Article

A Genome-Wide Association Study of Sprint Performance in Elite Youth Football Players

Pickering, Craig, Suraci, Bruce, Semenova, Ekaterina A, Boulygina, Eugenia A, Kostryukova, Elena S, Kulemin, Nickolay A, Borisov, Oleg V, Khabibova, Sofya A, Larin, Andrey K, Pavlenko, Alexander V, Lyubaeva, Ekaterina V, Popov, Daniil V, Lysenko, Evgeny A, Vepkhvadze, Tatiana F, Lednev, Egor M, Leońska-Duniec, Agata, Pająk, Beata, Chycki, Jakub, Moska, Waldemar, Lulińska-Kuklik, Ewelina, Dornowski, Marcin, Maszczyk, Adam, Bradley, Ben, Kana-Ah, Adam, Cięszczyk, Pawel, Generozov, Edward V and Ahmetov, Ildus I

Available at http://clok.uclan.ac.uk/29419/

It is advisable to refer to the publisher's version if you intend to cite from the work.
http://dx.doi.org/10.1519/JSC.0000000000003259

For more information about UCLan's research in this area go to
http://www.uclan.ac.uk/researchgroups/ and search for <name of research Group>.

For information about Research generally at UCLan please go to
http://www.uclan.ac.uk/research/

A genome-wide association study of sprint performance in elite youth football players

Running title: Genes and sprint performance

Craig Pickering[1], Eri Miyamoto-Mikami[2], Bruce Suraci[3], Ekaterina A. Semenova[4,5], Eugenia A. Boulygina[6], Elena S. Kostryukova[4], Hirofumi Zempo[7], Oleg V. Borisov[4,8], Daniil V. Popov[9], Agata Leońska-Duniec[10], Ben Bradley[11], Jakub Chycki[12], Marcin Dornowski[13], Noriko Ichinoseki-Sekine[14], Ryo Kakigi[15], Adam Kana-ah[11], Sofya A. Khabibova[4], Hiroyuki Kobayashi[16], Nickolay A. Kulemin[4], Hiroshi Kumagai[2,17], Andrey K. Larin[4], Egor M. Lednev[9], Ewelina Lulińska-Kuklik[18], Evgeny A. Lysenko[9], Ekaterina V. Lyubaeva[9], Adam Maszczyk[19], Waldemar Moska[18], Hisashi Naito[2], Beata Pająk[13], Alexander V. Pavlenko[4], Keisuke Shiose[20],Takuro Tobina[21],Takamasa Tsuzuki[22], Tatiana F. Vepkhvadze[9], Pawel Cięszczyk[19], Edward V. Generozov[4], Noriyuki Fuku[2], Ildus I. Ahmetov[4,23,24,25]

[1]School of Sport and Wellbeing, Institute of Coaching and Performance, University of Central Lancashire, Preston, United Kingdom

[2]Graduate School of Health and Sports Science, Juntendo University, Chiba, Japan

[3]Academy Coaching Department, AFC Bournemouth, Bournemouth, United Kingdom

[4]Department of Molecular Biology and Genetics, Federal Research and Clinical Center of Physical-Chemical Medicine, Moscow, Russia

[5]Department of Biochemistry, Kazan Federal University, Kazan, Russia

[6]Omics Technologies OpenLab, Kazan Federal University, Kazan, Russia

[7]Faculty of Health and Nutrition, Tokyo Seiei College, Tokyo, Japan

[8]Moscow Institute of Physics and Technology (State University), Moscow, Russia

[9]Laboratory of Exercise Physiology, Institute of Biomedical Problems of the Russian Academy of Sciences, Moscow, Russia

[10]Faculty of Physical Education, Gdansk University of Physical Education and Sport, Gdansk, Poland

[11]Academy Sports Science Department, AFC Bournemouth, Bournemouth, United Kingdom

[12]Department of Sports Training; Academy of Physical Education in Katowice, Poland

[13]Independent Laboratory of Genetics and Molecular Biology, Kaczkowski Military Institute of Hygiene Epidemiology, Poland

[14]Faculty of Liberal Arts, The Open University of Japan, Chiba, Japan

[15]Faculty of Medicine, Juntendo University, Tokyo, Japan

[16]Department of General Medicine, Mito Medical Center, Tsukuba University Hospital, Ibaraki, Japan

[17]Research Fellow of Japanese Society for the Promotion of Science, Tokyo, Japan

[18]Faculty of Tourism and Recreation, Gdansk University of Physical Education and Sport, Gdansk, Poland

[19]Department of Theory and Practice of Sport; Academy of Physical Education in Katowice; Poland

[20]Faculty of Sports and Health Science, Fukuoka University, Fukuoka, Japan

[21]Faculty of Nursing and Nutrition, University of Nagasaki, Nagasaki, Japan

[22]Faculty of Pharmacy, Meijo University, Aichi, Japan

[23]Laboratory of Molecular Genetics, Kazan State Medical University, Kazan, Russia

[24]Sports Genetics Laboratory, St Petersburg Research Institute of Physical Culture, St Petersburg, Russia

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

ABSTRACT

Sprint speed is an important component of football performance, with teams often placing a high value in sprint and acceleration ability. The aim of this study was to undertake the first genome-wide association study to identify genetic variants associated with sprint test performance in elite youth football players and to further validate obtained results in additional studies. Using micro-array data (600 K − 1.14 M single nucleotide polymorphisms; SNPs) of 1206 subjects, we identified 12 SNPs with suggestive significance after passing replication criteria. The polymorphism rs55743914 located in the *PTPRK* gene, was found as the most significant for 5 m sprint test ($P = 7.7$ x 10^{-7}). Seven of the discovered SNPs were also associated with sprint test performance in a cohort of 126 Polish females, and four were associated with power athlete status in a cohort of 399 elite Russian athletes. Six SNPs were associated with muscle fibre type in a cohort of 96 Russian subjects. We also examined genotype distributions and possible associations for 16 SNPs previously linked with sprint performance. Four SNPs (*AGT* rs699, *HSD17B14* rs7247312, *IGF2* rs680 and *IL6* rs1800795) were associated with sprint test performance in this cohort. Additionally, the G alleles of two SNPs in *ADRB2* (rs1042713 & rs1042714) were significantly over-represented in these players compared to British and European controls. These results suggest that there is a genetic influence on sprint test performance in footballers, and identifies some of the genetic variants that help explain this influence.

Key words: football, sprinters, training, muscle fibres, polymorphism

INTRODUCTION

Like many field-based team-sport games, association football is an intermittent sport, with match-play consisting of periods of walking interspersed with both low- and high-intensity running (4,25,26). It has previously been reported that the volume of high-speed running within a match is an important distinguisher between standards of player, with better players typically undertaking a higher proportion of high-intensity and sprint running (26). Indeed, in analysis of match play, straight line sprinting was the most frequent action carried out by players both prior to scoring or assisting a goal (10). Accordingly, football teams place a high value on sprint and acceleration ability, and, as a result, players have become faster over time (16).

A number of genetic variants have been associated with elite speed-power athlete status (2). Perhaps the most well-known of these occurs within *ACTN3*, which encodes for a protein found exclusively within type-II muscle fibres. A common polymorphism within this gene (R577X; rs1815739) results in a premature stop codon, denoted by the X allele. Individuals homozygous for the X allele (i.e. XX genotypes) cannot produce ACTN3. Such an outcome appears to be disadvantageous for speed-power performance (28), as demonstrated by the relative lack of XX genotypes reported in elite speed-power athletes (41). Other polymorphisms tentatively associated with elite speed-power athlete status include those within *ACE* (29)*, AGT* (14)*, AGTR2* (43) and others (reviewed in [2]).

As a result of this early research, there has been an increased interest in understanding the genetic underpinnings of elite football performance. For example, Santiago and colleagues (34) reported that the frequency of the *ACTN3* RR genotype was significantly higher in a cohort of elite players when compared to controls and elite endurance athletes, suggesting that elite

football players are potentially predisposed to possess speed-power genotypes, although this would be expected to vary with position and playing style. Furthermore, this same polymorphism has been explored with regards to the response to eccentric training, and as a potential modifier of injury risk in professional players (23,30). *ACTN3* is far from the only genetic variant to be explored with regards to football, with many other polymorphisms, such as those found in *ACE, PPARA,* and *PPARGC1A* also studied (13). The identification of associations between genetic polymorphisms and football performance has led to the development of multi-gene models to explain variations in performance. For example, Egorova and colleagues (8) utilised a Total Genotype Score (TGS), finding that football players scored significantly higher when compared to controls; such data suggests that there is a genetic underpinning to the achievement of elite status in football players.

Genome-wide association studies (GWAS) are a new approach in sport science that involves rapidly scanning several hundred thousand markers across the complete set of DNA of athletes and untrained subjects to find genetic variations associated with exercise and sport-related traits (1,31). 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.

Given that there is the potential for football performance to be partially determined by the genotype of a variety of gene polymorphisms, the aim of this work was to conduct a genome-wide association study of sprint performance in a cohort of youth academy football players, and to further validate obtained results in additional studies. We also sought to explore the relationship between 16 single nucleotide polymorphisms (SNPs), selected from the literature

and thought to affect speed-power performance, and sprint performance. Additionally, we hypothesized that alleles associated with better sprint times in football players should be over-represented in elite speed-power power athletes compared to controls or endurance athletes, associated with enhanced sprint performance in non-athletes, and associated with an increased proportion of fast-twitch muscle fibres. Accordingly, we sought to test these hypotheses across multiple cohorts.

METHODS

Experimental Approach to the Problem

This study utilized two different genetic assessment protocols, specifically a hypothesis-free, genome-wide association study (GWAS), and candidate gene analysis (design and results shown in Figure 1). The GWAS approach was selected as it represents a hypothesis free method of determining novel genetic associations, whilst the candidate gene analyse was employed in order to test associations with previously established genetic markers.

52 British soccer players (detailed below; "British Cohort") were recruited to this study. Following initial analysis, 4 players were removed due to their being of non-Caucasian ethnicity, confirmed using principal component analysis of genomic data. The remaining 48 Caucasian players underwent both GWAS and candidate gene analysis.

For the GWAS, we tested > 600 K SNPs for association with 5 and 20 m sprints in Caucasian British players. Next, we validated the obtained results in additional studies involving a separate cohort of Polish young women (5, 10 and 30 m sprints), and a cohort of

high-level Russian athletes in a case-control manner (power athletes vs controls / endurance athletes). The functionality of significant SNPs was further investigated in a cohort of Russian subjects who had undergone muscle biopsies. Finally, the suggestively significant SNPs from the GWAS were then compared to a sample of non-athletic Caucasian controls to test for associations between these SNPs and football player status.

For the candidate gene association study, genotype distributions for 16 gene polymorphisms (*ACE* rs4341 (alternative of Alu I/D), *ACTN3* rs1815739, *ADRB2* rs1042713, *ADRB2* rs1042714, *AGT* rs699, *AMPD1* rs17602729, *CKM* rs8111989, *GABRR1* rs282114, *HSD17B14* rs7247312, *IGF1* rs35767, *IGF2* rs680, *IL6* rs1800795, *MTHFR* rs1801131, *PPARA* rs4253778, *PPARG* rs1801282, *UCP2* rs660339), previously associated with power performance, were selected from the literature (Ahmetov et al., 2016). Associations between these genetic variants and sprint performance in football players were then calculated. Furthermore, allelic frequencies were compared for these SNPs between the football players and non-athletic Caucasian controls to determine associations with football player status.

Study participants

Ethical approval

The study was approved by the Ethics Committee of the University of Portsmouth, Ethics Committee of the Physiological Section of the Russian National Committee for Biological Ethics, Ethics Committee of the Federal Research and Clinical Center of Physical-chemical Medicine of the Federal Medical and Biological Agency of Russia, Ethics Committee of the Regional Medical Chamber in Szczecin. Written informed consent was obtained from each participant; where subjects were under the age of 18, parental or guardian informed consent was given. 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.

British cohort

The football study involved 52 elite youth (age 12-18) football players. After performing principal component analysis of microarray data, 48 athletes of only Caucasian origin remained. Each subject was currently signed to a Premier League youth contract as part of a Premier League football club's youth academy, under the governance of the Elite Player Performance Plan as defined by the Premier League. Publicly available genomic data of British (n = 81) and European (n = 364) populations from the 1000 Genomes Project was used to identify markers for football player status.

Russian cohorts

The Russian athletes study involved 176 Elite power (89 sprinters, 38 speed/strength athletes and 49 strength athletes) athletes, 223 endurance athletes and 173 controls. All athletes were Olympic team members who have tested negative for doping substances. The Russian muscle fibre study involved 96 physically active healthy Russian subjects (29 females, 67 males).

Polish cohort

One hundred and twenty-six healthy Polish (all Caucasian) young women (mean age 21 ± 1 years) completed sprint tests before and after aerobic training. None of these individuals had engaged in regular physical activity in the previous 6 months.

Genotyping

British and Polish studies

Upon enrollment into the studies, each participant volunteered a saliva sample, which was collected through sterile and self-administered buccal swabs. The samples were sent to AKESOgen, Inc (Peachtree Corners, GA, USA) within 36 hours, where DNA was extracted from the saliva samples using Qiagen chemistry on an automated Kingfisher FLEX instrument (Thermo Fisher Scientific, Waltham, MA, US), following the manufacturer's recommended protocols and standard operating procedures. PicoGreen and Nanodrop measurements are taken to measure the quality and quantity of the DNA. Input to the custom testing array occurs at 200 ng in 20 µL. Amplification, fragmentation, and resuspension was performed using Biomek FXP following Affymetrix's high throughput protocol for Axiom 2.0. Hybridization was performed for 24 hours at 48°C in a Binder oven, and staining and scanning of the arrays (> 600,000 SNPs; DNAFit's custom microchips) was performed using GeneTitan instrumentation (Thermo Fisher Scientific, Waltham, MA, US), all following the same Affymetrix high throughput Axiom 2.0 protocol. Data analysis was then performed using a raw CEL file data input into the Affymetrix Axiom Analysis Suite (Affymetrix, Santa Clara, CA, US).

Russian studies

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 athletes and controls and HumanOmniExpress BeadChips (Illumina Inc, USA) were used for genotyping of > 700,000 SNPs in participants of the Russian Muscle Fibre Study. 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.

Evaluation of muscle fibre composition

Samples of the vastus lateralis muscle of Russian athletes were obtained with the Bergström needle biopsy procedure under local anesthesia with 1% lidocaine solution. After this procedure, serial cross-sections (7 µm) were obtained from frozen samples using a microtome (Leica Microsystems, Wetzlar, Germany). The sections were thaw-mounted on Polysine glass slides, kept for 15 min at RT and incubated with in PBS (3 × 5 min). Then the sections were incubated at RT in primary antibodies against slow or fast isoform of the myosin heavy chains (M8421, 1:5000; M4276; 1:600, respectively; Sigma-Aldrich, USA) for 1 h and incubated in PBS(3 × 5 min). After this, the sections were incubated at RT in secondary antibodies conjugated with FITC (F0257; 1:100; Sigma-Aldrich) for 1 h. The antibodies were removed and the sections were washed in PBS (3 × 5 min), placed in mounting media and covered with a cover slip. The image was captured using a fluorescent microscope Eclipse Ti-U (Nikon, Japan). All analyzed image contained > 100 fibers. The ratio of the number of stained fibers to the total fiber number was calculated. Fibers stained in serial sections with antibodies against slow and fast isoforms were considered as hybrid fibers. The cross-sectional area of fast and slow fibers was evaluated using the Image J (NIH, USA).

Performance tests

5, 10, 20 and 30 m sprint

The 5/10/20/30-meter sprint test has been shown to be an accurate method to estimate acceleration (5 & 10 m) and speed (20 & 30 m) capabilities in invasion-type sports such as association football in adult males (42). Light gates were set up as per the manufacturer's instructions, and placed 5 meters apart. Participants started approximately 0.5 meters behind the light gate, and began the sprint at their own convenience. Football players completed 5 and 20 m sprints, with 3 minutes rest between each sprint, while Polish women completed 5, 10 and 30 m sprints. All attempts were recorded, with the highest score utilised for testing data.

Statistical analysis

Statistical analyses were conducted using PLINK v1.90, R (version 3.4.3), and GraphPad InStat (GraphPad Software, Inc., USA) software. All SNPs that did not pass the Hardy-Weinberg equilibrium test, genotyping rate and minor allele frequency (MAF) thresholds (options: MAF 0.05, genotyping rate 0.05, HWE 0.00001 in PLINK) were excluded from further analysis. Genotype distribution and allele frequencies between athletes and controls were compared using χ^2 tests. Quantitative trait (sprint tests) SNP association was tested in a linear additive model adjusted for age. Differences in phenotype between different genotype groups were analysed using ANOVA. Bonferroni's correction for multiple testing was performed by multiplying the *P* value with the number of tests where appropriate. Within the GWAS aspect of this study, we set our statistical threshold for further exploration as $p = 1.5 \times 10^{-5}$. This is below the standard GWAS threshold of $p < 5 \times 10^{-8}$; however, this approach has been shown to overcome the problem of "missing heritability", decrease the probability of false-negative results, and enhance the predictive ability of genotype scores in disease research. As

such, it represents a potentially worthwhile method to uncover novel genotype-phenotype associations. SNPs meeting this suggestive threshold were then assessed via additional criteria, including previous replication. In order to guard against false-positive results, additional criteria were applied to all significant SNPs, including at least one replication for suggestive SNPs, and at least two replications for less significant SNPs (i.e. those with p-values ranging from 0.049 to 1.51×10^{-5}).

RESULTS

Genome-wide association study and validation in the Caucasian cohorts

Following genome-wide association testing, 3029 SNPs associated ($P<0.05$) with both 5 and 20 m sprint times adjusted for age were identified; these and other results are summarized in figure 1. Of those, 13 SNPs reached suggestive significance level (P value from 1.5×10^{-5} to 3.1×10^{-7}), and four passed our previously outlined criteria. The polymorphism rs55743914 located in the *PTPRK* gene, was the most significant for the 5 m sprint test ($P = 7.7 \times 10^{-7}$). *SEMA4A* rs12401573 ($P = 7.8 \times 10^{-5}$), *NFATC2* rs4811192 ($P = 3.7 \times 10^{-5}$), and *TERT* rs33954691 ($P = 6.6 \times 10^{-5}$) were the most significant for the 20 m sprint test (Table 1).

Furthermore, of less significant SNPs (i.e. those with p-values ranging from 0.049 to 1.5×10^{-5}), eight passed our criteria, as described above. Overall, 12 SNPs remained (table 1), of which 7 genetic variants were also associated with sprint test performance in a cohort of Polish women (table 2), 4 polymorphisms with power athlete status in Russians (table 3), and 6 markers with proportion of fast-twitch muscle fibres in the Russian cohort (table 4).

INSERT FIGURE 1 AROUND HERE

INSERT TABLE 1 AROUND HERE

Table 2 displays seven SNPs of the top twelve that were associated with sprint test performance in a cohort of Polish females, either at baseline (rs55743914, rs17347590,rs3213537), post-training (rs4811192, rs62247016) or in terms of magnitude of improvements in sprint performance (rs33954691, rs8064257, rs62247016).

INSERT TABLE 2 AROUND HERE

In the Russian power athlete study, of the 12 top SNPs, four polymorphisms demonstrated a significant association with power athlete status within this cohort (Table 3). Specifically, rs12688220 G and rs3213537 G alleles were over-represented in sprinters compared to controls or endurance athletes. The frequency of the rs8064257 G allele was significantly higher in sprinters compared to endurance athletes. Furthermore, the rs1929877 G allele was significantly over-represented in elite Russian speed-strength athletes in comparison to controls.

INSERT TABLE 3 AROUND HERE

Finally, we tested for associations between these 12 SNPs and football player status. Upon analysis, we determined that the A allele of rs62247016 in *CNTN4* was significantly over-represented in football players (90.6%) compared to both British (81.5%, $P = 0.047$) and European (80.9%, $P = 0.019$) control populations.

Associations between six SNPs and muscle fibre type in males, females or the combined group are detailed in table 4. As mentioned above, sprint alleles discovered in the football study were associated with an increased proportion of fast-twitch muscle fibres in the Russian group of physically active subjects. The top SNP was the *CBLN2* gene rs8093502 polymorphism which was associated with the percentage of fast-twitch muscle fibres both in males ($P = 0.041$) and females ($P = 0.042$), as well as in the combined group ($P = 0.006$).

INSERT TABLE 4 AROUND HERE

Candidate gene association study

Genotype distributions of 16 gene polymorphisms (*ACE* rs4341 (alternative of I/D), *ACTN3* rs1815739, *ADRB2* rs1042713, *ADRB2* rs1042714, *AGT* rs699, *AMPD1* rs17602729, *CKM* rs8111989, *GABRR1* rs282114, *HSD17B14* rs7247312, *IGF1* rs35767, *IGF2* rs680, *IL6* rs1800795, *MTHFR* rs1801131, *PPARA* rs4253778, *PPARG* rs1801282, *UCP2* rs660339), selected from the literature and previously reported to be associated with power performance, were in Hardy-Weinberg equilibrium amongst all athletes. To assess the association between each polymorphism and performance parameters we used ANOVA. After Bonferroni's correction for multiple testing the results were considered significant with $P < 0.0031$ (i.e. 0.05/16). In accordance with the hypothesis, we found that athletes with the *AGT* Thr/Thr (rs699 CC) genotype demonstrated better results in both 5 m sprint test ($P = 0.0011$) and 20 m sprint tests ($P < 0.0001$) compared with the opposite genotype carriers. Furthermore, the *IL6* rs1800795 G allele was associated with faster sprint times ($P = 0.0031$ for a 5 m sprint test and $P < 0.0001$ for a 20 m sprint test). The *ADRB2* rs1042713 G ($P = 0.0124$), *HSD17B14*

rs7247312 G (P = 0.0268) and *IGF2* rs680 G (P = 0.0291) alleles were also associated with faster 5 m sprint test results, although this was not significant following Bonferroni correction.

Case-control analysis (football players vs British or European populations) identified an over-representation of *ADRB2* rs1042713 G (74.0 vs 62.2%, P = 0.0247 in comparison with European population; 74.0 vs 58.1%, P = 0.0099 compared to British controls) and *ADRB2* rs1042714 G (60.4 vs 40.8, P = 0.00026 in comparison with European population; 60.4 vs 41.4%, P = 0.0031 compared to British controls) alleles in players, in accordance with the hypothesis. No other significant associations were found.

DISCUSSION

To our knowledge, this is the first genome-wide association study for sprint performance. Previously, two GWASes were performed in relation to endurance performance (1, 31). In the current GWAS, in a cohort of youth footballers and after further validation steps, we identified 12 SNPs associated with sprint test performance of football players (table 1). Of those, 7 genetic variants were associated with sprint performance of Polish women (table 2), 4 polymorphisms with power athlete status in Russians (table 3), and 6 markers with proportion of fast-twitch muscle fibres in the Russian subjects (table 4).

These SNPs are located in genes that have multiple functions that may predispose to enhanced sprint performance. Four SNPs (*CPNE5* rs3213537 [influences the expression of *RP1-90K10.5* in the blood], *SEMA4A* rs12401573 [T allele increases the expression of *SEMA4A* in thyroid tissue], *RDH13* rs4806637 [G allele increases the expression of *RDH13* in skeletal muscles] and near *MORC4* rs12688220 [T allele increases *RNF128* expression in nerves; C allele increases the expression of *CLDN2* in thyroid tissue]) are functional and influence

expression of genes in various tissues, including skeletal muscles, nerves, blood and thyroid tissue – all important in terms of skeletal muscle fibre composition (11,15). Indeed, all of these SNPs were associated with the proportion of fast-twitch muscle fibres in our study.

One SNP, rs12688220 near *MORC4,* showed consistent associations across all cohorts in terms of sprint performance, elite sprint athlete status, and increased proportion of fast twitch muscle fibres in Russian subjects. This SNP has previously been associated with pancreatitis, with the G (sprint) allele protective against the development of both acute and chronic pancreatitis (22,39). This locus is implicated in the control of ion and water channels between endothelial cells (3), and expression appears to increase following both injury and stress (39). At present, the mechanism of its effects on sprint phenotype is poorly understood.

In our study, the G allele of the *CPNE5* (rs3213537) was associated with better 5 m (P=0.012) and 20 m (P=0.0075) sprint times in British football players, better 10 m sprint time in young Polish women (P=0.011), and over-represented in Russian sprinters compared to controls (P=0.0346) and endurance athletes (P=0.0045). This gene encodes a calcium-dependent protein expressed in both neural progenitor cells and the differentiated neurons during the neural development and is implicated in synaptic function, thus playing an important role in the development of central nervous system (7). Additionally, the G allele of rs4806637, located in the 3'-UTR region of the gene *RDH13,* was associated with an increased proportion of fast-twitch muscle fibres. These gene encodes a retinol dehydrogenase 13, which provides a protective role in the mitochondria against oxidative stress.

Of the other genetic variants associated with sprint performance in this study, the potential mechanisms are varied. *NFATC2* (rs4811192) forms part of the calcineurin-NFAT

pathway, which has been implicated hypertrophy of both cardiac (27) and skeletal (36) hypertrophy – factors which are positively correlated with sprinting performance. *TERT* (rs33954691) encodes for telomerase reverse transcriptase, which acts to maintain telomere end length, likely protecting against exercise-induced DNA damage (24). rs12401573 is located in the *SEMA4A* gene (encoding semaphorin 4A), and appears to play a role in immune function (21), potentially mediating the post-exercise immune response. rs8093502 is located in the *CBLN2* gene (encoding cerebellin 2 precursor), which plays a role in synapse formation (32). CBLN1-null mice are severely ataxic, walk with an irregular gait, and do not maintain their balance (18). rs62247016 is located in the *CNTN4* gene (encodes contactin 4), which plays a role in the formation of axon connections in the developing nervous system (12).

The main finding from the candidate gene analysis portion of this study was that there was a significant association between *AGT* (rs699) genotype and both 5- and 20m sprint test performance within academy footballers, with the CC genotype associated with faster sprint times. Additionally, we also found that the G allele of *IL6* (rs1800795), the G allele of *HSD17B14* (rs7247312) and the G allele of *IGF2* (rs680) were significantly associated with 5- and 20m sprint test performance. Furthermore, using a case-control design, we identified two SNPs within *ADRB2* as markers associated with football player status.

AGT encodes for angiotensinogen, a protein involved in the renin-angiotensin-aldosterone system (RAAS). The C allele for a common polymorphism in this gene, Met235Thr (rs699) results from a threonine-to-methionine substitution, and is associated with increased levels of AGT. The RAAS has been implicated with elite athlete status, with polymorphisms in its constituent genes, such as *ACE* and *AGT*, associated with elite power athlete status (14,29). The proposed mechanism is that these genetic variants predispose to increased levels of

angiotensin-II, which acts as a skeletal muscle growth factor (19). The increased levels of AGT seen in individuals with the C allele of the rs699 *AGT* polymorphism would therefore predispose to increased angiotensin-II concentrations; hence the association with both elite speed-power athlete status, and enhanced sprint testing scores in our cohort. Previously, Gomez-Gallego and colleagues (14) reported that the CC genotype of this SNP was over-represented in a cohort of elite power athletes, suggesting that the C allele may favor speed-power performance. Similar results have been reported in Polish speed-power athletes (45), as well as greater improvements in explosive strength following training (44). Our results further support these earlier findings.

IL6 encodes for interleukin-6, a multifunctional cytokine that is involved in immune function, muscle repair and hypertrophy. A common polymorphism within this gene, G-174C (rs1800795), has been associated with muscle damage following eccentric loading, with the C allele associated with greater increases in creatine kinase (CK) (40). As a result, C allele carriers are expected to have increased muscle damage following resistance training, and, as such, are proposed to be under-represented in elite cohorts that require such training. Such a hypothesis is supported by experimental data, with Ruiz and colleagues (33) reporting the C allele as significantly under-represented in a cohort of elite Spanish power athletes. These results were replicated in a Polish cohort (9). Our findings further strengthen the proposed influence of this polymorphism on speed-power phenotype.

IGF2 has growth-regulating, insulin-like and mitogenic activities and affects muscle mass and muscle function. The rs680 G allele is associated with significantly higher levels of *IGF2* mRNA compared with the A allele, suggesting a role for this polymorphism in *IGF2* transcription (37). Previous studies have demonstrated that the G allele is protective against the

loss of muscle strength directly after exertional muscle damage (6), greater back muscle strength (17) and over-represented in top-level Israeli sprinters and jumpers (5) and elite Japanese judo athletes (17). *HSD17B14* encodes Hydroxysteroid 17-Beta Dehydrogenase 14, which is involved in the metabolism of steroids, and the G allele of rs7247312 has been annotated to the group of the markers associated with power athlete status (2).

Finally, we found that *ADRB2* rs1042713 G and rs1042714 G alleles are over-represented in football players compared to British or European controls. These alleles have been previously associated with power athlete status in Polish population (35), consistent with our results. *ADRB2* encodes the β-2 adrenergic receptor, which is a member of the G protein-coupled receptor superfamily, expressed in many cell types throughout the body and playing a pivotal role in the regulation of the cardiac, pulmonary, vascular, endocrine and central nervous systems.

Given the importance of sprint speed within football (16,26), these results have the potential to assist football clubs in the identification of athletes with the potential to possess such a physiological trait. However, at present, the scientific consensus regarding the use of genetic information for talent identification purposes is that it lacks evidence, and is ethically troubling (38). Furthermore, the incorrect identification of individuals as future elite players, or vice versa, could be costly both economically and reputationally. However, there is the potential that, as the evidence base grows and evolves, such information might be used alongside other, more traditional talent identification tools. Additionally, genetic information may be used to identify those athletes with an increased predisposition to show a greater magnitude of improvements following a certain type of training; in this case, players could be matched to

their optimal training type. Early research suggests such an approach could be efficacious (20), although, again, more research is required.

As with any study, there are limitations to our findings that must be considered. Our initial sample size of 48 is modest, and we recognize the requirement for increased cohort sizes to enhance the findings within the sporting genomics sphere. However, whilst modest, our use of additional cohorts of athletes and controls and methodological approaches (including functional analysis) reduced the probability of false-positive findings. Further studies should aim to replicate these findings in athletes of different ethnicities, to better understand how wide-reaching these findings may be.

In summary, after undertaking a genome-wide analysis study on a cohort of youth football players, we identified 12 SNPs associated with sprint speed. Of these 12, 7 polymorphisms were associated with sprint performance of Polish women, 4 polymorphisms with power athlete status in the Russian cohort and 6 markers with proportion of fast-twitch muscle fibres. Our results also suggest that the C allele of *AGT* (rs699), the G allele of *IGF2* (rs680), the G allele of *HSD17B14* (rs7247312) and the G allele of *IL6* (rs1800795) are associated with improved sprint performance in a group of elite youth football players. These candidate gene-association findings replicate those of earlier studies. Finally, we found that two SNPs within *ADRB2* were over-represented in this group of elite football players compared to British and European controls. The combined results of both stages of this study suggest that there is a genetic influence on speed-power performance within football.

PRACTICAL APPLICATIONS

The findings of this study, along with others exploring the prevalence of various SNPs within soccer players (refs) suggest that there is a genetic component to the attainment of higher playing abilities. The results of this study specifically highlight the relationship between a number of well-established and novel genetic variants and sprint performance across cohorts, suggesting that there is a favourable genetic profile for sprint performance. At present, it's unclear how this information might be used; there is a lack of evidence supporting the use of genetic information as a talent identification tool, and the general scientific consensus is that it should not be used in this way (Webborn et al., 2015). However, the results of this study potentially represent tentative steps towards our understanding of which genetic variants may predispose to enhanced sprint performance, suggesting that, if this information could be replicated in other cohorts, in may hold value in identifying talented future performers.

REFERENCES

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

2. Ahmetov, II, Egorova, ES, Gabdrakhmanova, LJ, Fedotovskaya, ON. Genes and athletic performance: An update. *Med Sport Sci* 61: 41-54, 2016.

3. Amasheh, S, Meiri, N, Gitter, AH, et al. Claudin-2 expression induces cation-selective channels in tight junctions of epithelial cells. *J Cell Sci* 115: 4969-4976, 2002.

4. Bangsbo, J, Nørregaard, L, Thorsoe, F. Activity profile of competition soccer. *Can J Sport Sci* 16: 110-116, 1991.

5. Ben-Zaken, S, Meckel, Y, Nemet, D, Eliakim, A. High prevalence of the IGF2 rs680 GG polymorphism among top-level sprinters and jumpers. *Growth Horm IGF Res* 37: 26-30, 2017.

6. Devaney, JM, Hoffman, EP, Gordish-Dressman, H, et al. IGF-II gene region polymorphisms related to exertional muscle damage. *J Appl Physiol* 102: 1815-1823, 2007.

7. Ding, X, Jin, Y, Wu, Y, et al. Localization and cellular distribution of CPNE5 in embryonic mouse brain. *Brain Res* 1224: 20-28, 2008.

8. Egorova, ES, Borisova, AV, Mustafina, LJ, et al. The polygenic profile of Russian football players. *J Sports Sci* 32: 1286-1293, 2014.

9. Eider, J, Cieszczyk, P, Leońska-Duniec, A, et al. Association of the 174 G/C polymorphism of the IL6 gene in Polish power-orientated athletes. *J Sports Med Phys Fitness* 53: 88-92, 2013.

10. Faude, O, Koch, T, Meyer, T. Straight sprinting is the most frequent action in goal situations in professional football. *J Sports Sci* 30: 625-631, 2012.

11. Fehrmann, RS, Jansen, RC, Veldink, JH, et al. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet* 7: e1002197, 2011.

12. Fernandez, T, Morgan, T, Davis, N, et al. Disruption of contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. *Am J Hum Genet* 74: 1286-1293, 2004.

13. Gineviciene, V, Jakaitiene, A, Tubelis, L, Kucinskas, V. Variation in the ACE, PPARGC1A and PPARA genes in Lithuanian football players. *Eur J Sport Sci* 14(sup1): S289-295, 2014.

14. Gomez-Gallego, F, Santiago, C, González-Freire, M, et al. The C allele of the AGT Met235Thr polymorphism is associated with power sports performance. *Appl Physiol Nutr Metab* 34: 1108-1111,2009.

15. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* 550: 204-213, 2017.

16. Haugen, TA, Tønnessen, E, Hisdal, J, Seiler, S. The role and development of sprinting speed in soccer. *Int J Sports Physiol Perform* 9: 432-441, 2014.

17. Itaka, T, Agemizu, K, Aruga, S, Machida, S. G Allele of the IGF2 ApaI Polymorphism Is Associated With Judo Status. *J Strength Cond Res* 30: 2043-2048, 2016.

18. Ito-Ishida, A, Miura, E, Emi, K, et al. Cbln1 regulates rapid formation and maintenance of excitatory synapses in mature cerebellar Purkinje cells in vitro and in vivo. *J Neurosci* 28: 5920-5930, 2008.

19. Jones, A, Woods, DR. Skeletal muscle RAS and exercise performance. *Int J Biochem Cell Biol* 35: 855-866, 2003.

20. Jones, N, Kiely, J, Suraci, B, et al. A genetic-based algorithm for personalized resistance training. *Biol Sport* 33: 117-126, 2016.

21. Kumanogoh, A, Shikina, T, Suzuki, K, et al. Nonredundant roles of Sema4A in the immune system: defective T cell priming and Th1/Th2 regulation in Sema4A-deficient mice. *Immunity* 22: 305-316, 2005.

22. Masamune, A, Nakano, E, Hamada, S, et al. Common variants at PRSS1–PRSS2 and CLDN2–MORC4 loci associate with chronic pancreatitis in Japan. *Gut* 64: 1345-1346, 2015.

23. Massidda, M, Voisin, S, Culigioni, C, et al. ACTN3 R577X Polymorphism Is Associated With the Incidence and Severity of Injuries in Professional Football Players. *Clin J Sport Med* 2017, doi: 10.1097/JSM.0000000000000487.

24. Masutomi, K, Possemato, R, Wong, JM, et al. The telomerase reverse transcriptase regulates chromatin state and DNA damage responses. *Proc Natl Acad Sci USA* 102: 8222-8227, 2005.

25. Mayhew, SR, Wenger, HA. Time-motion analysis of professional soccer. *J Hum Movement Studies* 11: 49-52, 1985.

26. Mohr, M, Krustrup, P, Bangsbo, J. Match performance of high-standard soccer players with special reference to development of fatigue. *J Sports Sci* 21: 519-528, 2003.

27. Molkentin, JD. Calcineurin–NFAT signaling regulates the cardiac hypertrophic response in coordination with the MAPKs. *Cardiovasc Res* 63: 467-475, 2004.

28. Papadimitriou, ID, Lucia,A, Pitsiladis, YP, et al. ACTN3 R577X and ACE I/D gene variants influence performance in elite sprinters: a multi-cohort study. *BMC Genomics* 17: 285, 2016.

29. Papadimitriou, ID, Papadopoulos, C, Kouvatsi, A, Triantaphyllidis, C. The ACE I/D polymorphism in elite Greek track and field athletes. *J Sports Med Phys Fitness* 49: 459, 2009.

30. Pimenta, EM, Coelho, DB, Cruz, IR, et al. The ACTN3 genotype in soccer players in response to acute eccentric training. *Eur J Appl Physiol* 112: 1495-1503, 2012.

31. Rankinen, T, Fuku, N, Wolfarth, B, et al. No Evidence of a Common DNA Variant Profile Specific to World Class Endurance Athletes. *PLoS One* 11: e0147330, 2016.

32. Reiner,A, Yang, M, Cagle, MC, Honig, MG. Localization of cerebellin-2 in late embryonic chicken brain: implications for a role in synapse formation and for brain evolution. *J Comp Neurol* 519: 2225-2251, 2011.

33. Ruiz, JR, Buxens, A, Artieda, M, et al. The− 174 G/C polymorphism of the IL6 gene is associated with elite power performance. *J Sci Med Sport* 13: 549-553, 2010.

34. Santiago, C, González-Freire, M, Serratosa, L, et al. ACTN3 genotype in professional soccer players. *Br J Sports Med* 42: 71-73, 2008.

35. Sawczuk, M, Maciejewska-Karlowska, A, Cieszczyk, P, Skotarczak, B, Ficek, K. Association of the ADRB2 Gly16Arg and Glu27Gln polymorphisms with athlete status. *J Sports Sci* 31: 1535-1544, 2013.

36. Schulz, RA, Yutzey, KE. Calcineurin signaling and NFAT activation in cardiovascular and skeletal muscle development. *Dev Biol* 266: 1-6, 2004.

37. Vafiadis, P, Bennett, ST, Todd, JA, Grabs, R, Polychronakos, C. Divergence between genetic determinants of IGF2 transcription levels in leukocytes and of IDDM2-encoded susceptibility to type 1 diabetes. J Clin Endocrinol Metab 83: 2933-2939, 1998.

38. Webborn, N, Williams,A, McNamee, M, et al. Direct-to-consumer genetic testing for predicting sports performance and talent identification: consensus statement. *Br J Sports Med* 49: 1486-1491, 2015.

39. Whitcomb, DC, LaRusch, J, Krasinskas, AM, et al. Common genetic variants in the CLDN2 and PRSS1-PRSS2 loci alter risk for alcohol-related and sporadic pancreatitis. *Nat Genet* 44: 1349, 2012.

40. Yamin, C, Duarte, JA, Oliveira, JM, et al. IL6 (-174) and TNFA (-308) promoter polymorphisms are associated with systemic creatine kinase response to eccentric exercise. *Eur J ApplPhysiol* 104: 579, 2008.

41. Yang, N, MacArthur, DG, Gulbin, JP, et al. ACTN3 genotype is associated with human elite athletic performance. *Am J Hum Genet* 73: 627-631, 2003.

42. Young, W, Russell, A, Burge, P, et al. The use of sprint tests for assessment of speed qualities of elite Australian rules footballers. *Int J Sports Physiol Perform* 3: 199-206, 2008.

43. Yvert, T., Zempo, H, Gabdrakhmanova, LJ, et al. AGTR2 and sprint/power performance: A case-control replication study for rs11091046 polymorphism in two ethnicities. *Biol Sport* 35 (2): 105-109, 2018.

44. Zarębska, A, Jastrzębski, Z, Moska, W, et al. The AGT gene M235T polymorphism and response of power-related variables to aerobic training. *J Sports Sci Med* 15: 616, 2016.

45. Zarebska, A, Sawczyn, S, Kaczmarczyk, M, et al. Association of rs699 (M235T) polymorphism in the AGT gene with power but not endurance athlete status. *J Strength Cond Res* 27: 2898-2903, 2013.

Figure legends

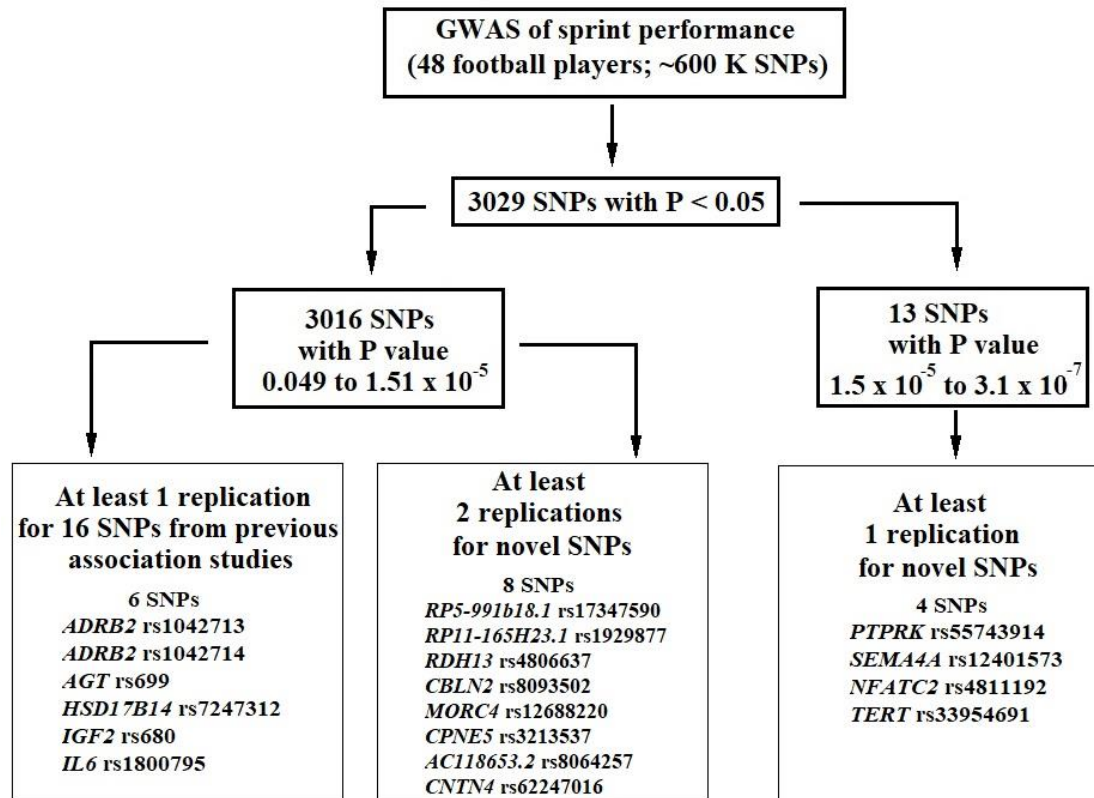**Figure 1.** A schematic overview of the study design and findings

Table 1. Top 12 SNPs associated with sprint test performance in football players.

Gene	SNP	Consequence	Sprint allele	5 m sprint test		20 m sprint test	
				β	*P*	β	*P*
PTPRK	rs55743914	intronic	C	-0.0809	**7.7x10⁻⁷**	-0.117	0.00175
SEMA4A	rs12401573	synonymous	T	-0.0478	0.0048	-0.133	**7.8x10⁻⁵**
NFATC2	rs4811192	intronic	G	-0.0477	0.011	-0.152	**3.7x10⁻⁵**
TERT	rs33954691	synonymous	G	-0.0528	0.015	-0.171	**6.6x10⁻⁵**
Near *RP5-991b18.1*	rs17347590	intergenic	C	-0.0448	0.04	-0.097	0.034
Near *RP11-165H23.1*	rs1929877	intergenic	G	-0.0435	0.041	-0.114	0.0096
RDH13	rs4806637	3'-UTR	G	-0.0638	0.026	-0.141	0.0181
CBLN2	rs8093502	upstream gene variant	G	-0.0516	0.014	-0.112	0.0099
Near *MORC4*	rs12688220	intergenic	G	-0.0584	0.013	-0.110	0.026
CPNE5	rs3213537	intronic	G	-0.0509	0.012	-0.113	0.0075
Near *AC118653.2*	rs8064257	intergenic	G	-0.0457	0.049	-0.102	0.035
CNTN4	rs62247016	intronic	A	-0.0804	0.006	-0.193	0.0014

Table 2. Associations between sprint alleles identified in the football cohort and sprint test performance in Polish females.

Gene	SNP	Sprint allele	Sprint test performance (*P* value)		
			5 m	**10 m**	**30 m**
PTPRK	rs55743914	C		0.015 (baseline)	
NFATC2	rs4811192	G		0.0049 (post-training)	
TERT	rs33954691	G		0.036 (increase)	
Near *RP5-991B18.1*	rs17347590	C	0.0126 (baseline)		
CPNE5	rs3213537	G		0.011 (baseline)	
Near *AC118653.2*	rs8064257	G			0.03823 (increase)
CNTN4	rs62247016	A			0.0074 (increase); 0.0096 (post-training)

Table 3. Associations between sprint alleles identified in the football cohort and power athlete status in the Russian groups of athletes and controls.

Gene	SNP	Sprint allele	Sprinters vs controls		Speed-strength athletes vs controls		Sprinters vs endurance athletes	
			OR	*P*	OR	*P*	OR	*P*
Near *RP11-165H23.1*	rs1929877	G			2.5	0.0066		
Near *MORC4*	rs12688220	G	2.9	0.02148			8.7	0.00049
CPNE5	rs3213537	G	2.5	0.0346			5.2	0.0045
Near *AC118653.2*	rs8064257	G					4.7	0.016

Table 4. Associations between sprint alleles identified in the football cohort and proportion of fast-twitch muscle fibres in the Russian group of physically active subjects.

Gene	SNP	Sprint allele	Proportion of fast-twitch muscle fibres					
			Males		Females		Combined	
			β	*P*	β	*P*	β	*P*
SEMA4A	rs12401573	T					4.371	0.032
Near *RP5-991B18.1*	rs17347590	C	11.23	0.0044			9.563	0.002
Near *RP11-165H23.1*	rs1929877	G	6.553	0.0021			4.665	0.0038
RDH13	rs4806637	G	10.59	0.028			9.137	0.024
CBLN2	rs8093502	G	7.06	0.041	10.37	0.042	7.772	0.006
Near *MORC4*	rs12688220	G					5.728	0.048