ORIGINAL ARTICLE

Sports talent identification based on motor tests and genetic analysis

JAROMÍR ŠIMONEK[1], RADOSLAV ŽIDEK[2]

Abstract

Introduction. Physical tests have long been used for determining a child's athletic abilities before age 9. At that age children are not yet physically mature and their motor skills are being developed. That is why the genetic test can come in [handy] for looking for early indicators of talent in performance areas. Aim of Study. The study aims at explaining the role of physical fitness testing and genetic analysis upon identifying sport talents. We hypothesized that using physical fitness tests will not bring the results which would match up with the ones of genetic analysis. This presumption was verified using motor tests battery and gene analysis in 7 year old population. Material and Methods. The research sample included 169 pupils (97 male; mean age = 7.438 y. and 72 female; mean age = 7.227 y.) attending 3 elementary schools in the region of Nitra, Slovakia. All pupils underwent 9 physical tests to determine their general physical abilities. Each performance of pupils in tests was allotted points. Subsequently, 30 best ranked pupils were selected to undergo 2 ml saliva sampling for genetic analysis. The values of individual genetic score are compared with histogram of genetic score distribution in European population. Results and Conclusions. The study showed that the results of genetic analysis did not match the ones of the fitness tests. Based on the analysis we offered parents and coaches valid information about their children's prerequisites for certain group of sports, type of muscle fiber, oxidative capacity, nutrition type, regeneration, injury prevention, injury susceptibility, etc.

KEYWORDS: sport talent, genotype, fitness tests, children, prerequisites for sport.

Corresponding author: jsimonek@ukf.sk

[1] *Constantine the Philosopher University in Nitra, Faculty of Education, Department of Physical Education & Sport, Nitra, Slovakia*
[2] *Slovak University of Agriculture in Nitra, Faculty of Biotechnology and Food Sciences, Department of Food Hygiene and Safety, Nitra, Slovakia*

Introduction

Sport performance is a multifactorial polygenic trait influenced by sports training, environmental and genetic predisposition. Elite athletes are more or less extreme products of genetic variability and their exceptionality consists in that they are markedly different from the common average. Among the most significant factors influencing muscular strength belong: muscle fiber proportions, nutrition, genetic predisposition, age, length of muscles, and an individual's somatotype. During the growth of an athlete series of important factors concur. Among them are: cultural background, motivation, social conditions, professional coaching and sports tradition in the given country.

Physical fitness of each individual is primarily limited by his/her genotype. One of the genes, which has crucial impact on physical performance of an individual is ACTN3 gene. As to the type of an individual, it is possible to specify, with a high degree of accuracy, whether the individual can reach through an intensive training top sports performance in endurance or explosive (speed-strength) sports events, or otherwise the training is only a redundant load without an expected effect.

The degree of inherent prerequisites is very variable. Distinctions can be found also within individual preferences or inclination to art or sport. Based on available information, it is possible to assume that there exists connection between genetics and sport performance in a certain sport event. By disclosing the relationship between endogenous prerequisites of an individual and inclination to some sport event, an individual who thanks to the better genetic outfit can reach better performance than the one with different genetic fund (skeletal muscle structure, regeneration, metabolism, injury prediction, function of organs etc.). As a matter of course, it is possible to assume that sport performance itself is not built by genetic predetermination, but consists of a wide range of factors (training, technical and tactical skills etc.).

When we focus our attention to elite athletes and their prerequisites for certain types of physical activity, we can come to the conclusion that they are just somatic parameters which are to a great degree genetically influenced (over 90%). Similarly, it is with physiological criteria. Within genetic factors influencing sport performance it is assumed that there exist gene variations which essentially affect human body composition and metabolism. Speed and reaction are the qualities which are affected by heritability to the highest degree, while psychic properties and coordinative prerequisites are less impacted.

The results obtained from classical motor predisposition tests indicate the current situation in motor skills and physical condition of subjects. At present, they are aimed at integrated measurements of the effect of various genes and environmental effects on one's genotype. Genetic tests differ in principle from the traditional motor tests, because DNA of an individual does not change during his/her life. genetic information advices us of an individual's genetic predispositions in his/her early childhood. Genetics plays an important role in determining the capacity of an individual to go in for sports at the top professional level. The question is which genetic elements influence motor abilities and what is their respective significance. Furthermore, it is necessary to know the related genes as well as the mechanisms and metabolic paths and their influence. Sports performance is not determined only by a single gene but concurrent interactions of several genes [28].

With regard to the fact that results of commonly used physical tests cannot be considered objective indicators of determining degree of sport potential in the future, genetic tests can be very beneficial for identification of sports talent. Bouchard et al. [8] state that genetic testing of young athletes offers a suitable method of identifying performance prerequisites just before their development. A genetic test can inform trainers and athletes whether one´s genotype is of endurance or of speed nature. This sort of information may be combined with results of the "classical" motor tests in developing individual training programs and in discovering talented children.

Oxidative capacity or maximal O_2 uptake (VO_2max), the maximal amount of O_2 per unit of time that can be delivered to peripheral organs, including skeletal muscle, where it is used to sustain muscular contraction at peak exercise, is considered the gold standard measure of cardiorespiratory fitness. VO_2max is characterized by wide inter-individual differences even among sedentary adults. It has been suggested that improvements in cardiorespiratory fitness in response to exercise training varies greatly between individuals, with some people responding well or very well to exercise training, whereas others only have mild increases in their cardiorespiratory fitness following similar exercise training [7, 23]. The ability to change cardiorespiratory fitness is a multifactorial trait influenced by environmental factors (such as exercise training) and genetic factors [26]. The gold standard measure for cardiorespiratory fitness is maximal oxygen uptake (VO_2max), which is quantified as the maximal amount of oxygen the body can use in 1 min, during dynamic work with large muscle mass. Research into human variation of VO_2max was first undertaken over forty years ago, with several authors identifying a strong genetic influence on VO_2max in twins [18]. The strongest evidence to date on this topic was found in the HEalth, RIsk factors, exercise Training And GEnetics (HERITAGE) [6]. Also skeletal muscle regeneration is a complex process that is mediated by satellite cells, and in which several factors are activated to regulate muscle remodelling [19]. Usually, satellite cells remain quiescent but are activated following damage [13]. They proliferate 24-48 h later and then do one of three things: (1) return to quiescence and restore the population of satellite cells; (2) migrate to the site of injury and support the repair process by increasing the nuclei to cytoplasm ratio; (3) fuse with other myogenic cells to form myotubes, thus generating new fibres to replace damaged myofibers [10, 29]. Insulin-like growth factor-IEb and IGF-IEc are also known as mechano-growth factor, because the mRNA is expressed in response to overload or damage in skeletal muscle. The expression of mechano-growth factor is enhanced shortly after muscle damage, which subsequently promotes satellite cell activation [15]. Risk of injuries is mostly related with tendons and ligaments, both representing dense

connective tissues and having comparably structural and mechanical properties. They are essential constructions for the musculoskeletal system. The major duty of tendons is to deliver force from muscles to skeletons, causing the connected joints in locomotion, whereas the general function of ligaments is to stabilize joints and guide joints through their normal range of motion [5, 12]. Typical musculoskeletal soft tissue injuries including anterior cruciate ligament (ACL) injuries, Achilles tendon pathology (ATP) [16], tennis elbow, together with rotator cuff tendon injuries usually occur to individuals who endure large amounts of exercise. Collagens, the foremost elements of connective tissues and extracellular matrix, are the most abundant protein in tendons and ligaments and comprise around 75% of their dry weight. Numerous diseases have been identified caused by mutations in genes for collagens or collagen-processing enzymes [24].

Material and Methods

The research sample included 169 pupils (97 male; mean age = 7.438 y. and 72 female; mean age = 7.227 y.) attending 3 elementary schools in the region of Nitra, Slovakia. All pupils underwent 9 physical tests to determine their general physical abilities. Each performance of pupils in tests was allotted points. 30 best ranked pupils were subsequently selected to undergo 2 ml saliva sampling (GeneFix Saliva Collectors) for genetic analysis. Samples were analysed using the apparatus HiScan (Illumina Inc., San Diego, USA), which allowed for analysing 400,000 polymorphisms in a human gene. The values of individual genetic score are compared with histogram of genetic score distribution in European population. Software Genomestudio (Illumina Inc., San Diego, USA) and software TANAGRA 1.4.50 were used for data analysis.

Three separate panels of genetic markers were selected for analysis. An unweighted "genetic score" based on contribution to oxidative capacity, regeneration speed and injury risk (Table 1) has been calculated. A score of '0' represented homozygote for the low-response variant; '1' represented heterozygous and '2' represented homozygous for the high-response allele. In certain variants with significantly higher impact was the genetic score increase to "2" for heterozygous and "4" for homozygous high-response allele. Sum of genetic score calculated for each haplotype has been enriched by probability to observe specific haplotype in European (Caucasian) population. The overall population risk has been visualized as a histogram of given genetic score based on frequencies of all possible

Table 1. Markers for oxidative capacity, regeneration speed and injury risk [2-4, 9, 11, 14, 17, 20, 21, 25, 27, 30, 31]

| ID	Gene	SNP	Source
Markers for oxidative capacity			
1	GRIN3A	rs1535628	(Bouchard et al. 2011)
2	KCNH8	rs4973706	(Bouchard et al. 2011)
3	C9orf27	rs12115454	(Bouchard et al. 2011)
4	ACSL1	rs6552828	(Bouchard et al. 2011)
5	ZIC4	rs11715829	(Bouchard et al. 2011)
6	RGS18	rs10921078	(Bouchard et al. 2011)
7	BIRC7	rs6090314	(Bouchard et al. 2011)
8	DBX1	rs10500872	(Bouchard et al. 2011)
9	AKR1B1P5	rs1956197	(Bouchard et al. 2011)
10	NDN	rs824205	(Bouchard et al. 2011)
11	TTC6	rs12896790	(Bouchard et al. 2011)
12	FLJ44450	rs4952535	(Bouchard et al. 2011)
13	CAMTA1	rs884736	(Bouchard et al. 2011)
14	BLR1	rs4938561	(Bouchard et al. 2011)
15	NFIA-AS2	rs1572312	(Ahmetov & Fedotovskaya 2015)
16	TSHR	rs7144481	(Ahmetov & Fedotovskaya 2015)
17	intron	rs11254160	(Comuzzie et al. 2012)
Markers for regeneration speed			
1	IL-6	rs1800795	(Baumert et al. 2016)
2	HIF1A	rs11549465	(Ahmetov & Fedotovskaya 2012)
3	AMPD1	rs17602729	(Lucia et al. 2006)
4	NAT2	rs1208	(Buxens et al. 2011)
5	PPARD	rs2016520	(Ahmetov & Fedotovskaya 2012)
6	IL-1B	rs16944	(Baumert et al. 2016)
7	IL-1B	rs1143634	(Baumert et al. 2016)
Markers for injury risk			
1	MMP3	rs591058	(Raleigh et al. 2009)
2	MMP3	rs679620	(Raleigh et al. 2009)
3	COL1A1	rs1800012	(Wang et al. 2017)
4	COL5A1	rs12722	(Lv et al. 2018)
5	CILP	rs2073711	(Kelempisioti et al. 2011)
6	COL5A1	rs3196378	(Heffernan et al. 2017)
7	CDF5	rs143384	(Rouault et al. 2010)

haplotypes in the European population. Predictive model based on 17 single-nucleotide polymorphisms (SNPs) (oxidative capacity); 7 SNPs (regeneration speed) and 7 SNPs (injury risk) has been visualised into histograms (Figures 1, 2 and 3).

The calculated rate of genetic score has been confronted to histogram of genetic score distribution of 30 control individuals (black bordered columns) genotyped through Illumina chip HumanOmniExpress-24 (23andMe v3, v4, v5) and derived from the project NU3Gen. To enrich maximal potential of used genetic chip, some of SNP markers has been imputed. In several situations (no call, missed SNP) we used LDLINK platform [22] to find suitable replacement SNP with highest possible pairwise R^2. For the graphical representation of the distribution of the genetic score in the population, a histogram of the model distribution of the genetic score in the European population based on the frequencies of the individual alleles obtained with the 1000 genome project [1] has been calculated.

Results

Oxidative capacity is a feature that is mainly associated with trainability. At the metabolic level, it is a response to the training load that is manifested by abidance in aerobic metabolism. The results of our study point to the fact that the model distribution of the genetic score in the European population ranges from genetic score (GS) 5 to GS19. Values outside of this range are extremely rare (Figure 1).

Figure 1. Distribution of genetic score for oxidative capacity
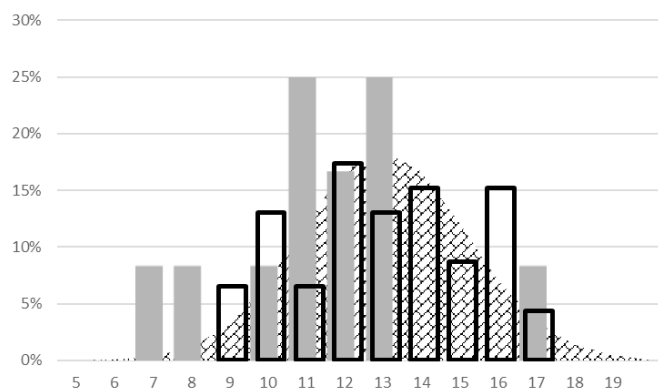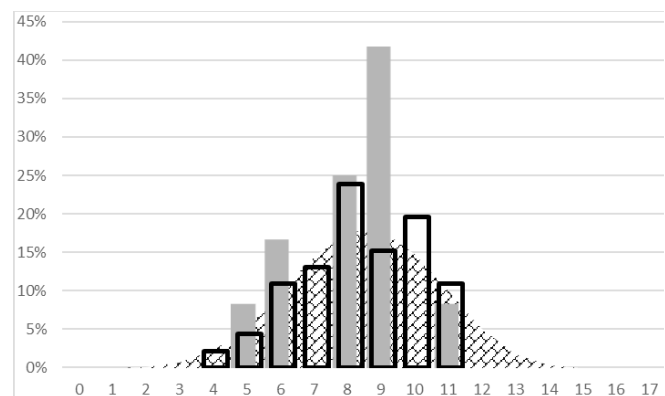

Explanatory notes: Background histogram represents distribution of oxidative capacity in European population. Black bordered columns represent distribution of GS in control group. Grey columns represent distribution of GS in selected group.

Although the maximum achievable GS value in the selected model is 34, this hypothetical value occurs in

European population with a probability of 1: 5.28×10^{21}. In contrast, the genetic score ranging from 12 to 14 can be seen in 51.9% of the European model population. The genetic score below 11 has been considered as a reduced oxidative capacity and a score above 15 as predisposition to increased oxidative capacity. As can be seen from Figure 1, the genetic score of the control population without selection is similar to the distribution of the calculated genetic score distribution in the European population.

The regeneration speed is one of the decisive factors that define the degree of potential training load of the individual. The calculated genetic score was based on the observations of several authors, with the cumulative effect of genes and the frequency of individual genes in the European population has being taken into account. The hypothetical value of the genetic score was from 0 to 17. The lowest value of 0 and the highest value of 17 in the European population occur with a probability of 1: 5.94×10^5 and 1: 1.27×10^6 respectively. The genetic score values ranging from 7 to 10 are typical for 64.1% of European populations. The genetic score below 7 was considered to be a reduced rate of regeneration and a value above 10 for an increased rate of regeneration. The values for the selected group of children were in all quartiles of the histogram (Figure 2).

Figure 2. Distribution of genetic score for regeneration speed
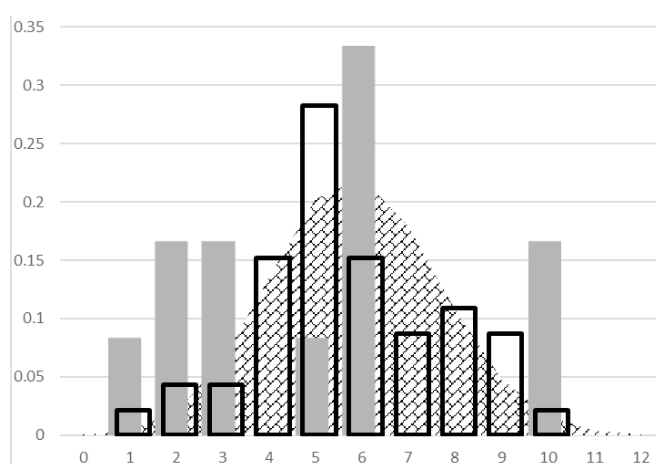

Explanatory notes: Background histogram represents distribution of regeneration speed in European population. Black bordered columns represent distribution of GS in control group. Grey columns represent distribution of GS in selected group.

To calculate the risk of injury, a model based on several studies identifying the genetic risk in athletes was constructed (Figure 3). The hypothetical range of the genetic score for risk of injury ranged from 0 to 12. The mean value in the population (59.9%) was in the range of the genetic score from 5 to 7. The genetic

score below 5 indicated the accumulation of such forms of genes that increase the risk of injuries. GS above 12 indicated a minimal manifestation of risk genes in the genotype of the individual. As outlined in Figure 3, selection of children aged 7 to 8 showed disrupted results. We identify individuals with higher risk (GS 1-3) and those with a significantly lower risk of injury (GS 10). Even in this case, we must state that selection of children based on phenotypic performance does not provide comparable results with genetic predisposition to important adult properties.

Figure 3. Distribution of genetic score for injury risk

Explanatory notes: Background histogram represents distribution of injury risk in European population. Black bordered columns represent distribution of GS in control group. Grey columns represent distribution of GS in selected group.

Discussion

The results of genetic analysis suggest that the selection of children aged 7-8 years does not provide results that would correspond to genetic predisposition to adult oxidative capacity. The selected group of children was characterized by a predominantly reduced (values of GS below 11) or average (values between 12 and 14) oxidative capacity. It is possible that some children with higher genetic potential have had poor results in test due to lack of physical development.

As for the regeneration speed, which is a crucial factor disclosing the degree of potential training load of the individual, the genetic score of the control population has been similar to predicted distribution in the European population. Distribution of genetic score between selected children shows no trend leading to increased (values above 10) genetic predisposition to better regeneration speed in adulthood. On the other hand, only 8% of the sample showed positive

values (above 11) signalling fair speed of regeneration processes.

Most genetic factors in our study were associated with an increased risk of tendon and ligament injuries. As outlined in Figure 3, selection of children aged 7 to 8 showed disrupted results. We identify individuals with higher risk (GS 1-3) and those with a significantly lower risk of injury (GS 10). Even in this case, we must state that selection of children based on phenotypic performance does not provide comparable results with genetic predisposition to important adult properties.

Conclusions

As the science of genomics progresses, there is increasing interest in the role that genetic testing might play in identifying individuals with genetic characteristics which might be advantageous in the sporting environment. There are several concerns about the effect of genetic testing on individual athletes using direct-to-consumer testing, particularly when this involves children. The lack of evidence-based interpretation of test results may result in athletes being provided with inappropriate advice about their suitability for specific sporting activities. Therefore, genetic testing is recommended mainly for predisposition to injury purpose.

Based on the results of genetic analysis and measurement of motor abilities of selected children aged 7-8 years, and in line with other authors [7, 8, 26] we can state that genetic testing of young athletes offers a suitable method of identifying performance prerequisites just before their development. Genetic tests can inform trainers and athletes on the type of physical activity (endurance or speed) suitable for the given individual. This sort of information may be combined with results of the "classical" motor tests in developing individual training programs and in discovering talented children. However, it has not been proved that identification of talent for sport is possible with the use of only one method (genetic analysis of predispositions and motor testing). Testing of children at the age of 6-7 years will not show all motor prerequisites for sport. The sole use of genetic analysis can offer only hereditary characteristic of an individual, which, however, need not be developed during the life of the person. Experts thus recommend using a combination of the above mentioned two methods for talent identification. Based on the analysis we offered parents and coaches valid information about their children´s prerequisites for certain group of sports, type of muscle fibre, oxidative capacity, nutrition type, regeneration, injury prevention, injury susceptibility, etc.

References

1. 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015; 526(7571): 68-74. doi: 10.1038/nature15393

2. Ahmetov II, Fedotovskaya ON. Current Progress in Sports Genomics. Adv Clin Chem. 2015; 70: 247-314. doi: 10.1016/bs.acc.2015.03.003

3. Ahmetov II, Fedotovskaya ON. Sports genomics: Current state of knowledge and future directions. Cell Mol Exerc Physiol. 2012; 1(1). doi: 10.7457/cmep.v1i1.e1

4. Baumert P, Lake MJ, Stewart CE, Drust B, Erskine RM. Genetic variation and exercise-induced muscle damage: implications for athletic performance, injury and ageing. Eur J Appl Physiol. 2016; 116(9): 1595-1625. doi: 10.1007/s00421-016-3411-1

5. Birch HL. Specialisation of extracellular matrix for function in tendons and ligaments. Muscles Ligaments Tendons J. 2013; 3(1): 12-22.

6. Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, et al. Familial aggregation of Vo2 max response to exercise training: results from the HERITAGE Family Study. J Appl Physiol. 1999; 87(3): 1003-1008. doi: 10.1152/jappl.1999.87.3.1003

7. Bouchard C, Sarzynski MA, Rice TK, Kraus WE, Church TS, Sung YJ, et al. Personalized Preventive Medicine: Genetics and the Response to Regular Exercise in Preventive Interventions. Prog Cardiovasc Dis. 2015; 57(4): 337-346.

8. Bouchard C, Sarzynski MA, Rice TK, Kraus WE, Church TS, Sung YJ, et al. Genomic predictors of the maximal O2 uptake response to standardized exercise training programs. J Appl Physiol. 2011; 110(5): 1160-1170. doi: 10.1152/japplphysiol.00973.2010

9. Buxens A, Ruiz JR, Arteta D, Artieda M, Santiago C, Gonzáles-Freire M, et al. Can we predict top-level sports performance in power vs endurance events? A genetic approach. Scand J Med Sci Sports. 2011; 21(4): 570-579. doi: 10.1111/j.1600-0838.2009.01079.x

10. Chazaud B. Inflammation during skeletal muscle regeneration and tissue remodeling: application to exercise-induced muscle damage management. Immunol Cell Biol. 2016; 94(2): 140-145. doi: 10.1038/icb.2015.97

11. Comuzzie AG, Cole SA, Laston SL, Voruganti S, Haack K, Gibbs RA, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. PLoS one. 2012; 7(12): e51954. doi: 10.1371/journal.pone.0051954

12. Frank CB. Ligament structure, physiology and function. J Musculoskelet Neuronal Interact. 2004; 4(2): 199-201.

13. Grobler L, Collins M, Lambert MI. Remodelling of Skeletal Muscle Following Exercise-Induced Muscle Damage. Internat SportMed J. 2004; 5(2): 67-83.

14. Heffernan SM, Kilduff LP, Erskine RM, Day SH, Stebbings GK, Cook CJ, et al. COL5A1 gene variants previously associated with reduced soft tissue injury risk are associated with elite athlete status in rugby. BMC Genomics. 2017; 18(S8): 820. doi: 10.1186/s12864-017-4187-3

15. Hill M, Goldspink G. Expression and Splicing of the Insulin-Like Growth Factor Gene in Rodent Muscle is Associated with Muscle Satellite (stem) Cell Activation following Local Tissue Damage. J Physiol. 2003; 549(2): 409-418. doi: 10.1113/jphysiol.2002.035832

16. Järvinen TAH, et al. Achilles tendon disorders: Etiology and epidemiology. Foot Ankle Clin. 2005; 10(2): 255-266.

17. Kelempisioti A, Eskola PJ, Okuloff A, Karjalainen U, Takatalo J, Daavittila L, et al. Genetic susceptibility of intervertebral disc degeneration among young Finnish adults. BMC Med Gen. 2011; 12(1): 153. doi: 10.1186/1471-2350-12-153

18. Klissouras V, Pirnay F, Petit JM. Adaptation to maximal effort: genetics and age. J Appl Physiol. 1973; 35(2): 288-293. doi: 10.1152/jappl.1973.35.2.288

19. Kurosaka M, Machida S. Exercise and skeletal muscle regeneration. J Phys Fitness Sports Med. 2012; 1(3): 537-540.

20. Lucia A, Martin MA, Esteve-Lanao J, San Juan AF, Rubio JC, Oliván J, Arenas J, et al. C34T mutation of the AMPD1 gene in an elite white runner. Br J Sports Med. 2006; 40(3): e7-e7. doi: 10.1136/bjsm.2005.019208

21. Lv Z-T, Gao ST, Cheng P, Liang S, Yu SY, Yang Q, Chen AM. Association between polymorphism rs12722 in COL5A1 and musculoskeletal soft tissue injuries: a systematic review and meta-analysis. Oncotarget. 2018; 9(20): 15365-15374. doi: 10.18632/oncotarget.23805

22. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics (Oxford, England). 2015; 31(21): 3555-7. doi: 10.1093/bioinformatics/btv402

23. Mann TN, Lamberts RP, Lambert MI. High Responders and Low Responders: Factors Associated with Individual

Variation in Response to Standardized Training. Sports Med. 2014; 44(8): 1113-1124. doi: 10.1007/s40279-014-0197-3

24. Myllyharju J, Kivirikko KI. Collagens and collagen-related diseases. Ann Med. 2001; 33(1): 7-21.

25. Raleigh SM, van der Merwe L, Ribbans WJ, Smith RK, Schwellnus MP, Collins M. Variants within the MMP3 gene are associated with Achilles tendinopathy: possible interaction with the COL5A1 gene. Br J Sports Med. 2009; 43(7): 514-520. doi: 10.1136/bjsm.2008.053892

26. Rankinen T, Bouchard C. Genetic Predictors of Exercise Training Response. Curr Cardiovasc Risk Rep. 2011; 5(4): 368-372. doi: 10.1007/s12170-011-0179-z

27. Rouault K, Scotet V, Autret S, Gaucher F, Dubrana F, tanguy D, et al. Evidence of association between GDF5 polymorphisms and congenital dislocation of the hip in a Caucasian population. Osteoarthritis and Cartilage. 2010; 18(9): 1144-1149. doi: 10.1016/j.joca.2010.05.018

28. Sessa F, Chetta M, Petito A, Franzetti M, Bafunno V, Pisanelli D, et al. Gene polymorphisms and sport attitude in Italian athletes. Genet Test Mol Biomarkers. 2011; 15(4): 285-290.

29. Tidball JG, Villalta SA. Regulatory interactions between muscle and the immune system during muscle regeneration. Am J Physiol Regul Integr Comp Physiol. 2010; 298(5): R1173-87. doi: 10.1152/ajpregu.00735.2009

30. Wang C, Li H, Chen K, Wu B, Liu H. Association of polymorphisms rs1800012 in COL1A1 with sports-related tendon and ligament injuries: a meta-analysis. Oncotarget. 2017; 8(16): 27627-27634. doi: 10.18632/oncotarget.15271

31. Williams CJ, Williams MG, Eynon N, Ashton KJ, Little JP, Wisloff U, et al. Genes to predict VO2 max trainability: A systematic review. BMC Genomics. 2017: 18(8).