Whole genome sequencing of elite athletes

Boulygina E.A., Borisov O.V., Valeeva E.V., Semenova E.A., Kostryukova E.S., Kulemin N.A., Larin A.K., Nabiullina R.M., Mavliev F.A., Akhatov A.M., Andryushchenko O.N., Andryushchenko L.B., Zmijewski P., Generozov E.V., Ahmetov I.I.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2020 Institute of Sport. All rights reserved. Whole genome sequencing (WGS) has great potential to explore all possible DNA variants associated with physical performance, psychological traits and health conditions of athletes. Here we present, for the first time, annotation of genomic variants of elite athletes, based on the WGS of 20 Tatar male wrestlers. The maximum number of high-quality variants per sample was over 3.8 M for single nucleotide polymorphisms (SNPs) and about 0.64 M for indels. The maximum number of nonsense mutations was 148 single nucleotide variants (SNVs) per individual. Athletes' genomes on average contained 18.9 nonsense SNPs in a homozygous state per sample, while non-Athletes' exomes (Tatar controls, n = 19) contained 18 nonsense SNPs. Finally, we applied genomic data for the association analysis and used reaction time (RT) as an example. Out of 1884 known genome-wide significant SNPs related to RT, we identified four SNPs (KIF27 rs10125715, APC rs518013, TMEM229A rs7783359, LRRN3 rs80054135) associated with RT in wrestlers. The cumulative number of favourable alleles (KIF27 A, APC A, TMEM229A T, LRRN3 T) was significantly correlated with RT both in wrestlers (P = 0.0003) and an independent cohort (n = 43) of physically active subjects (P = 0.029). Furthermore, we found that the frequencies of the APC A (53.3 vs 44.0%, P = 0.033) and LRRN3 T (7.5 vs 2.8%, P = 0.009) alleles were significantly higher in elite athletes (n = 107) involved in sports with RT as an essential component of performance (combat sports, table tennis and volleyball) compared to less successful (n = 176) athletes. The LRRN3 T allele was also over-represented in elite athletes (7.5%) in comparison with 189 controls (2.9%, P = 0.009). In conclusion, we present the first WGS study of athletes showing that WGS can be applied in sport and exercise science.

Keywords

Athletic performance, Genotype, Polymorphism, Reaction time, Wrestling

References

[1] Maciejewska-Skrendo A, Sawczuk M, Cięszczyk P, Ahmetov II. Genes and Power Athlete Status. In: Barh D, Ahmetov I, editors. Sports, Exercise, and Nutritional Genomics: Current Status and Future Directions. Academic Press; 2019. p. 41-72.

[2] Semenova EA, Fuku N, Ahmetov II. Genetic profile of elite endurance athletes. In: Barh D, Ahmetov I, editors. Sports, Exercise, and Nutritional Genomics: Current Status and Future Directions. Academic Press; 2019. p. 73-104.

[3] Valeeva EV, Ahmetov II, Rees T. Psychogenetics and sport. In: Barh D, Ahmetov I, editors. Sports, Exercise, and Nutritional Genomics: Current Status and Future Directions. Academic Press; 2019. p. 147-165.

[4] Mustafina LJ, Naumov VA, Cieszczyk P, et al. AGTR2 gene polymorphism is associated with muscle fibre composition, athletic status and aerobic performance. Exp Physiol. 2014; 99(8):1042-52.

[5] Guilherme JPLF, Egorova ES, Semenova EA, et al. The A-Allele of the FTO Gene rs9939609 Polymorphism Is Associated With Decreased Proportion of Slow Oxidative Muscle Fibers and Over-represented in Heavier Athletes. J Strength Cond Res. 2019; 33(3):691-700.

[6] Semenova EA, Miyamoto-Mikami E, Akimov EB, et al. The association of HFE gene H63D polymorphism with endurance athlete status and aerobic capacity: novel findings and a meta-Analysis. Eur J Appl Physiol. 2020; 120(3):665-673.

[7] Kusić D, Connolly J, Kainulainen H, et al. Striated muscle-specific serine/ threonine-protein kinase beta segregates with high versus low responsiveness to endurance exercise training. Physiol Genomics. 2020;52(1):35-46.

[8] Ahmetov I, Kulemin N, Popov D, et al. Genome-wide association study identifies three novel genetic markers associated with elite endurance performance. Biol Sport. 2015;32(1):3-9.

[9] Rankinen T, Fuku N, Wolfarth B, et al. No Evidence of a Common DNA Variant Profile Specific to World Class Endurance Athletes. PLoS One. 2016; 11(1):e0147330.

[10] Pickering C, Suraci B, Semenova EA, et al. A genome-wide association study of sprint performance in elite youth football players. J Strength Cond Res. 2019; 33:2344-2351.

[11] Al-Khelaifi F, Yousri NA, Diboun I, et al. Genome-wide association study reveals a novel association between MYBPC3 gene polymorphism, endurance athlete status, aerobic capacity and steroid metabolism. Front Genet. 2020. doi: 10.3389/fgene.2020.00595.

[12] Ackland TR, Elliott BC, Bloomfield J. Applied anatomy and biomechanics in sport. Human Kinetics; 2009.

[13] Gierczuk D, Bujak Z, Cieslinski I, et al. Response Time and Effectiveness in Elite Greco-Roman Wrestlers Under Simulated Fight Conditions. J Strength Cond Res. 2018;32(12):3433-3440.

[14] Kuntsi J, Rogers H, Swinard G, et al. Reaction time, inhibition, working memory and delay aversion' performance: genetic influences and their interpretation. Psychol Med. 2006;36(11):1613-24.

[15] Kikuchi N, Min SK, Ueda D, et al. Higher frequency of the ACTN3 R allele + ACE DD genotype in Japanese elite wrestlers. J Strength Cond Res. 2012; 26(12):3275-3280.

[16] Gabbasov RT, Arkhipova AA, Borisova AV, et al. The HIF1A gene Pro582Ser polymorphism in Russian strength athletes. J Strength Cond Res. 2013; 27(8):2055-2058.

[17] Kikuchi N, Ueda D, Min SK, Nakazato K, Igawa S. The ACTN3 XX genotype's underrepresentation in Japanese elite wrestlers. Int J Sports Physiol Perform. 2013;8(1):57-61.

[18] Kikuchi N, Fuku N, Matsumoto R, et al. The Association Between MCT1 T1470A Polymorphism and Power-Oriented Athletic Performance. Int J Sports Med. 2017;38(1):76-80.

[19] Boulygina EA, Lukianova E, Grigoryeva TV, et al. Lessons from the Whole Exome Sequencing Effort in Populations of Russia and Tajikistan. BioNanoSci. 2016;6:540-542.

[20] Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589-95.

[21] Kim S, Scheffler K, Halpern AL, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods. 2018;15(8):591-594.

[22] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-Throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

[23] Davies G, Marioni RE, Liewald DC. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N = 112 151). Mol Psychiatry. 2016; 21(6):758-67.

[24] Davies G, Lam M, Harris SE, et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun. 2018;9(1):2098.

[25] Wang J, Raskin L, Samuels DC, et al. Genome measures used for quality control are dependent on gene function and ancestry. Bioinformatics. 2015;31(3):318-23.

[26] Nagasaki M, Yasuda J, Katsuoka F, et al. Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. Nat Commun. 2015;6:8018.

[27] Sidore C, Busonero F, Maschio A, et al. Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. Nat Genet. 2015;47(11):1272-1281.

[28] Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet. 2014;46(8):818-25.

[29] Triska P, Chekanov N, Stepanov V, et al. Between Lake Baikal and the Baltic Sea: genomic history of the gateway to Europe. BMC Genet. 2017;18(1):110.

[30] GTEx Consortium. Genetic effects on gene expression across human tissues. Nature. 2017;550(7675):204-213.

[31] Coyle YM, Xie XJ, Lewis CM, et al. Role of physical activity in modulating breast cancer risk as defined by APC and RASSF1A promoter hypermethylation in nonmalignant breast tissue. Cancer Epidemiol Biomarkers Prev. 2007; 16(2):192-6.

[32] Katoh Y, Katoh M. Characterization of KIF7 gene in silico. Int J Oncol. 2004; 25(6):1881-6.

[33] Kang J, Bai R, Liu K, et al. Identification of significantly different modules between permanent and deciduous teeth by network and pathway analyses. Genet Mol Res. 2016;15(4).

[34] Yang J, Li F, Qiu L, et al. Role of LRRN3 in the cerebellum postnatal development in rats. Zhong Nan Da Xue Xue Bao Yi Xue Ban. 2011;36(5):424-9.

[35] Telenti A, Pierce LC, Biggs WH, et al. Deep sequencing of 10,000 human genomes. Proc Natl Acad Sci U S A. 2016;113(42):11901-11906.

[36] Stubbs A, McClellan EA, Horsman S, et al. Huvariome: A web server resource of whole genome next-generation sequencing allelic frequencies to aid in pathological candidate gene selection. J Clin Bioinforma. 2012;2(1):19.

[37] Rustagi N, Zhou A, Watkins WS, et al. Extremely low-coverage whole genome sequencing in South Asians captures population genomics information. BMC Genomics. 2017;18(1):396.

[38] Tanisawa K, Wang G, Seto J et al. Sport and exercise genomics: The FIMS 2019 consensus statement update. Br J Sports Med. 2020 Mar 22. doi: 10.1136/bjsports-2019-101532.