Genetic variants associated with exercise performance in both moderately trained and highly trained individuals

Harvey, N R; Voisin, S; Dunn, P J; Sutherland, H; Yan, X; Jacques, M; Papadimitriou, I D; Haseler, L J; Ashton, K J; Haupt, L M; Eynon, N; Griffiths, L R

Published in:
Molecular Genetics and Genomics

DOI:
10.1007/s00438-019-01639-8

GENETIC VARIANTS ASSOCIATED WITH EXERCISE PERFORMANCE IN BOTH MODERATELY TRAINED AND HIGHLY TRAINED INDIVIDUALS

Harvey NR[1, 2], Voisin S[3], Dunn PJ[2], Sutherland H[2], Yan X[3], Jacques M[3], Papadimitriou ID[3], Haseler LJ[4], Ashton KJ[1], Haupt LM[2], Eynon N[3] and *Griffiths LR[2]

[1]Health Sciences and Medicine faculty, Bond University, Robina, Australia, QLD 4226

[2]Genomics Research Centre, Institute of Health and Biomedical Innovation, School of Biomedical Sciences, Queensland University of Technology, Brisbane, Australia, QLD 4059

[3] Institute for Health and Sport (iHeS), Victoria University, Victoria, Australia, VIC 3011

[4]Faculty of Health Sciences, School of Physiotherapy and Exercise Science, Curtin University, Perth, WA 6102

***Corresponding author:** Prof. Lyn R. Griffiths, Genomics Research Centre, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane QLD, Australia

E-mail: lyn.griffiths@qut.edu.au

Phone: +61 7 3138 6102

Targeted journal Molecular Genetics and Genomics (SJR: 1.1 Q1))

NH: 0000-0003-2035-8475, **SV**: 0000-0002-4074-7083, **HS**: 0000-0002-8512-1498, **XY**: 0000-0001-8547-4210, **IP**: 0000-0001-5032-9937, **KA**: 0000-0001-6106-3425, **LH**: 0000-0002-7735-8110, **NE**: 0000-0003-4046-8276, **LG**: 0000-0002-6774-5475

Funding and Acknowledgements:

This research was chiefly supported by the Commonwealth Collaborative Research Network funding to Bond University CRN-AESS. Mr Nicholas Harvey was supported by a PhD stipend also provided by Bond University CRN-AESS. This research was also supported by infrastructure purchased with Australian Government EIF Super Science Funds as part of the Therapeutic Innovation Australia - Queensland Node project (LRG). Nir Eynon is supported by the National Health and Medical Research Council, Australia (NHMRC CDF# APP1140644). The

authors would like to acknowledge Dr Cassie Albury for assistance with the molecular methods used in this manuscript.

ABSTRACT

Adaptation to exercise training is a complex trait that may be influenced by genetic variants. We identified 36 single nucleotide polymorphisms (SNPs) that had been previously associated with endurance or strength performance, exercise-related phenotypes or exercise intolerant disorders. A MassARRAY multiplex genotyping assay was designed to identify associations with these SNPs against collected endurance fitness phenotype parameters obtained from 2 exercise cohorts (Gene SMART study; n=58 and Hawaiian Ironman Triathlon 2008; n=115). These parameters included peak power output (PP), a time trial (TT), lactate threshold (LT), maximal oxygen uptake (VO_2 max) in recreationally active individuals and a triathlon time to completion (Hawaiian Ironman Triathlon cohort only). A nominal significance threshold of $\alpha<0.05$ was used to identify 17 variants (11 in the Gene SMART population and 6 in the Hawaiian Ironman Triathlon cohort) which were significantly associated with performance gains in highly trained individuals. The variant rs1474347 located in *Interleukin 6* (*IL6*) was the only variant with a false discovery rate < 0.05 and was found to be associated with gains in VO_2 max (additional 4.016 mL/(kg·min) for each G allele inherited) after training in the Gene SMART cohort. In summary, this study found further evidence to suggest that genetic variance can influence training response in a moderately trained cohort and provides an example of the potential application of genomic research in the assessment of exercise trait response.

Key Words:

Genetics, exercise, MassARRAY, replication, endurance

INTRODUCTION

Currently, robust identification of genetic variants associated with exercise phenotypes is limited by a lack of reproducible results. Family and twin studies have estimated high heritability for various exercise performance metrics (e.g. muscle mass: 40%, anaerobic power: 70-80%, aerobic exercise: 50%) [1]. However, a wide variety of environmental (e.g. diet, sleep), psychological and epigenetic factors may also influence exercise responses [2]. In addition, within-subject variability (i.e. the variable response of a given individual to the same exercise training) considerably limits the identification of genetic variants with potentially small effects on exercise response [3].

To date, only two genetic signatures have consistently shown an association with exercise responses; the *Alpha-actinin-3* stop gain variant (*ACTN3)* p.Arg577Ter and the *Angiotensin converting enzyme (ACE)* Insertion/Deletion (I /D) in intron 16 [4, 5]. Genome Wide Association Studies (GWAS) have helped discern genomic loci associated with training response, however these usually contain a low number of participants, and/or evidence of association with exercise psychology related phenotypes [6]. Studies in metabolic and cardiovascular disorders such as diabetes or arterial hypertension have further complicated participant ability to perform exercise training at duration and intensity and as such participants can be classified as having exercise intolerant disorders [7]. As exercise training yields a host of health benefits, understanding which genetic and molecular processes contribute to these responses might be helpful to the development of personalised exercise therapeutics (e.g. exercise dosing to minimise risk of adverse response within exercise intolerant disorders) [8].

In this study, we investigated candidate genes previously implicated in exercise response in cohorts of varying fitness levels. We used a highly trained cohort (triathlon) and a moderately trained, longitudinal cohort of High-Intensity Interval endurance Training (HIIT) exercise training. We hypothesised that many, if not all, of the candidate SNPs would be found to be associated with triathlon performance and response to 4 weeks of endurance exercise training, regardless of age or baseline fitness level.

MATERIALS AND METHODS

The Gene SMART cohort

The Gene SMART (Skeletal Muscle Adaptive Response to Training) study design has previously been described [9]. The study is ongoing with currently > 100 moderately-trained participants who were sampled for blood and skeletal muscle (*vastus lateralis*) at several time points: before, immediately after and 3 hours after a single bout of high-intensity endurance exercise (HIIE), and after 4 weeks of High-Intensity Interval Training (HIIT) [9]. Exercise-related phenotypic measurements were collected before and after the completion of the exercise training intervention (e.g. Lactate Threshold (LT, in Watts), Peak Power output (PP, in Watts), maximal oxygen uptake (VO_2max, in mL/min/kg body weight, from graded exercise tests), and a Time Trial measurement (TT, in min). All participants gave informed consent and the study was approved by the Victoria University Ethics Committee (Approval number: HRE13-233). Subsequently, the study was also approved by the Queensland University of Technology (QUT) Human Research Ethics Committee (Approval number: 1600000342). All procedures performed in studies involving human participants were in accordance with the ethical standards of the respective institutions research committees, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. At the time of collection, n = 77 participants had participated in the study, with n = 58 completing the entire 4-week training program. Genomic DNA was extracted and purified from whole blood using the QIAamp DNA blood midi kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions for participants that completed the study. Samples that failed genotypic analysis or had a large amount of missing phenotypic data, were removed from further analysis, leaving a final sample size of n = 52 (Age = 30.95 ± 8.17 years). In the moderately trained cohort (Gene SMART, n = 58), we focused on the response to an exercise training program (longitudinal analysis). Specifically, we measured the change in (Δ = post - pre) measurement for each endurance fitness trait as a representation of response to exercise training.

Highly trained (Ironman) cohort

Ironman triathlons consist of a 3.86 km swim, a 180.25 km bike ride, followed by the completion of a full marathon (42.2 km). The 2008 Hawaiian Ironman Triathlon population has been previously described as an elite endurance cohort based on their eligibility and participation in the event [10, 11]. Due to the intensity of this endurance event only highly trained individuals that completed it were included in this study. To avoid genetic confounding, we analysed only the triathlon participants who self-identified as male and Caucasian (Age = 43.81 ± 11.39 years).

This is a post-peer-review, pre-copyedit version of an article published in Molecular Genetics and Genomics. The final authenticated version is available online at: https://doi.org/10.1007/s00438-019-01639-8

This was performed solely on the triathlon group as the Gene SMART population was already homogeneously male. All procedures performed in studies involving human participants were in accordance with the ethical standards of the QUT human Research Ethics Committee (approval number: 1300000499), and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Saliva samples (OG-250 Oragene Kit, DNA Genotek Inc.) and questionnaires were collected prior to the event; time to completion measurements for each event was collected from the publicly available online event webpage. Genomic DNA was extracted as per manufacturer instructions and described previously [11]. In the highly trained cohort (Ironman, n = 115), we focused on endurance performance, the result of months or years of training (cross-sectional analysis). Specifically, we used the time to completion of the running event, the biking event, the swimming event, and the total event.

Genotype method and SNP selection

The SNPs investigated in this study (Table I) were included based on conformity of 1 of 3 criteria. The first was that SNPs chosen had to have been previously associated with elite athletic status, exercise responses with reasonable replication, or exercise traits at baseline. This resulted in 11 SNPs chosen, though it should be noted that we were unable to genotype the *ACE* I/D variant (rs4340) using the MassARRAY and previous work failed to identify an association with baseline fitness levels in the Gene SMART cohort [4]. The second criteria encompassed SNPs previously investigated but less consistently associated with performance i.e. studies with equivalent numbers of negative studies or studies related to exercise psychology. The third criteria included SNPs associated with exercise intolerant disorders and non-exercise respiratory, muscular, or energy storage phenotypes such as hypertension (HT), cardiovascular disease (CVD), or Type 2 Diabetes Mellitus (T2DM). The experimental methodology for sample preparation and genotype analysis was performed using the Agena Biosciences MassARRAY, a Matrix Assisted Laser Desorption Ionisation Time of Flight (MALDI-TOF) mass spectrometer, which has been described elsewhere [12, 13]. An internal genotyping control SNP (rs17602729, *AMPD1*), previously validated in our endurance cohort, was used to ensure the MassARRAY system correctly identified genotypes [10].

This is a post-peer-review, pre-copyedit version of an article published in Molecular Genetics and Genomics. The final authenticated version is available online at: https://doi.org/10.1007/s00438-019-01639-8

127 **Table I: Details on the 36 SNPs included in the custom MassARRAY genotyping assay.**

CHR	Gene	SNP ID	A1	Phenotype(s)	Number of studies with positive results	Number of participants in studies with positive results	Number of studies with negative results	Number of participants in studies with negative results	Ref
Category 1: Well replicated or solely exercise associations									
1	*AMPD1*[†]	rs17602729	C T	END POW	2 3	231 510	1 0	84 0	[14]
6	*HFE*	rs1799945	G	END	2	148	-	-	[14]
6	*VEGFA*	rs2010963	C	END	1	942	-	-	[14]
11	*ACTN3*	rs1815739	C T	END POW	4 12	560 1,484	14 5	3,039 498	[14]
11	*UCP2*	rs660339	T C	END POW	1 1	694 29	- -	- -	[14]
19	*CKMM*	rs8111989	G	POW	2	233	-	-	[14]
21	*COL6A1*	rs35796750	T	END	1	661	-	-	[14]
22	*PPARα*	rs4253778	G C	END POW	5 2	740 260	- 1	- 81	[14]
Category 2: Mixed results									
1	*SGIP1*	rs9633417	C C	EBH POW	2 -	2,838 -	- 1	- 753	[15]
1	*LEPR*	rs1137101	A G G	END POW EBH	- 1 3	- 242 3676	1 - -	846 - -	[15-17]
4	*PGC1α*	rs8192678	A	END	4	849	3	508	[14]
4	*UCP1*	rs10440457	G	EBH POW	2 -	2,838 -	- 1	- 181	[15]
4	*PGC1α*	rs6821591	T	END	1	235	-	-	[18]
5	*ADRB2*	rs1042713	A	POW	1	100	-	-	[14]

This is a post-peer-review, pre-copyedit version of an article published in Molecular Genetics and Genomics.
The final authenticated version is available online at: https://doi.org/10.1007/s00438-019-01639-8

CHR	Gene	SNP ID	A1	Phenotype	Number of studies with positive results	Number of participants in studies with positive results	Number of studies with negative results	Number of participants in studies with negative results	Ref
Category 2 continued: Mixed results									
7	*NRF1*	rs6949152	G	END	1	102	1	75	[19, 20]
8	*ADRB3*	rs4994	C	END	1	100	1	81	[14]
14	*BDKRB2*	rs1799722	T	END	1	316	-	-	[14]
15	*NRF2*	rs7181866	G	END	2	129	1	89	[14]
15	*NRF2*	rs8031031	T	END	1	74	1	89	[14]
Category 3: Disease Associations and other									
1	*ATP1A2*[§]	rs28933400	T	HYP	-	-	1	388	[21]
1	*LEPR*	rs12405556	T	EBH	2	978	-	-	[15]
1	*DIO1*	rs2294512	A	THY	1	547	-	-	[22]
2	*MSTN*	rs1805086	G	POW	13	3,080	-	-	[10, 23]
4	*UCP1*	rs2270565	T	T2D	2	981	4	1,382	[24-26]
6	*EDN1*	rs5370	T	HYP	2	1,004	-	-	[27, 28]
6	*HLA-A*	rs1061235	T	END	1	32	-	-	[16]
7	*IL6*	rs1474347	A	T2D	1	10,775	-	-	[29, 30]
10	*ADRB1*	rs1801253	C	HYP T2D	1 1	61 947	- -	- -	[31, 32]
12	*IGF1*	rs121912430	T	OBE	1	502	-	-	[16]
15	*CYP19A1*[‡]	rs2470158	T	EBH	1	1,722	-	-	[15]
16	*Intronic*	rs238838	C	-	-	-	-	-	[16]
18	*MC4R*	rs9965495	A	T2D	2	6,657	-	-	[15, 33]

CHR	Gene	RS#	A1	Phenotype(s)	Number of studies with positive results	# Participants	Number of studies with negative results	#Participants	Ref
Category 3: Disease Associations and other									
19	*APOE*	rs7412	T	T2D END OBE	2 - -	9,314 - -	- 2 2	- 507 159	[34-39]
22	*PPARα*	rs1800206	G	HYP POW T2D	1 1 4	269 610 3,643	- - -	- - -	[40-46]
MT	*MTTT*	rs199474700	G	END	1	46	-	-	[16]
MT	*MTND5*	rs28359178	A	END	1	46	-	-	[16]

NB: Grey rows represent intergenic variants included from GWAS conducted by *Rankinen et al.* [16].

Phenotypes are the traits in which the SNP has been previously implicated. Previous studies and numbers of participants are shown and separated according to positive or negative results. SNPs are separated into three categories based on phenotype and replication (adapted from *Ahmetov I.I. et al*,.[14]). END: Endurance, POW: Power/Strength, T2D: Diabetes, HYP: Hypertension, EBH: Exercise behaviour, OBE: Obesity, †: Inbuilt genotyping control, A1: Tested Allele

‡ **excluded in Gene SMART population.**

§ **Excluded from highly trained population.**

This is a post-peer-review, pre-copyedit version of an article published in Molecular Genetics and Genomics.
The final authenticated version is available online at: https://doi.org/10.1007/s00438-019-01639-8

Data processing

The output files from the MassARRAY platform were converted to PLINK format and analysed for correct genotypic identification (calling). For the Gene SMART and Ironman populations respectively, SNPs were excluded from further analysis if they exceeded the following thresholds: 1) SNPs that had a calling rate < 80% (>20% missing data) (n = 5, n = 3); 2) SNPs with a minor allele frequency < 2% (n = 1, n = 2); 3) SNPs determined not in hardy Weinberg equilibrium (n=1, n=1) [12]. Subsequent analysis was performed on n = 29 SNPs for the Gene SMART population and n = 30 SNPs for the Ironman population.

Statistical analysis

We measured normality metrics (skewness and kurtosis) for each phenotype in both populations using the *ggplot2*, *tidyverse* and *moments* packages in R, to determine if data transformation was necessary from the raw phenotypic values. We used PLINK V1.90p to perform quantitative linear association tests (95% CI) with both dominant and recessive models for each cohort, adjusting for age. An additive model was considered but did not differ from the results obtained from the dominant model. As this was a candidate gene study, SNPs that had a raw *p*-value < 0.05 were considered nominally significant while variants that had an adjusted *p*-value (Benjamini-Hochberg False Discovery Rate (FDR)) < 0.05 were considered significant. This adjustment method represents a good balance between type I and type II errors and as such minimises false positive results. To avoid multiple testing burdens with phenotypic traits, we used a separate hypothesis for each quantitative trait. Effect sizes were determined using raw beta regression coefficient values interpreted as "how much a specific phenotype increased for each additional X allele at the SNP of interest".

RESULTS

The array genotyping control (*AMPD1*) was identified to be 100% concordant with the genotyping results from another method (RFLP) in our previous study with the same population, confirming the validity of the MassARRAY data.

Table II: Summary of nominally significant variants associated with gains in endurance fitness after exercise training in the Gene SMART cohort. *Tests were performed for both dominant and recessive models for each trait: Wpeak: maximum ergometer intensity at stop (Watts), LT: Lactate Threshold (Watts), VO_2max: maximum oxidative respiration uptake (mL/(kg·min)), TT: time trial completion (seconds).*

Trait	CHR	SNP	Allele	Gene Symbol	Type of SNP	Model	MAF	P-value[†]	FDR	Effect size (Beta)
Δ-Wpeak	1	rs17602729	A	*AMPD1*	Stop-gain	DOM	0.098	0.009	0.162	13.75
	16	rs238838	A	-	Intronic	DOM	0.11	0.01	0.162	-15.24
	1	rs2294512	C	*DIO1*	Intronic	DOM	0.042	0.043	0.393	-16.8
Δ-LT	14	rs1799722	T	*BDKRB2*	5`UTR variant	REC	0.33	0.027	0.24	-17.17
	15	rs8031031	T	*NRF2*	Intronic	REC	0.016	0.027	0.24	-37.34
	21	rs35796750	T	*COL6A1*	Intronic	REC	0.48	0.035	0.24	11.88
	19	rs7412	T	*APOE*	Missense	DOM	0.105	0.042	0.64	-11.95
	18	rs9965495	A	*MC4R*	Intronic	DOM	0.28	0.045	0.64	9.107
Δ-VO_2max	**7**	**rs1474347**	**C**	***IL6***	**Intronic**	**REC**	**0.45**	**0.00087**	**0.018**	**-4.016**
	11	rs660339	A	*UCP2*	Missense	REC	0.46	0.037	0.38	-2.835
	15	rs8031031	T	*NRF2*	Intronic	DOM	0.04	0.04	0.87	4.741
Δ-TT	6	rs1799945	G	*HFE*	Missense	DOM	0.21	0.019	0.58	101

CHR = Chromosome, SNP = Single Nucleotide Polymorphism, DOM = Dominant model, REC = Recessive model, MAF = Minor Allele Frequency, FDR = False Discovery Rate

[†]*P-value adjusted for age*

Table III: All nominally significant variants associated with different triathlon event finishing times in the highly trained endurance cohort. *SNPs were determined nominally significant under an arbitrary α<0.05 threshold. FDR adjusted results are shown for all nominal variants. All traits are shown in hours.*

Trait	CHR	SNP	Allele	Gene Symbol	Effect	Model	MAF	P-value[†]	FDR	Effect size (Beta)
Swim time	7	rs6949152	G	*NRF1*	Intronic	REC	0.14	0.019	0.47	0.2459
	19	rs8111989	G	*CKMM*	Downstream variant	DOM	0.34	0.019	0.31	-0.0903
	6	rs1061235	T	*HLA-A*	Non-coding transcript	DOM	0.061	0.02	0.31	0.1392
Cycle time	1	rs9633417	A	*SGIP1*	Intronic	DOM	0.1	0.036	0.89	0.3238
	4	rs8192678	A	*PGC1α*	Missense	REC	0.36	0.038	0.40	0.2965
	4	rs6821591	A	*PGC1α*	Non-coding transcript	REC	0.45	0.041	0.40	-0.2977
Run time	1	rs9633417	A	*SGIP1*	Intronic	DOM	0.1	0.019	0.61	0.424
Total-time	1	rs9633417	A	*SGIP1*	Intronic	DOM	0.1	0.012	0.37	0.846

CHR = Chromosome, SNP = Single Nucleotide Polymorphism, DOM = Dominant model, REC = Recessive model, MAF = Minor Allele Frequency, FDR = False Discovery Rate

[†]*P-value adjusted for age*

This is a post-peer-review, pre-copyedit version of an article published in Molecular Genetics and Genomics. The final authenticated version is available online at: https://doi.org/10.1007/s00438-019-01639-8

Six variants in five distinct genes were nominally associated with time-to-completion of Ironman events: *Nuclear Respiratory Factor 1 (NRF1: rs6949152), Myostatin (MSTN: rs1805086), Major Histocompatibility complex class 1A (HLA-A:* rs1061235), *Peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PPARGC1α: rs6821591, rs6821591),* and *SH3 Domain GRB2 Like Endophilin Interacting Protein (SGIP1: rs9633417).* The results of our association testing did not change significantly when age was used as a covariate.

In the Gene SMART cohort, eleven variants in nine distinct genes were shown to be nominally associated with gains in endurance fitness following exercise training (Table II): *Adenosine Monophosphate Deaminase 1 (AMPD1: rs17602729), Iodothyronine Deiodinase 1 (DIO1: rs2294512), Bradykinin receptor B2 (BDKRB2: rs1799722), Nuclear Respiratory Factor 2 (NRF2: rs7181866, rs8031031), (COL6A1, rs39796750), Apolipoprotein E (APOE: rs7412), Interleukin 6 (IL6: rs1474347), Mitochondrial uncoupling protein 2 (UCP2: rs660339)*, and *Homeostatic Iron Regulator (HFE: rs1799945).*

Interestingly, no variants were identified to be significantly associated with both time-to-completion in the Ironman cohort and the response to endurance exercise training in the Gene SMART cohort. Only rs1474347 in *IL6* passed correction for multiple testing using the BH-FDR method (FDR: 0.018). The C allele at rs1474347 was associated with VO_2max response within the Gene SMART study with an effect size of -4.016mL/(kg·min).

DISCUSSION

In the present study, we have successfully replicated previously associated exercise-related SNPs using the combined data from highly trained and moderately trained cohorts. Our main findings identified the rs1474347 in the *IL6* gene to be significantly associated with gains in VO_2 max in the Gene SMART cohort after multiple-testing statistical corrections. In addition, 17 genetic variants were found to be associated with either elite performance or responses to exercise, however, none of these variants were common between these cohorts.

Different genetic signatures likely confer different responses to exercise training via specific molecular pathways. Therefore, variants that influence pathways responsible for adaptation to moderate training may in part differ to those that confer response to high intensity endurance training. Additionally, moderately trained cohorts typically contain individuals with large variability in environmental factors such as diet, sleep and habitual physical activity

This is a post-peer-review, pre-copyedit version of an article published in Molecular Genetics and Genomics. The final authenticated version is available online at: https://doi.org/10.1007/s00438-019-01639-8

patterns, while the inter-individual variability in these measures is smaller in highly-trained cohorts and therefore less likely to confound results [47].

Association between genetic variants and exercise responses in the Gene SMART cohort

Located in an intron of the *IL6* gene, the rs1474347 variant has been previously associated with T2D traits in a large-scale study (n = 10,775). The IL6 protein is a pro-inflammatory cytokine with myokinetic (i.e. excreted from skeletal muscle) functions and is responsible for triggering and maintaining immune processes following post-exercise muscle damage [48]. We found that the C allele at this locus negatively affected the exercise response to the VO_2max phenotype (β = -4.016mL/kg·min) and therefore a homozygous C/C genotype would result in a VO_2max loss of -8.032 mL/kg·min.. The rs1800795 coding variant within the *IL6* gene has shown mixed evidence of exercise associations i.e. variant C = athleticism, G = power [14, 49]. Interestingly, further analysis identified the rs1474347 C allele to be in strong linkage disequilibrium (LD) with the C allele of rs1800795 (R^2 = 0.96). As such, it is feasible that the LD identified between these variants has contributed to the mixed evidence reported for association studies implicating the latter variant (rs1800795) in *IL6* for exercise traits. We propose that the rs1474347 variant may reduce the expression of IL6 during acute muscle damage and therefore cause a reduced local immune response leading to loss of skeletal muscle remodelling and repair. In addition this variant is also located 2kb upstream of an uncharacterised long non-coding RNA (lncRNA; *LOC541472*) and therefore, variants in this region may affect the IL-6 pro-inflammatory pathway or post-translational epigenetic and regulatory processes.

Association between genetic variants and Ironman performance

The run time (42.2 km marathon), and bike time (180.25 km ride) events in the triathlon are largely leg-based exercise activities, and therefore we expected a significant overlap of variants associated with these traits. In contrast, the triathlon swimming event utilises whole body muscle groups, therefore SNPs seen in this test were anticipated to only be associated with this particular trait. Our findings supported this as the variants associated with the swim time event were not seen in either of the other isolated finishing times, or indeed the total time to completion trait. This is also supported by the current literature where elite runners and swimmers are not analysed collectively [50]. We also note that the two variants nominally associated with the swim time trait are involved in hypoxic events characteristic of swimming [51]. It is possible that the rs6949152 G allele within the *NRF1* gene results in lower activity of NRF transcription factor and therefore increased levels of Hypoxia Inducible Factor 1

This is a post-peer-review, pre-copyedit version of an article published in Molecular Genetics and Genomics. The final authenticated version is available online at: https://doi.org/10.1007/s00438-019-01639-8

alpha (HIF1α). This would cause reduced oxidative metabolic processing and therefore lead to the increase in swim time that is associated with the variant (β = 0.2459 hours). Additionally, the CKMM protein has been shown to exhibit protective effects during mild hypoxia and therefore e hypothesised that the rs8111989 variant would increase the functionality of the CKMM protein, resulting in protection against re-oxygenation induce muscle damage and decreased swim time [52].

Although multiple SNPs examined in this study passed our nominal threshold for significance, which was unexpected given our relatively small sample sizes, all variants nominally significant in each cohort have previously been investigated as causative variants in multiple exercise studies.

Using a MassARRAY design of 36 SNPs, we found a significant association for the rs1474347 SNP in *IL6* with the change in VO2max trait in a cohort of moderately trained individuals. Furthermore, 16 other SNPs were shown to have nominal association with exercise response in the Gene SMART cohort, or Ironman performance in highly-trained athletes. As such, these markers may be useful in the development of tailored genetic panel screening and therapeutics in sports science and exercise intolerant disorders. However, to more fully exploit their applicability in this context, confirmation of the genotypic phenotype on gene function is required. Whilst this is outside the purview of this study, we have successfully replicated the significance of several exercise genes in two relatively small exercise study cohorts through nominally significant associations identified in the study cohorts. We were also able to implicate and ascertain directionality of SNPs between the different phenotypic traits. Additionally, the different variants associated with each cohort highlight the need to examine multiple cohorts of differing fitness levels and training capabilities. However, more replication studies are required in conjunction with functional transcriptomic/proteomic studies to confirm the genes and pathways associated with exercise adaptations. The use of multi-centre studies and consortia, such as the Athlome study consortium would be helpful to better facilitate these efforts to further develop the field of exercise genomics research [53].

Data Availability

The datasets generated and analysed in this study will be provided by the corresponding author upon reasonable request.

Conflicts of interest

This is a post-peer-review, pre-copyedit version of an article published in Molecular Genetics and Genomics. The final authenticated version is available online at: https://doi.org/10.1007/s00438-019-01639-8

The authors declare that they have no conflicts of interest

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (Gene SMART study: VU human research ethics committee: HRE13-233; QUT human research ethics committee: 1600000342; Hawaiian Ironman triathlon study: QUT Human Research Ethics Committee: 1300000499) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

REFERENCES

1. Issurin, V.B., *Evidence-Based Prerequisites and Precursors of Athletic Talent: A Review.* Sports Med, 2017. **47**(10): p. 1993-2010.
2. den Hoed, M., et al., *Heritability of objectively assessed daily physical activity and sedentary behavior.* Am J Clin Nutr, 2013. **98**(5): p. 1317-25.
3. Voisin, S., et al., *Statistical Considerations for Exercise Protocols Aimed at Measuring Trainability.* Exerc Sport Sci Rev, 2019. **47**(1): p. 37-45.
4. Yan, X., et al., *ACE I/D gene variant predicts ACE enzyme content in blood but not the ACE, UCP2, and UCP3 protein content in human skeletal muscle in the Gene SMART study.* J Appl Physiol (1985), 2018. **125**(3): p. 923-930.
5. Houweling, P.J., et al., *Is evolutionary loss our gain? The role of ACTN3 p.Arg577Ter (R577X) genotype in athletic performance, ageing, and disease.* Hum Mutat, 2018. **39**(12): p. 1774-1787.
6. Bouchard, C., *Overcoming Barriers to Progress in Exercise Genomics.* Exercise and Sport Sciences Reviews, 2011. **39**(4): p. 212-217.
7. McCoy, J., et al., *Pathophysiology of exercise intolerance in chronic diseases: the role of diminished cardiac performance in mitochondrial and heart failure patients.* Open Heart, 2017. **4**(2): p. e000632.
8. Buford, T.W., M.D. Roberts, and T.S. Church, *Toward exercise as personalized medicine.* Sports Med, 2013. **43**(3): p. 157-65.
9. Yan, X., et al., *The gene SMART study: method, study design, and preliminary findings.* BMC Genomics, 2017. **18**(Suppl 8): p. 821.
10. Grealy, R., et al., *Evaluation of a 7-Gene Genetic Profile for Athletic Endurance Phenotype in Ironman Championship Triathletes.* PLoS One, 2015. **10**(12): p. e0145171.
11. Grealy, R., et al., *The genetics of endurance: frequency of the ACTN3 R577X variant in Ironman World Championship athletes.* J Sci Med Sport, 2013. **16**(4): p. 365-71.
12. Ellis, J.A. and B. Ong, *The MassARRAY((R)) System for Targeted SNP Genotyping.* Methods Mol Biol, 2017. **1492**: p. 77-94.
13. Jurinke, C., et al., *Automated genotyping using the DNA MassArray technology.* Methods Mol Biol, 2002. **187**: p. 179-92.
14. Ahmetov, II and O.N. Fedotovskaya, *Current Progress in Sports Genomics.* Adv Clin Chem, 2015. **70**: p. 247-314.
15. De Moor, M.H., et al., *Genome-wide association study of exercise behavior in Dutch and American adults.* Med Sci Sports Exerc, 2009. **41**(10): p. 1887-95.
16. Rankinen, T., et al., *The human gene map for performance and health-related fitness phenotypes: the 2003 update.* Med Sci Sports Exerc, 2004. **36**(9): p. 1451-69.
17. Walsh, S., et al., *Leptin and leptin receptor genetic variants associate with habitual physical activity and the arm body composition response to resistance training.* Gene, 2012. **510**(1): p. 66-70.
18. He, Z.H., et al., *PGC-related gene variants and elite endurance athletic status in a Chinese cohort: a functional study.* Scand J Med Sci Sports, 2015. **25**(2): p. 184-95.
19. Garatachea, N., et al., *Genetic variants in the PPARD-PPARGC1A-NRF-TFAM mitochondriogenesis pathway are neither associated with muscle characteristics nor physical performance in elderly.* Ricyde-Revista Internacional De Ciencias Del Deporte, 2015. **11**(41): p. 196-208.
20. He, Z., et al., *NRF-1 genotypes and endurance exercise capacity in young Chinese men.* British Journal of Sports Medicine, 2008. **42**(5).
21. Montani, D., et al., *Pulmonary arterial hypertension in familial hemiplegic migraine with ATP1A2 channelopathy.* Eur Respir J, 2014. **43**(2): p. 641-3.

This is a post-peer-review, pre-copyedit version of an article published in Molecular Genetics and Genomics. The final authenticated version is available online at: https://doi.org/10.1007/s00438-019-01639-8

22. Panicker, V., et al., *A common variation in deiodinase 1 gene DIO1 is associated with the relative levels of free thyroxine and triiodothyronine.* J Clin Endocrinol Metab, 2008. **93**(8): p. 3075-81.
23. Fuku, N., et al., *Muscle-Related Polymorphisms (MSTN rs1805086 and ACTN3 rs1815739) Are Not Associated with Exceptional Longevity in Japanese Centenarians.* PLoS One, 2016. **11**(11): p. e0166605.
24. Chathoth, S., et al., *Association of Uncoupling Protein 1 (UCP1) gene polymorphism with obesity: a case-control study.* BMC Med Genet, 2018. **19**(1): p. 203.
25. Ramos, A.V., et al., *The contribution of FTO and UCP-1 SNPs to extreme obesity, diabetes and cardiovascular risk in Brazilian individuals.* BMC Med Genet, 2012. **13**: p. 101.
26. Brondani, L.A., et al., *The role of the uncoupling protein 1 (UCP1) on the development of obesity and type 2 diabetes mellitus.* Arq Bras Endocrinol Metabol, 2012. **56**(4): p. 215-25.
27. Rankinen, T., et al., *Effect of endothelin 1 genotype on blood pressure is dependent on physical activity or fitness levels.* Hypertension, 2007. **50**(6): p. 1120-5.
28. Panoulas, V.F., et al., *Polymorphisms of the endothelin-1 gene associate with hypertension in patients with rheumatoid arthritis.* Endothelium, 2008. **15**(4): p. 203-12.
29. Arora, P., et al., *Genetic polymorphisms of innate immunity-related inflammatory pathways and their association with factors related to type 2 diabetes.* BMC Med Genet, 2011. **12**: p. 95.
30. Liu, Y., et al., *IL-6 haplotypes, inflammation, and risk for cardiovascular disease in a multiethnic dialysis cohort.* J Am Soc Nephrol, 2006. **17**(3): p. 863-70.
31. Alves, A.J., et al., *GNAS A-1121G Variant is Associated with Improved Diastolic Dysfunction in Response to Exercise Training in Heart Failure Patients.* International Journal of Sports Medicine, 2013. **34**(3): p. 274-280.
32. Burguete-Garcia, A.I., et al., *Association of beta1 and beta3 adrenergic receptors gene polymorphisms with insulin resistance and high lipid profiles related to type 2 diabetes and metabolic syndrome.* Nutr Hosp, 2014. **29**(6): p. 1327-34.
33. Pan, Q., et al., *Variation at the Melanocortin 4 Receptor Gene and Response to Weight-Loss Interventions in the Diabetes Prevention Program.* Obesity, 2013. **21**(9): p. E520-E526.
34. Tejedor, M.T., et al., *The apolipoprotein E polymorphism rs7412 associates with body fatness independently of plasma lipids in middle aged men.* PLoS One, 2014. **9**(9): p. e108605.
35. Liang, K.Y., et al., *Exercise and Alzheimer's disease biomarkers in cognitively normal older adults.* Ann Neurol, 2010. **68**(3): p. 311-8.
36. Deeny, S.P., et al., *Exercise, APOE, and working memory: MEG and behavioral evidence for benefit of exercise in epsilon4 carriers.* Biol Psychol, 2008. **78**(2): p. 179-87.
37. Fredriksson, J., et al., *Variation in GYS1 interacts with exercise and gender to predict cardiovascular mortality.* PLoS One, 2007. **2**(3): p. e285.
38. Hennis, P.J., et al., *Genetic factors associated with exercise performance in atmospheric hypoxia.* Sports Med, 2015. **45**(5): p. 745-61.
39. Brown, B.M., et al., *Study protocol of the Intense Physical Activity and Cognition study: The effect of high-intensity exercise training on cognitive function in older adults.* Alzheimers Dement (N Y), 2017. **3**(4): p. 562-570.
40. Lin, Y., et al., *[Association between peroxisome proliferator-activated receptors gene polymorphism and essential hypertension].* Zhonghua Liu Xing Bing Xue Za Zhi, 2012. **33**(6): p. 597-601.
41. Uthurralt, J., et al., *PPARalpha L162V underlies variation in serum triglycerides and subcutaneous fat volume in young males.* BMC Med Genet, 2007. **8**: p. 55.
42. Andrulionyte, L., et al., *Single nucleotide polymorphisms of the peroxisome proliferator-activated receptor-alpha gene (PPARA) influence the conversion from impaired glucose tolerance to type 2 diabetes: the STOP-NIDDM trial.* Diabetes, 2007. **56**(4): p. 1181-6.

43. Doney, A.S., et al., *Association of common variation in the PPARA gene with incident myocardial infarction in individuals with type 2 diabetes: a Go-DARTS study.* Nucl Recept, 2005. **3**: p. 4.
44. Dong, C., et al., *Role of peroxisome proliferator-activated receptors gene polymorphisms in type 2 diabetes and metabolic syndrome.* World J Diabetes, 2015. **6**(4): p. 654-61.
45. Evans, D., et al., *A polymorphism, L162V, in the peroxisome proliferator-activated receptor alpha (PPARalpha) gene is associated with lower body mass index in patients with non-insulin-dependent diabetes mellitus.* J Mol Med (Berl), 2001. **79**(4): p. 198-204.
46. Flavell, D.M., et al., *Peroxisome proliferator-activated receptor alpha gene variation influences age of onset and progression of type 2 diabetes.* Diabetes, 2005. **54**(2): p. 582-6.
47. Bonetti, D.L. and W.G. Hopkins, *Variation in performance times of elite flat-water canoeists from race to race.* Int J Sports Physiol Perform, 2010. **5**(2): p. 210-7.
48. Munoz-Canoves, P., et al., *Interleukin-6 myokine signaling in skeletal muscle: a double-edged sword?* FEBS J, 2013. **280**(17): p. 4131-48.
49. Karanikolou, A., G. Wang, and Y. Pitsiladis, *Letter to the editor: A genetic-based algorithm for personalized resistance training.* Biol Sport, 2017. **34**(1): p. 31-33.
50. Woods, D., et al., *Elite swimmers and the D allele of the ACE I/D polymorphism.* Hum Genet, 2001. **108**(3): p. 230-2.
51. Galy, O., et al., *Influence of performance level on exercise-induced arterial hypoxemia during prolonged and successive exercise in triathletes.* Int J Sports Physiol Perform, 2008. **3**(4): p. 482-500.
52. Zervou, S., et al., *Increasing creatine kinase activity protects against hypoxia / reoxygenation injury but not against anthracycline toxicity in vitro.* PLoS One, 2017. **12**(8): p. e0182994.
53. Pitsiladis, Y.P., et al., *Athlome Project Consortium: a concerted effort to discover genomic and other "omic" markers of athletic performance.* Physiol Genomics, 2016. **48**(3): p. 183-90.