Института генетики и физиологии КН МНВО РК (Алматы, Казахстан, e-mail: zhantkm@gmail.com).

Ж. Сейсенова – старший лаборант лаборатории молекулярной генетики, РГП на ПХВ «Институт генетики и физиологии» КН МНВО РК (Алматы, Казахстан, e-mail: s_seisenova@mail.ru).

А. Загитов – магистрант Казахского национального университета имени аль-Фараби (Алматы, Казахстан, e-mail: almazzagit0420@gmail.com).

Received July 9, 2025
Accepted February 20, 2026