Ahmetov, II, Kulemin, NA, Popov, DV, Naumov, VA, Akimov, EB, Bravy, YR, Egorova, ES, Galeeva, AA, Generozov, EV, Kostryukova, ES, Larin, AK, Mustafina, LJ, Ospanova, EA, Pavlenko, AV, Starnes, LM, Zmijewski, P, Alexeev, DG, Vinogradova, OL and Govorun, VM

Genome-wide association study identifies three novel genetic markers associated with elite endurance performance

Original Paper DOI: 10.5604/20831862.1124568 Biol. Sport 2015;32:3-9

Genome-wide association study identifies three novel genetic markers associated with elite endurance performance

AUTHORS: Ahmetov II[1,2,3], Kulemin NA[3,4], Popov DV[5], Naumov VA[3], Akimov EB[6], Bravy YR[5], Egorova ES[2], Galeeva AA[1], Generozov EV[3], Kostryukova ES[3], Larin AK[3], Mustafina LJ[1,2], Ospanova EA[3], Pavlenko AV[3], Starnes LM[7], Żmijewski P[8], Alexeev DG[3], Vinogradova OL[5], Govorun VM[3]

Authors' affiliations are listed at the end of the paper

Correspondance author:
Ildus I. Ahmetov
Sport Technology Research Centre, Volga Region State Academy of Physical Culture, Sport and Tourism.
E-mail: genoterra@mail.ru

ABSTRACT: To investigate the association between multiple single-nucleotide polymorphisms (SNPs), aerobic performance and elite endurance athlete status in Russians. By using GWAS approach, we examined the association between 1,140,419 SNPs and relative maximal oxygen consumption rate ($\dot{V}O_2$max) in 80 international-level Russian endurance athletes (46 males and 34 females). To validate obtained results, we further performed case-control studies by comparing the frequencies of the most significant SNPs (with $P<10^{-5}$-10^{-8}) between 218 endurance athletes and opposite cohorts (192 Russian controls, 1367 European controls, and 230 Russian power athletes). Initially, six 'endurance alleles' were identified showing discrete associations with $\dot{V}O_2$max both in males and females. Next, case-control studies resulted in remaining three SNPs (*NFIA-AS2* rs1572312, *TSHR* rs7144481, *RBFOX1* rs7191721) associated with endurance athlete status. The C allele of the most significant SNP, rs1572312, was associated with high values of $\dot{V}O_2$max (males: $P=0.0051$; females: $P=0.0005$). Furthermore, the frequency of the rs1572312 C allele was significantly higher in elite endurance athletes (95.5%) in comparison with non-elite endurance athletes (89.8%, $P=0.0257$), Russian (88.8%, $P=0.007$) and European (90.6%, $P=0.0197$) controls and power athletes (86.2%, $P=0.0005$). The rs1572312 SNP is located on the nuclear factor I A antisense RNA 2 (*NFIA-AS2*) gene which is supposed to regulate the expression of the *NFIA* gene (encodes transcription factor involved in activation of erythropoiesis and repression of the granulopoiesis). Our data show that the *NFIA-AS2* rs1572312, *TSHR* rs7144481 and *RBFOX1* rs7191721 polymorphisms are associated with aerobic performance and elite endurance athlete status.

CITATION: Ahmetov II, Kulemin NA, Popov DV et al. Genome-wide association study identifies three novel genetic markers associated with elite endurance performance. Biol Sport. 2015;32(1):3–9.

Received: 2014-07-25; Reviewed: 2014-09-08; Re-submitted: 2014-09-11; Accepted: 2014-09-25; Published: 2014-10-21.

Key words:
GWAS
genotype
polymorphism
athletes
endurance
$\dot{V}O_2$max

INTRODUCTION

It has long been recognized that the interindividual variability of aerobic performance and the ability to become an elite endurance athlete have a strong genetic basis. A large body of evidence suggests that genetic markers may explain, in part, an interindividual variability of aerobic performance characteristics in response to endurance training [1-4]. With genotyping becoming widely available, a large number of genetic studies evaluating candidate gene variants have been published with largely unconfirmed associations with elite athlete status [5,6].

Case-control studies remain the most common study design in genomics of aerobic capacity and endurance and generally involve determining whether one allele of a DNA sequence (gene or non-coding region of DNA) is more common in a group of elite endurance athletes than it is in the general population, thus implying that the allele boosts performance [7–10]. To avoid false-positive results case-control studies should have at least one replication with additional athletic and non-athletic cohorts from different populations [6, 11,12]. Since endurance and power are located at opposite extremes of the muscle performance continuum, the comparison of allelic frequencies between endurance and power athletes is also the way to identify endurance/power markers [13,14].

Cross-sectional association studies are another type of study design in sports genomics and examine whether athletes with one genotype (or allele) of a particular DNA sequence show different measures of a trait (e.g. $\dot{V}O_{2max}$, percentage of slow-twitch muscle fibres, cardiac size, lactate, etc.) compared to the rest of the sample [15–20].

A genome-wide association study (GWAS) is a new approach that involves rapidly scanning several hundred thousand (up to 5 millions) markers across the complete sets of DNA of many people to find genetic variations associated with a particular trait. One of the advantages of the GWAS approach is that it is unbiased with respect to genomic structure and previous knowledge of the trait (hypothesis-free), in contrast to candidate gene studies, where knowledge of the trait is used to identify candidate loci contributing to the trait of interest [21].

The aim of the study was to investigate the association between multiple single-nucleotide polymorphisms, aerobic performance and elite endurance athlete status in Russians.

MATERIALS AND METHODS

Ethical approval. The study was approved by the Ethics Committee of the Physiological Section of the Russian National Committee for Biological Ethics. Written informed consent was obtained from each participant. The study complied with the guidelines set out in the Declaration of Helsinki. The experimental procedures were conducted in accordance with the set of guiding principles for reporting the results of genetic association studies defined by the STrengthening the REporting of Genetic Association studies (STREGA) Statement.

Study participants

The study involved 449 Russian athletes (267 males and 182 females) stratified into 3 groups according to type, intensity, and duration of exercise. The first group ('group I', n = 158) included long and middle endurance athletes (race duration > 5 min; rowers (n = 33), 3-10 km runners (n = 15), biathletes (n = 28), 5-10 km skaters (n = 3), cross-country skiers (n = 38), marathon runners (n = 2), 5-25 km swimmers (n = 14), 800-1500 m swimmers (n = 7), race walkers (n = 8), triathletes (n = 10). The second group ('group II', n = 61) comprised athletes involved in short endurance events (race duration 45 s – 5 min; 200-400 m swimmers (n = 10), 800-1500 m runners (n = 15), 1500-3000 m skaters and short-trackers (n = 15), 1000 m kayakers and canoers (n = 14), 1-4 km track cycling (n = 7). Overall, the athletes of long, middle and short endurance groups were considered as endurance athletes (n = 219). The third group ('group III' or 'power group'; n = 230) included sprinters, strength and explosive power athletes with predominantly anaerobic energy production (arm-wrestlers (n = 3), 50-100 m swimmers (n = 26), 200 m kayakers and canoers (n = 19), decathletes (n = 3), sprint cyclers (n = 22), 100-400 m runners (n = 29), 500-1000 m skaters and short-trackers (n = 20), track and field jumpers (n = 23), heptathletes (n = 2), powerlifters (n = 16), throwers (n = 10) and weightlifters (n = 57)).

There were 234 athletes classified as 'elite' (prize winners of major international competitions). There were 168 athletes classified as 'sub-elite' (participants in international competitions). The other athletes (n = 47) were classified as 'non-elite' athletes, being regional competitors with at least four years of experience participating in their sports.

Eighty endurance athletes (long endurance athletes: 19 biathletes, 13 cross-country skiers; middle and short endurance athletes: 29 rowers, 11 kayakers and 8 speed skaters) participated in the study of aerobic performance. All athletes were members of the National teams (46 males and 34 females) and 10 of them were prize winners of the Olympic Games.

Russian controls were 192 (141 males and 51 females) unrelated citizens of Russia without any competitive sport experience (all Caucasians of Eastern European descent). Data of European controls (n = 1367) were used from open sources (Illumina_iControl_DB).

Genotyping

Molecular genetic analysis 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 BeadChips (Illumina Inc, USA) were used for genotyping of 1,140,419 SNPs in Russian athletes and controls. The assay required 200 ng of DNA sample as input with a concentration of at least 50 ng·μl^{-1}. Exact concentrations of DNA in each sample were measured using a Qubit Fluorometer (Invitrogen, USA). All further procedures were performed according to the instructions of Infinium HD Assay.

Analysis of hematological parameters

Whole blood of 61 endurance athletes (41 males and 20 females) was collected and tested in hematological analyzer Sysmex XE-2100 (Sysmex Corporation, Japan) to assess hemoglobin content and the total number of erythrocytes, reticulocytes, leukocytes with subpopulations (neutrophils, eosinophils, basophils, monocytes and lymphocytes).

$\dot{V}O_2$max measurement

Maximal oxygen consumption rate ($\dot{V}O_{2\,max}$) in rowers was determined using an incremental test to exhaustion on a PM 3 rowing ergometer (Concept II, Morrisville, Vermont, USA). The initial workload was 150 W. The duration of exercise at each workload was 3 minutes, with a 30 s rest period between increments of 50 W. $\dot{V}O_2$ was determined breath by breath using a MetaMax 3B gas analysis system (Cortex, Leipzig, Germany). Using the MetaMax system, O_2 and CO_2 contents were measured using an electro-chemical cell and non-dispersive infrared sensor, respectively, and air flow was measured using a turbine transducer (Triple V). Two-point gas calibrations (first gas – 15% O_2, 5% CO_2; second gas – ambient air) were performed daily. A one-point gas calibration with ambient air was performed before each test as well as a flow transducer calibration using a 3 L syringe (Hans Rudolph, Kansas City, USA). The criteria used to confirm a maximal test were a decrease in power of more than 30 W from the target power despite strong verbal encouragement and a respiratory exchange ratio greater than 1.1 before cessation of exercise. $\dot{V}O_{2max}$ was recorded as the highest mean value observed over a 30 s period.

$\dot{V}O_{2max}$ in kayakers was determined using an incremental test to exhaustion on a kayaking ergometer (Efremov, Moscow, Russia). The

initial workload was 8 kg for men and 5 kg for women. The duration of exercise at each workload was 2 minutes, with a 30 s rest period between increments of 1 kg. $\dot{V}O_{2max}$ was determined breath by breath using a MetaLyzer II gas analysis system (Cortex, Leipzig, Germany). $\dot{V}O_{2max}$ was recorded as the highest mean value observed over a 30 s period.

$\dot{V}O_{2max}$ in speed skaters was determined using a ramp test to exhaustion on an electromagnetic cycle ergometer Ergoselect 200K (Ergoline, Germany). The initial workload was 60 W, the increment was 15 $W \cdot min^{-1}$, and the target cadence was 60-70 rpm. $\dot{V}O_{2max}$ was determined breath by breath using a MetaMax 3B gas analysis system. The criteria used to confirm a maximal test were a decrease in cadence of less than 50 rpm despite strong verbal encouragement and a respiratory exchange ratio greater than 1.1 before cessation of exercise. $\dot{V}O_{2max}$ was recorded as the highest mean value observed over a 30 s period.

$\dot{V}O_{2max}$ in biathletes and cross-country skiers was determined using an incremental test to exhaustion on a treadmill HP Cosmos (Germany). The initial speed was 7 $km \cdot h^{-1}$, the increment was 0.1 $km \cdot h^{-1}$ every 10 seconds. $\dot{V}O_{2max}$ was determined breath by breath using a MetaMax 3B-R2 gas analysis system. $\dot{V}O_{2max}$ was recorded as the highest mean value observed over a 30 s period.

Calculation of total genotype score

To quantify the combined influence of polymorphisms associated with endurance athlete status, an algorithm [22] was used to incorporate all favourable genotype scores for any given individual in a simple additive model. The total score was then transformed mathematically to lie within the range 0–100 and labelled the 'total genotype score' (TGS). A TGS of 100 represents a 'perfect' polygenic profile for endurance and a TGS of 0 represents the 'worst' possible profile for endurance.

Statistical analysis

Linear regression analysis of a quantitative trait ($\dot{V}O_{2max}$) with GWAS data of five discovery cohorts of athletes (all athletes, female athletes, male athletes, long endurance athletes, middle and short endurance athletes) was performed using PLINK software. The squared correlation coefficient R^2 was used as a measure of explained variance. Differences in phenotypes between groups were analysed using ANOVA or unpaired *t* tests (when variables of two genotypes were compared). All data are presented as mean (standard deviation). Genotype distribution and allele frequencies between athletes and controls were compared using χ^2 tests. *P* values < 0.05 were considered statistically significant. It has been estimated that the sample size required in this study to obtain a statistical power of 80% was sufficient. Statistical analyses were conducted using GraphPad InStat software (GraphPad Software, Inc., USA).

RESULTS

GWAS for $\dot{V}O_2max$. As expected, long endurance athletes had significantly greater values of relative $\dot{V}O_{2max}$ than middle and short endurance athletes of both genders. Furthermore, male athletes had greater aerobic capacity than female athletes (Table 1). We thus performed GWASes of five discovery cohorts of athletes (all athletes, female athletes, male athletes, long endurance athletes, middle and short endurance athletes) and identified fifteen suggestive 'endurance alleles' with $P < 1 \cdot 10^{-5} - 10^{-8}$. For those 15 SNPs we subsequently performed correlation analyses with all subgroups of athletes and remained six SNPs which were replicated in at least 1 female and 1 male subgroup (Table 2).

The C allele of one of the most significant SNPs, *NFIA-AS2* rs1572312 (discovery $P = 1.66 \cdot 10^{-6}$), was associated with high $\dot{V}O_{2\,max}$ in female long endurance athletes (AC genotype: 54.4 ± 2.5, CC genotype: 59.9 ± 3.3 $ml \cdot min^{-1} \cdot kg^{-1}$; $P = 0.0164$), male (AC: 55.3 ± 7.5, CC: 61.0 ± 4.5; $P = 0.0279$) and female (AA/AC: 50.1 ± 4.8, CC: 56.9 ± 5.8; $P = 0.0278$) middle and short endurance athletes, all male (AC: 55.3 ± 7.5, CC: 64.4 ± 6.4; $P = 0.0051$) and all female (AA/AC: 51.5 ± 4.5, CC: 58.5 ± 4.7; $P = 0.0005$) athletes. Male long endurance athletes were excluded from the analysis since there were no carriers of the A allele.

Case-control studies

To validate obtained results, we further performed case-control studies by comparing the frequencies of the six most significant SNPs between 218 endurance athletes (or 100 elite endurance athletes) and opposite cohorts (192 Russian controls, 1367 European controls, and 230 Russian power athletes). We assumed that 'endurance allele' should be under-represented in at least one opposite cohort (Russian controls or Russian power athletes) when compared to endurance athletes. This approach resulted in remaining three SNPs (*NFIA-AS2* rs1572312 C, *TSHR* rs7144481 C, *RBFOX1* rs7191721 G) associated with endurance athlete status (Table 3). In particular, the frequency of the *NFIA-AS2* rs1572312 C allele was significantly higher

TABLE 1. Characteristics of aerobic performance groups

Characteristics	Long endurance athletes		Middle and short endurance athletes	
	Males	Females	Males	Females
N	15	17	31	17
Age, years	21.3 ± 4.1	23.5 ± 3.5	22.9 ± 4.4	20.4 ± 3.6
$\dot{V}O_2max$ $ml \cdot kg^{-1} \cdot min^{-1}$	70.4 ± 4.6*	58.9 ± 3.8**	60.1 ± 5.4*	54.5 ± 6.3**

Note: Values are mean ± SD. *P<0.0001, significantly different between male subgroups of athletes; **P=0.019, significantly different between female subgroups of athletes.

TABLE 2. Summary of most significant SNPs identified by GWAS in different cohorts of athletes

Discovery and replication cohorts	SNP	Gene	'Endurance allele'	P value
All athletes (MLE, MA, FA)	rs7144481 T/C	TSHR	C	$8.92 \cdot 10^{-8}$
All athletes (FLE, MMSE, FMSE, MA, FA)	rs1572312 C/A	NFIA-AS2	C	$1.66 \cdot 10^{-6}$
Male athletes (MLE, MMSE, FMSE)	rs12893597 C/T	ESRRB	T	$7.44 \cdot 10^{-6}$
Middle and short endurance athletes (MMSE, FMSE)	rs9922134 C/T	CDH13	C	$8.85 \cdot 10^{-6}$
Middle and short endurance athletes (MMSE, FMSE, FA)	rs296000 G/T	SLIT3	G	$1.66 \cdot 10^{-5}$
Middle and short endurance athletes (MMSE, FMSE, FA)	rs7191721 G/A	RBFOX1	G	$1.73 \cdot 10^{-5}$

Note: MLE, male long endurance athletes; FLE, female long endurance athletes; MMSE, male middle and short endurance athletes; FMSE, female middle and short endurance athletes; MA, male athletes; FA, female athletes.

in elite endurance athletes (95.5%) in comparison with non-elite endurance athletes (89.8%, $P = 0.0257$), Russian (88.8%, $P = 0.007$) and European (90.6%, $P = 0.0197$) controls and power athletes (86.2%, $P = 0.0005$). Furthermore, all of prize winners of the Olympic Games amongst endurance athletes ($n = 20$) were carriers of the rs1572312 CC genotype (100 vs. 78.6%, $P = 0.0214$ in comparison with Russian controls).

Polygenic analyses

Three significant SNPs (*NFIA-AS2* rs1572312, *TSHR* rs7144481, *RBFOX1* rs7191721) in combination explained 24.6 and 48.8% of the variation in $\dot{V}O_{2max}$ of male and female endurance athletes, respectively. We also determined the total genotype score (TGS) composed from three SNPs in athletes and controls. The mean TGS was significantly higher in endurance athletes (58.9 ± 18.2) than in Russian controls (54.2 ± 15.7, $P = 0.0059$) or power athletes (53.6 ± 18.4), $P = 0.0022$). Elite endurance athletes had even higher TGS (59.4 ± 17.2); $P = 0.01$ in comparison with Russian controls). Compared with carriers of 0-4 'endurance alleles', the odds ratio (OR) of being an endurance athlete in carriers of 5-6 'endurance alleles' was 2.23 (95% confidence interval (CI): 1.215-4.091, $P = 0.0084$). Of note, none of controls had the ideal combination of genotypes (i.e. 6 'endurance alleles'), while there were four endurance athletes with such combination (OR = 8.03, $P = 0.06$).

NFIA-AS2 rs1572312 genotype and hematological parameters

Since *NFIA-AS2* gene may be involved in the regulation of erythropoiesis and granulopoiesis, we hypothesized that variation in this gene is associated with hematological parameters of endurance athletes (41 males and 20 females). Male athletes with *NFIA-AS2* CC genotype had greater hemoglobin content (161.7 ± 7.8 vs. 153.2 ± 7.6 $g \cdot l^{-1}$, $P = 0.0467$), number of erythrocytes per liter ($5.17 \pm 0.26 \cdot 10^{12}$ vs. $4.86 \pm 0.38 \cdot 10^{12}$, $P = 0.0403$) and number of reticulocytes per liter ($62.7 \pm 22.4 \cdot 10^9$ vs. $30.0 \pm 14.1 \cdot 10^9$, $P = 0.0072$) than carriers of the rare A allele. On the other hand, male CC homozygotes had lower number of neutrophils ($3.17 \pm 1.1 \cdot 10^9$ vs. $5.1 \pm 1.8 \cdot 10^9$, $P = 0.0029$) and lower leukocyte/erythrocyte ratio ($P = 0.0332$) than A allele carriers. The same interrelationship was observed in female athletes, but only as a tendency due to small sample size. Overall, these results indicate that *NFIA-AS2* rs1572312 C allele is associated with activation of erythropoiesis, while A allele with activated granulopoiesis.

DISCUSSION

This is the first study to demonstrate that polymorphisms of *NFIA-AS2*, *TSHR* and *RBFOX1* genes are associated with aerobic performance and elite endurance athlete status. Importantly, these results were obtained by two-stage approach using GWAS, which has been successfully used to identify genetic markers associated with relevant

TABLE 3. Allelic frequencies of three most significant SNPs identified by comparison between Russian endurance athletes and opposite cohorts (Russian and European controls, and Russian power athletes)

Frequency of 'endurance allele' in endurance cohorts, %	Allelic frequencies in opposite cohorts, %		
	Russian controls	European controls	Power athletes
NFIA-AS2 rs1572312 C	88.8	90.6	86.2
Endurance (92.4)	$P = 0.0737$	$P = 0.2164$	$P = 0.0028$*
Elite endurance (95.5)	$P = 0.007$**	$P = 0.0197$**	$P = 0.0005$**
TSHR rs7144481 C	17.4	15.7	20.7
Endurance (23.4)	$P = 0.0337$*	$P < 0.0001$*	$P = 0.3389$
Elite endurance (25.7)	$P = 0.0323$**	$P = 0.0014$**	$P = 0.3389$
RBFOX1 rs7191721 G	56.3	63.3	53.7
Endurance (60.8)	$P = 0.1964$	$P = 0.316$	$P = 0.0322$*
Elite endurance (58.6)	$P = 0.6008$	$P = 0.1868$	$P = 0.2473$

Note: *P < 0.05, in comparison with endurance athletes; **P < 0.05, in comparison with elite endurance athletes.

to athletic performance human traits like height, lean body mass and response to endurance training [1,2,23,24].

The most significant SNP (rs157231) was found in the nuclear factor I A antisense RNA 2 (*NFIA-AS2*) gene (localized to chromosome 1). More specifically, the CC genotype was associated with high $\dot{V}O_{2max}$ in three subgroups of athletes (male and female middle and short endurance athletes; female long endurance athletes). Additionally, the frequency of the C allele was significantly higher in elite endurance athletes in comparison with non-elite endurance athletes, Russian and European controls and power athletes.

The *NFIA-AS2* gene encodes long non-coding RNA (lnc-RNA) with undescribed function. Genes of antisense long non-coding RNAs are transcribed from either the same genomic site or a site distant from the gene locus where the sense transcript counterpart is produced. Antisense lnc-RNAs repress - and in some cases can also activate - transcription of the targeted protein coding genes via mechanisms such as DNA methylation and chromatin modification at the genomic loci of the targeted genes [25]. lncRNAs can also act as a sponge for microRNAs that affect differentiation programs [26]. Therefore, one might suggest that *NFIA-AS2* is involved in the regulation of expression of the nuclear factor I A (*NFIA*) gene or erythroid/myeloid specific RNAs. NFIA, as a transcription factor, induces erythropoiesis, whereas its silencing drives granulopoiesis [27]. Indeed, in the present study, we have shown that *NFIA-AS2* rs1572312 C allele is associated with activation of erythropoiesis (high level of hemoglobin, high number of reticulocytes and erythrocytes), while A allele with activated granulopoiesis (high number of neutrophils and greater leukocyte/erythrocyte ratio). Thus, it is probably partly via associations with erythropoiesis (factor influencing aerobic capacity) that the *NFIA-AS2* gene rs1572312 polymorphism contributes to a small but perhaps important component of the ability to perform at a high level in endurance events.

On the other hand, the discovered association might be explained by linkage disequilibrium (LD) in a functional variant of the same or a different gene. For instance, *NFIA-AS2* rs1572312 is in a strong (100%) LD with rs17304036 SNP. The use of the HaploReg (version 2) online program predicted the change of transcription factor-binding motifs in the rs17304036 locus of the *NFIA-AS2* gene for HOXA5 and MEF2A transcription factors. HOXA5 encodes a DNA-binding transcription factor which may regulate gene expression, morphogenesis, and differentiation, while MEF2A is involved in several cellular processes, including muscle development (it promotes the transformation of type II fast glycolytic fibres into type I slow oxidative fibres), neuronal differentiation, cell growth control, and apoptosis [28,29].

The rs7144481 polymorphism seems to be a significant predictor of elite endurance performance too (discovery *P* value - $8.92 \cdot 10^{-8}$). That is, rs7144481 C allele in the present study was associated with greater aerobic capacity in three subgroups of athletes (male long endurance athletes; all male and all female athletes). The significance of this SNP was confirmed in case-control studies (higher frequency of the C allele in elite endurance athletes in comparison with Russian and European controls). The rs7144481 polymorphism is located in the 3'-untranslated region of the thyroid stimulating hormone receptor (*TSHR*) gene (location: 14q31). The protein encoded by this gene is a membrane receptor for thyrothropin (produces thyroid hormones) and thyrostimulin (activates TSHR), and therefore, a major controller of thyroid cell metabolism. Thyroid hormones are known as determinants of metabolic and contractile phenotype of skeletal muscle [30]. TSHR also mediates the effect of thyrothropin on angiogenesis via cAMP-mammalian target of rapamycin signaling [31]. Although further *in vitro* functional studies are needed, one might speculate that the rs7144481 C allele is associated with greater expression of *TSHR* gene and increased angiogenesis and/or metabolic rate – factors that directly influence aerobic performance.

RNA binding protein, fox-1 homolog (C. elegans) 1 (RBFOX1) is an important splicing factor regulating developmental and tissue-specific alternative splicing in heart, muscle, and neuronal tissues [32]. Since RNA-binding proteins are key regulators of gene expression, Rbfox1 knockdown leads to the inhibition of muscle differentiation [33]. RBFOX1 (location: 16p13.3) is implicated in multiple medical conditions, including muscular dystrophies, cancers, neurodevelopmental and neuropsychiatric disorders [33,34]. We therefore speculate that *RBFOX1* rs7191721 G allele may be favourable for muscle function of endurance athletes. Indeed, our data show that middle and short endurance athletes of both genders, as well as all female endurance athletes with G allele have greater aerobic capacity. Additionally, the G allele is over-represented in endurance athletes when compared to power athletes.

The limitation of the study is the sample size of our discovery cohort with $\dot{V}O_{2max}$ data (n=80). Relatively small cohort did not allow to reach genome-wide significance level which was set at $P = 4.38 \cdot 10^{-8}$ (i.e. 0.05/1,140,419 SNPs). However, we believe that using additional approaches (validation of results in different gender groups, case-control study, biochemical study) had minimized false-positive results.

Our current progress towards understanding the molecular basis of athletic performance represents only the first steps. The next decade will be an exciting period for sports genomics, as we apply the new DNA technologies (whole genome sequencing, GWAS, epigenomic, transcriptomic and proteomic profiling etc.) and bioinformatics to further dissect and analyze the genetic effects on human physical ability [6]. Efforts to perform GWAS in the cohorts of athletes are presently underway (at least athletes from Australia, Canada, Ethiopia, Finland, Germany, Greece, Italy, Jamaica, Japan, Kenya, Poland, Russia and USA).

CONCLUSIONS

In conclusion, our data show that the *NFIA-AS2* rs1572312, *TSHR* rs7144481 and *RBFOX1* rs7191721 polymorphisms are associated with aerobic performance and elite endurance athlete status, but further replication studies using different ethnic cohorts are warranted.

Acknowledgements

This work was supported by grants from the Federal Medical-Biological Agency of the Russian Federation ("Sportgen project") and the Ministry of Education and Science of the Russian Federation (contract number 02.522.11.2004).

Conflict of interests: the authors declared no conflict of interests regarding the publication of this manuscript.

Authors' affiliations:
[1] Volga Region State Academy of Physical Culture, Sport and Tourism, Sport Technology Research Centre, Kazan, Russia
[2] Kazan State Medical University, Laboratory of Molecular Genetics, Kazan, Russia
[3] Research Institute for Physical-Chemical Medicine, Department of Molecular Biology and Genetics, Moscow, Russia
[4] Moscow Institute of Physics and Technology (State University), Moscow, Russia.
[5] SSC RF Institute for Biomedical Problems of the Russian Academy of Sciences, Laboratory of Exercise Physiology, Moscow, Russia
[6] Centre for Sports Innovation Technologies and National Teams of the Moscow Department of Physical Culture and Sport, Moscow, Russia
[7] University of Copenhagen, Novo Nordisk Foundation Center for Protein Research, Copenhagen, Denmark
[8] Institute of Sport, Department of Physiology, Warsaw, Poland

REFERENCES

1. Bouchard C, Sarzynski MA, Rice TK, Kraus WE, Church TS, Sung YJ, Rao DC, Rankinen T. Genomic predictors of the maximal O_2 uptake response to standardized exercise training programs. J Appl Physiol. 2011;110(5):1160–70.
2. Timmons JA, Knudsen S, Rankinen T, Koch LG, Sarzynski M, Jensen T, Keller P, Scheele C, Vollaard NBJ, Nielsen S, Akerström T, MacDougald OA, Jansson E, Greenhaff PL, Tarnopolsky MA, van Loon LJC, Pedersen BK, Sundberg CJ, Wahlestedt C, Britton SL, Bouchard C. Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans. J Appl Physiol. 2010;108(6):1487–96.
3. Brzeziańska E, Domańska D, Jegier A. Gene doping in sport – perspectives and risks. Biol Sport. 2014;31(4):251–9.
4. Pokrywka A, Kaliszewski P, Majorczyk E, Zembroń-Łacny A. Genes in sport and doping. Biol Sport. 2013;30(3):155–61.
5. Bray MS, Hagberg JM, Pérusse L, Rankinen T, Roth SM, Wolfarth B, Bouchard C. The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. Med Sci Sports Exerc. 2009;41(1):35–73.
6. Wang G, Padmanabhan S, Wolfarth B, Fuku N, Lucia A, Ahmetov II, Cieszczyk P, Collins M, Eynon N, Klissouras V, Williams A, Pitsiladis Y. Genomics of elite sporting performance: what little we know and necessary advances. Adv Genet. 2013;84:123–49.
7. Ahmetov II, Khakimullina AM, Popov DV, Missina SS, Vinogradova OL, Rogozkin VA. Polymorphism of the vascular endothelial growth factor gene (VEGF) and aerobic performance in athletes. Hum Physiol. 2008;34(4):477–81.
8. Ahmetov II, Hakimullina AM, Popov DV, Lyubaeva E V, Missina SS, Vinogradova OL, Williams AG, Rogozkin VA. Association of the VEGFR2 gene His472Gln polymorphism with endurance-related phenotypes. Eur J Appl Physiol. 2009;107(1):95–103.
9. Cieszczyk P, Krupecki K, Maciejewska A, Sawczuk M. The angiotensin converting enzyme gene I/D polymorphism in Polish rowers. Int J Sports Med. 2009;30(8):624–7.
10. Döring F, Onur S, Fischer A, Boulay MR, Pérusse L, Rankinen T, Rauramaa R, Wolfarth B, Bouchard C. A common haplotype and the Pro582Ser polymorphism of the hypoxia-inducible factor-1alpha (HIF1A) gene in elite endurance athletes. J Appl Physiol. 2010;108(6):1497–500.
11. Maciejewska A, Sawczuk M, Cieszczyk P, Mozhayskaya IA, Ahmetov II. The PPARGC1A gene Gly482Ser in Polish and Russian athletes. J Sports Sci. 2012;30(1):101–13.
12. Wang G, Mikami E, Chiu L-L, DE Perini A, Deason M, Fuku N, Miyachi M, Kaneoka K, Murakami H, Tanaka M, Hsieh L-L, Hsieh SS, Caporossi D, Pigozzi F, Hilley A, Lee R, Galloway SDR, Gulbin J, Rogozkin VA, Ahmetov II, Yang N, North KN, Ploutarhos S, Montgomery HE, Bailey MES, Pitsiladis YP. Association analysis of ACE and ACTN3 in elite Caucasian and East Asian swimmers. Med Sci Sports Exerc. 2013;45(5):892–900.
13. Drozdovska SB, Dosenko VE, Ahmetov II, Ilyin VN. The association of gene polymorphisms with athlete status in Ukrainians. Biol Sport. 2013;30(3): 163–7.
14. Eynon N, Ruiz JR, Bishop DJ, Santiago C, Gómez-Gallego F, Lucia A, Birk R. The rs12594956 polymorphism in the NRF-2 gene is associated with top-level Spanish athlete's performance status. J Sci Med Sport. 2013;16(2):135–9.
15. Akhmetov II, Popov D V, Mozhaĭskaia IA, Missina SS, Astratenkova I V, Vinogradova OL, Rogozkin VA. Association of regulatory genes polymorphisms with aerobic and anaerobic performance of athletes. Ross Fiziol Zh Im I M Sechenova. 2007;93(8):837–43.
16. Akhmetov II, Linde E V, Shikhova I V, Popov D V, Missina SS, Vinogradoba OL, Rogozkin VA. The influence of calcineurin gene polymorphism on morphofunctional characteristics of cardiovascular system of athletes. Ross Fiziol Zh Im I M Sechenova. 2008;94(8):915–22.
17. Fedotovskaya ON, Popov D V., Vinogradova OL, Akhmetov II. Association of muscle-specific creatine kinase (CKMM) gene polymorphism with physical performance of athletes. Hum Physiol. 2012;38(1):89–93.
18. Fedotovskaya ON, Mustafina LJ, Popov DV, Vinogradova OL, Ahmetov II. A common polymorphism of the MCT1 gene and athletic performance. Int J Sports Physiol Perform. 2014;9(1):173–80.
19. Lucia A, Gómez-Gallego F, Barroso I, Rabadán M, Bandrés F, San Juan AF, Chicharro JL, Ekelund U, Brage S, Earnest CP, Wareham NJ, Franks PW. PPARGC1A genotype (Gly482Ser) predicts exceptional endurance capacity in European men. J Appl Physiol. 2005;99(1):344–8.
20. Mustafina LJ, Naumov VA, Cieszczyk P, Popov DV, Lyubaeva EV, Kostryukova ES, Fedotovskaya ON, Druzhevskaya AM, Astratenkova IV, Glotov AS, Alexeev DG, Mustafina MM, Egorova ES, Maciejewska-Karłowska A, Larin AK, Generozov EV, Nurullin RE, Jastrzębski Z, Kulemin NA, Ospanova EA, Pavlenko AV, Sawczuk M, Akimov EB, Danilushkina AA, Zmijewski P, Vinogradova OL, Govorun VM, Ahmetov II. AGTR2 gene polymorphism is associated with muscle fibre composition, athletic status and aerobic performance. Exp Physiol. 2014;99(8):1042-52.

21. Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. Genetics. 2011;187(2):367–83.
22. Hughes DC, Day SH, Ahmetov II, Williams AG. Genetics of muscle strength and power: polygenic profile similarity limits skeletal muscle performance. J Sports Sci. 2011;29(13):1425–34.
23. Liu X-G, Tan L-J, Lei S-F, Liu Y-J, Shen H, Wang L, Yan H, Guo Y-F, Xiong D-H, Chen X-D, Pan F, Yang T-L, Zhang Y-P, Guo Y, Tang NL, Zhu X-Z, Deng H-Y, Levy S, Recker RR, Papasian CJ, Deng H-W. Genome-wide association and replication studies identified TRHR as an important gene for lean body mass. Am J Hum Genet. 2009;84(3):418–23.
24. Liu F, Hendriks AEJ, Ralf A, Boot AM, Benyi E, Sävendahl L, Oostra BA, van Duijn C, Hofman A, Rivadeneira F, Uitterlinden AG, Drop SLS, Kayser M. Common DNA variants predict tall stature in Europeans. Hum Genet. 2014;133(5):587–97.
25. Ling H, Fabbri M, Calin GA. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. Nat Rev Drug Discov. 2013;12(11):847–65.
26. Kallen AN, Zhou X-B, Xu J, Qiao C, Ma J, Yan L, Lu L, Liu C, Yi J-S, Zhang H, Min W, Bennett AM, Gregory RI, Ding Y, Huang Y. The imprinted H19 lncRNA antagonizes let-7 microRNAs. Mol Cell. 2013;52(1):101–12.
27. Starnes LM, Sorrentino A, Pelosi E, Ballarino M, Morsilli O, Biffoni M, Santoro S, Felli N, Castelli G, De Marchis ML, Mastroberardino G, Gabbianelli M, Fatica A, Bozzoni I, Nervi C, Peschle C. NFI-A directs the fate of hematopoietic progenitors to the erythroid or granulocytic lineage and controls beta-globin and G-CSF receptor expression. Blood. 2009;114(9):1753–63.
28. Theobald J, DiMario JX. Lineage-based primary muscle fiber type diversification independent of MEF2 and NFAT in chick embryos. J Muscle Res Cell Motil. 2011;31(5-6):369–81.
29. Wu H, Rothermel B, Kanatous S, Rosenberg P, Naya FJ, Shelton JM, Hutcheson KA, DiMaio JM, Olson EN, Bassel-Duby R, Williams RS. Activation of MEF2 by muscle activity is mediated through a calcineurin-dependent pathway. EMBO J. 2001;20(22):6414–23.
30. Simonides WS, van Hardeveld C. Thyroid hormone as a determinant of metabolic and contractile phenotype of skeletal muscle. Thyroid. 2008;18(2):205–16.
31. Balzan S, Del Carratore R, Nicolini G, Beffy P, Lubrano V, Forini F, Iervasi G. Proangiogenic effect of TSH in human microvascular endothelial cells through its membrane receptor. J Clin Endocrinol Metab. 2012;97(5):1763–70.
32. Kuroyanagi H. Fox-1 family of RNA-binding proteins. Cell Mol Life Sci. 2009;66(24):3895–907.
33. Fogel BL, Wexler E, Wahnich A, Friedrich T, Vijayendran C, Gao F, Parikshak N, Konopka G, Geschwind DH. RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. Hum Mol Genet. 2012;21(19):4171–86.
34. Sengupta N, Yau C, Sakthianandeswaren A, Mouradov D, Gibbs P, Suraweera N, Cazier J-B, Polanco-Echeverry G, Ghosh A, Thaha M, Ahmed S, Feakins R, Propper D, Dorudi S, Sieber O, Silver A, Lai C. Analysis of colorectal cancers in British Bangladeshi identifies early onset, frequent mucinous histotype and a high prevalence of RBFOX1 deletion. Mol Cancer. 2013;12:1.