

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](https://doi.org/10.1007/s00438-019-01639-8)

Licence:
Other

[Link to output in Bond University research repository.](#)

Recommended citation(APA):
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. (2020). Genetic variants associated with exercise performance in both moderately trained and highly trained individuals. *Molecular Genetics and Genomics*, 295(2), 515-523. <https://doi.org/10.1007/s00438-019-01639-8>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

For more information, or if you believe that this document breaches copyright, please contact the Bond University research repository coordinator.

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

3
4 **Harvey NR**^{1,2}, Voisin S³, Dunn PJ², Sutherland H², Yan X³, Jacques M³, Papadimitriou ID³, Haseler LJ⁴, Ashton
5 KJ¹, Haupt LM², Eynon N³ and *Griffiths LR²

6
7 ¹Health Sciences and Medicine faculty, Bond University, Robina, Australia, QLD 4226

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

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

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

12
13 ***Corresponding author:** Prof. Lyn R. Griffiths, Genomics Research Centre, Institute of Health and Biomedical
14 Innovation, Queensland University of Technology, Brisbane QLD, Australia

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

16 Phone: +61 7 3138 6102

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

18 **NH:** 0000-0003-2035-8475, **SV:** 0000-0002-4074-7083, **HS:** 0000-0002-8512-1498, **XY:** 0000-0001-8547-4210,

19 **IP:** 0000-0001-5032-9937, **KA:** 0000-0001-6106-3425, **LH:** 0000-0002-7735-8110, **NE:** 0000-0003-4046-8276,

20 **LG:** 0000-0002-6774-5475

21 **Funding and Acknowledgements:**

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

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

29 **ABSTRACT**

30 Adaptation to exercise training is a complex trait that may be influenced by genetic variants. We identified 36
31 single nucleotide polymorphisms (SNPs) that had been previously associated with endurance or strength
32 performance, exercise-related phenotypes or exercise intolerant disorders. A MassARRAY multiplex genotyping
33 assay was designed to identify associations with these SNPs against collected endurance fitness phenotype
34 parameters obtained from 2 exercise cohorts (Gene SMART study; n=58 and Hawaiian Ironman Triathlon 2008;
35 n=115). These parameters included peak power output (PP), a time trial (TT), lactate threshold (LT), maximal
36 oxygen uptake (VO₂ max) in recreationally active individuals and a triathlon time to completion (Hawaiian
37 Ironman Triathlon cohort only). A nominal significance threshold of $\alpha < 0.05$ was used to identify 17 variants (11
38 in the Gene SMART population and 6 in the Hawaiian Ironman Triathlon cohort) which were significantly
39 associated with performance gains in highly trained individuals. The variant rs1474347 located in *Interleukin 6*
40 (*IL6*) was the only variant with a false discovery rate < 0.05 and was found to be associated with gains in VO₂
41 max (additional 4.016 mL/(kg·min) for each G allele inherited) after training in the Gene SMART cohort. In
42 summary, this study found further evidence to suggest that genetic variance can influence training response in a
43 moderately trained cohort and provides an example of the potential application of genomic research in the
44 assessment of exercise trait response.

45

46 **Key Words:**

47 Genetics, exercise, MassARRAY, replication, endurance

48 **INTRODUCTION**

49 Currently, robust identification of genetic variants associated with exercise phenotypes is limited by a lack of
50 reproducible results. Family and twin studies have estimated high heritability for various exercise performance
51 metrics (e.g. muscle mass: 40%, anaerobic power: 70-80%, aerobic exercise: 50%) [1]. However, a wide variety
52 of environmental (e.g. diet, sleep), psychological and epigenetic factors may also influence exercise responses
53 [2]. In addition, within-subject variability (i.e. the variable response of a given individual to the same exercise
54 training) considerably limits the identification of genetic variants with potentially small effects on exercise
55 response [3].

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

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

71

72

73 **MATERIALS AND METHODS**

74 **The Gene SMART cohort**

75 The Gene SMART (Skeletal Muscle Adaptive Response to Training) study design has previously been described
76 [9]. The study is ongoing with currently > 100 moderately-trained participants who were sampled for blood and
77 skeletal muscle (*vastus lateralis*) at several time points: before, immediately after and 3 hours after a single bout
78 of high-intensity endurance exercise (HIIE), and after 4 weeks of High-Intensity Interval Training (HIIT) [9].
79 Exercise-related phenotypic measurements were collected before and after the completion of the exercise training
80 intervention (e.g. Lactate Threshold (LT, in Watts), Peak Power output (PP, in Watts), maximal oxygen uptake
81 (VO_2max , in mL/min/kg body weight, from graded exercise tests), and a Time Trial measurement (TT, in min).
82 All participants gave informed consent and the study was approved by the Victoria University Ethics Committee
83 (Approval number: HRE13-233). Subsequently, the study was also approved by the Queensland University of
84 Technology (QUT) Human Research Ethics Committee (Approval number: 1600000342). All procedures
85 performed in studies involving human participants were in accordance with the ethical standards of the respective
86 institutions research committees, and with the 1964 Helsinki declaration and its later amendments or comparable
87 ethical standards. At the time of collection, $n = 77$ participants had participated in the study, with $n = 58$ completing
88 the entire 4-week training program. Genomic DNA was extracted and purified from whole blood using the
89 QIAamp DNA blood midi kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions for
90 participants that completed the study. Samples that failed genotypic analysis or had a large amount of missing
91 phenotypic data, were removed from further analysis, leaving a final sample size of $n = 52$ (Age = 30.95 ± 8.17
92 years). In the moderately trained cohort (Gene SMART, $n = 58$), we focused on the response to an exercise training
93 program (longitudinal analysis). Specifically, we measured the change in ($\Delta = \text{post} - \text{pre}$) measurement for each
94 endurance fitness trait as a representation of response to exercise training.

95 **Highly trained (Ironman) cohort**

96 Ironman triathlons consist of a 3.86 km swim, a 180.25 km bike ride, followed by the completion of a full marathon
97 (42.2 km). The 2008 Hawaiian Ironman Triathlon population has been previously described as an elite endurance
98 cohort based on their eligibility and participation in the event [10, 11]. Due to the intensity of this endurance event
99 only highly trained individuals that completed it were included in this study. To avoid genetic confounding, we
100 analysed only the triathlon participants who self-identified as male and Caucasian (Age = 43.81 ± 11.39 years).

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

111

112 **Genotype method and SNP selection**

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

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	<i>AMPD1</i> [†]	rs17602729	C	END	2	231	1	84	[14]
			T	POW	3	510	0	0	
6	<i>HFE</i>	rs1799945	G	END	2	148	-	-	[14]
6	<i>VEGFA</i>	rs2010963	C	END	1	942	-	-	[14]
11	<i>ACTN3</i>	rs1815739	C	END	4	560	14	3,039	[14]
			T	POW	12	1,484	5	498	
11	<i>UCP2</i>	rs660339	T	END	1	694	-	-	[14]
			C	POW	1	29	-	-	
19	<i>CKMM</i>	rs8111989	G	POW	2	233	-	-	[14]
21	<i>COL6A1</i>	rs35796750	T	END	1	661	-	-	[14]
22	<i>PPARα</i>	rs4253778	G	END	5	740	-	-	[14]
			C	POW	2	260	1	81	
Category 2: Mixed results									
1	<i>SGIP1</i>	rs9633417	C	EBH	2	2,838	-	-	[15]
			C	POW	-	-	1	753	
1	<i>LEPR</i>	rs1137101	A	END	-	-	1	846	[15-17]
			G	POW	1	242	-	-	
4	<i>PGC1α</i>	rs8192678	G	EBH	3	3676	-	-	[14]
			A	END	4	849	3	508	
4	<i>UCP1</i>	rs10440457	G	EBH	2	2,838	-	-	[15]
			-	POW	-	-	1	181	
4	<i>PGC1α</i>	rs6821591	T	END	1	235	-	-	[18]
5	<i>ADRB2</i>	rs1042713	A	POW	1	100	-	-	[14]

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	<i>NRF1</i>	rs6949152	G	END	1	102	1	75	[19, 20]
8	<i>ADRB3</i>	rs4994	C	END	1	100	1	81	[14]
14	<i>BDKRB2</i>	rs1799722	T	END	1	316	-	-	[14]
15	<i>NRF2</i>	rs7181866	G	END	2	129	1	89	[14]
15	<i>NRF2</i>	rs8031031	T	END	1	74	1	89	[14]
Category 3: Disease Associations and other									
1	<i>ATPIA2</i> [§]	rs28933400	T	HYP	-	-	1	388	[21]
1	<i>LEPR</i>	rs12405556	T	EBH	2	978	-	-	[15]
1	<i>DIO1</i>	rs2294512	A	THY	1	547	-	-	[22]
2	<i>MSTN</i>	rs1805086	G	POW	13	3,080	-	-	[10, 23]
4	<i>UCP1</i>	rs2270565	T	T2D	2	981	4	1,382	[24-26]
6	<i>EDN1</i>	rs5370	T	HYP	2	1,004	-	-	[27, 28]
6	<i>HLA-A</i>	rs1061235	T	END	1	32	-	-	[16]
7	<i>IL6</i>	rs1474347	A	T2D	1	10,775	-	-	[29, 30]
10	<i>ADRB1</i>	rs1801253	C	HYP T2D	1 1	61 947	- -	- -	[31, 32]
12	<i>IGF1</i>	rs121912430	T	OBE	1	502	-	-	[16]
15	<i>CYP19A1</i> [‡]	rs2470158	T	EBH	1	1,722	-	-	[15]
16	<i>Intronic</i>	rs238838	C	-	-	-	-	-	[16]
18	<i>MC4R</i>	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	<i>APOE</i>	rs7412	T	T2D END OBE	2 - -	9,314 - -	- 2 2	- 507 159	[34-39]
22	<i>PPARα</i>	rs1800206	G	HYP POW T2D	1 1 4	269 610 3,643	- - -	- - -	[40-46]
MT	<i>MTT</i>	rs199474700	G	END	1	46	-	-	[16]
MT	<i>MTND5</i>	rs28359178	A	END	1	46	-	-	[16]

128

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

130 **Phenotypes are the traits in which the SNP has been previously implicated. Previous studies and numbers of participants are shown and separated according to**
131 **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,

132 POW: Power/Strength, T2D: Diabetes, HYP: Hypertension, EBH: Exercise behaviour, OBE: Obesity, †: Inbuilt genotyping control, A1: Tested Allele

133 ‡ excluded in Gene SMART population.

134 § Excluded from highly trained population.

135

136 **Data processing**

137 The output files from the MassARRAY platform were converted to PLINK format and analysed for correct
138 genotypic identification (calling). For the Gene SMART and Ironman populations respectively, SNPs were
139 excluded from further analysis if they exceeded the following thresholds: 1) SNPs that had a calling rate < 80%
140 (>20% missing data) (n = 5, n = 3); 2) SNPs with a minor allele frequency < 2% (n = 1, n = 2); 3) SNPs determined
141 not in Hardy Weinberg equilibrium (n=1, n=1) [12]. Subsequent analysis was performed on n = 29 SNPs for the
142 Gene SMART population and n = 30 SNPs for the Ironman population.

143 **Statistical analysis**

144 We measured normality metrics (skewness and kurtosis) for each phenotype in both populations using the *ggplot2*,
145 *tidyverse* and *moments* packages in R, to determine if data transformation was necessary from the raw phenotypic
146 values. We used PLINK V1.90p to perform quantitative linear association tests (95% CI) with both dominant and
147 recessive models for each cohort, adjusting for age. An additive model was considered but did not differ from the
148 results obtained from the dominant model. As this was a candidate gene study, SNPs that had a raw *p*-value <
149 0.05 were considered nominally significant while variants that had an adjusted *p*-value (Benjamini-Hochberg
150 False Discovery Rate (FDR)) < 0.05 were considered significant. This adjustment method represents a good
151 balance between type I and type II errors and as such minimises false positive results. To avoid multiple testing
152 burdens with phenotypic traits, we used a separate hypothesis for each quantitative trait. Effect sizes were
153 determined using raw beta regression coefficient values interpreted as “how much a specific phenotype increased
154 for each additional X allele at the SNP of interest”.

155

156 **RESULTS**

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

160

161 **Table II: Summary of nominally significant variants associated with gains in endurance fitness after exercise training in the Gene SMART cohort.** Tests were
 162 performed for both dominant and recessive models for each trait: *Wpeak*: maximum ergometer intensity at stop (Watts), *LT*: Lactate Threshold (Watts), *VO₂max*: maximum
 163 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	<i>AMPD1</i>	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	<i>DIO1</i>	Intronic	DOM	0.042	0.043	0.393	-16.8
Δ-LT	14	rs1799722	T	<i>BDKRB2</i>	5'UTR variant	REC	0.33	0.027	0.24	-17.17
	15	rs8031031	T	<i>NRF2</i>	Intronic	REC	0.016	0.027	0.24	-37.34
	21	rs35796750	T	<i>COL6A1</i>	Intronic	REC	0.48	0.035	0.24	11.88
	19	rs7412	T	<i>APOE</i>	Missense	DOM	0.105	0.042	0.64	-11.95
	18	rs9965495	A	<i>MC4R</i>	Intronic	DOM	0.28	0.045	0.64	9.107
Δ-VO₂max	7	rs1474347	C	<i>IL6</i>	Intronic	REC	0.45	0.00087	0.018	-4.016
	11	rs660339	A	<i>UCP2</i>	Missense	REC	0.46	0.037	0.38	-2.835
	15	rs8031031	T	<i>NRF2</i>	Intronic	DOM	0.04	0.04	0.87	4.741
Δ-TT	6	rs1799945	G	<i>HFE</i>	Missense	DOM	0.21	0.019	0.58	101

164

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

166 *Rate*

167 [†]*P*-value adjusted for age

168

169

170 **Table III: All nominally significant variants associated with different triathlon event finishing times in the highly trained endurance cohort. SNPs were determined**
 171 *nominally significant under an arbitrary $\alpha < 0.05$ 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	<i>NRF1</i>	Intronic	REC	0.14	0.019	0.47	0.2459
	19	rs8111989	G	<i>CKMM</i>	Downstream variant	DOM	0.34	0.019	0.31	-0.0903
	6	rs1061235	T	<i>HLA-A</i>	Non-coding transcript	DOM	0.061	0.02	0.31	0.1392
Cycle time	1	rs9633417	A	<i>SGIP1</i>	Intronic	DOM	0.1	0.036	0.89	0.3238
	4	rs8192678	A	<i>PGC1α</i>	Missense	REC	0.36	0.038	0.40	0.2965
	4	rs6821591	A	<i>PGC1α</i>	Non-coding transcript	REC	0.45	0.041	0.40	-0.2977
Run time	1	rs9633417	A	<i>SGIP1</i>	Intronic	DOM	0.1	0.019	0.61	0.424
Total-time	1	rs9633417	A	<i>SGIP1</i>	Intronic	DOM	0.1	0.012	0.37	0.846

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

173 *Rate*

174 [†]*P-value adjusted for age*

175 Six variants in five distinct genes were nominally associated with time-to-completion of Ironman events: *Nuclear*
176 *Respiratory Factor 1 (NRF1: rs6949152)*, *Myostatin (MSTN: rs1805086)*, *Major Histocompatibility complex*
177 *class 1A (HLA-A: rs1061235)*, *Peroxisome proliferator-activated receptor gamma coactivator 1-alpha*
178 *(PPARGC1a: rs6821591, rs6821591)*, and *SH3 Domain GRB2 Like Endophilin Interacting Protein (SGIP1:*
179 *rs9633417)*. The results of our association testing did not change significantly when age was used as a covariate.

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

186 Interestingly, no variants were identified to be significantly associated with both time-to-completion in the
187 Ironman cohort and the response to endurance exercise training in the Gene SMART cohort. Only rs1474347 in
188 *IL6* passed correction for multiple testing using the BH-FDR method (FDR: 0.018). The C allele at rs1474347
189 was associated with VO₂max response within the Gene SMART study with an effect size of -4.016mL/(kg·min).

190

191 **DISCUSSION**

192 In the present study, we have successfully replicated previously associated exercise-related SNPs using the
193 combined data from highly trained and moderately trained cohorts. Our main findings identified the rs1474347 in
194 the *IL6* gene to be significantly associated with gains in VO₂ max in the Gene SMART cohort after multiple-
195 testing statistical corrections. In addition, 17 genetic variants were found to be associated with either elite
196 performance or responses to exercise, however, none of these variants were common between these cohorts.

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

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

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

204 Located in an intron of the *IL6* gene, the rs1474347 variant has been previously associated with T2D traits in a
205 large-scale study (n = 10,775). The IL6 protein is a pro-inflammatory cytokine with myokinetic (i.e. excreted from
206 skeletal muscle) functions and is responsible for triggering and maintaining immune processes following post-
207 exercise muscle damage [48]. We found that the C allele at this locus negatively affected the exercise response to
208 the VO₂max phenotype ($\beta = -4.016\text{mL/kg}\cdot\text{min}$) and therefore a homozygous C/C genotype would result in a
209 VO₂max loss of -8.032 mL/kg·min.. The rs1800795 coding variant within the *IL6* gene has shown mixed evidence
210 of exercise associations i.e. variant C = athleticism, G = power [14, 49]. Interestingly, further analysis identified
211 the rs1474347 C allele to be in strong linkage disequilibrium (LD) with the C allele of rs1800795 ($R^2 = 0.96$). As
212 such, it is feasible that the LD identified between these variants has contributed to the mixed evidence reported
213 for association studies implicating the latter variant (rs1800795) in *IL6* for exercise traits. We propose that the
214 rs1474347 variant may reduce the expression of IL6 during acute muscle damage and therefore cause a reduced
215 local immune response leading to loss of skeletal muscle remodelling and repair. In addition this variant is also
216 located 2kb upstream of an uncharacterised long non-coding RNA (lncRNA; *LOC541472*) and therefore, variants
217 in this region may affect the IL-6 pro-inflammatory pathway or post-translational epigenetic and regulatory
218 processes.

219 **Association between genetic variants and Ironman performance**

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

229 alpha (HIF1 α). This would cause reduced oxidative metabolic processing and therefore lead to the increase in
230 swim time that is associated with the variant ($\beta = 0.2459$ hours). Additionally, the CKMM protein has been shown
231 to exhibit protective effects during mild hypoxia and therefore e hypothesised that the rs8111989 variant would
232 increase the functionality of the CKMM protein, resulting in protection against re-oxygenation induce muscle
233 damage and decreased swim time [52].

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

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

251

252 **Data Availability**

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

255 **Conflicts of interest**

256 The authors declare that they have no conflicts of interest

257 **Ethical Approval**

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

263

264 **REFERENCES**

265 1. Issurin, V.B., *Evidence-Based Prerequisites and Precursors of Athletic Talent: A Review*. Sports
 266 Med, 2017. **47**(10): p. 1993-2010.

267 2. den Hoed, M., et al., *Heritability of objectively assessed daily physical activity and sedentary*
 268 *behavior*. Am J Clin Nutr, 2013. **98**(5): p. 1317-25.

269 3. Voisin, S., et al., *Statistical Considerations for Exercise Protocols Aimed at Measuring*
 270 *Trainability*. Exerc Sport Sci Rev, 2019. **47**(1): p. 37-45.

271 4. Yan, X., et al., *ACE I/D gene variant predicts ACE enzyme content in blood but not the ACE,*
 272 *UCP2, and UCP3 protein content in human skeletal muscle in the Gene SMART study*. J Appl
 273 Physiol (1985), 2018. **125**(3): p. 923-930.

274 5. Houweling, P.J., et al., *Is evolutionary loss our gain? The role of ACTN3 p.Arg577Ter (R577X)*
 275 *genotype in athletic performance, ageing, and disease*. Hum Mutat, 2018. **39**(12): p. 1774-
 276 1787.

277 6. Bouchard, C., *Overcoming Barriers to Progress in Exercise Genomics*. Exercise and Sport
 278 Sciences Reviews, 2011. **39**(4): p. 212-217.

279 7. McCoy, J., et al., *Pathophysiology of exercise intolerance in chronic diseases: the role of*
 280 *diminished cardiac performance in mitochondrial and heart failure patients*. Open Heart, 2017.
 281 **4**(2): p. e000632.

282 8. Buford, T.W., M.D. Roberts, and T.S. Church, *Toward exercise as personalized medicine*. Sports
 283 Med, 2013. **43**(3): p. 157-65.

284 9. Yan, X., et al., *The gene SMART study: method, study design, and preliminary findings*. BMC
 285 Genomics, 2017. **18**(Suppl 8): p. 821.

286 10. Grealy, R., et al., *Evaluation of a 7-Gene Genetic Profile for Athletic Endurance Phenotype in*
 287 *Ironman Championship Triathletes*. PLoS One, 2015. **10**(12): p. e0145171.

288 11. Grealy, R., et al., *The genetics of endurance: frequency of the ACTN3 R577X variant in Ironman*
 289 *World Championship athletes*. J Sci Med Sport, 2013. **16**(4): p. 365-71.

290 12. Ellis, J.A. and B. Ong, *The MassARRAY((R)) System for Targeted SNP Genotyping*. Methods Mol
 291 Biol, 2017. **1492**: p. 77-94.

292 13. Jurinke, C., et al., *Automated genotyping using the DNA MassArray technology*. Methods Mol
 293 Biol, 2002. **187**: p. 179-92.

294 14. Ahmetov, II and O.N. Fedotovskaya, *Current Progress in Sports Genomics*. Adv Clin Chem,
 295 2015. **70**: p. 247-314.

296 15. De Moor, M.H., et al., *Genome-wide association study of exercise behavior in Dutch and*
 297 *American adults*. Med Sci Sports Exerc, 2009. **41**(10): p. 1887-95.

298 16. Rankinen, T., et al., *The human gene map for performance and health-related fitness*
 299 *phenotypes: the 2003 update*. Med Sci Sports Exerc, 2004. **36**(9): p. 1451-69.

300 17. Walsh, S., et al., *Leptin and leptin receptor genetic variants associate with habitual physical*
 301 *activity and the arm body composition response to resistance training*. Gene, 2012. **510**(1): p.
 302 66-70.

303 18. He, Z.H., et al., *PGC-related gene variants and elite endurance athletic status in a Chinese*
 304 *cohort: a functional study*. Scand J Med Sci Sports, 2015. **25**(2): p. 184-95.

305 19. Garatachea, N., et al., *Genetic variants in the PPAR α -PPARGC1A-NRF-TFAM*
 306 *mitochondriogenesis pathway are neither associated with muscle characteristics nor physical*
 307 *performance in elderly*. Ricyde-Revista Internacional De Ciencias Del Deporte, 2015. **11**(41): p.
 308 196-208.

309 20. He, Z., et al., *NRF-1 genotypes and endurance exercise capacity in young Chinese men*. British
 310 Journal of Sports Medicine, 2008. **42**(5).

311 21. Montani, D., et al., *Pulmonary arterial hypertension in familial hemiplegic migraine with*
 312 *ATP1A2 channelopathy*. Eur Respir J, 2014. **43**(2): p. 641-3.

- 313 22. Panicker, V., et al., *A common variation in deiodinase 1 gene DIO1 is associated with the*
314 *relative levels of free thyroxine and triiodothyronine.* J Clin Endocrinol Metab, 2008. **93**(8): p.
315 3075-81.
- 316 23. Fuku, N., et al., *Muscle-Related Polymorphisms (MSTN rs1805086 and ACTN3 rs1815739) Are*
317 *Not Associated with Exceptional Longevity in Japanese Centenarians.* PLoS One, 2016. **11**(11):
318 p. e0166605.
- 319 24. Chathoth, S., et al., *Association of Uncoupling Protein 1 (UCP1) gene polymorphism with*
320 *obesity: a case-control study.* BMC Med Genet, 2018. **19**(1): p. 203.
- 321 25. Ramos, A.V., et al., *The contribution of FTO and UCP-1 SNPs to extreme obesity, diabetes and*
322 *cardiovascular risk in Brazilian individuals.* BMC Med Genet, 2012. **13**: p. 101.
- 323 26. Brondani, L.A., et al., *The role of the uncoupling protein 1 (UCP1) on the development of obesity*
324 *and type 2 diabetes mellitus.* Arq Bras Endocrinol Metabol, 2012. **56**(4): p. 215-25.
- 325 27. Rankinen, T., et al., *Effect of endothelin 1 genotype on blood pressure is dependent on physical*
326 *activity or fitness levels.* Hypertension, 2007. **50**(6): p. 1120-5.
- 327 28. Panoulas, V.F., et al., *Polymorphisms of the endothelin-1 gene associate with hypertension in*
328 *patients with rheumatoid arthritis.* Endothelium, 2008. **15**(4): p. 203-12.
- 329 29. Arora, P., et al., *Genetic polymorphisms of innate immunity-related inflammatory pathways*
330 *and their association with factors related to type 2 diabetes.* BMC Med Genet, 2011. **12**: p. 95.
- 331 30. Liu, Y., et al., *IL-6 haplotypes, inflammation, and risk for cardiovascular disease in a multiethnic*
332 *dialysis cohort.* J Am Soc Nephrol, 2006. **17**(3): p. 863-70.
- 333 31. Alves, A.J., et al., *GNAS A-1121G Variant is Associated with Improved Diastolic Dysfunction in*
334 *Response to Exercise Training in Heart Failure Patients.* International Journal of Sports
335 Medicine, 2013. **34**(3): p. 274-280.
- 336 32. Burguete-Garcia, A.I., et al., *Association of beta1 and beta3 adrenergic receptors gene*
337 *polymorphisms with insulin resistance and high lipid profiles related to type 2 diabetes and*
338 *metabolic syndrome.* Nutr Hosp, 2014. **29**(6): p. 1327-34.
- 339 33. Pan, Q., et al., *Variation at the Melanocortin 4 Receptor Gene and Response to Weight-Loss*
340 *Interventions in the Diabetes Prevention Program.* Obesity, 2013. **21**(9): p. E520-E526.
- 341 34. Tejedor, M.T., et al., *The apolipoprotein E polymorphism rs7412 associates with body fatness*
342 *independently of plasma lipids in middle aged men.* PLoS One, 2014. **9**(9): p. e108605.
- 343 35. Liang, K.Y., et al., *Exercise and Alzheimer's disease biomarkers in cognitively normal older*
344 *adults.* Ann Neurol, 2010. **68**(3): p. 311-8.
- 345 36. Deeny, S.P., et al., *Exercise, APOE, and working memory: MEG and behavioral evidence for*
346 *benefit of exercise in epsilon4 carriers.* Biol Psychol, 2008. **78**(2): p. 179-87.
- 347 37. Fredriksson, J., et al., *Variation in GYS1 interacts with exercise and gender to predict*
348 *cardiovascular mortality.* PLoS One, 2007. **2**(3): p. e285.
- 349 38. Hennis, P.J., et al., *Genetic factors associated with exercise performance in atmospheric*
350 *hypoxia.* Sports Med, 2015. **45**(5): p. 745-61.
- 351 39. Brown, B.M., et al., *Study protocol of the Intense Physical Activity and Cognition study: The*
352 *effect of high-intensity exercise training on cognitive function in older adults.* Alzheimers
353 Dement (N Y), 2017. **3**(4): p. 562-570.
- 354 40. Lin, Y., et al., *[Association between peroxisome proliferator-activated receptors gene*
355 *polymorphism and essential hypertension].* Zhonghua Liu Xing Bing Xue Za Zhi, 2012. **33**(6): p.
356 597-601.
- 357 41. Uthurralt, J., et al., *PPARalpha L162V underlies variation in serum triglycerides and*
358 *subcutaneous fat volume in young males.* BMC Med Genet, 2007. **8**: p. 55.
- 359 42. Andrulionyte, L., et al., *Single nucleotide polymorphisms of the peroxisome proliferator-*
360 *activated receptor-alpha gene (PPARA) influence the conversion from impaired glucose*
361 *tolerance to type 2 diabetes: the STOP-NIDDM trial.* Diabetes, 2007. **56**(4): p. 1181-6.

- 362 43. Doney, A.S., et al., *Association of common variation in the PPARA gene with incident*
363 *myocardial infarction in individuals with type 2 diabetes: a Go-DARTS study*. Nucl Recept,
364 2005. **3**: p. 4.
- 365 44. Dong, C., et al., *Role of peroxisome proliferator-activated receptors gene polymorphisms in*
366 *type 2 diabetes and metabolic syndrome*. World J Diabetes, 2015. **6**(4): p. 654-61.
- 367 45. Evans, D., et al., *A polymorphism, L162V, in the peroxisome proliferator-activated receptor*
368 *alpha (PPARalpha) gene is associated with lower body mass index in patients with non-insulin-*
369 *dependent diabetes mellitus*. J Mol Med (Berl), 2001. **79**(4): p. 198-204.
- 370 46. Flavell, D.M., et al., *Peroxisome proliferator-activated receptor alpha gene variation influences*
371 *age of onset and progression of type 2 diabetes*. Diabetes, 2005. **54**(2): p. 582-6.
- 372 47. Bonetti, D.L. and W.G. Hopkins, *Variation in performance times of elite flat-water canoeists*
373 *from race to race*. Int J Sports Physiol Perform, 2010. **5**(2): p. 210-7.
- 374 48. Munoz-Canoves, P., et al., *Interleukin-6 myokine signaling in skeletal muscle: a double-edged*
375 *sword?* FEBS J, 2013. **280**(17): p. 4131-48.
- 376 49. Karanikolou, A., G. Wang, and Y. Pitsiladis, *Letter to the editor: A genetic-based algorithm for*
377 *personalized resistance training*. Biol Sport, 2017. **34**(1): p. 31-33.
- 378 50. Woods, D., et al., *Elite swimmers and the D allele of the ACE I/D polymorphism*. Hum Genet,
379 2001. **108**(3): p. 230-2.
- 380 51. Galy, O., et al., *Influence of performance level on exercise-induced arterial hypoxemia during*
381 *prolonged and successive exercise in triathletes*. Int J Sports Physiol Perform, 2008. **3**(4): p.
382 482-500.
- 383 52. Zervou, S., et al., *Increasing creatine kinase activity protects against hypoxia / reoxygenation*
384 *injury but not against anthracycline toxicity in vitro*. PLoS One, 2017. **12**(8): p. e0182994.
- 385 53. Pitsiladis, Y.P., et al., *Athlome Project Consortium: a concerted effort to discover genomic and*
386 *other "omic" markers of athletic performance*. Physiol Genomics, 2016. **48**(3): p. 183-90.

387