Detection of epistasis between *ACTN3* and *SNAP-25* with an insight towards gymnastic aptitude identification

ACTN3 – SNAP-25 interaction in the context of athleticism

Łukasz Płóciennik[1,6¶*], Jan Zaucha[2], Jan Maciej Zaucha[3], Krzysztof Łukaszuk[4], Marek Jóźwicki[5], Magdalena Płóciennik[6], Paweł Cięszczyk[1]

[1] Department of Physical Education, Academy of Physical Education and Sport in Gdansk, Gdansk, Pomorskie Voivodeship, Poland,

[2] Department of Bioinformatics, Wissenschaftszentrum Weihenstephan, Technische Universität München, Freising, Germany,

[3] Department of Haematology and Transplantation, Medical University of Gdansk, Gdansk, Pomorskie Voivodeship, Poland,

[4] Faculty of Health Sciences with Institute of Maritime and Tropical Medicine, Medical University of Gdansk, Gdansk, Pomorskie Voivodeship, Poland,

[5] Department of Architecture and Design, Academy of Fine Arts, Gdansk, Pomorskie Voivodeship, Poland,

[6] FitnessFitback, Pomorskie Voivodeship, Poland.

[*]Corresponding Author: Łukasz Płóciennik (lukaszplociennik.awf@wp.pl)

Abstract

In this study, we performed an analysis of the impact of performance enhancing polymorphisms (PEPs) on gymnastic aptitude while considering epistatic effects. Seven PEPs (*rs1815739*, *rs8192678*, *rs4253778*, *rs6265*, *rs5443*, *rs1076560*, *rs362584*) were considered in a case (gymnasts) – control (sedentary individuals) setting. The study sample comprised of two athletes' sets: 27 elite (aged 24.8 ± 2.1 years) and 46 sub-elite (aged 19.7 ± 2.4 years) sportsmen as well as a control group of 245 sedentary individuals (aged 22.5 ± 2.1 years). The DNA was

derived from saliva and PEP alleles were determined by PCR, RT-PCR. Following Multifactor Dimensionality Reduction, logistic regression models were built. The synergistic effect for *rs1815739 x rs362584* reached 5.43%. The *rs1815739 x rs362584* epistatic regression model exhibited a good fit to the data (Chi-squared = 33.758, $p \approx 0$) achieving a significant improvement in sportsmen identification over naïve guessing. The area under the receiver operating characteristic curve was 0.715 (Z-score = 38.917, $p \approx 0$). In contrast, the additive *ACTN3 – SNAP-25* logistic regression model has been verified as non-significant.

We demonstrate that a gene involved in the differentiation of muscle architecture – *ACTN3* and a gene, which plays an important role in the nervous system – *SNAP-25* interact. From the perspective originally established by the Berlin Academy of Science in 1751, the matter of communication between the brain and muscles via nerves adopts molecular manifestations. **Further in-vitro investigations are required to explain the molecular details of the *rs1815739 – rs362584* interaction.**

Introduction

By 1798, Luigi Galvani discovered two phenomena: muscle stimulation by extrinsic electricity and a genuine potential difference between the nerve and the muscle. These findings lead his successors to investigate the details of the electrical influence on nerve function in the context of muscle movement. By now, the scientific community has reached the molecular level of understanding the mechanisms involved and have already honed in on the genomic loci affecting athleticism. As a result, multiple single nucleotide polymorphisms (SNPs) have been implicated in affecting the aptitude for gymnastics. To move beyond simple SNP associations, genetic epistasis modeling may enhance the understanding of sports performance. Authors investigating genetic interactions typically rely only on genotype frequency odds ratios [1-3] or perform Genome-Wide Interaction Analyses (GWIA) employing tests visualized by pseudo-Manhattan plotting. So far, the matter of epistasis has been investigated for: (a) the Body Mass

Index (BMI) [4]; (b) physical activity in mice [5]; (c) medical disorders in clinical studies [6]; in ischemic stroke susceptibility [7].

Variant interactions including synergy or redundancy have not yet been considered in the context of predicting athletic performance. [8-9]. Instead, the total genotype score (TGS) for distinguishing athletes has been calculated several times in different research projects [10-11]. Unfortunately, TGS models do not consider interactions between polymorphisms, i.e., their synergy and redundancy [11]. The main strength of pure epistatic models is their potential for deciphering the genetic variation of predisposed athletes *ab initio*. Interestingly, ensemble-based classifiers [12], which are free of external attributes, have so far yielded better predictions than alternative approaches incorporating environmental effects into the model.

The genetic foundations of muscle performance are explored by mathematical modeling. While parametric techniques, such as logistic regression (LR) are limited in their ability to characterize the multivariate architecture of complex phenotypes, information theory provides a solution for quantifying the information gain between different statistical models of inference. The relative difference in Shannon entropy i.e. the Kullback-Leibler divergence (also known as information gain - IG) allows selecting the optimal approach for modeling the genetic effects on phenotype. Additionally, Multifactor Dimensionality Reduction (MDR), a non-parametric statistical technique allows detecting interactions between attributes of the model. In this work, we applied this method to detect epistasis in a set of candidate genes,

Artistic gymnastics is one of many sport disciplines, which has not been extensively studied with regard to its genetic underpinnings. Notwithstanding the exact definition of the proportion of speed and strength to power output, gymnastics is definitely a highly polygenic anaerobic event, dependent on multiple, potentially interacting genetic variants.

The seven PEPs that were evaluated in this study include: (1) *rs1815739*, located within the *ACTN3* gene is involved in muscle contractions [13]; (2) *rs8192678*, located within the

PPARGC1A gene is responsible for the variability in power output; the substitution of glycine for serine at position 428 was reported to hinder performance in endurance activities [14]; (3) *rs4253778,* located within the *PPARα* gene appears to be associated with the hypertrophic effect due to its effects on the cardiac and skeletal muscle substrate utilization [15]; (4) *rs6265,* located within the *BDNF-AS* gene is highly correlated with learning and the development of memory-related hippocampal neurons; (5) *rs5443,* located within the *GNB3* gene seems to be a candidate for explaining the variability in exercise phenotypes [16, 17]. Specifically, the proportion of the *TT* genotype is more pronounced in the top-level endurance athletes as compared with the sprinter group. Hence, G protein activity may affect the likelihood of becoming a top-level endurance athlete [17]; (6) *rs1076560,* located within the *DRD2* gene can predispose athletes to better performance in Australian Rules Football; it allows for specific talent identification and has been linked with motor coordination and learning [18]; (7) *rs362584,* located within the *SNAP-25* gene was found to be associated with cognitive ability [19] and with the cognitive disorder [20]. Furthermore in 2015, Islamov et al. [21] have shown that *SNAP-25* is synthesized in the motor nerve endings, and affects motor neurons of the spinal cord. The aforementioned PEPs were analyzed with regard to epistasis in the context of gymnastics and evaluated in terms of their ability to discriminate between athletes and non-athletic individuals.

Results

Quality control of SNPs called

The minor allele frequency (MAF) for every candidate SNP was no less than 16.5%, which was the lowest value for the case of *rs4253788* (*PPARα*) – in the control group (S2 Supplementary Material, Table S2). All of the seven genetic polymorphisms were in Hardy-Weinberg equilibrium (HWE; H_0: $\chi^2 \leq 6.635_{(0.01;\,1)}$).

Models adjustment according to genetic markers

All SNPs under consideration were coded according to the values of the odds ratios for heterozygote, homozygote of major allele and for homozygote of minor allele (odd_{Mm}, odd_{MM}, and odd_{mm}) extracted from contingency tables [22] (S3 Supplementary Material, p. 2). Data in Table 1 indicates the odds ratios obtained for different genetic models.

Table 1. Model adjustment according to examined SNPs.

SNP / OR	*ACTN3*	*PPARGC1A*	*PPARα*	*BDNF-AS*	*GNB3*	*DRD2*	*SNAP-25*
OR1[a]	1.043	1.225	0.475	1.139	0.999	0.749	0.808
OR2[b]	1.083	1.114	0.949	0.990	1.309	0.570	1.319
Model	Multiplicative	Additive	Dominant	Dominant	Recessive	Multiplicative	Over-dominant

[a] OR1 = odds ratio for heterozygote (Mm)/odds ratio for homozygote of major allele (MM);

[b] OR2 = odds ratio for homozygote of minor allele (mm)/ odds ratio for heterozygote (Mm).

Entropy analysis

Next, the statistical significance has been calculated for each polymorphism's ability to distinguish between the case (athletes) and control (non-athletes) groups. The strongest effect observed for any single locus was for *PPARGC1A*. Its normalized information gain (IG) reached the value of 0.0065 bits (0.65%). It was the largest univariate factor reducing entropy with a borderline significance at p = 0.07 (at $\chi^2 = 5.317$). Table 2 presents IGs and p-values of all genetic markers in the performed analysis:

Table 2. Information gain values of studied genetic attributes.

Measures	*ACTN3*	*PPARGC1A*	*PPARα*	*BDNF-AS*	*GNB3*	*DRD2*	*SNAP-25*
IG[a] [bit]	0.0017	0.0065	0.0043	0.0017	0.0020	0.0023	0.0001
G^2	0.665	5.317	1.681	0.665	0.782	0.899	0.039
p-value	0.717	0.070	0.431	0.717	0.676	0.638	0.981

[a] IG – information gain (S3 Supplementary Material, Eq. 3);

[b] G^2 – G-square statistics (S3 Supplementary Material, Eq. 4).

Multifactor dimensionality reduction

Next, a genetic dendrogram has been constructed, using Rajski's distance, Ward's method and Lance and Williams recursive algorithm (S3 Supplementary Material, pp. 3-4). As a consequence, synergistic (red connections) and redundant effects have been determined (Figure 1). The analysis shows that polymorphisms are grouped into two clusters and two independent genetic pools of variants, namely: *PPARα*, *PPARGC1A – GNB3* and *BDNF*, *DRD2 – ACTN3 – SNAP-25*.

Figure 1. A gene-gene interaction dendrogram in sports gymnastics performance.[a]

[a]Orange line indicates weak positive interaction between clusters. Golden connections suggest the independence of *PPARα*, *BDNF*.

Epistasis between pairs of SNPs was evaluated in terms of the interaction information (I) between SNPs A and B in the context of class C: I(A; B; C), with positive values corresponding to synergy while negative values indicating a redundancy (correlation) of the markers [23]. The only strong synergistic effects were found between *ACTN3 – SNAP-25* and *PPARGC1A – GNB3*, represented by 0.0543 bits of interaction information (5.43%) and 0.0364 bits (3.64%), respectively. However, little evidence corroborates other possible two-way interactions. A positive moderation has been detected for twenty out of twenty-one combinations. The highest values regard *PPARGC1A – SNAP-25* (0.0523 bits - 5.23%), *ACTN3 – PPARα* (0.298 bits - 2.98%) and *GNB3 – BDNF* (0.027 bits - 2.70%). The only negative interaction was between *SNAP-25* and *PPARα*; this pair of SNPs diminishes 0.0001 bits of information about sports gymnastics. The results presented above support the alternative hypothesis stipulating the existence of a synergistic effect (e.g. for *ACTN3* and *SNAP-25*) in the set comprised of twenty-one possible two-way interactions between *rs1815739*, *rs8192678*, *rs4253778*, *rs6265*, *rs5443*, *rs1076560*, *rs362584*.

Next, a filtering technique (S3 Supplementary Material, Eq. 8) has been applied to identify the best epistatic framework The optimal model has been obtained for the combination of *ACTN3 – PPARGC1A – PPARα – SNAP-25*. Its performance is summarized in table 3.

Table 3. Test set results obtained for the *ACTN3 – PPARGC1A – PPARα – SNAP-25* epistatic model selected to maximize balanced accuracy in 10-fold cross validation.

BAL. ACC.[a]	ACC.	SENSIT.	SPECIF.	OR /CI	χ^2	χ^2 p-val.	PRE.[b]	KAPPA	F[c]	CVC[d]
0.712	0.692	0.75	0.674	6.211/ 0.840 ; 45.938	3.652	= 0.056	0.403	0.326	0.525	10/ 10

[a] BAL. ACC. – balanced accuracy; [b] PRE. – test precision; [c] F – F1-statistics;
[d] CVC – cross validation consistency (count).

MDR analysis confirmed the statistical significance (p = 0.001) of the model by comparing the value of the sign test against 1000 random permutations of the data, assuming no association under the null hypothesis. The model achieved a balanced accuracy (weighting case and control samples so as to simulate an equal sample size in each group) of 0.712. The odds ratio of positivity within the gymnasts' group relative to the controls is equal to 6.2. Interestingly, the p-value of the model estimated from the χ^2-test achieved only borderline significance, confirming previous concerns about the reliability of the p-value obtained from the MDR analysis sign-test [24]. Nevertheless, the precision is above 40% and Cohen's Kappa at 0.326 indicates a performance, which significantly surpasses naïve guessing. With regard to perfect precision and recall, the classifier is positioned in the middle of the achievable spectrum: F1-measure = 0.525. The training and whole data models are even more convincing (S2 Supplementary Material, Table S5, S6), since χ^2 p–values retained significance after Bonferroni's correction for multiple hypothesis testing. Nevertheless, we do not have definitive evidence that that the null hypothesis can be rejected.

Logistic regression analysis

For a simultaneous examination of the first and second order effects in the *ACTN3 – PPARGC1A – PPARα – SNAP-25* interaction, logistic regression with backward variable selection has been adopted. Since this analysis yielded empty combinations, two-way interactions were considered first. Contrasts between genotype categories were expressed in terms of cross-partial derivatives. To ensure the interpretability of the results for unbalanced classes, we used weighted effect coding (WEC). Interestingly, none of the other known mathematical and statistical coding structures apart from WEC allows detecting pure genetic interaction (S2 Supplementary Material, Table S1). In particular, such phenomenon has been confirmed between *ACTN3* and *SNAP-25*, when setting the homogenous derived (alternative) allele category as the reference (Table 4):

Table 4. The full *ACTN3 – SNAP-25* model with the derived allele reference category.

Constant / Genotypes	*b* weights	CI 0.95 ±	St. errors	χ^2	p-values
Intercept	-1.445	0.337	0.171	8.448	0.004**
$b_{(ACTN3)}$ *heterozygous* (*RX*)	-0.082	0.313	0.159	0.518	0.471
$b_{(ACTN3)}$ *ancestral* (*RR*)	0.006	0.524	0.266	0.024	0.876
$b_{(SNAP\text{-}25)}$ *heterozygous* (*GA*)	-0.064	0.388	0.197	0.326	0.568
$b_{(SNAP\text{-}25)}$ *ancestral GG*	0.089	0.372	0.189	0.473	0.492
$b_{1(ACTN3),1(SNAP\text{-}25)}$ *heterozygous – heterozygous*	-0.805	0.317	0.161	4.351	0.037*
$b_{(ACTN3),(SNAP\text{-}25)}$ *heterozygous – ancestral*	0.674	0.305	0.155	4.351	0.037*
$b_{(ACTN3),(SNAP\text{-}25)}$ *ancestral – heterozygous*	1.39	0.74	0.376	3.694	0.055*
$b_{(ACTN3),(SNAP\text{-}25)}$ *ancestral – ancestral*	-0.876	0.386	0.196	4.479	0.034*

b_i – SNP marginal effect; b_{ii} – 2-way G-G interaction product term,** Significant at p≤ 0.01, * significant at p≤ 0.05 to second decimal place.

The baseline OR for being a highly qualified gymnast equals 0.24, when carrying the most common genotype. Maximal log-likelihood for the estimated model totalled -133.857 with χ^2-score of 34.344 (df = 8) and p-value ≈ 0.000. Although the model explains genetic

foundations for sub-elite versus elite gymnasts' recognition in just 11% (pseudo $R^2 = 0.114$), we accept the global alternative hypothesis – $H1_e$, which states that at least one product term between PEPs is significantly different than zero. Considering the WEC data arrangement, the main effects of the model can be considered as non-significant being an order of magnitude less than the interaction weights, which are all below or equal 0.05*. Thus, individual beta weights (b_i) for *ACTN3* and *SNAP-25* are ≈ 0 and obeying statistical parsimony, we reject the null hypothesis. Next, we performed logistic regression for *rs1815739* and *rs362584* without first-order effects. Typically, in WEC, weights of regression coefficients do not change when the reference category is switched. The same applies to maximal log-likelihood statistics. Hence, we present different models (grouped according to reference genotype category) of interactions between genotypes in Table 5:

Table 5. The *ACTN3 – SNAP-25* interaction models.

Constant / Genotypes	*b* weights	CI 0.95 ±	St. errors	χ^2	p-values
Intercept	-1.445	0.337	0.171	8.448	0.004**
The model for the minor (*XX*, *AA*) allele reference category					
$\boldsymbol{b_{1,1}}$ *heterozygous – heterozygous*	-0.805	0.317	0.161	4.351	0.037*
$\boldsymbol{b_{1,2}}$ *heterozygous – ancestral*	0.674	0.305	0.155	4.351	0.037*
$\boldsymbol{b_{2,1}}$ *ancestral – heterozygous*	1.39	0.74	0.376	3.694	0.055*
$\boldsymbol{b_{2,2}}$ *ancestral – ancestral*	-0.876	0.386	0.196	4.479	0.034*
The model for the heterozygous reference category					
$\boldsymbol{b_{1,1}}$ *derived – derived*	-1.377	0.854	0.434	3.171	$0.075^{\dagger}$
$\boldsymbol{b_{1,2}}$ *derived – ancestral*	-0.099	1.323	0.672	0.147	0.701^{ns}
$\boldsymbol{b_{2,1}}$ *ancestral – derived*	2.089	1.726	0.877	2.382	$0.123^{\ddagger}$
$\boldsymbol{b_{1,2}}$ *ancestral – ancestral*	-0.876	0.386	0.196	4.479	0.034*
The model for the ancestral (*RR*,*GG*) reference category					
$\boldsymbol{b_{1,1}}$ *derived – derived*	-1.377	0.854	0.434	3.171	$0.075^{\dagger}$
$\boldsymbol{b_{1,2}}$ *derived – heterozygous*	0.809	0.535	0.272	3.179	$0.085^{\dagger}$
$\boldsymbol{b_{2,1}}$ *heterozygous – derived*	1.000	1.01	0.513	2.974	$0.163^{\ddagger}$
$\boldsymbol{b_{2,2}}$ *heterozygous – heterozygous*	-0.805	0.317	0.161	4.351	0.037*

b_{ii} – 2-way G-G interaction product term,** Significant at $p \leq 0.01$, * significant at $p \leq 0.05$ to second decimal place, † significant at $p < 0.1$, ‡ significant at $p \leq 0.1$ to first decimal place, ^{ns} – non significant.

In agreement with previous results, all interaction effects from the model for *ACTN3 – SNAP-25*, with the derived (minor allele) genotype set as the weighted reference category are significant. Moreover, G-G homogenous derived genotype, ancestral-derived and heterozygous (*XX*,GA) interaction genotypes also show considerable effects, at the edge of the p-value threshold for statistical significance. Maximal log-likelihood for the interaction model for the homogenous derived allele reference category has reached the value of -134.150. The χ^2 statistic was equal to 33.758 (df = 4) and pseudo R^2=0.112 giving a p-value < 0.00001. According to the model, the pure minor allele (*XX,AA*) genotype has the strongest negative influence. Thus, it determines the context for the other interactions. In our analysis, $b_{1,1}$, $b_{1,2}$, $b_{2,1}$, $b_{2,2}$ reached the p-value of 0.05 for the derived allele reference category (Table 5). The statistical significance was retained after applying Bonferroni's correction for multiple tests (p-value$_{\alpha/2}$ = 0.001). **In the light of this fact, three-way and multi-way interactions have not been examined**.

Particularly noteworthy is that the pure epistatic logistic regression model achieved much better performance as compared with the additive-only model. When removing all second-order derivatives, the maximal log-likelihood for the *rs1815739* + *rs362584* combination is -150.688 and becomes non-significant with a p-value of 0.409.

The results obtained from the MDR and LR analyses revealed a remarkable crosstalk between *ACTN3 – SNAP-25* polymorphisms. Disappointingly, the $b_{heterozygous,heterozygous}$ and $b_{ancestral,ancestral}$ coefficients are attributed with negative weights; presumably, in both cases a low ratio of gymnasts to sedentary individuals (5/49 and 6/70, respectively) cause these effects (S2 Supplementary Material, Table S4). Nevertheless, homogenous minor allele (*XX,AA*) genotype hosts represent the lowest chance of classification to the gymnast group: 0.059. Taking this genotype as the reference, the modeled *ACTN3 – SNAP-25* interaction effects allow rejecting the null hypothesis of no interaction.

Based on the training set, the classification performance for the interaction model without additive terms, with the *XX – AA* allele reference category and multiplicative entries arranged according to WEC achieved the area under the ROC curve (AUC-ROC) of 0.715 (95% CI: 0.647 – 0.782; Z-score = 38.917, p-value ≈ 0.000) with a standard error (Se) of AUC-ROC = 0.034. The cut-off point was selected by maximizing the Youden index = TPF-FPF and was equal to 0.379 (Figure 2). Although the achieved classification accuracy offers good specificity and is already satisfactory to aid gymnasts' recognition, the Cohen's Kappa statistic is fair (27.2%) and F1-measure totals 0.498.

Figure 2. The area under the curve (AUC-ROC) and cut-off point for the epistatic *rs1815739 * rs362584* model based on the training dataset.

When applied to the test hold-out dataset (n = 36), our classifier has correctly classified four athletes and fifteen sedentary individuals, yielding an accuracy of 52.78%. This is unsatisfactory for the purpose of supporting decision-making in sub-elite or elite gymnasts' identification. The observed AUC-ROC (0.715) and measure of Se AUC-ROC (0.034), despite being highly significant (p-value ≈ 0.000) has limited potential to confer these genetic variants as predictors for athlete's discrimination in the light of the obtained Kappa statistics and F1-measure. Further studies comprising larger samples may assert the status of these variants as informative for the task of gymnasts' identification. However, our results do not allow rejecting the null hypothesis.

Worth reporting are other insights shed by the LR and WEC data organization for the *ACTN3 – PPARα*, *PPARGC1A – SNAP-25*, *PPARGC1A – GNB3*, *GNB3 – BDNF* interactions. The contingency table for *ACTN3 – PPARα* and *GNB3 – BDNF* exposed empty cell or singular representatives in genotype categories. Consequently, data were not processed any further for these models. Fortunately, the same did not apply, when *PPARGC1A – SNAP-25* and *PPARGC1A – GNB3* were considered. Both pairs of SNPs were

247 annotated with four statistically significant weights (p-value ≤ 0.05) for the same second-order
248 product terms: *PPARGC1A – SNAP-25*: $b_{GlyGly,GA}$ (*SerSer,GG* reference (ref.) genotype:
249 favorable), *PPARGC1A – GNB3*: $b_{GlyGly,CT}$ (*SerSer,CC* ref. group: favorable),
250 *PPARGC1A – SNAP-25*: $b_{GlyGly,GG}$ (*GlySer,GA* ref. heterozygous), *PPARGC1A – GNB3*:
251 $b_{GlyGly,CC}$ (*GlySer,CT* heterozygous reference group), *PPARGC1A – SNAP-25*: $b_{GlySer,GG}$,
252 $b_{SerSer,GA}$ (*GlyGly,AA* ref. disfavorable), *PPARGC1A – GNB3*: $b_{GlySer,CC}$, $b_{SerSer,CT}$
253 (*GlyGly,TT* reference group: disfavorable). The maximal log-likelihood value was -129.97 and
254 -139.52, respectively. Nevertheless, the first-order effects remain insignificant for all possible
255 pairwise combinations of SNPs. Further non-trivial effects of cross-partial G-G interactions
256 obtained from eighteen other coding schemes applied to LR are in Supplementary Material S2.

257 ## Discussion

258 ## The biological and sport science perspective

259 The ultimate goal in sport is the athletic outcome, which correlates strongly with the
260 level of physical fitness (with psychological effects playing a secondary role). An important
261 theoretical aspect of predicting, which individuals are genetically predisposed to athleticism
262 regards establishing which allele encoding schemes allow for the most faithful discrimination
263 between athletically-gifted and ungifted individuals. Apart from fundamental, molecular types
264 of genotype ordering, we evaluated nineteen classic (statistical and mathematical) notations to
265 describe SNPs (list available in Supplementary Material S2). On the basis of planned contrasts
266 [25], taking the trend and non-trend approaches [25], all possible ways of raw genetic data
267 encoding have been processed to detect epistatic interactions. So far, there have been no studies
268 in which genetic epistasis has been investigated using so many different encoding schemes.
269 Most authors do not recognize this possibility and are reporting G-G interactions by means of
270 LR but without considering cross-partial derivatives and using unspecified coding schemes
271 [26-27]. Nonetheless, a growing body of literature has discussed ways of combining non-

parametric and parametric techniques with the goal of examining epistasis. A comprehensive attempt at investigating molecular interactions has been performed by Manuguerra et al. [28]. Similar to our research, these authors have presented, apart from a measure of CVC and p-values, a prediction error percentage of low and-high risk instances for given G-G models and odds ratio reports to determine the probability of false-positive predictions. Besides, it is worth noting that Wu et al. [29] have performed an analysis considering relationships between genotypes internally but also with environmental variables. Unfortunately, no information has been given on the categorical coding scheme. Only a general linear assignment was presented, which enabled us to determine the class that was used as the reference. Also, Dasgupta et al. [30], inform on gene – environmental interaction odds ratios based on MLR without considering regression coefficients. Nevertheless the essential result summarizing protective and risk-conferring alleles has been delineated. Bottema et al. applied LR to confirm interactions identified by means of MDR. Of the epistatic interactions they identified, MDR indicated that most interactions were synergistic [31]. However, the negative gene – gene interactions in the logistic regression of two-locus models suggest that polymorphisms of these genes counteract the effect of one another.

In this study we provide multiple lines of evidence indicating an interaction between *ACTN3* and *SNAP-25*. To the best of our knowledge, no previous study has reported such a relationship. Furthermore, notwithstanding the context of gymnast recognition, no data suggesting any kind of interaction between *ACTN3* and *SNAP-25* is available in String-db [32]. However, based on the outcome of the multidimensional stimulation therapy - MST intervention, neurophysiological studies have indicated the possibility of epistatic interactions between *APOE* and *SNAP-25* [33]. Interestingly, the interaction between *ACTN3* and *APOE* has been studied to explain the potential for exceptional longevity [34]. So far, with regard to

296 sports science, an epistasis of *ACE ID* and *ACTN3 R577X* polymorphisms has been determined,
297 e.g. in swimmers – sprint and endurance performance [2].

298 In order to detect epistatic interactions Wei et al. [4] applied MLR and demonstrated
299 two-way G-G effects affecting the body mass index (BMI) based on a genome-wide analysis.
300 Specifically, interactions between the 19 shared epistatic genes (defined as these, which
301 represent significant SNP interactions across cohorts) and those involving BMI candidate loci
302 were tested across five populations (p-value < 5.0E-08). Ultimately, eight replicated SNP pairs
303 were found in at least one cohort (p-value < 0.05) and no beta coefficients were detailed.

304 An interaction can also be recognized as product term, e.g. second-order parameter in
305 logistic model under the assumption of linear coding. This technique has been used by
306 Lee et al. [35] for testing the interaction between *EOT-2* and *CCR3* genes. The authors found
307 that an *EOTAXIN-2* gene variant: *EOT-2+304C>A* (29L>I), was significantly associated with
308 blood eosinophilia (p = 0.0087) by the effect of *CCR3* = -0.68. Nevertheless, no information
309 was presented on logistic regression main effects. Potentially, an analysis of first-order
310 parameters in the LR model may be essential to verify pseudo R^2 performance. In comparison
311 all marginal weights of the full *ACTN3 – SNAP-25* model are insignificant and the benefit from
312 applying the additive – multiplicative paradigm to gymnasts recognition is just 2‰.
313 Likewise, the subject of interaction has been studied for the *rs12722* and the *rs13946* in
314 *COL5A1* gene to assess a risk of the anterior cruciate ligament rupture in soccer players and
315 controls [36]. Unfortunately, with regard to sportsman diagnosis or prognosis no details have
316 been given on classification accuracy.

317 The *ACTN3 – SNAP-25* interaction allows explaining 11% of the variance between
318 high-level sports gymnasts. Bearing in mind that genetic factors typically explain between 20%
319 - 80% variation in a wide variety of traits relevant to athletic performance [37], the G-G epistasis
320 detailed in this paper should not be neglected in future investigations.

Methodological aspects

Several details of our analysis deserve particular attention. Firstly, considering the multiplicative – over-dominant scheme of epistasis between *ACTN3* and *SNAP-25*, the theoretically desirable ancestral – ancestral ($b_{ancestral,ancestral}$) or heterozygous – heterozygous ($b_{heterozygous,heterozygous}$) genotype carries a negative value. However, assuming disordinal interactions, there may be a region of non-significance [38], wherein there is a range of values for which no epistatic effect occurs. Secondly, possible signs change might occur for non-linear models even in the absence of an interaction [39]. These exist rational explanations for our results concerning $b_{heterozygous,heterozygous}$ and $b_{ancestral,ancestral}$. The third aspect concerns the data distribution. There were very few instances of gymnasts, who carried two heterozygous or dominant alleles for *ACTN3* and *SNAP-25*. An additional corroboration of our results is the fact that the *gene * gene* interaction at the *rs1815739* and *rs362584* loci was detected by means of both: non-parametric and parametric tests. Here, after correction for multiple testing, statistical significance was far below the restrictive threshold. Finally, in terms of probability calculus, an additive only model: *ACTN3 + SNAP-25* is not significant. **Consequently, our results have interesting implications, which explain the underlying molecular details coordinating the neuromuscular system, which has been first studied by Luigi Galvani in the 18th century.** Finally, we would like to stress that further studies concerning the *ACTN3 * SNAP-25* interactions should be conducted while considering two other levels of epistasis (suppressive, co-suppressive) [40].

The gymnasts identification context

Despite significant results corroborating the identified genetic interaction, the resultant model for discriminating between athletes and non-athletes does not yet allow for making fully reliable predictions (Figure 2). In terms of prognosis, even a single genotype of a genetic polymorphism may be introduced as a biomarker of prevalence risk, like has been done for

ischemic stroke [7]. Similarly, in our opinion, the *PPARGC1A* gene (Table 2) might be considered for diagnostic purposes. However, its usefulness in the context of gymnasts recognition has not been so far confirmed. Finally, we also observed a nominal statistical G*G partial interaction of *PPARGC1A* – *SNAP-25* and *PPARGC1A* – *GNB3* based on the gymnast status, which is interesting in the context of the studies that have associated these loci with effects relating to sport [14, 16-17, 19-21].

Conclusions

Our analysis of seven PEPs (*ACTN3*, *PPARGC1A*, *PPARα*, *BDNF-AS*, *DRD2*, *GNB3*, *SNAP-2*), allows us to state with 93% confidence that the *rs819267* provides as much as 0.0065 bit of information on sports gymnastics. The molecular dendrogram of gymnastics aptitude indicated the strongest connection between *rs1815739* and *rs362584*: 5.43% with a significant threshold of ≈ 0.000, when the homogenous derived allele category is set as the reference group. According to the findings, the best MDR epistatic model of sports gymnastics comprises of: *ACTN3* – *PPARGC1A* – *PPARα* – *SNAP-25* (the cross validation consistency equals 100%). Manifestly, when considering all pairwise combinations between *ACTN3*, *PPARGC1A*, *PPARα*, *BDNF-AS*, *DRD2*, *GNB3*, *SNAP*-25, the results confirm that only the second order terms of sports gymnastics epistatic models are non-zero. Lastly, out of the set of *ACTN3*, *PPARGC1A*, *PPARα*, *BDNF-AS*, *DRD2*, *GNB3* and *SNAP-25* genes, the most informative epistatic classifier – *rs1815739 x rs362584* is statistically significant in the context of sportsman recognition.

Materials and Methods

Ethic Committee

The study was approved by The Pomeranian Medical University Ethics Committee, Poland (Approval number 09/KB/IV/2011). Research procedures were run according to the World Medical Association Declaration of Helsinki. An informed consent form was completed

371 by each participant or obtained from a parent / legal guardian (in the case of minors) in
372 accordance with current Polish, Italian and Lithuanian law.

373 **Participants**

374 A Seventy three sportsman and two hundred forty five sedentary, non-active individuals
375 met the inclusion criteria and comprised a group for this study. They had no records of
376 metabolic, cardiovascular diseases or musculoskeletal injuries. The subjects were non-smokers
377 and did not take any medications. The cohort participants volunteered in Poland, Italy,
378 Lithuania between 2012 and 2017. All participants were unrelated European men (59.4%)
379 or women (40.6%), and all of European descent (as self-reported) for ≥ 3 generations.
380 Therefore, the influence of an ethnically-induced genetic skew has been minimized and
381 the potential population stratification issues have been controlled (S1 Study Protocol, p. 4, 5).
382 The study sample included 34 females and 39 males in two homogenous athletes groups – elite
383 (25.2 ± 2.8 years old): $n_{gymnasts\ (1,1)}$ = 18 (24.7%), who had competed at an international level
384 (European or World Championships or Olympic Games) and sub-elite – national-level athletes
385 (19.4 ± 3.5 years old): $n_{gymnasts\ (1,2)}$ = 55 (75.3%), who performed sports gymnastics at a national
386 level only. Contestants were classified according to the highest-level contest they had appeared
387 in. The gymnasts were only included if they had never been tested positive by an anti-doping
388 agency. A control group of healthy individuals $n_{controls}$ = 245; 150 males and 95 females; 22.6
389 ± 2.5 years old was also selected from the Polish, Italian and Lithuanian population
390 (college students) with no background in the sport.

391 Controls were matched to gymnasts in ca. 1:4 ratio; adjustment consideration has been
392 specified in the Study Protocol (S1).

393

394

395

Methods, aims and hypotheses

In the paper, a quantitative approach to analyses has been conducted. The methods of observation and diagnostic survey were used. To gather the molecular data, PCR and RT-PCR techniques have been applied.

The goals of the research were: (a) **to measure the magnitude of informative entropy of sport PEPs in artistic gymnastics with subsequent analysis of synergistic effects or redundancy between genetic variants**; (b) **to determine marginal effects and cross-partial derivatives at the level of 2-way gene-gene interactions**; and (c) **to investigate quality measures of MDR and logistic regression epistatic models for athletes recognition.**

The aims implicate the following questions: (a) How much information will be gained on artistic gymnastics after quantifying Shannon entropy of a single genetic variant? (b) Does at least one two-attribute synergistic or redundant effect exist between sport performance enhancing polymorphisms? (c) Will the best MDR epistatic model of sports gymnastics achieve an outcome greater than 55% in cross validation consistency test? (d) For which combination of gene-gene models are the first and second order terms different than zero? (e) Are genetic classifiers statistically significant in the context of sportsman recognition? These questions concern six alternative hypotheses H_1:

(a) $H\ (S_{max}) < 1$; (b) $\bigvee_{IG(A;B;C)\in \boldsymbol{IG}(\boldsymbol{A};\boldsymbol{B};\boldsymbol{C})} I(A;\ B;\ C) \neq 0$; (c) $CVC_{max} > 55\%$; (d) $\bigvee_{b_i\in \boldsymbol{b_i}} b_i \neq 0$ and; (e) $\bigvee_{b_{ii}\in \boldsymbol{b_{ii}}} b_{ii} \neq 0$ when two SNPs are investigated in 2-way interaction model; (f) $AUC_i > 0{,}7\ for\ i = 1, \dots, m$; i-th Kappa statistic > 0.6,

where:

$H\ (S_{max})$ is the maximal value of Shannon entropy in the set of genetic polymorphisms $j = 1, \dots, k$, IG is the information gain; $\boldsymbol{I}(\boldsymbol{A};\boldsymbol{B};\boldsymbol{C})$ is the vector of multiple mutual information results from all possible combinations in the analysis; CVC_{max} – the highest value obtained in cross-validation consistency (count) for epistatic models; b_i – SNP marginal effect; b_{ii} is

a 2-way G-G interaction product term; AUC_i – area under the curve for model i; $\bigvee_{IG(A;B;C)\in \boldsymbol{IG}(\boldsymbol{A};\boldsymbol{B};\boldsymbol{C})}$ is the existential quantifier.

Biological sample collection and DNA extraction

The buccal cells donated by the participants were acquired using the Oragene – DNA isolation kit (DNA Genotek, Kanata, ON, Canada). The subjects abstained from drinking, and eating for 2 hours prior to saliva collection. Each participant was asked to perform a 2-min mouth rinse with water 30 min before retrieving the DNA sample. Samples were collected by passive drooling in sterile 50 ml tubes. Tubes were filled up to 4 ml, then vigorously mixed and transported to a laboratory for further processing. All samples were stored in the same conditions at −25°C until subsequent steps were performed.

DNA was extracted according to the producer's protocol. Briefly, the DNA material located in the Oragene tubes was incubated at 50°C overnight. Afterward, the probes were opened and divided into four equal parts. Each one was treated with 40 µl of buffer solution supplied by the manufacturer. After a period of 10 minutes of ice incubation, centrifugation for 3 minutes at 13,000 rpm was performed. The resulting supernatant (DNA) was assessed for both purity and integrity by using spectrometric and electrophoretic methods, respectively.

Determination of genotypes

DNA isolation and genotyping were performed in the molecular laboratory of Gdansk University of Physical Education and Sport, Poland. The genotyping error was assessed as 1%, while the call rate was above 95%. Details on PEPs genotyping can be verified in Supplementary Material – S1. Briefly, six gene variants (*ACTN3 – rs1815739*, *PPARGC1A – rs8192678*, *PPARα – rs4253778*, *BDNF-AS – rs6265*, *GNB3 – rs5443*, *DRD2– rs1076560*) were assessed by PCR. In accordance with [2], amplification was performed in a total volume of 10 µl PCR reaction mix containing 1.5 mM $MgCl_2$, 0.75 nM of each deoxynucleoside triphosphate – dNTP (Novazym, Poland), 4 pM of specific primer (Genomed, Poland) in TE

(pH= 8.0; Thermo Fisher Scientific), 0.5 U DNA recombinant Taq polymerase in buffer (pH= 8.0; Sigma, Germany), 1x PCR buffer (pH=8.7; Sigma, Germany) and 1 μl (30–50 ng) of template DNA (isolate). The thermal-time PCR amplification cycling profile conditions consisted of 10 min of preincubation at 95°C (activation of the Taq DNA polymerase), followed by 40 cycles of denaturation at 95°C for 15 s, and primer annealing, and extension for 1 min at 60°C, followed by a final elongation cycle at 72 °C for 3 min. The PCR fragments were subsequently digested with the appropriate restriction enzyme. The PCR products were separated by electrophoresis at 80mV on a 2% agarose gel, stained with ethidium bromide (250ng / ml), and visualized in UV light. The *SNAP-25* (*rs362584*) was genotyped in two replicates with TaqMan fluorescent oligonucleotide probes. Likewise, following [41], a BioRad CFX96 Touch™ RT-PCR Detection System in tandem with the Bio-Rad CFX Manager Software was used to detect the fluorescent signals and to produce a graphical representation which allowed for A / G allelic discrimination. Freshly purified / sterile water was used as a negative control for PCR.

Statistical analyses

From 318 observations, 36 (roughly 10%) of instances were included into the test set (hold-out dataset). Minor allele frequencies were computed for each of the seven SNPs and Hardy-Weinberg equilibrium was tested. In the standard – linear approach, genotypes were coded as '1': potentially disfavorable for strength / power sports activities, '2': heterozygotes, or '3' (S2 Supplementary Material, p. 8). Next, the most commonly used six subject-level gene models including: recessive, multiplicative, additive / harmonic, dominant, and over-dominant models [22] were computed to select the best one to the given data distribution of each SNP. After quality control of alleles and model selection, the information gain (IG) of every SNP was computed with standard coding and with the adjustment for the optimal genetic model.

Next, the Multifactor Dimensionality Reduction (MDR) and logistic regression algorithms were applied.

All statistical analyses were run in MS Excel on a standard PC and in MDR program available on the Internet (https://www.multifactordimensionalityreduction.org/). The threshold for statistical significance was set to p-value ≤ 0.05, with two-sided Bonferroni correction for multiple comparisons. Formulae used for data processing have been compiled in Supplementary Material (S3 Theoretical Background – Data Analysis), for further inspection.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

1. Eynon N, Alves AJ, Sagiv M, Yamin C, Sagiv M, Meckel Y. Interaction between SNPs in the NRF2 gene and elite endurance performance. Physiol Genomics. 2010; 41(1): 78-81. doi:10.1152/physiolgenomics.00199.2009.
2. Grenda A, Leońska-Duniec A, Kaczmarczyk M, Ficek K, Król P, Cięszczyk P. Interaction Between ACE I/D and ACTN3 R557X Polymorphisms in Polish Competitive Swimmers. J Hum Kinet. 2014; 42: 127-36. doi:10.2478/hukin-2014-0067.
3. O'Connell K, Knight H, Ficek K, Leonska-Duniec A, Maciejewska-Karlowska A, Sawczuk M, et al. Interactions between collagen gene variants and risk of anterior cruciate ligament rupture. Int J Sports Sci. 2015; 15(4): 341-350. doi: 10.1080/17461391.2014.936324.
4. Wei W-H, Hemani G, Gyenesei A, Vitart V, Navarro1 P, Hayward C, et al. Genome-wide analysis of epistasis in body mass index using multiple human populations. Eur J Hum Genet. 2012; 20(8): 857-862. doi: 10.1038/ejhg.2012.17.
5. Leamy LJ, Pomp D, Lightfoot JT. An epistatic genetic basis for physical activity traits in mice. J Hered. 2008; 99(6): 639-46. doi:10.1093/jhered/esn045.

6. Hsu P-C, Yang U-C, Shih K-H, Liu C-M, Liu Y-L, Hwu H-G. A protein interaction based model for schizophrenia study. BMC Bioinform. 2008; 9 (Suppl 12): S23: 1-9. doi:10.1186/1471-2105-9-S12-S23.

7. Liu, D., Liu, L., Song, Z., Liu, J., & Hou, D. (2017). Genetic Variations of Oxidative Stress Related Genes ALOX5, ALOX5AP and MPO Modulate Ischemic Stroke Susceptibility Through Main Effects and Epistatic Interactions in a Chinese Population. Cell Physiol Biochem. 43, 1588 – 1602. doi.org/10.1159/000482023.

8. Ruiz JR, Arteta D, Buxens A, Artieda M, Gómez-Gallego F, Santiago C, et al. Can we identify a power-oriented polygenic profile? J Appl Physiol. 2010 a; 108(3): 561-566. doi: 10.1152/japplphysiol.01242.2009.

9. Hughes DC, Day SH, Ahmetov II, Williams AG. Genetic of strength and power: polygenic profile similarity limits skeletal muscle performance. J Sports Sci. 2011; 29(13): 1425-34. doi: 10.1080/02640414.2011.597773.

10. Williams AG, Folland JP. Similarity of polygenic profiles limits the potential for elite human physical performance. J Physiol. 2008; 586(1): 113-21. doi:10.1113/jphysiol.2007.141887.

11. Grealy R, Herruer J, Smith CLE, Hiller D, Haseler LJ, Griffiths LR. Evaluation of a 7-Gene Genetic Profile for Athletic Endurance Phenotype in Ironman Championship Triathletes. PloS One. 2015; 10(12): e145171: 1-20. doi: 10.1371/journal.pone.0145171.

12. Sun YV. Multigenic Modeling of Complex Disease by Random Forest. In: Moore JH, Dunlap JC, editors. Computational Methods for Genetics of Complex Traits. San Diego, CA: Elsevier; 2010. pp. 73-97.

13. Ben-Zaken S, Eliakim A, Nemet D, Meckel Y. Genetic Variability Among Power Athletes: The Stronger vs. the Faster. J Strength Cond Res. 2019; 33(6): 1505-1511. doi:

10.1519/JSC.0000000000001356.

14. Tharabenjasin P, Pabalan N, Jarjanazi H. Association of PPARGC1A Gly428Ser (rs8192678) polymorphism with potential for athletic ability and sports performance: A meta-analysis. PLoS ONE. 2019; 14(1): e0200967: 1-18. doi. org/10.1371/journal.pone.0200967.

15. Ahmetov II, Fedotovskaya ON. Sports genomics: Current state of knowledge and future directions. CMEP. 2012; 1(1): 1-24. doi:10.7457/cmep.v1i1.e1.

16. Ruiz J, Eynon N, Meckel Y, Fiuza-Luces C, Santiago C, Gómez-Gallego F, et al. GNB3 C825T Polymorphism and Elite Athletic Status: A Replication Study with Two Ethnic Groups. Int J Sports Med. 2010 b; 32:151-153. doi: 10.1055/s-0030-1268438.

17. Eynon N, Oliveira J, Meckel Y, Sagiv M, Yamin C, Sagiv M, et al. The Guanine Nucleotide Binding Protein Beta Polypeptide 3 Gene C825T Polymorphism Is Associated With Elite Endurance Athletes. Exp Physiol. 2009; 94: 344-349.

18. Jacob Y, Chivers P, Anderton RS. Genetic predictors of match performance in sub-elite Australian football players: A pilot study. J Exerc Sci Fit. 2019; 17(2): 41-46. doi: 10.1016/j.jesf.2018.10.007. Epub 2018 Nov 1.

19. Gosso MF, De Geus EJC, van Belzen MJ, Polderman TJC, Heutink P, Boomsma DI, et al. The SNAP-25 Gene Is Associated With Cognitive Ability: Evidence From a Family-Based Study In Two Independent Dutch Cohorts. Mol Psychiatr. 2006; 11(9): 878-886. doi: 10.1038/sj.mp.4001868.

20. Liu Y-S, Dai X, Wu W, Yuan F-F, Gu X, Chen J-G, et al. The Association of SNAP25 Gene Polymorphisms in Attention Deficit / Hyperactivity Disorder: A Systematic Review and Meta-Analysis. Mol Neurobiol. 2017; 54(3): 2189-2200. doi: 10.1007/s12035-016-9810-9.

21. Islamov RR, Samigullin DV, Rizvanov AA, Bondarenko NI, Nikolskiy EE. Synaptosome-associated protein 25 (SNAP25) synthesis in terminal buttons of mouse motor neuron. Dokl Biochem Biophys. 2015; 464: 272-274. doi: 10.1134/S1607672915050026. Epub 2015 Oct 31.

22. Horita N, Kaneko T. Genetic model selection for a case-control study and a meta-analysis. Meta Gene. 2015; 5: 1-8.

23. Hu T, Chen Y, Kiralis JW, Collins RL, Wejse Ch, Sirugo G, et al. An information-gain approach to detecting three-way epistatic interactions in genetic association studies. J Am Med Inform Assoc. 2013; 20: 630-636. doi: 10.1136/amiajnl-2012-00152.

24. Milne RL, Fagerholm R, Nevanlinna H, Benítez J. The importance of replication in gene–gene interaction studies: multifactor dimensionality reduction applied to a two-stage breast cancer case-control study. Carcinogenesis. 2008; 29(6): 1215-1218. doi: 10.1093/carcin/bgn120.

25. Sundström S. Coding in Multiple Regression Analysis: A Review of Popular Coding Techniques. U.U.D.M. Project Report. Uppsala University. 2010. Available from: http://uu.diva-portal.org/smash/get/diva2:325460/FULLTEXT01.pdf

26. Lekman M, Hössjer O, Andrews P, Källberg H, Uvehag D, Charney D, et al. The genetic interacting landscape of 63 candidate genes in Major Depressive Disorder: an explorative study. BioData Min. 2014; 7(19): 1-18. doi: 10.1186/1756-0381-7-19.

27. Jyothi KU, Reddy BM. Gene–gene and gene–environment interactions in the etiology of type 2 diabetes mellitus in the population of Hyderabad, India. Meta Gene. 2015; 5: 9-20. doi: 10.1016/j.mgene.2015.05.001.

28. Manuguerra M, Matullo G, Veglia F, Autrup H, Dunning AM, Garte S, et al. Multifactor dimensionality reduction applied to a large prospective investigation on gene-gene and gene-environment interactions. Carcinogenesis. 2007; 28(2): 414-22. doi:

570 10.1093/carcin/bgl159.

571 29. Wu Y, Zhang L, Liu L, Zhang Y, Zhao Z, Liu X, et al. A multifactor dimensionality
572 reduction-logistic regression model of gene polymorphisms and an environmental
573 interaction analysis in cancer research. Asian Pac J Cancer Prev. 2011; 12(11): 2887-
574 2892.

575 30. Dasgupta S, Reddy BM. The role of epistasis in the etiology of Polycystic Ovary
576 Syndrome among Indian women: SNP-SNP and SNP-environment interactions. Ann
577 Hum Genet. 2013; 77(4): 288-298. doi: 10.1111/ahg.12020.

578 31. Bottema RW, Postma DS, Reijmerink NE, Thijs C, Stelma, FF, Smit HA, et al. Interaction
579 of T-cell and antigen presenting cell co-stimulatory genes in childhood IgE. Eur Respir J.
580 2010; 35(1): 54-63. doi: 10.1183/09031936.00018909.

581 32. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING
582 v11: protein–protein association network swith increased coverage, supporting functional
583 discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019; 47
584 (D1):D607-D613: doi: 10.1093/nar/gky1131.

585 33. Guerini FR, Farina E, Costa AS, Baglio F, Saibene FL, Margaritella N. et al. ApoE and
586 SNAP-25 Polymorphisms Predict the Outcome of Multidimensional Stimulation Therapy
587 Rehabilitation in Alzheimer's Disease. Neurorehab Neural Re. 2016; 30(9): 883-893. doi:
588 10.1177/1545968316642523.

589 34. Fuku N, Díaz-Peña R, Arai Y, Abe Y, Zempo H, Naito H. et al. Epistasis, physical
590 capacity-related genes and exceptional longevity: *FNDC5* gene interactions with
591 candidate genes *FOXOA3* and *APOE.* BMC Genomics. *2017;* 18(Suppl 8): 803: 126-131.
592 doi: 10.1186/s12864-017-4194.

593 35. Lee J-H, Jang A-S, Park S-W, Kim D-J, Park C-S. Gene-Gene Interaction Between CCR3
594 and Eotaxin Genes: The Relationship With Blood Eosinophilia in Asthma. Allergy

595 Asthma Immunol Res. 2014; 6(1); 55-60. doi: 10.4168/aair.2014.6.1.55.

596 36. Lulińska-Kuklik E, Rahim M, Domańska-Senderowska D, Ficek K, Michałowska-
597 Sawczyn M, Moska W, et al. Interactions between *COL5A1* Gene and Risk of the Anterior
598 Cruciate Ligament Rupture. J Hum Kinet. 2018; 62: 65-71. doi: 10.1515/hukin-2017-
599 0177.

600 37. MacArthur, DG, North KN. ACTN3: A Genetic Influence on Muscle Function and
601 Athletic Performance. Exerc Sport Sci Rev. 2007; 35(1):
602 30-34. doi: 10.1097/JES.0b013e31802d8874.

603 38. Jakulin A. Attribute Interactions in Machine Learning. M.Sc. Thesis. Second Edition.
604 University of Ljubljana. 2003. p. 37. Available from:
605 http://stat.columbia.edu/~jakulin/Int/interactions_full.pdf

606 39. Ai Ch, Norton EC. Interaction terms in logit and probit models. Econ Lett. 2003; 80(1):
607 123-129. doi: 10.1016/S0165-1765(03)00032-6.

608 40. Kim J, Yum S, Kang C, Kang S-J. Gene-gene interactions in gastrointestinal cancer
609 susceptibility. Oncotarget. 2016; 7(41): 67612-67625. doi:10.18632/oncotarget.11701.

610 41. Leońska-Duniec A, Jastrzębski Z, Jażdżewska A, Moska W, Lulińska-Kuklik E, Sawczuk
611 M, et al. Individual Responsiveness to Exercise-Induced Fat Loss and Improvement of
612 Metabolic Profile in Young Women is Associated with Polymorphisms of Adrenergic
613 Receptor Genes. J Sport Sci Med. 2018; 17(1): 134-144.

614 Supporting information

615 **S1 Study Protocol.** This protocol has been provided by the authors to give readers additional
616 information about the research work (.docx).

617 **S2 Supplementary Material 1.** This work contains all supplemental text, figures, and tables.
618 (.docx).

619 **S3 Supplementary Material 2.** Theoretical Background - Data Analysis.

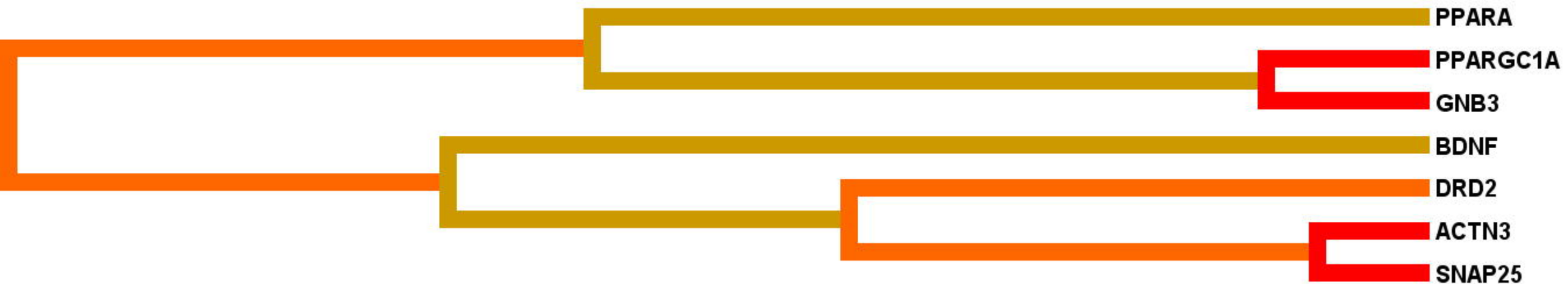
620 **S4 Data Input.**

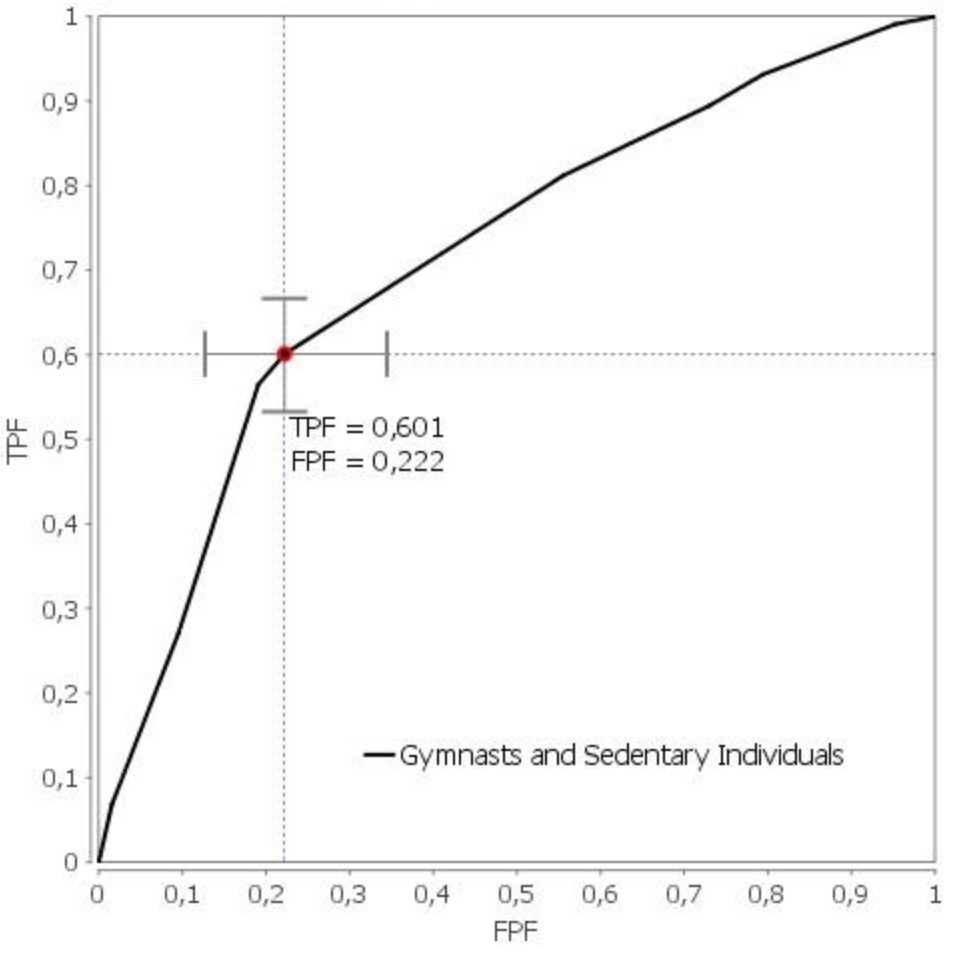
621 **S5 Author Summary.**

622 **S6 Figure 1.**

623 **S7 Figure 2.**

624 **S8 Graphical Abstract.**


PPARA
PPARGC1A
GNB3
BDNF
DRD2
ACTN3
SNAP25


1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
TPF
TPF = 0,601
FPF = 0,222
Gymnasts and Sedentary Individuals
0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1
FPF