DOI 10.18699/SBB-2022-219

Loss-of-function mutations in athletes

Boulygina E.A.[1]*, Valeeva E.V.[1, 2], Semenova E.A.[3, 4], Larin A.K.[3], Nabiullina R.M.[2], Mavliev F.A.[4], Akhatov A.M.[4], Generozov E.V.[3], Ahmetov I.I.[2, 3, 5]

[1] *Kazan (Volga region) Federal University, Kazan, Russia*

[2] *Kazan State Medical University, Kazan, Russia*

[3] *Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia*

[4] *Volga Region State Academy of Physical Culture, Sport and Tourism, Kazan, Russia*

[5] *Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, United Kingdom*

* *boulygina@gmail.com*

Key words: LoF mutation, athletes, muscle mass, peak power, reaction time, speed performance

Motivation and Aim: Whole-genome and whole-exome sequencing data reveal that individuals carry many variants predicted to inactivate genes (knockouts). Some of naturally-occurring knockout or loss-of-function (LoF) variants may potentially influence athletic performance. The aim of the study was to identify loss-of-function variants in the whole genome of athletes and establish an association between such variants with exercise-related traits.

Methods and Algorithms: Twenty Tatar sub-elite male wrestlers participated in the study. DNA was extracted from fasting blood, and whole-genome libraries were sequenced on the Illumina HiSeq 2500 platform. Raw reads were mapped to the human genome hg19 reference using BWA [1], and the variant calling was performed using Strelka2 [2] followed with variants filtering and annotation [3]. Whole-genome variants were validated using the microarray data (HumanOmniExpress BeadChip, 900K SNVs).

Athletes have been subjected to several functional tests. The anaerobic lower-limb peak power was assessed during the Wingate 30-sec test. The reaction time was evaluated using the visual computer "Traffic light" test. Specifically, the participants were asked to press the mouse button when the green signal appeared on the screen, and the reaction time was calculated as an average of 3 best attempts out of 5. The muscle mass of wrestlers was measured by the bioimpedance method using a Tanita MC 980 MA electronic scales after fasting for at least 8 hours. In addition, athletes were asked to evaluate their ability to run short distances as "good", "average" or "poor".

The list of known loss-of-function variants was extracted from the gnomAD v.2.1.1 database [4] (variants annotated as "stop gained", "frameshift", "splice donor", "splice acceptor") and filtered to exclude low confidence LoF and multi-allelic variants. To test the association between high-quality LoF variants in athletes and their functional characteristics, Spearman rank correlation coefficient was calculated using the additive, dominant, and recessive genetic model. P value less than 0.05 was considered as statistically significant.

Results: Over 580 loss-of-function variants (SNPs and indels) were detected in athletes' genomes. Among them, 263 polymorphisms were frameshift mutations, 170 variants caused a premature stop codon, other mutations occurred in acceptor (n = 105) and donor (n = 50) splicing site. Fifteen high-confidence polymorphisms were found to be

associated with at least one evaluated exercise-related parameter. Variants in genes *BCKDHA* (rs3217385), *RHBG* (rs2245623), *ATP13A5* (rs74437357), *SLC6A18* (rs7447815), *ANKDD1B* (rs34358), *OR2J1* (rs2394517), and *KIAA1755* (rs41282820) showed an association with lower-limb peak power. Muscle mass values were shown to have correlation with polymorphisms in genes *PRSS48* (rs77216366), *ZACN* (rs1043149), *FUT2* (rs601338), and *H2BW2* (rs2301384). Three SNPs were demonstrated to reduce reaction time, namely, *VWDE* (rs17165936), *APOBEC1* (rs34275479), and *STK31* (rs6945306). Finally, self-reported speed running performance was associated with variants in genes *ANKDD1B* (rs34358), *IDO2* (rs4503083), and *BCKDHA* (rs3217385).

Conclusion: In conclusion, we identified 15 LoF mutations associated with exercise-related traits in athletes. Further studies involving larger cohorts of athletes are warranted to replicate and extend these findings.

Acknowledgements: This work was part of Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030) and funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities (project No. 0671-2020-0058).

References

1. Li H., Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589-595.
2. Kim S. et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018;15(8):591-594.
3. Boulygina E.A. et al. Whole genome sequencing of elite athletes. *Biol Sport*. 2020;37(3):295-304.
4. Karczewski K.J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443.