Genome-wide association study reveals novel genetic markers associated with endurance athlete status

Fatima Al-Khelaifi

Anti-Doping Lab Qatar

Noha Yousri

Weill Cornell Medical College in Qatar

Omar Albagha

Qatar Biomedical Research Institute

Ekaterina A. Semenova

FGBU Federal'nyj naucno-kliniceskij centr specializirovannyh vidov medicinskoj pomosi i medicinskih tehnologij FMBA Rossii

Elena Kostryukova

FGBU Federal'nyj naucno-kliniceskij centr specializirovannyh vidov medicinskoj pomosi i medicinskih tehnologij FMBA Rossii

Nickolay A Kulemin

FGBU Federal'nyj naucno-kliniceskij centr specializirovannyh vidov medicinskoj pomosi i medicinskih tehnologij FMBA Rossii

Oleg V Borisov

FGBU Federal'nyj naucno-kliniceskij centr specializirovannyh vidov medicinskoj pomosi i medicinskih tehnologij FMBA Rossii

Liliya B Andryushchenko

FGBU Federal'nyj naucno-kliniceskij centr specializirovannyh vidov medicinskoj pomosi i medicinskih tehnologij FMBA Rossii

Andrey K Larin

FGBU Federal'nyj naucno-kliniceskij centr specializirovannyh vidov medicinskoj pomosi i medicinskih tehnologij FMBA Rossii

Edward V Generozov

FGBU Federal'nyj naucno-kliniceskij centr specializirovannyh vidov medicinskoj pomosi i medicinskih tehnologij FMBA Rossii

Francesco Donati

FMSI

Francesco Botrè

FMSI

David Abraham

University College London

Aron Hingorani

University College London

Costas Georgakopoulos

Anti-Doping Lab Qatar

Karsten Suhre

Weill Cornell Medical College in Qatar

Ildus Ahmetov

FGBU Federal'nyj naucno-kliniceskij centr specializirovannyh vidov medicinskoj pomosi i medicinskih tehnologij FMBA Rossii

Mohamed Elrayess (✉ maelrayess@hotmail.com)

Anti-Doping Lab Qatar <https://orcid.org/0000-0003-3803-4604>

Research article

Keywords: GWAS, SNP, Metabolomics, Metabolites, Elite athletes, Endurance

Posted Date: September 8th, 2019

DOI: <https://doi.org/10.21203/rs.2.14107/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: The genetic predisposition to elite athletic performance has been a controversial subject due to the underpowered studies and the small effect size of identified genetic variants. The aims of this study were to investigate the association of common single-nucleotide polymorphisms (SNPs) with endurance athlete status in a large cohort of elite athletes using GWAS approach, followed by a functional validation of significant SNPs by metabolomics analysis. **Results:** The association of 476,728 SNPs of Illumina DrugCore Gene chip and endurance athlete status was investigated in 753 European international-level athletes (594 males, 159 females) by comparing allelic frequencies between athletes specialized in sports with high (n=630) and low/moderate (n=123) aerobic component. Validation of results was performed by comparing the frequencies of the most significant SNPs between 176 elite Russian endurance athletes and 173 Russian controls or 43 sprinters. Two novel SNPs showed significant associations with endurance athlete status at Bonferroni level of significance (rs56330321 in ATP2B2, $p=1.46E-7$) and FDR level of significance (rs2635438 in SYNE1, $p=2.54E-7$), respectively. Replication study using Russian cohorts and a subsequent meta-analysis have confirmed the association of rs56330321 and rs2635438 SNPs with endurance athlete status at genome-wide significance ($P=5.13E-09$ and $1.91E-08$, respectively). Metabolomics analysis revealed several amino acids and lipids associated with the identified SNPs with potential roles in performance enhancement. **Conclusions:** This is the first report of GWAS significant SNPs and related metabolites associated with elite athlete status. Further investigations of the functional relevance of the identified SNPs and metabolites in relation to enhanced athletic performance are warranted.

Background

Elite athletic performance is a multi-factorial trait with input from both genetic and environmental factors. The superior performance of elite athletes has been historically considered an outcome of a special talent shaped by intensive training. The talent is now believed to be a product of additive genetic components predisposing the athlete to endurance, speed, strength, flexibility and coordination trainability under the control of strong environmental cues including exercise and nutrition. In this model, the genetic predisposition together with ability to respond to training are the keys to the superior physical performance of elite athletes (1).

Sports can be classified according to the type and intensity of the exercise required to perform during competition. The percentage of maximal oxygen uptake (VO_{2max}) is a detrimental factor in the categorization of endurance sports, as it reflects the maximal cardiac output, the oxygen transport capacity, and the blood volume (2). Accordingly, sports can be divided into sport events with low, moderate and high aerobic (dynamic) component (3). Similarly, the percent of maximal voluntary contraction (MVC), which reflects the greatest amount of tension a muscle can generate and hold, is used to classify sports into sporting disciplines with low, moderate and high power component (3).

Classical twin and family genetic studies have suggested that VO_{2max} is up to 94% inherited (4, 5). Genome-wide association studies (GWAS) in athletes versus non-athletes have uncovered many new loci in association with VO_{2max} (6, 7) and elite endurance performance (8). A more recent review of genetic predisposition to elite athletic endurance has highlighted 93 endurance variants (9). However, research into the genetics of athletic performance has been hindered by a small sample size and complex phenotype (10). One of the first GWAS in athletes using 143 K single-nucleotide polymorphisms (SNPs) and subsequent meta-analysis of 45 promising genetic markers in 1520 endurance athletes and 2760 controls has revealed only one statistically significant marker (rs558129 at *GALNTL6*) associated with endurance status in world class athletes, although the association did not reach the genome-wide level of significance (11). Therefore, the genetic predisposition to endurance traits remains unclear, largely due to the relatively underpowered elite athletes' cohorts.

Metabolomics analysis has presented a novel tool to validate genomics data by providing an intermediate phenotype (metabolites) in association with the identified genetic variants (12, 13). Pilot metabolomics studies have revealed differences in the metabolic signature of moderate and high endurance elite athletes, such as steroid biosynthesis, fatty acid metabolism, oxidative stress and energy-related molecular pathways (14, 15).

In this study, we aimed to investigate the association of multiple SNPs and endurance athlete status in a relatively large cohort of European elite athletes specialized in sports with high and low/moderate aerobic component using GWAS approach and replicate our findings in elite Russian athletes and matched controls. We also aimed to perform functional validation using metabolomics analysis by identifying metabolites that are associated with significant endurance-related SNPs.

Methods

The aim of this study is to investigate the genetic predisposition to elite athletic endurance through conducting the largest GWAS in elite athletes to date, followed by functional validation through metabolomics study to shed light on the underlying mechanisms of genetic associations.

Participants

Discovery study

Seven hundred and fifty-three consented European international-level athletes (594 males, 159 females) from different sports disciplines who participated in national or international sports events and tested negative for doping substances at anti-doping laboratories in Qatar (ADLQ) and Italy (FMSI) were included in this study. No other information of participants was available due to the strict anonymization process undertaken by the anti-doping laboratories. This study was performed in line with the World Medical Association Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. All protocols were approved by the Institutional Research Board of ADLQ (F2014000009). Athletes were dichotomized into groups with different aerobic (dynamic) and power (static) components (Table 1) based on their sport types as described previously (3). Table 1 further lists the number of participants based on various analyses as per sport type in each class/group and their genders.

Replication study

The Russian athletes' study involved 219 athletes (95 females, age 21.9 (3.5) years, 124 males, age 22.1 (4.2) years; 43 sprinters, 120 middle-distance athletes, 56 long-distance athletes). Sprinters included 8 100-400 m runners, 5 sprint cyclers, 10 500-1000 m speed skaters / short trackers, 19 50-100 m swimmers, 1 200 m kayaker. Middle-distance athletes comprised 59 rowers, 10 0.8-1.5 km runners, 7 middle-distance cyclers, 21 middle-distance kayakers / canoers, 15 1.5-3.0 km speed skaters, 8 200-400 m swimmers. Long-distance athletes included 3 3-10 km runners, 1 marathon runner, 14 biathletes, 12 cross-country skiers, 14 0.8-25 km swimmers, 6 triathletes, 6 race walkers. All athletes were Olympic team members (International level) who have tested negative for doping substances. Russian controls were 173 (126 males and 47 females) unrelated citizens of Russia without any competitive sport experience (all Caucasians of Eastern European descent). The Russian 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 standards in sport and exercise science research. 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.

Genotyping

Discovery study

DNA was extracted from leukocytes (venous blood) samples from all participants using DNeasy Blood & Tissue kit (Qiagen) following manufacturer's instructions. The concentration and the quality of DNA were assessed using the Nanodrop (Thermo Fisher) and Qubit Fluorometer (Invitrogen) to ensure sufficient amount and quality of DNA were obtained for genotyping. Illumina Drug Core array-24 BeadChips was chosen for the genotyping of 476,728 SNPs in the 753 European elite athletes collected for Anti-Doping analysis (discovery cohort). This array contains over 240,000 highly-informative genome-wide tag SNPs and a novel ~200,000 custom marker set designed to support studies of drug target validation and treatment response. The assay required 200 ng of DNA sample as input with a concentration of at least 50 ng/μl. All further procedures were performed according to the instructions of Infinium HD Assay according to manufacturer's instructions. Briefly, 4 μl of obtained DNA was mixed with Illumina amplification reagents and incubated overnight at 37°C in hybridization oven. On the second day, enzymatic reagents were used to fragment the amplified DNA then precipitated by centrifugation. Subsequently, re-suspended pellet was loaded in the beadchip then incubated overnight at 48°C in hybridization oven. On third day, beadchips underwent enzymatic base extension and fluorescent staining. Lastly, after coating, the beadchips were imaged using iScan.

Replication study

Molecular genetic analysis in Russian cohorts 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 athletes and controls. The assay required 200 ng of DNA sample as input with a concentration of at least 50 ng/μl. 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.

Data Extraction and SNP Identification

Raw data was extracted, peak-identified and QC processed using Illumina iScan hardware and software. These systems are built on a web-service platform utilizing Microsoft's NET technologies, which run on high-performance application servers and fiber-channel storage arrays in clusters to provide active failover and load-balancing.

Metabolomics

Screening of serum metabolites was performed in 490 elite athletes (Table S1) using protocols established at Metabolon, Durham, NC, USA. The platform utilizes Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. Detailed protocol and QC measures were previously published (14, 42).

Statistical analysis

Following genotyping using Illumina's Drug Core SNP array, analysis was performed using Plink v1.9. Quality control measures were applied to the genotype data set to exclude samples with low genotype call rate or excess heterozygosity. Accordingly, SNPs with a genotype call rate < 98%, minor allele frequency < 1%, or deviating from Hardy-Weinberg equilibrium ($P < 10E-6$) were excluded. After filtering the data with the above criteria, 341,385 SNPs were used in analysis. Population background was determined using principal component analysis (PCA) in comparison to samples from 1000 genome project and only samples with European ancestry were included in the analysis. The analysis in European and Russian cohorts was performed using linear or logistic regression models. A model incorporating sports grouped by training modalities (i.e. sports with high vs. low/moderate aerobic component) was used for the discovery cohort after incorporating gender and PCA components 1, 2, 3 & 4 as covariates in the model. A stringent Bonferroni level of significance of $p < 0.05/341385 = 1.46E-7$ was used to define significant associations. To perform the meta-analysis, the Cochrane Review Manager version 5.3 was used. Random and fixed effect models were applied. The heterogeneity degree between the studies was assessed with the I^2 statistics. Associations between SNPs and metabolite levels were computed using `lm` function in R (version 3.3.1) while correcting for gender, hemolysis and PCA. An additive inheritance model was used (SNPs were coded as 0,1,2 according to their genotype group). Pathway enrichment analyses were carried out using Chi square tests to identify pathways with enriched metabolites ranked by p-value from the linear model since Bonferroni level of significance was not observed.

Results

Genome-Wide Association study

Athletes from the discovery cohort were classified into different groups of sports following previously published sports classification criteria (3), as shown in Table 1.

The PCA of the genotyping data revealed no influence of sport disciplines (Figure 1A) or training modality (i.e. sports with low/moderate vs high aerobic component) (Figure 1B) on genotype distribution. The principle components from the PCA were used in the subsequent association analysis to correct for any possible population stratification.

Following quality control data processing, genotyping of 341385 SNPs in 753 European elite athletes has revealed several variants associated with endurance athlete status, with one novel SNP (rs56330321 in ATPase Plasma Membrane Ca^{2+} Transporting 2 gene, *ATP2B2*) reaching Bonferroni level of significance ($p=1.47E10-7$) and another novel SNP (rs2635438 in Spectrin Repeat Containing Nuclear Envelope Protein 1 gene, *SYNE1*) reaching FDR level of significance of 5% ($p=2.54E10-7$). Table 2 shows top 10 SNPs ($P < E-5$) with their odd ratios (OR) in relation to elite athletic endurance, location according to function genome variation server (GVS), gene name and minor allele frequency (MAF) in sports with high and low/moderate aerobic component. MAF in non-elite athletes from 1000 genome project were used as a reference. Figure 2 shows Manhattan (A) and quartile-quartile (QQ) plots (B) of GWAS hits associated with endurance.

Validation of endurance SNPs in a second cohort

Validation of results was performed by comparing the frequencies of the most significant SNPs ($P < E-5$) in 219 elite Russian athletes (120 middle-distance athletes, 56 long-distance athletes, 43 sprinters) and 173 Russian controls. The rs56330321 A allele was under-represented (i.e. unfavorable) in Russian middle-distance athletes ($n=120$) compared to 173 Russian controls (0.8 vs 3.8%; OR=0.2036; SE=0.6435; $P=0.013$). The rs2635438 G allele was under-represented in 56 Russian long-distance athletes compared to 43 elite Russian sprinters (3.6 vs 8.1%; OR=0.132; SE=0.9004; $P=0.024$). A subsequent meta-analysis has confirmed the under representation of the rs56330321 A and rs2635438 G alleles in high endurance sports at genome-wide level of significance ($5.13E-09$ and $1.91E-08$, respectively).

Functional metabolic validation of GWAS significant SNPs

To validate the potential functionality of the identified GWAS SNPs, metabolomics of 750 metabolites was carried out in a subset of the discovery cohort (n=490) and enriched metabolic pathways associated with the rs56330321 and rs2635438 were determined (Table 3). Among the metabolic pathways associated with rs56330321, ceramides, fatty acid (Acyl Carnitine), polyamine and creatine metabolites were significantly altered by rs56330321 genotype (Table 3, Figure 3). Whereas, gamma-glutamyl amino acid and glutamate metabolic pathways were significantly changed with rs2635438 (Table 3, Figure 4).

Discussion

Genetic predisposition into cardiorespiratory fitness and response to exercise training has been previously described (5, 16-20). Since endurance sports are characterized by increased cardiorespiratory capacity, genetic predisposition into elite endurance performance is also expected to be genetically influenced (21). However, genetic studies of elite athletic endurance showed inconsistent results (10, 21-23). The aims of this study were to carry out the largest GWAS study of elite European athletes to date using a unique SNP microarray that is enriched with genes involved in different metabolic pathways with direct influence on various physiological pathways characteristic of elite athletes. GWAS results have revealed two novel SNPs (rs56330321 and rs2635438) associated with endurance at Bonferroni and FDR level of significance, respectively. Validation of the results in an independent cohort of elite Russian athletes and controls has confirmed the association of rs56330321 and rs2635438 with endurance athlete status. Subsequent meta-analysis of the two cohorts has shown for the first time that both SNPs were associated with endurance athlete status at genome-wide level of significance.

The two novel SNPs (rs56330321 and rs2635438) are located within genes *ATP2A2* and *SYNE2*, respectively. Although these two genes have not been previously implicated directly in physical performance, their potential roles in cell signaling and cytoskeletal structure of skeletal muscle cells were previously established (24, 25). Investigation of the functional relevance of these SNPs in relation to enhanced athletic performance was sought using metabolomics analysis. The metabolic pathways associated with the two significant SNPs included various ceramides, fatty acids, acyl carnitine, polyamines, creatine/creatinine, gamma-glutamyl amino acids and glutamate metabolites. The functional relevance of these associations remains to be further validated.

The top GWAS significant SNP (rs56330321) is located within the intron of *ATP2B2*. This gene codes for the plasma membrane Ca²⁺ ATPase 2 (PMCA2) that belongs to the P-type primary ion transport ATPases. These enzymes can remove bivalent calcium ions from the cell against high gradients, providing a pivotal role in intracellular calcium homeostasis (24). PMCA2 is mainly expressed in the inner ear, the cerebellum and the mammary gland with an established role in hearing and balance in mice (26) and humans (27). The expression of different isoforms and splice variants is highly regulated following the physiological demand of the cell (28). The association between PMCA2 and physical performance has not been previously described. The under representation of the rs56330321 A allele in athletes specialized in sports with high aerobic component in the discovery and replication cohorts, compared to other athletes and controls, may suggest that carrying the A allele is disadvantageous for endurance athletes.

The rs56330321 A allele is associated with higher levels of several ceramides, fatty acids, acyl carnitine, polyamines, except for acisoga (N-(3-acetamidopropyl)pyrrolidin-2-one), and creatine/creatinine (Figure 3). Ceramides tend to accumulate in skeletal muscles and promote insulin resistance. Chronic endurance exercise lowers muscle ceramides and promotes the insulin-sensitivity in exercising muscle (29). Since the A allele is associated with higher ceramides levels, it could be compromising the beneficial effect of exercise in carriers on improving insulin sensitivity (30). The A allele is also associated with higher levels of fatty acids, acyl carnitines, a hallmark of active fatty acid oxidation. During endurance exercise, fatty acids oxidation increases, sparing glycogen and delaying muscle fatigue (31). Despite the beneficial effect of fatty acid oxidation in endurance athletes, the elevated fatty acid acyl carnitines in A allele carriers may represent a compensatory mechanism to counteract ceramides-induced impairment of fatty acid oxidation (32). The A allele was also associated with higher polyamine accumulation, except for acisoga. The increase in polyamines concentration in exercising skeletal muscle after physical exercise reflects oxidative processes related to muscle adaptation to exercise (33). The elevated polyamines in A allele carriers may therefore reflect higher oxidative mechanisms, also suggested by the increased acyl carnitines, in response to endurance exercise. The elevated creatine/creatinine levels in A allele carriers may suggest worse renal functions compared to GG individuals, perhaps contributing to their lower prevalence in high endurance athletes (34). The direct link between rs56330321 and levels of these metabolites is yet to be determined.

The second GWAS significant SNP (rs2635438) is located within the intron of *SYNE1*. This gene encodes a spectrin repeat containing protein expressed in skeletal and smooth muscle, and peripheral blood lymphocytes, that localizes to the nuclear membrane. Mutations in this gene have been associated with autosomal recessive spinocerebellar ataxia 8, Emery-Dreifuss muscular dystrophy type 4, dominant muscular dystrophy and Emery-Dreifuss muscular dystrophy-like (35-40). Both discovery and replication cohorts have shown that the G allele is under represented in endurance athletes compared to other athletes, suggesting that carriers of the G allele may have lower endurance ability, perhaps through replacement of healthy muscle tissue by fibrosis and fatty infiltration described in recessive arthrogryposis families carrying mutations in *SYNE1* gene (40). The G allele is associated with lower gamma-glutamyl amino acids and glutamine but higher glutamate metabolism. Glutamine has various ergogenic benefits including increased muscle strength and better recovery (41). The lower levels of glutamine in G allele carriers may

partially explain their lower prevalence in endurance athletes. The direct link between rs2635438 and levels of these and other metabolites remains to be determined.

Conclusions

This study reports the first GWAS significant SNPs associated with endurance athlete status in genes with no previous association with physical performance. This study also shows levels of metabolites associated with these SNPs and suggests potential role in performance enhancement. Further investigations of the functional relevance of the identified SNPs and associated metabolites in relation to enhanced athletic performance are warranted.

Abbreviations

Anti-doping laboratories in Qatar (ADLQ)
ATPase Plasma Membrane Ca ²⁺ Transporting 2 gene (ATP2B2)
False discovery rate (FDR)
Genome variation server (GVS)
Genome-wide association studies (GWAS)
High resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II)
Laboratorio Antidoping, Federazione Medico Sportiva Italiana (FMSI)
Maximal oxygen uptake (VO ₂ max)
Maximal voluntary contraction (MVC)
Minor allele frequency (MAF)
N-(3-acetamidopropyl)pyrrolidin-2-one) (acisoga)
Odds Ratio (OR)
Plasma membrane Ca ²⁺ ATPase 2 (PMCA2)
Principal component analysis (PCA)
Quartile-quartile (QQ)
Single-nucleotide polymorphisms (SNPs)
Spectrin Repeat Containing Nuclear Envelope Protein 1 gene (<i>SYNE1</i>)
STrengthening the REporting of Genetic Association studies (STREGA)
Ultra-performance liquid chromatography (UPLC)

Declarations

Ethics approval and consent to participate: This study was performed in accordance with the World Medical Association Declaration of Helsinki. All protocols were approved by the Institutional Research Board of anti-doping lab Qatar (F2014000009) and participants have given consent to participate.

Consent for publication: Not applicable.

Availability of data and materials: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests: The authors declare no competing interests.

Funding: This study was funded by Qatar National Research Fund (QNRF), Grant number NPRP7-272-1-041 (MAE, KS, CG and FB). The funding body had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions: FK, NAY, OA, EAS, ESK, NAK, OVB, LBA, AKL, EVG, DA, AH, FD, FB, CG, KS, IA, MAE collected samples, carried out analysis, wrote the paper, reviewed and accepted its final version. MAE (corresponding) is responsible for the integrity of the work as a whole.

Acknowledgements: Authors would like to thank Qatar National Research Fund (QNRF) for funding this project. Grant number NPRP7-272-1-041 (MAE, KS, CG and FB).

References

1. Georgiades E, Klissouras V, Baulch J, Wang G, Pitsiladis Y. Why nature prevails over nurture in the making of the elite athlete. *BMC genomics*. 2017;18(Suppl 8):835.
2. Bergh U, Ekblom B, Astrand PO. Maximal oxygen uptake "classical" versus "contemporary" viewpoints. *Medicine and science in sports and exercise*. 2000;32(1):85-8.
3. Mitchell JH, Haskell W, Snell P, Van Camp SP. Task Force 8: classification of sports. *Journal of the American College of Cardiology*. 2005;45(8):1364-7.
4. Peeters MW, Thomis MA, Beunen GP, Malina RM. Genetics and sports: an overview of the pre-molecular biology era. *Medicine and sport science*. 2009;54:28-42.
5. Bouchard C, Daw EW, Rice T, Perusse L, Gagnon J, Province MA, et al. Familial resemblance for VO₂max in the sedentary state: the HERITAGE family study. *Medicine and science in sports and exercise*. 1998;30(2):252-8.
6. Bouchard C, Sarzynski MA, Rice TK, Kraus WE, Church TS, Sung YJ, et al. Genomic predictors of the maximal O₂ uptake response to standardized exercise training programs. *Journal of applied physiology*. 2011;110(5):1160-70.
7. Rankinen T, Roth SM, Bray MS, Loos R, Perusse L, Wolfarth B, et al. Advances in exercise, fitness, and performance genomics. *Medicine and science in sports and exercise*. 2010;42(5):835-46.
8. Ahmetov I, Kulemin N, Popov D, Naumov V, Akimov E, Bravy Y, et al. Genome-wide association study identifies three novel genetic markers associated with elite endurance performance. *Biology of sport*. 2015;32(1):3-9.
9. Ahmetov, II, Egorova ES, Gabdrakhmanova LJ, Fedotovskaya ON. Genes and Athletic Performance: An Update. *Medicine and sport science*. 2016;61:41-54.
10. Pitsiladis YP, Tanaka M, Eynon N, Bouchard C, North KN, Williams AG, et al. Athlome Project Consortium: a concerted effort to discover genomic and other "omic" markers of athletic performance. *Physiological genomics*. 2016;48(3):183-90.
11. Rankinen T, Fuku N, Wolfarth B, Wang G, Sarzynski MA, Alexeev DG, et al. No Evidence of a Common DNA Variant Profile Specific to World Class Endurance Athletes. *PloS one*. 2016;11(1):e0147330.
12. Kastenmuller G, Raffler J, Gieger C, Suhre K. Genetics of human metabolism: an update. *Human molecular genetics*. 2015;24(R1):R93-R101.
13. Tanaka M, Wang G, Pitsiladis YP. Advancing sports and exercise genomics: moving from hypothesis-driven single study approaches to large multi-omics collaborative science. *Physiological genomics*. 2016;48(3):173-4.
14. Al-Khelaifi F, Diboun I, Donati F, Botre F, Alsayrafi M, Georgakopoulos C, et al. A pilot study comparing the metabolic profiles of elite-level athletes from different sporting disciplines. *Sports medicine - open*. 2018;4(1):2.
15. Al-Khelaifi F, Donati F, Botre F, Latiff A, Abraham D, Hingorani A, et al. Metabolic profiling of elite athletes with different cardiovascular demand. *Scand J Med Sci Sports*. 2019.
16. Lortie G, Bouchard C, Leblanc C, Tremblay A, Simoneau JA, Theriault G, et al. Familial similarity in aerobic power. *Human biology*. 1982;54(4):801-12.
17. Bouchard C, Tremblay A, Despres JP, Theriault G, Nadeau A, Lupien PJ, et al. The response to exercise with constant energy intake in identical twins. *Obes Res*. 1994;2(5):400-10.
18. Hamel P, Simoneau JA, Lortie G, Boulay MR, Bouchard C. Heredity and muscle adaptation to endurance training. *Medicine and science in sports and exercise*. 1986;18(6):690-6.
19. Prud'homme D, Bouchard C, Leblanc C, Landry F, Fontaine E. Sensitivity of maximal aerobic power to training is genotype-dependent. *Medicine and science in sports and exercise*. 1984;16(5):489-93.
20. Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, et al. Familial aggregation of VO₂(max) response to exercise training: results from the HERITAGE Family Study. *Journal of applied physiology*. 1999;87(3):1003-8.
21. Guth LM, Roth SM. Genetic influence on athletic performance. *Curr Opin Pediatr*. 2013;25(6):653-8.
22. Wang G, Tanaka M, Eynon N, North KN, Williams AG, Collins M, et al. The Future of Genomic Research in Athletic Performance and Adaptation to Training. *Medicine and sport science*. 2016;61:55-67.
23. Ahmetov, II, Fedotovskaya ON. Current Progress in Sports Genomics. *Advances in clinical chemistry*. 2015;70:247-314.
24. Brini M, Cali T, Ottolini D, Carafoli E. The plasma membrane calcium pump in health and disease. *FEBS J*. 2013;280(21):5385-97.
25. Zhang Q, Ragnauth CD, Skepper JN, Worth NF, Warren DT, Roberts RG, et al. Nesprin-2 is a multi-isomeric protein that binds lamin and emerin at the nuclear envelope and forms a subcellular network in skeletal muscle. *J Cell Sci*. 2005;118(Pt 4):673-87.

26. Street VA, McKee-Johnson JW, Fonseca RC, Tempel BL, Noben-Trauth K. Mutations in a plasma membrane Ca²⁺-ATPase gene cause deafness in deafwaddler mice. *Nature genetics*. 1998;19(4):390-4.
27. Smits JJ, Oostrik J, Beynon AJ, Kant SG, de Koning Gans PAM, Rotteveel LJC, et al. De novo and inherited loss-of-function variants of ATP2B2 are associated with rapidly progressive hearing impairment. *Human genetics*. 2019;138(1):61-72.
28. Strehler EE, Zacharias DA. Role of alternative splicing in generating isoform diversity among plasma membrane calcium pumps. *Physiological reviews*. 2001;81(1):21-50.
29. Bergman BC, Brozinick JT, Strauss A, Bacon S, Kerege A, Bui HH, et al. Muscle sphingolipids during rest and exercise: a C18:0 signature for insulin resistance in humans. *Diabetologia*. 2016;59(4):785-98.
30. Borghouts LB, Keizer HA. Exercise and insulin sensitivity: a review. *International journal of sports medicine*. 2000;21(1):1-12.
31. In: Marriott BM, editor. *Food Components to Enhance Performance: An Evaluation of Potential Performance-Enhancing Food Components for Operational Rations*. Washington (DC) 1994.
32. Fucho R, Casals N, Serra D, Herrero L. Ceramides and mitochondrial fatty acid oxidation in obesity. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2017;31(4):1263-72.
33. Turchanowa L, Rogozkin VA, Milovic V, Feldkoren BI, Caspary WF, Stein J. Influence of physical exercise on polyamine synthesis in the rat skeletal muscle. *Eur J Clin Invest*. 2000;30(1):72-8.
34. Hodgson LE, Walter E, Venn RM, Galloway R, Pitsiladis Y, Sardat F, et al. Acute kidney injury associated with endurance events-is it a cause for concern? A systematic review. *BMJ Open Sport Exerc Med*. 2017;3(1):e000093.
35. Yoshinaga T, Nakamura K, Ishikawa M, Yamaguchi T, Takano K, Wakui K, et al. A novel frameshift mutation of SYNE1 in a Japanese family with autosomal recessive cerebellar ataxia type 8. *Hum Genome Var*. 2017;4:17052.
36. Sandra M, Maria Pia L, Stefano C, Pietro P, Crociani P, Aldo R, et al. Emery-Dreifuss muscular dystrophy type 4: A new SYNE1 mutation associated with hypertrophic cardiomyopathy masked by a perinatal distress-related spastic diplegia. *Clin Case Rep*. 2019;7(5):1078-82.
37. Izumi Y, Miyamoto R, Morino H, Yoshizawa A, Nishinaka K, Udaka F, et al. Cerebellar ataxia with SYNE1 mutation accompanying motor neuron disease. *Neurology*. 2013;80(6):600-1.
38. Fanin M, Savarese M, Nascimbeni AC, Di Fruscio G, Pastorello E, Tasca E, et al. Dominant muscular dystrophy with a novel SYNE1 gene mutation. *Muscle & nerve*. 2015;51(1):145-7.
39. Chen Z, Ren Z, Mei W, Ma Q, Shi Y, Zhang Y, et al. A novel SYNE1 gene mutation in a Chinese family of Emery-Dreifuss muscular dystrophy-like. *BMC medical genetics*. 2017;18(1):63.
40. Baumann M, Steichen-Gersdorf E, Krabichler B, Petersen BS, Weber U, Schmidt WM, et al. Homozygous SYNE1 mutation causes congenital onset of muscular weakness with distal arthrogryposis: a genotype-phenotype correlation. *European journal of human genetics : EJHG*. 2017;25(2):262-6.
41. Phillips GC. Glutamine: the nonessential amino acid for performance enhancement. *Current sports medicine reports*. 2007;6(4):265-8.
42. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Analytical chemistry*. 2009;81(16):6656-67.

Tables

Table 1. Classification of GWAS participants according to sports classes. Distribution of elite athletes in various categories based on sport type-associated peak dynamic (maximal oxygen uptake percentage; VO_{2max}) and peak static (maximal voluntary muscle contraction percentage; MVC) components achieved during competition as described previously (3).

	Low/moderate (< 70% VO _{2max})		High (>70% VO _{2max})			Total
High (>50% MVC)	Wrestling and Judo (6M)	Skate boarding (2M)	Modern Pentathlon (1F)			273
			Kayaking (1F)	Rowing (9M/8F)	Biathlon (2M/1F)	
	Weightlifting (8M/7F)		Boxing (3M/12F)	Cycling (149M/48F)	Triathlon (9M/7F)	
Moderate (20-50% MVC)	Jumping (athletics) (1F)		Handball (15M/3F)	Middle distance running (1M)	Basketball (2M)	156
	Rugby (15M)	Aquatics (8M/5F)				
	Athletics other (31M/29F)		Hockey (1F)	Skiing Cross Country (3M/1F)	Swimming (24M/17F)	
Low (<20% MVC)	Baseball (2M)		Long-Distance running and marathon (35M/12F)		Tennis (2M/3F)	324
	Volleyball (2M)					
	Table tennis (7M)		Soccer (243M)	Ultra-running (1F)	Football (16M/1F)	
	123		630			753

Table 2. Top GWAS SNPs associated with Endurance athlete status from the discovery study.

rsID	Chromosome	Position	Reference Base	N	OR	Standard Error	P value	Function GVS	Gene List	MAF-High aerobic	MAF-Moderate/Low aerobic	MAF-nonathletes
rs56330321	3	10684812	A	751	0.2501	0.2614	1.15E-07	intron	ATP2B2	0.03822	0.1179	A=0.055
rs2635438	6	152506470	G	753	0.1636	0.3512	2.54E-07	intron	SYNE1	0.01508	0.0813	C=0.06
rs225902	14	30459451	A	748	0.41	0.1893	2.48E-06	intergenic	none	0.1366	0.2459	A=0.18
rs146654270	9	123798492	A	749	0.1621	0.3917	3.38E-06	intron	C5	0.01276	0.05738	A=0.01655
rs1969772	1	59205102	A	752	0.4928	0.1538	4.23E-06	intergenic	none	0.2913	0.4426	A=0.26
rs77471963	15	69245158	G	746	0.3671	0.2182	4.38E-06	intron	NOX5	0.08106	0.1626	G=0.068
rs10011584	4	127213903	T	750	0.4572	0.1721	5.41E-06	intergenic	none	0.1467	0.2561	T=0.13
rs142155779	7	126252753	C	751	0.2247	0.3284	5.47E-06	intron	GRM8	0.02146	0.08197	C=0.015
rs7584904	2	106131036	T	750	0.4047	0.2003	6.30E-06	intergenic	none	0.0925	0.187	T=0.14
rs7599151	2	59308101	A	753	0.5097	0.1507	7.75E-06	intergenic	none	0.3619	0.4634	A=0.33

Table 3. Metabolites that belong to the significantly enriched phospholipids pathway Top metabolites associated with significant SNPs.

SNP	Metabolite	Beta	SE Beta	P value	Super-pathways	Enriched Sup-pathways	Enrichment p value
rs56330321	N-Stearoyl-Sphingosine (D18:1/18:0)*	0.33	0.14	0.01	Lipid	Ceramides	0.04
	Ceramide (D18:1/20:0, D16:1/22:0, D20:1/18:0)*	0.31	0.14	0.03			
	N-Palmitoyl-Sphingosine (D18:1/16:0)	0.29	0.13	0.03			
	Ceramide (D18:2/24:1, D18:1/24:2)*	0.28	0.14	0.04			
	Ceramide (D16:1/24:1, D18:1/22:1)*	0.26	0.13	0.05			
	Adrenoylcarnitine (C22:4)*	0.5	0.17	0		Fatty Acid Metabolism (Acyl Carnitine)	0.04
	Dihomo-Linolenoylcarnitine (20:3N3 Or 6)*	0.37	0.14	0.01			
	3-Hydroxybutyrylcarnitine (2)	0.36	0.14	0.01			
	3-Hydroxybutyrylcarnitine (1)	0.36	0.14	0.01			
	Palmitoylcarnitine (C16)	0.3	0.13	0.02			
	Acetylcarnitine (C2)	0.3	0.14	0.03			
	Oleoylcarnitine (C18:1)	0.28	0.13	0.04			
	Cerotoylcarnitine (C26)*	0.27	0.13	0.04			
	4-Acetamidobutanoate	0.38	0.14	0.01	Amino Acid	Polyamine Metabolism	0.03
	Acisoga	-0.46	0.2	0.02			
	N-Acetylputrescine	0.3	0.14	0.03		Creatine Metabolism	0.03
	Creatine	0.32	0.14	0.02			
	Creatinine	0.25	0.13	0.05			
rs2635438	Gamma-Glutamyl-2-Aminobutyrate	-0.58	0.21	0.006	Peptide	Gamma-glutamyl Amino Acid	0.00002
	Gamma-Glutamylhistidine	-0.5	0.19	0.01			
	Gamma-Glutamylglutamine	-0.53	0.21	0.012			
	Gamma-Glutamylalanine	-0.54	0.22	0.012			
	Gamma-Glutamylthreonine	-0.52	0.21	0.015			
	Gamma-Glutamylmethionine	-0.44	0.22	0.042			
	Gamma-Glutamyl-Alpha-Lysine	-0.42	0.21	0.045			
	N-Acetyl-Aspartyl-Glutamate (NAAG)	-0.68	0.22	0.002	Amino Acid	Glutamate Metabolism	0.0002
	Glutamate	0.43	0.18	0.021			
	Glutamine	-0.47	0.21	0.026			
	Beta-Citrylglutamate	0.45	0.22	0.041			
	Carboxyethyl-GABA	0.41	0.21	0.047			

Figures

Figure 1

PCA shows no difference in the genotype distribution among sport disciplines (A) or between groups (sports with low/moderate vs high aerobic component) (B).

Figure 2

Manhattan (A) and quantile-quantile (B) plots illustrating GWAS results in association with endurance (red line indicates the Bonferroni level of significance, $P=1.46E-7$).

Figure 3

Boxplots representing levels of metabolites affiliated with enriched metabolic pathways in association with the rs56330321 genotype groups.