Whole genome sequencing of elite athletes

AUTHORS: Eugenia A. Boulygina1, Oleg V. Borisov2,3, Elena V. Valeeva4,5, Ekaterina A. Semenova2,4, Elena S. Kostryukova2, Nikolay A. Kulemin7, Andrey K. Larin7, Roza M. Nabiullina2, Fanis A. Mavliev8, Azat M. Achtenov9, Oleg N. Andryushchenko10, Liulya B. Andryushchenko10, Piotr Zmijewski11, Edward V. Generozov11, Ildus I. Ahmetov2,3,10,12

1 “Omics Technologies” OpenLab, Kazan Federal University, Kazan, Russia
2 Department of Molecular Biology and Genetics, Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia
3 Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn, Bonn, Germany
4 Department of Biochemistry, Biotechnology and Pharmacology, Kazan Federal University, Kazan, Russia
5 Laboratory of Molecular Genetics, Central Research Laboratory, Kazan State Medical University, Kazan, Russia
6 Department of Biochemistry, Kazan State Medical University, Kazan, Russia
7 Sport Technology Research Center, Volga Region State Academy of Physical Culture, Sport and Tourism, Kazan, Russia
8 Department of Theory and Methodology of Combat Sports, Volga Region State Academy of Physical Culture, Sport and Tourism, Kazan, Russia
9 Department of Physical Education, Financial University under the Government of the Russian Federation, Moscow, Russia
10 Department of Physical Education, Plekhanov Russian University of Economics, Moscow, Russia
11 Institute of Sport - National Research Institute, Warsaw, Poland
12 Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, United Kingdom

ABSTRACT: 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 associated with RT, we identified four SNPs (rs7783359, rs10125715, TMEM229A rs7783359, LRRN3 rs80054135) associated with RT in wrestlers. The cumulative number of favourable alleles 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.

CITATION: Boulygina EA, Borisov OV, Valeeva EV et al. Whole genome sequencing of elite athletes. Biol Sport. 2020;37(3):295–304.

Received: 2020-05-16; Reviewed: 2020-05-26; Re-submitted: 2020-05-29; Accepted: 2020-06-01; Published: 2020-06-10.

INTRODUCTION

Sports genomics is a relatively new scientific discipline focusing on the organization and functioning of the genome of elite athletes. It postulates that genetic and epigenetic factors play a key role in athletic performance and related phenotypes such as power, strength, aerobic capacity, flexibility, height, muscle mass, coordination, and personality traits. Despite a relatively high heritability of athlete status and performance related phenotypes, the search for genetic variants contributing to predisposition to success in certain types of sport has been a challenging task. So far, 185 DNA polymorphisms associated with athlete status have been identified in the last 21 years [1–3].
Among common tools for the detection of performance-associated DNA polymorphisms researchers use case-control or genotype-phenotype studies based on a candidate gene design [4–7]. The limitation of this approach is that one cannot detect the polymorphic variant which lies within a non-coding (possibly, regulatory) genome region. Another approach is a genome-wide association study (GWAS) using micro-array analysis which proved to be extremely successful to uncover genetic association in sport-related phenotypes [8–11]. However, micro-array analysis covers only a limited number (up to 5 M) of DNA polymorphisms (< 0.2% of the genome). Although these polymorphisms are designed to evenly cover most of the genome regions, linkage disequilibrium differences in various populations restrict the generalizability of such an approach. To overcome this, the whole-genome based technique can be effectively used. Whole genome sequencing (WGS) refers to the construction of the complete nucleotide sequence of a genome (~3.2 billion base pairs in humans) and provides a powerful tool to obtain greater insight into the genetic variability that could produce a range of benefits for sport and exercise science.

Here we performed, for the first time, a low coverage whole-genome analysis in a group of athletes which was homogeneous in terms of ethnicity (Tatars), sex (males) and sport discipline (belt wrestling). A number of factors determine athletic performance in wrestling, among them both physiological (strength and endurance, muscle mass, dexterity, displacement speed, flexibility, coordination, balance) and psychological (reaction time, decision-making speed, ingenuity, patience) [12]. For example, most successful wrestlers show a significantly quicker reaction time during fights [13]. The heritability of reaction time has been shown to reach 60% [14]. At least 5 genetic markers (located within ACE, ACTN3, FTO, HIF1A and MCT1 genes; involved in metabolism, skeletal muscle structure and function) have previously been shown to be linked to wrestler status [5, 15–18]. However, genetic association analysis in wrestlers using DNA polymorphisms previously associated with reaction time has not been performed.

The aim of our study was, therefore, to characterize the whole genome sequence of wrestlers and apply this information to establish an association between DNA variants and reaction time.

**MATERIALS AND METHODS**

**Ethical approval**

The study was approved by the Ethics Committee of the Federal Research and Clinical Center of Physical-chemical Medicine of the Federal Medical and Biological Agency of Russia. Written informed consent was obtained from each participant. The study complied with the guidelines set out in the Declaration of Helsinki and ethical consent was obtained from each participant. The study complied with the recommendations, with pair-end 125-bp read length at an average sequencing depth of 9.9x (ranging from 2.6x to 16.8x). Raw reads were mapped to the human reference genome hg19 using BWA [20]. The low coverage whole genome variant calling was performed using Strelka v. 2 [21]. Hard filtering was applied to the detected raw single nucleotide variants with parameters as follows: MQ < 40, LowDepth > 3, HighSNVS < 10. Variants were annotated using Annovar [22] equipped with additional databases (ClinVar, COSMIC, dbSNP, ESP6500, ExAC). The whole genome variants were validated by the microarray technology with HumanOmniExpress Bead-Chips (Illumina Inc, USA) for genotyping of ~900,000 SNPs.

Sample collection, DNA sequencing, SNV calling and SNP genotyping

Fasting venous blood samples (a total of 9 ml of blood) of wrestlers were collected in the morning in tubes containing K2-EDTA and stored at −20°C until DNA extraction. DNA extraction was performed using a Wizard Genomic DNA Purification Kit according to the manufacturer’s instructions (Promega, USA). DNA libraries were sequenced by Illumina HiSeq 2500 platform using the HiSeq SBS Kit v4 (Illumina, San Diego, USA) according to the manufacturer’s recommendations, with pair-end 125-bp read length at an average sequencing depth of 9.9x (ranging from 2.6x to 16.8x). Raw reads were mapped to the human reference genome hg19 using BWA [20]. The low coverage whole genome variant calling was performed using Strelka v. 2 [21]. Hard filtering was applied to the detected raw single nucleotide variants with parameters as follows: MQ < 40, LowDepth > 3, HighSNVS < 10. Variants were annotated using Annovar [22] equipped with additional databases (ClinVar, COSMIC, dbSNP, ESP6500, ExAC). The whole genome variants were validated by the microarray technology with HumanOmniExpress Bead-Chips (Illumina Inc, USA) for genotyping of ~900,000 SNPs.

SNP genotyping (micro-array analysis) of physically active participants, athletes and controls was performed with DNA samples obtained from leukocytes (venous blood). Four ml of venous blood were collected in tubes containing EDTA (Vacuette EDTA tubes, Greiner Bio-One, Austria). Blood samples were transported to the laboratory at 4°C and DNA was extracted on the same day. DNA extraction and purification were performed using a commercial kit according to the manufacturer’s instructions (Technoclon, Russia) and included chemical lysis, selective DNA binding on silica spin columns and ethanol washing. Extracted DNA quality was assessed by agarose gel electrophoresis at this step. HumanOmni1-Quad Bead-Chips (Illumina Inc, USA) were used for genotyping of 1,140,419 SNPs.

**Participants**

Twenty professional male belt (kurash) wrestlers (age 20 ± 4.4 years; height 173.0 ± 10.0 cm, weight 73.6 ± 10.6 kg) volunteered for the WGS study. The athletes were all Caucasian Tatars from the Republic of Tatarstan (Russian Federation). Exomes (protein-coding regions of genes in a genome) of wrestlers were compared with exomes of ethnicity-matched controls (Tatars, n = 19) from a previous study [19]. The second cohort consisted of 43 physically active participants (27 males, age 35.8 ± 7.9 years, height 178.4 ± 6.2 cm, weight 77.1 ± 11.0 kg; 16 females, age 29.4 ± 8.7 years, height 168.8 ± 6.4 cm, weight 57.3 ± 5.2 kg) and was used for validation of findings from the association study (reaction time). The case-control study involved 283 athletes (110 females, 173 males; age 24.1 ± 3.6 years) from the following sporting disciplines: boxing (n = 101), wrestling (n = 82), karate (n = 21), taekwondo (n = 24), volleyball (n = 45), table tennis (n = 10). All athletes were international-level competitors who represented Russia in international competitions (107 elite (prize winners) and 176 sub-elite) and have been tested negative for doping substances. The control group (n = 189, 38 females, 151 males; age 45 ± 4.3 years) included unrelated citizens, without any competitive sport experience.
Athletes’ genome

in 283 athletes and 189 controls, while HumanOmniExpress BeadChips (Illumina Inc, USA) were used for genotyping of ~900,000 SNPs in 43 physically active participants. Reaction time related DNA variants (n = 1884; including leading and tag SNPs) for validation in wrestlers and physically active participants were selected from published studies [23, 24].

Reaction time measurement

Visual reaction time was evaluated using the computer test ‘Traffic light’. Laboratory-based testing was carried out under same conditions for participants (i.e. in morning, in the resting state, using the same computer, under supervision of the same test administrator). Subjects sat in front of a table with the palm of the dominant hand supported and their index finger on a computer mouse. The participants were consistently presented with light signals in the centre of the monitor screen, and were asked to press the button when the green signal appeared. The duration of the intervals between the red and green signals ranged from 0.5 to 5 s. The first 5 signals were trial and were not recorded. The best three attempts from the following 5 signals were recorded and the average reaction time was calculated. All attempts were observed by the test administrator.

Statistical analysis

Statistical analyses were conducted using PLINK v1.90, R (version 3.4.3), and GraphPad InStat (GraphPad Software, Inc., USA) software. The chi-square test ($\chi^2$) was used to test for the presence of the Hardy-Weinberg equilibrium in the genotype distribution, to compare the proportions of subjects with a high number of reaction time improving alleles or allelic frequencies between groups. To evaluate the associations between polygenic profiles and reaction time, the Spearman rank correlation coefficient was calculated. P values < 0.05 were considered statistically significant.

RESULTS

The metrics of genomic variants

As a genomic data quality metric, we used the transition/transversion (Ts/Tv) ratio, which was detected to be consistent with previous studies ($\approx 2$) [25]. For 20 wrestlers’ genomes, using a low coverage protocol we detected over 11.5 million raw genomic variants in total. Taking into account only sufficiently covered genomes ($\geq 12x$), for the 12 most deeply covered samples, average numbers of SNVs and indels were 3.8 million and 0.64 million per sample, correspondingly. About 11 million raw variants passed the quality filters (3.6 mil-

### TABLE 1. Basic statistics of raw WGS data and genomic variants in wrestlers (n = 20)

| Sample # | Raw reads number | Mapped reads number | Mapped reads, % | Mean coverage | Ts/Tv ratio | Raw SNPs number | Raw indels number | Filtered SNPs number | Filtered indels number |
|----------|------------------|---------------------|-----------------|---------------|-------------|----------------|------------------|---------------------|-----------------------|
| 1        | 135419040        | 134973712           | 99.67           | 2.65          | 2.01        | 1170490        | 94591            | 1132716             | 92585                 |
| 2        | 152651625        | 152180976           | 99.69           | 3.00          | 2           | 1338431        | 115651           | 1296133             | 113359                |
| 3        | 135166473        | 134674889           | 99.64           | 2.64          | 1.98        | 1156025        | 100851           | 1119445             | 98894                 |
| 4        | 138380470        | 137897091           | 99.65           | 2.71          | 2.01        | 1187357        | 97416            | 1149187             | 95394                 |
| 5        | 138247447        | 137870773           | 99.73           | 2.72          | 1.98        | 1197711        | 105756           | 1160354             | 103601                |
| 6        | 170747873        | 170301203           | 99.74           | 3.36          | 1.98        | 1504990        | 139290           | 1455811             | 136558                |
| 7        | 156619056        | 156211771           | 99.74           | 3.08          | 1.98        | 1389524        | 126783           | 1345519             | 123477                |
| 8        | 167468957        | 167046801           | 99.75           | 3.29          | 1.99        | 1468924        | 133239           | 1422379             | 124277                |
| 9        | 349634959        | 349479692           | 99.96           | 14.13         | 1.98        | 3845412        | 638443           | 3646325             | 619174                |
| 10       | 354831314        | 354646812           | 99.95           | 14.27         | 1.98        | 3864805        | 644140           | 3662651             | 624830                |
| 11       | 419085260        | 418925257           | 99.97           | 16.81         | 1.98        | 3381380        | 559592           | 3210289             | 539752                |
| 12       | 370391660        | 370236765           | 99.96           | 14.93         | 1.99        | 3899130        | 644534           | 3703033             | 624312                |
| 13       | 305447291        | 305279572           | 99.95           | 12.27         | 1.99        | 3785574        | 617903           | 3587268             | 599639                |
| 14       | 318191054        | 318037831           | 99.95           | 12.85         | 1.98        | 3789480        | 628848           | 3586741             | 609087                |
| 15       | 362707535        | 362532795           | 99.95           | 14.62         | 1.98        | 3879120        | 651822           | 3666744             | 631928                |
| 16       | 384884409        | 384695366           | 99.95           | 15.52         | 1.98        | 3887156        | 660029           | 3676958             | 639700                |
| 17       | 330816726        | 330647548           | 99.95           | 13.31         | 1.98        | 3809749        | 634330           | 3603943             | 615521                |
| 18       | 366875825        | 366706680           | 99.95           | 14.78         | 1.98        | 3852608        | 645399           | 3649519             | 625418                |
| 19       | 366788900        | 366617355           | 99.95           | 14.78         | 1.98        | 3893521        | 654992           | 3691772             | 635218                |
| 20       | 412081693        | 411940287           | 99.97           | 16.62         | 1.98        | 3933194        | 659427           | 3727201             | 636988                |
Reaction times (RT) did not differ between 20 wrestlers and 43 physically active subjects (0.286 (0.015) s vs 0.274 (0.059) s; \( P = 0.372 \)). RT between males and females in the group of physically active subjects was not significantly different (\( P = 0.891 \)); therefore in the association analysis we used the combined group.

In the discovery phase, out of 1884 known genome-wide significant SNPs (including leading and tag SNPs) related to RT, 24 SNPs (four leading and 20 tag SNPs) were associated with RT in wrestlers. Of those, four alleles (\( \text{KIF27} \text{rs10125715 A} \), \( \text{APC} \text{rs518013 A} \), \( \text{TMEM229A} \text{rs7783359 T} \), \( \text{LRRN3} \text{rs80054135 T} \)) were found to be independently associated with the best RT in wrestlers (Table 2).

To assess the combined impact of all 4 gene polymorphisms, we classified wrestlers and physically active subjects according to the number of ‘short reaction time’ alleles they possessed (e.g. carriers of genotype \( \text{KIF27} \text{rs10125715 A} \), \( \text{APC} \text{rs518013 A} \), \( \text{TMEM229A} \text{rs7783359 T} \), \( \text{LRRN3} \text{rs80054135 T} \) had 0 ‘short reaction time’ alleles). The cumulative number of favourable (i.e. leading to a short reaction time) alleles was significantly cor-

| Closest gene | Favourable allele | Reaction time, s | Genotype 1 | Genotype 2 | Genotype 3 | \( r \) | \( P \) |
|--------------|------------------|-----------------|------------|------------|------------|-------|-------|
| \( \text{APC} \text{rs518013 A} \) | GG (n = 4) | 0.287 ± 0.023 | GA (n = 10) | 0.286 ± 0.017 | AA (n = 6) | -0.52 | 0.028* |
| \( \text{KIF27} \text{rs10125715 A} \) | TT (n = 1) | 0.301 | AT (n = 12) | 0.289 ± 0.016 | AA (n = 7) | -0.49 | 0.034* |
| \( \text{TMEM229A} \text{rs7783359 T} \) | – | 0.292 ± 0.016 | AT (n = 10) | 0.292 ± 0.016 | TT (n = 10) | -0.44 | 0.048* |
| \( \text{LRRN3} \text{rs80054135 T} \) | AA (n = 17) | 0.289 ± 0.013 | AT (n = 2) | 0.277 ± 0.004 | TT (n = 1) | -0.56 | 0.012* |

* \( P < 0.05 \), significant correlation. Values are mean ± SD.

| Groups | \( n \) | Frequencies of favourable alleles, % |
|--------|-------|-------------------------------------|
| \( \text{APC} \text{rs518013 A} \) | \( \text{KIF27} \text{rs10125715 A} \) | \( \text{TMEM229A} \text{rs7783359 T} \) | \( \text{LRRN3} \text{rs80054135 T} \) |
| Elite athletes | 107 | 71.5 | 65.4 | 7.5** |
| Sub-elite athletes | 176 | 70.2 | 63.6 | 2.8 |
| Russian controls | 189 | 69.6 | 66.7 | 2.9 |

* \( P = 0.033 \), statistically significant differences between elite and sub-elite athletes. ** \( P = 0.009 \), statistically significant differences between elite and sub-elite athletes or controls.

Genetic association analysis

Reaction times (RT) did not differ between 20 wrestlers and 43 physically active subjects (0.286 (0.015) s vs 0.274 (0.059) s; \( P = 0.372 \)). RT between males and females in the group of physically active subjects was not significantly different (\( P = 0.891 \)); therefore in the association analysis we used the combined group. In the discovery phase, out of 1884 known genome-wide significant SNPs (including leading and tag SNPs) related to RT, 24 SNPs (four leading and 20 tag SNPs) were associated with RT in wrestlers. Of those, four alleles (\( \text{KIF27} \text{rs10125715 A} \), \( \text{APC} \text{rs518013 A} \), \( \text{TMEM229A} \text{rs7783359 T} \), \( \text{LRRN3} \text{rs80054135 T} \)) were found to be independently associated with the best RT in wrestlers (Table 2).

To assess the combined impact of all 4 gene polymorphisms, we classified wrestlers and physically active subjects according to the number of ‘short reaction time’ alleles they possessed (e.g. carriers of genotype \( \text{KIF27} \text{rs10125715 TT} \), \( \text{APC} \text{rs518013 GG} \), \( \text{TMEM229A} \text{rs7783359 AA} \), \( \text{LRRN3} \text{rs80054135 AA} \) had 0 ‘short reaction time’ alleles, and subjects with \( \text{KIF27} \text{rs10125715 AA} \), \( \text{APC} \text{rs518013 AA} \), \( \text{TMEM229A} \text{rs7783359 TT} \), \( \text{LRRN3} \text{rs80054135 TT} \) genotype had 8 ‘short reaction time’ alleles). The cumulative number of favourable (i.e. leading to a short reaction time) alleles was significantly cor-

10 million SNVs and 0.62 million indels on average per mostly covered samples). 47.8% of variants were annotated as synonymous SNV and 46.2% as nonsynonymous SNV; about 1.2% were frameshift and non-frameshift indels. The average numbers of stop-gain and stop-loss mutations were 128 and 14, respectively, for deeply covered samples; the maximum number of these nonsense mutations was 148 SNVs per individual. As expected, the vast majority of variants localized in intergenic and intronic regions (≈56% and ≈34%, respectively). About 2.7% of variation lay within exons, upstream and downstream, and in 3' and 5' UTRs. Basic statistics of raw WGS data and genomic variants in wrestlers are given in Table 1 and Supplementary Tables 1–7.

The fraction of variants that were predicted to be ‘benign’ and ‘likely benign’ was the highest (about 91.6%), followed by the fraction that had ‘uncertain significance’ (2.2%). Variants annotated as ‘pathogenic’ and ‘likely pathogenic’ represented 0.66% of total variation. We next compared the number of homozygous stop-gain mutations between wrestlers and 19 Tatar controls from our previous study [19]. Athletes’ genomes on average contained 18.9 nonsense SNPs in a homozygous state per sample, while non-athletes’ exomes contained 18 nonsense SNPs (\( P > 0.05 \)).
related with RT in wrestlers ($r = 0.73$, $P = 0.0003$). This finding was also validated in the independent cohort of physically active subjects ($r = 0.33$, $P = 0.029$).

Next, we compared allelic frequencies of four SNPs between elite athletes ($n = 107$) involved in sports with RT as an essential component of performance (combat sports, table tennis and volleyball), sub-elite athletes ($n = 176$) and 189 controls (Table 3 and Supplementary Table 8). The genotypes distributions of four SNPs met the Hardy-Weinberg equilibrium expectations in athletes and controls. We found that the frequencies of the $APC$ rs518013 A (53.3 vs 44.0%, $P = 0.033$) and $LRRN3$ rs80054135 T (7.5 vs 2.8%, $P = 0.009$) alleles were significantly higher in elite compared to non-elite athletes. The $LRRN3$ rs80054135 T allele was also over-represented in elite athletes (7.5%) in comparison with controls (2.9%, $P = 0.009$). Using the 1000 Genomes database (http://www.ensembl.org), we identified that East Asian populations for most SNPs have the highest frequency of favourable alleles compared to other populations (Supplementary Table 9).

**DISCUSSION**

To our knowledge, this is the first paper on whole genome sequencing of athletes. We found that the mutational load per Tatar athlete (the number of stop-loss and stop-gained mutations), SNV localization and clinical relevance statistics were comparable with those in Eurasian populations [26–28]. We also found that athletes’ genomes on average contained 18.9 nonsense SNPs in a homozygous state per sample, while Tatar non-athletes’ exomes contained almost the same (18) number of nonsense SNPs. These observations suggest that the obtained sequencing data have an adequate quality and may serve as a good starting point for further research in sports genomics. Tatars are one of the major Turkic-speaking groups in the Volga-Ural region of the Russian Federation. It was believed that, due to the geographic position of the region and the complex ethnic history of the population, Tatars have an extremely high genetic diversity in which the Asian (Mongolian) component had a significant contribution. The latest studies showed that the trace of East Asian or Central Siberian ancestry in the genomes of Volga Tatars is less than expected (approximately 5%) [29], but nevertheless, it still allows us to evaluate the genome data quality using European and Asian genomic data as a reference.

We also applied genomic information to establish an association between DNA variants and reaction time. The high quality of WGS was confirmed by micro-array analysis of 1884 SNPs previously reported to be associated with RT in the UK Biobank cohorts [23, 24]. As a result, we confirmed the association between four independent SNPs ($APC$ rs518013, $KIF27$ rs10125715, $TMEM229A$ rs7783359, $LRRN3$ rs80054135) and RT in two cohorts (wrestlers and physically active subjects). We also found that the frequencies of $APC$ rs518013 A and $LRRN3$ rs80054135 T alleles were significantly higher in elite athletes involved in sports with RT as an essential component of performance (combat sports, table tennis and volleyball) compared to non-elite athletes and/or controls, indicating that these DNA polymorphisms may play a role in the natural selection process of athletes. Interestingly, for most of the SNPs the highest frequency of favourable alleles compared to other populations was found in East Asians – populations with a long history of cultivation of martial arts and dominance in judo, karate, taekwondo and table tennis. Our approach thus provided for the first time sufficient power of WGS to detect a wide range of candidate alleles that may lead to athletic success.

According to the GTEx Portal [30], three of those SNPs are functional and may influence expression of several genes in the brain and other tissues ($APC$ rs518013 influences expression of the $SRP19$ gene; $KIF27$ rs10125715 influences expression of $GKAP1$ and $RM1$ genes; $TMEM229A$ rs7783359 influences expression of the $RP5–921G16.1$ gene). The $APC$ (adenomatous polyposis coli protein) gene encodes a tumour suppressor protein that acts as an antagonist of the Wnt signalling pathway and is involved in other processes including cell migration and adhesion, transcriptional activation, and apoptosis. Interestingly, the hypermethylation of the $APC$ gene was reported to be inversely associated with physical activity [31]. The $KIF27$ (kinesin family member 27) gene encodes a protein involved in ATPase activity and microtubule motor activity [32]. The $TMEM229A$ (transmembrane protein 229A) gene encodes a protein involved in DNA-binding transcription factor activity and developmental processes [33]. The $LRRN3$ (leucine rich neuronal 3) gene encodes a protein which plays an important role in cerebellum postnatal development [34].

Among the limitations of the current study are the sample size of the wrestler cohort ($n = 20$) and the low overall sequencing depth. Despite the fact that the total number of SNVs per genome did not reach the level that was observed before [35, 36] (probably due to insufficient sequencing depth), still, such low-coverage sequencing was shown to allow genotyping variants with confidence [37], and this was also confirmed by micro-array analysis in our study. Other efforts to sequence hundreds of genomes of elite athletes are presently underway [38].

**CONCLUSIONS**

In conclusion, we present the first WGS study of athletes showing that WGS can be applied in sports genomics. By replicating previous findings from non-athletic populations, we demonstrate that the $APC$ rs518013 A and $LRRN3$ rs80054135 T alleles are associated with the best reaction time in wrestlers and physically active subjects and over-represented in elite athletes involved in sports with reaction time as an essential component of performance.

**Conflict of interest**

The authors declare no conflict of interest.
REFERENCES

1. Maciejewska-Skrendo A, Sawczuk M, Cieszczyk 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-allelic of the FTO Gene n9939609 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. J Strength Cond Res. 2019; 33(3):691–700.

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, Wolforth 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, Youssi 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/fgen.2020.00595.

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

13. Gierczuk D, Bujak Z, Ciesliński 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. Kukuchi J, Rogers H, Swindar 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. Kukuchi 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. Kukuchi 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. Kukuchi N, Fuku N, Matsumoto R, et al. The Association Between MCT1 T1470A Polymorphism and Power-Oriented Athletic Performance. Int J Sports Med. 2013;34(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 Burrows–Wheeler transform. Bioinformatics. 2010;26:589–95.

21. Telenti A, Pierce LC, Biggs WH, et al. Rare variant discovery by deep sequencing in South Asians captures extremely low-coverage whole genome sequencing data. Nucleic Acids Res. 2016;113(42):11901–11906.

22. Strelka: fast and accurate calling of germline and somatic variants. Nat Methods. 2010;7:540–542.

23. Kim S, Scheffler K, Halpern AL, et al. Strelka2: fast and accurate calling of permanent and deciduous teeth by network and ancestry. Bioinformatics. 2016;15(8):591–594.

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

25. Davies G, Marioni RE, Liewald DC. Extremely low-coverage whole genome sequencing in South Asians captures population genomics information. BMC Genomics. 2017;18(1):396.

26. Nagasak M, Yasuda J, Katsuoka F, et al. Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. Nat Commun. 2015;6:6081.

27. Sidore C, Busonero F, Masciaio 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, Stepانov 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 non-malignant 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, Qi L, et al. Role of LRRN3 in the cerebellum postnatal development in rats. Zhongguo 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. Huvario.me: a web server resource of whole genome next-generation sequencing allele 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. Taniwasaka K, Wang G, Seto J et al. Sport and exercise genomics: the FIMS consensus statement update. Br J Sports Med. 2020 Mar 22. doi: 10.1136/bjsports-2019–101532.
### SUPPLEMENTARY TABLE 1. Variants localization

| Sample # | Exonic | Splicing | ncRNA | UTR5 | UTR3 | Intrinsic | Upstream | Downstream | Intergenic |
|----------|--------|----------|-------|------|------|-----------|----------|------------|------------|
| 1        | 7322   | 46       | 82943 | 1512 | 9541 | 423608    | 6817     | 8045       | 685167     |
| 2        | 7956   | 61       | 93985 | 1575 | 10745| 485026    | 7518     | 8950       | 793325     |
| 3        | 6251   | 44       | 81324 | 1163 | 9318 | 416904    | 6021     | 7510       | 689536     |
| 4        | 7220   | 51       | 83069 | 1487 | 9786 | 430918    | 7000     | 8218       | 696527     |
| 5        | 6551   | 47       | 82924 | 1265 | 9526 | 433249    | 6493     | 7910       | 715721     |
| 6        | 8466   | 63       | 105903| 1723 | 11928| 546717    | 8271     | 9833       | 899139     |
| 7        | 7674   | 49       | 97758 | 1471 | 10934| 503454    | 7439     | 9229       | 831484     |
| 8        | 8388   | 57       | 102768| 1619 | 11666| 532930    | 7998     | 9781       | 877571     |
| 9        | 23368  | 216      | 284400| 6596 | 31451| 1453515   | 26561    | 28486      | 2408633    |
| 10       | 23550  | 197      | 284216| 6571 | 31695| 1465213   | 26999    | 28594      | 2418169    |
| 11       | 19231  | 159      | 249751| 5598 | 27122| 1298338   | 22436    | 24082      | 2103214    |
| 12       | 24072  | 192      | 285008| 6671 | 32160| 1482433   | 27346    | 29089      | 2438070    |
| 13       | 22444  | 191      | 281279| 6460 | 30682| 1422247   | 25857    | 27787      | 2368005    |
| 14       | 22666  | 186      | 279610| 6364 | 31345| 1435265   | 26285    | 27783      | 2364281    |
| 15       | 23366  | 201      | 287954| 6652 | 31917| 1473141   | 26928    | 28946      | 2417142    |
| 16       | 23674  | 205      | 290871| 6556 | 31944| 1472548   | 27057    | 28951      | 2432393    |
| 17       | 22957  | 205      | 280664| 6477 | 31228| 1440480   | 26498    | 28157      | 2380668    |
| 18       | 23328  | 201      | 282964| 6592 | 31685| 1460742   | 26779    | 28494      | 2411895    |
| 19       | 23345  | 215      | 284624| 6676 | 32164| 1473218   | 27066    | 29128      | 2448248    |
| 20       | 24045  | 206      | 288257| 6851 | 32382| 1493547   | 27650    | 29073      | 2459687    |

### SUPPLEMENTARY TABLE 2. Variants effect

| Sample # | Frameshift insertion | Frameshift deletion | Stopgain | Stoploss | Non-fameshift insertion | Non-fameshift deletion | Non-synonymous SNV | Synonymous SNV | Unknown |
|----------|----------------------|---------------------|----------|----------|-------------------------|------------------------|--------------------|---------------|---------|
| 1        | 30                   | 32                  | 29       | 7        | 19                      | 16                     | 3460               | 3542          | 195     |
| 2        | 35                   | 35                  | 43       | 12       | 20                      | 31                     | 3724               | 3857          | 214     |
| 3        | 22                   | 23                  | 35       | 5        | 24                      | 18                     | 2943               | 3009          | 176     |
| 4        | 30                   | 28                  | 35       | 3        | 18                      | 23                     | 3394               | 3488          | 221     |
| 5        | 17                   | 24                  | 21       | 6        | 16                      | 24                     | 3020               | 3228          | 208     |
| 6        | 38                   | 37                  | 38       | 7        | 30                      | 30                     | 4008               | 4049          | 240     |
| 7        | 21                   | 29                  | 33       | 4        | 24                      | 24                     | 3673               | 3665          | 213     |
| 8        | 26                   | 41                  | 48       | 4        | 32                      | 25                     | 3914               | 4067          | 242     |
| 9        | 127                  | 159                 | 133      | 15       | 169                     | 176                    | 10780              | 11137         | 730     |
| 10       | 128                  | 162                 | 128      | 14       | 151                     | 168                    | 10872              | 11248         | 725     |
| 11       | 96                   | 150                 | 101      | 12       | 112                     | 140                    | 8906               | 9112          | 647     |
| 12       | 121                  | 155                 | 129      | 16       | 163                     | 185                    | 11110              | 11519         | 730     |
| 13       | 127                  | 164                 | 108      | 19       | 143                     | 164                    | 10350              | 10766         | 662     |
| 14       | 116                  | 153                 | 118      | 16       | 138                     | 171                    | 10448              | 10900         | 663     |
| 15       | 118                  | 161                 | 126      | 12       | 145                     | 167                    | 10856              | 11135         | 701     |
| 16       | 127                  | 150                 | 118      | 14       | 167                     | 171                    | 10957              | 11342         | 685     |
| 17       | 132                  | 152                 | 116      | 15       | 132                     | 164                    | 10517              | 11054         | 726     |
| 18       | 114                  | 143                 | 127      | 16       | 140                     | 182                    | 10780              | 11150         | 732     |
| 19       | 114                  | 144                 | 117      | 16       | 159                     | 165                    | 10803              | 11116         | 766     |
| 20       | 134                  | 196                 | 129      | 13       | 145                     | 193                    | 11072              | 11493         | 726     |
SUPPLEMENTARY TABLE 3. Variants clinical relevance

| Sample # | Benign | Likely benign | Pathogenic | Likely pathogenic | Uncertain significance | Drug response | Other | Not provided |
|----------|--------|---------------|------------|------------------|------------------------|---------------|-------|--------------|
| 1        | 2339   | 506           | 17         | 0                | 37                     | 28            | 92    | 26           |
| 2        | 2463   | 560           | 17         | 3                | 47                     | 26            | 117   | 40           |
| 3        | 1998   | 490           | 13         | 2                | 48                     | 19            | 77    | 28           |
| 4        | 2254   | 490           | 15         | 2                | 46                     | 18            | 94    | 34           |
| 5        | 2179   | 466           | 14         | 2                | 37                     | 24            | 104   | 33           |
| 6        | 2671   | 633           | 26         | 4                | 61                     | 34            | 148   | 44           |
| 7        | 2391   | 541           | 16         | 4                | 52                     | 22            | 97    | 37           |
| 8        | 2591   | 603           | 18         | 5                | 58                     | 28            | 119   | 42           |
| 9        | 7003   | 1873          | 57         | 6                | 217                    | 73            | 339   | 118          |
| 10       | 6453   | 1736          | 48         | 10               | 212                    | 71            | 301   | 125          |
| 11       | 5834   | 1566          | 41         | 6                | 194                    | 53            | 311   | 67           |
| 12       | 6960   | 1834          | 55         | 16               | 217                    | 69            | 382   | 134          |
| 13       | 6541   | 1664          | 50         | 8                | 240                    | 61            | 314   | 125          |
| 14       | 6792   | 1738          | 44         | 10               | 215                    | 62            | 350   | 120          |
| 15       | 6767   | 1804          | 54         | 13               | 219                    | 63            | 373   | 125          |
| 16       | 6768   | 1760          | 55         | 7                | 255                    | 62            | 256   | 119          |
| 17       | 6857   | 1827          | 54         | 14               | 214                    | 63            | 318   | 116          |
| 18       | 6786   | 1742          | 45         | 15               | 214                    | 59            | 344   | 133          |
| 19       | 6915   | 1812          | 49         | 10               | 242                    | 66            | 344   | 150          |
| 20       | 7030   | 1838          | 60         | 11               | 236                    | 72            | 373   | 135          |

SUPPLEMENTARY TABLE 4. The average of basic statistics

| Statistics               | 20 genomes | 12 most covered genomes |
|--------------------------|------------|-------------------------|
| Raw reads number         | 276821878.4| 361811385.5             |
| Mapped reads number      | 276547024.8| 361648606.6             |
| Mapped reads,%           | 99.85      | 99.95                   |
| Mean coverage            | 9.9        | 14.6                    |
| Ts/Tv ratio              | 2.0        | 2.0                     |
| Raw SNPs number          | 2811729.1  | 3818427.4               |
| Raw indels number        | 427651.8   | 636621.6                |
| Filtered SNPs number     | 2674789.4  | 3617853.7               |
| Filtered indels number   | 414847.5   | 616797.3                |

SUPPLEMENTARY TABLE 5. Variants with localization (N = 61763469)

| Localization   | 20 genomes average | 12 most covered genomes average | % based on 20 genomes |
|----------------|---------------------|---------------------------------|-----------------------|
| Intergenic     | 1741943.8           | 2387533.8                       | 56.4                  |
| Intronic       | 1057174.7           | 1447557.3                       | 34.2                  |
| ncRNA          | 205513.6            | 281633.2                        | 6.7                   |
| UTR3           | 22961.0             | 31314.6                         | 0.7                   |
| Downstream     | 20402.3             | 28214.2                         | 0.7                   |
| Upstream       | 18751.0             | 26455.2                         | 0.6                   |
| Exonic         | 16793.7             | 23003.8                         | 0.5                   |
| UTR5           | 4494.0              | 6505.3                          | 0.1                   |
| Splicing       | 139.6               | 197.8                           | 0.005                 |
### SUPPLEMENTARY TABLE 6. Variants with predicted effect (n = 336619)

| Type of mutation             | 20 genomes average | 12 most covered genomes average | % based on 20 genomes |
|------------------------------|--------------------|---------------------------------|-----------------------|
| Synonymous SNV               | 8043.9             | 10997.7                         | 47.8                  |
| Non-synonymous SNV           | 7779.4             | 10620.9                         | 46.2                  |
| Unknown                      | 510.1              | 707.8                           | 3.0                   |
| Non-frameshift deletion      | 111.9              | 170.5                           | 0.7                   |
| Frameshift deletion          | 106.9              | 157.4                           | 0.6                   |
| Non-frameshift insertion     | 97.4               | 147.0                           | 0.6                   |
| Stopgain                     | 86.6               | 120.8                           | 0.5                   |
| Frameshift insertion         | 83.7               | 121.2                           | 0.5                   |
| Stoploss                     | 11.3               | 14.8                            | 0.1                   |

### SUPPLEMENTARY TABLE 7. Variants with clinical significance (n = 136609)

| Clinical effect              | 20 genomes average | 12 most covered genomes average | % based on 20 genomes |
|------------------------------|--------------------|---------------------------------|-----------------------|
| Benign                       | 4979.6             | 6725.5                          | 72.9                  |
| Likely benign                | 1274.2             | 1766.2                          | 18.7                  |
| Other                        | 242.7              | 33.8                            | 3.6                   |
| Uncertain significance       | 153.1              | 222.9                           | 2.2                   |
| Not provided                 | 87.6               | 122.3                           | 1.3                   |
| Drug response                | 48.7               | 64.5                            | 0.7                   |
| Pathogenic                   | 37.4               | 51.0                            | 0.5                   |
| Likely pathogenic            | 7.4                | 10.5                            | 0.1                   |
**SUPPLEMENTARY TABLE 8.** Frequencies of the favourable alleles in athletes and controls

| Groups                  | n    | APC rs518013 A % | KIF27 rs10125715 A % | TMEM229A rs7783359 T % | LRRN3 rs80054135 T % | P value Elite vs non-elite | P value Elite vs controls | P value Elite vs non-elite | P value Elite vs controls | P value Elite vs non-elite | P value Elite vs controls | P value Elite vs non-elite | P value Elite vs controls | P value Elite vs non-elite | P value Elite vs controls | P value Elite vs non-elite | P value Elite vs controls |
|-------------------------|------|-----------------|----------------------|------------------------|-----------------------|-------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Elite boxers            | 41   | 54.9            | 0.107                | 0.172                  | 69.5                  | -                        | 0.958                      | 0.991                      | 70.7                      | 0.184                      | 0.477                      | 5.0                      | 0.346                     | 0.340                     | -                        | -                        | -                        |
| Sub-elite boxers        | 60   | 43.3            | -                    | -                      | 69.2                  | -                        | -                         | -                         | -                         | -                         | 2.5                      | -                        | -                        | -                        | -                        | -                        | -                        |
| Elite wrestlers         | 34   | 55.9            | 0.205                | 0.157                  | 69.1                  | -                        | 0.813                      | 0.939                      | 57.6                      | 0.718                      | 0.152                     | 5.9                      | 0.202                     | 0.211                     | -                        | -                        | -                        |
| Sub-elite wrestlers     | 48   | 45.8            | -                    | -                      | 70.8                  | -                        | -                         | -                         | -                         | 60.4                      | -                        | 2.1                      | -                        | -                        | -                        | -                        | -                        |
| Elite karate athletes   | 5    | 40.0            | 0.834                | 0.681                  | 80.0                  | 0.899                    | 0.478                      | 60.0                      | 0.359                     | 0.659                     | 20.0                     | 0.071                    | 0.0026*                   | -                        | -                        | -                        |
| Sub-elite karate athletes| 16   | 43.8            | -                    | -                      | 78.1                  | -                        | -                         | 75.0                      | -                         | -                         | 3.1                      | -                        | -                        | -                        | -                        | -                        |
| Elite taekwondo athletes| 5    | 30.0            | 0.400                | 0.299                  | 50.0                  | 0.278                    | 0.186                      | 70.0                      | 0.816                     | 0.825                     | 30.0                     | 0.0053*                  | 0.0039*                   | -                        | -                        | -                        |
| Sub-elite taekwondo athletes| 19   | 44.7            | -                    | -                      | 68.4                  | -                        | -                         | 73.7                      | -                         | -                         | 2.6                      | -                        | -                        | -                        | -                        | -                        |
| Elite volleyball players| 17   | 50.0            | 0.667                | 0.723                  | 82.4                  | 0.315                    | 0.117                      | 64.7                      | 0.478                     | 0.816                     | 6.3                      | 0.862                    | 0.301                     | -                        | -                        | -                        |
| Sub-elite volleyball players| 28   | 44.6            | -                    | -                      | 71.4                  | -                        | -                         | 57.1                      | -                         | -                         | 5.4                      | -                        | -                        | -                        | -                        | -                        |
| Elite table tennis players| 5    | 70.0            | 0.074                | 0.143                  | 80.0                  | 0.159                    | 0.478                      | 90.0                      | 0.531                     | 0.121                     | 10.0                     | 0.305                    | 0.201                     | -                        | -                        | -                        |
| Sub-elite table tennis players| 5    | 30.0            | -                    | -                      | 50.0                  | -                        | 80.0                      | -                         | -                         | 0                         | -                        | -                        | -                        | -                        | -                        | -                        |
| Elite athletes          | 107  | 53.3            | 0.033*               | 0.117                  | 71.5                  | 0.737                    | 0.624                      | 65.4                      | 0.564                     | 0.877                     | 7.5                      | 0.009*                   | 0.009*                    | -                        | -                        | -                        |
| Sub-elite athletes      | 176  | 44.0            | -                    | -                      | 70.2                  | -                        | -                         | 63.6                      | -                         | -                         | 2.8                      | -                        | -                        | -                        | -                        | -                        |
| Russian controls        | 189  | 46.6            | -                    | -                      | 69.6                  | -                        | 66.7                      | -                         | -                         | 2.9                      | -                        | -                        | -                        | -                        | -                        | -                        |

*P< 0.05, statistically significant differences

**SUPPLEMENTARY TABLE 9.** Frequencies of the favourable alleles in different populations

| Groups                  | n    | APC rs518013 A % | KIF27 rs10125715 A % | TMEM229A rs7783359 T % | LRRN3 rs80054135 T |
|-------------------------|------|-----------------|----------------------|------------------------|-------------------|
| Tatar wrestlers         | 20   | 55.0            | 65.0                 | 75.0                   | 10.0              |
| Elite Russian athletes  | 107  | 53.3            | 71.5                 | 65.4                   | 7.5               |
| Russian population      | 189  | 46.6            | 69.6                 | 66.7                   | 2.9               |
| African (1000 Genomes)  | 661  | 7.7             | 52.1                 | 63.6                   | 9.8               |
| Admixed American (1000 Genomes) | 347 | 60.7            | 77.8                 | 59.1                   | 3.5               |
| East Asian (1000 Genomes) | 504 | 68.6            | 71.2                 | 74.9                   | 25.9              |
| European (1000 Genomes) | 503  | 47.2            | 71.3                 | 66.4                   | 4.8               |
| South Asian (1000 Genomes) | 489 | 54.2            | 62.7                 | 72.3                   | 13.1              |