ORIGINAL RESEARCH ARTICLE

Journal of Cellular Physiology WILEY

Genomic analysis reveals association of specific SNPs with athletic performance and susceptibility to injuries in professional soccer players

Raffaele La Montagna[1]* | Raffaele Canonico[2]* | Luigi Alfano[3] | Enrico Bucci[1] |
Silvia Boffo[1] | Leopoldo Staiano[4] | Beniamino Fulco[5] | Enrico D'Andrea[1] |
Antonio De Nicola[1] | Patrizia Maiorano[5] | Costantino D'Angelo[5] |
Andrea Chirico[1,6] | Alfonso De Nicola[1] | Antonio Giordano[1,5]

[1]Sbarro Health Research Organization, Wayne, Pennsylvania

[2]Unità Operativa Complessa di Dietetica, Medicina dello Sport e Benessere Psico-Fisico, Università degli Studi della Campania Luigi Vanvitelli, Naples, Italy

[3]Cell Biology and Biotherapy Unit, Istituto Nazionale Tumori - IRCCS, Fondazione G. Pascale, Naples, Italy

[4]Laboratory of Cellular and Developmental Biology, Stazione Zoologica Anton Dohrn, Naples, Italy

[5]Department of Medical Biotechnologies, University of Siena, Italy

[6]Department of Psychology of Development and Socialization Processes, "Sapienza" University of Rome, Italy

Correspondence

Antonio Giordano MD PhD, Sbarro Health Research Organization, BioLife Science Bldg. Suite 431 1900 N 12th Street, Philadelphia, PA 19122.

Email: giordano@temple.edu

Abstract

The development of specific and individualized training programs is a possible way to improve athletic performance and minimize injuries in professional athletes. The information regarding the sport's physical demands and the athletes' physical profile have been, so far, considered as exhaustive for the design of effective training programs. However, it is currently emerging that the genetic profile has to be also taken into consideration. By merging medical and genetic data, it is thus possible to identify the athlete's specific attitude to respond to training, diet, and physical stress. In this context, we performed a study in which 30 professional soccer players, subjected to standard sport medical evaluation and practices, were also screened for genetic polymorphism in five key genes (*ACTN3*, *COL5A1*, *MCT1*, *VEGF*, and *HFE*). This genetic analysis represents the central point of a multidisciplinary method that can be adopted by elite soccer teams to obtain an improvement in athletic performance and a concomitant reduction of injuries by tailoring training and nutritional programs. The genetic fingerprinting of single athletes led to the identification of two performance-enhancing polymorphisms (ACTN3 18705C>T, VEGF-634C>G) significantly enriched. Moreover, we derived a genetic model based on the gene set analyzed, which was tentatively used to reduce athletes' predisposition to injuries, by dictating a personalized nutrition and training program. The potential usefulness of this approach is concordant with data showing that this team has been classified as the healthiest and least injured team in Europe while covering the highest distance/match with the highest number of high-intensity actions/match.

KEYWORDS

athletic performance, genes, injuries, professional soccer players, single-nucleotide polymorphisms

1 | INTRODUCTION

Training for elite athletes is a complex task that needs to be adapted not only to specific sports but also to the individual predisposition of every single athlete. Soccer is considered a total and very demanding sport that requires the concomitant training of different body systems, such as muscular, cardio-circulatory, respiratory, and nervous system often supported by specific nutrient supplementation. This multiparametric training fulfils different physical and nutritional requirements that, in turn, might improve muscle size and strength, endurance, resistance to fatigue, and coordination. Although the physical and metabolic features of each athlete are considered to be the main parameters to design specific training programs, the impact of genetic traits on the success of a given training is currently accepted. Advances in science and technology could in principle change the way in which soccer players are subjected to training programs designed from a personalized perspective. The inclusion of genetics in the design of training programs (PEP, performance-enhancing polymorphisms) was triggered by the identification of a large number of genetic variants, linked to athletic performance (Ostrander, Huson, & Ostrander, 2009). The correlation between specific gene polymorphisms and the response to physical and nutritional cues can then be explored to evaluate athletes' profiles (Ferioli et al., 2019). This would allow for the development of tailored programs that maximize athletic performance and reduce the occurrence of indirect injuries. The medical staff decided to test this new scientific approach and has defined a new physical, genetic, and nutritional multiparametric approach for soccer training aiming for: (a) Improvement of athletic performance; (b) reduction of injuries; and (c) amelioration of athletes' health. Since the program started in 2013, players trained by this medical staff are considered the "healthiest" in terms of injury susceptibility and performance (Figure 1) by both scientific and sport

communities (*Lega Serie A and UEFA*). As a unique physical-genetic entity, each athlete was subjected to genetic screening and accurate physical monitoring during the entire season. The purpose of these analyses was to explore the connection between gene polymorphisms considered to be important for athletic performance and the actual training efficacy, with a particular focus on the effect that genetic differences have on athletes' competitiveness, endurance, and resistance to injuries.

For athletes' profiling and coaching, several professionals were involved including physicians, geneticists, nutritionists, and physiotherapists. Review of data in the literature allowed us to identify a set of five genes that were previously correlated with athletic performance, such as endurance capacity, lactic acid clearance, iron intake and clearance, and muscle fatigue. These genes have also specific polymorphisms that are responsible for increased or decreased gene activity. Specifically, we focused our attention on: MCT1 (*alias SLC16A1, Solute Carrier Family 16 Member 1*), a lactate transporter protein (Merezhinskaya, Fishbein, Davis, & Foellmer, 2000); HFE (*Homeostatic Iron Regulator*) involved in blood and tissue iron homeostasis (Luszczyk et al., 2017); ACTN3 (*Actin alpha 3*) a protein responsible for fast muscle fiber functioning (Lee, Houweling, North, & Quinlan, 2016), and COL5A1 (*Collagen Type V Alpha 1 Chain*) mainly found in ligaments and tendons and associated with severity of musculoskeletal injuries (Massidda et al., 2015a). Moreover, we analyzed a polymorphism in the *VEGF* gene, encoding the vascular endothelial growth factor, associated with an improvement in aerobic performance (Prior et al., 2006). The aim of the present study was to analyze the concordance between the information available on the effects of the selected PEPs on athletic performance and the actual performance of the team. Focusing on a set of genes, which were already validated by other authors on large samples of professional athletes, is a way to avoid the risk of finding spurious genetic correlations in the limited population target of the present study.
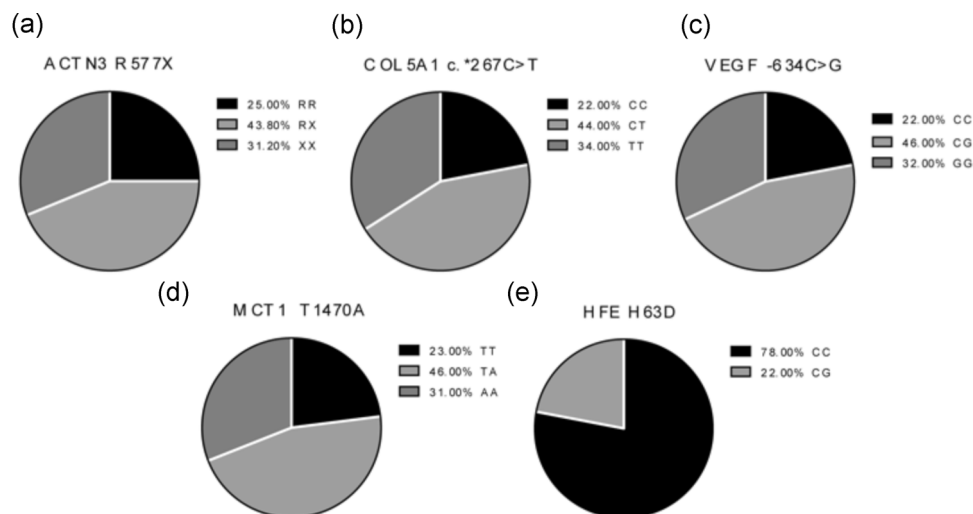
FIGURE 1 Distribution and frequency of performance-enhancing polymorphisms in S.S.C. Naples soccer players. (a–e) Genes, alleles, and genotype frequency in the 30 S.S.C. Naples soccer players. In black, the homozygous wild type genotype; in light gray, the heterozygous and in the dark gray, the homozygous polymorphic genotype

The data obtained by the genetic analysis and the physical information of each athlete were correlated, to see whether single genetic polymorphisms or a combination thereof could be used to retrospectively predict the injury rate of each athlete.

2 | METHODS

The soccer team's medical staff authorized the use of samples for screening, as described in the protocol of this project. Athletes authorized the use of the results for scientific scope.

2.1 | Study participants

In total, 30 soccer players with different nationalities were recruited and included in the study.

2.2 | Genomic DNA extraction

The genomic DNA was extracted from saliva, by using the Oragene-DNA (OG-500) of the DNA Genotek Inc., following manufactory instructions. Briefly, 2 ml were collected from a single soccer player; 500 µl of the mixed sample was transferred to the microcentrifuge tube followed by adding of 20 µl (1/25th volume) of PT-L2P (DNA Genotek), mixed by vortexing and incubated for 10 min on ice. The sample was centrifuged for 5 min at $15,000 \times g$ to pellet the impurities. About 600 µl of 95% EtOH was added to the 500 µl of cleared supernatant and mixed by inversion ten times. The sample was allowed to stand at room temperature for 10 min for the DNA precipitation, followed by centrifugation for 2 min at $15,000 \times g$. The pellet was washed by adding 250 µl of 70% EtOH and centrifugated for 2 min at $15,000 \times g$. The genomic DNA was resuspended in 100 µl of TE solution and stored at −20°C.

2.3 | Genotyping

For the evaluation of SNP in the *MCT1* gene, we amplified the genomic DNA by using the PfU ultra HF as manufactory instructions. Primers containing the desired mutation were designed between 20

and 40 bases in length (sequences provided in Table 1). The PCR reaction was performed as follows: 94°C for 3 min (1 cycle); 94°C for 30 s , −55°C for 1 min, −72°C for 1 min (for 35 cycles), and 72°C for 10 min. The PCR product was analyzed on the 1% agarose gel and extracted with the QIAquick Gel Extraction Kit (Qiagen) per the manufactory's instructions. The samples were then sequenced at the Thomas Jefferson Genomic Facility: Sanger sequencing was performed in 100 ng of PCR product, using 3.2 µM the forward primer.

For the evaluation of SNPs in the HFE, COL5A1, ACTN3, and VEGF, we performed an allelic discrimination by TaqMan assay using the following specific reagents: two unlabeled PCR primers (forward and reverse, Table 1) at 900 nM final concentration; Applied Biosystems™ VIC™ and Applied Biosystems™ FAM™ dyes (MGB-labeled probes), both at 200 nM final concentration. Thermocycling was performed in real-time PCR instrument platforms [Roche, Light Cycler 480II], using the specific Applied Biosystems™ TaqMan™ Genotyping Master Mix following standard conditions. In particular, the reaction is divided into two steps. Initial steps: AmpErase UNG activation (2 min at 50°C); second step: AmpliTaq Gold DNA polymerase activation (10 min at 95°C), followed by 40 cycles of PCR (Melt 15 s at 95°C and Anneal/Extend 1 min at 60°C).

2.4 | Statistical analysis

To exclude sampling biases, we analyzed the consistency between the tested genotype frequencies with a Hardy–Weinberg equilibrium by using a χ^2 test, verifying whether the probability of obtaining the calculated χ^2 value by chance only was larger than 0.05.

To test the significance of the difference in frequency of any single genotype compared to a given reference frequency, we used the χ^2 test as recommended by Campbell (2007) and Richardson (2011), assuming alpha = 0.05.

To obtain a quantitative predictor of injury risk, based on the genotype of the five selected genes, we fitted the number of muscular or ligament injuries occurred for each team player with available genotype information ($n = 27$) by a logistic (or logit) multinomial model.

This is a generalization of the logistic model with a variable response on more than two categories. We used this method to understand the effect of a qualitative variable series (the five selected

TABLE 1 List of specific primers used for the allelic discrimination by TaqMan assay

SNP ID	Gene	Context sequence [VIC/FAM]
rs1800562	HFE C28Y	CCTGGGGAAGAGCAGAGATATACGT[**G/A**]CCAGGTGGAGCACCCAGGCCTGGAT
rs1799945	HFE H63D	TGACCAGCTGTTCGTGTTCTATGAT[**C/G**]ATGAGAGTCGCCGTGTGGAGCCCCG
rs1800730	HFE S65C	GCTGTTCGTGTTCTATGATCATGAG[**A/T**]GTCGCCGTGTGGAGCCCCGAACTCC
rs12722	COL5A1	CCGCCCCACGCTCTGTCCACACCCA[**C/T**]GCGCCCCGGGAGCGGGGCCATGCCT
rs1815739	ACTN3 R577X	CAAGGCAACACTGCCCGAGGCTGAC[**T/C**]GAGAGCGAGGTGCCATCATGGGCAT
rs2010963	VEGF	CGCGCGGGCGTGCGAGCAGCGAAAG[**C/G**]GACAGGGGCAAAGTGAGTGACCTGC
	MCT1	Forward: 5′-ACACATACTGGGCATGTGGCGTCGT Reverse: 5′-AAATCCCATCAATGAACAACTGGTATGATTTCCAC

genotypes) on the response variable (the number of muscular or ligament injuries occurring to each included player).

The obtained model fitting to the available data was good, with a R^2(McFadden) = 0.953, R^2(Cox and Snell) = 0.876 and R^2(Nagelkerke) = 0.986.

With a χ^2 value of 56.446, testing the null hypothesis of getting such a fitting by chance with the available degrees of freedom yielded a $p = .016$.

All calculations were performed using the software NodeXL – v. 2018.5.

3 | RESULTS

Here we report a detailed analysis of the incidence of PEP in the genes of interest in 30 soccer players, highlighting the predicted effects of a given PEP on athletic performance.

Among the five genes selected for this study, *ACTN3*, *COL5A1*, and *VEGF* influence athletic performance and predisposition to indirect injury, independently of any nutritional supply. The effect of these PEPs may be enhanced or counteracted, perhaps developing personalized training. However, there are two other genes (*MCT1* and *HFE*) in which PEPs could be used in the coordination of a nutritional intervention along with the development of specific training programs to modify their positive or negative influence.

What follows, is a detailed analysis of the PEP genetic makeup found for each gene in the soccer team, and then a discussion of their combined predicting power when a suitable quantitative biomarker model is applied.

3.1 | Actin alpha-3

ACTN3 encodes α-actinin-3, a protein that participates to the composition of sarcomeres, the functional units of striated muscles; it is particularly abundant in the fast type II muscle fibers, which are involved in the muscle sprinting and strength activities (Blanchard, Ohanian, & Critchley, 1989). The ACTN3 R577X (rs1815739) polymorphism has been suspected for several years to be associated with particular athletic performances in power-oriented elite athletes of different ethnicities (Yang et al., 2003). This polymorphism is a genomic C to T transition in position 18705 coding for an arginine to X substitution in position 577 (R577X). The polymorphism can be present in homozygosis (DNA: TT or protein: XX), in heterozygosis (DNA: CT or Protein: RX) or absent (DNA: CC or protein: RR). The X-allele encodes a shorter variant of ACTN3 that stops at amino acid 577 instead of 901 resulting in a protein with lower performance capabilities. The CC (RR at protein level) and CT (RX at protein level) genotypes were initially found to be increased in frequency in elite strength athletes, in agreement with the fact that the TT (XX) mutation was previously associated with impaired muscle performance (Roth et al., 2008). The enrichment in CC and CT genotypes, however, was not confirmed by different large-scale studies (Massidda et al., 2015; Rankinen et al., 2016), casting doubts on the original interpretation

and on the differences between elite athletes and the general population. However, the exact genotype plays a role in differentiating among athletes of the same discipline: in a study on 200 Brazilian soccer players, the presence of a CC genotype was associated with players fastest in short distances and with higher jump potential, while the TT genotype was associated with higher aerobic capacity (Pimenta et al., 2013). The soccer team analyzed in this project is characterized by a high frequency of TT homozygotes, which is statistically different from controls or athletes enrolled in all the above-mentioned studies. In line with results reported for soccer players (Pimenta et al., 2013), this might indicate that, on average, athletes have a higher aerobic capacity. Furthermore, very recently, the TT genotype has been strongly associated also with a higher incidence of muscle injuries (Massidda et al., 2017). The overall athletic performances, recognized by the *Lega Calcio Serie A*, agreed with a possible effect of the optimized training program (Figure 2a).

3.2 | Collagen Type V alpha 1

Collagens are among the main components of connective tissues and the extracellular matrix. Many diseases, collectively known as collagenopathies, are associated with mutations and/or polymorphisms in different types of collagen (Jobling et al., 2014). Among the 13 types of collagen, Collagen V is mainly enriched in tendons and ligaments where it accounts for 2% of total collagen (mainly type I). Although Collagen V is quantitatively a minor component, it has a fundamental role in the regulation of the assembly of functional collagen fibers. It is composed of two α1 chains and two α2 chains, and *COL5A1* encodes specifically for the Collagen type V alpha 1 chain. A single-nucleotide polymorphism C to T (rs12722) of *COL5A1* gene occurs in the 3′-untranslated region of mRNA improving its stability and increasing the protein amount, consequently (Abrahams, Posthumus, & Collins, 2014; Bertuzzi et al., 2014). The higher amount of Collagen V alters the mechanical properties of tendons resulting in an increased predisposition to tendon and ligament injuries. *COL5A1* rs12722 polymorphism distribution in the cohort analyzed in this study highlights a prevalence of the heterozygous (CT) genotype, which accounts for the majority of athletes (44%) compared with the homozygous genotypes CC (22%) and TT (34%) (Figure 1b). Recently, Heffernan et al. (2017) described a small frequency of the CC allele associated with better performance in rugby players. The frequency of CC genotype in the analyzed soccer players is similar to that of the professional rugby players included in the study by Heffernan et al. However, the number of athletes included in the present study is too limited for the detection of the expected small effect (insufficient sample power), so that is not possible to draw any significant conclusion on whether the soccer players deviate significantly from the general population (i.e., the control sample reported in the study by Heffernan et al., 2017).

3.3 | Vascular endothelial growth factor

VEGF is a member of the growth factor family that plays a pivotal role in the formation of new and functional blood vessels (Hoeben
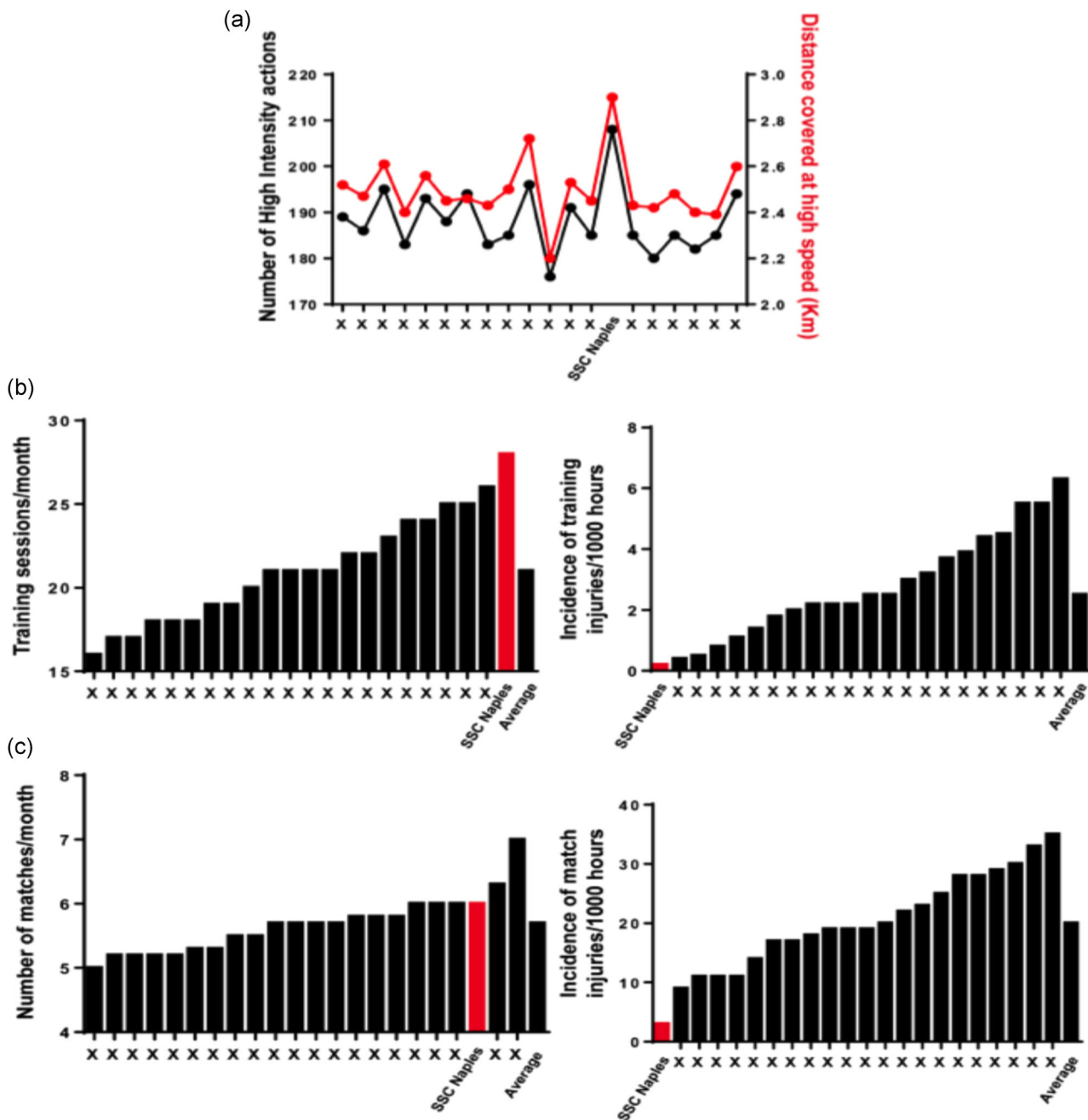
FIGURE 2 Analysis of athletic performances and injury rate of S.S.C. Naples athletes (adapted from Ranking Serie A-Report Fisico (a) and UEFA elite club injury study (b), (c)). (a) In the graph the report of the athletic performance in the 38 matches of the Italian championship 2017–2018 is presented. The black line indicates the number of high-intensity actions and the red line indicates the distance covered at high speed (> 15 km/hr) by the 20 teams of Serie A. (b), (c) The graphs presented are adapted from the mid-season report (2017–2018) of the UEFA Elite club injury study. In (b) the data regarding the number of training sessions/month and the incidence of injuries/1,000 hr of training are presented. It is clear that S.S.C. Naples is the Team, among the other participants to the UEFA Champions League, with the highest number of training session/month and with the lowest incidence of indirect injuries. In (c) the data regarding the number of matches/month and the incidence of injuries/1,000 hr of match are presented. It is clear that S.S.C. Naples is one of the teams with the highest number of matches/month and with the lowest incidence of indirect injuries. These analysis were conducted by Professor Jan Ekstrand and are reported in The UEFA Elite Club Injury Study (ECIS), funded and supported by UEFA. Data reported above correspond to 2017/2018 middle season analysis. This trend has also been confirmed in the 2017/18 Final Season Report performed by Professor Jan Ekstrand (data not shown)

et al., 2004). VEGF is expressed in a variety of human tissues, including skeletal muscles, and it has been recently demonstrated to control the angiogenic processes that follow aerobic exercises (Gustafsson, Puntschart, Kaijser, Jansson, & Sundberg, 1999; Richardson et al., 2000). Aerobic exercise causes the development of a hypoxic environment in the muscles (Richardson, Noyszewski, Kendrick, Leigh, & Wagner, 1995; Thompson, Pratley, & Ossowski, 1996) that induces the synthesis of hypoxia-inducible-factor 1 (HIF1) which, in turn, upregulates the expression of VEGF, through the prolonged mRNA stability, resulting in an increased protein function (Liu et al., 2002). Being tightly linked to aerobic exercise and athletic performance, angiogenesis thus represents a critical factor in the adaptation to the training-induced increased oxygen consumption (VO2 Max) since it can foster the formation of new capillaries in the muscles. Neo-angiogenesis facilitates the delivery of oxygen by the red blood cells to the muscles, making it available to foster muscle metabolism. Previous data show that the amount of capillaries per muscle fiber, whose formation is triggered by the activity of VEGF, is increased in endurance athletes, indicating the requirement for an improved oxygen transport system to support aerobic activity (Brodal, Ingjer, & Hermansen, 1977). Although some aerobic training has been demonstrated to control the generation of new capillaries through VEGF upregulation, it is worth mentioning that genetic factors may also contribute to the success of angiogenesis in elite athletes (Wahl et al., 2014). Variations of the DNA sequence in the promoter of VEGF have been correlated with different levels of VEGF expression (Al-Habboubi, Sater, Almawi, Al-Khateeb, & Almawi, 2011). It has been reported that a C to G polymorphism in the VEGF promoter at position −634 is causative, when in homozygosis (GG), of a reduced expression of VEGF in hypoxic conditions and this causes, in turn, a reduced maximal oxygen consumption (Prior et al., 2006). However, in the same study, the authors have also demonstrated that whenever the polymorphism occurs in heterozygosis (CG), the VEGF expression under hypoxia is null and VO2 Max is not affected. In light of these findings, we screened 30 soccer players for the rs2010963 polymorphism (−634C > G). The three possible genotypes CC, CG, and GG are represented as follows: the heterozygous (CG) is the single most diffuse genotype (46%), as compared with the homozygous geno-types CC (22%) and GG (32%; Figure 1c). In a large study on 670 Russian athletes of different disciplines, a small but significant excess of the C allele was correlated to the higher aerobic performance of athletes and to a better lactic acid metabolism (Akhmetov et al., 2008). As for the case of the ACTN gene, soccer players exhibit a significant excess of haplotype conferring better aerobic capacities, with the CC homozygotes exceeding in a statistically significant proportion the frequency measured both for the control and the athlete populations in the Russian study, indirectly confirming the association of G-634C polymorphism with physical performance of the athletes. This genetic makeup of the team agrees with the fact that the soccer team under investigation has been the Italian soccer team, which covered the highest distance/match with the highest number of high-intensity actions/match (Figure 2a). This supports the

tight connection that is needed among athletes, medical staff and geneticists to allow for a tailored training program for each athlete that takes into account both physical and genetic features.

3.4 | Monocarboxylate transporter 1

Lactate is widely considered as a negative influencer of athletic performance and it is often correlated with muscle fatigue and pain after exercise (Westerblad, Allen, & Lannergren, 2002). However, it is becoming more and more clear that lactate is an extremely important fuel for a particular type of muscle fibers: the type I oxidative muscle fibers. In general, during intense exercise, the glycolytic fibers produce lactate and protons as end products of glycolysis. These by-products need to be excreted from the muscle fibers and released into the blood. The excretion of lactate and protons mainly occur through the monocarboxylate transporters (MCTs), a family of 14 members with different tissue localizations and expression levels (Halestrap & Wilson, 2012). MCTs can mediate the inward/outward transfer of lactate/proton according to their chemical gradient (Juel & Halestrap, 1999). The most characterized MCTs are MCT1 and MCT4, with the latter controlling the export of lactate from glycolytic fibers (Cupeiro et al., 2016). MCT1 is, by contrast, responsible of the uptake of lactate by type I oxidative fibers and its activity has the double beneficial effect of (a) reducing the circulating levels of lactate and (b) increasing the levels of lactate in the muscle fibers that can use it as fuel (Van Hall, 2000). Interestingly, MCT1 expression levels are increased upon high-intensity training, enforcing its role in enhancing the athletic performance through lactate reabsorption. A PEP (rs1049434) in the MCT1 gene has been identified in position 1470, corresponding to A to T transversion (A1470T) which results in a glutamic acid to aspartic acid substitution in position 490 of the protein (Glu490Asp). Compared with carriers of the A allele, the presence of the minor allele (T) is causative of a two-fold reduction in lactate transport by MCT1 (Merezhinskaya et al., 2000) and has been correlated with higher levels of blood lactate during high-intensity training (Cupeiro, Benito, Maffulli, Calderon, & Gonzalez-Lamuno, 2010). These findings highlight the impact that MCT1 A1470T polymorphism may have on endurance performances and on athlete's response to specific training. It has been demonstrated however that the frequency of the T allele of MCT1 was high in elite sprinters, suggesting that reduced lactate uptake by type I oxidative fibers is not adverse to anaerobic activities (Kikuchi et al., 2017; Sawczuk et al., 2015). By contrast, the T-allele may be deleterious for the aerobic and endurance activities (such as soccer players). We analyzed the occurrence of MCT1 A1470T polymorphism in the 30 soccer players; the three possible genotypes AA, AT, and TT turned out to be distributed with a prevalence of heterozygous (AT) genotype, that accounts for most athletes (46%) as compared to the homozygous genotypes AA (23%) and GG (31%; Figure 1d). In a large study on 323 Russian athletes and 467 control individuals, a small excess of the AA genotype frequency in the athlete population has been associated with higher in endurance (Fedotovskaya, Mustafina, Popov, Vinogradova, & Ahmetov, 2014). However, the

limited sample used in the present study did not allow us to find any significant difference between the AA frequency among the soccer players and the control population used in that study, thus preventing any conclusion to be drawn. However, in a recent study on a population on 73 male soccer players, the AA genotype was found to be associated with a higher risk of muscular injuries (Massidda et al., 2015). The frequency of the AA genotype found in the published study involving 73 soccer players is statistically identical to that observed in the team, meaning that the genetic risk of muscular problems due to this variant is not different in the soccer team as compared with other teams in the same country. Thus, the observed lower frequency of muscular accidents for this team must be due to reasons other than the genetic MCT1 makeup. Among the possible reasons, the superior performance of the studied team can be connected to the special training programs and nutritional intervention (i.e., the specific formulation of sports drinks to carbohydrate and fluid supplementation) developed by the medical staff with the help of professional nutritionists to improve the athletic performance of the athletes (Figure 2a).

3.5 | Hemostatic iron regulator

The hereditary hemochromatosis gene (*HFE*; High-Fe) encodes a transmembrane protein that plays a key role in iron absorption (Barton, Edwards, & Acton, 2015). Although HFE is not an iron transporter, it interacts and regulates the activity of other proteins such as: the transferrin receptor (being a negative regulator of it), that directly controls iron uptake from blood circulation (Parkkila et al., 2001); the ferroportin, a transmembrane protein that clears iron from the cell, excreting it into the blood (Ganz, 2005). HFE is thus able to control the homeostasis of iron (Barton et al., 2015). Loss of function of HFE results in a genetic disease known as hemochromatosis, characterized by excessive accumulation of iron in the organism (Barton & Edwards, 1993; Barton et al., 2015). However, high iron content, below the pathological threshold, may be beneficial for athletic performance. The HFE C187G polymorphism results, at the protein level, in the substitution of a Histidine with an aspartic acid in position 63 (H63D; rs1799945) with the subsequent reduction of HFE function (Yang, Ferec, & Mura, 2011). HFE H63D has a reduced ability to bind to the transferrin receptor resulting in increased transport of iron in blood and cells (Feder et al., 1998). Although a clear association of HFE H63D polymorphism with elite athletic performance has not been established, it is worth noting that in several studies a higher frequency of HFE H63D has been found in endurance athletes such as cyclists (Chicharro et al., 2004; Deugnier et al., 2002). Iron is extremely important for oxygen transport and delivery to the muscles; thus, it is reasonable that any impairment of iron uptake would result in an increase of circulating iron and an increase in oxygen availability, conditions that are both important for the execution of aerobic performance. The HFE H63D polymorphism, exclusively in heterozygosis with the wild type allele (CG genotype) has been found in 22% of the soccer players (Figure 1e). The players

carrying the HFE H63D polymorphism have been carefully followed by the medical staff that periodically checked the serum iron, transferrin, and ferritin levels to control the iron homeostasis and execute nutritional interventions to avoid the variation of these parameters above a certain threshold detrimental for the health and the performance of the athletes. For HFE polymorphism, we obtained a nonsignificant difference with the data reported for a large, general population study and for a smaller control population in a different study (Chicharro et al., 2004; Katsarou et al., 2016). In contrast, the G haplotype resulted significantly less present in S.S.C. Naples athletes than in a sample of 65 endurance athletes, with a significant difference between the overall number of individuals carrying at least one H63D C > G mutation [$p = .0067$; χ^2 test as recommended by (Campbell, 2007) and (Richardson, 2011)]. In this respect, we conclude that S.S.C. Naples soccer players do not differ from the general control population.

3.6 | Combined analysis

From the single-gene analysis discussed so far, it appeared that the S.S.C. Naples team is predicted to be negatively influenced at the genetic level by the measured frequency of ACTN3 haplotypes, which predisposes the players to muscular injuries, and by specific MCT1 haplotypes positively affecting the iron transport, which turned out to be less abundant in comparison to other teams playing in Italy. However, the team is also well disposed because of its VEGF superior genetic makeup, predicted to confer high aerobic capacities and in agreement with the reported team's highest distance/match covered and the highest number of intensity actions/match performed.

To unravel the combined effects of the studied polymorphisms on the number of muscular injuries occurring to each player, we resorted to a multinomial fitting procedure. The overall number of muscular injuries occurred to each player is reported in Table 2.

Complete genetic information was available only for 27 players. Assuming as a regression variable the number of muscular incidents, we performed a multinomial fitting of a logistic model based on the player' genotype. After the regression, an almost complete separation of observations was detected, with the following very good determination coefficients: R^2(McFadden) = 0.953, R^2(Cox and Snell) = 0.876, R^2(Nagelkerke) = 0.986. The goodness of the fit is demonstrated by a $\chi^2 = 56.446$; with the available degrees of freedom, the probability of finding a value of χ^2 equal or superior in case the null hypothesis is true (i.e., there is no correlation between the genetic makeup and the number of injuries for any given player) is $p = .016$, thus below the usual significance alpha threshold of $p = .05$.

In agreement with such a good and significant fitting, the players' reclassification table obtained by applying the identified model to the 27 considered players is reported in Table 3.

TABLE 2 Overall number of muscular injuries occurred in S.S.C. Naples soccer players (30) during the season 2014–2018

Player	Injuries occurred	Player	Injuries occurred
1. xxxxx	0	16. xxxxx	2
2. xxxxx	0	17. xxxxx	0
3. xxxxx	1	18. xxxxx	0
4. xxxxx	3	19. xxxxx	1
5. xxxxx	0	20. xxxxx	0
6. xxxxx	0	21. xxxxx	4
7. xxxxx	2	22. xxxxx	1
8. xxxxx	0	23. xxxxx	0
9. xxxxx	1	24. xxxxx	0
10. xxxxx	0	25. xxxxx	0
11. xxxxx	0	26. xxxxx	0
12. xxxxx	1	27. xxxxx	0
13. xxxxx	1	28. xxxxx	0
14. xxxxx	0	29. xxxxx	2
15. xxxxx	1	30. xxxxx	0

As evident from the table, the prediction of the number of injuries based on the obtained model is wrong in only a single case (a player who was injured only once is back-predicted to have incurred no injuries).

This is the most accurate athletic injury predictive model we are aware of, relying on the thoughtful selection of five genes already related to the athletic performance, which were never tested simultaneously for their predictive power. With the help of this model, the S.S.C. Naples medical staff or the staff of other soccer teams can now in principle focus their efforts on players at risk, tailoring their training programs and giving them useful insights, so as to optimize the performance of the athletes.

As a note of caution, we should, however, mention that the obtained regressive model, albeit very good in retro-predicting the occurrence of injuries for each considered player in the S.S.C. Naples soccer team, is promising but vexed by a potential overfitting problem, due to the limited sample size used and to the lack of an external, large sample of genetic information on soccer players to test it. A thorough validation of the predictor needs further testing, which is however not in the scope of the present study.

TABLE 3 Prediction of number of injuries for S.S.C. Naples soccer players (27), based on the new model

From / to	0	1	2	3	4	Total	% correct
0	**16**	0	0	0	0	16	100.00
1	1	**6**	0	0	0	7	85.71
2	0	0	**2**	0	0	2	100.00
3	0	0	0	**1**	0	1	100.00
4	0	0	0	0	**1**	1	100.00
Total	17	6	2	1	1	27	97.14

4 | CONCLUSION

This study is based on the hypothesis that the combination of appropriate training, adapted to the genetic and nutritional profile of each athlete, can in principle be a way to achieve elite athletic status and performance.

Five genes were selected for the screening of performance-enhancing polymorphisms in elite athletes, so as to cover the polymorphisms in genes that intrinsically predispose toward a specific activity (sprint vs. endurance) or to injuries and in genes that can be externally modulated, for instance through specific nutrient supplementation.

The evaluation of the status of each of the selected five genes in the S.S.C. Naples soccer team uncovered two potential genetic hindrances and one genetic enhancer of athletic performance. Given the robust association of these genes to performance and injury risk, it is possible to infer that the potential genetic hindrances were minimized at least in part by the training and nutritional program designed by the team's medical staff, thus explaining the excellent performance obtained by the team (as recognized by *Lega Calcio Serie A* and *UEFA*) in terms of athletic performance and injury rate.

Despite the overall reduced injuries occurring for the team's players, a residual genetic effect on the individual risk can be observed when combining all the five considered genes. The simultaneous effect of these five genes is indeed strongly correlated to the number of injuries experienced by the single components of the S.S.C. Naples soccer team. The obtained regression indicates that the identified five genes-performance association could, at least in part, explain why an athlete is more susceptible to injuries than another, so that, after a proper validation in a larger population, it can in principle be used to identify athletes who need special care during training sessions and special attention when taking the field and could support the development of tailored injury prevention programs and/or targeted therapeutic interventions. Moreover, our data can be helpful for the soccer team to define the peculiar role of the players, as defender or attacker, respect to their genetic profile. Finally, an interesting point of view can be the analysis of the possible correlation of the gene SNPs variability and the role predisposition of the female soccer players.

CONFLICT OF INTERESTS

The authors have no conflict of interests.

AUTHOR CONTRIBUTIONS

ETHICS STATEMENT

We obtained the signed informed consent from all the involved subjects; after performing the DNA extraction and analysis, all data were anonymized. Dott. De Nicola and Dott. Canonico are responsible for the processing of all personal data. This study did not involve any additional pharmacological or therapeutic intervention, as compared with the routine procedures adopted by the SSC Calcio Napoli (Naples soccer team).

ORCID

Luigi Alfano http://orcid.org/0000-0002-4516-9956
Enrico Bucci http://orcid.org/0000-0002-3317-8003
Andrea Chirico http://orcid.org/0000-0001-9955-1926
Antonio Giordano http://orcid.org/0000-0002-5959-016X

REFERENCES

Abrahams, S., Posthumus, M., & Collins, M. (2014). A polymorphism in a functional region of the COL5A1 gene: Association with ultraendurance-running performance and joint range of motion. *International Journal of Sports Physiology and Performance, 9*, 583–590.

Akhmetov, A. M., II, Khakimullina, D. V., Popov, S. S., Missina, O. L., Vinogradova, & Rogozkin, V. A. (2008). Polymorphism of the vascular endothelial growth factor gene (VEGF) and aerobic performance in athletes. *Fiziologiia Cheloveka, 34*, 97–101.

Al-Habboubi, H. H., Sater, M. S., Almawi, A. W., Al-Khateeb, G. M., & Almawi, W. Y. (2011). Contribution of VEGF polymorphisms to variation in VEGF serum levels in a healthy population. *European Cytokine Network, 22*, 154–158.

Barton, J.C., & Edwards, C.Q. 1993. HFE hemochromatosis. In gene reviews((R)). M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J.H. Bean, K. Stephens, and A. Amemiya, editors, Seattle (WA).

Barton, J. C., Edwards, C. Q., & Acton, R. T. (2015). HFE gene: Structure, function, mutations, and associated iron abnormalities. *Gene, 574*, 179–192.

Bertuzzi, R., Pasqua, L. A., Bueno, S., Lima-Silva, A. E., Matsuda, M., Marquezini, M., & Saldiva, P. H. (2014). Is the COL5A1 rs12722 gene polymorphism associated with running economy? *PLoS One, 9*, e106581.

Blanchard, A., Ohanian, V., & Critchley, D. (1989). The structure and function of α-actinin. *Journal of Muscle Research and Cell Motility, 10*, 280–289.

Brodal, P., Ingjer, F., & Hermansen, L. (1977). Capillary supply of skeletal muscle fibers in untrained and endurance-trained men. *American Journal of Physiology-Heart and Circulatory Physiology, 232*, H705–H712.

Campbell, I. (2007). Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine, 26*, 3661–3675.

Chicharro, J. L., Hoyos, J., Gomez-Gallego, F., Villa, J. G., Bandres, F., Celaya, P., Jimenez, F., Alonso, J. M., Cordova, A., & Lucia, A. (2004). Mutations in the hereditary haemochromatosis gene HFE in professional endurance athletes. *British Journal of Sports Medicine, 38*, 418–421.

Cupeiro, R., Benito, P. J., Maffulli, N., Calderón, F. J., & González-Lamuño, D. (2010). MCT1 genetic polymorphism influence in high intensity circuit training: A pilot study. *Journal of Science and Medicine in Sport, 13*, 526–530.

Cupeiro, R., Pérez-Prieto, R., Amigo, T., Gortázar, P., Redondo, C., & González-Lamuño, D. (2016). Role of the monocarboxylate transporter MCT1 in the uptake of lactate during active recovery. *European Journal of Applied Physiology and Occupational Physiology, 116*, 1005–1010.

Deugnier, Y., Loreal, O., Carre, O., Duvallet, E., Zoulim, F., Vinel, J. P., … Guinot, M. (2002). Increased body iron stores in elite road cyclists. *Medicine & Science in Sports & Exercise, 34*, 876–880.

Feder, J. N., Penny, D. M., Irrinki, A., Lee, V. K., Lebron, J. A., Watson, N., … Schatzman, R. C. (1998). The hemochromatosis gene product complexes with the transferrin receptor and lowers its affinity for ligand binding. *Proceedings of the National Academy of Sciences United States of America, 95*, 1472–1477.

Fedotovskaya, O. N., Mustafina, L. J., Popov, D. V., Vinogradova, O. L., & Ahmetov, I. I. (2014). A common polymorphism of the MCT1 gene and athletic performance. *Int J Sports Physiol Perform, 9*, 173–180.

Ferioli, M., Zauli, G., Maiorano, P., Milani, D., Mirandola, P., & Neri, L. M. (2019). Role of physical exercise in the regulation of epigenetic mechanisms in inflammation, cancer, neurodegenerative diseases, and aging process. *Journal of Cellular Physiology, 234*, 14852–14864.

Ganz, T. (2005). Cellular iron: Ferroportin is the only way out. *Cell Metabolism, 1*, 155–157.

Gustafsson, T., Puntschart, A., Kaijser, L., Jansson, E., & Sundberg, C. J. (1999). Exercise-induced expression of angiogenesis-related transcription and growth factors in human skeletal muscle. *American Journal of Physiology-Heart and Circulatory Physiology, 276*, H679–H685.

Halestrap, A. P., & Wilson, M. C. (2012). The monocarboxylate transporter family--role and regulation. *IUBMB Life, 64*, 109–119.

Van Hall, G. (2000). Lactate as a fuel for mitochondrial respiration. *Acta Physiologica Scandinavica, 168*, 643–656.

Heffernan, S. M., Kilduff, L. P., Erskine, R. M., Day, S. H., Stebbings, G. K., Cook, C. J., … Williams, A. G. (2017). COL5A1 gene variants previously associated with reduced soft tissue injury risk are associated with elite athlete status in rugby. *BMC Genomics, 18*, 820.

Hoeben, A., Landuyt, B., Highley, M. S., Wildiers, H., Van Oosterom, A. T., & De Bruijn, E. A. (2004). Vascular endothelial growth factor and angiogenesis. *Pharmacological Reviews, 56*, 549–580.

Jobling, R., D'souza, R., Baker, N., Lara-Corrales, I., Mendoza-Londono, R., Dupuis, L., … Kannu, P. (2014). The collagenopathies: Review of clinical phenotypes and molecular correlations. *Current Rheumatology Reports, 16*, 394.

Juel, C., & Halestrap, A. P. (1999). Lactate transport in skeletal muscle - role and regulation of the monocarboxylate transporter. *Journal of Physiology (Cambridge, United Kingdom), 517*(Pt 3), 633–642.

Katsarou, M. S., Latsi, R., Papasavva, M., Demertzis, N., Kalogridis, T., Tsatsakis, A. M., … Drakoulis, N. (2016). Population-based analysis of the frequency of HFE gene polymorphisms: Correlation with the susceptibility to develop hereditary hemochromatosis. *Molecular Medicine Reports, 14*, 630–636.

Kikuchi, N., Fuku, N., Matsumoto, R., Matsumoto, S., Murakami, H., Miyachi, M., & Nakazato, K. (2017). The association between MCT1

T1470A polymorphism and power-oriented athletic performance. *International Journal of Sports Medicine*, 38, 76–80.

Lee, F. X. Z., Houweling, P. J., North, K. N., & Quinlan, K. G. R. (2016). How does α-actinin-3 deficiency alter muscle function? Mechanistic insights into ACTN3, the 'gene for speed. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1863, 686–693.

Liu, L. X., Lu, H., Luo, Y., Date, T., Belanger, A. J., Vincent, K. A., … Jiang, C. (2002). Stabilization of vascular endothelial growth factor mRNA by hypoxia-inducible factor 1. *Biochemical and Biophysical Research Communications*, 291, 908–914.

Luszczyk, M., Kaczorowska-Hac, B., Milosz, E., Adamkiewicz-Drozynska, E., Ziemann, E., Laskowski, R., … Antosiewicz, J. (2017). Reduction of skeletal muscle power in adolescent males carrying H63D mutation in the HFE gene. *BioMed Research International*, 2017, 1–7.

Massidda, M., Eynon, N., Bachis, V., Corrias, L., Culigioni, C., Piras, F., … Calò, C. M. (2015). Influence of the MCT1 rs1049434 on indirect muscle disorders/injuries in elite football players. *Sports Medicine - Open*, 1, 33.

Massidda, M., Bachis, V., Corrias, L., Piras, F., Scorcu, M., Culigioni, C., … Calò, C. M. (2015). ACTN3 R577X polymorphism is not associated with team sport athletic status in Italians. *Sports Medicine - Open*, 1, 6.

Massidda, M., Voisin, S., Culigioni, C., Piras, F., Cugia, P., Yan, X., … Calo, C. M. (2017). ACTN3 R577X polymorphism is associated with the incidence and severity of injuries in professional football players. *Clinical Journal of Sport Medicine*, 29(1), 57–61.

Massidda, M., Bachis, V., Corrias, L., Piras, F., Scorcu, M., & Calo, C. M. (2015a). Influence of the COL5A1 rs12722 on musculoskeletal injuries in professional soccer players. *Journal of Sports Medicine and Physical Fitness*, 55, 1348–1353.

Merezhinskaya, N., Fishbein, W. N., Davis, J. I., & Foellmer, J. W. (2000). Mutations in MCT1 cDNA in patients with symptomatic deficiency in lactate transport. *Muscle and Nerve*, 23, 90–97.

Ostrander, E. A., Huson, H. J., & Ostrander, G. K. (2009). Genetics of athletic performance. *Annual Review of Genomics and Human Genetics*, 10, 407–429.

Parkkila, S., Niemelä, O., Britton, R. S., Fleming, R. E., Waheed, A., Bacon, B. R., & Sly, W. S. (2001). Molecular aspects of iron absorption and HFE expression. *Gastroenterology*, 121, 1489–1496.

Pimenta, E. M., Coelho, D. B., Veneroso, C. E., Barros Coelho, E. J., Cruz, I. R., Morandi, R. F., … De paz fernández, J. A. (2013). Effect of ACTN3 gene on strength and endurance in soccer players. *Journal of Strength and Conditioning Research*, 27, 3286–3292.

Prior, S. J., Hagberg, J. M., Paton, C. M., Douglass, L. W., Brown, M. D., McLenithan, J. C., & Roth, S. M. (2006). DNA sequence variation in the promoter region of the VEGF gene impacts VEGF gene expression and maximal oxygen consumption. *American Journal of Physiology-Heart and Circulatory Physiology*, 290, H1848–H1855.

Rankinen, T., Fuku, N., Wolfarth, B., Wang, G., Sarzynski, M. A., Alexeev, D. G., … Bouchard, C. (2016). No evidence of a common DNA variant profile specific to world class endurance athletes. *PLoS One*, 11, e0147330.

Richardson, R. S., Wagner, H., Mudaliar, S. R. D., Saucedo, E., Henry, R., & Wagner, P. D. (2000). Exercise adaptation attenuates VEGF gene expression in human skeletal muscle. *American Journal of Physiology-Heart and Circulatory Physiology*, 279, H772–H778.

Richardson, M., & Reischman, D. (2011). The magical number 7. *Teaching Statistics*, 33, 17–19.

Richardson, R. S., Noyszewski, E. A., Kendrick, K. F., Leigh, J. S., & Wagner, P. D. (1995). Myoglobin O2 desaturation during exercise. Evidence of limited O2 transport. *Journal of Clinical Investigation*, 96, 1916–1926.

Roth, S. M., Walsh, S., Liu, D., Metter, E. J., Ferrucci, L., & Hurley, B. F. (2008). The ACTN3 R577X nonsense allele is under-represented in elite-level strength athletes. *European Journal of Human Genetics*, 16, 391–394.

Sawczuk, M., Banting, L. K., Cięszczyk, P., Maciejewska-Karłowska, A., Zarębska, A., Leońska-Duniec, A., … Eynon, N. (2015). MCT1 A1470T: A novel polymorphism for sprint performance? *Journal of Science and Medicine in Sport*, 18, 114–118.

Thompson, D. B., Pratley, R., & Ossowski, V. (1996). Human primary myoblast cell cultures from non-diabetic insulin resistant subjects retain defects in insulin action. *Journal of Clinical Investigation*, 98, 2346–2350.

Wahl, P., Jansen, F., Achtzehn, S., Schmitz, T., Bloch, W., Mester, J., & Werner, N. (2014). Effects of high intensity training and high volume training on endothelial microparticles and angiogenic growth factors. *PLoS One*, 9, e96024.

Westerblad, H., Allen, D. G., & Lännergren, J. (2002). Muscle fatigue: Lactic acid or inorganic phosphate the major cause? *Physiology*, 17, 17–21.

Yang, N., MacArthur, D. G., Gulbin, J. P., Hahn, A. G., Beggs, A. H., Easteal, S., & North, K. (2003). ACTN3 genotype is associated with human elite athletic performance. *The American Journal of Human Genetics*, 73, 627–631.

Yang, Y., Férec, C., & Mura, C. (2011). SNP and haplotype analysis reveals new HFE variants associated with iron overload trait. *Human Mutation*, 32, E2104–E2117.