DOCTORAL THESIS

Identifying the genomic predictors of exercise-induced adaptation.

Harvey, Nicholas

Award date:
2022

IDENTIFYING THE GENOMIC PREDICTORS OF EXERCISE-INDUCED ADAPTATION

Nicholas Robert Harvey

BBiomedSc (Hons)

Associate Professor Kevin Ashton, Professor Lyn Griffiths, Associate Professor Nir Eynon, Associate Professor Larisa Haupt, Professor Nuala Byrne

Submitted in total fulfilment of the requirements for the degree of

Doctor of Philosophy (Research)

Faculty of Health Sciences and Medicine

Bond University

September 2020

This research was supported by an Australian Government Research Training Program Scholarship.

Abstract

Exercise remains the most efficient way to maintain health in normative states and increases the likelihood of recovery from a host of disease states. Even so the exact mechanisms by which these changes occur within skeletal muscle remain unknown. Whilst there have been many exercise studies that examine endurance exercise and consequently identified genes (*PPARGC1A*, *HIF1A*) and processes (mitochondrial biogenesis/function, lipid metabolism, angiogenesis) involved in adaptation, very few were centred around high intensity interval training (HIIT). Of these, even less included more than ten participants, analysed young/lean/healthy individuals, examined Omic level data, or aimed to tie together multiple molecular datasets to more robustly identify molecular processes involved in adaptation to HIIT. A recent meta-analysis aimed to address the limitations within the current literature and was able to discern many genes and molecular processes associated in various forms of training. However molecular changes regarding HIIT remained elusive with only two studies examining large-scale gene expression changes. As such, the Genes and Skeletal Muscle Adaptive Response to Training (Gene SMART) study was designed to alleviate these limitations and further the knowledge of molecular exercise adaptation processes. The Gene SMART study represents a moderately trained, longitudinal HIIT cohort. Briefly, ~60 participants completed the study, wherein muscle and blood samples were taken at baseline (PRE), immediately following a single bout of high intensity interval exercise (HIIE) (P0), three-hours post HIIE (P3), and four-week post HIIT (4WP). This thesis represents a collaborative research scheme between Bond University, Victoria University, and Queensland University of Technology. The aim of this thesis was to identify robust genetic and molecular determinants of acute response to HIIE and chronic responses to HIIT. For the purposes of this thesis, the study contains four related projects, each centred around a different molecular aspect of adaptation: nuclear genetics, mitochondrial genetics, transcriptomics, and epigenetics. Therefore, this thesis has achieved and reported key findings from each of these projects and is detailed in the following.

Firstly, the study aimed to identify associations between implicated genetic variants with response phenotypes from the Gene SMART cohort. Further, we utilised a highly trained (2008

Konoha Ironman Triathlon) exercise population to identify variants solely associated with either moderate or highly trained individuals. A MassARRAY SNP genotyping plex was developed that contained 36 genetic variants within three categories; 1) previously strongly associated with training, 2) previously associated but with both positive and negative results, and 3) associated with exercise intolerant disorders,. The *rs1474347* G allele within the *IL6* gene was found to be associated with an increased aerobic exercise capacity following training. Further, there was no overlap between the nominally significant variants associated with the highly trained and moderately trained cohorts indicating a large difference in the adaptation processes between fitness levels.

The next focus of the study was to sequence mitochondrial genomes to discover point variants and mitochondrial regional haplogroups associated with exercise response. As a secondary aim of this study, the design and optimisation of a high throughput method for the sequencing of mitochondrial genomes was utilised. Point variations were also examined within nuclear encoded mitochondrial related genes to ensure any additional variation in mitochondrial function was considered. Unfortunately, the study lacked significant statistical power to identify associations between mitochondrial variation and exercise response phenotypes. When regional haplogroups were examined, the number of individual haplogroups was identified to be equivalent to the number of participants and prevented association testing. Whilst this did not yield any significant results, this work highlighted the requirement for accurate representation of the mitochondrial genome, rather than just hypervariable region sequencing, to obtain haplogroup information. A total of 4,325 nuclear encoded variants were identified within mitochondrial related genes that passed a nominal threshold of $\alpha<0.05$. Nine nuclear encoded variants in eight separate genes were found to be associated with exercise responses (FDR<0.05) *(rs11061368: DIABLO, rs113400963: FAM185A, rs6062129 and rs6121949: MTG2, rs7231304: AFG3L2, rs2041840: NDUFAF7, rs7085433: TIMM23, rs1063271: SPTLC2, rs2275273: ALDH18A1).* The data generated from this study suggest novel nuclear-encoded SNPs and mitochondrial pathways associated with exercise response phenotypes.

The third aim of this project utilised Ion Proton transcriptome sequencing to identify genes and molecular pathways differentially regulated in skeletal muscle following a single

bout of HIIE and four-weeks of HIIT. The transcription factors, *MYC*, *FOS*, and *JUN*, were identified to be highly upregulated immediately (P0) following a single bout of HIIE. Further, these remained upregulated after three-hours (P3), but the effect was attenuated following four-weeks of training (4WP). These findings replicated upregulation of previously discovered exercise inducible genes (*PPARGC1a, MSTN, CKM*). To discern biological meaning, Gene Set Enrichment Analysis (GSEA) was performed to determine enriched biological processes (BPs) within each exercise time point. Immune BPs were highly upregulated in immediate and three-hour acute HIIE. Interestingly, the number of differentially expressed genes did not correlate with the number of enriched BPs as chronic HIIT (4WP) contained the largest number of enriched gene ontology terms within the data. A longitudinal switch from immediate stress response inducible BPs to prolonged biological function was observed with protein targeting and energy metabolism more evident after HIIT (4WP). The next aim was to address the transcriptomic responses pertaining to previous chapters within this thesis framework. This analysis found that repression of *IL6* at chronic HIIT may be prevalent within BPs corresponding to cytokine production. The *MICAL-L2* gene was identified as an important mediator between insulin signalling, actin filament polymerisation, and glucose transport into skeletal muscle. Lastly, the *METT-L12* and *JMJD6* genes were identified as likely important mediators of epigenetic chromatin remodelling in chronic HIIT.

To assess global DNA methylation levels in skeletal muscle, 850K EPIC arrays were utilised. A subset (n=19) of individuals from the Gene SMART study were used to assess the immediate (P0), and four-weeks of HIIT (4WP). The global DNA hypomethylation event commonly described at immediate exercise time points was observed and provided confidence in the findings. Following differential methylation analysis (delta Beta ($\Delta\beta$) $\pm$2%, FDR<0.05), we identified a total of 1,138 (746 hypermethylated and 392 hypomethylated differentially methylated probes (DMPs)) for acute HIIE. For chronic HIIT, 7,470 probes were differentially methylated, of which, 2,371 DMPs were hypermethylated and 5,099 hypomethylated. To prevent the use of arbitrary thresholds that may induce false positive data, *mCSEA* (methylation set enrichment analysis) was performed to discern coordinated shifts in methylation within gene promoter regions and their negative correlation with gene expression (i.e. upregulated gene expression with corresponding promoter hypermethylation). We identified 122, and 276 significant (FDR<0.05) genes that were likely to be regulated via promoter methylation for acute HIIE (P0) and chronic HIIT (4WP) respectively. The correlated genes for acute HIIE

were extremely similar to the sole transcriptome data, where transcription factors such as *JUNB, EGR1, FOS, FOSB, EGR3,* and *MYC*, were upregulated with corresponding promoter hypomethylation. For chronic HIIT, the correlated genes seemed to suggest a more functional adaptive response involving upregulation/hypermethylation of RNA processing (*BRUNOL6*) and protein localisation genes (*DZIP1L*), and downregulation/hypermethylation of genes involved in calcium signalling (*MYOZ1, HRC*) and muscle differentiation and growth (*SPEG, PPAPDC3*). The methylation regulated BPs associated with acute HIIE overlapped with the clusters identified from the sole transcriptome analysis (response to reactive oxygen species, processes involved in development, skeletal muscle tissue development, and response to peptide). For chronic HIIT, both up and downregulated were identified. Upregulated/hypermethylated genes mapped to clusters contained similar immune BPs to the sole transcriptome data (regulation of leukocyte differentiation, regulation of myeloid differentiation, negative regulation of leukocyte differentiation, neutrophil activation, negative regulation of leukocyte apoptosis, and regulation of leukocyte proliferation). Whilst downregulated/hypomethylated genes mapped to clusters associated with muscle contraction and muscle tissue development, which represented a novel finding for response to sustained HIIT.

Whilst this study generated large-scale datasets, there were several limitations within the study design and analysis methodologies utilised. 1) The sample size used for the study, while comparatively large, was too underpowered to gain genome wide significant findings from the genomic data generated. This was addressed through limiting the genomic data to the mitochondrial related genes of interest. 2) The Gene SMART sub-cohort utilised was all male, and as such, only representative of half of the available healthy population. 3) There were constraints within the analysis approaches utilised for the study. As yet, no robust analysis pipeline for MultiOmics applications has been developed. 4) Skeletal muscle cellular complexity, specifically myonucleation (more than one myonuclei per fibre) and microenvironment cell heterogeneity (endothelial cells, mesenchymal progenitors (fibro-adipogenic progenitors (FAP) cells), and resident immune cells).

Future directions for this research study are as follows. 1) The results gained from Omics analysis must be replicated within other exercise cohorts. 2) Female populations should be

analysed to discern biological differences in exercise response between sexes. 3) Novel data analysis techniques such as data imputation and deconvolution analysis should be employed to increase the statistical power of the studies, and ensure findings are representative of skeletal muscle myonuclei respectively. 4) Proteomic approaches should be utilised to functionally assess the findings within this thesis.

The findings from this thesis represent novel and exciting results for the field of exercise science at large. Further, the results discussed within the framework of this thesis represent a summary of the overall findings from each of the projects. The Omic level data generated from this thesis will continue to be useful for the exercise genetics community and will assist with future meta-analyses that aim to delineate the adaptive response to training. The projects within this study were able to identify novel genes and biological processes, and replicate previous findings, emphasizing the requirement and statistical power gained from strong studies within the field of exercise science (longitudinal, paired design, healthy population, physiological and biochemical data).

Keywords

Genomics, predicting, exercise, outcomes, adaptations, transcriptomics, epigenetics, genes, methylation, RNA, DNA, Omics, MassARRAY, high-intensity, training, study, Gene SMART, strength, endurance, VO_2peak, mitochondrial genome, respiration, SNPs, fitness, ability, health, predictive

Declaration by Author

This thesis is submitted to Bond University in fulfilment of the requirements of the degree of Doctor of Philosophy (Research).

This thesis represents my own original work towards this research degree and contains no material that has previously been submitted for a degree or diploma at this University or any other institution, except where due acknowledgment is made.

Full Name Nicholas Robert Harvey

Signature

Research Outputs

Publications arising from this thesis:

1. **Harvey NR**, Voisin S, Dunn PJ, Sutherland H, Yan X, Papadimitriou ID, Ashton K, Haseler LJ, Haupt LM, Eynon N, and *Griffiths LR, (2020) Multiple genetic variants associated with exercise performance in both moderately trained and highly trained individuals. *Molecular Genetics and Genomics Jan 2020, DOI: 10.1007/s00438-019-01639-8*

2. **Harvey NR**, Voisin S, Lea RA, Yan X, Benton MC, Haupt LM, Ashton KJ, Eynon N[#] and Griffiths LR[#], (2020) Investigating the influence of mtDNA and nuclear encoded mitochondrial variants on high intensity interval training outcomes [#] Co-last authors. *Scientific Reports June 2020, DOI: 10.1038/s41598-020-67870-1*

Related Publications:

1. Voisin S, **Harvey NR**, Haupt LM, Griffiths LR, Ashton KJ, Coffey VG, Doering TM, Thompson JM, Benedict C, Cedernaes J, Lindholm ME, Craig JM, Rowlands DS, Sharples AP, Horvath S, Eynon N. (2020) An epigenetic clock for human skeletal muscle. *Journal of Cachexia, Sarcopenia, and Muscle Jan 31, 2020, DOI: 10.1002/jcsm.12556*

2. **N.R. Harvey [*], C.L. Albury[*],** M.C. Benton, D.A. Eccles, S. Stuart, Connell J, H.G. Sutherland, R.J.N. Allcock, R.A. Lea, L.M. Haupt and L.R. Griffiths. (2019) Ion torrent high throughput mitochondrial genome sequencing (HTMGS). ***** Co-first authors. *Plos One, Nov 15, 2019, DOI: 10.1371/journal.pone.0224847*

Contributing Author:

1. Camilla J. Williams, Brendon J. Gurd, Jacob T. Bonafiglia, Sarah Voisin, Zhixiu Li, **Nicholas Harvey**, Ilaria Croci, Jenna Taylor, Trishan Gajanand, Joyce S. Ramos, Robert G. Fassett, Jonathan P. Little, Monique E. Francois, Christopher M. Hearon Jr, Satyam Sarma, Sylvan L.J.E. Janssen, Emeline M. Van Craenenbroeck, Paul Beckers, Véronique A. Cornelissen, Nele Pattyn, Erin J. Howden, Shelley E. Keating, Anja Bye, Dorthe Stensvold, Ulrik Wisloff, Ioannis Papadimitriou, Xu Yan, David J Bishop, Nir Eynon, *Jeff S. Coombes (2019) A Multi-Center Comparison ofVO2peakTrainability Between Interval Training and Moderate Intensity Continuous Training, *Frontiers Physiology. DOI: 10.3389/fphys.2019.00019, ISSN1664-042X*

2. **Camilla J Williams*, Zhixiu Li*, Nicholas Harvey***, Rodney A Lea, Brendon J Gurd, Jacob T Bonafiglia, Ioannis Papadimitriou, Macsue Jacques, Ilaria Croci, Dorthe Stensvold, Ulrik Wisloff, Jenna L Taylor, Trishan Gajanand, Emily R Cox, Joyce S Ramos, Robert G Fassett, Jonathan P Little, Monique E Francois, Christopher M Hearon, Satyam Sarma, Sylvan LJE Janssen, Emeline M Van Craenenbroeck, Paul Beckers, Véronique A Cornelissen, Erin J Howden, Shelley E Keating, Xu Yan, David J Bishop, Anja Bye, Larisa M Haupt, Lyn R Griffiths, Kevin J Ashton, Matthew A Brown, Luciana Torquati, Nir Eynon, Jeff S Coombes. (2021) Genome wide association study of response to interval and continuous training: the Predict-HIIT study. **Co-first authors. Journal of Biomedical Science, May 13, 2021, DOI: 10.1186/s12929-021-00733-7*

3. Anna Jaeschke, **Nicholas R Harvey**, Mikhail Tsurkan, Carsten Werner, Lyn R Griffiths, Larisa M Haupt, Laura J Bray. (2021) Techniques for RNA extraction from cells cultured in starPEG-heparin hydrogels, *Open Biology, June 2, 2021, DOI: 10.1098/rsob.200388*

Conference Abstracts

1. 2019, **Harvey N.R.,** Thompson J.L., Voisin S., Benton M., Haupt L.M., Eynon N., Griffiths L.R., Ashton K.J., Novel transcriptional pathways associated with response to high intensity training in the Gene SMART study. Bond University Postgraduate and

Medical student conference, Gold Coast, QLD, Australia, 16th October, **Oral Presentation**

2. 2019, **Harvey NR,** Voisin S, Lea RA, Yan X, Benton MC, Papadimitriou ID, Eccles DA, Haupt LM, Ashton KJ, Eynon N and Griffiths LR, IHBI Inspires Annual scientific Conference, Brisbane, QLD, Australia, 12-13th August

3. 2018, **Harvey N.R.**, Albury C.L., Benton M.C., Eccles D., Lee R.A., Haupt L.M., Ashton K.J, Eynon N., and Griffiths L.R. Identifying genetic variants associated with phenotypic traits using high throughput mitochondrial genome sequencing in the Gene SMART study. IHBI Inspires Annual scientific Conference, Brisbane, QLD, Australia, 16th-17th August

4. 2018, **Harvey N.R.**, Albury C.L., Benton M.C., Eccles D., Lee R.A., Haupt L.M., Ashton K.J, Eynon N., and Griffiths L.R. Identifying variants associated with exercise response using whole mitochondrial genome sequencing in the GeneSMART cohort**.** Human Genetics Society of Australasia Annual Scientific Meeting, Inborn Errors of Metabolism Special Interest Group, Sydney, NSW, Australia, 4th-7th August, **Oral Presentation**

5. 2017 **Harvey N.R.,** Haupt L.M., Sutherland H., Albury C.L., Ashton K.J., Eynon N., Griffiths L.R Significant SNPs associated with elite endurance performance in the 2008 Hawaiian Ironman Triathlon population, IHBI Inspires Annual scientific Conference, Brisbane, QLD, Australia, 23rd-24th August, **Oral Presentation**

6. 2017, **Harvey N.R.,** Haupt L.M., Sutherland H., Albury C.L., Ashton K.J., Eynon N., Griffiths L.R. The MassARRAY system reveals significant polymorphisms associated with endurance phenotype and exercise response, Human Genetics Society of Australasia Annual Scientific Meeting, Brisbane, QLD, Australia, 5th-8th August

7. 2017, **Harvey N.R.,** Haupt L.M., Sutherland H., Albury C.L., Ashton K.J., Eynon N., Griffiths L.R Significant SNPs associated with elite endurance performance in the 2008 Hawaiian Ironman Triathlon population, Australian Society of Medical Research Postgraduate Student Conference, Brisbane, QLD, Australia, 31st May

8. 2016, **Harvey N.R.,** Haupt L.M., Eynon N., Ashton K.J., & Griffiths L.R. Identifying the transcriptomic predictors of exercise adaptations: Utilising the Gene SMART study, Australian Society of Medical Research National Conference, Gold Coast, QLD, Australia, 13th-15th November

9. 2016, **Harvey N.R**., Haupt L.M., Eynon N., Ashton K.J., & Griffiths L.R. Analysing the skeletal muscle transcriptome in a moderately trained cohort to ascertain the genetics of exercise adaption. IHBI Inspires Higher Degree by Research Conference, Robina, QLD, Australia, 10th-11th November

10. 2016, **Harvey N.R.,** Haupt L.M., Eynon N., Ashton K.J., & Griffiths L.R. Analysing the skeletal muscle transcriptome in a moderately trained cohort to ascertain the genetics of exercise adaption. Bond University Higher Degree by Research Conference, Robina, QLD, Australia, **Oral Presentation**

Other:

1. 2017, Human Genetics Society of Australasia, QLD Branch Meeting, Oral presentation "Identifying the genomic predictors of exercise adaptations: The Gene SMART study"

2. 2016, Bond University CRN for Advancing Exercise and Sports Science, Closing Meeting, Oral presentation "The Gene SMART study progress report and closing updates"

3. 2016, Bond University CRN for Advancing Exercise and Sports Science, Annual Meeting, Oral presentation "The Gene SMART study project update and timeline"

Awards and Scholarships Arising from PhD

1. 2016-2019, Bond University CRN-AESS Scholarship

2. 2018, Human Genetics Society of Australasia (HGSA), Annual Scientific Meeting, Sydney (Australia), Travel Scholarship

Copyright Declaration

This thesis makes careful note of all sections which have been previously published, along with relevant copyright information.

Copyright permission has been obtained from the publisher for the one (1) manuscript accepted for publication from the publishers. Copies of the permissions are referenced in the preface of **Chapter 4** within the thesis.

Harvey NR, Voisin S, Dunn PJ, Sutherland H, Yan X, Jacques M, et al. Genetic variants associated with exercise performance in both moderately trained and highly trained individuals. Mol Genet Genomics. 2020;295(2):515-23.

Acknowledgements

Firstly, I would like to acknowledge and thank the Bond University Collaborative Research Network for Advancing Exercise and Sports Science (CRN-AESS). This project would not have been possible without large-scale collaboration, facilitated by the CRN-AESS between Bond University, Victoria University, and Queensland University of Technology. Secondly, the CRN-AESS provided funding for the individual research projects, and my PhD stipend until 2019. I would also like to acknowledge Victoria University and in particular Associate Professor Nir Eynon and Dr Sarah Voisin for the design and execution of the exercise study that formed the basis of these thesis projects. This project would also not have been possible without the time and significant amount of effort provided by the research participants for the Gene SMART study.

A massive thank you to my principal supervisor, Associate Professor Kevin Ashton, who supported my candidature throughout my project and especially provided much needed time and input to deal with my thesis writing antics in the last year of my PhD. You have showed me that thinking outside the box is a must in order to progress with research, and one's own mindset. A special thank you and acknowledgment to Associate Professor Larisa Haupt, whose unwavering support allowed me to explore multiple avenues of research and provided numerous networking opportunities. I like to think that our banter kept me sane during the more stressful situations that I found myself in. Thank you to Associate Professor Nir Eynon for your support and knowledge of the field which has been invaluable during my candidature. You have shown me the importance and value of networking and collaboration within the life sciences. And last but certainly not least, thank you to Professor Lyn Griffiths, who has supported both me and my work, especially during the stress of a depleted stipend, and provided me with a networking opportunity that resulted in a postdoctoral research position at King's College London. Lyn, you have been an extremely important mentor to me, and I have no idea how to repay you for your kindness and understanding. I have been extremely fortunate to have learnt and been supported by such experienced and passionate people throughout my candidature. Your collective love of scientific pursuits has inspired me to continue in my

research career with confidence and I look forward very much to maintaining the collaborations and research networks that we have all developed during the course of my candidature.

I have been fortunate enough during my PhD journey to develop long-lasting friendships with several of my colleagues. To Mr Paul Dunn and Mr Omar Ibrahim, it is extremely hard to cover everything you two have done for me during the time we have known each other. It is amazing how much you can learn from bouncing ideas off each other in the office. Please know that I wholeheartedly love and appreciate you both and look forward to seeing where we all are in 5-10 years' time. To Ms Melinda Mitchell and Ms Jasmine Connell, you have both provided me with so much support and insightful conversations throughout our journeys. I am really happy that you both have been successful considering the big interruption in changing from Griffith to QUT. Lastly, to Ms Jamie-Lee Thompson, you have been a wonderful friend and an appreciated constant friendly face down at Bond. It has been great to commiserate over the difficulties of publishing large-scale data. Thank you everyone; I am really proud of each of you and respect you all. I hope that we can all stay connected and continue to collaborate/bounce ideas off of each other in the future.

The support of the Genomics Research Centre (GRC) has been greatly appreciated during my candidature. I would like to thank the GRC for funding the majority of the consumables for my projects. I would also like to acknowledge and thank the researchers within the GRC who keep the group running. A special thanks to Dr Rachel Okolicsanyi for some really great discussions. I appreciate how you cut through the nonsense and tell it how it is.

To my family, Dale, Sonja, and Josie, I know it has been hard for me to articulate what I do, but I am extremely grateful for your unwavering support during all of my studying years. I think we are all ecstatic that I did not turn out to be a perpetual student. The timing for going overseas (even to pursue a really great job) is a bit of a risk but I know you will continue to support me through the stress of it all. I have seen both of my parents work extremely hard without sacrificing their principles and have no doubt that this is where I gained my work ethic. A special thank you to Josie, whose sense of humour kept me going during the more stressful time of my study. I love all of you very much and looking forward to celebrating this achievement with you all.

Last but certainly not least, to my enchanting partner Ms Lisa Osborne who has supported me for the last two years. A massive thanks for letting me rant about my work to you and to always keep bragging about me to your family and friends. Imposter syndrome is all too common, and you have helped me realise that what I have achieved over the last few years has been an incredibly large amount of work that I can be proud of. I have no doubt that you will continue to support both me and my career over in London. Love you honey.

At the risk of sounding like a pretentious researcher… I would like to end this acknowledgment section with a quote that I think summarises my experience over the last few years.

"Our doubts are traitors and make us lose the good we oft might win by fearing to attempt."

- William Shakespeare, *Measure for Measure*

Table of Contents

List of Tables

List of Figures

List of Abbreviations

1000G	1000 genomes database
3`UTR	Three-prime untranslated region
4WP	Four-weeks post exercise time point
ACE	Angiotensin converting enzyme
ACTN3	Alpha actinin 3
AGE	Agarose gel electrophoresis
AGT	Angiotensinogen
AMPD1	Adenosine monophosphate deaminase
APOE	Apolipoprotein E
ATP	Adenosine triphosphate
BMI	Body mass index
bp	Base pairs
BP	Biological process
BWA	Burrows wheeler alignment
Ca^{2+}	Calcium ion
cAMP	Cyclic adenosine monophosphate
CARDIA	Coronary Risk Development in Young Adults
CHR	Chromosome
CI/II	Mitochondrial complex 1 and 2
CKMM	Muscle specific creatine kinase
COL4A1	Collagen type IV alpha I chain
COL4A2	Collagen type IV alpha II chain

COXI-III	Cyclooxygenase I-III
CpG	Cytosine residue directly adjacent to a Guanine
CS	Citrate synthase
CVD	Cardiovascular disease
CYTB	Cytochrome B
dH_2O	Distilled water
DIABLO	Diablo IAP-binding mitochondrial protein
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
EDTA	Ethylenediaminetetraacetic acid
EGR1	Early growth response 1
EGR2	Early growth response 2
ER	Endoplasmic reticulum
ETS2	Ets proto-oncogene 2
FAM185A	Family with sequence similarity 185 member A
FC	Fold change
FDR	False discovery rate
FOS	Proto-Oncogene c-Fos
FOSB	FBJ murine osteosarcoma viral oncogene homolog B
GATK	Genome Analysis toolkit
GDF8	Growth differentiation factor 8
gDNA	Genomic DNA
Gene SMART	Genes and ‘Skeletal Muscle Adaptive Response to Training’
GO	Gene ontology

GOBP	Gene ontology biological process
GRC	Genomics Research Centre
GSEA	Gene set enrichment analysis
GWAS	Genome wide association study
GXT	Graded exercise test
H3K27	Histone H3 lysine 27
H3K37	Histone H3 lysine 37
H3K4	Histone H3 lysine 4
H3K9	Histone H3 lysine 9
HET	Heterozygote
HFE	High FE^{2+}
HIF1a	Hypoxia inducible factor 1 alpha
HIIE	High intensity interval exercise
HIIT	High intensity interval training
HOM	Homozygote
HREC	Human research ethics committee
HT	Hypertension
HTMGS	High throughput mitochondrial genome sequencing
HVI-III	Hyper variable region I-III
HWE	Hardy-Weinberg equilibrium
IHBI	Institute of Health and Biomedical Innovation
IL6	Interleukin 6
INDEL	Insertions/deletion
K+	Potassium ion
KDM4C	Lysine-specific demethylate 4C

LD	Linkage disequilibrium
lncRNA	Long non-coding RNA
LT	Lactate threshold
MAF	Minor allele frequency
MALDI-TOF	Matrix-assisted Laser Desorption Ionisation-Time of Flight
Mb	Mega base
Mitoseq	Mitochondrial sequencing
mRNA	Messenger RNA
MSTN	Myostatin
MT-CYB	Mitochondrially encoded cytochrome B
mtDNA	Mitochondrial DNA
MTG2	Mitochondrial ribosome associated gtpase2
MT-ND4	Mitochondrially encoded NADH dehydrogenase 4
MTND5	Mitochondrially encoded NADH dehydrogenase 5
MYC	MYC Proto-Oncogene, BHLH transcription factor
NADH	Nicotinamide adenine dinucleotide + hydrogen
NATA	National Association of Testing Authorities
NCBI	National Center for Biotechnology Information
NDUFAF7	NADH:Ubiquinone Oxidoreductase Complex Assembly Factor 7
NEMP	Nuclear encoded mitochondrial protein
NGS	Next generation sequencing
NR4A2	Nuclear receptor subfamily 4 group a member 2
NRF1	Nuclear respiratory factor 1
NRF2	Nuclear respiration factor 2

NTC	No template control
OR	Odds ratio
P0	Post exercise time point
P3	Three-hours post exercise time point
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
PGM	Personal genome machine
pM	Pico Molar
PPARGC1A	Peroxisome Proliferator-Activated Receptor Gamma Coactivator 1 Alpha
PPARα	Peroxisome proliferator activated receptor alpha
PRE	Preliminary time point
QC	Quality control
QFS	Quebec family study
QUT	Queensland University of Technology
rCRS	Revised Cambridge Reference Sequence
RFLP	Restriction fragment length polymorphism
RIN	RNA integrity number
RNA	Ribonucleic acid
RNAseq	RNA sequencing
ROS	Reactive oxygen species
RSRS	Reconstructed sapiens reference sequence
SAP	Shrimp alkaline phosphatase
SC	Satellite cell

SD	Standard deviation
SDC4	Syndecan-4
SE	Standard error
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variation
SVD	Single value decomposition
SWGDAM	Scientific Working Group on DNA Analysis Methods
T2DM	Type 2 diabetes mellitus
TAE	Tris acetate EDTA
TEM	Technical error of measurement
TGS	Total genotype score
TIGER	Training Interventions and Genetics of Exercise Response
tRNA	Transfer RNA
TT	Time trial
VO_{2peak}	Maximal oxygen respiration
VU	Victoria University
Wpeak	Peak power output
Δ	Delta change

Chapter 1: Introduction

This introductory chapter will outline the background and context for the purposes of the projects in Sections **1.1** and **1.2,** respectively. The thesis outline, including all projects, aims and hypotheses are outlined in Section **1.3**. Lastly, the significance, scope, and definitions section (**1.4**) will address the overall importance of this research to the field of exercise science.

1.1 BACKGROUND

Exercise remains the most efficient way to maintain health in normative states and increases the likelihood of recovery from a host of disease states. Even so the exact mechanisms by which these changes occur within skeletal muscle remain poorly understood. Whilst there have been many exercise studies that examine endurance exercise and consequently identified genes and processes involved in adaptation, very few were centred around high intensity interval training (HIIT). Of these few, even less included more than ten participants, analysed young/lean/healthy individuals, examined comprehensive Omic level data, or aimed to tie together multiple molecular datasets to robustly identify molecular processes involved in adaptation to HIIT. Therefore, there remains a need within the current exercise science framework to 1) identify whether other exercise related genes are differentially regulated following long term HIIT 2), identify novel genes and molecular pathways that are responsive to long term HIIT, and 3) replicate previously identified markers in a larger HIIT cohort. This provides the opportunity to assess trends in well-known genes for other data sets. For instance, a study that assesses many genetic factors and replicates some previously identified genes may be more confident in the secondary findings gained.

The Gene SMART study was designed to address the aforementioned knowledge gaps within the literature. The study contained a relatively large (n=77) cohort when compared with existing exercise training cohorts. Participants were included based on strict inclusion criteria to ensure lean and healthy individuals and limit the number of confounding variables to training. The study was based on a longitudinal design and included sampling time points prior to exercise, immediately post a single bout of high intensity interval exercise (HIIE), three hours following HIIE, and at rest, 48 hours after completion of the four-week HIIT program. In addition to the rigorous phenotypic response data collected during this study, the study performed 1) replication analysis based on previously associated exercise genetic markers, 2) large scale genotyping of genomic variants and high throughput sequencing of the

mitochondrial genome, 3) large-scale transcriptome analysis, and 4) large-scale epigenetic analysis. Whilst the analysis of any single one of these data sets offers a great deal of information on the possible biological basis of complex phenotypes, the integration of these technologies and analysis techniques offers a far more complete picture on the complete molecular architecture.

1.2 CONTEXT

Of the current literature in the field of exercise response, no studies have co-analysed genomic, epigenetic, and gene expression data. The discovery of novel targets is needed to identify accurate pathways, functional outcomes, and actionable mechanisms that influence complex phenotypic traits such as exercise adaptation. This thesis will examine in detail the mechanistic changes in response to strenuous exercise in healthy participants. The study also highlights the crucial need for large scale replication studies in the field of exercise genomics. Additionally, there is no requirement to stringently stratify participants based on arbitrary thresholds where multiple exercise phenotypes are considered. This problem has been addressed using strict significance thresholds and quantitative trait analyses. However, appropriate participant stratification may yield increased statistical power given significant differences in phenotypes between exercise response groups.

1.3 PURPOSES

MultiOmics refers to the combination of two or more large-scale data such as genetic sequencing, RNA sequencing, or epigenetic typing technologies. As previously mentioned, exercise studies are extremely targeted to specific genes of interest and very few have utilised more than one Omic approach. Further, MultiOmics analysis in fields outside of cancer are extremely rare. As such, the work within this thesis represents a large, novel, and exciting contribution to the field of exercise science as a whole and may be further explored with functional studies to discern novel exercise response targets.

1.3.1 Thesis Aim 1

Replicate previously discovered exercise associated genetic variants within the Gene SMART participants. Specifically:

a. Research previously discovered exercise SNP markers to develop a list of candidate SNPs to genotype within the Gene SMART participants.

b. Design a MassARRAY with the highest multiplex possible to ensure cost effective genotyping

c. Perform associations between genotypes and quantitative traits within the Gene SMART study to discover significant exercise genetic variants

Based on the specific sub-aims of Aim 1, several outcomes are expected. Firstly, the development of an extensive list of previously associated genetic markers for use in the project and for future projects. Secondly, a designed MassARRAY to be used for further studies that examine exercise based genetic markers in other populations. Lastly, replication of several exercise markers within the current study as proof that the previous literature was successful in identifying appropriate markers of exercise adaptation processes.

1.3.2 Thesis Aim 2

Identify mitochondrial genetic markers that govern response to exercise response phenotypes. The specific sub-aims are as below:

a. Adapt and optimise a mitochondrial genome sequencing protocol to sequence Gene SMART participants

b. Identify mtDNA regional haplogroups associated with exercise response

c. Identify mtDNA genetic variants associated with exercise response

d. Identify mitochondrial related variants associated with exercise response

As mitochondrial number and function correlate directly with energy homeostasis within skeletal muscle, changes in mitochondrial genetics were expected to be correlated with training outcomes. In addition, multiple nuclear encoded genes interact with mitochondria to influence biogenesis and continued function. As such nuclear encoded genes were expected to elicit similar responses to the mitochondrial genome and energy homeostasis phenotypes within the Gene SMART participants. Firstly, the assessment of haplogroup with exercise outcomes has previously been extremely limited to control region sequences. As such, in this study it was expected to identify specific and possibly novel mitochondrial haplogroups associated with response. Any findings from this sub-aim should assist future studies in the examination of mitochondrial population-based response outcomes. As a secondary development of this aim, we hope to optimise a cost-effective protocol for the tandem sequencing of mitochondrial genomes. This would improve practicality of the examination of mtDNA variants for numerous fields including energy homeostasis, population-based genetics, mitochondrial based diseases, and exercise science. Lastly, sub-aims **(2c)** and **(2d)** were expected to allow the identification of specific genetic markers pertaining to exercise adaptation processes involving mitochondrial function.

1.3.3 Thesis Aim 3

Identify longitudinal global gene expression changes and pathways in response to training. The sub-aims utilised to achieve this aim are detailed as follows:

a. QC RNA samples prior to experimentation to address any bias in gene expression values

b. Utilise targeted RNA sequencing using the AmpliSeq Transcriptome methodology unique to Ion Torrent platforms

c. Perform differential gene expression analysis using a paired design for increased statistical power

d. Perform Gene Set Enrichment Analysis to assess coordinated changes in gene expression corresponding to molecular pathways

e. Assess the link between differentially expressed genes and pathway changes to previous genomic chapters and aims

Several novel outcomes were expected to result from this thesis aim. Firstly, sub-aim **(3a)** will demonstrate the utility of RNA integrity assessment for downstream sequencing applications. Secondly, this aim will demonstrate the utility of Ion Torrent platforms for sequencing applications in Omics studies. Further, the targeted form of sequencing utilised, may be beneficial to future large-scale exercise studies due to the low cost per sequence when compared with Total RNAseq. Thirdly, to discover novel genes and molecular pathways associated with exercise time points. The potential of these results may be beneficial to the exercise science community at large as novel exercise inducible pathways may be functionally assessed in other cohorts. Lastly, assess the link between results from previous chapters with exercise inducible pathways. This would demonstrate the utility of MultiOmics for the discovery of molecular pathways within any complex system.

1.3.4 Thesis Aim 4

Ascertain the global methylation changes corresponding to immediate and long-term longitudinal exercise time points in a specific subset of Gene SMART participants.

a. Discover differentially methylated probes and loci at each exercise time point

b. Associate the differentially methylated probes to promoters and genes

c. Perform pathway analysis to discover the molecular pathways epigenetically regulated at each exercise time point

d. Discern common significant molecular pathways changes between epigenomic and transcriptomic data sets

For the last aim of this thesis, several "real world" outcomes were proposed. Previous studies that examined methylation changes in response to exercise have been limited by sample size. Further, the association of differentially methylated probes of genes and molecular pathways has been lacking. To address this within the longitudinal Gene SMART cohort, novel analysis methodologies were utilised. Further, the assessment of coordinated changes in methylation to link probes to specific genes and promoter regions represents a novel analysis methodology. The utilisation of gene set enrichment analysis for epigenomic data has not been well characterised. As such, the findings from this analysis will represent an extremely important and significant outcome to the field of exercise science. Further, the assessment of MultiOmics between transcriptome and epigenome data sets is extremely rare within the established literature. The findings from this section of Aim 4 represent the culmination of the contribution of this thesis and the Gene SMART study to the field of exercise science. Further, the findings from this aim represent a significant amount of novel work and as such the outcomes from this project will significantly enhance the standard and breadth of future studies pertaining to exercise science.

1.4 THESIS OUTLINE

This section outlines the chapters within this thesis framework that correspond to the aforementioned aims and hypotheses. **Chapter 2** will outline a critical review of the current literature in exercise physiology, genetics, transcriptomics, and epigenetics. **Chapter 3** will outline the Gene SMART study as it relates to this thesis. The chapter will discuss inclusion criteria, participant recruitment, sampling methodologies, phenotypic tests, and the analysis of phenotypic data and the building of composite traits. **Chapter 4** will discuss thesis aim 1 and details the designed MassARRAY and replication of pre-known exercise SNPs within the Gene SMART study. Further, an elite exercise population was included to identify any differences in performance-based association tests. Thesis Aim 2 is addressed in **Chapter 5**, where mitochondrial and mitochondrial related genetics in relation to exercise response phenotypes were assessed. **Chapter 6** corresponds to thesis Aim 3 and outlines the use of RNA sequencing to identify genes and molecular pathways differentially regulated in longitudinal exercise time points. **Chapter 7** examined the epigenetic marks that correspond to longitudinal exercise time points following immediate training and longer-term HIIT. This chapter also outlines the aim to assess similarities between transcriptomics and epigenetics for greater scrutiny in analyses. Finally, **Chapter 8** provides an overview of all the studies undertaken in this research, including discussion of the limitations and potential future directions of this research.

Chapter 2: Literature Review

2.1 EXERCISE: DISEASE PREVENTION AND HEALTHCARE COSTS

The benefits of regular exercise or physical activity on health and performance are well-documented. These broad benefits include improved fitness, well-being, and mental health. Exercise is also known to reduce the risk of developing a large number of chronic diseases including, type II diabetes, metabolic syndrome, cardiovascular disease, stroke and depression.(1, 2). Indeed it has been estimated that physical inactivity leads to increased risk of approximately 40 disease states (3). An increased cardiorespiratory fitness level (commonly measured by maximal oxygen uptake [VO_{2peak}]) has been associated with a large (~50%) decreased mortality risk (4). This is perhaps unsurprising as exercise induces changes in multiple biological systems including skeletal muscle, cardiac muscle, lung capacity, blood vessel capacity, and brain function (5, 6).

Aerobic exercise has also been shown to improve cognitive and physical function in older individuals (4, 7). Further, training has been utilised as an adjuvant therapy alongside chemotherapy and transplantation techniques (8, 9). With only a small portion (9.6%) of the adult population reaching the necessary physical activity guidelines (150-300 minutes of moderate or 75-150 minutes of vigorous aerobic physical activity per week) (10) it has been estimated by the World Health Organisation, that approximately 3.3 million individuals die as a result of sedentary behaviour. Further, the direct and indirect healthcare costs associated with lack of physical activity are large but vary significantly in different populations. For instance, the estimated costs in the United States of America are approximately $30 billion USD (11) whereas Canada reported the sum of indirect and direct medical costs to be approximately $5.3 billion CAD (12). In Australia, it has been estimated that a 10% reduction in physically inactive adults would deliver a $96 million AUD reduction in healthcare costs, which represents approximately 7.4% of current Gross domestic product (13). Further, this would equate to approximately 6,000 less cases of inactivity induced disease and 2,000 less deaths.

These data suggest that widespread uptake of exercise training will increase the health of both young and older populations, increase response to various therapy, and drastically lower the economic impact of sedentary behaviours on national healthcare systems.

2.2 EXERCISE TRAINING AND SKELETAL MUSCLE

Skeletal muscle is a highly specialised tissue (multinucleated, shared sarcolemma) that is commonly described as plastic due to its ability to adapt to extremely stressful physiological states, including large-scale inflammation, muscle damage, and oxidative stress (14). Studies relating to exercise adaptations have characterised the abundance and type of skeletal muscle fibres as a physiological response to repeated bouts of training, however the molecular processes behind these changes are not well understood. Indeed, the molecular changes within each skeletal muscle cell are largely unknown. Broadly, skeletal muscle fibres are classified based on their functional and metabolic capabilities. For example, slow twitch muscle fibres (type I) are aerobic and associated with responsiveness to long term (>12 weeks) endurance training. In contrast, type II fibres are classified as glycolytic fast twitch and are associated with more power-oriented exercise traits such as sprinting or weightlifting (15, 16). The combination of each of these fibres contribute to the overarching phenotype resulting from repeated bouts of training. Of note, a key difference between human and rodent skeletal muscle glycolytic fibres is the absence of *MYH4* mRNA. In humans, this is not translated into functional protein even though this type of myosin may still be measured through transcriptional studies (17).

The skeletal muscle microenvironment is complex and may involve several different cell types. Specifically, skeletal muscle may contain muscle specific cells such as myocytes and satellite cells (muscle specific stem cells), cell types commonly associated with connective tissues (endothelial cells), mesenchymal progenitors (fibro-adipogenic progenitors (FAP) cells), and resident immune cells (18). In addition, exercise training may induce immune responses, immune cell extravasation into the skeletal muscle niche, and therefore a greater mixture of cell types leading to adaptation. It has been estimated from murine models that ~25% of nuclei within the skeletal muscle microenvironment are not myonuclei (19). In addition, myonuclei have been shown to exhibit diverse transcriptional signatures, which may further exacerbate the inter and intra-individual transcriptional variability observed within exercise studies (19). This has yet to be demonstrated within human studies and should be considered contextually within human skeletal muscle biopsies.

2.2.1 Strength Training

Generally, muscle responds to extremely intense exercise in two stages: 1) muscle overload leading to damage, and 2) muscle repair, leading to either enlargement of the muscle (hypertrophy) or molecular restructuring for improved oxygen usage. For regular muscle function, normal cellular pathways need to be active such as steady state actin polymerisation and myosin/myostatin recruitment and expression (20, 21). Energy producing pathways also need to be active, these pathways will differ between muscle types as previously discussed, and this will correlate to changes in vasculature and the renin/angiotensin pathway (22). Once a muscle is overloaded, micro-damage along the fibres develop, this activates inflammatory pathways and results in the recruitment of satellite cells to the site of the break. These muscle specific stem cells fuse to the muscle and cause increased myonucleation and expansion. These processes are repeated with every bout of exercise and therefore differential regulation in any of these pathways will extensively change the response to exercise and may explain some of the variation that can be seen in exercise study participants. Further, detraining, or sedentary behaviours often lead to a loss of muscle mass, known as atrophy (23). It has been well reported that atrophic muscle retains the myonucleation gained from previous training and therefore re-training will induce faster recovery to the original trained state (24). This process has been outlined in **Figure 2-1**. The increased myonucleation following detraining further increases the disparity between exercise participants as there is no non-invasive way to gauge previous training status. Interestingly, this specialised tissues propensity for multinucleation may have drastic effects on the biological implications of previous and continued scientific findings.

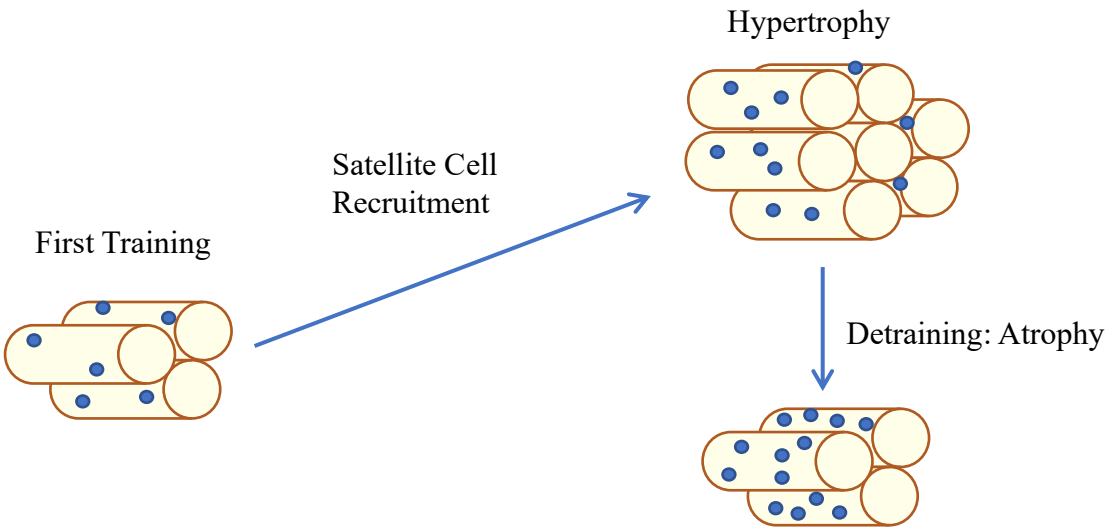


Figure 2-1: *Muscle fibre response to strength-based training and detraining. Muscle Fibres are shown in cream, myonuclei are shown in dark blue.*

2.2.2 Aerobic Training

Aerobic exercise is associated with larger quantities of type I skeletal muscle fibres. A hallmark of endurance training is the ability to trigger pro-angiogenic pathways leading to increased muscle oxidative capacity. Endurance training is the primary form of training within the existing literature and may be further stratified into moderate intensity continuous training (MICT), as seen in long distance running or high intensity interval training (HIIT), for example several bouts of sprinting. HIIT consists of repeated sessions of high intensity interval exercise (HIIE) that are approaching maximal oxygen consumption (typically 70-90% VO_{2peak}). In addition, HIIE sessions are much shorter (typically 20 minutes) than MICT (typically 45 minutes) proving the time efficiency of HIIT compared to other aerobic training modalities (25-29). HIIT contains several different modalities, which lead to differing physiological and biological responses. High volume HIIT, classified by longer interval-rest periods (usually five minutes), leads to larger molecular responses than low volume HIIT (intervals of approximately 1 minute). Of note, low volume HIIT has been shown to elicit extremely similar responses to MICT (30). In addition, sprint interval training (SIT), which represents a supramaximal effort, has been shown to elicit molecular changes more akin to that of strength exercise and will not be addressed further (31). As such, high volume HIIT has the greatest potential to enhance VO_{2peak}, while delivering greater molecular responses than that of low volume HIIT, MICT, or SIT.

Due to the high intensity and repeated bouts necessary for HIIT, the training is often associated with a decreased time period for the same physiological results such as VO_{2peak} and mitochondrial respiration (32). Further, it has been described to elicit larger aerobic fitness improvements than MICT (33). Perhaps the largest difference in these training methodologies is the potential to elicit an increase in mitochondrial density and function (34). HIIT induces mitochondrial biogenesis in skeletal muscle and therefore increases oxidative capacity (35). HIIT has been shown to activate the AMP-activated protein kinase (AMPK) cascade, resulting in the mitochondrial biogenesis seen in endurance training (36). Relatively short sessions (30 seconds Vs minutes) of HIIT have been shown to increase the enzymatic activity of mitochondrial enzymes such as citrate synthase (37). This represents a strong mechanism by which skeletal muscle adapts to HIIT by increasing machinery for continued oxidative respiration and therefore larger aerobic capacity. The largest changes from aerobic training occur at the molecular level and so will be discussed in the following section.

2.3 MOLECULAR RESPONSE TO TRAINING

The vast majority of literature examining molecular response to training has focused on a clear distinction between obvious endurance versus strength phenotypes. For example, Egan and Zierath et al, have thoroughly examined the molecular response to endurance and strength exercise. The intersection between the molecular response to these extreme phenotypes and HIIT was not explored. Given the complexity of molecular signals resulting from these differing types of training, this section will aim to summarise the key findings for each training type and delineate these molecular responses in a HIIT context. As strength training does not relate to this thesis, the section will provide a brief overview of the molecular responses.

2.3.1 Strength Training

As previously discussed, strength training may be characterised by skeletal muscle overload or training occurring at or near maximal oxygen respiration (VO_2peak). As such, the largest trigger for molecular induction from strength training is skeletal muscle contractile processes. Contractile function may cause mechanical stress leading to the activation of p38, ERK1/2 and JNK, which in turn modulate skeletal muscle gene expression and adaptation to training (38). Differences in calcium ion flux has also been shown to influence adaptation to

strength-based training through CaMKII and lead to phosphorylation of metabolic (CREB) and epigenetic modifying proteins (HDACs) to modify gene expression (39). Mechanosensation from these contractile processes are sensed by the FAK and PA proteins. PA stimulates a signalling cascade through PI3K, resulting in AKT signalling, mTOR stimulation and FOX repression. This results in a molecular switch from protein degradation to protein synthesis to support muscle growth.

2.3.2 Aerobic Training

Adaptations resulting from endurance training are largely triggered by molecular processes rather than physiological modulation. As such, the molecular processes triggered by endurance training are more numerous than strength-based training and are more targeted to the regulation of gene expression. Aerobic training may induce hypoxia and therefore HIF1a expression leading to increased gene expression of HIF inducible pathways. In addition, skeletal muscle metabolic processes during training lead to altered redox state leading to enrichment of NAD^+, ATP reduction leading to enrichment in AMP, and induction of reactive oxidative species (ROS). The enrichment in NAD^+ leads to induction of SIRT epigenetic modifying proteins and deacetylation of key adaptive genes such as *PGC1α, FOXO1*, and *p53*. The increased AMP triggers the AMPK pathway leading to phosphorylation of epigenetic modifying proteins (HDACs, SIRTs), activation of key exercise inducible genes (*PGC1α, HIF*), and metabolic genes (*CREB*).

2.3.3 HIIT

HIIT is considered a subset of aerobic training both in terms of physiological results (increases in oxidative capacity of the individual) and due to similarities in many molecular processes. Even so, training programs for high volume HIIT are similar to those for strength training in that both use a threshold of >80% VO_{2peak}. In addition, HIIT exhibits the hallmarks of aerobic training (i.e. mitochondrial biogenesis and angiogenesis triggered largely through transient PGC1α and HIF expression) as well as molecular signalling indicative of strength training (calcium ion flux leading to phosphorylation of CREB, HDACs and SRF; myofibrillar contraction leading to mechanosensation and protein synthesis) (38). As such, whilst considered under the endurance training umbrella, HIIT represents a hybrid training regime and may contain the benefits of both types of training. The majority of the current literature focuses on candidate proteins involved in mitochondrial function, myokine secretion, or angiogenic processes (40). Further, acute bouts of HIIE (41-44) rather than HIIT (45-47) are examined more frequently within the established literature. In a recent review article by *Torma et al*, the molecular adaptations to HIIT were discussed in great detail (48). Interestingly, the main summary of the current literature pertaining to molecular pathways in response to HIIT were limited to the stimuli and pathways discussed previously with a few key adaptions as outlined below.

Oxidative stress

High intensity training may induce oxidative stress in skeletal muscle microenvironments and in extreme cases may lead to local hypoxia. This oxidative stress has been shown to increase reactive oxygen species (ROS) and increased levels of oxidative stress sensors such as NOX, NO, and XO (49-51). HIIT has specifically been shown to exhibit excess amounts of ROS when compared to MICT, however these levels were not at the excess amounts associated with metabolic damage. ROS has been shown to elevate hydrogen peroxide within skeletal muscle, which in turn increases the promoter activity of the *PPARGC1A* gene. PGC1α is a transcription factor for the antioxidant genes *SOD2* and *GPX (52)*. This represents a mechanism by which skeletal muscle adaption processes may be triggered by oxidative stress. This mechanism is not as effective in MICT as a steady state of respiration is maintained and the oxidative stress is not as pronounced. The majority of studies pertaining to ROS in skeletal

muscle were based in animal models and therefore the findings are suggestive within the context of HIIT in human studies.

Oxidative capacity

In conjunction to oxidative stress, HIIT induces mitochondrial biogenesis in type II skeletal muscle fibres. MICT has been shown to specifically influence this process in type I fibres, corresponding to the aerobic and steady state phenotype generated (53). Mitochondrial biogenesis is directly linked with the common fitness marker VO_{2peak}. As previously stated, high volume HIIT may induce HIF1α and therefore trigger pro-angiogenic pathways that further distinguish this training from MICT. Of note, capillary density has been shown to increase following MICT but not HIIT (54). Potential contributors to this phenotype include hypoxia (leading to release of reactive oxidative species), increased blood saturation in skeletal muscle leading to greater access for pro-angiogenic molecules, and lactic acidosis (55). Of these, hypoxia has been robustly associated with angiogenesis following training as the levels of HIF1α and the pro-angiogenic factor VEGF are strongly correlated (56). Levels of HIF1α have been shown to increase in skeletal muscle three hours following a single session of HIIE (57, 58). More recently, VEGF has been shown in animal models to be stimulated by PGC1α in an ERRα dependant manner (59). Even so, the exact molecular pathways that are involved in these angiogenic processes remain unknown. Low volume HIIT however has been shown to be more involved with the elevation of mitochondrial complex enzymes (COX IV) regardless of fibre type (60). These differences highlight the inter-training variability that may occur between exercise studies. The large energy demand triggered by high volume HIIT leads to a significant drop in the cellular ATP/AMP ratio. This in turn triggers the AMPK pathway, leading to stimulation of the PGC1α mitochondrial biogenesis mediator (61). This represents another mechanism unique to HIIT when compared with MICT, as mitochondrial biogenesis is triggered within multiple fibre types and will therefore have a more profound effect on skeletal muscle adaptive aerobic response.

Calcium flux

Similarly to strength training, high volume HIIT causes muscle contraction-initiated calcium ion release from the sarcoplasmic reticulum through the fragmentation of RyR calcium channels (62). This fragmentation has been shown to be heavily involved in SIT with a total

fragmentation rate of 85%. The resulting calcium ion release triggers the calcium mediated proteins CAMKII and CAMKKβ. These proteins specifically target and activate the α2 isoform of AMPK, which leads to the activation of sarcoplasmic PGC1α and again triggers mitochondrial biogenesis through expression of NRF1/2, and an oxidative phenotype within type II fibres (63). This in turn creates a positive feedback loop with the AMPK pathway whereby available energy pool of NAD^+ increases in response to mitochondrial biogenesis, causing activation of the epigenetic protein Sirt1, and continued activation of AMPK and PGC1α (36, 64-67). Of note, HIIT has been shown to cause optimal biological responses at varying levels of intensity for different individuals, leading to the inherent inter-individual and inter-study variability commonly observed in HIIT studies.

Metabolism

Further to the oxidative adaptations outlined above, HIIT elicits responses to carbohydrate and lipid metabolic processes (68, 69). Specifically, pyruvate dehydrogenase E1α content as a measure of carbohydrate metabolism and 3-hydroxyacyl CoA dehydrogenase activity for lipid metabolism. A decrease in carbohydrate metabolism has been demonstrated following both HIIT and MICT, whereas lipid metabolism was increased (69). Interestingly, it has been shown that MICT is able to influence the levels of blood lipids following training more so than HIIT (70). In summary, skeletal muscle adapts to endurance training by downregulating carbohydrate metabolic process but upregulating lipid metabolism.

Inflammation

HIIT has also been shown to induce skeletal muscle inflammation. Interestingly, skeletal muscle is able to secrete cytokines (myokines) following training, and therefore may regulate inflammation in an autocrine manner, immune cell extravasation, and may affect metabolic pathways in other tissues (71, 72). Common myokines include myostatin (MSTN), several interleukins (IL6, IL7, IL8, IL15), fibroblast growth factor 21 (FGF21), Bone derived neurotropic factor (BDNF), and Insulin-like growth factor 1 (IGF1) (72). When assessing the inflammatory markers (IL-1β, IL-6, IL-10, MCP-1, IGF-1, CRP), it was found HIIT (6x30 seconds supramaximal cycling) elicited the same responses as a single bout of endurance training (45 minutes cycling at 62.5% maximal heart rate), further supporting HIIT as a timesaving viable approach to endurance training (42).

In summary skeletal muscle adapts to the stress induced by HIIE thorough a number of mechanisms; specifically, calcium influx and oxidative stress leading to an increase in mitochondrial number and function through AMPK and PGC1α pathways; a shift from carbohydrate metabolism to lipid metabolism; induction of HIF1α to stimulate glycolysis and lactate transport; secretion of myokines (such as IL-6) to stimulate inflammation; and expression of VEGF leading to angiogenesis. These processes are summarised in **Figure 2-2**.

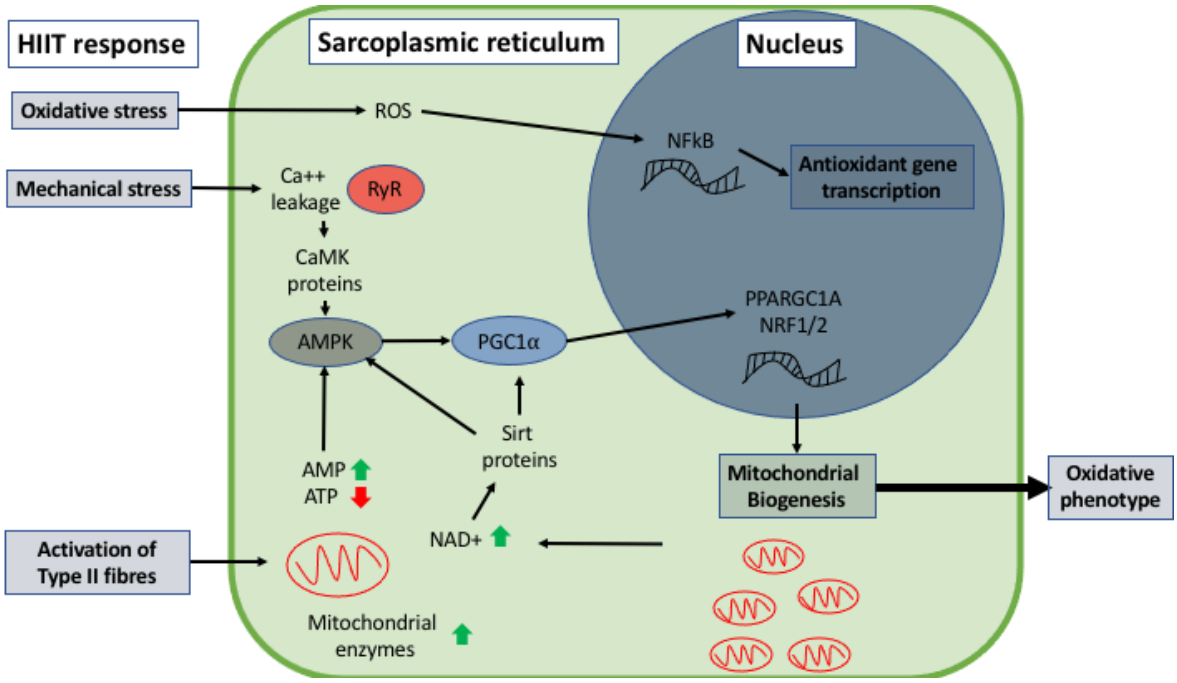


Figure 2-2: *The molecular adaptations to HIIT as summarised from the existing literature. Adapted from Torma et al (48).*

The above results, outline several key findings from the literature and further highlight the quantity and quality of work on exercise adaptability. Even so there are several knowledge gaps within the current literature. Firstly, the vast majority of literature pertaining to HIIT contained participant numbers of n<10 and/or were based in animal models and have not been confirmed in human studies. As such the findings from these studies (whilst similar functionally) should be replicated in larger human cohorts. In addition, inter-individual variability for participants performing the same training has been well documented (73, 74). This may be due to several variables (typically adjusted for within exercise studies) such as age, sex, ethnicity, and genetics. Indeed, studies examining the effects of genetic factors on

training outcomes estimated that aerobic exercise capacity was ~50% heritable (75, 76). Secondly, the studies outlined within this section were specifically focussed to each aspect of these molecular pathways and therefore represent a narrow snapshot of the molecular mechanisms in response to HIIT. These past findings were extremely important for the understanding of potential mechanisms for skeletal muscle adaptation following HIIE and HIIT, however the larger genetic architecture and molecular pathway changes at HIIT longitudinal time points has yet to be explored in any detail.

2.4 CANDIDATE GENE STUDIES

To date, hundreds of gene polymorphisms have been implicated in exercise response and adaptability, with much focus on the multifactorial nature of exercise genomics (77). Studies focusing on a genotype-phenotype relationship have primarily focused on gene polymorphisms with subsequent functional studies. Many of the studies have lacked statistical power to identify functional outcomes due to the limited number of people with elite athlete status. The most prominent of these functional variants include *alpha-actinin-3* (*ACTN3)* p.R577X (rs1815739), *angiotensin-converting enzyme (ACE)* Ins/Del (rs4340) and *myostatin (MSTN)* (rs1805086). As each of these SNPs partially predicts individual response, an additive model of exercise genomics would likely be predictive of response. Previous studies have tried to model this with little success, indicating that:

1. Exercise response is poorly explained by an additive model.
2. There are many unknown SNPs responsible for exercise adaption that have not yet been discovered.

Concurrent exercise studies have focused on strength versus endurance outcomes associated with different variants. Functional variants associated with endurance phenotype have been shown to trigger angiogenesis and therefore VO_{2peak}. Increased sarcoplasmic reticulum structure and function, as well as stimulated growth of type I, aerobic, slow twitch muscle fibres have also been implicated in response to endurance type exercise (78). Variants shown to coincide with a strength phenotype have been associated with increasing blood pressure, rate of muscle damage and satellite cell recruitment leading to increased type II, anaerobic, fast twitch muscle fibres (79). A total genotype score (TGS) is the theory that a series of mutations

that are individually beneficial, would collectively show an additive effect to a complex trait. Many studies have implicated that a total genotype score if present in a population of elite athletes would correlate highly with athletic success (80). While this is a theoretically sound principle, it may not be practically applicable to general populations. For example, *Williams and Folland* assessed 23 genetic variants associated with exercise response including *ACE* I/D and *ACTN3* R577X and suggested that the chance of an individual positively predisposed in all 23 of their studied gene variants was <0.0005%. Therefore, they concluded that examining a TGS would not be of relevance to the overwhelming majority of the population (81). Subsequent studies examining this principle have shown non-significant correlations between TGSs and performance in athletic events (82). However, if an athlete were predisposed to a certain phenotypic trait, they could display the opposite, due to a more power-oriented training regimen and *vice versa*.

Despite multiple genetic studies examining exercise adaptability only two significant genetic variants (*ACTN3* R577X, *ACE* I/D) have been robustly replicated in exercise cohorts.

2.4.1 Alpha Actinin 3 (ACTN3) R577X polymorphism

The *ACTN3* gene is commonly referred to as the speed gene in exercise genetics as the variant was first discovered by examining elite athletes in sprinting events (83, 84). ACTN3 is expressed primarily in type II muscle fibres, and functions to anchor actin to intracellular structures. The activation of the mTOR pathway has been shown to be greater in individuals with a functional ACTN3 protein (85). A common genetic variant within the gene results in the replacement of arginine (R) with a premature stop codon (X) at amino acid 577; summarised as p.R577X. The RR genotype of this variant has been shown to influence the response to power-oriented training and therefore greater propensity for fast twitch muscle fibres and a power-oriented phenotype following repeated bouts of exercise training (86, 87). The XX genotype has been shown to increase the chance of a greater propensity for endurance training (88).

2.4.2 Angiotensin Converting Enzyme (ACE) Insertion/Deletion

ACE is a key component of the Renin-Angiotensin pathway of aerobic respiration (89). The enzyme catalyses the conversion of angiotensin I to angiotensin II, thereby increasing the ability of the substrate as a vasoconstrictor. A common INDEL (insertion/deletion polymorphism) within the *ACE* gene has been implicated to have a functional impact on numerous exercise states, specifically in cases where endurance training has been the primary form of training (90, 91). The deletion form of this gene is missing 287 base pairs, resulting in a knockout phenotype for this gene, and therefore the angiotensin pathway at large (92). There has been mixed evidence for this gene in response to training as several studies have found negative results when assessing comparatively large cohorts (n=698) of elite endurance athletes (93, 94).

2.4.3 Mitochondrial Genetics

Mitochondria are intracellular organelles responsible for energy generation and homeostasis, nitric oxide removal and apoptosis and as such are pertinent to the correct function of cellular processes. Mitochondria contain multiple copies of the 16,569 bp circularised DNA aptly named the mitochondrial genome (**Figure 2-3**), which remains the only form of genetic material that is solely maternally inherited. This genome is extremely small in comparison to the nuclear genome, however, there remains little to no obsolescence of expression from the mitochondrial genome. The 37 genes encoded on the mitochondrial genome remain important not just for continuing function of the organelle, but the coding of multiple tRNAs for the correct translation of nuclear mRNA into protein. Despite the importance of mitochondria to the continuing function of skeletal muscle, the current literature in relation to exercise is limited. Many reports to date have focussed on the regulation of mitochondrial genes in response to exercise with only three studies outlining possible exercise intolerance due to mitochondrial DNA mutations (95-97). Apart from mutations in the mtDNA that cause myopathies and exercise intolerance, there is no evidence from the mitochondrial genome to ascertain whether novel SNPs or SNVs are associated with exercise response.

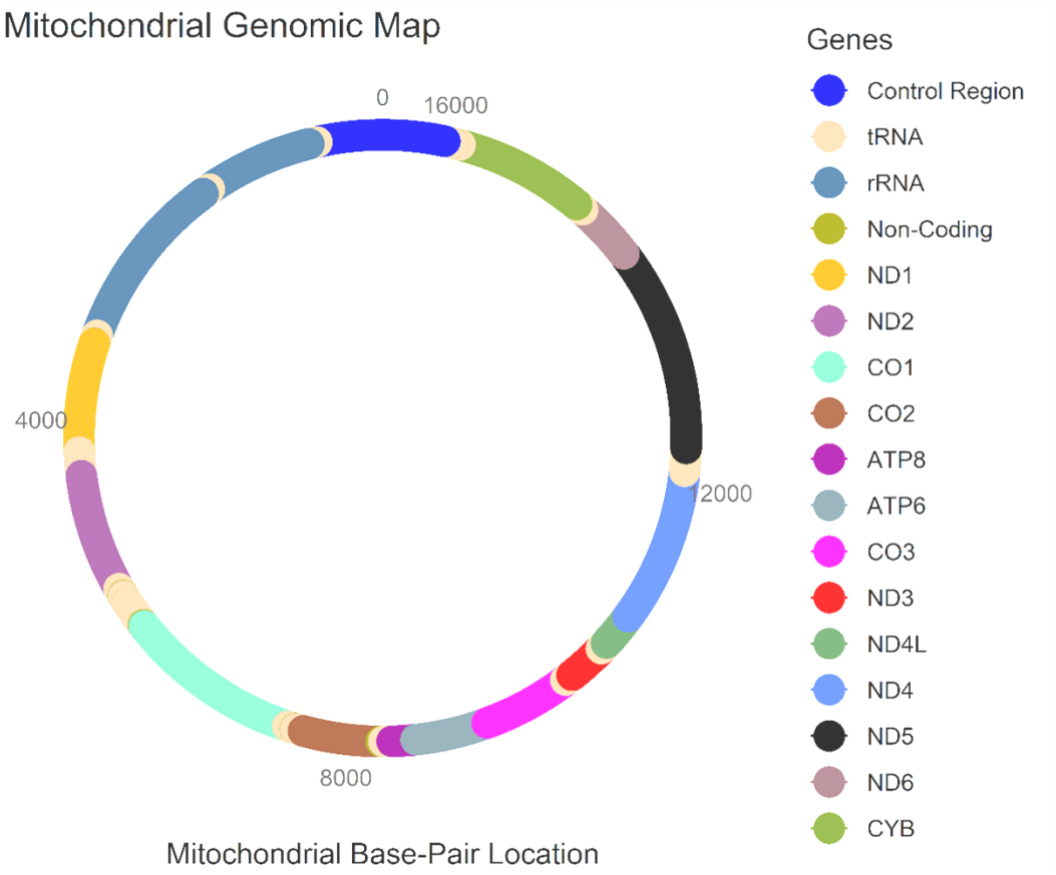


Figure 2-3: *A depiction of the human mitochondrial genome. Mitochondrial gene symbols and regions are shown in the legend (right). The mitochondrial genomic base pair location is arranged counterclockwise. The mitochondrial gene regions are colour-labelled according to the specific gene at that genomic location. The mitochondrial control region is depicted as the blue section of the genome at the top of the figure.*

2.5 LARGE-SCALE EXERCISE GENETICS STUDIES TO DATE

The first studies including phenotypic data such as BMI and respiratory fitness associated with better prognosis for cardiovascular disease (CVD) or hypertension include the Quebec Family (1978), CARDIA fitness (1985), and HERITAGE family studies (1992). Two consortium type exercise studies have made use of a GWAS approach to ascertain the impact of SNPs on elite endurance status (GENATHLETE (1993), and TIGER (2003)). Whilst the study design for the GENATHLETE project incorporated elite athletes from five separate nations, it produced confounding results as to whether commonly known polymorphisms such as the *ACE I/D* (rs4340) or *CKMM* (rs8111989) variants, are associated with training response. The TIGER study focused on the response traits (body fat, aerobic fitness) rather than genetic factors that contribute to exercise response. An outline of the studies is described below:

2.5.1 Quebec Family Study (QFS):

The QFS commenced in 1978 as a way to track obese families of French descent and followed 767 subjects from 207 families (98). Whilst the chief aim of the study was to assess body fat traits, follow up studies with this population have aimed to determine genetic factors that influence physical activity level. The study found that physical activity was segregated in families, indicating a genetic reason for physical aptitude (99). The study then performed a "genome wide linkage scan" of 432 genetic loci and was able to determine suggestive linkages between chromosomal regions and differing levels of physical activity (100). Lastly, the study assessed the impact of neurotransmitter gene variants on physical activity. The study found that the *MC4R* C2745T TT genotype was associated with physical inactivity (101). Whilst the QFS was useful for determining the genetics governing the risk to CVD, there were several confounding variables that limited the study when assessing physical activity levels. Firstly, the study was originally designed to address the risk of CVD in French families and as such any results gained from the study were only applicable in a diseased population (or individuals with close family ties to diseased individuals). Secondly, the follow up studies were focussed on candidate genes. These results and limitations indicate that the QFS was not applicable when assessing genes responsible for exercise response. Further, the study did not consist of a healthy population and any results gained were not applicable to the wider population.

2.5.2 CARDIA Fitness Study:

Similar to the QFS, the CARDIA fitness study also assessed CVD risk and included 2,663 participants of both Caucasian and African American ethnicities. The study assessed SNPs within seven candidate exercise genes: *ACE, AGT, BDKRB2, EDN1, GNB3, NOS3, PPARGC1A* (102). Subsequent studies using this cohort (20 years following) have implicated a number of variants from 17 candidate genes that showed association with BMI, fitness, and hypertension. The study found that SNPs within the *PPARGC1A*, *EDN1*, *HIF1A*, and *CKM* genes, were nominally associated with baseline exercise testing in Caucasians. In addition, a follow up study used 3,960 participants from the original CARDIA study with treadmill data at baseline and 20 years later. The participants performed a graded exercise test and the delta change in performance over 20 years was used to associate genetic variants with fitness. The study was able to implicate SNPs within the *AGT*, *AMPD1*, *PPARGC1A*, and *ANG* genes with decreased exercise ability in the following 20 years (103). Unfortunately, the study focused on candidate genes and did not expand to genome wide levels of association. Therefore, the results from this study were extremely limited and prevented the discovery of novel genes associated with exercise ability. In addition, the primary focus of the CARDIA study was to assess risk profiles of CVD. As such, the findings from this study, whilst promising for candidate gene-based exercise genetics, were not applicable to healthy populations.

2.5.3 HERITAGE Family Study:

The initial HERITAGE study began in 1992 and originally aimed to address the familial basis of risk to CVD traits. The study design included both Caucasian (n=469) and African American (n=224) individuals from 17-65 years of age. All participants were sedentary for 3 months prior to the initiation of the study. The participants were required to complete 20 weeks of 3/week sessions (50 minutes) of aerobic training on a cycle ergometer (104-110). The intensity was adjusted every 2 weeks to ensure adequate training progression. Phenotypic data was collected before and following the 20-week study and included VO_{2peak}, BMI, lipids plasma levels, glucose levels, insulin, steroid hormone levels, and body fat distribution. There have been many follow up studies focussing on the genomic implications of the research rather than the phenotypic data acquired with CVD risk factors. This study was the first consortium type exercise study to describe the wide range of variation seen in response to exercise, in a prognostic approach. Recent papers from this study have begun to link the GWAS data to the genetic components underlying body shape (111). There have been many follow up studies that

assessed genetic associations relating to the various phenotypic data collected prior and following the study. Namely, GWAS within this population found that the *UCP3* gene was associated with change in body composition (112) and segregation analysis implicated that genes involving heart rate and blood pressure contributed to 30% of the variance within VO_{2peak} responses (113). In addition, the study implicated well-known exercise variants (*CKM, ACTN3, PPARGC1A*) used to classify exercise study participants, by increasing the power and changing the nature of the study to include RNA expression analysis from skeletal muscle tissue (104-110). Lastly, a GWAS conducted on the HERITAGE participants to determine the heritability of VO_{2peak} response found 39 SNPs associated with responses (76). Further, 21 of these were used to build a total genotype score (TGS), which accounted for 49% of the variance in VO_{2peak} training response. The study then assessed the genes closest to the SNPs of interest and found that genes such as *PRDM1* (transcription factor for adaptive immune response), *ACSL1* (lipid metabolism), *CXCR5* (cytokine receptor triggering B-cell migration), and *CD44* (cell surface receptor involved in immune cell activation) were proximal to the associated variants within the TGS. Methods such as these must be carefully scrutinised as genetic variants identified from GWAS have been shown to be associated with genes a long way (>100Mb) from the identified loci.

The studies utilising the HERITAGE study contained several limitations that prevented the analysis of genetic variants contributing to exercise response within healthy individuals. Similar to the previous, the study was originally designed for the analysis of genetics contributing to the risk of CVD and the associated phenotypic traits and as such the results obtained were not applicable to a wider healthy population. In addition, the majority of the participants within the study were related, meaning the heritability of the assessed traits was able to be analysed. Whilst this was useful for determining the contribution of genetic factors on the phenotypic response traits, this was not assessed within a healthy population. The GWAS performed were robust at the time of conception, however the number of SNPs assessed (~320,000) and the number of participants utilised (n=473) for the study were not enough to discover genome wide significant findings ($P<1x10^{-8}$) (114). As previously mentioned, total genotype scores are usually only applicable to the study in which they are generated, and the HERITAGE study was no exception. The study found that 49% of the variance within the VO_{2peak} response trait was accounted for by the generated TGS consisting of 21 SNPs. Many replication studies (including our own (115)) have attempted to replicate these findings in

healthy populations with little success (75), indicating that the TGS generated within the HERITAGE study was likely overfitted and not applicable to the wider exercise field.

2.5.4 GENATHLETE Study:

The aim of the GENATHLETE study was designed as a case-control study between elite endurance athletes (n=316) and sedentary controls (n=299). This study had a candidate gene focus and aimed to replicate previously associated exercise genetic markers such as *ACE I/D* (116), *ACTN3 R577X* (117), *CKM* (*rs3791783, rs11681628, rs7570532*), *MSTN* (*rs344816, rs10410448, rs432979, rs1133190, rs7260359, rs7260463, rs4884*) (118), and *NOS rs1799983* (119). Unfortunately, none of these variants were found to be associated with the elite endurance population when compared with sedentary controls. It should be noted that whilst some of the variants were not significantly associated with endurance athleticism in this cohort, they have been consequently associated with strength-based exercise cohorts. For example, the associations with *ACTN3* RR and elite strength events, and *MSTN* have been robustly associated with strength training (120, 121).

2.5.5 TIGER study:

The Training, Interventions and Genetics of Exercise Response (TIGER) study commenced in 2003 (122, 123). The aim of the study was to assess the effect of 30 weeks of MICT exercise on a primarily sedentary population of university age students. The cohort included over 1,567 participants (39% male) of different ethnicities (Caucasian, African American, Asian) and aged between 18-35 years old. Participants were required to perform 40 minutes of MICT, three days per week for a total of 30 weeks. The main purpose of the study was to determine the effect of 30 weeks of MICT on weight outcomes from sedentary populations. Overall weight retention was found to be 20% following the training regime. The findings of the study were used to measure the health benefits from regular exercise, important insights into currently known gene variants were absent.

In summary, these studies showed that there is a necessity for large-scale genomic studies within healthy individuals. The studies discussed, contained multiple limitations that prevented the discovery of novel exercise adaptation genes. Namely, the majority of previous studies

assessed diseased populations and therefore any results gained, whilst applicable to the disease groups within the respective studies, was not respective of healthy populations. In addition, these studies contained comparatively large (n>1,000) cohorts but the assessment of large-scale genomic associations was not addressed. While each of these studies had benefits, no studies implicated SNPs in exercise response within healthy populations.

2.6 GENE EXPRESSION STUDIES

Understanding which genes are being expressed within a cell or tissue may provide insight into the biological process contributing to cellular function, and any adaptation to homeostasis. As such, gene expression studies have been extremely useful for the determination of biological context within all fields of molecular biology. Previously, the assessment of mRNA transcript levels has been limited to candidate genes of interest through quantitative PCR technologies. With the advent of large-scale technologies including microarray, and more recently RNA sequencing, it is feasible to assess whole transcriptomes (sum of mRNA transcripts). Even so candidate gene expression studies are more common within the existing literature. The gene expression response of several key exercise genes (*PPARGC1A, ACE, MSTN*, *HIF1A*) have been well characterised in response to endurance training and are discussed in the following.

The *PPARGC1A* gene has been shown to increase in expression following acute endurance (see public Series (GSE) data held in the Gene Expression Omnibus (GEO): GSE44818, GSE87748, GSE43856) exercise, but is not differentially regulated following prolonged training or HIIT (124). This was expected as the expression of *PPARGC1A* has been robustly associated in acute exercise to elicit responses in mitochondrial biogenesis. The *ACE* gene showed the opposite effect, where acute bouts of endurance exercise did not elicit a response in gene expression. However; longer-term training (GSE35661: n=24, 6 weeks, supervised MICT cycling, GSE9103: n=42, self-reported 4 years, 6 days/week MICT) and HIIT (GSE97084: n=15, 4x4 minutes cycling, 3 days/week, 12 weeks HIIT, GSE109657: n=11, 6x20 seconds cycling, 6 weeks HIIT) produced a slight upregulation in expression. The *MSTN* gene was found to be slightly downregulated in response to both acute and long-term training. Lastly, *HIF1A* expression was found to be upregulated in response to acute endurance exercise but did not change following MICT or HIIT (124). Of note, ACE expression in kidney has been shown to be reduced in individuals with an I/I genotype (125).

A number of studies have now investigated the global gene expression changes within skeletal muscle using high-density technologies including microarrays and RNA sequencing. To date 29 studies have assessed global gene expression responses to aerobic exercise – 13 acute aerobic and 16 training/chronic, summarised in a recent meta-analysis by *Pillon et al*, (124). Specifically, there are two studies to date that have investigated HIIT in human participants (GSE109657, GSE97084), detailed in the following.

2.6.1 GSE109657

This study by *Miyamoto-Mikami et al,* assessed global gene expression response to HIIT training (126). The cohort consisted of n=11 healthy males and participants were required to complete 6 weeks of low volume HIIT (6-7 bouts of 20 seconds). Muscle biopsies were taken from the vastus lateralis muscle prior to the training and following the 6 weeks of HIIT. The study utilised Affymetrix microarrays (GeneChip Human Gene 2.0 ST Array) to assess global gene expression levels of 24,838 genes and validated the findings using quantitative PCR. The expression of 152 genes (79 upregulated, 73 downregulated) were identified to be differentially regulated following HIIT. Furthermore, the study performed gene ontology over-representation analysis and found glucose metabolism, extracellular matrix, angiogenesis, and mitochondrial membrane processes to be upregulated following the training.

Whilst this study was able to identify gene expression changes and molecular pathway changes in response to HIIT, there were several limitations that may have contributed to confounding information within the analyses. Firstly, the study was limited by small sample numbers and therefore the lack of statistical power within the study would have contributed to false negative results and these results may not be applicable to the wider population Secondly, it remains more costly on a per sample basis to perform microarray when compared with RNA sequencing, therefore limiting the number of available transcriptomes to analyse (127). In addition, the probes utilised for microarray can be based on sequences with cross-reactivity and/or low specificity (128). Thirdly, the study did not utilise paired analysis (i.e., estimates within pair differences to increase statistical power), which would have assisted to control the inter-individual variability in response, and therefore drastically improved the statistical power within the study (129). Last, the study did perform gene ontology analysis, but chose to only assess the upregulated genes and therefore omit half of the available data. This study assessed

the gene level and molecular pathway responses to HIIT, however the limitations within the study design and analysis prevented more significant findings and further highlight the requirement for robust exercise studies within healthy populations.

2.6.2 GSE97084

This study, generated by *Robinson et al*, represents the most robust study within the current literature and assessed gene expression, proteomics, and epigenetics in response to HIIT. The epigenetic findings from this study are discussed in **Section 2.8.3.** and will not be addressed here. The study assessed HIIT, resistance training, and combined training in young and old individuals. 72 participants were recruited for the study; however, these participants were separated according to the training regimen and the old/young status. Therefore, the number of participants performing HIIT was n=11 young (~25yo) individuals and n=8 old (~70) individuals. Of note, the VO_{2peak} measures between the young and old HIIT groups was significantly different prior to the study. The older individuals were omitted from the remainder of this discussion as they were not respective of a lean healthy population. The HIIT training within the study consisted of 4 bouts of 4-minute-high intensity training, with 3sessions per week for 12 weeks. Muscle samples from the vastus lateralis were taken at baseline and completion of the study. In addition, physical (BMI, weight, body fat) and biochemical (insulin, glucose) measurements were taken prior to and at completion of the study. Following total transcriptome Illumina sequencing, 274 genes were differentially regulated within the young HIIT group, with 211 of these unique to the young individuals and not shared with the old or resistance groups. Ingenuity pathway analysis (IPA) was utilised to assess gene level regulators of the entire dataset. The *VEGFA, VEGF, AGT, FGFR2, CTNNB1, SP1,* and *IL10RA* genes were found to have an upstream regulatory effect on the entire differentially regulated gene dataset. The authors concluded that genes upregulated due to HIIT in young individuals were associated with age, and any dissemination of genes associated with training were not examined. The study then assessed the molecular pathway changes in older individuals and did not examine these changes in the young HIIT cohort.

As stated previously, this study was the most robust within the field of exercise training relating to HIIT, and yet contained several weaknesses as outlined in the following. The study examined transcriptome, proteome, and epigenetic changes pertaining to HIIT in older

individuals. Whilst this study contributed a large amount of knowledge to the field of exercise science, there were several limitations within the study design and analysis that may have prevented more significant findings. Firstly, the sample size for the original study was n=72, representing a comparatively large exercise cohort. The study design then split this into young and old individuals, and subsequently three different training modalities. As such, what would have been a significant cohort was reduced to n=11 for young individuals that performed HIIT. Therefore, the study lacked statistical power when assessing Omic level data and may have excluded true positive results. Secondly, the study aimed to identify the molecular processes that were significantly differentially regulated as a result of training however the conclusions were based solely on age related changes. Indeed, the examination of pathways solely regulated in each of the training cohorts was not discussed. The results from this study indicated that even with a comparatively large cohort, a study design overloaded with research questions (young versus old, HIIT versus resistance, resistance versus combined, single Omics versus MultiOmics) may significantly reduce the statistical power of each of the questions being addressed. These findings represent the continued need for large-scale exercise studies examining HIIT within healthy populations.

The limitations outlined previously highlight important gaps within the established literature on response to HIIT. A recent (Jan 2020) meta-analysis by *Pillon et al*, aimed to address some of these limitations by incorporating 66 published exercise transcriptomic data sets. Further, the authors produced a database with an intuitive user interface to assess the gene expression signals for different genes, in different exercise training programs (http://www.metamex.eu/) (124). The primary findings from this meta-analysis are discussed in the following.

Firstly, the meta-analysis was able to replicate findings from the established literature as described previously. The aerobic training studies elicited increased levels in all mitochondrial respiratory chain subunits (I-V), increased levels of genes involved in lipid metabolism (*ACAD, ACAT, ACOT, ACOX*), did not elicit a large response in cytokine genes (*IL6, CCL2, CXC* genes), and increased the levels of markers associated with type I aerobic muscle fibres (*MYH7*). Secondly, the meta-analysis was able to identify a novel mediator of acute training response. The *Nuclear receptor subfamily 4 group A member 3 (NR4A3)* gene was found to be overexpressed in both acute resistance and aerobic training. Whilst there was no evidence for increased expression of this gene in aerobic training, the acute response showed a mechanism by which skeletal muscle may adapt to each bout of aerobic training. Lastly, the meta-analysis

was able to implicate several molecular pathways to each type of training. Interestingly, 153 gene ontology terms were enriched in healthy individuals that completed aerobic exercise. Particularly, metal ion transport was positively enriched (number of DEGs significantly changed in this term) in aerobic training whereas oxidative phosphorylation and nucleotide biosynthesis were negatively enriched. Whilst this is an extremely pertinent finding and contributes large amounts of novel knowledge to the field of exercise science, the molecular pathways were not fully explored within the manuscript as only the top ~20 pathways were shown, of which several had redundant biological functions due to the hierarchal nature of describing gene ontology terms and pathways.

2.7 EPIGENETIC STUDIES

Epigenetics is an overarching term for molecular processes that influence gene expression and protein stability without effecting DNA sequence (130). DNA methylation is the most widely studied epigenetic mark whereby a methyl group is covalently bound to Cytosine phosphate Guanidine (CpG) sites (131) to affect a wide variety of processes across the genome (imprinting, X-chromosome silencing) (132, 133). Gene promoter methylation is now widely accepted to repress gene expression, influencing the molecular landscape of cells or tissues (134, 135). In contrast, methylation of enhancer regions has been shown to permit expression of a target gene (136). The influence of these epigenetic marks to exercise adaptations has been previously examined in candidate gene studies relating to acute bouts of endurance training. Therefore, there has been consistent replication of epigenetic promoter silencing of several genes (*PPARGC1A, PDK4, PPAR-δ, TFAM*), discussed in the following (38, 137).

As previously discussed, *PPARCG1A* is the most heavily studied and consistently replicated gene within exercise populations. Methylation of the *PPARGC1A* promoter region at -260bp to the transcription start site has been shown to be hypomethylated in response to acute bouts of endurance training (138, 139), leading to an increase in *PPARGC1A* gene expression at three hours post training (140). Of interest, non-CpG related DNA methylation of this promoter regions has been found to be influenced by the *de novo* epigenetic modifying enzyme DNMT3B (141). Contrary to the hypomethylation event observed following acute training, sedentary behaviour has been shown to increase the level of promoter methylation (142).

Similarly to the findings observed for the *PPARGC1A* gene, *PDK4* was also found to be hypomethylated immediately following acute exercise with a corresponding upregulation in expression at three hours post exercise (143). Expression of *PDK4* has been previously identified to change in response both HIIT and MICT (144). Interestingly, PPARGC1A has been shown to coactivate PDK4 with PPAR-δ in skeletal muscle, highlighting a further adaptation mechanism that may be epigenetically mediated (145).

The promoters of the *TFAM* and *PPAR-* δ genes were also shown to be hypomethylated following acute exercise, however the upregulated changes in gene expression following this were also immediate indicating distinct differentiating mechanisms between the epigenetic regulation of the previously discussed genes (*PPARGC1A*, *PDK4*) (137).

No candidate gene studies have found further methylation changes within promoter regions associated with responses to training. High density microarray technology such as Illumina 850K EPIC BeadChips have allowed for rapid typing of epigenetic marks. Briefly, these technologies utilise probes designed to target ~850,000 CpG sites across the genome representing 4% of the known CpG sites. Whilst this appears to be narrow coverage of the genomic CpG sites, the probes are designed to target areas that may be reflective of epigenetic regulation of gene expression, and therefore 25.4% are located within promoter regions (146). A recent literature review summarised previous exercise studies in which DNA methylation was investigated in healthy subjects in response to acute and chronic exercise. Of the 22 studies included, three were classified as both chronic and endurance based (147-150). Only one of the studies assessed long term HIIT, further outlining the large gap of HIIT based endurance training studies within the literature (150). The primary findings from each of these studies are outlined in the following.

2.7.1 GSE60655

A study by *Lindholm et al*, aimed to determine longitudinal genome wide methylation changes in single leg-based exercise following three months of endurance training (147). The study contained n=23 participants, who were required to complete one legged cycling exercise with increasing loads (2 minutes 10-20W increased by 3-5W every 30 seconds). The training load was kept as high as possible to keep pace at 60rpm for 15 minutes. The study design for this project incorporated Illumina 450K methylation arrays and total RNA sequencing. Whilst this study was not based on HIIT, the findings highlight several genes and molecular mechanisms that respond to endurance training. The authors observed ~5,000 DMPs (differentially methylated probes) within ~4,000 genes over the genome in the trained leg when compared to the untrained leg. This study found that 4,076 genes were differentially expressed and largely inversely correlated with the epigenetic marks. Some of the primary findings include DMP enrichment in binding motifs within the *myelin regulatory factor* (*MRF*), *myocyte*

enhancer factor 2 (*MEF2*), and *ETS proto-oncogene* (*ETS*) genes. Interestingly, the study found that only 7.5% and 7.0% of the upregulated and downregulated genes respectively were also significantly and inversely differentially methylated. This study further investigated gene networks and found that processes involving skeletal muscle morphology, cell energy utilisation, and transcription were differentially regulated following the long-term endurance training. In summary the results obtained from this study represent unique and valuable insights into the molecular processes of skeletal muscle adaptation to endurance training.

2.7.2 Nitert et al,.

Secondly, a study by *Nitert et al,* assessed methylation changes following six months of endurance training (148). The participants contained a mixture of healthy (n=13) and T2DM (type II diabetes mellitus) related (n=15) young individuals. Participants were required to complete a 1 hour spin class and 2x1 hour aerobic classes per week for 6 months. VO_{2peak} was assessed via a max energy cycling test. The study design incorporated methylated DNA immunoprecipitation for microarray. The authors found that 134 genes were differentially methylated following the training, of which the large majority (n=115) were hypomethylated. Specifically, the authors identified known exercise response genes to be suppressed due to promoter methylation such as *myocyte enhancer factor 2A (MEF2A: -0.8 ± 0.01%), RUNX family transcription factor 1 (RUNX1: 20.4 ± 9.4%), NADH: Ubiquinone oxidoreductase subunit C2 (NDUFC2: 16.3 ± 3.7%),* and *THADA armadillo repeat containing (THADA: 3.6 ± 2.7%)*. Molecular pathways such as retinol metabolism and calcium signalling were also implicated within this study. It should be noted that the participants of this study were all first-degree relatives of T2DM patients, which may have resulted in confounding variables and therefore results that were not respective of a healthy exercise population.

2.7.3 GSE97084

As previously discussed in **Section 2.7.2**, *Robinson et al,* published the only HIIT study to incorporate healthy subjects, longitudinal sampling, and both gene expression and epigenomic analysis (150). Briefly, the study included 72 participants, separated into six exercise categories resulting in a sample size of 11 healthy individuals completing the HIIT. The study then assessed various physiological and biochemical measures relating to the influence of training on age. The gene expression results from this study were discussed in the

previous section and therefore the epigenetic results are examined here in isolation. The study did not see large-scale (>10%) changes in skeletal muscle methylation as is common in the existing epigenomic literature (137). This may have been due to several factors such as intervention (resistance versus MICT versus HIIT), heterogeneity of methylation signatures within skeletal muscle myofibres, or environmental lifestyle factors unknown to the authors. As discussed previously, this study contained several limitations (multi-hypothesis testing leading to lack of significant sample size) likely preventing robust findings. Of note, the study found 3,874 CpGs within promoter regions were differentially methylated between the age groups, however a need for high-density methylation analyses within healthy populations performing HIIT remains.

2.8 KNOWLEDGE GAPS

In summarising the findings pertaining to molecular skeletal muscle adaptations, several current knowledge gaps were identified and addressed within the current thesis. A recent study by *Halperin et al*, described several general limitations within current exercise studies (151). The limitations outlined within the study included 1) inadequate validation of clinically meaningful results, 2) few longitudinal studies, 3) underreporting of non-significant results, 4) few replication studies, and 5) insufficient scientific transparency. This was noted throughout the examination of the current literature relating to exercise genetics, transcriptomics, and epigenetics. Whilst there have been many exercise studies that examine endurance exercise and consequently identified genes and processes involved in adaptation, very few were centred around HIIT. Of these few, even less included more than ten participants, analysed young/lean/healthy individuals, examined Omic level data, or aimed to tie together multiple molecular datasets to robustly identify molecular processes involved in adaptation.to HIIT. Therefore, there remains a need within the current exercise science framework to 1), identify novel genes and molecular pathways that are responsive to long term HIIT, 2) identify whether other exercise related genes are differentially regulated following long term HIIT, and 3) investigate the results from the *Robinson et al*, manuscript in a larger HIIT cohort.

Exercise studies incorporating these factors will add invaluable knowledge to the field of exercise science. Molecular technologies such as RNA sequencing and mass genotyping and epi-typing through array technology allow for interpretation of omic level findings, which

offers far greater information when compared with previous technologies such as microarray This accuracy may be further enhanced by integrated analyses with all three Omics technologies through reduced redundancy from single Omics based errors (i.e., sequence artifacts, cross reactive probes). Longitudinal studies that utilise these technologies and a paired study design would allow for much greater statistical power and may allow for novel findings that have the potential to be functionally relevant to exercise adaptation. In addition, more recent analysis approaches such as gene set enrichment analysis and gene ontology further increase functional interpretation as they assess small but coordinated shifts in gene expression within a given pathway.

Chapter 3: The Gene SMART Study

The Gene SMART study was conducted at Victoria University by A.Prof. Nir Eynon, Dr. Sarah Voisin, Dr. Xu Yan, and Prof. David Bishop, and the methodology has been published within *BMC Genomics (152)*. As such, my original contribution to this chapter lies within the analysis of the phenotypic data provided, and not the participant recruitment or collection of original data. This chapter contains the necessary information from the study for the understanding of this thesis.

3.1 BACKGROUND

Due to the number of complex systems involved as a part of the response to exercise, unravelling the molecular processes that govern adaptations to training will assist with the understanding of many normative and diseased complex disorders. Previous exercise studies have been able to implicate several genes (*PPARGC1A, ACE, ACTN3, PDK4, AMPK*) and molecular pathways induced by transcription factors (VEGF, HIF1A) following acute bouts of HIIE (**Sections 2.6-2.7**), however long-term molecular changes in response to HIIT have yet to be explored.

Previous limitations pertaining to exercise studies have been summarised in a study by Halperin et al, (151). The study outlined five key limitations that must be addressed in order to progress the field of exercise science. Briefly, these were 1) inadequate validation of clinically meaningful results, 2) few longitudinal studies, 3) underreporting of non-significant results, 4) few replication studies, and 5) insufficient scientific transparency. Previous exercise studies have been characterised by small participant numbers, older technologies, cross-sectional data, lack of appropriate phenotype collection, or some combination of these (151). This issue is further exacerbated in exercise studies as inter-individual variability in training measures is consistently reported (153-155). The variability seen between exercise participants may be partially explained by genetic factors, however many of the previously identified genes and variants have been identified using underpowered genome wide association studies (GWAS) with diseased populations. In addition, it remains difficult to separate true physiological variability and noise generated from the exercise testing (156).

As such, there remains a need within the field for an improved study design (longitudinal, well phenotyped) and well controlled and phenotyped replication cohorts to further the understanding of exercise-based genomics. The Genes and Skeletal Muscle Adaptive Response to Training (Gene SMART) study was established in 2016 to address these challenges in exercise genetics and represents a moderately trained, longitudinal cohort of High-Intensity Interval endurance Training (HIIT) exercise training. This study will allow for the generation of novel data pertaining to genetics, transcriptomics, and epigenetics in response to HIIT and will provide significant insight into the molecular processes that govern the skeletal muscle adaptive response to training.

3.2 METHODS

A detailed description of the Gene SMART study regarding design, recruitment and methodologies has previously been reported by **Yan et al, 2017** (152). It is beyond the scope of this chapter to repeat this information. **Below is a summary of these details as they pertain to this sub-study.**

3.2.1 Participant Recruitment

Ethical clearance for this study was provided by the Human Research Ethics Committee at Victoria University (Approval Number: HRE13-233), and the clearance was transferred to and also provided by the QUT Human Research Ethics Committee (Approval Number: 1600000342). All participants provided informed consent prior to the study and all methods were carried out in accordance with relevant guidelines and regulations. All participants were recruited at Victoria University, Australia, of which to date the study has recruited over one hundred of male and female participants (aged 18-45 years old) to participate in the four-week HIIT program. As a number of covariates, including sex, ethnicity and weight are known to additionally contribute to varied exercise response, strict inclusion criteria for this study was applied. A sub-population of the Gene SMART study were selected based on the inclusion criteria in Table 3-1, resulting in the final inclusion of 77. The participants. (n=77) for the Gene SMART study were included based on multiple criteria (Table 3-1) to limit the number of covariates contributing to unwanted variation within the data.

Table 3-1: *Inclusion criteria for the participants of the Gene SMART study*

Trait	Inclusion Criteria Range
VO_2peak	35-60mL/min/kg
Sex	Male
Age	18-45 Years
Ethnicity	Caucasian
BMI	20-30kg/m^2
Body fat percentage	<25%

3.2.2 Study Design

Training for all stages of the Gene SMART study was performed on electronically braked cycle ergometers. The Gene SMART study was stratified into four main stages: Screening and Familiarisation, Baseline testing, HIIT training, and Post training testing (**Figure 3-1**).

Firstly, and to control for previous exercise training, the physical activity of participants was measured through an ActiGraph GT3X+ monitor (*ActiGraph LLC*) for the one-week screening and familiarisation stage. The baseline testing phase was utilised to assess the starting point for all participants for future normalisation. Participants were required to perform a graded exercise test (GXT) and a 20 km time trial (TT) on a cycle ergometer (Velotron, Racer Mate Inc., Seattle, USA). The results from these tests were used to normalise the testing across participants and establish baseline measures for further testing. The HIIT program was the main sampling phase of the study and consisted of four weeks of high intensity interval cycling exercise with three HIIE sessions per week. Each HIIE session consisted of a five-minute warm up period (60W) followed by 6-14 intervals where intervals involved two minutes of high intensity cycling followed by one minute of rest. A five-minute cooldown was performed at the conclusion of each training session. The number of intervals ranged from 6-14 to ensure progression through training. Further, cycling intensity was initially set at predetermined threshold plus 40%Δ resistance (Δ = difference between peak power (PP) and lactate threshold (LT)). Resistance was then increased by 10%Δ each week to final resistance of 70%Δ by the fourth week. It should also be noted that all HIIE sessions were separated by 48-hour intervals and nutrition was standardised 2 hours prior to each session. Diet was also controlled 48 hours prior to and following the four-week HIIT program. The post-training testing phase was initiated at the conclusion of the four-weeks of training. Participants reported to the lab for four visits, the first of which muscle and blood samples were taken at rest. As previous, a 48-hour controlled diet was administered prior to the biopsy. All visits were separated by a minimum period of 48 hours. GXT and VO_{2peak} tests were conducted during the second and fourth visits as described previously and conducted at the same time as the baseline testing phase to eliminate confounding variables. The average of the GXT measures was used where the difference was less than 5%, otherwise the highest value obtained for this test was used as the post-training measurement. Participants conducted a last 20Km time trial on the third visit.

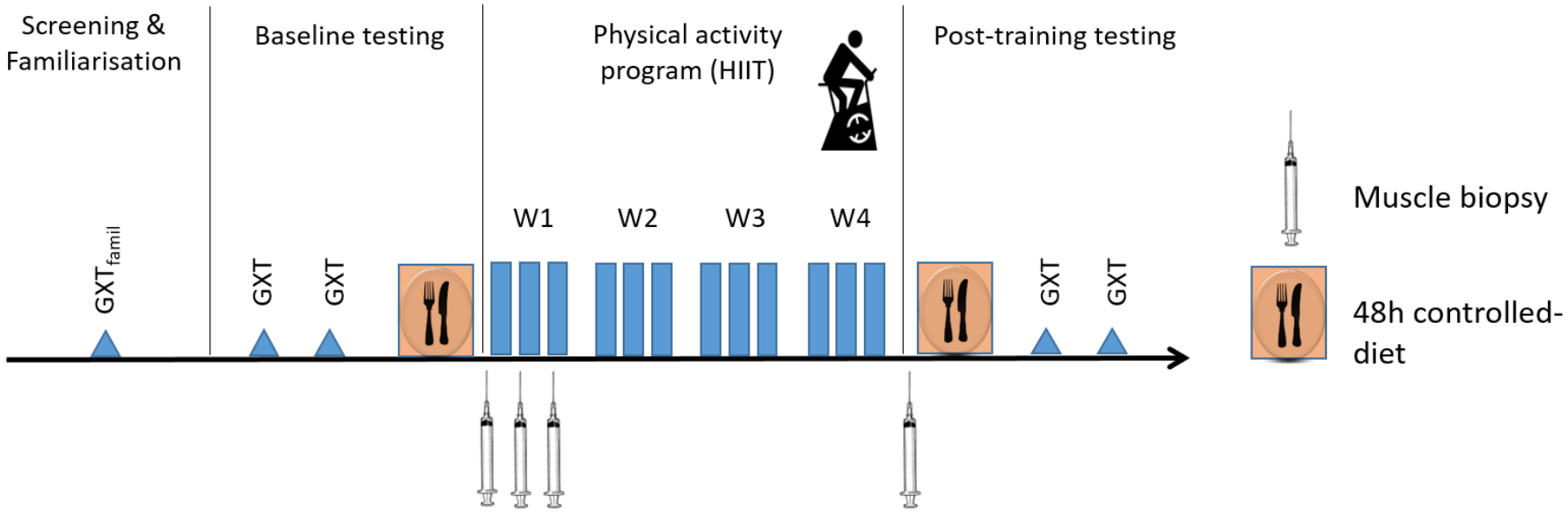


Figure 3-1: *Linear timeline depicting the Gene SMART study design. Phases of the study are separated by vertical lines. Blue triangles represent graded exercise tests (GXTs) and blue rectangles represent a session of high intensity interval exercise (HIIE). 48h control diet is shown as a plate symbol on the timeline. Sample extraction is depicted as needles at the bottom of the figure.*

3.2.3 Phenotypic and biochemical testing

Three methods were utilised for the collection of phenotypic data: 20 km TT, GXT and VO_{2peak} test. Of particular note, these tests were utilised at different stages and times during the study framework to ensure appropriate data collection over the timeframe.

An electronically braked cycle ergometer (Lode-excalibur sport, Groningen, Netherlands) was used in the familiarisation, baseline testing, and final stages of the study for a 20Km time trial test. A warmup period of cycling at 60W for 5 minutes preceded the time trial. The participants then completed 20Km in the fastest possible time. Participants were only permitted to see distance completion measures with no time or power output from the ergometer visible. This phenotypic test contributed solely to the "Time Trial" response phenotypes generated. The cycle ergometer was again utilised during the familiarisation, baseline testing, and final stages of the study to perform a GXT. The 20Km TT test results for each participant were used to determine whether the test started at 60, 90, or 120W speed. The test was performed in 4-minute stages (30sec rest), with a 30W load added with each additional stage. Participants continued the test until exhaustion and individualised exercise intensities were set according to performance during the test. This test was used to calculate the "Lactate Threshold" (modified DMAX method (157, 158)) and "Peak power (W_{peak})" measures utilised. The VO_{2peak} phenotypic test was measured during baseline testing, and at completion of the study. Participants were required to wear COSMED face masks following 5 minutes of rest after GXT. VO_{2peak} was measured at rest for 2 minutes and then while exercising at the same rate as the GXT. The highest value obtained in a single minute of testing was taken as the VO_{2peak} value.

3.2.4 Sampling

Muscle biopsies and matched blood samples were taken at four major time points following fasting:

- **PRE time point:** collected prior to the first HIIE session. Participants were required to perform the session of HIIE immediately following the biopsy.
- **P0 time point:** collected immediately after the first HIIE session.
- **P3 time point:** collected three hours after the first HIIE session.
- **4WP time point:** collected at rest at least 48h following the four-week training phase, during the first post-training visit.

Muscle biopsies were taken from the vastus lateralis of the participant's dominant leg using a modified Bergström needle (68). Briefly, the procedure was performed under local anaesthetic (5mL, 1% Xylocaine), after which an incision in the leg was made to enable the insertion of the Bergström biopsy needle. Muscle biopsies (50-200 mg) were collected under manual suction, blotted to remove residual blood, then immediately snap-frozen in liquid nitrogen and stored at -80°C. For this study, an aliquot (~20 mg) was removed from the original biopsy on dry ice. Venous blood samples (5 mL) were collected in EDTA blood collection tubes, centrifuged, the plasma supernatant removed, and the remaining cellular portion stored at -80°C for subsequent DNA extraction.

As previously discussed, (**Section 3.2.1**), the initial inclusion criteria for this study resulted in the selection of 77 participant from the Gene SMART cohort. As shown in **Figure 3-2**, all 77 participants were genotyped (gDNA and mtDNA) using venous blood samples (**Chapter 4 and 5**). Muscle biopsies from at least three of the four timepoints were available for 54 participants and underwent subsequent transcriptomic analysis (**Chapter 6**). Finally, a subset of 19 participants (with exact matching transcriptome time point samples) were used to investigate the skeletal muscle epigenome at PRE, P0 and 4WP time points (**Chapter 7**). *Barres et al*, reported that methylation was changed for exercise responsive genes immediately following acute exercise (137). As such, and due to cost constraints, the three-hour timepoint was omitted from epigenetic analysis.

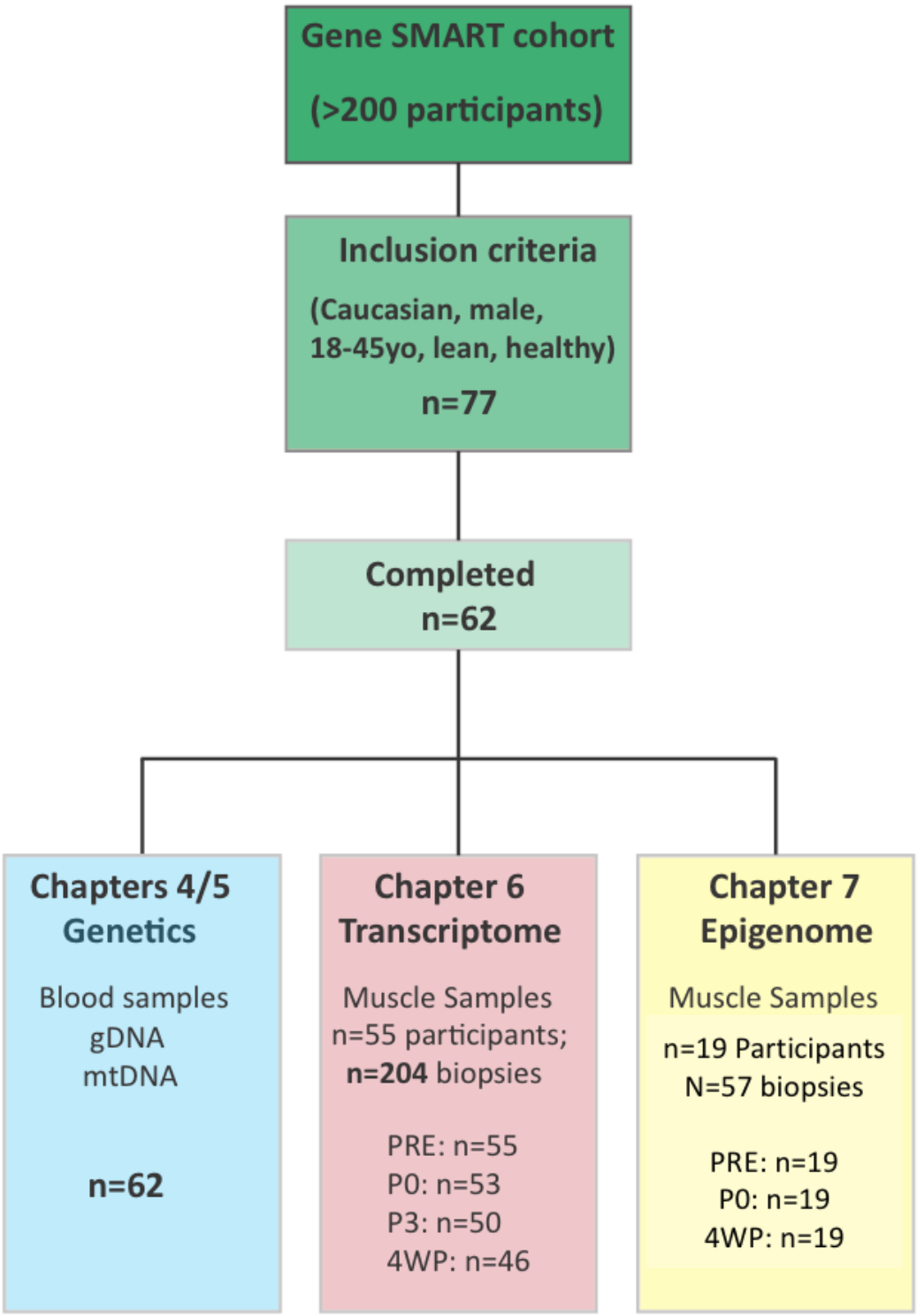


***Figure 3-2**: Flowchart detailing the number of participants and samples utilised for each project within the Gene SMART study. At the start of this thesis project, 77 participants had initiated the study, of which 62 had completed. All 62 participants were utilised for **Chapter 4** and 5. At least three timepoint muscle biopsies were available for 55 of the 62 participants and underwent transcriptomic analysis (**Chapter 6**). Finally, 19 participants underwent epigenetic analysis (**Chapter 7**).*

3.2.5 Data processing

Accompanying phenotypic data was obtained from collaborators at Victoria University for outlined pre and post data of 4 phenotypic tests (Power output, Lactate threshold, VO_{2Peak}, and 20Km time trial). To ascertain variants associated with response for the various phenotypic traits, the delta (Post phenotype – Pre phenotype) quantitative trait data for the Δ-Weight, Δ-BMI, Δ-power output (Δ-Wpeak), Δ-maximal oxygen uptake (Δ-VO_{2max}), and Δ-time to completion measurement for a 20Km time trial (Δ-TT) was used. Analysis of the response traits was performed in SPSS (version 17.0) using a paired samples t-test.

To determine the importance of variants in multiple phenotypic traits; Principal Components Analysis (PCA) was performed using the R package *FactoMineR* (159) and the top principle components (PCs) were added to the quantitative traits for testing. Of note, the inclusion of the BMI and weight measures to the PCA was to ensure these were not skewing the top PCs and these measures were not used for further association testing. PCA is a dimensionality reduction method that computes linear combinations of the multiple response phenotypes into PCs so that the variance between individuals is maximised. Every individual is then represented by one value for each PC, considered a composite trait of the different response phenotypes. Missing phenotypic values were excluded from the phenotype table prior to PCA to prevent skewing of data and to maintain appropriate PCs. Single Value Decomposition (SVD) analysis was also used to determine which quantitative variables were most represented within the top PCs (**Figure 3-3**).

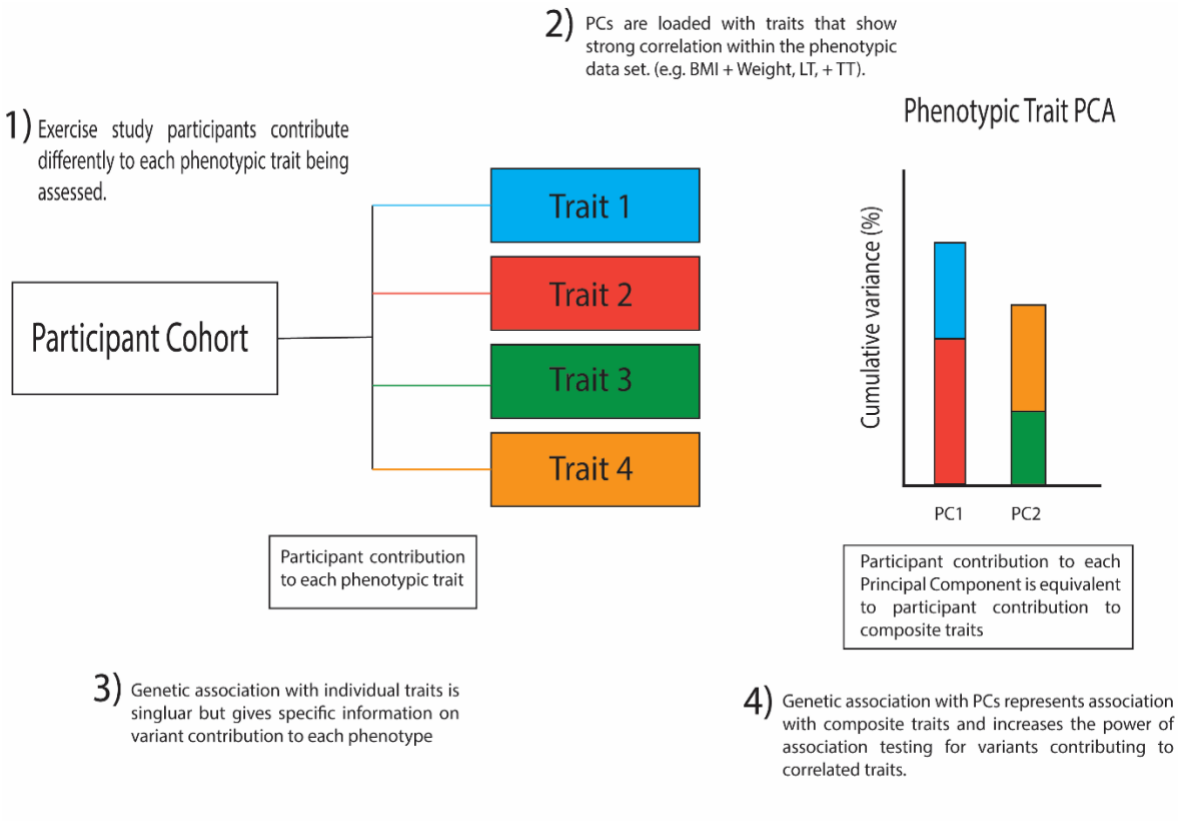


Figure 3-3: *Building composite traits with principal component analysis and determining the contribution of each participant outcome to each Principal Component. Each trait reduces into the variability within each PC that we can gauge with loadings from an SVD analysis. Then each PC may be explained by a contribution from each participant. Therefore, association with the participant contributions is equal to the association of composite phenotypic traits. (160)*

3.2.6 Power analysis

Statistical power analysis prior to study design ensures appropriate sample sizes for the discovery of significant results. This study was limited by the sample size, however multiple factors addressed in this study allowed for the potential discovery of statistically significant results. This study contained several individual projects and as such the power analyses varied with each project.

A priori power analysis for the genetic component of the study was conducted in R (3.6.2.) and indicated that the linear modelling approach with additive genotypic effects for our sample size (n=62) was sufficient for at least 80% power to detect SNP-based heritability of 13% or more at the relaxed alpha level of 0.05. Of note, the Gene SMART cohort is a tightly controlled study

with rigorous physiological measures, all performed in duplicate, which significantly increased the power of the detected *a priori*.

For the transcriptomic analyses, sample size estimates were generated using a model by *Hart et al,* (161). Briefly, the model uses non-binomial distribution to determine optimal study design for RNA-seq experiments. Further, the authors were able to derive a formula based on the relationship between technical and biological variability. Briefly, the sequencing technology utilised for this study generated an average of 10 million reads per sample. A transcript cut-off of 0.5 counts per million reads was used to ensure all transcripts were present at least 20 times for further analysis. A total of n=55 participants were utilised for this study and therefore the power calculations resulted in 80% power to detect 1.3-fold gene expression changes. This power was further enhanced through a paired design, which allowed for subject level confounders (age, BMI) to be estimated out of the analyses (129). *Stevens et al,* found that a study with a fully paired design and n=10 participants achieved greater statistical significance than a cross-sectional study with n=20 samples. In addition, the study found that a lack of sample pairing resulted in fold change underestimation. The Gene SMART study has controlled for various confounding phenotypes (outlined in **Section 3.2.1**) however the exercise phenotypes were found to be varied. As such, a high level of sample pairing (80-90%) would achieve a moderate statistical power enhancement.

Power analyses for genome wide epigenetic changes have typically relied on similar significance thresholds to GWA studies ($P<5x10^{-8}$) (162). The epigenetic aspect of this study utilised a subset (n=25) of Gene SMART participants and would therefore not reach this level of significance. The study was adapted to assess coordinated changes in DNA methylation over promoter regions and increase the power of the study for the determination of epigenetically regulated molecular pathway changes. The statistical power for this study was similar to that of the transcriptome analysis due to the paired design and set enrichment analysis utilised. In addition, the potential MultiOmic analysis would further increase the scrutiny of the discovered molecular pathways and lead to greater confidence in the findings.

3.3 RESULTS AND DISCUSSION

After applying the selection criteria to the Gene SMART cohort, a total of 77 (62 completed) participants remained and were included in this study. This participant population had an average age of 31 ± 8.2 years and an average weight of 82 ±12 kg prior to training. Following the four-week HIIT program the response to a number of phenotypes were assessed as shown in **Table 3-2**.

Table 3-2: *Summary statistics for the Gene SMART response phenotypes. (160)*

Phenotype (units)	**Time point**	**Mean** (L95%CI-U95%CI)	**Min-Max**	**SD**	**D**	**P-value**
BMI (kg/m^2)	PRE	25.06 (24.23-25.89)	17.77-35.05	±3.20		
	POST	25.12 (24.29-25.94	18.37-35.36	±3.27		
	Δ	0.04 (-0.04-0.15)	-1.10-1.00	±0.37	0.02	0.114
Peak Power (Watts)	PRE	296.88 (279.11-314.66)	184.00-490.5	±70.57		
	POST	315.84 (298.78-332.91)	197.00-504.00	±67.77		
	Δ	18.96 (14.81-23.12)	-16.00-65.5	±16.49	0.27	$2.28e^{-13}$
Lactate Threshold (Watts)	PRE	209.22 (194.19-224.26)	112.40-385.80	±59.70		
	POST	224.91 (209.64-240.20)	130.65-398.55	±60.68		
	Δ	15.69 (11.60-19.78)	-21.00-50.70	±16.24	0.26	$7.47e^{-11}$
VO_{2peak} (mL/min·kg)	PRE	46.34 (44.49-48.20)	28.14-70.89	±7.36		
	POST	47.46 (45.69-49.23)	28.04-59.40	±7.04		
	Δ	1.12 (0.15-2.08)	-11.49-8.59	±3.84	0.16	0.012
Time Trial (seconds)	PRE	2295.99 (2212.61-2360.29)	1788.44-2979.16	±292.95		
	POST	2194.13 (2131.56-2253.97)	1749.08-2965.05	±246.91		
	Δ	-101.86 (-131.38--55.99)	-489.02-463.05	±144.64	-0.35	$2.81e^{-6}$

Δ: Delta change, Min: Minimum value, Max: Maximum value, L95%CI-U95%CI: Lower and upper bound 95% confidence interval, *SD: Standard Deviation, D: Cohen's D*

Interestingly, there has been very little evidence for the applicability of HIIT for weight change in healthy populations (163, 164). Further, another study examining the effects of HIIT on weight loss found that a period of 12-15 weeks of training was required to elicit responses in weight (165). The weight of the Gene SMART participants was expected not to fluctuate for two reasons. Firstly, the participants were active and healthy and therefore any fluctuation in weight would not be significant enough to measure accurately. Secondly, the four-week training program would not be enough to elicit a weight response, even for overweight individuals. Lastly, the dietary interventions utilised for the Gene SMART study were not designed to elicit weight responses but to ensure consistency in testing.

An increase in peak power, lactate threshold and VO_{2peak} were observed following the four-weeks of HIIT. A small improvement (18.96 ± 16.49 Watts) in peak power demonstrated that participants were now able to exert more force during the GXT. Whilst the increase in lactate threshold showed that participants were able to maintain exercise intensity for longer prior to accumulation of lactate. A significant increase (1.12 ± 3.84 mL/min·kg) in VO_{2peak} was also observed in most individuals, thus demonstrating an overall improvement in maximal aerobic capacity. Finally, on average participants were able to complete the 20 km TT faster (Δ TT = -101.86 seconds).

Taken together this data shows that following the four-week HIIT program most participants improved with respect to aerobic fitness.

3.3.1 Composite trait building

Participant stratification into high and low response groups leads to a loss of statistical power in testing. As such, and to avoid classifying responders and non-responders via arbitrary thresholds, the phenotypes were kept as continuous variables for association testing in Chapter 4 and 5 (166).

Following PCA on the response traits, the first four PCs (PC1: 35.49%, PC2: 28.46%, PC3: 16.51%, PC4: 12.74%) cumulatively explained 93.2% of the total variance within the response phenotypes (**Figure 3-4**); therefore, only these first four PCs were included in subsequent analyses. In utilising the SVD function of the PCA, the contributions of each Δ trait in each of

these PCs was identified (**Table 3-3**). PC1 was loaded with traits related to weight response, PC2 with the peak power output, lactate threshold and VO_{2peak} response measures, PC3 contained the lactate threshold, VO_{2peak}, and time trial response phenotypes, and PC4 was composed of the same phenotypes as PC3, with VO_{2peak} was more heavily weighted (52% *vs* 24%).

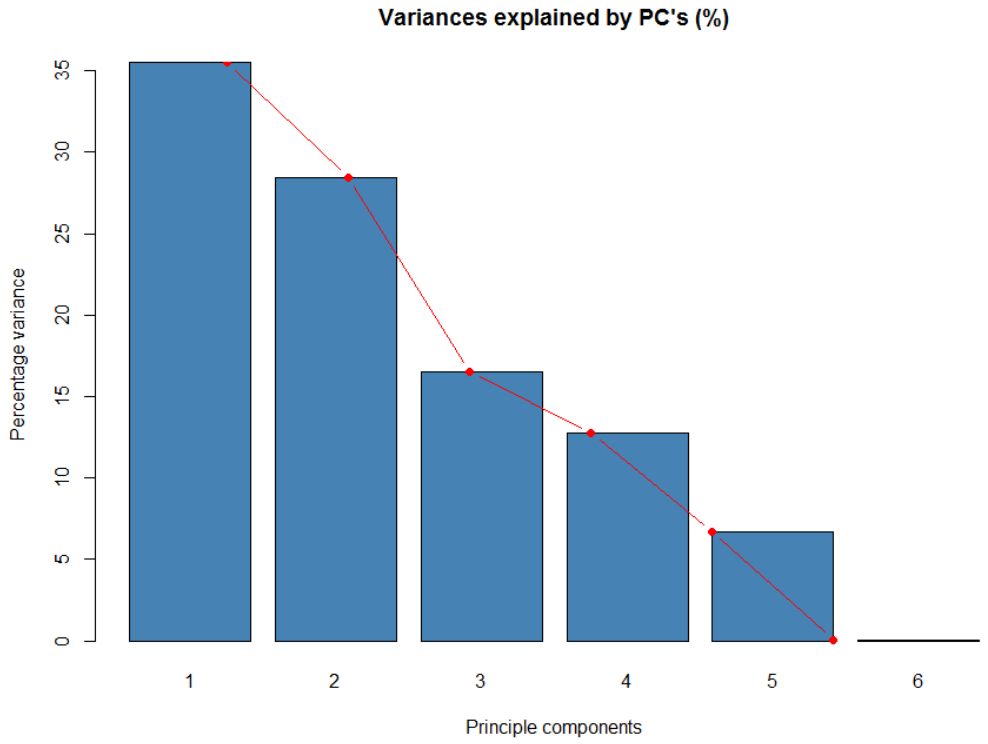


Figure 3-4: *PCA scree plot of the variance explained by each principal component (PC). PCs are shown on the X-axis and percentage of overall variance in shown on the Y-axis. (160)*

Table 3-3: *Percent contribution of each response trait to the top four principle components (160)*

	Δ-Weight	**Δ-BMI**	**Δ-Wpeak**	**Δ-LT**	**Δ-VO_{2max}**	**Δ-TT**
PC1	44.74	44.93	1.40	0.29	2.18	6.45
PC2	0.35	0.46	43.02	37.70	17.76	0.71
PC3	0.83	0.80	0.69	11.27	23.71	62.70
PC4	3.69	3.03	2.03	9.33	52.06	29.86

PC: Principal Component #, Δ: Change in phenotype, Green shading highlights which response traits are most loaded in each PC, green shading represents non-exercise traits loaded into respective PCs

3.4 CONCLUSION

Previous exercise studies have characteristically lacked the combination of large participant numbers, longitudinal time points, collection of rigorous phenotype and biochemical information, and samples from skeletal muscle and blood. The Gene SMART study incorporates all of these aspects and represents a large-scale collaborative contribution to the field of exercise science.

Following the four weeks of HIIT, the participants significantly improved in all tests, however there was no significant difference in the weight or BMI measures. This was expected, as the four weeks of training was not designed to elicit a response to weight. Of note, the inclusion of both weight and BMI within the model was expected not to skew the PCA. Each metric contributed extremely small amounts (0.35-0.83%) to the exercise relevant PCs. Further, the inclusion criteria ensured that all participants were fitter individuals, and as such any possible weight change due to training would not be varied enough to observe a response. When composite trait building was assessed, the weight measures were included into the phenotypes table to ensure no variation in these measures was contributing to other phenotypes. Successful PCA found that the top four PCs explained the majority (93.2%) of the variance within the total data set with PC2-4 loaded with the response phenotypes of interest (PP/LT/VO_{2peak}/TT) and as such used as composite traits of interest.

The Gene SMART study represents a unique opportunity to uncover the molecular basis of the response to HIIT through its longitudinal study design, comparatively large participant numbers, and collection of rigorous phenotypic and biochemical data. On average, the participants improved in each phenotypic metric apart from weight/BMI which was expected. In spite of this tightly controlled study, variability in the responses was observed indicating that some of the participants did not respond to some or all aspects of the training. Composite trait building allows for increased statistical power when considering biologically related variables. These composite traits were generated for use in **Chapter 5** to uncover mitochondrial genetic variants that influence multiple physiological outcomes.

Chapter 4: SNP Genotyping (MassARRAY)

The purpose of this chapter was to investigate previously associated exercise SNPs in our exercise cohorts using a designed MassARRAY for multi-allele genotyping. Findings from this study prove the need for accurate and reproducible replication studies in exercise genetics. This chapter follows the formatting required for manuscript submission. Section **4.1** outlines the abstract pertaining to this project of the thesis. The background to this study is described in Section **4.2**. Section **4.3** (Methods) discuss the different exercise populations (**4.3.1, 4.3.2**), the MassARRAY design (**4.3.3**), data processing for downstream applications (**4.3.4**), and the statistical analysis (**4.3.5**) used in this project. Section **4.4** details the association results for each cohort and Section **4.5** discusses the results and possible mechanisms why these genetic variants are influencing training. The work from this chapter has been published in the journal *Molecular Genetics and Genomics* as outlined below:

Manuscripts resulting from this work:

- **Harvey NR**, Voisin S, Dunn PJ, Sutherland H, Jacques M, Yan X, Papadimitriou ID, Ashton KJ, Haseler LJ, Haupt LM, Eynon N and *Griffiths LR, (2019) Multiple genetic variants associated with exercise performance in both moderately trained and highly trained individuals. *Published Jan 2, 2020, Molecular Genetics and Genomics, DOI: 10.1007/s00438-019-01639-8*

Authors Contributions:

Harvey NR designed the MassARRAY, performed the laboratory component, performed the analysis, wrote the manuscript, and was involved in conception and drafting of the final version. Dunn PJ was involved in conception of the manuscript and drafted the final version. Voisin S, Yan X, Papadimitriou ID, Jacques M, and Eynon N were involved in conception of the original Gene SMART study design, recruitment of participants, and drafted the final manuscript. Haseler LJ was involved in the recruitment of the Ironman participants and drafted the final manuscript. Sutherland H, Ashton KJ, Haupt LM, and Griffiths LR were involved in the conception of the study design and drafted the final manuscript.

The work within this thesis chapter addresses the first aim of this thesis as detailed below:

To genotype previously discovered SNP markers in the Gene SMART participants using a MassARRAY platform;

a. Research previously discovered exercise SNP markers to develop a list of potential candidates for the Gene SMART population

b. Design a MassARRAY with the highest multiplex possible to ensure cost effective genotyping

c. Perform a basic association of the Gene SMART genotypes to discover any significant exercise genetic variants

4.1 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 A 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.

4.2 BACKGROUND

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%) (167). However, a wide variety of environmental (e.g. diet, sleep), psychological and epigenetic factors may also influence exercise responses (168). 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 (166).

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 (169, 170). 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 (171). 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 (172). 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) (173).

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 four weeks of endurance exercise training, regardless of age or baseline fitness level.

4.3 METHODS

4.3.1 The Gene SMART moderately trained cohort

The Gene SMART (Skeletal Muscle Adaptive Response to Training) study design has previously been described (152). 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 three hours after a single bout of high-intensity endurance exercise (HIIE), and after four weeks of High-Intensity Interval Training (HIIT) (152). 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 = 62 completing the entire four-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.

4.3.2 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 (82, 174). Due to the energy demands 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 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 (174). 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.

4.3.3 SNP selection

The SNPs investigated in this study (**Table 4-1**) 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 (169). 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).

4.3.4 MassARRAY protocol

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 (175, 176). An internal genotyping control SNP (rs17602729, *AMPD1*), previously

validated in our endurance cohort, was used to ensure the MassARRAY system correctly identified genotypes (82). The protocol has been detailed below with specific reference to any adaptations to the protocol.

Prior to starting the laboratory protocol, the extension primers were pooled in equimolar amounts and analysed on the Mass spectrometer to ensure even pooling of the primer masses. Different additions of the uneven primers were added to the pooled mix and flown again until the primers were all within 1.5 standard deviations of the mean heights.

Day 1:

Firstly, a PCR was used to amplify the genomic regions surrounding the SNPs of interest on the designed MassARRAY. The dNTPs and *Taq* polymerase concentrations were modified to 500µM and 1U respectively to suit the level of MassARRAY (n=36). The PCR cycling conditions were as follows: first denaturation was 94ºC for 15 min. This was followed by 45 cycles of 94ºC for 20 seconds, 56ºC for 30 seconds, and 72ºC for 1 minute. The PCR was finished at 72ºC for 3 minutes followed by a holding step at 4ºC.

For the second step, Shrimp Alkaline Phosphatase enzyme was prepared with a corresponding SAP buffer and diluted with dH_2O. 2µl of the SAP mixture was added to each sample well on the 96 well plate and each well was observed to have ~10ul total volume. The plate was incubated in a thermocycler at 37ºC for 40 minutes and then 85ºC for 5 minutes followed by a holding step at 4ºC.

Thirdly, a master mix containing iPLEX buffer plus, termination mix, and iPLEX enzyme was made and 2µl of the solution was added to each sample well on the 96 well plate. The plate was incubated in a thermocycler using the following conditions: 94ºC for 30 seconds followed by 40X (94ºC for 5 seconds, then 5X (52ºC for 5 seconds and 80ºC for 5 seconds)) then 72ºC for 3 minutes and a hold step at 4ºC.

Day 2:

For the fourth step, Clean Resin was added to a 96 well dimple plate and scraped with a flat plate to smooth the resin into each dimple. dH_2O was added to each sample well on the 96 well plate up to 50µl. the plate was inverted and placed onto the dimple plate with resin so that the wells were directly over the dimples. The dimple plate was inverted and tapped lightly to ensure all resin fell into the sample plate. The sample plate with added resin was sealed and spun on a rotator for at least 30 minutes prior to loading onto a SpectroCHIP.

The sample plate was centrifuged to collect the resin at the bottom of the wells. The cleaned iPLEX assays were added to the corresponding plate place on a Nanodispenser to add samples onto SpectroCHIPs. 4-point calibration buffer was added to the calibration chamber in the Nanodispenser. A SpectroCHIP was added to the Nanodispenser and the barcode was noted for downstream chip linking processes. The SpectroCHIP was loaded using the 96 plate to 96 chip protocol and then loaded into the MassARRAY analyser.

Table 4-1: *Details on the 36 SNPs included in the custom MassARRAY genotyping assay. (177)*

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	(178)
6	*HFE*	rs1799945	G	END	2	148	-	-	(178)
6	*VEGFA*	rs2010963	C	END	1	942	-	-	(178)
11	*ACTN3*	rs1815739	C T	END POW	4 12	560 1,484	14 5	3,039 498	(178)
11	*UCP2*	rs660339	T C	END POW	1 1	694 29	- -	- -	(178)
19	*CKMM*	rs8111989	G	POW	2	233	-	-	(178)
21	*COL6A1*	rs35796750	T	END	1	661	-	-	(178)
22	*PPARα*	rs4253778	G C	END POW	5 2	740 260	- 1	- 81	(178)
Category 2: Mixed results									
1	*SGIP1*	rs9633417	C C	EBH POW	2 -	2,838 -	- 1	- 753	(179)
1	*LEPR*	rs1137101	A G G	END POW EBH	- 1 3	- 242 3676	1 - -	846 - -	(179-181)
4	*PGC1α*	rs8192678	A	END	4	849	3	508	(178)
4	*UCP1*	rs10440457	G	EBH POW	2 -	2,838 -	- 1	- 181	(179)
4	*PGC1α*	rs6821591	T	END	1	235	-	-	(182)
5	*ADRB2*	rs1042713	A	POW	1	100	-	-	(178)

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	(183, 184)
8	*ADRB3*	rs4994	C	END	1	100	1	81	(178)
14	*BDKRB2*	rs1799722	T	END	1	316	-	-	(178)
15	*NRF2*	rs7181866	G	END	2	129	1	89	(178)
15	*NRF2*	rs8031031	T	END	1	74	1	89	(178)
				Category 3: Disease Associations and other					
1	*ATP1A2*§	rs28933400	T	HYP	-	-	1	388	(185)
1	*LEPR*	rs12405556	T	EBH	2	978	-	-	(179)
1	*DIO1*	rs2294512	A	THY	1	547	-	-	(186)
2	*MSTN*	rs1805086	G	POW	13	3,080	-	-	(82, 187)
4	*UCP1*	rs2270565	T	T2D	2	981	4	1,382	(188-190)
6	*EDN1*	rs5370	T	HYP	2	1,004	-	-	(191, 192)
6	*HLA-A*	rs1061235	T	END	1	32	-	-	(180)
7	*IL6*	rs1474347	A	T2D	1	10,775	-	-	(193, 194)
10	*ADRB1*	rs1801253	C	HYP T2D	1 1	61 947	- -	- -	(195, 196)
12	*IGF1*	rs121912430	T	OBE	1	502	-	-	(180)
15	*CYP19A1*‡	rs2470158	T	EBH	1	1,722	-	-	(179)
16	*Intronic*	rs238838	C	-	-	-	-	-	(180)
18	*MC4R*	rs9965495	A	T2D	2	6,657	-	-	(179, 197)

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	(198-203)
22	*PPARα*	rs1800206	G	HYP POW T2D	1 1 4	269 610 3,643	- - -	- - -	(204-210)
MT	*MTTT*	rs199474700	G	END	1	46	-	-	(180)
MT	*MTND5*	rs28359178	A	END	1	46	-	-	(180)

NB: Grey rows represent intergenic variants included from GWAS conducted by *Rankinen T. et al.* (180).

Phenotypes represent 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,.(178)). E: Endurance, P: Power/Strength, T2D: Diabetes, HYP: Hypertension, *: Other including dementia and behavioural studies, #: Inbuilt genotyping control

 # excluded in Gene SMART population.

Excluded from highly trained population.

4.3.5 Data processing

The output files from the MassARRAY platform were converted to PLINK (version 1.07p) 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) (175). Subsequent analysis was performed on n = 29 SNPs for the Gene SMART population and n = 30 SNPs for the Ironman population.

4.3.6 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 (211). 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”.

4.4 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 (82, 174).

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 4-2**): *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).*

Six variants in five distinct genes were nominally associated with time-to-completion of Ironman events (**Table 4-3**): *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.

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).

Table 4-2: *Summary of nominally significant variants associated with gains in endurance fitness after exercise training in the Gene SMART cohort.(177)*

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

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).

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 4-3: *All nominally significant variants associated with different triathlon event finishing times in the highly trained endurance cohort. (177)*

Trait	CHR	SNP	Gene Symbol	Effect	Model	MAF	P-value*	FDR	Effect size (Beta)
Swim time	7	rs6949152	*NRF1*	Intronic	REC	0.14	0.019	0.47	0.2459
	19	rs8111989	*CKMM*	Downstream variant	DOM	0.34	0.019	0.31	-0.0903
	6	rs1061235	*HLA-A*	Non-coding transcript	DOM	0.061	0.02	0.31	0.1392
Cycle time	1	rs9633417	*SGIP1*	Intronic	DOM	0.1	0.036	0.89	0.3238
	4	rs8192678	*PPARGC1α*	Missense	REC	0.36	0.038	0.40	0.2965
	4	rs6821591	*PPARGC1α*	Non-coding transcript	REC	0.45	0.041	0.40	-0.2977
Run time	1	rs9633417	*SGIP1*	Intronic	DOM	0.1	0.019	0.61	0.424
Total-time	1	rs9633417	*SGIP1*	Intronic	DOM	0.1	0.012	0.37	0.846

Tests were performed for both dominant and recessive models for each trait:

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*

4.5 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 patterns, while the inter-individual variability in these measures is smaller in highly-trained cohorts and therefore less likely to confound results (212).

4.5.1 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 (213). 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 (178, 214). 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.

4.5.2 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 (215). We also note that the two variants nominally associated with the swim time trait are involved in hypoxic events characteristic of swimming (216). 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 alpha (HIF1α). This would cause reduced oxidative metabolic processing and therefore lead to the increase in swim time that is associated with the variant ($\beta = 0.2459$ hours). Additionally, the CKMM protein has been shown to exhibit protective effects during mild hypoxia and therefore we 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 (217).

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.

4.6 SUMMARY

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 (218).

Chapter 5: Mitochondrial variants associated with exercise response

This chapter will define the methodology, techniques, and analysis underpinning the mitochondrial sequencing project. The first section of this chapter (**5.1**), will summarise the project in the form of an abstract. The background and basis for this project will be outlined in Section **5.2**. Section **5.3** will outline the development and validation of project specific methods and analysis. Section **5.4** will discuss the results of the quality control following the development of our methodology, and Section **5.5** will outline the results of the mitochondrial association. The discussion section (**5.6**) will describe in greater detail the results from the QC and association. The conclusion of this chapter will discuss the contribution of the mitochondrial genome-sequencing project to the thesis (**5.7**).

The work within this chapter contributed to published manuscripts as outlined below:

- **N.R. Harvey [*], C.L. Albury[*],** M.C. Benton, D.A. Eccles, S. Stuart, J. Connell, H.G. Sutherland, R.J.N. Allcock, R.A. Lea, L.M. Haupt and L.R. Griffiths. (2019) Ion torrent high throughput mitochondrial genome sequencing (HTMGS). ***** Co-first authors. *Published PloS One, Nov 15, 2019, DOI: 10.1371/journal.pone.0224847*

Author Contribution:

NR Harvey was involved with conceptualisation of the above manuscript, performed analysis, performed preliminary lab work, validated the protocol, wrote the original draft, and drafted the final version. CL Albury contributed equally to this manuscript, performed the lab component, and drafted the final version of the manuscript. MC Benton, DA Eccles, and RA Lea contributed to the analysis pipeline. S Stuart and J Connell assisted with the lab component. HG Sutherland, RJN Allcock, LM Haupt, and LR Griffiths drafted the final manuscript.

- **Harvey NR**, Voisin S, Lea RA[·], Yan X, Benton MC, Papadimitriou ID, Jacques M, Haupt LM., Ashton KJ, Eynon N[#] and Griffiths LR[#], (2020) Investigating the influence of mtDNA and nuclear encoded mitochondrial variants on high intensity interval training outcomes. [#] Co-last authors. *Published, Scientific Reports, May 2020, DOI: 10.1038/s41598-020-67870-1*

Author Contribution:

Harvey NR was involved in conceptualisation of the manuscript, performed the lab work, performed the analysis, drafted the original version, and drafted the final version. Voisin S, Yan X, Papadimitriou I, Jacques M, and N Eynon were involved in participant recruitment for the original Gene SMART study. Lea RA and Benton MC assisted with analysis and drafted the final version. Haupt LM, Ashton KJ, and Griffiths LR were involved in conceptualisation and drafted the final version.

The work within this chapter summarised the work performed as per the second aim of this thesis as described in detail below:

Identify mtDNA haplogroups and mitochondrial variants that may affect exercise response.

a. Adapt and optimise a mitochondrial genome sequencing protocol to sequence the Gene SMART participants

b. Identify mtDNA haplogroups that are associated with exercise response

c. Identify mtDNA point variants that are associated with exercise response

d. Identify nuclear encoded mitochondrial variants that are associated with exercise response

5.1 ABSTRACT

Mitochondria supply intracellular energy requirements during exercise. Specific mitochondrial haplogroups and mitochondrial genetic variants have been associated with athletic performance, and exercise responses. However, these associations were discovered using underpowered, candidate gene approaches, and consequently have not been replicated. Here, we used whole-mitochondrial genome sequencing, in conjunction with high-throughput genotyping arrays, to discover novel genetic variants associated with exercise responses in the Gene SMART (Skeletal Muscle Adaptive Response to Training) cohort (n=62 completed). We performed a Principal Component Analysis of cohort aerobic fitness measures to build composite traits and test for variants associated with exercise outcomes. None of the mitochondrial genetic variants but nine nuclear encoded variants in eight separate genes were found to be associated with exercise responses (FDR<0.05) *(rs11061368: DIABLO, rs113400963: FAM185A, rs6062129 and rs6121949: MTG2, rs7231304: AFG3L2, rs2041840: NDUFAF7, rs7085433: TIMM23, rs1063271: SPTLC2, rs2275273: ALDH18A1).* Additionally, we outline potential mechanisms by which these variants may be contributing to exercise phenotypes. Our data suggest novel nuclear-encoded SNPs and mitochondrial pathways associated with exercise response phenotypes. Future studies should focus on validating these variants across different cohorts and ethnicities.

5.2 BACKGROUND

Response to exercise training depend on the type of exercise stimulus and varies considerably between individuals (154, 219, 220). This variability is tissue-specific and may be explained by a combination of genetic variants, epigenetic signatures, other molecular and lifestyle factors (221, 222). Mitochondria are the key mediators of intracellular energy and are involved in many essential cell metabolism and homeostasis processes (223) with exercise training improving mitochondrial function and content (223-226).

The mitochondrial genome encodes 37 genes that are highly conserved but differ slightly amongst different regional isolates (haplogroups) (227). Mitochondrial haplogroups and SNPs, in conjunction with SNPs in mitochondrial-related genes (nuclear encoded mitochondrial proteins: NEMPs) have previously been associated with athletic performance in highly trained populations and response to exercise training in the general population (228). While these studies have advanced our understanding, they have primarily utilised targeted genotyping technology such as candidate gene approaches, or Sanger sequencing to investigate specific mitochondrial coding regions and NEMPs, such as *NRF2* and *PGC1α* (229-232). Many of these studies also lacked robust technical measures on aerobic fitness measures (226). As such, many of the identified variants have not been replicated, and exercise-related genetic variants remain unknown (233). To date, studies assessing mitochondrial DNA (mtDNA) variants and NEMPs pertaining to exercise training have focused on protein-coding variants, with no studies looking at the more subtle effects of synonymous and non-coding changes (228, 234-237). Further, these studies have often based haplogroup analyses on sequencing or genotyping of the mitochondrial hypervariable region(s) (~500-1,000bp), with no consideration for the remaining mitochondrial genome (~15,000bp) and the specific haplogroup of exercise participants. For instance, 3`UTR (untranslated regions) variants that do not directly affect protein function may however affect translation, mRNA shuttling to specific organelles, or epigenetic modification such as microRNA silencing (238). Intronic variants may also lead to splice site changes directly contributing altered protein structure and function (239). As next generation sequencing (NGS) has become more widely available and affordable, sequencing of the whole mitochondrial genome (16,569 bp) is now feasible to uncover genetic variants associated with physical fitness phenotypes. When used in combination with SNP genotyping arrays, it is possible to examine, not only the 37 mitochondrially-encoded genes, but variants within all nuclear NEMP genes simultaneously.

Genetics may influence exercise response in conjunction with environmental factors such as diet, repeated exercise bouts, and age. Whilst these are modifiable, it is difficult to gauge the contribution of these factors to exercise response within short term exercise studies. Further, the additive effects of genetic variants to exercise response are not well understood as only a few genetic variants have been consistently replicated in the field.

Therefore, the aim of the present study was to examine the association between genetic variants (i.e. mitochondrial variants and NEMPs), and aerobic fitness measures in the well-characterised Gene SMART cohort. We hypothesise that by utilising whole-mitochondrial sequencing, we will uncover novel genetic variants associated with exercise responses.

5.3 METHODS

5.3.1 Participants

At the time of analysis, 77 participants had taken part in the study, 62 of whom successfully completed four weeks of High-Intensity Interval Training (HIIT) intervention protocol in the Gene SMART (Skeletal Muscle Adaptive Response to Training) study (152) at Victoria University, Australia. Ethical clearance for this study was provided by the Human Research Ethics Committee at Victoria University (Approval Number: HRE13-233), and the clearance was transferred to and also provided by the QUT Human Research Ethics Committee (Approval Number: 1600000342). We analysed the 62 participants who did not drop out of the study and all had healthy BMI and were moderately trained with an age range of (31.33 ± 7.94 years).

The Gene SMART study design has been previously reported (152). Briefly, participants were required to provide medical clearance to satisfy the inclusion criteria. Following familiarisation, baseline exercise performance was determined on a cycle ergometer during a 20 km time trial (TT), and two graded exercise tests (GXTs); these tests were administered a few days apart and no more than two weeks apart to limit temporal variability in performance.

5.3.2 Long Range PCR

Initial amplification of targeted mtDNA is most commonly performed using multiple sets of primers. It is generally accepted that shorter products amplify more efficiently than longer products (240), such as in the PrecisionID mitochondrial genome panel supplied by ThermoFisher Scientific. Long range PCR remains the most effective way of limiting the introduction of PCR bias and nuclear contamination into the later sequencing steps. Using two long range PCRs with two overlapping primer sets we generated large products (9,250bp and 8,985bp) suitable for uniform fragmentation. (Fragment 1 F: 5'-AACCAAACCCCAAAGACACC-3'; Fragment 1 R: 5'-GCCAATAATGACGTGAAGTCC-3'; and Fragment 2 F: 5'-TCCCACTCCTAAACACATCC-3'; Fragment 2 R: 5'-TTTATGGGGTGATGTGAGCC-3'). In addition, this also introduced time and cost advantages over multiple fragment amplification and significantly reduced the inherent complications of nuclear encoded mitochondrial pseudogene amplification (241). The GoTaq Long PCR Master Mix (Promega, Madison, WI) was used to complete this step.

Thermocycling conditions included denaturing: 94.5ºC for 2 minutes; cycling: 92 ºC for 20 seconds, 60 ºC for 20 seconds, 68 ºC for 9 minutes; and the final elongation stage: 72 ºC for 10 minutes. Fragment 2 amplified more efficiently than fragment 1 with the number of PCR cycles reduced from 35 to 28 for the former to obtain equivalent yields. Interestingly, we found that a low DNA input of 20ng in a 40µl reaction volume significantly enhanced overall product yield compared with the kit-recommended gDNA input of 0.1-0.5µg in a reaction volume of 50µl (242). Reaction components included 20ng Input genomic DNA, 20µl GoTaq Long Master Mix, 200pM of each primer, dH_2O up to the required 40µl reaction volume.

PCR products for both fragments were visualised on 1.0% agarose gels (60V, 60 minutes) with a 1kb ladder (New England Biolabs, Ipswich, MA). An example gel of the long-range PCR fragments is shown in **Figure 5-1**.

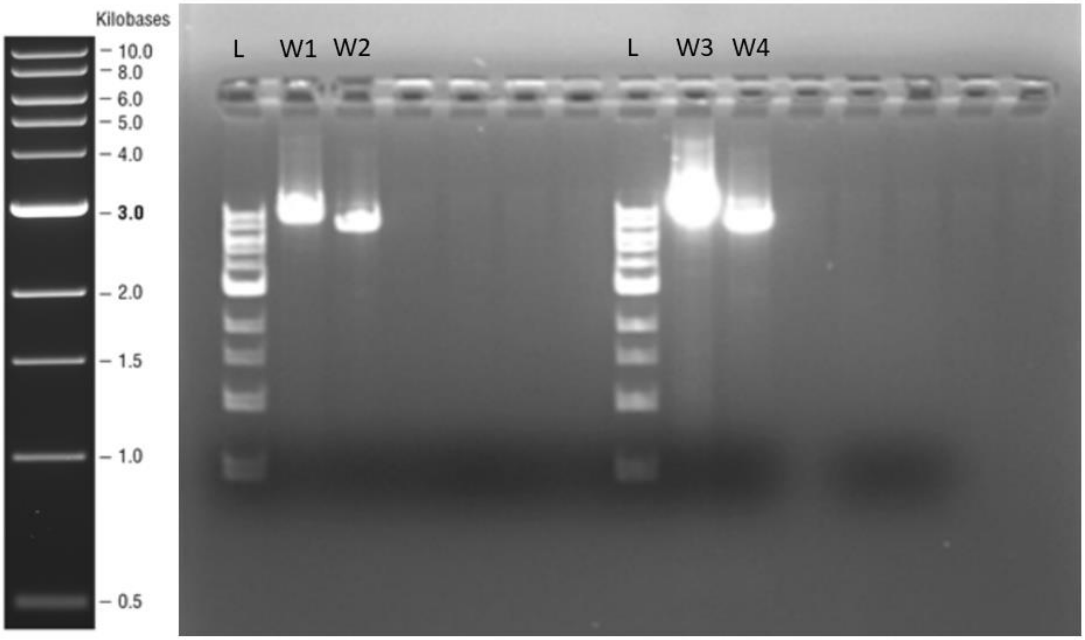


Figure 5-1: *1% Agarose gel showing long range PCR fragments of the mitochondrial genome. Ladder lanes contain a 1Kb ladder. Lanes W1 and W2 contain 1µl of PCR product from fragment 1 and fragment 2 of the mitochondrial genome for one individual. Lanes W3 and W4 contain 3µl of the same PCR products as W1 and W2.*

Fragments were then purified with QIAquick post PCR clean-up columns from (QIAGEN, Hilden, Germany) and quantified using Agilent DNA 12000 chips on the

Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA). An Example of the Bioanalyser trace is shown below in **Figure 5-2.**

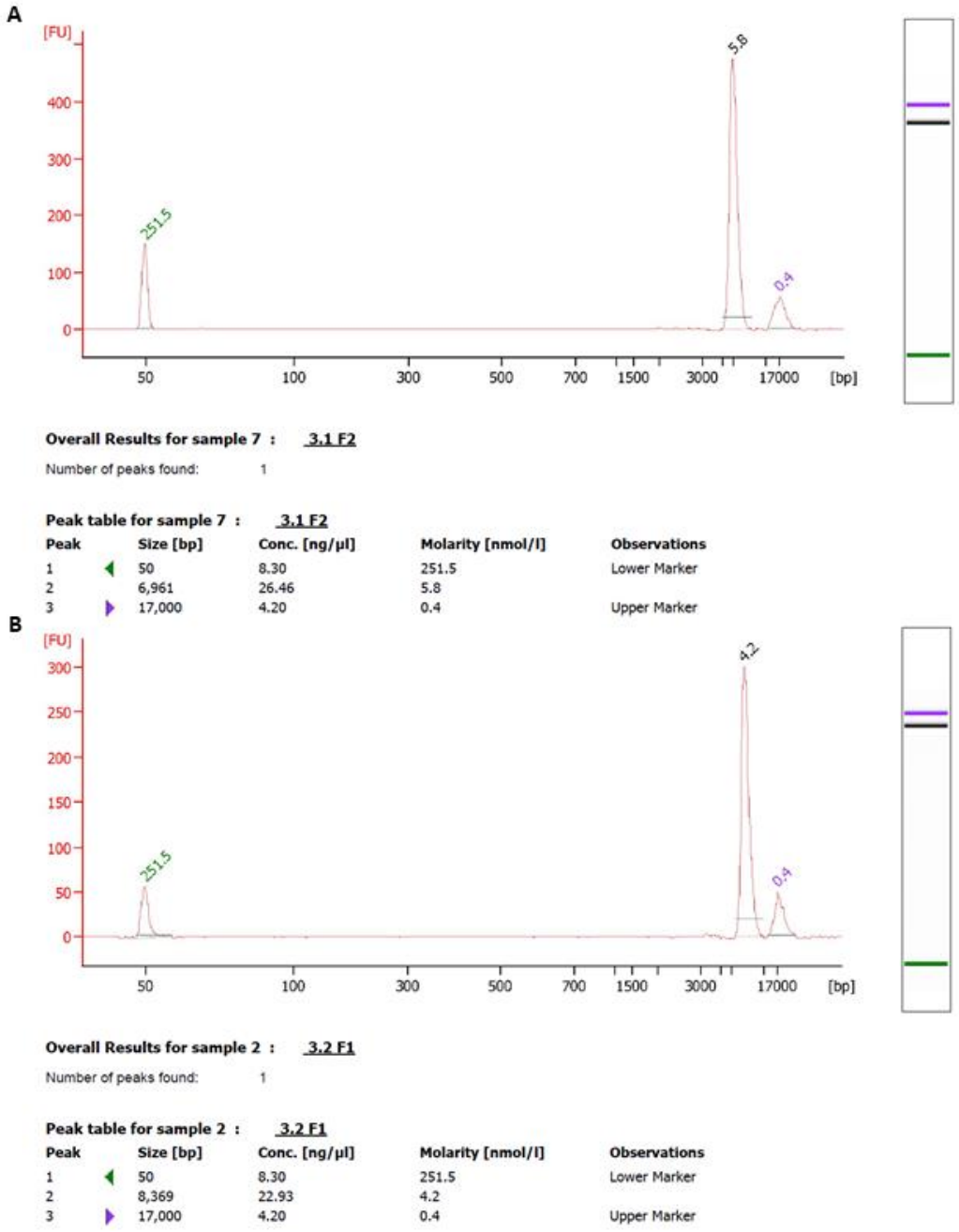


Figure 5-2: *Bioanalyser spectrographs of the mitochondrial long range fragment 1 (**A**) and 2 (**B**) for the same exercise participant. X-axis shows size of the DNA in base pairs (bp) and y-axis shows the Fluorescence units for each peak detected. Sizes of each long-range fragment peak are detailed in the panels immediately under each spectrograph between the lower and upper markers.*

The highest concentrated fragment of each corresponding sample was diluted to match the other and both F1 and F2 were pooled at equimolar concentrations to a final amount of 100ng in 51µl to ensure evenly distributed amplicons for subsequent library preparation methods. This ensured optimised control and therefore even mtDNA genome coverage.

5.3.3 Physical Fragmentation and End-Repair Methods

Following sequence amplification, fragmentation may be performed through enzymatic digestion or through the physical shearing of DNA using acoustic shearing, hydrodynamic shearing, or sonication methods. Some evidence has demonstrated enzymatic fragmentation using restriction endonucleases or transposases are consistent yet futile when discerning bias and detecting insertions and deletions (INDELS) (243). To add to ease and cost effectiveness of the protocol, we utilised physical shearing with sonication using the Bioruptor (Diagenode, London) system due to its selective design for small sample volumes and effectiveness in maintaining sample integrity. Considering our desired fragmentation profile (50 – 500bp, peak at 200bp), 200ng of pooled mtDNA in a final volume of 51µl was sheared using the following parameters: 5x (15 cycles of 30secs high intensity sonication followed by 30secs off). All 200ng of fragmented mtDNA was end repaired using the NEB end repair module kit (New England Biolabs, Ipswich, MA) according to the manufacturer's instructions.

5.3.4 Barcoding and size selecting methods

NEXTflex DNA Barcodes-Ion Torrent compatible (Bioo Scientific, Austin, TX), were selected as a cost-effective alternative to standard IonXpress barcodes (ThermoFisher Scientific, Waltham, MA). Barcode ligation was achieved with the T4 DNA ligase enzyme component of the NEBNext fast DNA library prep set for Ion Torrent (New England Biolabs, Ipswich, MA). Barcoded products were then subjected to a purification process using 180µl Agencourt AMPure XP mix (Beckman Coulter, Brea, CA) and subsequent 80% (v/v) Ethanol wash. Following this, the barcoded mtDNA fragments were examined on an E-gel system (ThermoFisher Scientific, Waltham, MA) utilising an iBase unit to visualise the running gel. This allowed us to identify and recover appropriately sized fragments between 250-350 base pairs. Libraries were size selected at a target peak of 330bp thus maximising efficiency under the limitations of 200bp sequencing chemistry. It should be noted here that as each barcode is ~40bp in length, we must size select for 200bp + (40bp X 2) for each end of the read.

5.3.5 Library amplification of mtDNA

Size selected amplicons underwent final library amplification using the NEBNext Fast DNA Library Prep Set for Ion Torrent (New England Biolabs, Ipswich, MA). We found that 100ng of input mtDNA combined with the manufacturer defined protocol aided in high quality library generation. Amplified libraries were again subjected to a purification process using 140µl Agencourt AMPure XP magnetic beads and subsequent 80% (v/v) Ethanol wash. Purified products were quantified using Bioanalyzer system with Agilent DNA 1000 chips (Agilent Technologies, Santa Clara, CA)

5.3.6 Template Preparation and Semiconductor Sequencing

For library preparation, all 77 mitochondrial libraries were diluted to achieve a final equimolar concentration of 26pM. To achieve maximum efficiency, we pooled 38 and 39 mitochondrial libraries for each chip respectively. The libraries then underwent template preparation using the Ion Hi-Q View Chef Kit (ThermoFisher Scientific, Waltham, MA) on the automated Ion chef system (ThermoFisher Scientific, Waltham, MA). The mitochondrial libraries were loaded onto Ion 316 V2 BC chips (ThermoFisher Scientific, Waltham, MA). We ran two chips at once to prevent wasted reagents from the Ion Chef and sequencing kits. Semiconductor sequencing was performed using the Ion PGM Hi-Q view Sequencing 200 Kit (ThermoFisher Scientific, Waltham, MA) on the Ion Torrent PGM platform (ThermoFisher Scientific, Waltham, MA). All protocols followed standard manufacturer's instructions. Following successful amplification and sequencing of mitochondrial libraries for NGS, read trimming, base calling, and mapping (Revised Cambridge Reference Sequence [rCRS]) was completed using the Ion Torrent Suite platform (ThermoFisher Scientific, Waltham, MA USA).

5.3.7 Data filtering

A bioinformatics pipeline (*SAMtools, BCFtools*) was utilised to generate variant call files (VCF) for all samples as described previously (244). VCF files were then aligned to the *revised Cambridge Reference Sequence* (rCRS) and all sequences were stringently left aligned back to this reference genome to account for the single end (SE) reads generated from Ion Torrent sequence information. FASTA files were generated for all samples and then merged VCF and

FASTA files were produced for the entire data set. The merged FASTA files were annexed using Mitomaster, a mitochondrial sequence database, to call haplogroups and obtain variant annotation information for all samples (245, 246). The merged VCF file was converted to PLINK (v1.90p) format using the function '*--make-bed'* for further association analysis. A detailed description of the analysis pipeline may be found in markdown format in the GRC computational genetics GitHub account (https://github.com/GRC-CompGen/mitochondrial_seq_pipeline), including all necessary files (NEMP locations BED file) and scripts (mitochondrial Solarplot R script) to replicate our analyses within other data sets.

The *ped* file generated from Illumina GenomeStudio v2.0 software was converted into binary format. We did not impute any genotypes to prevent false positive associations and a larger multiple testing burden. There were 551,839 typed SNPs; subsequent SNP and individual filtering and trimming was based on **1)** SNPs with > 20% missing data (239 removed), **2)** individuals with > 20% missing data (0 removed), **3)** minor allele frequency < 0.01 to remove rare variant associations (260,269 removed), **4)** SNPs out of Hardy Weinberg equilibrium for quantitative traits (58 removed due to $P<1e^{-6}$) (247). All samples passed kinship and heterozygosity thresholds after the filtering outlined above, leaving 62 samples and 291,273 SNPs to analyse. A BED file containing the genomic locations (GRCh37) of all known NEMPs was obtained from the Broad Institute's human MitoCarta2.0 website (248-251). PLINK was used to extract the SNPs within the genomic locations from the Omni Express SNP chip data of the same participants. In total, 4,806 SNPs were within the NEMP genomic regions detailed by the Broad Institute MitoCarta2.0 bed file and considered to be mitochondrially related variants.

5.3.8 Exercise response phenotypes

Participant stratification into high and low response groups lead to a loss of statistical power in association testing. As such, and to avoid classifying responders and non-responders via arbitrary thresholds, we chose to keep the phenotypes as continuous variables for association testing (166).

To ascertain variants that were associated with exercise response for key physiological traits, we utilised the delta (Δ) change (Post phenotype – Pre phenotype) quantitative trait data for; peak power output (ΔWpeak in Watts); power at lactate threshold (ΔLT in Watts); peak oxygen

uptake (ΔVO_{2peak} in mL/min/kg body weight); and time to completion measurement for a 20 km time trial (ΔTT in seconds). As the quantitative traits were all continuous and to keep maximal statistical power, we did not use arbitrary response thresholds. With multiple, correlated response phenotypes, we conducted a Principal Component Analysis (PCA) of the response phenotypes using the R package *FactoMineR (159).* PCA is a dimensionality reduction method that computes linear combinations of the multiple response phenotypes into principal components (PCs) so that the variance between individuals is maximised. Every individual is then represented by one value for each PC, considered a composite trait of the different response phenotypes. A more detailed description of PCA for composite trait association testing was discussed in **Chapter 3, Section 3.3.2**.

Missing phenotypic values were excluded from the phenotype table prior to PCA to prevent skewing of data and to maintain appropriate PCs. Following the PCA, these variables were set as "missing" for the association analysis. We also tested the individual response phenotypes and compared the significance levels of variants between the composite traits with those within each PC. This resulted in 4 PCs that cumulatively explained > 90% of the variance between participants.

5.3.9 Statistical analysis

Analysis of the response traits was performed in SPSS (version 17.0) using a paired samples t-test. SPSS was also used to test associations between mitochondrial haplogroups and exercise response with a Wald test. Analyses for the mitochondrial SNPs and NEMP SNPs were kept separate for analysis using different association models. We used PLINK V1.90p to perform quantitative linear association tests (95% CI) with both dominant and recessive models. An additive model was also attempted but yielded the same results as our dominant model. We adjusted all association results for age and effect sizes were determined using raw beta regression coefficient values (i.e. genotype X is associated with β [unit specific to trait of interest] changes in the phenotype). Variants that passed a nominal P value threshold of $P<0.05$ were considered for further analysis whereas variants that passed multiple testing adjustment using the Benjamini-Hochberg False Discovery Rate ($FDR<0.05$) method were considered significant associations. We performed adjustment for multiple testing for each phenotype separately. Performance of an a*priori* power calculation for this study indicated that the linear modelling approach with additive genotypic effects for our sample size (n=62) was sufficient for at least 80% power to detect SNP-based heritability of 13% or more at the relaxed alpha

level of 0.05. SNP-based heritability estimates were approximated by genotype Vs outcome R^2 values from linear regressions. We also note that the Gene SMART cohort is a tightly controlled study with rigorous physiological measures, all performed in duplicate, which by itself significantly increase the power of the detected a*priori*.

All variants from the association tests were plotted in R using the *tidyverse, ggplot2,* and *qqman* packages. Locus zoom plots were generated with the online locus zoom software (v0.12) using a compilation of the dominant and recessive nominal results from our association tests (252). As the participants were all Caucasian, SNP linkage r^2 values were calculated with the HapMap CEU database. As intronic SNPs may affect genes far away, we termed genes within 200Kb of the SNP of interest as "proximal" regardless of NEMP status.

UCSC genome browser (hg19/GRCh37) was utilised to ascertain the genomic effect and therefore consequence of all statistically significant variants. The GeneHancer track was utilised to determine the regulatory element affect for each variant (253). We chose to postulate molecular mechanisms for variants that were purely intronic and did not show affinity for epigenetic or transcription factor binding. However, these should be interpreted with caution and therefore we note that the variants in this category were found to be association based only and should be confirmed through replication analysis in larger exercise cohorts.

5.4 RESULTS: QUALITY CONTROL

Output sequencing produced approximately $3x10^6$ of 124bp-single-end reads per chip. The mean sequencing depth across all 77 samples was 615X. Each sample was identified to have 99.97% bases mapped to the reference sequence covered by approximately 81K reads. Three samples out of the total 77 did not amplify on the beads during template preparation and so were excluded from further comparisons. The average reads per sample was 80,155, with a standard deviation of 52,183. 73% (n=54/74) of all samples on the chip had total mapped reads above 50K (384X).

Direct comparisons were made between the overall sequencing qualities of our modified protocol with the suggested sequencing quality output from Ion Torrent. Chip loading and enrichment measurements (**Table 5-1**) yielded similar results for our protocol compared with what was expected from a commercial method. However, the percentage of polyclonality and low-quality reads were under the recommended thresholds for a good quality run. This in turn resulted in a high percentage of usable reads (79.0%), high percentage of mapped usable reads (99.7%) and higher quality base calling (>AQ17 = 374Mb, and 331Mb) in our protocol.

Table 5-1: *Quality metric comparison between suggested sequencing output metrics and HTMGS sequencing chips*

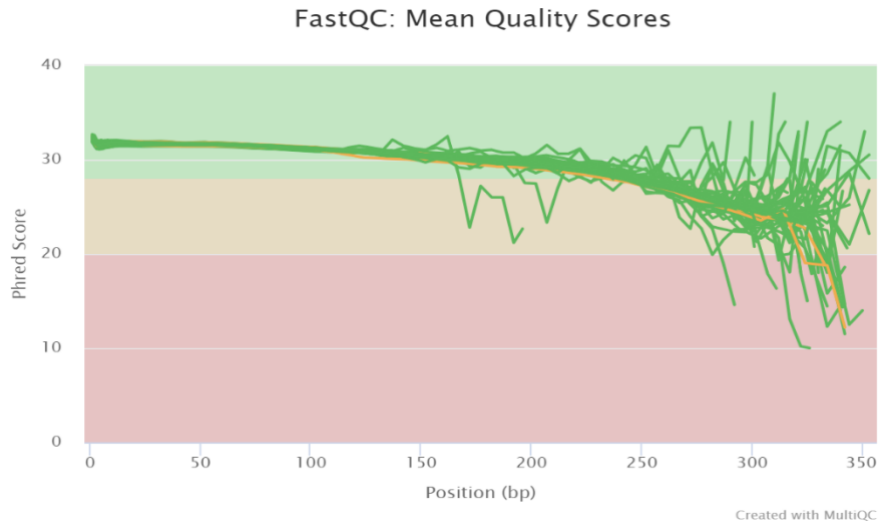
	Ion Torrent suggested quality output	**HTMGS Chip 1 (n=38)**	**Chip 2 (n=39)**
Chip loading %	>60	65	61
Enrichment %	100	100	100
Polyclonal %	20-30	18	19
Low quality %	<5	3	4
Primer Dimer %	<5	1	5
Usable reads (%)	2-3M	3,182,960 (79.0%)	2,839,445 (73%)
Mapped usable reads %	>90	99.7	99.3
Mean Read length bp	<200	126	123
Total bases (>AQ17)	NA*	374Mb	331Mb

**This value is dependent on the overall sequencing quality for the rest of the measurements*

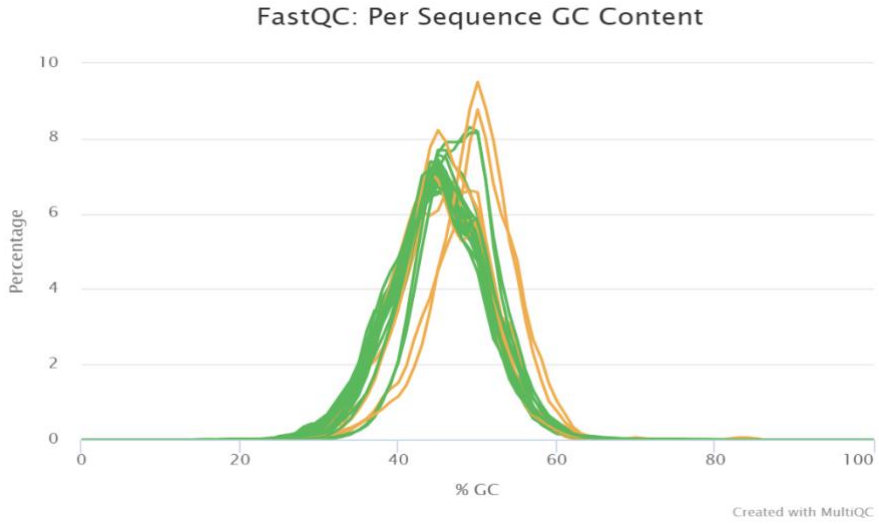
An important consideration is the number of samples loaded onto a single sequencing Ion chip which influences the number of total reads per sample. In addition, sample number per chip should not limit the number of high-quality bases (Mb >AQ17), the total number of reads per chip, the level of polyclonality or the degree (%) of low-quality reads generated. All samples were subjected to quality control analysis using the multiQC function in the Galaxy web platform. Phred scores remained consistent at >25 for all samples at the median read length (125bp) and per sequence GC content followed a normal distribution as expected for NGS data. Sequence length distribution was inconsistent which was expected given the size selection method utilised for this study. (**Figure 5-3**)

A density versus median coverage plot was generated using the R packages *tidyverse* and *ggplot2* (**Figure 5-4**). As can be viewed in **Figure 5-4**, the high throughput HTMGS method for both chips 1 and 2 reached a high median depth of ~500X. The HTMGS method appeared to have a significant fraction of the mitochondrial genome with sequence of high quality at >100X sequencing depth.

A



B



C

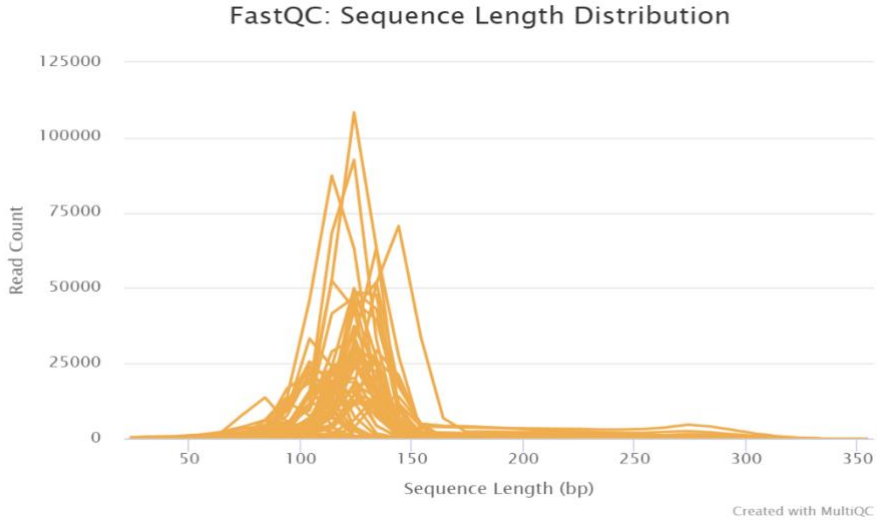


Figure 5-3: *MultiQC summary of all 38 multiplexed samples showing* ***A)*** *Phred quality scores (Y-axis) per base in read (X-axis****),*** ***B)*** *sequence GC content (X-axis) following a normal distribution,* ***C)*** *read length in base pairs (X-axis) versus the read count (Y-axis)*

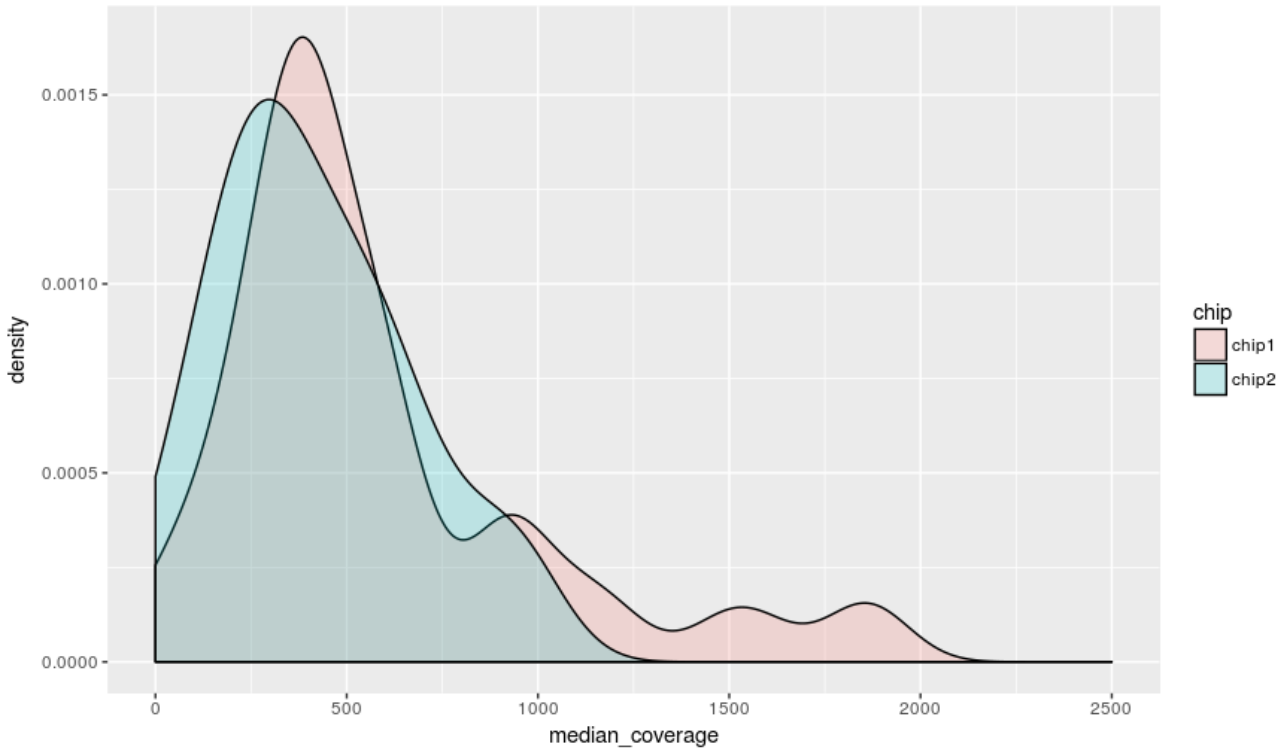


Figure 5-4: *The HTMGS median coverage estimates. The x-axis illustrates the level of median coverage reached at a genomic position fraction reaching coverage x at a given point (y-axis, density).*

5.5 RESULTS: ASSOCIATION

From the mtDNA sequencing, an average depth coverage of 615X was obtained over the mitochondrial genome. Following sample annotation with Mitomaster, 60 distinct haplogroups were identified within the Gene SMART completed cohort of 62 participants. Association testing maintains the assumption that genetic differences may be identified between groups of individuals. Due to the distinct haplogroups observed, there were no statistically significant associations between individual mitochondrial haplogroups with exercise response traits. When reduced to broad mitochondrial haplogroups, Wald testing was unable to discern significant associations with mitochondrial haplogroups and exercise response phenotypes. A summary table of the mitochondrial haplogroups found within the Gene SMART participants is shown in **Table 5-2**. The confidence scores (0-1) represent the number of mtDNA variants found in each participant that belong to their respective haplogroup.

***Table 5-2**: Summary of mitochondrial Haplogroups within the Gene SMART study.(160)*

Participant ID	MtDNA Haplogroup	Confidence	Participant ID	MtDNA Haplogroup	Confidence
SG100	H1c2a	0.9505	SG140	H1c7	0.9581
SG102	C1b10	0.9305	SG141	H2a2b3	0.9386
SG103	K1a1b2b	0.9648	SG142	H+152	0.8534
SG104	H6a1b2	0.9438	SG143	U4a1a	1
SG105	H3g	0.9353	SG144	T2b4+152	0.9535
SG106	H94	0.8164	SG145	H24a	1
SG107	K1a1b1a	0.968	SG146	U5b3e	0.9818
SG108	J1c3g	0.9366	SG147	U5a1a1	1
SG109	W3a1c	0.9804	SG148	I1a1e	0.9762
SG110	H1e1a3	0.9486	SG149	H6a1a3	0.9958
SG111	H1t	0.9336	SG150	HV	0.7231
SG112	K1a4f1	0.9641	SG151	U5a2b4	0.9481
SG113	T2b+152	0.9795	SG152	J1c2f	0.9805
SG114	U5b1b1+@16192	0.9924	SG153	K1a4a1	0.9783
SG115	T2b13a	0.9827	SG154	U2e1a1	0.94
SG116	J1c3g	0.9639	SG155	H1a1	0.9505
SG117	H10	0.9356	SG156	H1a	0.9898
SG118	H16b	1	SG157	H3u1	0.8918
SG119	U3a1c1	0.9499	SG158	H1e1a2	0.9243
SG120	T2b1	0.9904	SG159	U4b1a2	0.9924
SG121	H15a1a1	0.9175	SG160	U8a1a	0.9319
SG122	K1a	0.9508	SG161	K1a	0.9204
SG123	K1a4a1a+195	0.9941	SG162	U4b1a2	0.9924
SG124	H3	0.9852	SG163	H4a1a2a	0.9818
SG125	L0d2a1a	0.9839	SG164	H2a2b4	0.9037
SG126	H5a1	1	SG165	T2f1a1	0.9306
SG127	H2b	0.8848	SG166	H1a1	1
SG128	H1	0.8676	SG167	U5a1b1d+16093	0.9791
SG129	H24a2	0.9202	SG168*	H2a2a1	0.5
SG130	J2a1a1	0.9726	SG169	T2b	0.9918
SG131	U8b1a1	0.9258	SG170	J1b1a1a	0.9857
SG132	V10a	0.9673	SG171	H6a1b3	0.985
SG133	HV1a1a	0.9296	SG172	I2	0.9222
SG134	J1c7a	0.9841	SG173	I2c	0.9577
SG135	R1a1a2	0.9875	SG174	M1a	0.905
SG136	HV6a	0.951	SG175	W5	0.9513
SG137	H2a1e1a	0.9591	SG176	T2f1a1	0.8887
SG138	H1b1+16362	1	SG177	K1a16	0.9932
SG139	J2b1a2a	0.9655			

***Sample SG168 contained sequence identical to the rCRS reference genome and therefore stratification into mtDNA haplogroup was not based on genetic variation but sequence homology to the reference.**

Following quality control, 170 mitochondrial and 4,124 NEMP genetic variants were included in association testing. A cumulative total of 4,325 NEMP variants and 28 mitochondrial variants passed the nominal threshold of significance ($P_{unadjusted}$ <0.05) for all tests. A solar plot showing the clustering of mitochondrial genomic variants for each trait is shown in **Figure 5-5** (254)**.** The exonic variants passing the nominal threshold from the mitochondrial association results are summarised in **Table 5-3**.

17 variants passed the nominal significance threshold of $\alpha<0.05$ in various delta traits and principle components. Of these, 8 were located within the hypervariable control region and therefore discounted from further analyses. A further 2 were located within an rRNA gene, 1 within the *$tRNA^{Leu}$* gene, 1 within the *mitochondrially encoded ATP synthase membrane subunit 6* (*ATP6)* gene, 2 within the *mitochondrially encoded NADH: ubiquinone oxidoreductase core subunit 4* (*ND4)*, 2 in *mitochondrially encoded NADH: ubiquinone oxidoreductase core subunit 5* (*ND5)* and 1 in mitochondrially encoded cytochrome B (*CYB)*. An amalgamation of both the dominant and recessive models for all quantitative variables was used to create a Manhattan plot (**Figure 5-6**).

28 variants passed the nominal significance threshold of $P_{unadjusted}$ <0.05 in various delta traits and principal components. Of these, 8 were located within the hypervariable control region and therefore discounted from further analyses. A further 2 genetic variants were located within a mitochondrial rRNA gene, 1 within the *$tRNA^{Leu}$* gene, 1 within the *mitochondrially encoded ATP synthase membrane subunit 6* (*ATP6)* gene, 2 within the *mitochondrially encoded NADH: ubiquinone oxidoreductase core subunit 4* (*ND4)*, 2 in *mitochondrially encoded NADH: ubiquinone oxidoreductase core subunit 5* (*ND5)* and 1 in mitochondrially encoded cytochrome B (*CYB)*. None of the mitochondrial genomic variants were associated with composite response traits or individual response traits at FDR < 0.05. A Manhattan plot of the NEMP variants is shown in **Figure 5-7**. A summary of the association statistics for the variants passing a nominal threshold of $P_{unadjusted}$ $<1e^{-4}$ in both the NEMP associations are shown in **Table 5-4**.

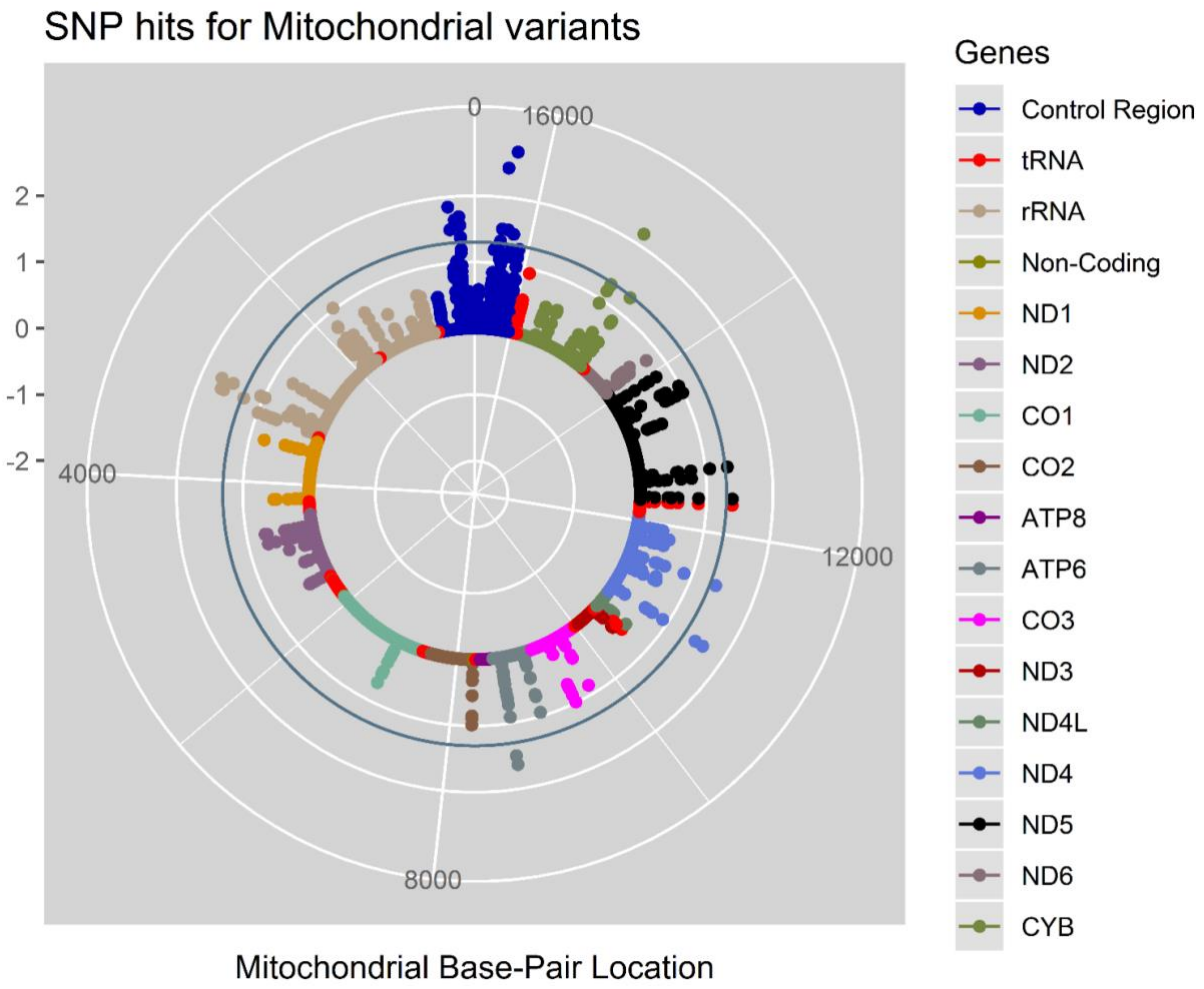


Figure 5-5: *Solar plot showing significant hits from mitochondrial association testing. Each dot represents a detected variant. The inner ring of the plot represents the mitochondrial genome and is coloured based on genomic region as summarised in the plot legend. The X-axis represents the mitochondrial base pair location. The Y-axis represents the significance level [-log10 (P-value)] in the Gene SMART population over multiple traits. The significance threshold was set at P<0.05 and is represented by the circular blue line. The concentric white rings surrounding the genome represent the P-value thresholds -log10 (0.01) and -log10 (0.001) respectively.(160)*

Table 5-3: *Exonic mitochondrial SNPs associated with phenotypic traits and PCs.* ***(160)***

Trait	CHR	SNP	Allele	Gene	Consequence	Model	GenBank MAF	MAF	SE (95% CI)	P-value*	FDR	Effect size (beta)
Δ-LT	MT	rs2000975	G	*ATP6*	Missense	ADD	0.019	0.032	11.24 (-48.2 - -4.2)	0.023	0.39	-26.19
	MT	rs2857284	C	*ND4*	Synonymous	ADD	0.021	0.032	11.24 (-48.2 - -4.2)	0.023	0.39	-26.19
	MT	rs193302956	T	*ND5*	Synonymous	ADD	0.12	0.081	7.25 (-29 - -0.5)	0.046	0.49	-14.75
	MT	rs193302985	A	*CYB*	Synonymous	ADD	0.044	0.113	6.08 (-29 - -5.2)	0.0067	0.28	-17.10
PC3	MT	rs2857284	C	*ND4*	Synonymous	ADD	0.021	0.032	0.72 (0.17 – 3.0)	0.032	0.54	1.57
	MT	rs2000975	G	*ATP6*	Missense	ADD	0.019	0.032	0.72 (0.17 – 3.0)	0.032	0.54	1.57
PC4	MT	rs2853493	G	*ND4*	Synonymous	ADD	0.28	0.258	0.25 (-1.0 - -0.03)	0.041	0.69	-0.53
	MT	rs2853498	G	$tRNA^{Leu}$	-	ADD	0.23	0.258	0.25 (-1.0 - -0.03)	0.041	0.69	-0.53
	MT	rs2853499	A	*ND5*	Synonymous	ADD	0.28	0.258	0.25 (-1.0 - -0.03)	0.041	0.69	-0.53

CHR: Chromosome #, SNP: Single Nucleotide Polymorphism, MAF: Minor Allele Frequency, SE: Standard Error, CI: Confidence Interval, FDR = False Discovery Rate, Δ: delta change, ADD: Additive model

**P-value adjusted for age*

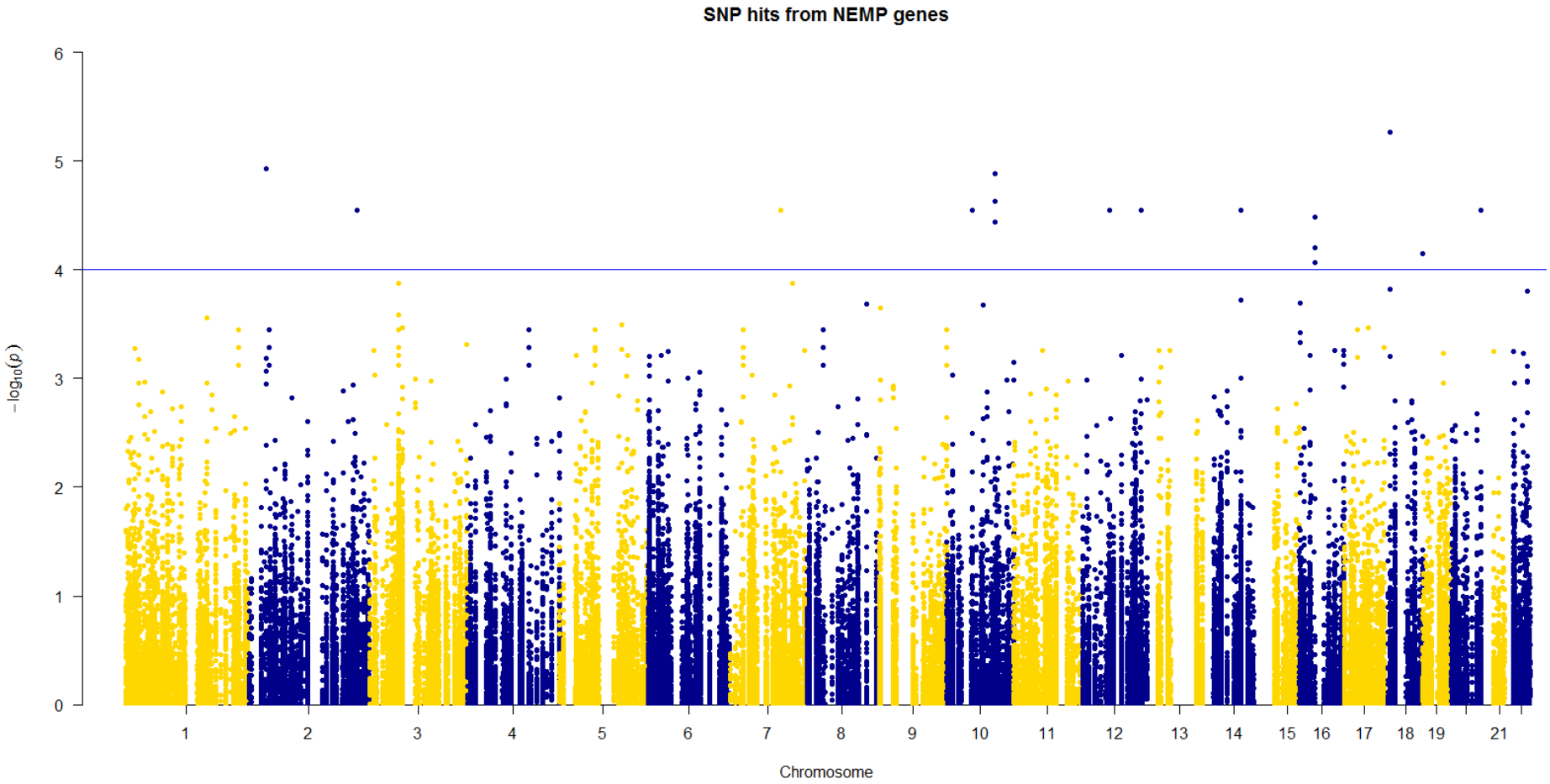


Figure 5-6: *Manhattan plot for all hits from all response phenotypes, biochemical measures, and PCs in the linear dominant and recessive association models. Suggestive significance was set at –log10($P_{unadjusted}$ = 0.0001, blue line). As all traits were included clusters of variants represent association across multiple traits rather than one significant locus commonly associated with GWAS.(160)*

Table 5-4: *Summary statistics for exonic variants in the nuclear encoded, mitochondria-related genes. (160)*

Trait	CHR	SNP	Response Allele	Gene	Consequence	Model	gnomAD MAF	MAF	H^2	Effect size (beta)	SE (95% CI)	P-value*	FDR
Δ-LT	18	rs12964779	A	*RBFA*	Intronic	DOM	0.48	0.49	0.136	-16.67	3.94 (-24.4 - -8.9)	8.25E-05	0.288
Δ-TT	2	rs41272687	A	*CYP27A1*	Missense	DOM	0.019	0.008	0.244	587.7	127.2 (338.5 - 837)	2.23E-05	**0.052**
	12	rs73338162	A	*SHMT2*	Missense	DOM	0.007	0.008	0.244	587.7	127.2 (338.5 - 837)	2.23E-05	**0.052**
	7	rs113400963	G	*FAM185A*	Intronic	REC	0.088	0.096	0.105	587.7	127.2 (338.5 - 837)	2.23E-05	**0.013**
	10	rs7085433	T	*TIMM23*	Noncoding transcript	REC	0.11	0.096	0.027	587.7	127.2 (338.5 - 837)	2.23E-05	**0.013**
	12	rs11061368	G	*DIABLO*	Intronic	REC	0.049	0.088	0.161	587.7	127.2(338.5 - 837)	2.23E-05	**0.013**
	14	rs1063271	C	*SPTLC2*	3`UTR	REC	0.15	0.16	0.069	587.7	127.2(338.5 - 837)	2.23E-05	**0.013**
	20	rs6062129	C	*MTG2*	Intronic	REC	0.33	0.29	0.164	587.7	127.2 (338.5 - 837)	2.23E-05	**0.013**
	20	rs6121949	G	*MTG2*	Intronic	REC	0.17	0.14	0.076	587.7	127.2 (338.5 - 837)	2.23E-05	**0.013**
Δ-VO_{2max}	2	rs2041840	T	*NDUFAF7*	Intronic	DOM	0.36	0.48	0.147	4.257	0.965 (-7.6 - -3.1)	4.52E-05	0.184
	7	rs322820	T	*SND1*	Intronic	REC	0.37	0.36	0.086	-5.346	1.168 (2.4 – 6.1)	2.54E-05	0.091
Δ-PP	2	rs2041840	T	*NDUFAF7*	Intronic	DOM	0.36	0.48	0.193	17.3	4.066 (9.3 – 25.3)	7.56E-05	0.173
PC2	2	rs2041840	T	*NDUFAF7*	Intronic	DOM	0.36	0.48	0.211	1.737	0.309 (1.1 – 2.3)	5.45E-07	**0.002**
	9	rs4742213	T	*GLDC*	Intronic	REC	0.46[#]	0.45	0.069	-1.471	0.3517 (-0.9 - -4.7)	9.73E-05	0.348
	18	rs7231304	C	*AFG3L2*	Intronic	DOM	0.11	0.14	0.231	-1.564	0.3298 (-0.8 - -4.2)	1.38E-05	**0.028**

CHR: Chromosome number, SNP: Single Nucleotide Polymorphism, MAF: Minor Allele Frequency, SE: Standard Error, CI: Confidence Interval, FDR = False Discovery Rate, Δ: delta change, DOM: Dominant model, REC: Recessive model, gnomAD MAF: publicly reported MAF values for the gnomAD-Genomes database (European)

[#] gnomAD MAF not reported, (1000 genomes European value used)

**P-value adjusted for age, All variants were assessed within the hg19/GRCh37 reference genome*

Six SNPs in 5 distinct genes were found to be associated with ΔTT and 2 SNPs in 2 distinct genes with PC2. The most significant variant was rs2041840 associated with PC2 and located within *NDUFAF7;* the rs12712528 variant was also found to be within *NDUFAF7* and had a moderate correlation with rs2041840 (R^2=0.5) **Figure 5-7e**. This variant was also trending towards significance in the Δ-Weight and Δ-VO_{2max} response phenotypes (**Table 5-4**). The T allele at rs2041840 was associated with a better response to exercise. The Locus Zoom plot (**Fig 5-7d**) surrounding the *MTG2* gene was also gene-rich with 11 proximal genes. The two associated variants (rs6062129 and rs6121949) were moderately correlated (R^2=0.5), however there were no SNPs found within the proximal genes. The locus zoom plot for the variants found within the *AFG3L2* gene (**Fig 5-7f**) was proximal to 6 genes within 200Kb. There was also a proximal SNP within the *SLMO1* gene however this was not in linkage with the variants identified within the *AFG3L2* gene.

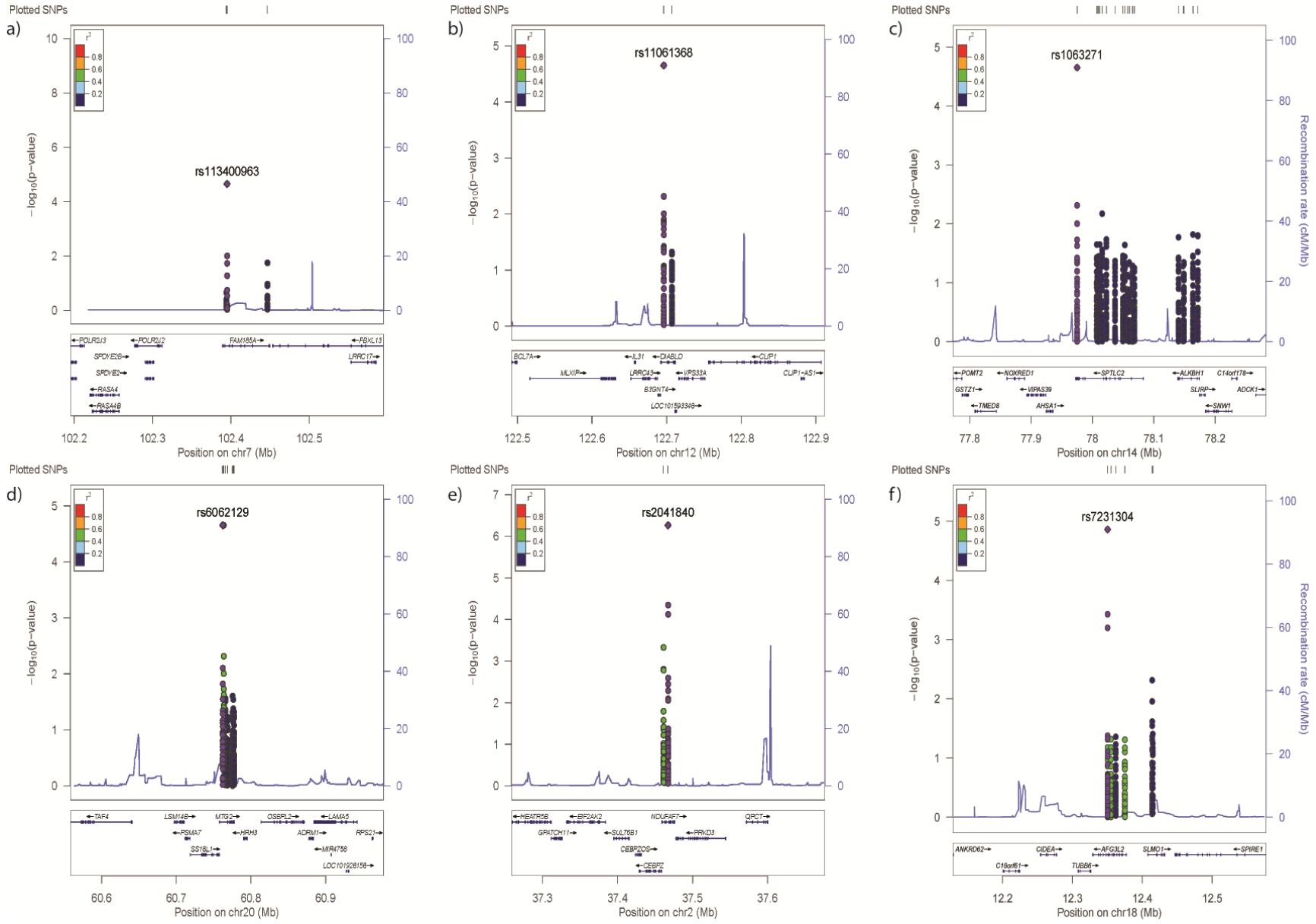

a)
rs113400963
Position on chr7 (Mb)
b)
rs11061368
Position on chr12 (Mb)
c)
rs1063271
Position on chr14 (Mb)
Recombination rate (cM/Mb)
d)
rs6062129
Position on chr20 (Mb)
e)
rs2041840
Position on chr2 (Mb)
f)
rs7231304
Position on chr18 (Mb)
Recombination rate (cM/Mb)
Plotted SNPs
−log10(p−value)

Figure 5-7: *Locus Zoom plots of significant intronic SNPs from the nuclear mitochondrial association testing. Each panel shows the locus surrounding* ***a)*** *rs6969054 variant within the FAM185A gene,* ***b)*** *rs11061368 variant within the DIABLO gene,* ***c)*** *rs1063271 variant within the SPTLC2 gene,* ***d)*** *rs6062129 variant within the MTG2 gene,* ***e)*** *rs2041840 variant within the NDUFAF7 gene, and* ***f)*** *the rs7231304 variant within the AFG3L2 gene. All panels show the variant of interest ±200Kb. Left y-axis shows –log10(p-value) of association results for all traits and right y-axis shows recombination rate across the locus in relation to the variant of interest. X-axis shows genomic position across the respective chromosomal regions. All variants were plotted according to the hg19/GRCh37 reference build and recombination rate was calculated from the default parameters within the locus zoom software. (160)*

5.6 DISCUSSION

In this study, state-of-the-art mitochondrial sequencing was utilised, along with high-throughput targeted genotyping of mitochondrial-related variants encoded by the nucleus (NEMPs) to discover novel genetic variants associated with responses to exercise. A total of 28 mitochondrial and 4,325 nuclear encoded mitochondrial associated variants passed the nominal significance thresholds for the various candidate gene association tests. While no mitochondrial variants were found to be significantly associated with exercise response, we uncovered nine variants in eight distinct genes found to be associated with exercise training response. The arguably largest contributors to exercise response (diet, age, repeated bouts) were carefully scrutinised in order to ascertain the genetic contribution to each phenotypic trait. Further, the genetic variants contributing to multiple phenotypic traits was assessed using composite traits built with PCA data reduction.

Novel exercise loci associated with training response

The most significant variant was associated with the composite exercise response phenotype and located within an intron of *NDUFAF7* (rs2041840). The T allele was associated with better exercise response as shown by the positive beta values. *NDUFAF7* encodes an arginine methyltransferase that is essential for mitochondrial complex I assembly (255). We have showed that this variant was in a gene rich region with 8 proximal genes (**Fig 3a**), indicating possible effects for this variant in any of the proximal genes or indeed for genes that may be further away from the loci. Specifically, the interaction between this variant and the *Glutaminyl-Peptide Cyclotransferase (QPCT) and Protein Kinase D3 (PRKD3)* genes has been previously described in a recent GWAS study (256). In a recent RNAseq profiling study of exercise training, it was demonstrated that the *QPCT* gene was upregulated following 12 weeks of training (257). As such, we expected the variant seen within the *NDUFAF7* gene to be associated with differing levels of the *QPCT* transcript following prolonged exercise training.

The two intronic variants within the *MTG2* gene were found to be associated with the change in time trial measures and appeared to be moderately linked (**Figure 5-7b**). The *MTG2* gene resides in a gene rich locus with 11 proximal genes. When assessed for functionality within the UCSC genome browser, we noted that both the MTG2 variants were located in a regulatory element (GH20J062181) that interacts with the MTG2 transcription start site. Further, there was a large amount of layered H3K27 acetylation at the variant site, and the linked MTG2

promoter region. The MTG protein regulates the assembly and function of the mitochondrial ribosome. As such, differential regulation of the gene could result in the downregulation of mitochondrial translation, and therefore a lower response to exercise training. The variants also showed a 20% recombination rate with the 5` regions of the *TAF4* gene. The TAF4 protein forms part of the transcription factor II D (TFIID) complex and has a central role in mediating promoter responses to transcriptional activators and repressors. Differential regulation of this gene could introduce global translational repression and therefore lack of response to HIIT training. This is supported by the positive effect size for the C and G alleles of the *MTG2* variants rs6062129 and rs6121949 respectively ($\beta = 587.7$ seconds).

An intronic variant within *AFG3L2* was also shown to be associated with the composite exercise response phenotype (rs7231304), but this gene has not previously been associated with exercise response. Mutations in *AFG3L2* have been shown to cause spinocerebellar ataxia through the development of mitochondrial proteotoxicity (258, 259). As such, the intronic variation within this gene might inhibit exercise response through differential regulation of mitochondrial structure and function. Further, this variant is in a locus with 6 proximal genes (**Fig 5-7c**), however no genes within this locus shared a recombination rate above 10%. There were two proximal SNPs with a moderate correlation to the SNP of interest also within the *AFG3L2* genic region. When assessed for functionality through the UCSC genome browser, the SNP was identified to be in a DNase I hypersensitivity region, and therefore may have affected the mRNA half-life rather than protein functionality in response to training.

The T allele at the exonic rs7085433 variant in the *TIMM23* gene was associated with the change in time trial phenotype (Δ-TT) causes a non-coding transcript of the *TIMM23* gene. This gene is one of the targets of transcriptional activators NRF-1 and GA binding protein (GABP/NRF-2) (260), in which we have previously shown genetic variants associated with athletic performance (261, 262). TIMM23 is one of the mitochondrial transmembrane subunits that form the mitochondrial protein import (TIM23) complex. Therefore, this subunit is essential for the transport of peptide containing proteins across the inner mitochondrial membrane. The non-coding transcript resulting from the variant would render the complex non-functional and as such impaired transport of biomolecules across the inner mitochondrial membrane may impair exercise potential. The effect size of this variant was positive ($\beta = 587.7$ seconds) and therefore, this non-coding transcript may result in a slower time to complete the time trial.

The rs1063271 variant lies within the 3` Untranslated Region (UTR) of the *SPTLC2* gene. UTR variants have been shown to influence transcript half-life; through the differentially regulated binding of transcript shuttle proteins; or change the binding site of miRNAs resulting in epigenetic silencing of the gene (263). As many current miRNA binding site analysis tools require targeted sequences, we examined the interaction between specific miRNAs previously found within exercise training with the 3`UTR of the *SPTLC2* gene (264). We noted that all of the miRNAs included in our STarMir curation were able to bind to the 3`UTR of the SPTLC2 gene in both seed and seedless sites. As such, it was not possible to indicate a specific miRNA mechanism within the context of this study although we note that this genetic variant in *SPTLC2* should be computationally explored in future studies. The SPTLC2 protein is involved in the de novo biosynthesis of sphingolipids by forming a complex with its counterpart; SPTLC1 (265). Overexpression of this protein has also been shown to cause elevated sphingolipid formation and therefore mitochondrial autophagy (266). Much like the *TIMM23* rs7085433 variant, the effect size for time to completion in Time Trial ($\beta = 587.7$ seconds) indicated that carriers of T allele/genotype have slower TT and therefore poorer response to exercise when compared to carriers of the C allele/genotype. The T allele for this variant was hypothesised to induce a novel miRNA binding site in the transcript resulting in the silencing of the *SPTLC2* gene.

The rs11061368 variant lies within an intronic region of the *DIABLO* gene. The protein encoded by this gene functions to induce apoptotic processes through the activation of caspases in the Cytochrome C/Apaf-1/caspase-9 pathway. When viewed in UCSC genome browser, the SNP was not able to be affiliated with any regulatory elements and therefore the true functionality of this intronic variant was unable to be determined. However, a molecular mechanism that should be explored in future exercise related studies was postulated. Although the associated variant does not show functionality within this gene, the differential regulation of the *DIABLO* gene may prevent adequate muscle remodelling resulting in the lack of response to training. Further, the variant also lies ~50Kb away from the *Interleukin 31* (*IL31)* gene, a pro-inflammatory cytokine associated with the activation of Signal Transducer and Activator of Transcription 3 (STAT3) pathways, which have already been extensively studied and implicated in exercise training responses.

The *FAM185A* gene was associated with the change in time trial measure with an effect size of 587.7 seconds. The gene has had limited previous research and as such any specific molecular function within the context of exercise training was unable to be elucidated. However, the gene has been previously associated with plexus-forming angiogenesis within the context of foetal lung tissue (267). It is plausible that the gene is involved in angiogenic processes outside embryonic development. Further, the gene is proximal to 9 other genes within 200Kb. There was no evidence of high recombination rates with any of the proximal genes, and there were no proximal SNPs correlated with the rs113400963 polymorphism. The polymorphism was assessed using the UCSC Genome Browser (https://www.genome.ucsc.edu/) to determine functional consequence and no link was identified between the intronic variant with any regulatory or epigenetic regions.

Mitochondrial genetic variants associated with training response

None of the mitochondrial genetic variants identified in this study were associated with exercise response in the present study to a threshold of FDR<0.05. Additionally, we lacked enough statistical power to associate mitochondrial haplogroup with exercise responses as the cohort was extremely heterogeneous. Although nominal significance was achieved, due to the hypervariable nature of the control region, SNPs within this region were not scrutinised.

The g.A8701G variant within the *ATP6* gene causes a missense change within its respective protein (p.Thr59Ala) and has been well characterised in hypertensive cases (268). This variant was nominally significant in both the Δ-LT phenotype and the PC3 composite trait within the cohort. As the Δ-LT trait provided a smaller contribution to PC3, the variant was assumed to be partially associated with a mixture of the Δ-TT and Δ-VO_{2max} phenotypes. The effect size of this variant indicated a poor response to exercise training (β = -26.19 LT).

Interestingly, all the variants found to be associated with PC4 were related to the utilisation of the amino acid Leucine. Firstly, the g.A12308G variant within the mitochondrial coding region for the tRNA for Leucine. Whilst the effect of this variant was unclear, it appeared to have influenced the composite phenotypes within PC4. Mutations within tRNA genes have previously been associated with reduction in organelle quantity and downregulation of protein synthesis (269). Secondly, both synonymous variants in the ND4 (g.A11467G) and ND5 (g.G12372A) genes result in a codon that is used far less frequently ($CUA_{[276]}$ > $CUG_{[42]}$) in mitochondrial translation processes (270). As the biosynthesis of tRNAs is costly with respect

to intracellular energy levels, it is possible that the combination of the differential regulation of the $tRNA^{leu}$ and the codon usage frequency change in two subunits of the mitochondrial membrane respiratory chain NADH dehydrogenase (complex I) may result in premature intracellular energy (ATP) deficiency and contribute to the poor response to exercise training associated with these traits. Of note, the stringent thresholds for association in the mitochondrial association tests could also have resulted in false negative results. Additionally, mitochondrial genetic variants rarely influence phenotypic traits in isolation.

Novel nuclear-encoded and mitochondrial-related SNPs and loci were found to be associated with adaptations to High Intensity Interval Training. Additionally, we have postulated the mode of action for different molecular mechanisms that may be responsible for the variability in response to exercise intervention. It should be noted that performing mitochondrial sequencing on muscle tissue as opposed to blood may yield more informative results with heteroplasmic associations due to the high concentration of mitochondria. Comprehensive sequencing and high throughput arrays in combination with robust exercise phenotypes were utilised for this study, however the variants found to be associated with responses in this study need to be replicated in larger cohorts of both the general population and elite athletes. Further, the variants assessed within the current study were tag SNPs within the genotyping arrays and further information may be gained from the imputation of additional SNPs within the regions we have discovered. This could be achieved by leveraging on large multi-centre initiatives such as the Athlome consortium (218). Additionally, functional genomic analyses are required to determine the effect of these variants on the molecular pathways commonly involved in exercise response. Such studies could include transcriptomics, epigenetics, and functional cell work in a multi-omics approach.

As a secondary part of this project, a mitochondrial sequencing protocol for Ion Torrent applications was optimised and utilised. The modified HTMGS protocol provided comparable quality sequencing data and mitochondrial genome coverage/sample compared to the manufacturer's recommended output. Components of the protocol were assessed including the total cost of the library preparation reagents used in each protocol with their respective reaction amounts to determine the cost of reagents on a per sample basis. Additionally, as sequencing reagent costs are relatively similar, the higher multiplexing of samples (n=48) shown in the protocol should further increase the price gap between the methodologies. Of note, the use of

an Ion Chef instrument for templating in our method likely contributed to the observed high yield and sequencing quality.

This method may be further adjusted for higher throughput capabilities, which will decrease coverage per sample and deliver a reduced cost per sample for sequencing. Multiplexing of n=96 samples onto an Ion 316 chip was expected to deliver an approximate coverage of 200 times per sample (assuming chip utilisation at full capacity). For studies that require large amounts of samples and an increased coverage depth we recommend utilising the scaling capabilities of the Ion Torrent instruments and the capacity of other available chips.

The Ion One Touch system may have potentially introduced user error and increased assay time, an important factor for high throughput applications which may be minimised through integration of automation into the HTMGS protocol. For forensic and diagnostic applications, coverage of the mitochondrial genome is not strictly specified and is more stringent on the concomitant replication of identification results via Sanger sequencing (Scientific Working Group on DNA Analysis Methods [SWGDAM]/National Association of Testing Authorities [NATA] guidelines). Further optimisation of total coverage should be performed when considering heteroplasmic associations.

A new NGS mitochondrial sequencing protocol was developed for whole mitochondrial genome sequencing. While other current techniques produce comparative sequencing qualities and time to completion, we have shown that our protocol delivered increased overall read quality due to the differences in library preparation. The HTMGS protocol utilises multiplexing of samples onto the same Ion chip and delivers comparable high-quality results whilst being more cost effective than current commercial methodologies.

Chapter 6: Transcriptomics of Exercise Response

The data presented within this chapter contributed to the planned publication as outlined below:

- **Harvey NR**, Thompson J, Voisin S, Haupt LM, Eynon N, Griffiths LR, Ashton KJ (2021). **Integrative analysis of epigenetic and transcriptomic data from skeletal muscle microenvironments reveals key biological terms associated with adaptations to acute and chronic HIIT**, *Planned submission September 2021 Molecular Systems Biology*

The data produced and discussed during this chapter relates to Aim 3 of this thesis. The Aim is reiterated below, and the project specific sub-aims stated in greater detail.

3. Utilise targeted RNA sequencing to identify longitudinal global gene expression changes and pathways in response to training.
    a. QC RNA samples prior to experimentation to address any bias in gene expression values
    b. Perform targeted RNA sequencing using the AmpliSeq Transcriptome methodology
    c. Perform differential gene expression analysis based on the results from the sequencing
    d. Perform pathway analysis from the differentially regulated genes
    e. Link molecular pathways changes to previous chapters and aims

6.1 ABSTRACT

Exercise training remains the most efficient way of maintaining health by reducing the risk of metabolic disease. Previous studies have been limited by low participant numbers and lack of longitudinal time points during training. Recent meta-analyses have implicated multiple genes and molecular pathways at early exercise time points however HIIT studies remain under-represented. In addition, there is much to be gained from exploring transcriptomic responses in later exercise time points as studies are heavily biased towards acute responses. In the present study, 55 participants (corresponding to 204 muscle biopsy samples) were used from the Genes SMART (Skeletal Muscle Adaptive Response to Exercise) study conducted at Victoria University. All muscle samples were subjected to targeted transcriptome sequencing using an Ion Torrent Proton platform. Following bioinformatic approaches to QC and normalisation of samples, 12,682 genes were expressed across all samples, of which 96 genes were significantly differentially regulated immediately following HIIE (P0), 3,939 genes three hours post HIIE (P3), and 2,812 genes following four-weeks of HIIT (4WP). The transcription factors *MYC*, *FOS*, and *JUN* were highly upregulated immediately following a single bout of HIIE. Further, these remained upregulated after three-hours, but the effect was limited to acute exercise training response. This study was also able to replicate upregulation of previously discovered exercise inducible genes (*PPARGC1A MSTN, CK-MM*). To discern biological meaning, Gene Set Enrichment Analysis (GSEA) was performed to determine enriched biological processes within each exercise time point. Immune terms were found to be highly upregulated in the immediate and three-hour time points. Interestingly, the number of differentially expressed genes did not correlate with the number of enriched terms as the four-week HIIT time point contained the largest number of enriched gene ontology terms within the dataset. A longitudinal switch occurred from immediate stress response inducible processes to prolonged biological function with protein targeting and energy metabolism more evident after four-weeks post HIIT.

This chapter also assessed the transcriptomic responses of candidate genes previously identified in Chapters 4 and 5. Repression of *IL6* at four-weeks post training was found to be prevalent within pathways corresponding to cytokine production, as per the first aim within this thesis. The *MICAL-L2* gene was identified as an important mediator between insulin signalling, actin filament polymerisation, and glucose transport into skeletal muscle likely mediated by GLUT4 as per thesis Aim 2. In addition, the *METT-L12* and *JMJD6* genes were identified as important mediators of epigenetic chromatin remodelling at four-weeks post HIIT

as per Aim 4. Within this chapter, we have demonstrated the utility of transcriptomic approaches and identified likely biological processes that are temporally differentially regulated in response to prolonged HIIT.

6.2 BACKGROUND

As previously stated within Chapter 2, several molecular processes that govern adaptation to endurance training are well understood. Briefly, skeletal muscle adapts to the stress induced by HIIE thorough a number of mechanisms; specifically, calcium influx and oxidative stress leading to an increase in mitochondrial number and function through AMPK and PGC1α pathways; a shift from carbohydrate metabolism to lipid metabolism; induction of HIF1α to stimulate glycolysis and lactate transport; secretion of myokines (such as IL-6) to stimulate inflammation; and expression of VEGF leading to angiogenesis. The transcriptomics behind these adaptations have been well assessed within exercise studies relating to acute endurance and MICT (174, 271-279). Exercise training induces several physiological responses and therefore global transcriptional changes as a response. For example, regular exercise training at high intensities (>80% VO_{2max}) will lead to hypoxia and therefore activation of HIF inducible molecular pathways changes as the microenvironment reacts to physiological stress (280-282). *Sato* and colleagues reported multiple molecular pathways unique to sedentary individuals and fitter participants. In particular, rRNA processing, mitochondrial translation and electron transport were shown to be significantly downregulated in the active phase of exercise training . Whereas, significant upregulation of glucose to acetyl-CoA and cytokinetic pathways were observed in response to exercise (283, 284).

Longitudinal exercise studies are challenging due to the increased participant compliance and time commitment, often resulting in high attrition rates and increased funding and resource requirements (151). Further, there remains a large gap between studies examining muscular response to exercise training and studies examining the response to exercise training as a therapeutic option for relevant myopathies (285). As such, significant 'Omic' level findings are difficult to obtain without large scale collaboration and/or meta-analyses. However, caution must be taken during the compilation of multiple small-scale studies as this may lead to confounding variables when analysed together. For instance, multiple anthropomorphic covariates (ethnicity, country where study was performed, differing ways to measure the same phenotypic data) may explain different transcriptional effects within each data set. Further, as many exercise studies were performed prior to the advent of RNA sequencing technologies, the technology utilised to measure transcriptomic changes may be different and therefore contain significant variation as well as limitations. Specifically, the compilation of studies

utilising gene expression microarrays, RT-qPCR, and RNA sequencing technologies should be heavily scrutinized as the significant results must be limited to the smallest collection of data. In addition, many exercise studies have examined the multiple disease state interactions, which may be contributing to large variances within the transcriptomic data across normative control studies.

One large scale study aimed to address these limitations and was able to compile the transcriptomics of 66 published exercise studies in a meta-analysis. However, the study was unable to examine or implicate specific molecular pathways that change in response to HIIT training (124). The study did, however, confirm the effect of induced *NR4A3* expression on multiple exercise inducible pathways. It should be noted that further genes were also implicated but did not correlate to the expression of downstream genes that influenced exercise capacity, and are not discussed here. The authors were able to perform KEGG pathway analysis and as such implicated several molecular pathways that correlated with acute resistance or aerobic training. More specifically these were '*ribonucleoprotein complex biogenesis', 'generation of precursor metabolites and energy',* and *'cellular respiration'*. Resistance training pathways included '*regulation of protein complex assembly'*, '*regulation of mRNA metabolism'*, '*ECM organisation'*, and '*RNA splicing'*. In addition, mitochondrial complex proteins were upregulated in aerobic training but did not change with resistance training. *Robinson et al,* (as described in Section 2.6.2 and Section 2.7.3) published (to date) the only HIIT study to incorporate healthy subjects, longitudinal sampling, and both gene expression and epigenomic analysis (150).. A key knowledge gap within the existing literature is the lack of studies pertaining to HIIT, with specific reference to large-scale transcriptomics.

The Gene SMART study is a tightly controlled longitudinal exercise study examining the acute and chronic response to HIIT in a relatively large cohort (n=55) of healthy individuals. It was hypothesised that the assessment of transcriptomic signatures within the Gene SMART study would allow for the discovery of novel exercise responsive genes and molecular pathways. In addition, it was expected that several of the known exercise inducible genes (*PPARGC1A, NR4A3, HIF1A*) would be replicated within the acute bout time points and therefore validate previous studies and provide confidence in the other findings from the Gene SMART cohort.

6.3 METHODS

6.3.1 Participant Samples

As previously described in **Section 3.2.4**, muscle biopsies for three or more timepoints were available for n=55 participants who completed the four-week HIIT program.

6.3.2 RNA Isolation

Simultaneous co-extraction of DNA and RNA was performed using the AllPrep DNA/RNA/miRNA Universal kit (*Qiagen*). This ensured that the low amount of sample was fully utilised, along with collection of miRNAs for future projects. Briefly, ~20mg ± 5mg of skeletal muscle tissue was first homogenised in β-Mercaptoethanol reagent with 0.9-2.0mm RNase-free steel beads in a Bullet Blender Gold (*NextAdvanced*) at 4°C to minimise enzymatic degradation of RNA or DNA. Homogenised muscle samples were then processed using the AllPrep kit according to manufacturing protocols. The RNA quantity and quality were assessed via UV spectrophotometry Nanodrop (*Thermofisher Scientific*) and Qubit fluorometry (*Thermofisher Scientific*). RNA integrity was also assessed using the RNA 6000 Nano kit on a 2100 Bioanalyser (*Agilent*). Total RNA was stored at -80°C prior to RNA sequencing.

6.3.3 RNA sequencing

RNA sequencing was conducted using the Ion AmpliSeq Transcriptome Human Gene Expression kit (*Thermofisher Scientific*). This kit utilises a primer pool targeting ~95% of RefSeq genes (hg19: 18,574 mRNAs, 2,228 ncRNAs), generating a single amplicon from each gene, for subsequent sequencing and quantitation (286). The method was conducted according to manufacturer's protocols, with the following amendments. For transcriptome amplification, 14 PCR cycles were used compared rather than the recommended 12 cycles, resulting in a notable increase (~1.5fold) in library prep efficiency and yield. This kit was used in conjunction with IonXpress barcode adapters (#1-16) for Ion Torrent applications (*ThermoFisher Scientific*, Cat#4471250). As such, all transcriptome libraries were produced in batches of 16 samples corresponding to the number of barcodes available. Each barcoded amplicon was diluted to 100pM and an equimolar library pool (25µL total) was then loaded onto Ion PI chips using an Ion Chef System (*ThermoFisher Scientific*) as per the manufacturer's instructions (Pub. No. MAN0010967, Rev. B.0). Sequencing was performed with an Ion Proton™ semi-

conductor sequencer (*ThermoFisher Scientific*) and Ion PI Hi-Q Sequencing 200 kits (*ThermoFisher Scientific*) according to manufacturer's instructions (Pub. No. MAN0010947, Rev. C.0). All samples were sequenced using a total of 30 Ion PI chips, including repeat sequencing of 18 samples due to low initial read numbers.

6.3.4 Bioinformatics

Linux based pre-processing and alignment

Raw binary alignment metric (BAM) sequence files were downloaded from the Ion Torrent Proton server and placed onto a Linux based external server with pre-loaded bioinformatics software (*SAMtools (v1.6-56-glc5c508), FastQC, BBMAP, bgzip*) (287-289). The raw files were converted to FASTQ files using the *SAMtools bam2fq* function and then compressed with *bgzip*. *FastQC* analysis was performed on the raw FASTQ files to determine the quality of the sequencing, as well as the presence of any contaminating adapter sequence (290, 291).

The laboratory method and library prep outlined in Section 6.2 resulted in some sequence contamination, specifically concatomer reads (i.e. twice the mean read length) or sequence artefacts (i.e. less than half the desired read length). The FASTQ files were therefore trimmed based on size distribution using the *BBMAP* package to retain reads between 50-150bp. Following trimming, FastQC analysis was repeated to validate the removal of these contaminating sequences.

A compressed FASTA file containing the entire human RefSeq mRNA sequence was downloaded from the ENSEMBL database (www.ensembl.org: *GRCh37_latest_rna.fna.gz, version 91*) and uploaded to the Linux based server. A quasi-index was built from this file with the *salmon, -index* (v-0.8.2) function. *Salmon* was used to align and quasi-map the reads to the reference index and then quantify the number of transcripts per gene. The resulting quantitation files were then exported to R/Bioconductor for further quality control and data handling, along with differential expression analysis.

QC analysis and differential regulation analysis in R/Bioconductor

Differential expression analysis was performed using R/Bioconductor version 3.6.1 and closely followed the analysis guidelines outlined by Law *et al* (292). Firstly, ENSEMBL transcript IDs from the *Salmon* quasi-mapping were consolidated to the matching ENSEMBL gene IDs using the *tx2gene* function of the *tximport* package (293). Genes were annotated via

the *biomaRt* package using ENSEMBL version 91. Transcripts were filtered to exclude those with a median count per million (CPM) below 0.5 across all 204 samples, as this low level of expression was considered too low for accurate quantification. All count values were then normalised using trimmed mean of M-values (TMM) normalisation to adjust for library size and therefore transcript count bias between samples.

Following this data processing, a model matrix was constructed with the sample time points and included the necessary confounding variables (RNA Integrity Number (RIN) and sequencing batch). The *voom* function within the *limma* package was then used to normalise the expression values for each gene according to the model matrix (294, 295). Secondly, as samples were longitudinal time points for the same participants, a paired design matrix was used to perform a second round of *voom* normalisation and correct the model for samples with the same participant ID. Briefly, paired design increases the statistical power in related samples (such as longitudinal samples) and does not carry the same assumptions (normal distribution/ no difference between comparison groups) as studies with an unpaired design. The corrected model was fit into a linear model using the *lmFit* function in *limma*. A contrast matrix comparing all exercise time points to the baseline samples was constructed and then fit to the linear model to discover differentially expressed genes following training. Lastly, the normalised and contrasted model was fit to a Bayesian model in order to gauge the magnitude of differential regulation necessary for each contrast within the contrast matrix. The final numbers of differentially expressed genes for each time point was summarised using the *decideTests* function in *limma* and Venn diagrams. Multiple testing burden was addressed by correcting raw P-values using the Benjamini-Hochberg-False Discovery Rate (FDR) correction and significance was set at FDR<0.05.

Gene set enrichment and gene ontology analysis

Gene set enrichment analysis (GSEA) was used to assess coordinated shifts in gene expression associated with biological processes and molecular pathways. Briefly, all genes within the normalised expression data were ranked according to the t-statistic (*limma*) from most upregulated to most downregulated. Multiple gene set enrichment databases were employed to gauge an accurate number of pathway changes across the exercise time points. These included the *reactome.db (v1.68.0)(296)* package for *reactome* annotated pathways, as well as the *ClusterProfiler (v3.12.0)(297)* package to detail gene ontology (GO) changes for

the biological processes (BP) domain. To gauge significantly enriched processes, 10,000 permutations were used in conjunction with a dual statistical threshold of FDR <0.05 and P <0.001 Additionally, the set size ranges were limited to between 25 and 500 genes to ensure appropriately sized pathway gene sets for further analysis.

Significant ontology terms for each time point were visualised in *Cytoscape v3.7.2* using the *EnrichmentMap* application (298). The Jaccard overlap (similarity threshold) was set to 0.375 to limit the number of overly connected BP nodes in the enrichment maps. The *AutoAnnotate* package was utilised to cluster the BP nodes into clusters based on similar function (299). The clustered BP terms were manually named to accurately describe the biological function of the respective grouped BP nodes. BP nodes were firstly assessed for differential regulation status and coloured by normalised enrichment score generated from the gene set enrichment. To assess the BPs corresponding to the previous chapters of this thesis, the *grep* function was used to separate the GO BP terms based on "cytokine", "mitochondrial", "histone", and "epigenetic" terminologies. These were visualised as previously described and coloured by FDR to gauge specific terms of interest to the respective clusters. The *heatplot* function of the *ClusterProfiler* package was then used to extract the genes within the GO terms of interest. The gene sets were assessed in *Cytoscape* within the *GeneMania* package to visualise the co-expression of the genes within the BP (300). Individual gene nodes were coloured by log fold change (logFC) to ascertain the directionality of the genes at each time point. Genes with the most connectivity to other gene nodes within the BPs were termed 'Hub' genes and modified in order to be emphasised in the *GeneMania* plots.

6.4 RESULTS AND DISCUSSION: DIFFERENTIALLY REGULATED GENES

6.4.1 Quality Control

Integrity of the RNA extracted from the muscle samples using was assessed via Bioanalyser RNA chips. The average RIN was 6.8 ± 2.1 across all 204 samples. A representative electropherogram (RIN=6) is shown in **Figure 6-1**. However, as previously discussed in the methods, **Section 6.3.2**, the inclusion of the small RNA fraction in the final elution during RNA extraction resulted in low estimations of RNA integrity even though the ribosomal peaks were consistent with high sample integrity. Therefore, RIN was used as a covariate in the model matrix to adjust for this variance.

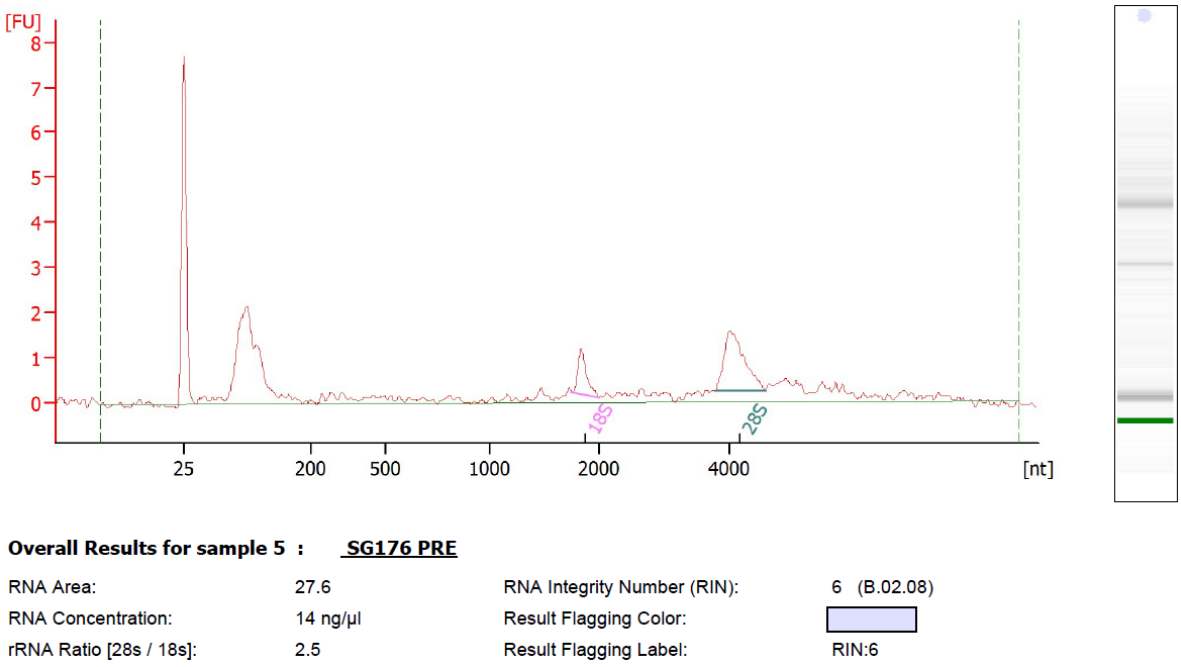


Figure 6-1: *Representative Bioanalyser trace showing RNA starting material prior to sequencing. RIN value was calculated with the bioanalyser software using the peak ratios between the ribosomal RNA regions. Y-axis shows fluorescence units and x-axis shows size of RNA fragments in nucleotides.*

Following RNA isolation and library construction, Bioanalyser traces showed all samples generated libraries of consistently high yield (210 ± 83nM) and within the protocol's specified size (150-300bp). An example of the Bioanalyser traces for the final prepared libraries (stock and diluted) are shown in **Figure 6-2**.

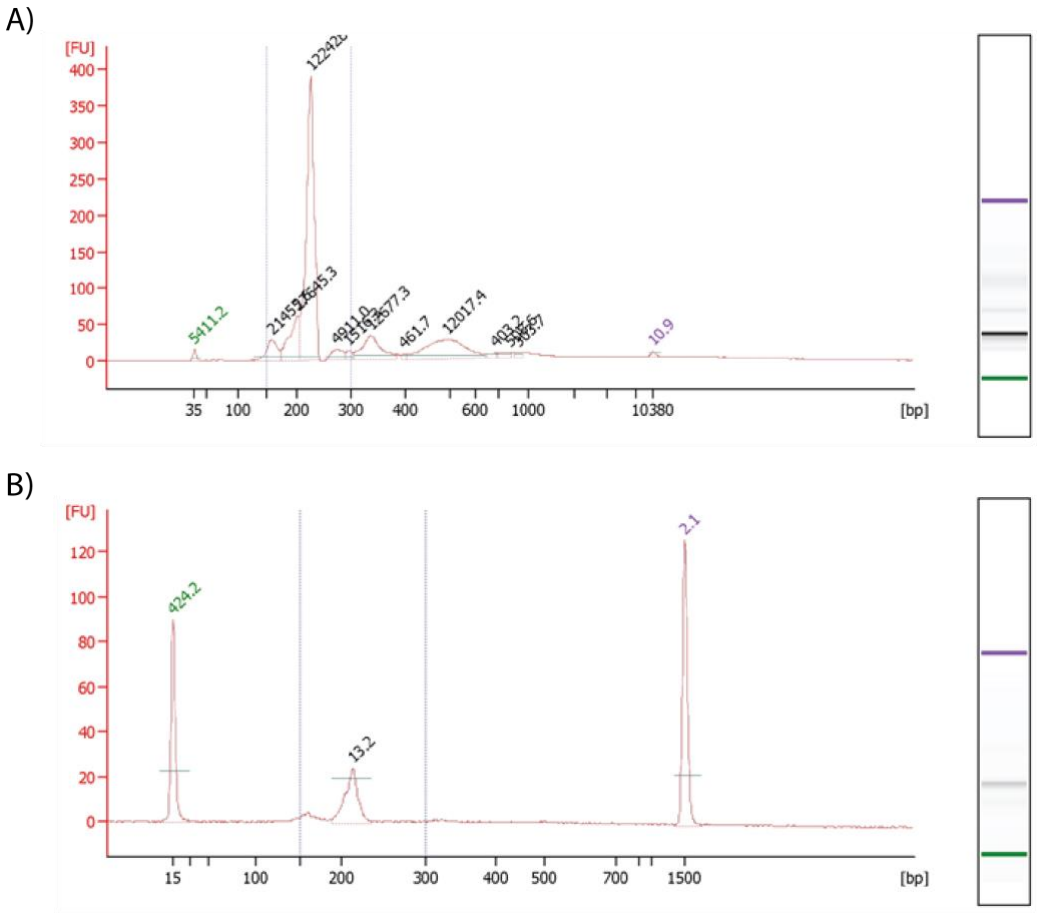


Figure 6-2: *Bioanalyser traces showing the* ***A****) stock final transcriptome library and a* ***B****) 1:10 dilution to more accurately quantify prior to sequencing on a high sensitivity DNA chip. Library size was approximately 200bp (130bp amplicon + (2X 30bp barcodes)). For each panel, X-axis represents library size in base pairs, Y-axis shows fluorescence unit intensity at specified base pair size. Electropherogram traces are shown to the right of each trace.*

Trimming of adapter contaminants with *BBMAP* had a minor effect, improving the mapping rate by ~1%. On average, 97.3% of the generated sequencing reads mapped accurately to the human transcriptome. *FastQC* and *MultiQC* analysis was then performed on the trimmed files to determine both sequence quality and contaminating sequences. All the *MultiQC* metrics were consistent with high quality sequencing except for the GC content metric (**Figure 6-3**).

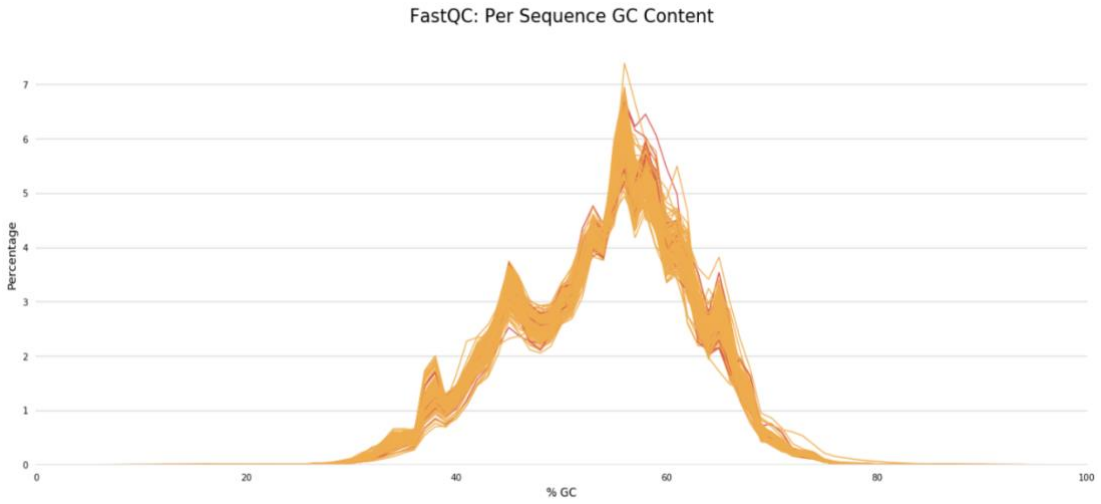


Figure 6-3: *MultiQC GC-content metric from all transcriptome samples. GC content should follow a normal distribution however the targeted sequencing resulted in spiking at GC locations on overrepresented reads with high basal expression within skeletal muscle tissue. X-axis shows percent of pyrimidine residues within reads, Y-axis represents percentage of reads reaching the specified GC content within the data.*

Considering the sequencing chemistry used utilises small regional sequences over many genes, we noted that any overrepresented sequences would result in consistently placed GC rich regions within the sequencing reads. None of the samples contained overrepresented sequence above 9% of the total sequences within the transcriptome data. As such, the spikes seen within the GC content plot **(Figure 6-3),** were most likely due to highly expressed transcripts. To examine this, the overrepresented sequences from the FastQC analysis were identified using the NCBI mRNA sequence BLAST tool. Of the top 29 overrepresented transcripts, 23 mapped to muscle specific transcripts including *ACTA1*, *TNNT1*, *ACTN2*, *TNNC2* and *CKM*, whereas 6 mapped to genes commonly thought to be constitutively expressed in every cell type: *GAPDH*, *TPT1*, *EIF1*, *AK1* and *ATP5F1B*. This provided some evidence that the overrepresented sequences were biologically accurate and not due to PCR bias or single-end read generation.

The average number of sequencing reads for each sample was 10.8M ± 4.6M, with 11 samples outside the 5-15M read range, specifically two samples below and nine samples above (**Figure 6-4**), with the lowest read count of 2.6M, and the largest read count of 47.2M.

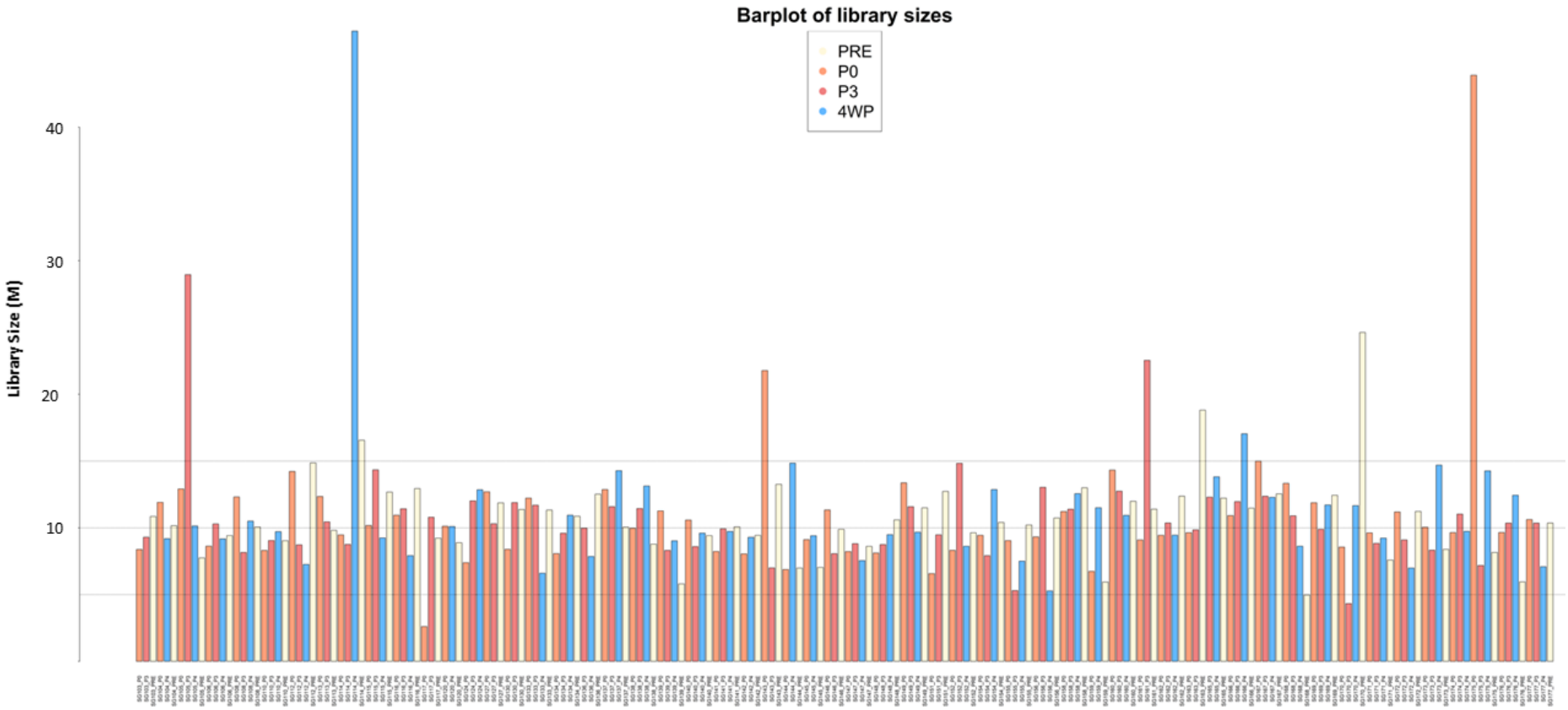


Figure 6-4: *Summary of the total library sizes across all transcriptome samples. Threshold lines were set at 5, 10, and 15 million reads to assess (non)uniformity between samples. Samples were coloured by time point (PRE: cream, P0: orange, P3: red, 4WP: blue) to assess library size discrimination between sample groups. X-axis shows sample names, Y-axis shows sequenced library size in millions (M).*

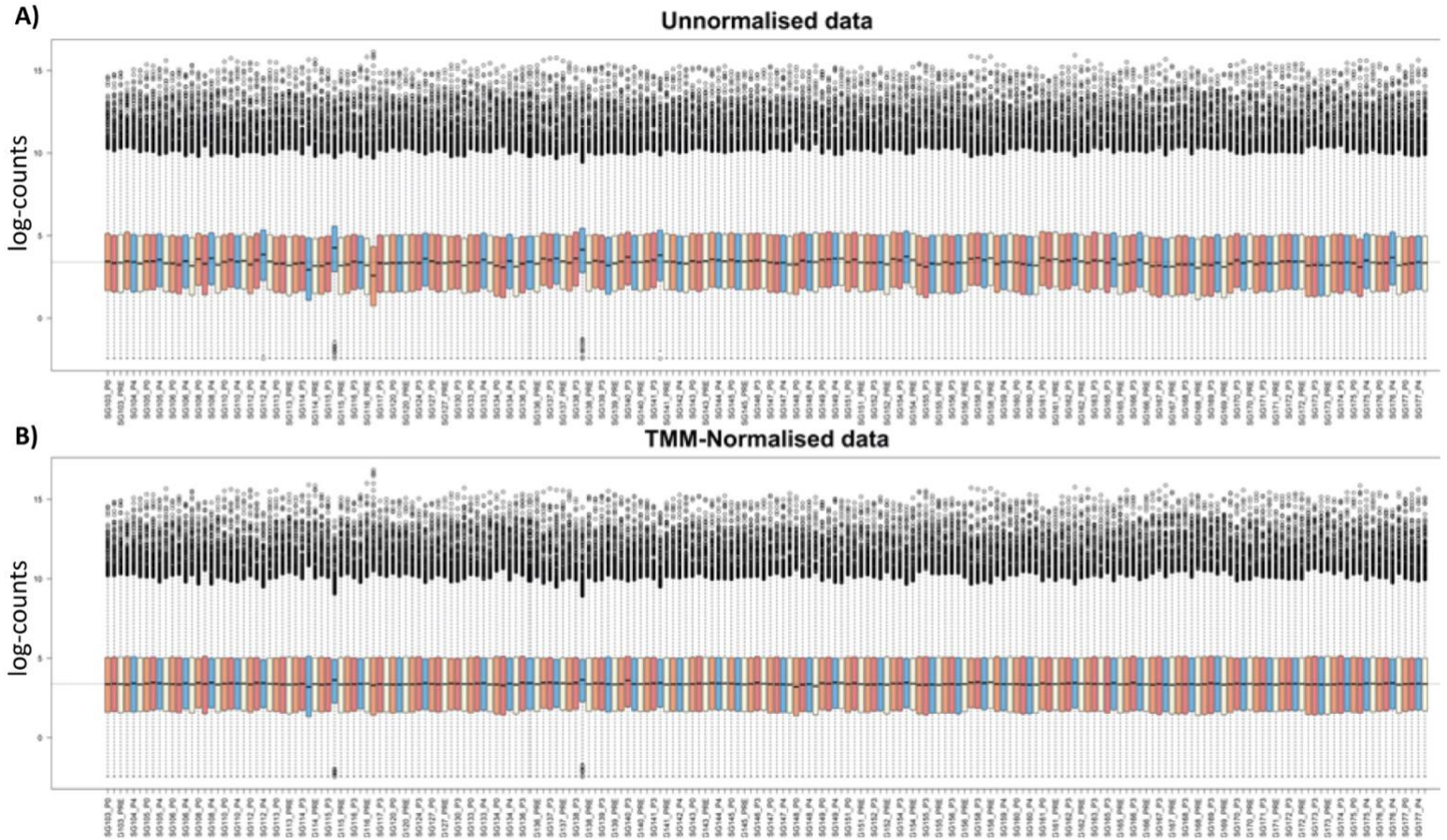


Figure 6-5: *Boxplots of raw and TMM-normalised log counts across transcriptome samples. Boxplots show interquartile range of sequencing data, whiskers represent the 95% confidence intervals with outlying data depicted as black points.* ***A)*** *raw unadjusted data with the median log odds values for each sample depicted as the black line at the centre of each box.* ***B)*** *TMM normalised data with similar median values. Samples were coloured by time point (PRE: cream, P0: orange, P3: red, 4WP: blue) to assess library size discrimination between sample groups. X-axis shows sample names, Y-axis shows log Odds read count values.*

The data in **Figure 6-5A** shows the varied distribution of expression values in the pre-normalised samples. Following TMM normalisation, the read counts are scaled based on library size resulting in more uniform range and median per sample (**Figure 6-5B**). Genes were then filtered based on a median of <0.5 CPM to remove non- and very lowly expressed genes which would have unreliable expression levels (**Figure 6-6 A).** Of the 12,682 genes remaining, most had a low to medium (0-5 $\log_2$[CPM]) expression level (**Figure 6-6B).**

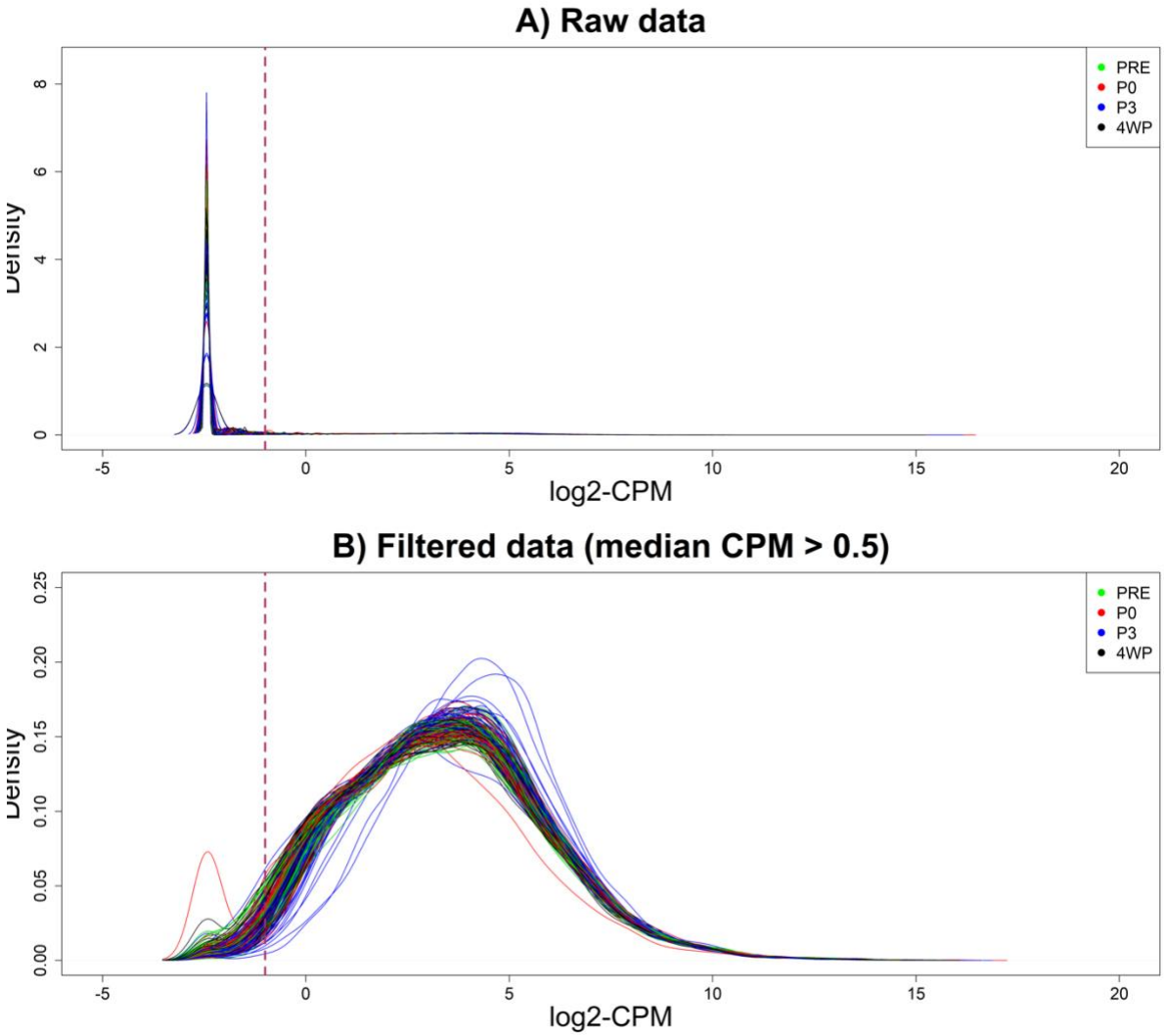


Figure 6-6: *Transcript filtering based on expression level.* ***A)*** *Raw unfiltered data, and* ***B)*** *data following removal of unexpressed and lowly expressed transcripts. Samples were coloured based on exercise time point (PRE: green, P0: red, P3: blue, 4WP: black) to ensure no discrepancy between read counts for each sample group.*

Prior to differential expression analysis, the unsupervised clustering of samples with PCA plots was examined to discern any unexplained variance within the data set. The unadjusted and covariate adjusted PCA plots are shown in **Figure 6-7.** No sample clustering that clearly corresponded with each of the four exercise time points prior to adjusting for covariates (RIN, batch) was identified (**Figure 6.7A**). The unadjusted correlation matrix (**Figure 6-7B**) suggested that the variance in the top PCs was more strongly associated with participant, library size and RIN. This unwanted variance was largely removed following normalisation and the paired sample design (to adjust for inter-participant variation) with the model adjusted for RIN (**Figure 6-7C-D**).

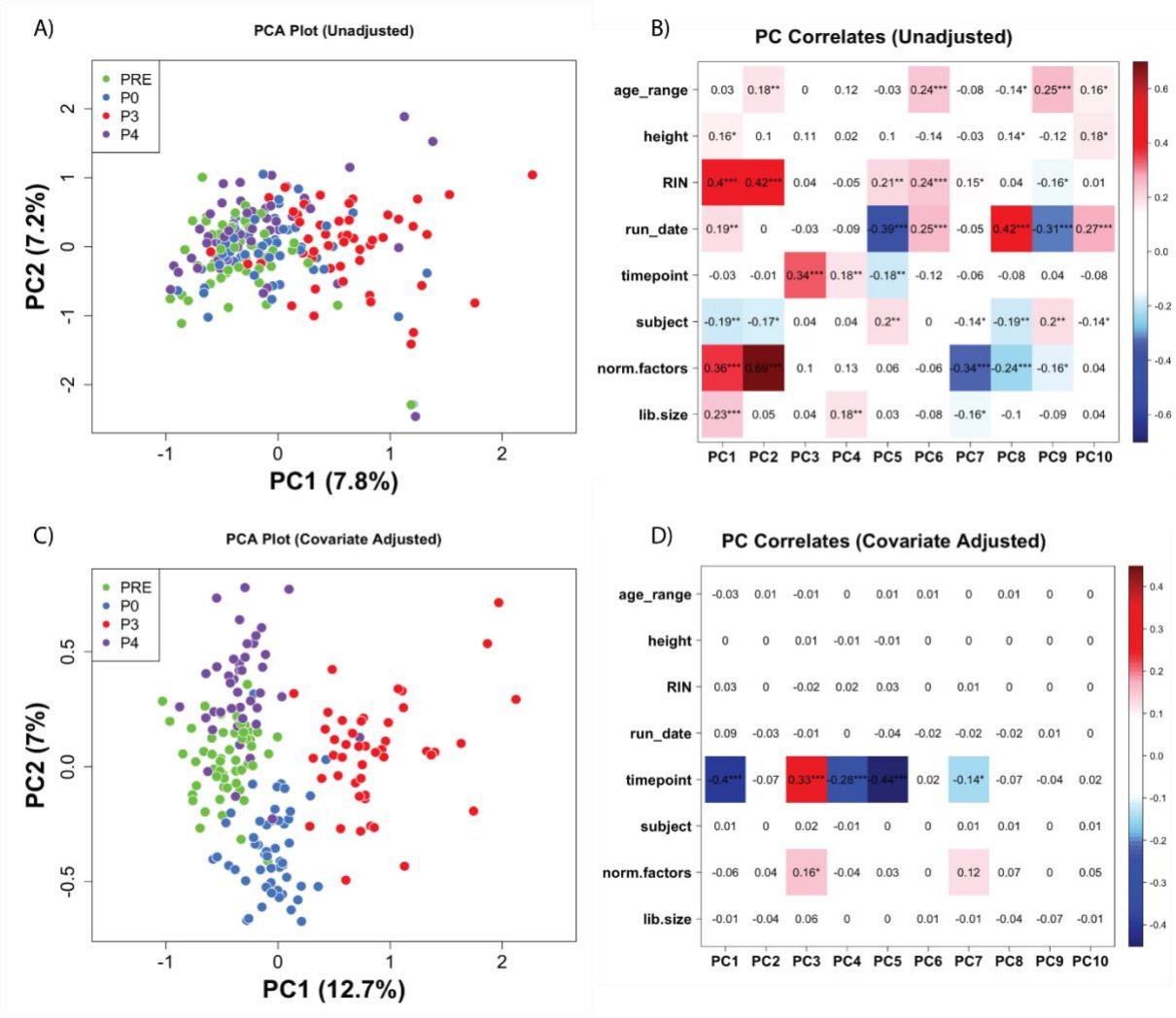


Figure 6-7: *PCA correlation plot and eigen correlation plot of the top 500 most variable genes within the Gene SMART samples.* ***A)*** *unadjusted PCA plot showing no separate clustering of time points across samples.* ***B)*** *Correlation matrix of the contributing phenotypes to PCs 1-10 showing multiple confounding phenotypes.* ***C)*** *PCA plot adjusted for the most contributing phenotypes (RIN, batch) from panel B showing correct clustering of samples into individual exercise time points.* ***D)*** *Correlation matrix of the adjusted contribution of phenotypes to each of the top 10 PCs showing heavily weighted timepoint phenotype in the early PCs. Panels B and D show correlation values used to determine directionality of each covariate to each PC.*

6.4.2 Differentially expressed genes

Following analysis using a robust analysis pipeline (**Section 6.3.4**), the number of genes significantly differentially expressed in each of the exercise time points when compared with the baseline (PRE) transcriptome was assessed. A total of 5,858 genes were found to be differentially regulated over the entire data set (**Figure 6-8**). Five genes (*ACTB, SLC25A25, CXCR4, GEM, CD83*) were consistently upregulated in each time point. The P0 and P3 time points shared a total of 68 differentially regulated genes, 58 of which were upregulated and 10 downregulated. The 4WP and P3 time points contained the largest number of overlapping genes with 521 upregulated and 398 downregulated. It should be noted that the intersection of differentially regulated genes between time points was not explored within this thesis. All the results discussed refer to the detailed exploration of the individual time points. All differentially regulated genes in each time point were assessed for magnitude and significance using volcano plots (**Figure 6-9**).

Immediately after HIIE (P0) a small number (96 genes) of genes were shown to be differentially regulated; of which the majority (80.2%) were up-regulated. These genes corresponded to the upregulation of transcription factors such as *Early Growth Response 1* (*EGR1*), *JunB Proto-Oncogene, AP-1 Transcription Factor Subunit (JUNB), Early Growth Response 2* (*EGR2), Fos Proto-Oncogene, AP-1 Transcription Factor Subunit (FOS)*, and *FosB Proto-Oncogene, AP-1 Transcription Factor Subunit (FOSB)* (**Table 6-1**). This was unsurprising as the cellular response to stress has been extensively characterised and found to be a feature at early exercise time points (301-303).

At three hours post exercise (P3-PRE), the magnitude of differentially expressed gene changes greatly increased with 3,939 genes identified to change from the preliminary time point. A more even proportion of the genes (52.7%) were identified to be upregulated in response to training. This time point contained the largest number of upregulated genes at a level of more than 2-fold to baseline. Whilst this accurately describes the magnitude of differential expression at this time point, fold changes must be interpreted contextually with transcriptome data as large fold change values may be observed for lowly expressed transcripts. Interestingly, the differentially expressed genes at the three-hour s post-HIIE have also been previously shown to be involved in shock response, however it was also evident that immune response genes were also highly upregulated (**Table 6-2**). Specifically, the *FOSB*, *EGR1*, and

MYC proto-oncogene, bHLH transcription factor (MYC) genes remained upregulated at three hours. The epigenetic modifying protein genes *AT-rich interaction domain 5B (ARID5A*), and *tet methylcytosine dioxygenase 1* (*TET1*) were identified to be downregulated at three hours. The expression of the common exercise gene *myostatin* (*MSTN*) was attenuated at the three-hour time point, which was unexpected due to the common association with strength training. This has been demonstrated previously for endurance exercise activity (304, 305), however the functional effects from this change in gene expression should be explored in further studies.

The four-week HIIT (4WP-PRE) time point contained a large number (919 genes) of overlapping genes with the three-hour time point. A total of 2,812 genes were found to be differentially regulated, with 54.3% upregulated in response to four weeks of HIIT. Most notably, a large number of histone protein subunit genes were identified to be upregulated at 4WP-PRE, with no evidence of these transcription factors at both of the early time points (**Table 6-3**).

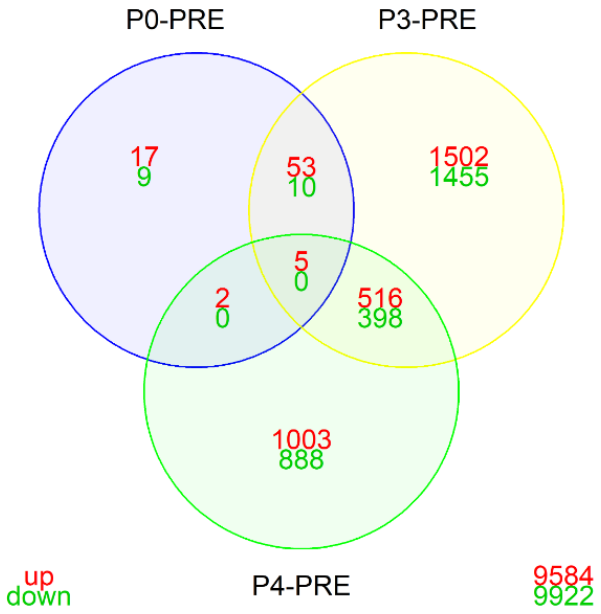


Figure 6-8: *Venn-diagram showing the number of differentially regulated genes in each exercise time point (P0, P3 or 4WP) compared with baseline (PRE). Numbers within the circles represent upregulated (red) and downregulated (green) genes. The numbers in the right-hand bottom corner of the plot represent total numbers of genes that were not significantly differentially expressed (FDR<0.05) in any of the three time points.*

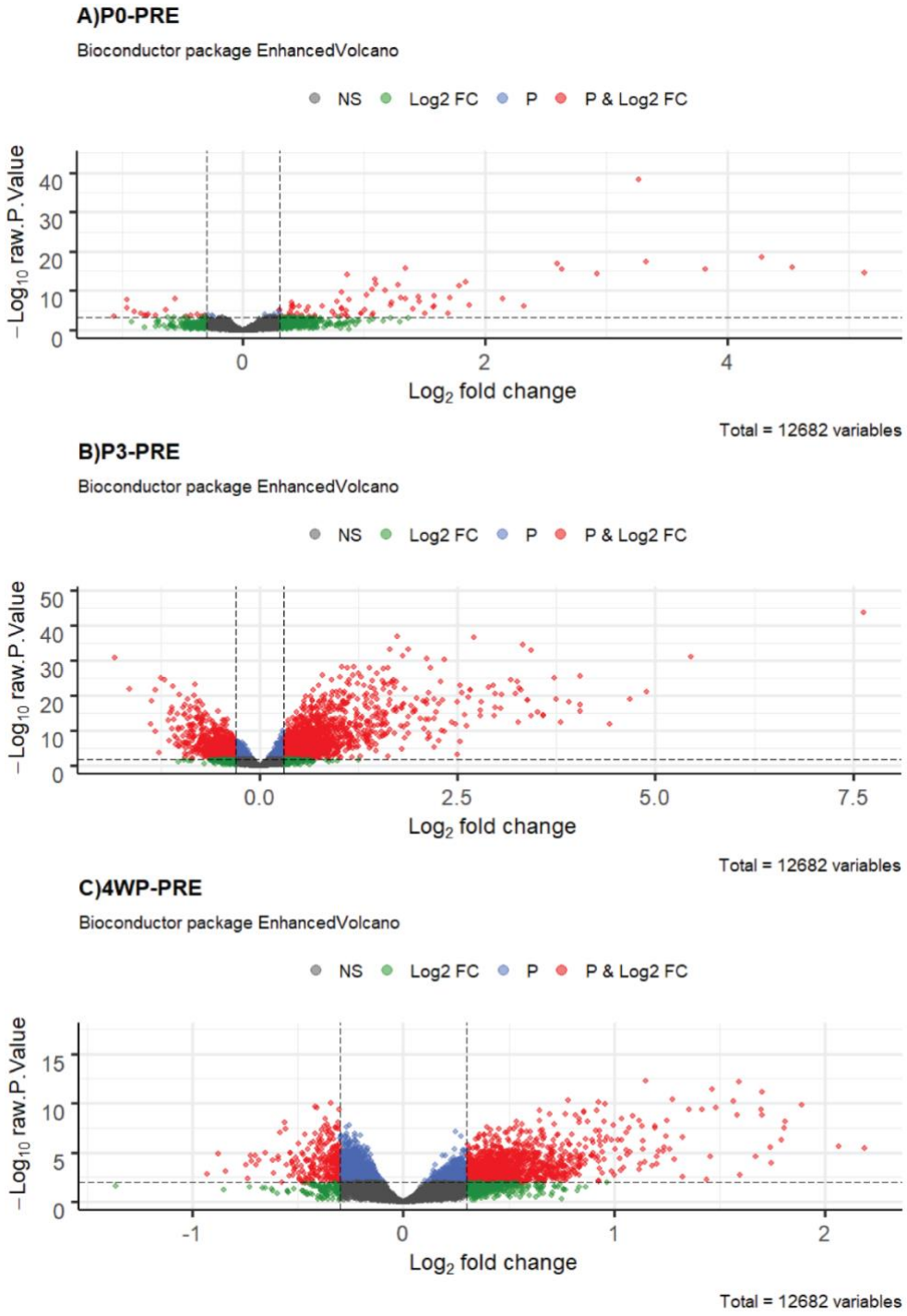


Figure 6-9: *Volcano plots showing significance and fold change of differentially regulated genes in each exercise time point.* ***A****) differentially regulated genes that are significant in the immediately post exercise time point (P0),* ***B****) significantly differentially regulated genes in the three-hour time point (P3),* ***C****) significantly differentially regulated genes after four-weeks of HIIT (4WP).*

Table 6-1: *Top 10 up and downregulated genes in the immediately post exercise time point (P0) ranked by fold change*

	P0-PRE		
SYMBOL	**GENENAME**	**FDR**	**FC**
FOSB	*FosB proto-oncogene, AP-1 transcription factor subunit*	3.37E-12	35.05
FOS	*Fos proto-oncogene, AP-1 transcription factor subunit*	2.69E-13	23.17
EGR1	*early growth response 1*	1.39E-15	19.48
EGR2	*early growth response 2*	4.10E-13	14.08
CXCL2	*C-X-C motif chemokine ligand 2*	1.47E-14	10.03
RSRP1	*arginine and serine rich protein 1*	5.08E-35	9.63
EGR3	*early growth response 3*	7.06E-12	7.59
JUNB	*JunB proto-oncogene, AP-1 transcription factor subunit*	4.10E-13	6.21
NR4A2	*nuclear receptor subfamily 4 group A member 2*	3.44E-14	6.04
SOCS3	*suppressor of cytokine signaling 3*	2.82E-04	5.00
GZMA	*granzyme A*	2.60E-02	-1.72
CD247	*CD247 molecule*	2.14E-02	-1.72
PPP2R3B	*protein phosphatase 2 regulatory subunit beta*	7.66E-01	-1.75
PTPRC	*protein tyrosine phosphatase, receptor type C*	2.52E-02	-1.75
CTSW	*cathepsin W*	9.82E-03	-1.79
PLAC8	*placenta specific 8*	4.07E-03	-1.85
USP17L16P	*ubiquitin specific peptidase 17-like family member 16, pseudogene*	2.95E-01	-1.89
KG7	*natural killer cell granule protein 7*	9.61E-06	-1.92
GZMB	*granzyme B*	5.93E-04	-1.96
ETFB	*electron transfer flavoprotein beta subunit*	4.81E-02	-2.08

FC: fold change, FDR: False discovery rate

Table 6-2: *Top 10 up and downregulated genes in the three-hour post exercise time point (P3) ranked by fold change*

	P3-PRE		
SYMBOL	**GENENAME**	**FDR**	**FC**
NR4A3	*nuclear receptor subfamily 4 group A member 3*	3.06E-40	198.46
ANGPTL4	*angiopoietin like 4*	1.29E-28	43.57
MT1A	*metallothionein 1A*	1.52E-19	29.64
SOCS3	*suppressor of cytokine signaling 3*	1.77E-17	25.62
FOSB	*FosB proto-oncogene, AP-1 transcription factor subunit*	6.42E-11	21.44
MYC	*MYC proto-oncogene, bHLH transcription factor*	1.20E-23	16.56
ANKRD1	*ankyrin repeat domain 1*	2.42E-14	16.54
EGR1	*early growth response 1*	4.50E-16	16.49
CHAC1	*ChaC glutathione specific gamma-glutamylcyclotransferase 1*	5.55E-17	15.07
FOS	*Fos proto-oncogene, AP-1 transcription factor subunit*	1.51E-11	13.97
CC2D2A	*coiled-coil and C2 domain containing 2A*	4.20E-12	-2.22
ARID5B	*AT-rich interaction domain 5B*	1.46E-22	-2.27
TET1	*tet methylcytosine dioxygenase 1*	5.08E-23	-2.38
STK11IP	*serine/threonine kinase 11 interacting protein*	1.59E-03	-2.44
TRIM2	*tripartite motif containing 2*	6.01E-09	-2.50
HS3ST5	*heparan sulfate-glucosamine 3-sulfotransferase 5*	4.54E-20	-2.50
HMGCLL1	*3-hydroxymethyl-3-methylglutaryl-CoA lyase like 1*	4.27E-17	-2.56
CRHBP	*corticotropin releasing hormone binding protein*	5.60E-11	-2.56
MSTN	*myostatin*	3.11E-20	-3.13
NR1D1	*nuclear receptor subfamily 1 group D member 1*	2.05E-28	-3.57

FC: fold change, FDR: False discovery rate

Table 6-3: Top 10 up and downregulated genes in the four-week exercise time point ranked by fold change

4WP-PRE			
SYMBOL	**GENENAME**	**FDR**	**FC**
TNC	*tenascin C*	2.05E-04	4.56
ANKRD1	*ankyrin repeat domain 1*	1.66E-04	4.19
HIST1H2BO	*histone cluster 1 H2B family member o*	1.88E-07	3.71
HIST1H3B	*histone cluster 1 H3 family member b*	2.48E-06	3.52
TYMS	*thymidylate synthetase*	6.96E-06	3.50
MYL4	*myosin light chain 4*	5.22E-05	3.47
POSTN	*periostin*	1.96E-03	3.36
HIST1H3J	*histone cluster 1 H3 family member j*	1.78E-04	3.34
HIST1H2AM	*histone cluster 1 H2A family member m*	8.27E-07	3.26
HIST1H2AH	*histone cluster 1 H2A family member h*	2.15E-08	3.25
METTL12	*methyltransferase like 12*	1.49E-03	-1.64
SLC6A20	*solute carrier family 6 member 20*	6.70E-04	-1.67
PDAP1	*PDGFA associated protein 1*	1.07E-01	-1.67
LRRC3B	*leucine rich repeat containing 3B*	2.62E-02	-1.67
LINC01091	*long intergenic non-protein coding RNA 1091*	2.85E-03	-1.67
CALML6	*calmodulin like 6*	7.60E-03	-1.79
ADIPOQ	*adiponectin, C1Q and collagen domain containing*	1.55E-01	-1.82
PPDPFL	*pancreatic progenitor cell differentiation and proliferation factor like*	4.91E-04	-1.85
PLIN1	*perilipin 1*	1.20E-02	-1.92
PPP2R3B	*protein phosphatase 2 regulatory subunit B''beta*	9.23E-02	-2.56

FC: fold change, FDR: False discovery rate

6.4.3 Known Exercise & Muscle Specific Genes

The next objective was to establish a link between the data from this study with the commonly associated exercise genes reported within the literature (*ACTN3*, *MSTN*, *CKM*, *PPARGC1α*, *MYC*, and *HIF1α*). Firstly, these genes were examined at each time point to confirm any change in these genes in transient post exercise response. Secondly, a recent meta-analysis of exercise transcriptomics implicated the *NR4A3*, *KLHL40*, and *DNAJA4* genes as significantly differentially regulated in immediate post exercise skeletal muscle for a number of different exercise types (endurance versus strength, HIIT versus prolonged, long-term training versus transient) (124).

The *MSTN* gene was one of the top 10 downregulated genes within the three-hour post exercise time point and has already been discussed in **Section 6.4.2**. The expression of the *ACTN3, MSTN* and *PGC1a*, genes over the different exercise time points is shown in **Figure 6-10A, B, and D,** respectively. We did not observe a change in *ACTN3* expression across exercise time points, which was consistent with the literature. Previous studies have identified a dosage effect for the *ACTN3* gene; however it has been theorised that the expression profile of *ACTN2* compensates for the lack of ACTN3 due to the premature termination induced by the p.R577X (rs1815739) polymorphism (87).

Plasma levels of CKM have been directly correlated with exercise induced muscle damage and therefore a change in expression for this transcript was unexpected within an endurance phenotype (306). Interestingly, *CKMM* was found to be significantly downregulated at both the three-hour (logFC: -0.26) and four-week (logFC: -0.31) time points but did not change following an immediate bout of exercise (**Figure 6-10C**). This confirmed previous findings for a single bout of exercise training showing no immediate effect, however the downregulation at prolonged time points indicates a change from muscle damage response to continued muscular function.

Muscle overload has been previously shown to induce *MYC* expression and as a result induce MYC dependent pathway activation such as via STAT3 and epigenetic modifying mechanisms. This study found *MYC* expression was indeed elevated following early exercise as a notable upregulation was seen at the P0 (logFC: 1.58) and P3 (logFC: 4.05) time points

(**Figure 6-10E**). This expression dropped back to baseline levels (logFC: -0.16) at the four-week time point indicating *MYC* expression to be transient post exercise.

HIF1α is a transcription factor activated through oxidative stress that induces multiple molecular pathways including stress response, angiogenic processes, epigenetic regulation through hypomethylation and activation of epigenetic modifying proteins (307-309). As HIF1*α* is activated in hypoxic environments and may lead to oxidative stress, it was expected to be differentially expressed in the Gene SMART cohort. Interestingly, the expression of *HIF1α* was found to be upregulated (logFC: 0.72) at three hours post-exercise as expected, however it was also shown to remain upregulated (logFC: 0.43) following four weeks of HIIT (**Figure 6-10I**). This showed that the molecular pathways chiefly mediated by HIF1α may have played a role in prolonged response to HIIT. Further studies should examine the functional role of HIF1*α* in chronic HIIT time points.

Our data also support the findings of *Pillon et al's* meta-analysis with respect to the nuclear receptor *NR4A3* (124), specifically *NR4A3* was the most highly expressed gene (logFC: 7.63) at the three-hour post exercise time point. Changes in the *KLHL40* (logFC: 1.05) and *DNAJA4* (logFC: 0.60) genes were also confirmed at the three-hour time point. These gene-level changes support previous findings in exercise transcriptomics for early changes to exercise training and in particular, HIIT.

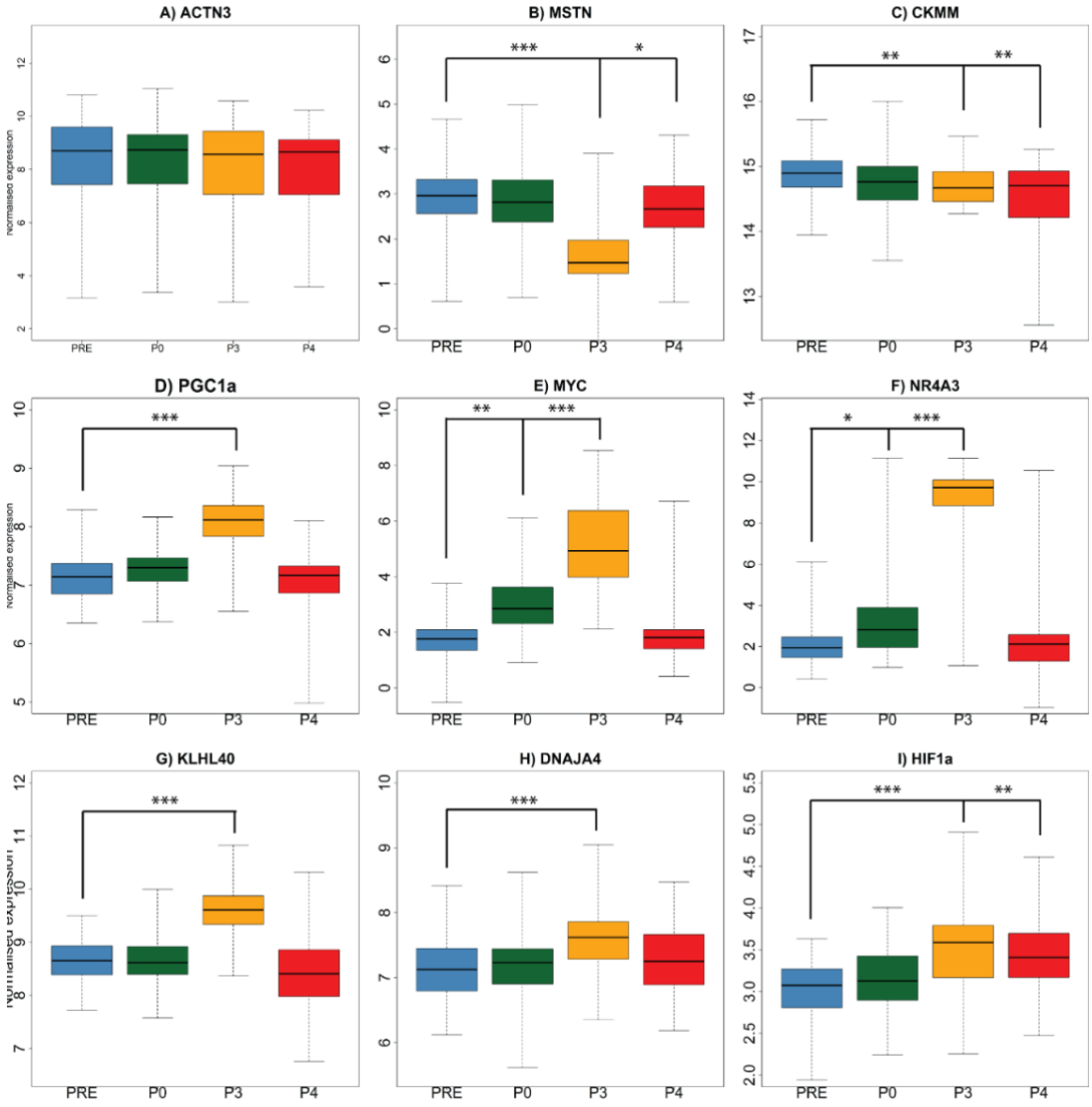


Figure 6-10: *Expression of known exercise or muscle specific genes across each exercise time point. Y-axis represents normalised expression values across samples, X-axis represents time point and individual groups are coloured separately (PRE: blue, P0: dark green, P3: orange, 4WP: red). Asterisks between time points represent varying levels of statistically significant changes from the PRE time point (*: FDR<0.01, **: FDR<0.001, ***: FDR<0.0001).*

6.5 RESULTS AND DISCUSSION: DIFFERETIALLY REGULATED BIOLOGICAL PROCESSES

Gene Set Enrichment Analysis is a method used to obtain information regarding coordinated gene changes within transcriptomic data sets. All 12,682 genes within the data were ranked based on their t-statistic and then assessed against the gene ontology (GO) biological process (BP) domain. As part of this, visualisation of global BP changes for each exercise time point in the form of enrichment maps was performed. The most significant BP nodes within each cluster are discussed in greater detail.

6.5.1 Immediate Biological Process Changes

We observed 62 positively and 11 negatively enriched BP nodes, to reduce redundancy in GO terms based on similar biological function, these were consolidated into 15 clusters (**Figure 6-11)**. The major themes for positive enrichment were consistent with the existing literature and encompassed RNA processing (nuclear transport, mRNA splicing) (150), angiogenesis (310), response to oxidative stress (14, 311-315), skeletal muscle cell differentiation (316), and response to peptide (48). Similarly, the major negative enrichments corresponded to protein targeting to membrane and icosanoid metabolic process, and regulation of morphogenesis of epithelium, which have been shown previously (317).

The immediate transcriptional profile to HIIE suggests that the skeletal muscle is responding to the strenuous exercise bout for each participant. As the majority of current literature focusses on the early exercise time points, these findings strongly support previous research (318-321). It does remain difficult to discern the exact molecular mechanisms at extremely early exercise time points as multiple molecular processes are activated simultaneously. It was expected that the three-hours post, and four-weeks post exercise time points were more biologically informative and important to continued cellular adaptation to exercise training.

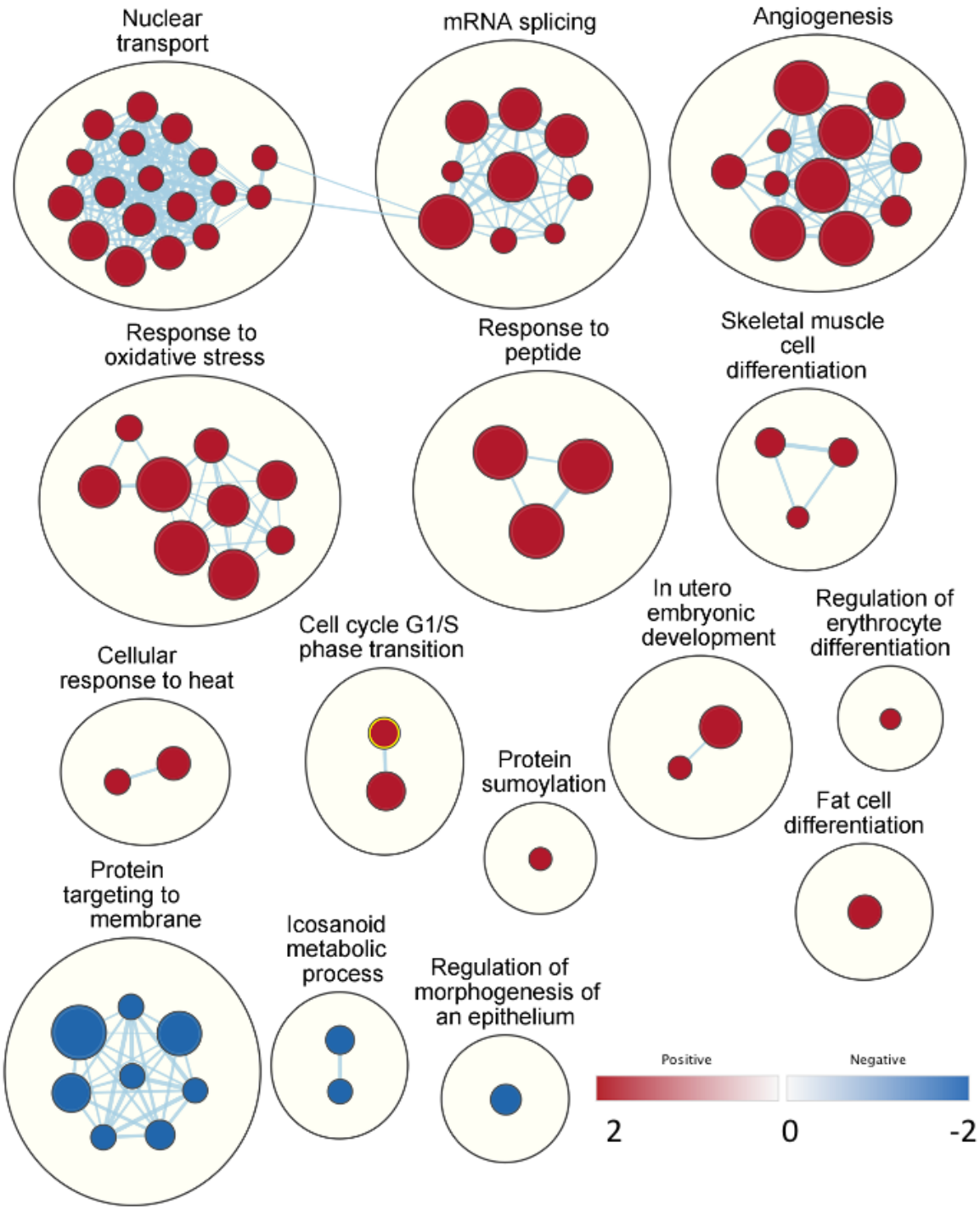


Figure 6-11: *Enrichment map showing global biological process (BP) changes immediately following high intensity interval exercise (HIIE). Each node represents a BP term that is significantly differentially regulated (FDR<0.05 and P<0.001). Jaccard overlap between nodes was set to 0.375. Size of the nodes correlate with number of genes within the BP and colours represent normalised enrichment scores (red is upregulated, and blue is downregulated). Circles encompassing the molecular nodes are an auto annotation of the larger BP of the clustered nodes.*

6.5.2 Biological Process Changes after Three Hours

At three hours post HIIE (P3), we observed 366 nodes grouped into 26 BP clusters, with no evidence of downregulated nodes (**Figure 6-13)**. The changes observed at P0 were also observed at P3 and supported by more related gene sets, indicating that the processes were initiated immediately following HIIE and built upon after three-hours. The largest BP clusters unique to P3 are detailed in the following.

Six clusters were found to be strongly linked and contributed to similar immune cell processes (leukocyte mediated immunity, regulation of protein transport, cytokine production, acute inflammatory response, response to wounding, and positive regulation of cell migration), which was consistent with existing meta-analyses (124). Within these, the six largest and most significant BP nodes (*'GO:0002443 – Leukocyte mediated immunity'*, '*GO:1904951 – Positive regulation of establishment of protein localization'; 'GO:0001816 – Cytokine production'; GO:0006954 – Inflammatory response; 'GO:0009611 – Response to wounding'; and 'GO:0042330 – Taxis'*) were significantly differentially regulated when compared with the other BP nodes within these clusters. Interestingly, the implication for genetic variants in cytokine genes within response to training within the present study has been discussed in **Chapter 4** (177). Muscle tissue may release cytokines to increase the level of leukocyte recruitment to the tissue as a result of stress induced by oxidative or contraction-based means (71, 213, 257). The cytokines released from muscle tissue during this process are aptly named myokines due to the tissue of origin/excretion. It should be noted that we were unable to determine whether the extremely high level of immune processes seen upregulated at this time point was due to an increase in immune cell extravasation into the skeletal muscle microenvironment, or due to an increase of immune gene expression from resident immune cells or myonuclei. In addition, the skeletal muscle microenvironment contains many cell types, including resident immune cells, that may have partially contributed to the immune response at acute HIIE (322).

Within the apoptosis BP cluster, the most significant node with the highest number of genes was *'GO:0043605 – Positive regulation of apoptotic process'*. This indicated that some of the molecular processes observed within the immediate exercise response were still activated. Nuclear transport was still upregulated at three hours indicating prolonged acute

functions for the BP. Reactive oxygen species (ROS) metabolism terms were upregulated, signalling a change from oxidative stress immediately post exercise, to metabolism of ROS and clearing of the stress inducing metabolites at three hours post HIIE, which was consistent with existing literature for acute exercise response (49-51). This was further supported by the continued upregulation of response to oxidative stress. In addition, there was a cluster of upregulated terms within response to hypoxia and most noticeably, the '*Response to oxygen levels (GO:0070482)*' BP was the most significant node within this cluster. Together these results indicate a global switch involving metabolism induced on oxygen level response.

Other clusters also seen in the immediate time point include angiogenesis, RNA processing, fat cell differentiation, and skeletal muscle development, which has been shown previously (126, 318). Interestingly, there was a shift from regulation of skeletal muscle development to hypertrophic processes at three-hours post HIIE. In addition, epigenetic mechanisms were starting to become evident at this time point as BP clusters involving miRNA transcription and non-coding RNA processing were positively enriched. A response to nutrient levels cluster was also evident, which was expected as energy depletion following this type of acute exercise has been shown previously (323, 324).

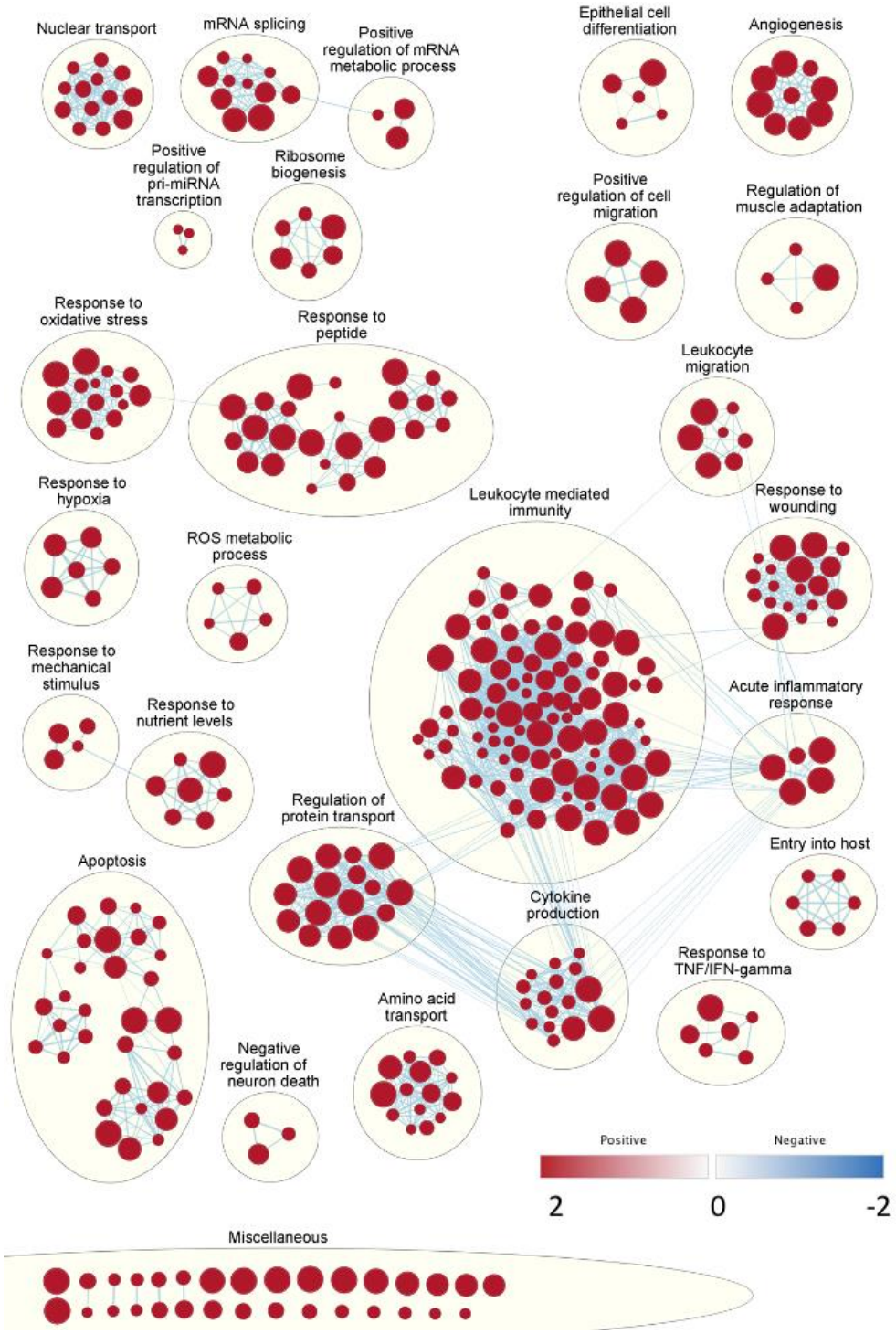


Figure 6-12: *Enrichment map showing global biological process (BP) changes following high intensity interval exercise (HIIE) with a three-hour rest period. Each node represents a BP that is significantly differentially regulated (FDR<0.05 and P<0.001). Jaccard overlap between nodes was set to 0.375. Size of the nodes correlate with number of genes within the BP and colours represent normalised enrichment scores (red is upregulated, and blue is downregulated). Circles encompassing the nodes are an auto annotation of the larger BP of the clustered nodes.*

6.5.3 Biological Process Changes after four weeks of HIIT

At four-weeks post HIIT (4WP – at rest), we observed 581 gene sets, forming 42 BP clusters according to larger biological function **(Figure 6-13)**. Similarly, to the P0 time point, the majority (79.7%) of the differentially expressed BP nodes were found to be positively enriched. Despite the lower number of significant genes in the four-week time point when compared with three hours (2,812 versus 3,939), there were more differentially regulated BP gene sets identified following four-weeks of training. As the four-week sample was taken at rest, we found that differentially regulated genes were more consistently changing in the same directions within each BP, which likely led to the larger BP term versus significant genes ratio at this time point when compared with three-hours post HIIE.

Like the acute exercise time points, leukocyte mediated immunity, angiogenesis, apoptosis, mesenchyme development, and cytokine production terms were all upregulated and formed the largest clusters within the enrichment maps. Each of these clusters contained larger numbers of BP nodes than the three-hours post HIIE (P3), again suggesting that these biological processes were initiated following a single bout of HIIE but maintained and amplified with continuous HIIE training. When analysing the transcriptomic GO terms following four-weeks of HIIT, a particularly large response in BPs changing when compared to the preliminary time point was observed. This was unexpected as the sampling was taken from the exercise participants when at rest (at least 48 hours post-exercise), compared to the other time points where muscle biopsies were taken immediately or three-hours post-exercise. Further, the four-weeks post HIIT response contained the largest number of negatively enriched GOBPs of any time point, which indicated that the BPs were more likely changing for adaptation in response to training rather than a shock response to oxidative stress, or skeletal muscle remodelling.

As there were many positively enriched clusters, an effort was made to summarise the larger biological function of each cluster and the implications for exercise response. Firstly, it was noted that extracellular matrix processes (ECM organisation, cell-matrix adhesion, and glycosaminoglycan metabolic process) were enhanced at four-weeks post HIIT (4WP). Specifically, the '*Glycosaminoglycan catabolic process (GO:0006027)*' and '*Collagen metabolic process (GO:0032963)*' nodes were markedly upregulated when compared to other nodes within their respective groups.

A similar study in which participants performed 6-weeks of HIIT (muscle biopsy taken at rest 48h following) identified upregulation of glucose metabolism, ECM organisation, angiogenesis, and mitochondrial membrane (126). We observed similar ECM and angiogenesis processes within our HIIT data, however glucose metabolism was not enriched, and mitochondrial processes were identified to be downregulated in response to HIIT. Downregulated terms included contractile fibre, regulation of synaptic transmission, mitochondrial matrix, and cytoskeletal binding. We observed similar patterns in skeletal muscle contraction and mitochondrial organisation GO terms, however, no such observation was identified for synaptic transmission, and actin cytoskeleton organisation was positively enriched in our data. Of note, the authors used only upregulated genes to inform positively enriched GO terms and downregulated genes to inform negatively enriched pathways, with no indication of whether these genes were in similar biological processes. In contrast, we performed GO in unison with gene set enrichment to discover significant biological processes regardless of gene level biases.

Negatively enriched BPs were found to be associated with altered gene expression, histone modification, ubiquitin-dependant protein catabolic process, and ATP metabolism. Interestingly, mRNA splicing, gene expression, and mitochondrial processes (mitochondrial organisation, mitochondrial translation) were all negatively enriched BP clusters, indicating a large molecular change from acute exercise response (upregulations in response to oxidative stress, large scale increases in translational processes) to downregulation of these processes as the microenvironment adapts to HIIT. Specifically, the *'Mitochondrial translational termination (GO:0070126)', 'Protein acylation (GO:0043543)', 'rRNA processing (GO:0006364)', 'Lipid oxidation (GO:0034440)', 'Interleukin-1-mediated signalling pathway (GO:0070498)', Mitochondrial electron transport, NADH to ubiquinone (GO:0006120)', 'Transcription-coupled nucleotide-excision repair (GO:0006283)',* and *'Striated muscle contraction (GO:0006941)'* BP nodes were the most significant nodes within their respective clusters.

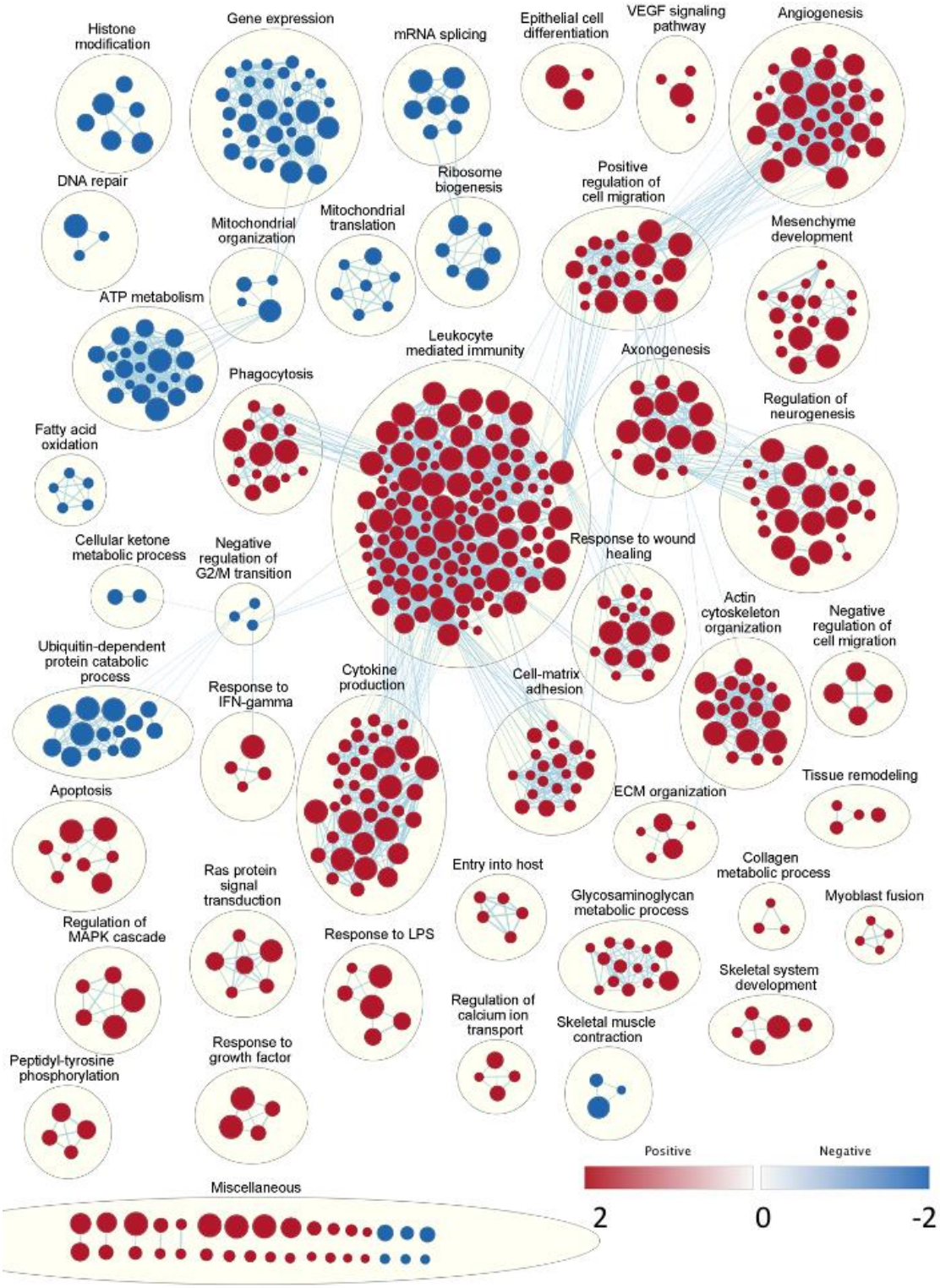


Figure 6-12: *Enrichment map showing global biological process (BP) changes in response to four weeks of HIIT training (4WP). Each node represents a BP that is significantly differentially regulated (FDR<0.05 and P<0.001). Size of the dots correlate with number of genes within the GO term and colours represent normalised enrichment scores (red is upregulated, and blue is downregulated). Nodes without connections were classed as miscellaneous regardless of size or level of significance.*

6.6 RESULTS AND DISCUSSION: PATHWAYS OF INTEREST

Due to the large-scale nature of this transcriptomic study, we elected to select specific biological process clusters for in depth, gene level, analysis. These three key processes (cytokine production, mitochondrial regulation, and chromatin modification) were selected from the significant pathway clusters at four-weeks post HIIT (4WP). Briefly, this section aimed to determine gene level changes particular to sustained HIIT, and so acute responses to HIIE were discussed only minimally. In some instances, the findings from these analyses were reminiscent of earlier chapters within the thesis framework, which has been clearly stated within the relevant sections.

6.6.1 Biological process changes pertaining to cytokine production and regulation

To discern the importance of cytokines and immune processes on exercise training, the cytokine BPs were extracted from the four-week time point global dataset. '*Regulation of cytokine production'* (GO: 0001817) was the most statistically significant GO term within this cluster and was selected for GeneMANIA analysis across exercise time points **(Figure 6-14)**.

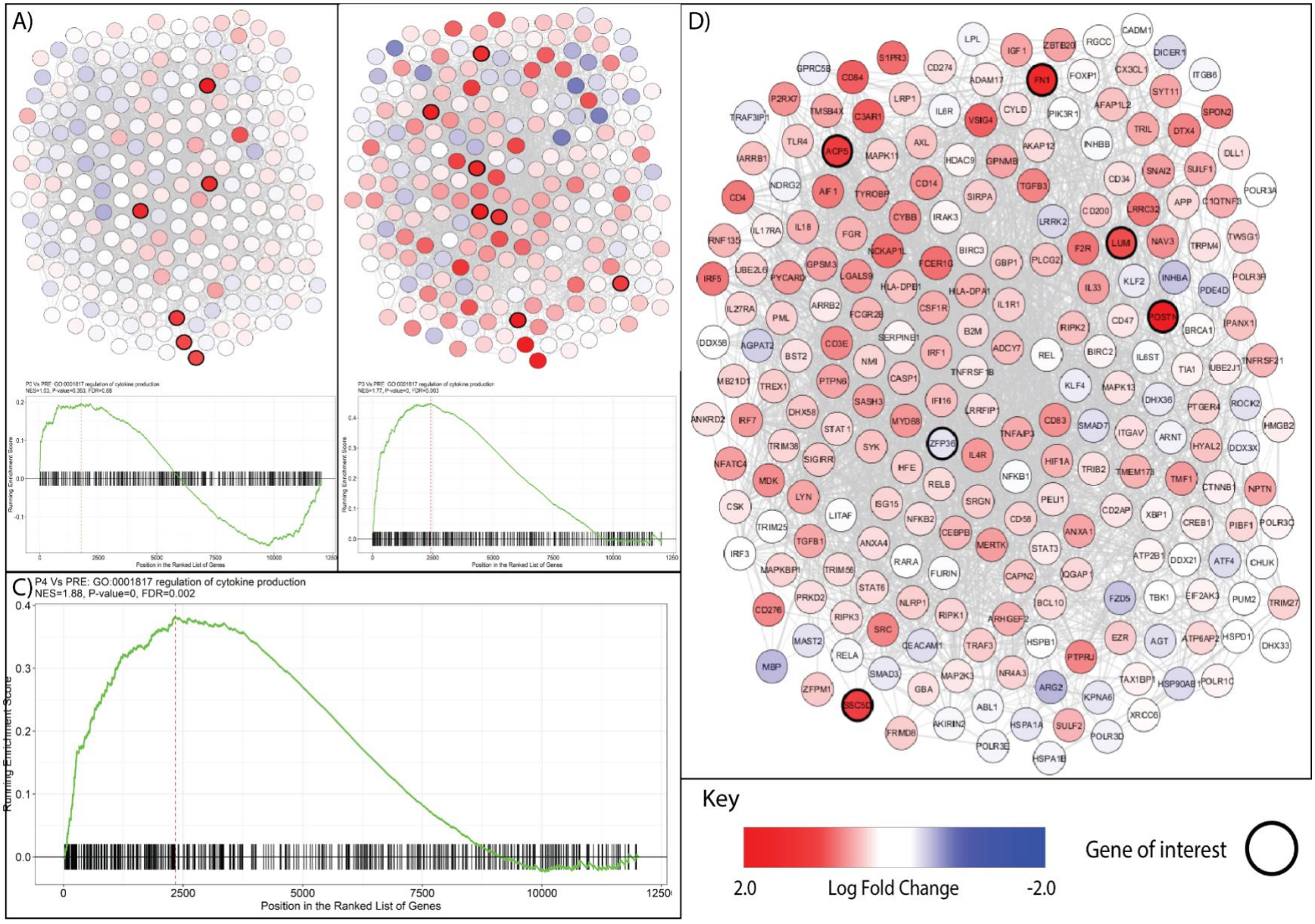

A)
C)
P4 Vs PRE: GO:0001817 regulation of cytokine production
NES=1.88, P-value=0, FDR=0.002
Running Enrichment Score
Position in the Ranked List of Genes
D)
Key
2.0
Log Fold Change
-2.0
Gene of interest

Figure 6-13: *Gene Set Enrichment Analysis of 'Regulation of cytokine production (GO:0001817)'.* ***A)*** *Co-expression network plot and GSEA plot for P0,* ***B)*** *Co-expression network plot and GSEA plot for P3,* ***C)*** *GSEA plot for 4WP, and* ***D)*** *Co-expression network for 4WP. Nodes correspond to individual genes (FDR<0.05, p<0.001). Edge lines between two genes represent a co-expression relationship. Colour intensity shows differential regulation (red= upregulated, blue=downregulated) and black borders highlight genes of interest within each co-expression network. For each GSEA plot, vertical black lines (barcodes) indicate the position of each gene. A running enrichment score is depicted by the green line. The enrichment score with the maximum deviation from zero is shown by the vertical red dotted line (NES).*

Post HIIE (P0)

Firstly, the differentially regulated genes immediately post HIIE were also differentially regulated after three-hours. A clear distinction emerged in genes that were differentially regulated in the BP between the acute exercise time points and the four-week post time points. At the P0 time point, seven genes were found to be highly upregulated in response to acute exercise including: *Heat Shock Protein Family A Member 1B* (*HSPA1B: FC=2.2*), *Heat Shock Protein Family A Member 1A* (*HSPA1A: FC=2.5*), *Nuclear Receptor Subfamily 4 Group A Member 3* (*NR4A3: FC=2.5*), *ZFP36 Ring Finger Protein* (*ZFP36: FC=2.8), Kruppel Like Factor 4* (*KLF4: FC=2.4*), and *Early Growth Response 1* (*EGR1: FC=22.8*).

Zinc finger proteins are induced through cytokine signalling and represent a rapid response to inflammation. Their purpose is to halt the expression of inflammatory markers through the binding of cytokinetic mRNA (325, 326). Most notably, the ZFP36 protein has been shown to influence *interleukin 6* (*IL-6*) expression (327). Additionally, the strong influence of this gene on the negative regulation of *IL-6* expression, in conjunction with the *IL-6* genetic variant (rs1474347) identified in **Chapter 4** of this thesis, may have contributed to the apparent lack of overall IL-6 expression within the exercise study. This possible mechanism should be further explored within functional exercise studies.

KLF4 is a transcriptional activator that has been strongly associated with ageing and cancer as the protein mediates telomerase expression (328). The role of *KLF4* in smooth muscle has been extensively studied. For instance, expression of *KLF4* is directly correlated with the dedifferentiation of smooth muscle cells (329). Further, *KLF4* expression has been shown to spike following vascular damage (330). Interestingly, *KLF4* has not been previously implicated

in skeletal muscle following exercise training. Instead, a similar KLF family member (KLF3) has been shown to interact directly with the *CK-MM* promoter to influence muscle differentiation (331).

Three-hours post HIIE (P3)

At the three-hour time point, 20 genes within the GO:0001817 *(Regulation of cytokine production)* process were shown to be highly upregulated in response to exercise. This was observed in the gene set enrichment plot for this time point with the enrichment score at the upper end of the ranked gene list (**Figure 6-15B**). Most notably, *CD14 Antigen* (*CD14: FC=2.33)*, *Interleukin 6 Receptor* (*IL6R: FC=3.09*), *Interleukin 4 Receptor (IL4R: FC=2.30), NR4A3: FC=127.60, Interferon Regulatory Factor 1* (*IRF1: FC=2.76*), and *Serpin Family E Member 1* (*SERPINE1: FC=2.83*) genes were strongly upregulated. Due to the size of this chapter with respect to the thesis document, only the specific functions of key genes as they pertain to exercise training are discussed further.

The IL6R receptor functions similarly to IL4R as outlined previously, however interactions between T-helper 1 (Th1) immune cells, macrophages, and muscle tissue require IL-6 signalling for proliferation rather than differentiation (332). Interestingly, and as described previously in **Chapter 4**, a SNP in the promoter region of the *IL6* gene was found to be significantly associated with lower performance in this study. *ZFP36* negatively regulates the expression of *IL6* as an early response to exercise. Interestingly, the expression of *ZFP36* was attenuated at three-hours post-exercise supporting the data that IL-6 has a role at three hours to influence the proliferation of skeletal muscle following HIIE. The *F2R* gene contributes to coagulation, inflammation, and angiogenesis, and therefore plays an important role in early exercise time points and has been shown to induce the expression of *IL-6* in endothelial cells (333). A recent transcriptomic study was able to identify the *F2R* gene as being upregulated following 12 weeks of training (257), replicated in this study after four-weeks and provides confidence in the involvement of this process within the current study.

The SERPIN family of proteins are commonly associated with protease activity and have been extensively studied in cancer as regulators of cell movement and metastasis (334). *SERPINE1* has not been as extensively studied in the field of exercise science, however one

study found that expression of this gene was downregulated in rats that performed exercise prior to diving (335). Further, the study found that cytokine, and cytokine receptor expression was upregulated in the exercised rats. Interestingly, expression of the *SERPINE1* gene was upregulated at three-hours following HIIE. This data is perhaps an example of the limited biological similarities between rodents and humans, in particular with large scale interventions such as exercise training.

Four-weeks post (HIIT)

At the four-week time point following HIIT, only five genes within the '*Regulation of cytokine production'* process were strongly upregulated. Both *IL4R* and *CD14* remained upregulated but to a much lesser extent to the three-hour time point. The most highly expressed genes at the four-week time point were *Periostin (POSTN: FC=2.64), Lumican (LUM: FC=1.99), Acid phosphatase 5, tartrate resistant (ACP5: FC=1.79), Scavenger receptor cysteine rich family member with 5 domains (SSC5D: FC=2.05),* and the *Fibronectin 1 (FN1: FC=2.81)* genes.

Several of the most upregulated genes following chronic HIIT (4WP) were reported to be involved in ECM remodelling (specifically collagen binding) such as *LUM*, *FN1*, and *SSC5D*. This signalled a major shift for this process from immediate responses being heavily involved in immune function and response to inflammation, to ECM remodelling for the recently remodelled skeletal muscle. The protein (Periostin) produced from the *POSTN* gene has been associated with cell adhesion and has been suggested to enhance the function of fibroblasts in different physiological states. Varying levels of gene expression for this gene have been linked to several disease states including breast cancer prognosis and asthma outcomes (336, 337). The gene has also been linked to bone mineral density and may play a role in the stimulation of Bone Morphogenetic Protein 1 (BMP1) (338). Expression levels of *POSTN* have been shown to increase in rodent models following swim-based exercise under induced aortic constriction (339). The study correlated exercise induced *POSTN* expression with a marked increase in activated cardiac fibroblasts. Interestingly, expression levels of the Periostin gene have been shown to directly influence skeletal muscle regeneration and differentiation in rats (340). Whilst this was shown in rodents and largely cardiac based, it is

feasible that the *POSTN* gene may be important for skeletal muscle remodelling after HIIT in humans.

Interestingly, the *ACP5* gene has previously only been associated with disease states involving muscle wasting. Further, rodent based studies utilising knock-out models for *dystrophin* have implicated *ACP5* downregulation when compared with healthy mice (341). The *ACP5* gene appears may be required for macrophage based inflammatory responses, indicating a role for continued immune processes even after four-weeks of training (342). In addition, a study examining the basal expression levels of *ACP5* found that skeletal muscle did not express the transcript (343). These results indicate that *ACP5* plays a role in long term exercise adaptations and is only expressed in skeletal muscle in response to large scale interventions rather than being constitutively expressed.

The gene level results obtained from the '*Regulation of cytokine production'* pathway indicated that, whilst replication of some previous findings was achieved, the *POSTN* gene was identified as the likely mediator causing the pathway to become enriched following four weeks of training. In particular, upregulation of the *POSTN* gene at this time point indicates a strong and novel trend for this gene in exercise training.

6.6.2 Molecular pathway changes pertaining to mitochondrial function

We identified multiple mitochondrial genes and molecular pathways that were differentially regulated after four-weeks of HIIT (4WP). Interestingly, the mitochondrial pathways were all downregulated with those at 4WP clustered into three BP clusters: Mitochondrial organisation, Energy metabolism, and mitochondrial gene expression. The '*Mitochondrial gene expression (GO:0140053)'* process was selected for GeneMANIA analysis as it showed a great deal of comparative differential regulation when considered amongst the other mitochondrial pathways **(Figure 6-17).**

The mitochondrial associated genes within the '*Mitochondrion gene expression (GO:0140053)'* BP was not differentially regulated immediately post exercise, however there was no change after three hours, followed by downregulation after chronic HIIT (4WP). At three-hours post exercise, there appeared to be an even number of downregulated and upregulated genes (Figure 6B), which likely contributed to the non-significant enrichment of the BP three-hour post HIIE (P3). As neither the immediate nor three-hours transcriptional response to HIIE were found to be significantly enriched in the data set, only the differentially regulated genes within the four-week time point have been examined in more detail. It is conceivable that the differentially regulated genes within the early time points may still be important within other molecular pathways following HIIE, however this is out of context of this thesis and should be the pursuit of subsequent studies.

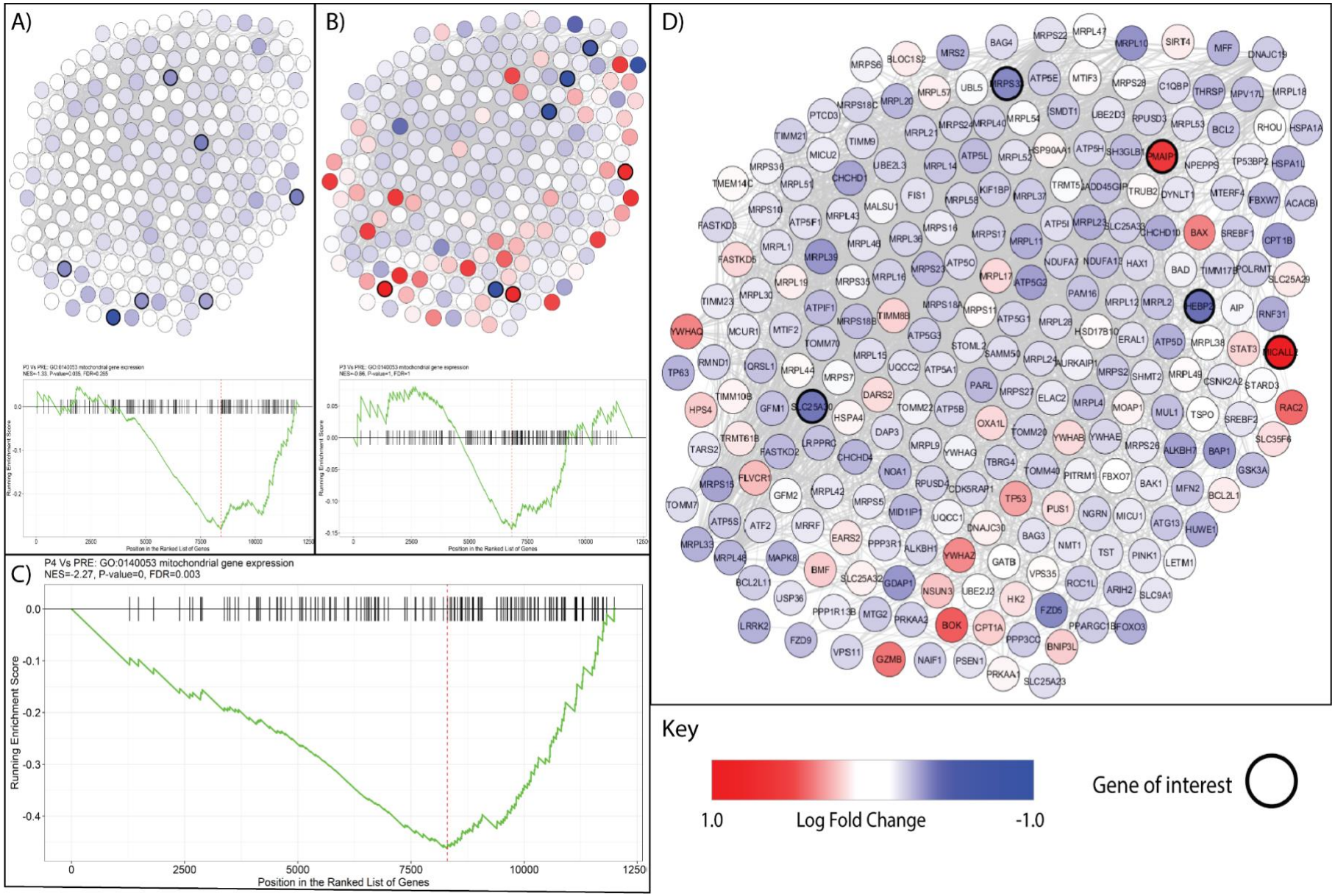

A)
B)
C)
D)
P4 Vs PRE: GO:0140053 mitochondrial gene expression
NES=-2.27, P-value=0, FDR=0.003
Running Enrichment Score
Position in the Ranked List of Genes
Key
Log Fold Change
1.0
-1.0
Gene of interest

Figure 6-14: *Gene Set Enrichment Analysis of 'Mitochondrial gene expression (GO:0140053)'.* ***A)*** *Co-expression network plot and corresponding GSEA plot at P0,* ***B)*** *Co-expression network plot and corresponding GSEA plot at P3,* ***C)*** *GSEA plot at 4WP, and* ***D)*** *Co-expression network for 4WP. Co-expression nodes correspond to individual genes that are significantly enriched (FDR<0.05, p<0.001) in each exercise time point. Colour intensity represents the magnitude of differential regulation (red= upregulated, blue=downregulated) and black borders highlight genes of interest within each co-expression network. For each GSEA plot, vertical black lines (barcodes) indicate the position of each gene within the GO:0140053 gene-set. A running enrichment score is depicted by the green line. The enrichment score with the maximum deviation from zero is shown by the vertical red dotted line (NES).*

Four-weeks post (HIIT)

It was evident at the four-week time point that the majority of genes within the mitochondrial gene expression pathway were downregulated. This contrasted with the phenotypic data, where improved lactate threshold and VO_{2peak} measures were observed following HIIT. We hypothesise that this difference was due to the sampling time utilised for each time point. For instance, the four-week time point was taken at rest following the training time frame, and so we would expect lactate clearing and aerobic respiration to be improved following acute HIIE (344, 345), but lower mitochondrial function required at rest, leading to the apparent negative enrichment of mitochondrial processes following HIIT. Moreover, mitochondrial function following HIIT has been shown to differ between skeletal muscle types (346).

The '*Mitochondrial gene expression'* biological process had a negative enrichment score and was strongly downregulated with five significantly enriched genes identified; the three downregulated genes were *Solute carrier family 25 member 30* (*SLC25A30*), *Mitochondrial ribosomal protein S33* (*MRPS33*), *Heme binding protein 2* (*HEBP2*), the two upregulated genes were *MICAL like 2* (*MICAL-L2*), and *Phorbol-12-myristate-13-acetate-induced-protein 1* (*PMAIP1*).

The SLC family of transporter molecules assist with the modulation of signalling molecules for appropriate molecular pathway function in response to extracellular intervention (347). Interestingly, downregulated expression of the *SLC25A30* gene has been associated with old *vs* young skeletal muscle murine studies (348). As this study focussed on basal expression and not exercise intervention, it is possible that the downregulation of the *SLC25A30* gene following exercise may assist with long-term adaptations to strenuous exercise. The exact function of the *MRPS33* gene is unclear, however previous studies have found evidence of its upregulation following irradiation (349). Therefore, this gene may have a role in modulating stress response following acute bouts of training and therefore the downregulation following four weeks of training would be expected.

The upregulated genes at this time point included *MICAL-L2* and *PMAIP1*, both associated with insulin stimulated pathways within muscle. *MICAL-L2* is a cytoskeletal protein involved in tight junction formation and a recent study has implicated the protein in insulin activation of the Rab13 protein in muscle cells (350). Briefly, Rab13 binding changes the conformation of the Mical-L2 protein and allows binding of ACTN4. The GLUT4 protein is able to tether to this complex, enabling vesicle fusion and import (351). The strong upregulation of this transcript at the four-week time point may have led to continued Insulin-based signalling and enabled skeletal muscle remodelling and hormone-based function. The PMAIP1 protein has been previously strongly associated with cell mediated apoptosis (352). Interestingly, *PMAIP1* upregulation was seen in skeletal muscle of individuals with T2DM, in conjunction with a host of insulin resistance, amino acid metabolism, and immune response genes (353). Taken together, these data indicate that insulin signalling remains an important result of exercise induced adaptations at longer exercise time points following HIIT.

6.6.3 Molecular pathway changes pertaining to epigenetic regulation

Next the molecular mechanisms by which epigenetic changes may occur within exercise response was assessed. Of the plotted GO terms, the *'Macromolecule methylation (GO:0043414)*' pathway node was the found to be the most significant with the largest gene-set size and was selected for GeneMANIA analysis.

Firstly, and most notably, the majority of the genes within the *'macromolecule methylation*' process changed over time in response to exercise. As such, the genes with a significant log fold change within this pathway (±1.1) were the focus of more detailed analysis. The function of each epigenetic modifying protein was also considered separately as many epigenetic functions occur in unison to perform global changes to cellular function.

Post HIIE (P0)/ Three-hours post HIIE (P3)

Similar to the mitochondrial pathway outlined previously, the *'Macromolecule methylation'* process contained approximately equivalent numbers of downregulated and upregulated genes at acute exercise time points. As such, the pathway was identified to be solely enriched after chronic HIIT (4WP).

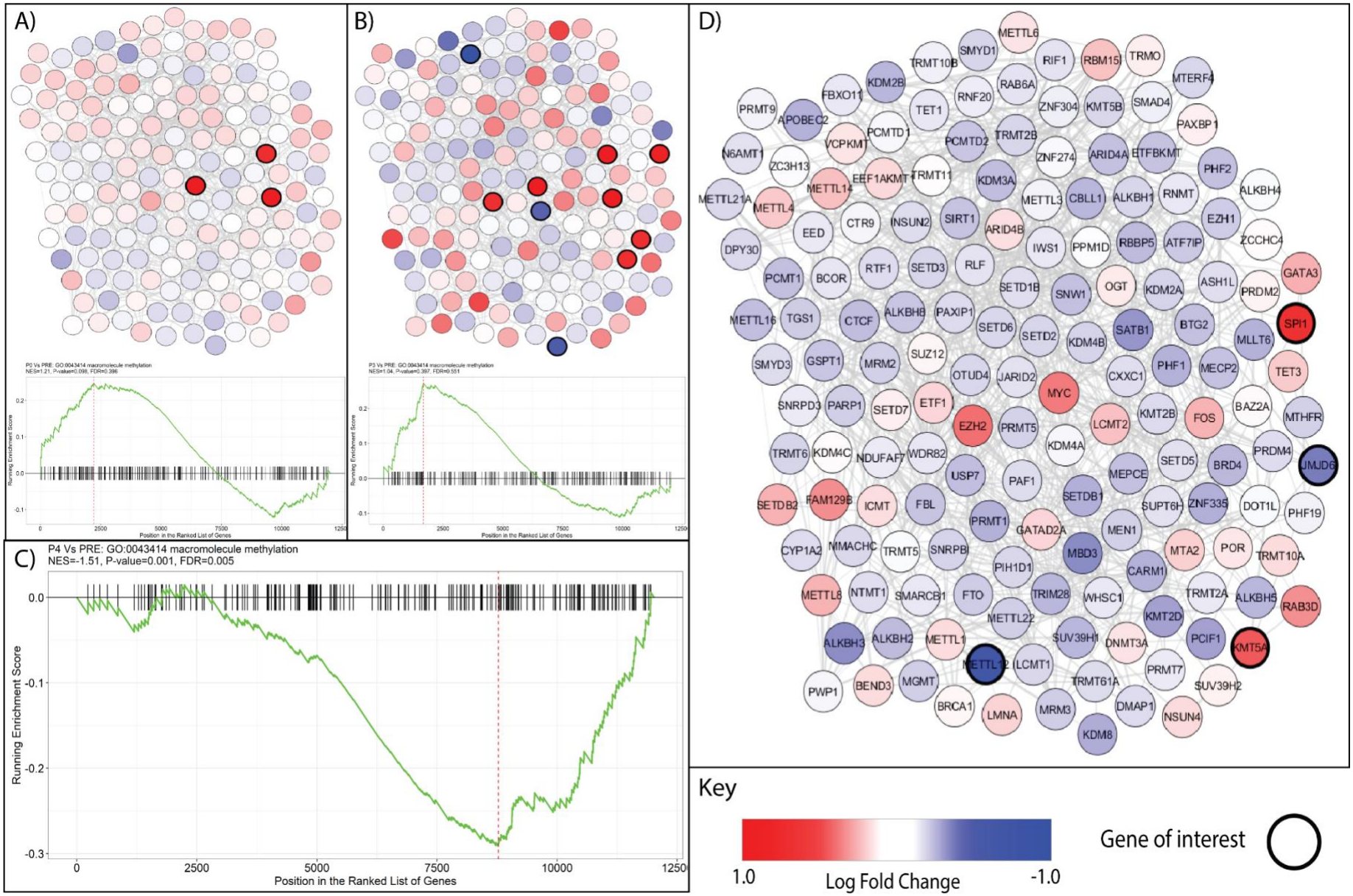

A)
B)
C)
P4 Vs PRE: GO:0043414 macromolecule methylation
NES=-1.51, P-value=0.001, FDR=0.005
Running Enrichment Score
Position in the Ranked List of Genes
D)
Key
1.0
Log Fold Change
-1.0
Gene of interest

Figure 6-15: *Gene Set Enrichment Analysis of the 'Macromolecule methylation (GO:0043414)' gene-set.* ***A)*** *Co-expression network plot and corresponding GSEA plot at P0,* ***B)*** *Co-expression network plot and GSEA plot for at P3,* ***C)*** *GSEA plot at 4WP, and* ***D)*** *Co-expression network at 4WP. Co-expression nodes correspond to individual genes that are significantly enriched (FDR<0.05, p<0.001). Colour intensity represents the magnitude of differential regulation (red= upregulated, blue=downregulated) and black borders highlight genes of interest within each co-expression network. For each GSEA plot, vertical black lines (barcodes) indicate the position of each gene within the GO:0043414 gene-set. A running enrichment score is depicted by the green line. The enrichment score with the maximum deviation from zero is shown by the vertical red dotted line (NES).*

Four-weeks post (HIIT)

At the four-week post exercise time point, the majority of the genes within this pathway were found to be non-significantly downregulated correlating with the negative enrichment score for this pathway. The two upregulated genes were identified as *SPI1* and *Lysine methyltransferase 5A* (*KMT5A*). The genes showing the largest level of downregulation were *METTL12*, and *JMJD6*. Interestingly, the expression of *BTG2* was appended at four-weeks indicating that it is required for epigenetic remodelling at early exercise time points, but different epigenetic processes seem to be required for adaptation to extended training.

The KMT5A protein is a histone specific demethylase protein that specifically regulates the monomethylation of lysine 20 on histone H4. KMT5A has been shown to heavily influence the proliferation of myocytes. Further, miR-127-3p has recently been shown to attenuate these effects in rodent models (354). This interaction was unable to be assessed within the current study due to the targeted nature of the RNA sequencing chemistry utilised. The *SPI1* gene has previously been studied within exercise genetics and was found to strongly influence immune function following exercise training (355, 356). This represents an interesting mechanism by which epigenetic regulation at early time points may lead to the high expression *SPI1*, and therefore immune adaptation to exercise training.

The *METTL12* gene has not yet been explored within the context of exercise training however the magnitude of downregulation at four-weeks indicates a mechanism for response to HIIT. Interestingly, this gene has been characterised as a mitochondrial methyltransferase

that specifically methylates the well-known mitochondrial energy protein Citrate Synthase (357). As such, this represents a novel and fascinating mechanism by which training may modulate adaptations to energy production and homeostasis. The JMJD6 epigenetic modifying protein catalyses the demethylation of histone protein H3 at Arg2 and histone protein H4 at Arg3. Expression changes of the *JMJD6* gene have been seen in response to endurance exercise in Alaskan sled dogs (358). The time course within the study extended to 24 hours however the slight upregulation described in the dogs was also evident in our study.

6.7 SUMMARY

Unravelling the molecular landscape of exercise training with transcriptomic studies has been difficult for many reasons. Primarily, small scale exercise studies and extremely targeted gene expression analysis have resulted in few (but strong and replicated) identification of differentially regulated genes and biological processes. Namely, PGC1α and HIF1α inducible BPs showed consistent replication, particularly across exercise studies where high intensity exercise was the primary form of training. Larger scale exercise studies focusing on transcriptomics are starting to become popular within the sports medicine field however there remains a need for replication and in-depth analysis of next generation sequencing data.

In the present study targeted RNA sequencing was utilised to sequence the transcriptomes of 55 exercise participants at different exercise time points, equating to a total of 204 transcriptome samples. Innovative analysis methodologies were used to explore specific gene level changes over different exercise time points and revealed a total of 5,858 differentially regulated genes across the entire data set with the most differentially regulated genes seen within the three-hour post HIIE time point.

Initially, the similarity between biological functions of the most differentially regulated genes in response to training was assessed within the study cohort at the different exercise time points. Well known transcription factors such as *FOS, MYC*, and *JUN* were highly upregulated immediately following a single bout of HIIE. At three hours, transcription factors remained upregulated however there were also large-scale changes in immune genes. The downregulated genes at this time point were largely from epigenetic modifying proteins. Further, the *MSTN* gene was downregulated at three-hours indicating a possible mechanism for skeletal muscle hypertrophy. Examination of the genes differentially regulated at the four-week time point (following HIIT), found histone protein subunit genes to be highly upregulated, indicating a large chromatin remodelling event.

Next, examination of the expression of previously discovered exercise response genes replicated previous findings and found most transcripts (*MSTN*, *CK-MM, PGC1α, MYC, NR4A3, HIF1α, DNAJA4, KLHL40*) to be significantly changed at three hours, indicating that most previous research was focussed on transient exercise responses. Further, the replication

of these results in this study provided support for the findings in this study and that the data obtained is also likely applicable outside of the current study and to the broader exercise field.

Following differential expression analysis enrichment in molecular pathways in each of the time points was assessed. Gene Set Enrichment Analysis is able to assess small but coordinated changes in expression within individual gene-sets relating to biological processes and pathways was used to discern biological meaning of the transcriptomic data. This analysis identified 73 GOBPs differentially regulated in the immediate time point, 366 in the three-hour time point, and 581 at four-weeks post exercise. The immediate pathway changes suggested stress response and immune mediated inflammation processes, consistent with the current literature (14, 308, 311-313). The GO terms enriched at the three-hour time point were shown to be predominantly immunity based, however angiogenic and protein metabolism were also evident. Interestingly, the number of enriched processes was larger at the four-week mark, despite the number of differentially regulated genes being fewer than three-hour HIIE. This was likely due to the coordinated nature of GSEA. Immune response GO terms were still evident following chronic HIIT (4WP), however BPs involved in epigenetic remodelling, mitochondrial function, and development were also identified. GO terms involving glucose metabolism, extracellular matrix organisation, angiogenesis, and mitochondrial membrane processes have been previously identified as upregulated following six-weeks of low volume HIIT (126).

Following this, a link between the transcriptomic data, and the genetic results seen in the previous chapters within this thesis were considered. To do this, cytokine, mitochondrial, and epigenetic/methylation pathways were extracted from the bulk dataset in order to analyse the most significant pathway within each group. GeneMANIA was used to determine the genes of interest within each process at the separate time points.

Gene sets involving cytokine production and usage were highly enriched in both the three-hour and four-week exercise time points. In **Chapter 4**, a SNP proximal to the *IL6* gene was significantly associated and negatively correlated with higher VO_{2peak} threshold measurements within the study cohort. Interestingly, there was no evidence of *IL6* expression within the cohort, indicating a probable repression of the gene, either due to the SNP, or

increased prevalence of other myokine mechanisms within the context of response to training. The *ZFP36* gene was found to be upregulated in response to training, a negative regulator of *IL6* expression at the mRNA level, also representing a possible mechanism by which repression of *IL6* may contribute to adaptations to exercise in both short and long term HIIT.

In **Chapter 5** of this thesis, several genetic markers were found to be associated with exercise response. Specifically, an intronic SNP (rs2041840) within the *NDUFAF7* gene was shown to be significantly associated with better responses within the VO_{2max}, peak power, and composite trait response phenotypes. When the enriched molecular pathways pertaining to mitochondrial function was assessed, a negative enrichment and therefore downregulation of mitochondrial biological function at the four-week exercise time point was observed. Interestingly, the levels of *NDUFAF7* did not change over the exercise time points however the differences between expression based on genotype was not specifically assessed. From the GSEA, the *'Mitochondrial gene expression'* process was found to be the most significantly enriched at the four-week time point with the largest gene set size. Of note the most enriched genes within this pathway at the four-week time point were found to be associated with Insulin signalling through the *MICAL-L2* gene. Briefly, the product of this gene forms a complex with the Rab13 protein, which allows for the binding of ACTN2 in an insulin dependent manner. As stated previously, ACTN2 expression may rescue the lack of functional ACTN3 due to the R577X genotype.

Identifying links between epigenetics and transcriptomics remains elusive, particularly in cases with strong interventions such as exercise training. The focus of this study was on the mechanisms by which methylation signatures may occur in response to training over the exercise time points examined. This approach identified *'Macromolecule methylation'* gene set to be highly enriched at the four-week time point. The well-known transcription factors *MYC*, *FOS*, and *BTG2* were highly upregulated following a single bout of training at both the immediate and three-hour time points. In addition, large scale changes in gene expression within the three-hour time point even though the molecular pathway was non-significant at this time point. There also appeared to be highly upregulated histone protein methyltransferases, in conjunction with extremely downregulated levels of histone demethylase proteins. In particular, the *EZH2* gene was proposed as a central epigenetic mechanism of early stress

response pathways. Interestingly, similar processes were noted in the four-week time point although different methyltransferase and demethylation genes appear to be involved. The differentially regulated genes within this pathway at four-weeks have previously been largely uncharacterised in exercise training. Based on these observations, Chapter 7 will further explore the epigenetic response to training through assessment of global DNA methylation, which will provide a novel and informative link between gene expression and DNA methylation for acute and chronic HIIT. These findings represent novel mechanisms for epigenetic regulation of prolonged adaptation to training and should be explored in more detail in future studies.

Within this chapter, differentially regulated gene expression, pathway analysis using gene set enrichment, and results pertaining to previous genetic findings over longitudinal exercise time points were explored. Briefly, we identified similar gene expression profiles and biological processes to existing studies for acute HIIE (P0, P3). In particular, BPs associated with acute HIIE showed the skeletal muscle microenvironment was responding to oxidative stress. The novel aspect of our findings showed that large-scale gene expression and BP changes were maintained in chronic HIIT (4WP). BPs associated with chronic HIIT suggested that the skeletal muscle microenvironment was remodelling the ECM, and amplifying protein localisations processes for continued adaptation. It should be noted that inter-time point comparisons were out of context within the framework of the current thesis, however they should be explored further in future studies. Indeed, significant gene expression changes did not always correlate with enriched pathways due to the coordinated nature of GSEA. Therefore, enriched pathways within each exercise time point should be examined further within the other time points (such as in **Sections 1.6.1-1.6.3**) to deconvolute the molecular mechanisms behind adaptations to strenuous exercise.

Chapter 7: Epigenetic Regulation of Gene Expression

The data generated as a part of this project contributed to a publication as outlined below:

- Voisin S, **Harvey NR**, Haupt LM, Griffiths LR, Ashton KJ, Coffey VG, Doering TM, Thompson JM, Benedict C, Cedernaes J, Lindholm ME, Craig JM, Rowlands DS, Sharples AP, Horvath S, Eynon N. (2020) An epigenetic clock for human skeletal muscle. *Journal of Cachexia, Sarcopenia, and Muscle Jan 31, 2020, DOI: 10.1002/jcsm.12556*

This chapter outlines the background, methodology, results, and significance of **Aim 4** of this thesis (epigenetic regulation of gene expression in exercise time points). Section 7.1 outlines the entirety of the chapter in abstract form. Section 7.2 discusses the background and justification for the epigenetic project. Section 7.3 describes the aims and specific methodology unique to the epigenetic typing of the Gene SMART participants. Sections 7.4.1 and 7.4.2 outline the results for the quality control and the differential methylation, respectively. Section 7.5 discusses the differentially methylated promoter regions identified within the study. Rudimentary MultiOmics analysis is outlined in Sections 7.6 and 7.7. Finally, Section 7.8 summarises the key findings from the epigenetic study.

This chapter pertains to thesis Aim 4 as outlined in greater detail below:

Thesis Aim 4:

4. Ascertain the methylation changes in a specific subset of Gene SMART participants to;
    a. Ascertain the quality of the methylation signals across arrays, samples, and time points
    b. Discover differentially methylated probes and loci at each exercise time point
    c. Discover differentially methylated genes at each exercise time point
    d. Discern common significant molecular pathways changes between the epigenomic and transcriptomic data sets

7.1 ABSTRACT

Epigenetics is defined as the study of heritable changes to DNA that do not involve alterations to the DNA sequence. More recently, epigenetics has been assessed in response to biological changes including stress, disease and exercise. The main epigenetic mechanisms include DNA methylation, histone modifications and regulation by non-coding RNAs, with DNA methylation the most commonly understood epigenetic mechanism to date. Previous exercise studies have utilised a candidate gene approach to assess the DNA methylation changes to acute bouts of exercise. As such, there has been strong and consistent replication of key genes and promoter regions (*PPARGC1A*) that are methylated in response to training; however, the effects of long-term HIIT on the epigenetic landscape has not yet been explored. To assess global DNA methylation levels within the study, Illumina 850K EPIC BeadChips were utilised. A subset (n=19) of individuals from the Gene SMART study were selected and assessed at baseline (PRE), at acute HIIE (P0), and at chronic HIIT (4WP). Following differential methylation analysis, we identified a total of 1,138 (746 hypermethylated and 392 hypomethylated differentially methylated probes (DMPs)) for acute HIIE. For chronic HIIT, 7,470 probes were differentially methylated, of which, 2,371 DMPs were hypermethylated and 5,099 hypomethylated. To gain biological insight, mCSEA was used to gauge coordinated DMP signatures over gene promoter regions. We then assessed gene expression patterns that were negatively correlated and likely mediated by DNA methylation of promoter regions using *mCSEA.* We noted that correlated upregulated/hypomethylated genes for acute HIIE were the same transcription factor genes (*JUNB, EGR1, FOS, FOSB, EGR3,* and *MYC*) that were upregulated in the sole transcriptome findings. These results were obtained with multiple Omic level data sets and added extra weight to our findings. Unfortunately, no such correlation was identified for were identified for chronic HIIT, where the upregulation/hypermethylation of RNA processing (*BRUNOL6*) and protein localisation genes (*DZIP1L*), and downregulation/hypermethylation of genes involved in calcium signalling (*MYOZ1, HRC*) and muscle differentiation and growth (*SPEG, PPAPDC3*), was observed. We then used the upregulated and downregulated subsets of correlated genes for each time point for over-representation analysis, which identified 84 and 118 significant BPs for acute HIIE and chronic HIIT respectively. For acute HIIE, we found that only overrepresented BPs were associated with correlated genes, and the BPs were similar to BPs identified for both immediate acute (P0) and three-hours post HIIE (P3) in **Section 6.5.1** and **6.5.2**. This confirmed our findings that the skeletal muscle microenvironment was adjusting to stressful intervention. For chronic HIIT,

both up and downregulated clusters were identified. Upregulated/hypomethylated genes mapped to clusters previously identified in the sole transcriptome data. ECM organisation was noted as the most significant BP cluster that was likely to be regulated by promoter methylation of chronic HIIT responsive genes, which was consistent from previous findings pertaining to HIIT. BP clusters containing downregulated genes were associated with muscle contraction and muscle tissue development, which represented a novel finding for response to sustained HIIT. For this study, we identified that a four-week period of moderate volume HIIT is sufficient to trigger similar biological responses to low volume HIIT, and indeed demonstrated that other, novel BPs are epigenetically regulated in acute HIIE and chronic HIIT.

7.2 BACKGROUND

Environmental stimuli such as diet and exercise may heavily influence biological processes through the epigenetic regulation (DNA methylation, chromatin remodelling, non-coding RNA binding) of transcriptional processes. The most common and therefore most heavily studied epigenetic mechanism is DNA methylation, which may repress or activate gene expression, depending on the genomic location of the methylated site. For instance, gene promoter methylation is typically negatively correlated with expression of the gene. Methylation at these sites have been shown to block binding of transcription factors and therefore silence gene expression (134, 135). In contrast, methylation of enhancer regions has been shown to permit expression of a target gene (136). Methylation of DNA typically occurs at CpG-dinucleotides, where clusters of these are classified as CpG islands (359, 360). CpG islands show strong overlaps with a number of promoter regions throughout the genome and as such the epigenetic regulation of these islands strongly influences gene expression (361). Other regions within the genome are also known to be methylated (non-island associated CpGs, non-CpG methylation), however the change in methylation of these sites is poorly characterised (362).

Barres et al, previously reported the hypomethylation of promoter regions of *PPARGC1A*, *PDK4*, *TFAM*, and *PPAR-δ* immediately following HIIE (137). In addition, gene expression levels of *PPARGC1A* and *PDK4* were negatively correlated with promoter methylation levels, with a corresponding peak in gene expression at three-hours post HIIE. Multiple studies have replicated this interaction between methylation and gene expression for these response genes (363, 364). The expression of *TFAM* and *PPAR-δ* were found to be increased immediately following HIIE. The study also found that high intensity exercise was required to elicit changes in methylation and expression of these response genes.

Limited studies to date have utilised high density methylation analysis to discern genome wide methylation responses to training. A study utilising methylated DNA immunoprecipitation (MeDIP) found 134 genes were differentially methylated in skeletal muscle tissue following six months of training, with 86% being hypomethylated (reduced CpG methylation). This datum supported the findings that exercise induces a global hypomethylation event for exercise

responsive genes (365, 366). To date, there is only one study that has investigated genome wide DNA methylation changes in response to HIIT. The study by *Robinson et al*, investigated transcriptome and DNA methylation changes in response to twelve weeks of HIIT (150). They observed no statistically significant changes in promoter methylation and that methylation changes detected were also relatively small (<10%). *Robinson et al* concluded that there were no robust differences in DNA methylation by 72 hours post training. Suggesting that DNA methylation may be largely an acute response to exercise. However, as discussed in **Section 2.8.3**, this study contained n=11 young healthy participants and as such was likely underpowered. Specifically the sample numbers utilised did not reach the n=20 as required by their *a priori* power calculation. and the study did not observe (what is deemed to be) large significant (±10% beta) changes in DNA methylation following training. As such, *Robinson et al,* did not investigate the correlation between differentially methylated promoter regions and corresponding gene expression.

Recently, there has been debate surrounding the classification of a substantial methylation change in terms of a biological context. For instance, diseases characterised by genomic instability and large-scale biological differences such as cancer often yield very clear differences in methylation (~60%) when compared with healthy phenotypes (359, 360). However, studies that assess subtle differences in phenotypes, as observed in many complex diseases may observe methylation differences that are far smaller (1-10%) in magnitude (361). As such, previous exercise studies assessing methylation differences may have been overly strict with data interpretation and therefore excluded true positive results from analyses. As previously discussed, only one study to date has investigated the epigenomic response to HIIT in young healthy individuals, however the sample size for the study was small (n=11) and the authors were unable to identify epigenetic changes in response to training (150). The current study aimed to discern epigenetic marks associated with acute (P0) and chronic (4WP) post exercise time points within a comparatively large (n=19) training cohort. In addition, this study aimed to utilise novel bioinformatic approaches such as mCSEA to assess small to moderate, yet consistent and coordinated shifts in methylation and therefore discover differentially methylated promoter regions. This study also aimed to integrate this epigenetic approach with transcriptomic data at both gene and pathway level to increase the scrutiny of molecular processes governing the adaptation to HIIT.

7.3 METHODS

7.3.1 Participant Samples

A subset of 19 participants were selected from the Gene SMART cohort for genome wide epigenetic analyses at the following three time points: baseline (PRE), immediately post (P0), and four-weeks post (4WP) exercise. Briefly, *Barres et al,* reported that promoter methylation for exercise responsive genes was changed immediately following acute training, and transcription of the same genes was changed at three-hours post exercise (137). As such, and due to limited funding, we omitted the three-hours post time point from the methylation analysis. In addition, the analysis was limited to the number of samples with matching transcriptomic sequencing time points, as per details in **Section 3.2.4**.

7.3.2 DNA Isolation

As detailed previously (**Section 5.3.2**), a dual RNA/DNA extraction was performed for all muscle biopsy samples using the Allprep DNA/RNA/miRNA universal kit (*Qiagen*). The DNA quantity and quality was assessed via Nanodrop spectrophotometry (*ThermoFisher*), and Qubit fluorometry (*ThermoFisher*). All DNA samples were found to be of high quality ($A_{260/280}$ > 1.8) with approximately 3.8 ± 1.7µg of gDNA yielded per biopsy sample. Following extraction and quantitation, all stock gDNA was then stored at -20ºC prior to use in the methylation chip protocol.

7.3.3 EPIC 850K Methylation array

Briefly, 1µg of gDNA was bisulfite converted using the EZ DNA methylation kit (*Zymo Research*) according to manufacturer's instructions. Samples were randomised across all chips to limit the large-scale batch effects commonly associated with methylation chip protocols. The overnight (16 hours) bisulfite conversion method was used for subsequent DNA methylation BeadChip applications using the Illumina 850K EPIC BeadChip kit (*Illumina*). Briefly, this BeadChip contained ~850,000 probes, targeting 4% of the known CpG sites over the entire human genome, of which, 25.4% were located within promoter regions (146). The method was conducted over four days per 32 samples according to manufacturer's instructions with the following amendments. All 57 bisulfite converted DNA samples were randomised across ten

BeadChips for methylation analysis. An Illumina HiScan (*Illumina, USA*) was used to image the stained BeadChips. Samples on the plate were limited to 32 samples (four BeadChips) due to the capacity requirements of the instrument used.

7.3.4 Bioinformatics

7.3.4.1 QC analysis

Differential methylation analysis was conducted using R/Bioconductor version 3.6.1. Firstly, both red and green IDAT files were imported using the *minfi* package, then converted into methylated and unmethylated signals, respectively. Following this, probes were filtered based on significant detection, presence in at least 5% of samples, probes with non-CpG status, probes with SNPs, and identified cross-reactive probes (367-369). Following filtering, methylation data was normalised using the Beta-mixture quantile (BMIQ) method to correct for type II probe bias within the arrays (370). Batch correction was performed via the *RUVm* package (k=40) which uses the 411 negative controls on the EPIC BeadChip to identify a set of empirical control probes (bottom 10% of ranked probes) to accurately estimate the unwanted variation, this variation is then adjusted for in the *limma* linear model for differential analysis. (371).

7.3.4.2 Differential methylation analysis

After normalisation and batch correction, the *limma* package was used to generate a discovery differentially methylated probes (DMPs) set at each exercise time point. The median beta values were calculated within the individual time points and used to inform the magnitude of differential methylation between time points. Only probes that passed FDR<0.05 and had a methylation change of $\Delta\beta > 2\%$ between time points were considered for further analyses. Following DMP analysis, mCSEA was performed to identify enriched promoters. Briefly, the *mCSEA* package was used detect moderate, but consistent DNA methylation changes in gene promoters (372).

7.3.4.3 Integration analysis

We aimed to correlate the transcriptome data from **Chapter 6**, with the DNA methylation data from the present Chapter. Briefly, common genes between transcriptome and enriched methylated promoter regions were identified via gene symbol. We then utilised the '*mCSEAIntegrate*' function of the mCSEA package (372), which averages the methylation of

promoter CpGs identified from the '*mCSEATest*' function. A threshold of >5 CpG probes per promoter was used to ensure robust methylation changes in gene promoter regions. The function then uses a Pearson correlation (<-0.3; moderate linear relationship) between these CpGs and gene expression changes for nearby (±1,500bps) genes. Negative correlations (i.e., classic form of DNA methylation regulated gene expression) were considered statistically significant at FDR<0.05.

Gene ontology (GO) was then conducted via Over Representation Analysis (ORA), which assesses over or under representation of a gene set relative to genes within individual BPs. To ensure focussed BP signals, we limited the ORA gene sets to a minimum of 15 and maximum of 500. As per Chapter 6, conservative thresholds (P<0.001, FDR<0.05) were set to minimise the chance of false positive findings within the integration analysis. GO:ORA analysis was conducted for upregulated (hypomethylated) and downregulated (hypermethylated) genes separately for both acute HIIE (P0) and HIIT (4WP) respectively. Biological processes were visualised in Cytoscape as per **Section 3.6.4**. Beta values were used for visualisation as per best practice for methylation data (373).

7.4 RESULTS AND DISCUSSION: DIFFERENTIALLY METHYLATED PROBES

7.4.1 QC Results

All BeadChip experiments require stringent QC thresholds to ensure robust and reproducible data. To do this, several QC steps were employed to ensure probes were filtered based on detection, raw beadcount, CpG status, probes with proximal SNPs, and cross-reactive probes. Following filtering 744,000 total probes remained, representing ~86% of the total probes on the arrays (**Table 7-1**). The remaining probes were then BMIQ normalised to adjust for the bias in type II probes. Following normalisation, the different distributions of the type-I and type-II probes has been appropriately adjusted for (**Figure 7-1).** Sample level barplots were used to assess probe variability before and after quantile normalisation, and no samples were removed for failing QC. Prior to differential methylation analysis, unsupervised clustering analysis was performed of samples with PCA plots to discern any unexplained variance within the data set. The unadjusted and covariate adjusted PCA plots are shown in **Figure 7-2.** As shown in the unadjusted PCA plot (**Figure 7-2A**) no discernible clustering was observed that clearly corresponded with each of the exercise time points. The unadjusted correlation matrix

(**Figure 6-2B**) suggested the variance in the top PCs was strongly correlated (r ± 0.7) with experimental batch (and therefore slide number), and array position on each slide. This unwanted variance was almost completely removed (r±0.2) following batch correction with the sample group contributing the most variance to PC1 (**Figure 7-2C-D**).

Table 7-1: *Pipeline detailing filtering of CpG probes from raw data to final number of analysed sites.*

Filtering step	**# of probes removed**	**% probes removed**	**# probes remaining**	**% probes remaining**
Prior to filtering	0	0	866238	100
Detection P-value > 0.01	13126	1.52	853112	98.48
Beadcount <3 in at least 5% samples	1863	0.21	851249	98.27
Non-CpG probes	2825	0.33	848424	97.94
Probes with SNPs (367)	101239	11.68	747185	86.26
Cross-reactive probes (368, 369)	3185	0.37	744000	85.89

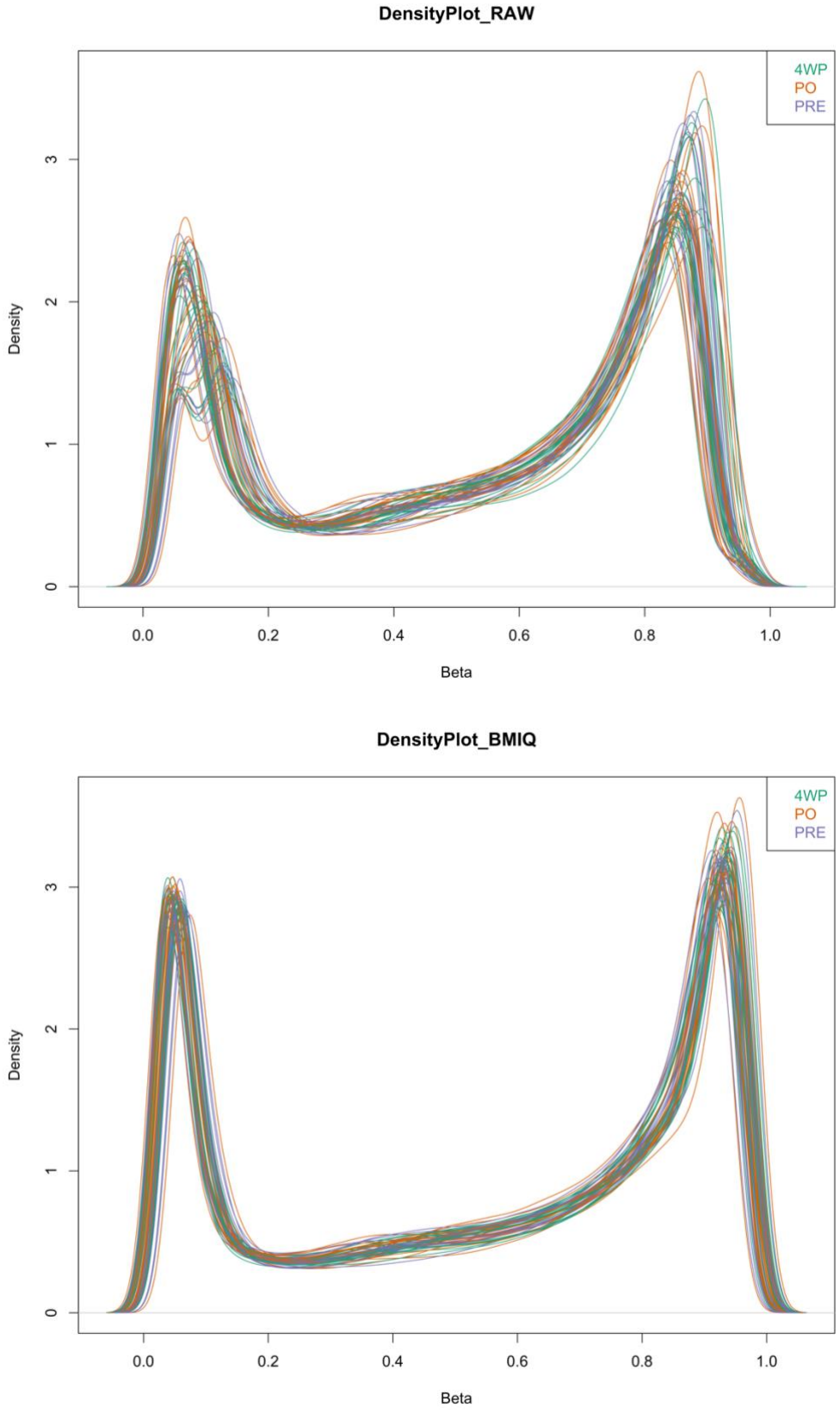


Figure 7-1: *Density plots showing* ***A)*** *unadjusted beta values showing inconsistent curves, and* ***B)*** *BMIQ adjusted beta values showing normalised results. X-axis shows beta values ranging from 0-1, Y-axis represents density of probes reaching the specified beta value.*

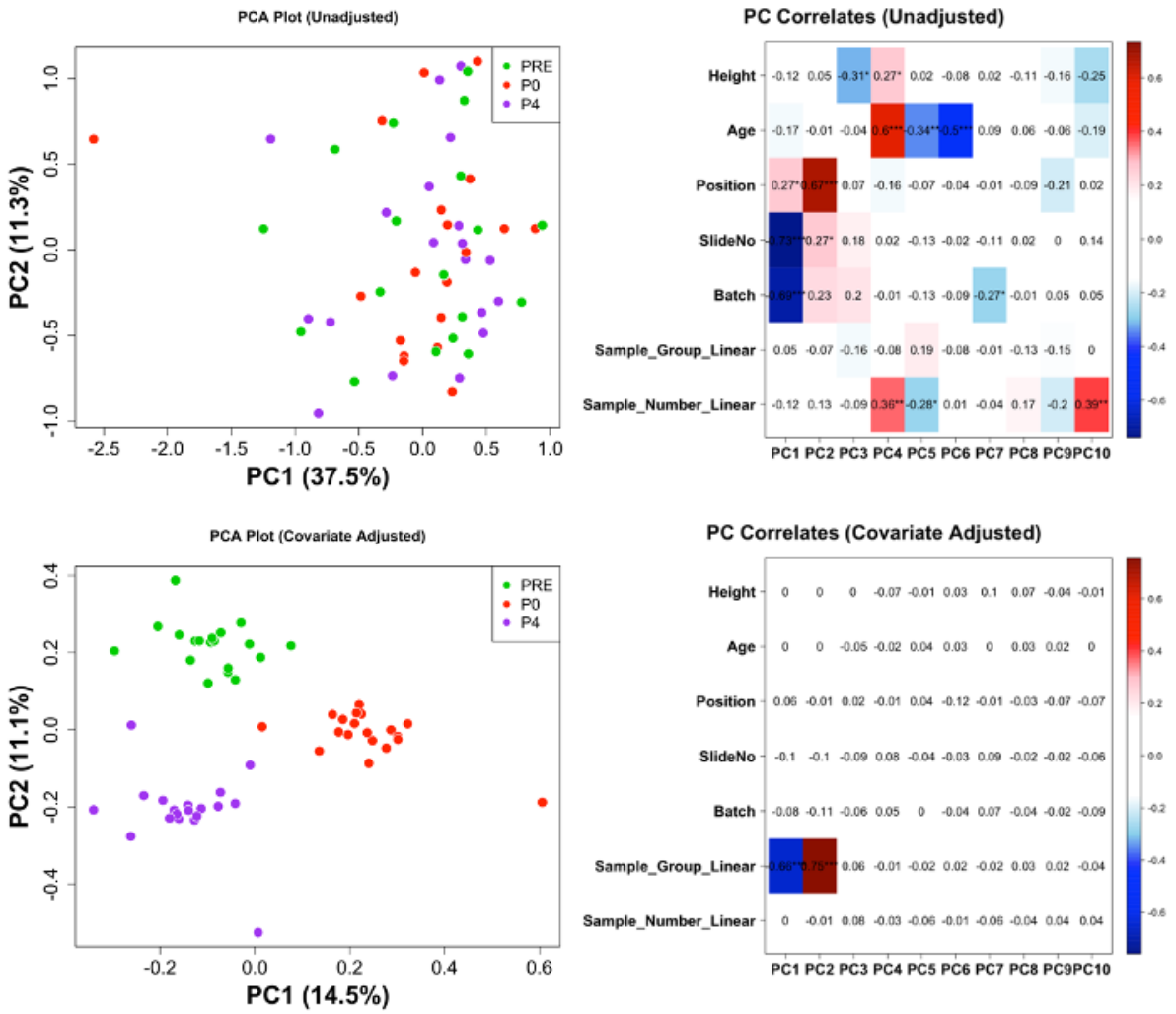


Figure 7-2: *PCA correlation plot and single value decomposition of the top 500 probes within the Gene SMART samples.* ***A)*** *unadjusted PCA plot showing no separate clustering of time points across samples.* ***B)*** *Correlation matrix of the contributing phenotypes to PCs 1-10 showing multiple confounding phenotypes.* ***C)*** *PCA plot adjusted for the most contributing phenotypes (covariates) from panel B showing correct clustering of samples into individual exercise time points.* ***D)*** *Correlation matrix of the adjusted contribution of phenotypes to each of the top 10 PCs showing heavily weighted timepoint phenotype in the early PCs. Panels B and D show correlation values used to determine directionality of each covariate to each PC.*

7.4.2 Differentially Methylated Probes (DMPs)

To visualise the magnitude of changes in differential methylation between exercise time points (PRE, P0 and 4WP), probes that passed FDR<0.05 and had a delta Beta (Δβ) ±2% were considered. This identified a total of 8,414 DMPs across all exercise time points (**Figure 7-3**). Interestingly, much fewer probes were identified to be differentially methylated immediately following HIIE (P0) when compared with chronic HIIT (4WP). From the data, a total of 1,138 (746 hypermethylated and 392 hypomethylated) DMPs were identified at P0. At 4WP, 7,470 probes were differentially methylated, of which, 2,371 DMPs were hypermethylated and 5,099 hypomethylated.

When we assessed the DMPs in common between acute HIIE and chronic HIIT, we identified 96 hypermethylated and 98 hypomethylated probes (**Figure 7-4A**). Interestingly, and despite the small overlap of DMPs between time points, we did not observe a significant difference in probe location across time points for either CpG islands (**Figure 7-4B**) or gene locations (**Figure 7-4C**). When considering the promoter regions (as determined by *mCSEA*), we found that TSS1500 and 5`UTR regions showed an initial drop in DMPs following HIIE (P0), but these returned to baseline for chronic HIIT (4WP). In contrast, DMP levels for TSS200 and 1st exon regions dropped immediately following HIIE (P0), and the drop was maintained in chronic HIIT (4WP). This slight drop in CpG methylation has been shown previously for acute exercise (137), however, the authors focussed on targeted genes of interest and did not investigate global promoter methylation changes in response to training. More recently, *Robinson et al*, aimed to discover DNA methylation in response to HIIT, but was unable to identify any methylation changes (150). This was likely due to small sample sizes, and bioinformatic approach. These limitations further highlight the utility of assessing small but coordinated changes to biological systems (*mCSEA*), especially regarding large-scale interventions such as exercise training.

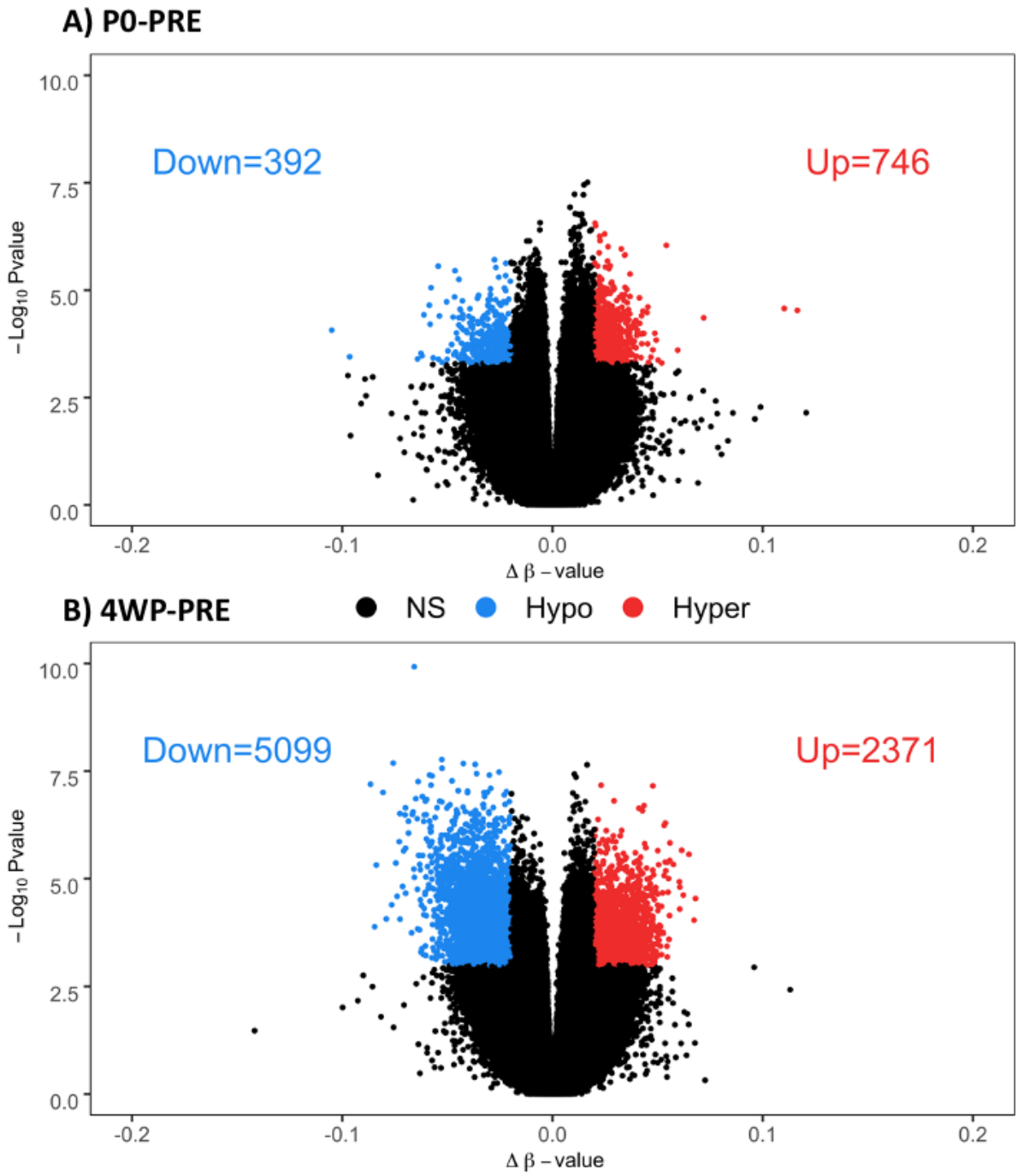


Figure 7-3: *Volcano plots showing significant DMPs following* ***A)*** *acute HIIE (P0), and* ***B)*** *chronic HIIT (4WP). X-axis shows the median Δβ value for each test and Y-axis represent nominal P-value for the DMPs. Hyper- (red) and hypomethylated (blue) CpGs for each time point are coloured based on Δβ±2%; FDR<0.05.*

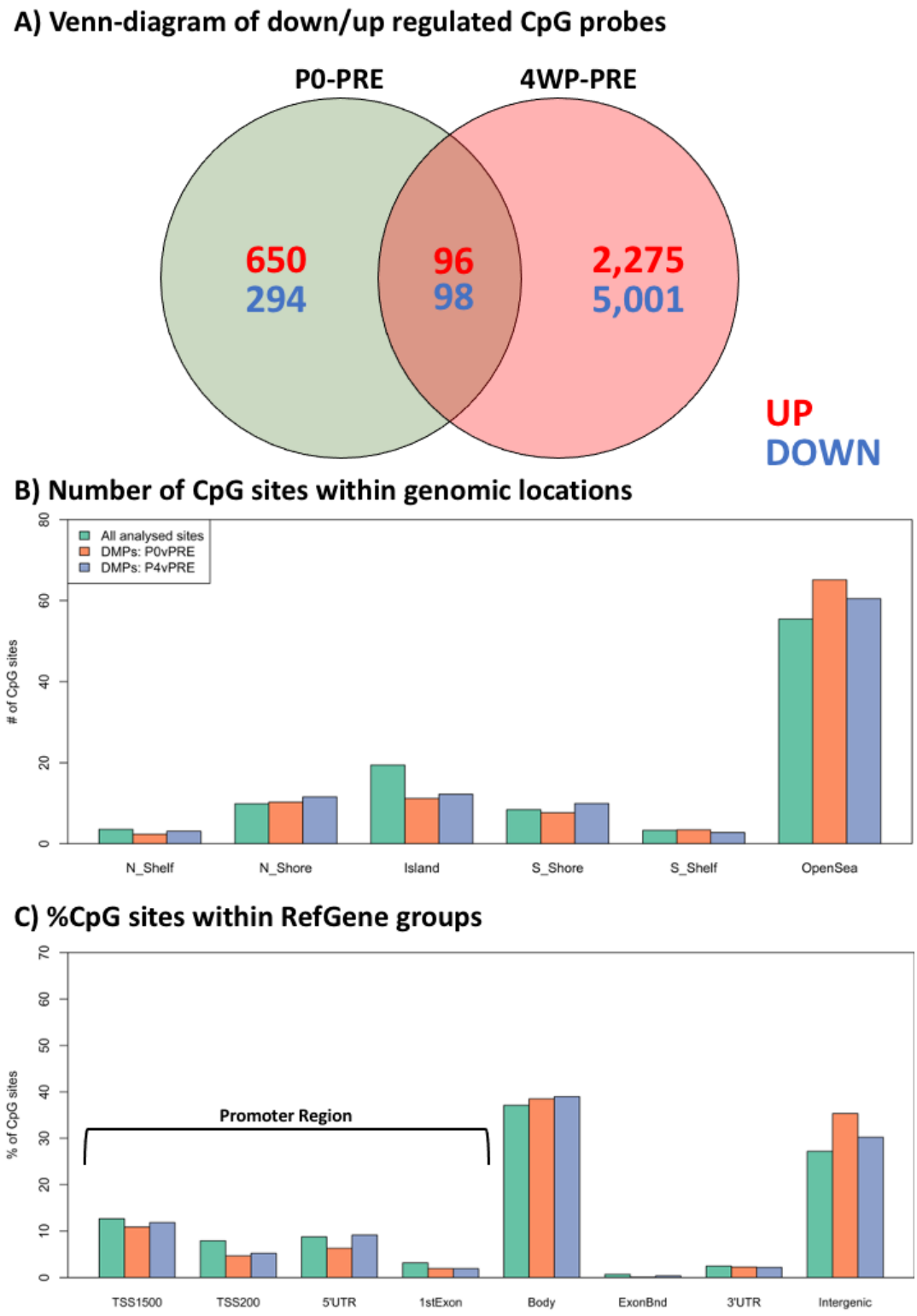


Figure 7-4: *Summary of significant DMPs identified in acute and chronic HIIT*. **A)** Venn-Diagram of significant (Δβ±2%; FDR<0.05) DMPs in each exercise test detailing level of overlapping DMPs between time points**.** B) Number of significant DMPs within different genomic locations. C) Percentage of significant CpG sites within genomic sites in relation to genic regions.

7.5 RESULTS AND DISCUSSION: INTEGRATED ANALYSIS

As previously discussed, promoter methylation is strongly associated with the repression of gene expression. We identified 10,697 genes that were present in both the transcriptome and promoter methylation data sets. Following data integration using *mCSEAIntegrate*, we identified 122, and 276 significant (FDR<0.05) genes that were likely to be regulated via promoter methylation for acute HIIE (P0) and chronic HIIT (4WP) respectively (**Figure 7-5**). Of these, nine genes were in common between exercise time points (*CCNE1*, *FAM111A*, *IRF1*, *PLAU*, *RAD9A*, *RGS16*, *STK17A*, *ZC3HAV1L*, and *ZNF546*).

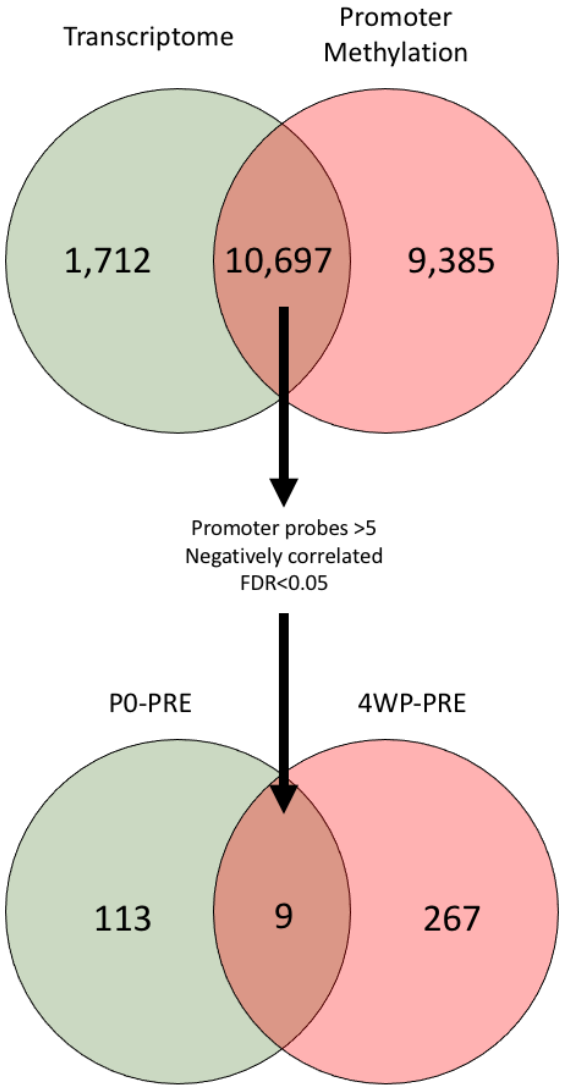


Figure 7-5: *Venn Diagram of overlapping genes and then correlated promoters between each exercise time point. (>5 CpG probes in promoter; FDR<0.05).*

7.5.1 Previous promoter methylation regulated exercise response genes

We aimed to confirm previous findings for exercise responsive genes (*PGC1α, PDK4, MYC, HIF1α*) as identified by *Barres et al,* (2012) for acute exercise response (137). We observed similar patterns for promoter hypomethylation and upregulated gene expression for these genes within our study. Except for *MYC*, these response genes were not identified within the top 10 upregulated/hypomethylated or downregulated/hypermethylated genes identified within our acute HIIE or chronic HIIT data. These findings indicated that novel exercise responsive genes, especially in HIIE and HIIT, have yet to be identified.

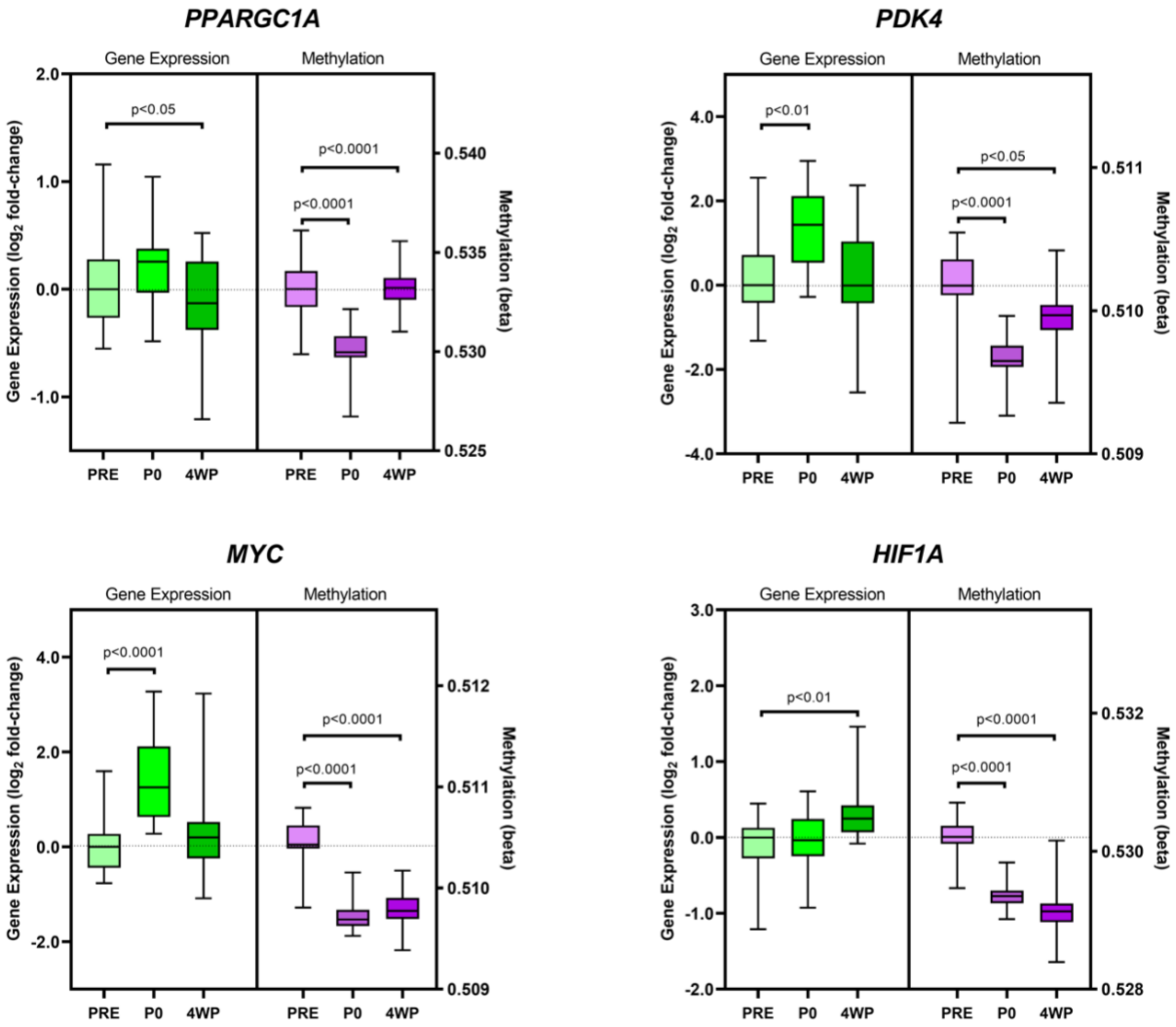
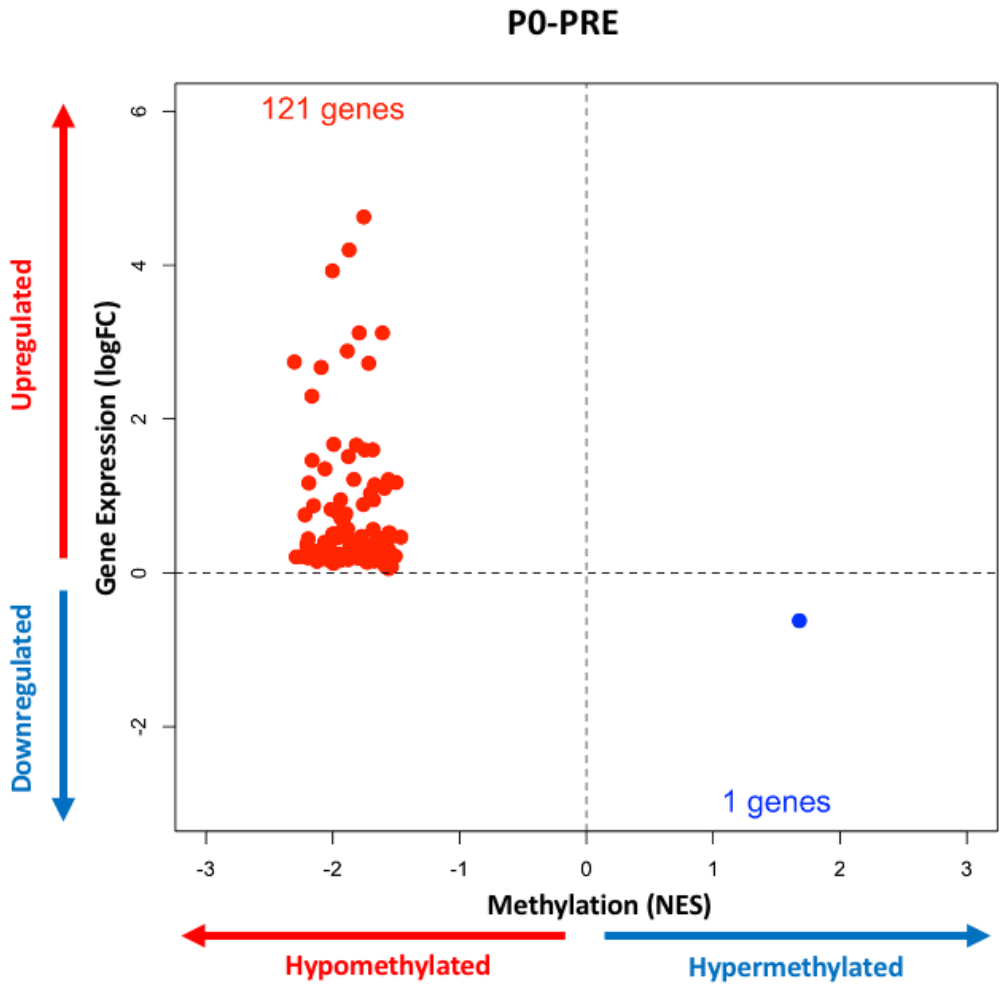


Figure 7-6: *Gene expression and DNA promoter methylation for previously identified exercise responsive genes. Green boxplots show longitudinal gene expression changes (logFC) identified within* ***Chapter 6*** *(colour intensity corresponds to increase in time). Purple boxplots show longitudinal promoter methylation changes (beta value – colour intensity corresponds to increase in time).*

7.5.2 Epigenetically regulated DEGs following acute HIIE (P0)

Immediately post HIIE (P0), most negatively correlated genes (99.1%) were identified to be upregulated, with a corresponding hypomethylated promoter region. We also noted that these upregulated, methylation responsive genes were associated with higher correlation coefficients ($r < -0.58$) than the singular downregulated gene (*GLNY:* $r = -0.38$) at this time point.

Early response to HIIE resulted in hypomethylation in the promoter regions of transcription factors such as *JUNB, EGR1, FOS, FOSB, EGR3,* and *MYC*, leading to the upregulation of these genes at acute HIIE. As per **Section 6.5.1**, the upregulation of these genes in response to early HIIE has been shown previously and suggests the skeletal muscle microenvironment is responding to a stressful environmental change (301-303). The epigenetic regulation of these transcription factors at acute HIIE represents a novel finding within exercise science and should be confirmed with replication studies. The reoccurrence of these transcription factors in multiple Omics analyses further illustrates the significance of the findings from **Chapters 6 and 7**.



Gene Symbol	Correlation coefficient	FDR	Promoter methylation (NES)	Gene expression (LogFC)
Upregulated/Hypomethylated				
C1orf63	-0.80	1.90E-07	-1.79	3.12
JUNB	-0.78	3.49E-07	-2.30	2.74
TMEM50A	-0.76	1.55E-06	-1.61	3.12
EGR1	-0.74	3.41E-06	-1.87	4.20
FOS	-0.72	9.25E-06	-2.00	3.92
ADAMTS1	-0.66	1.25E-04	-1.50	1.17
FOSB	-0.66	1.40E-04	-1.76	4.63
EGR3	-0.62	4.90E-04	-1.88	2.88
MYC	-0.62	5.36E-04	-2.16	1.46
CYR61	-0.58	1.38E-03	-1.75	1.60
Downregulated/Hypermethylated				
GNLY	-0.38	0.0408	1.68	-0.62

Figure 7-7: *Correlation assessment of gene expression and promoter DNA methylation for acute HIIE (P0). Red data points show upregulated genes with accompanied promoter hypomethylation, blue data points show downregulated genes with accompanied promoter hypermethylation (r<-0.3; FDR<0.05).*

To visualise the correlation of the genes of interest (GOIs: *EGR1*, *FOSB*, *NR4A2*, *JUNB*), we constructed boxplots of the gene expression (logFC) and promoter methylation (beta) changes for each (**Figure 7-8**). All immediate changes to gene expression observed for HIIE were correlated with significant changes to promoter methylation. Interestingly, gene expression levels for the GOIs reverted to baseline for chronic HIIT, however, the DNA promoter methylation remained significantly hypomethylated. This finding suggested a mechanism whereby promoter methylation is reduced for acute exercise responsive genes in a chronic manner for sustained response to future bouts of HIIE.

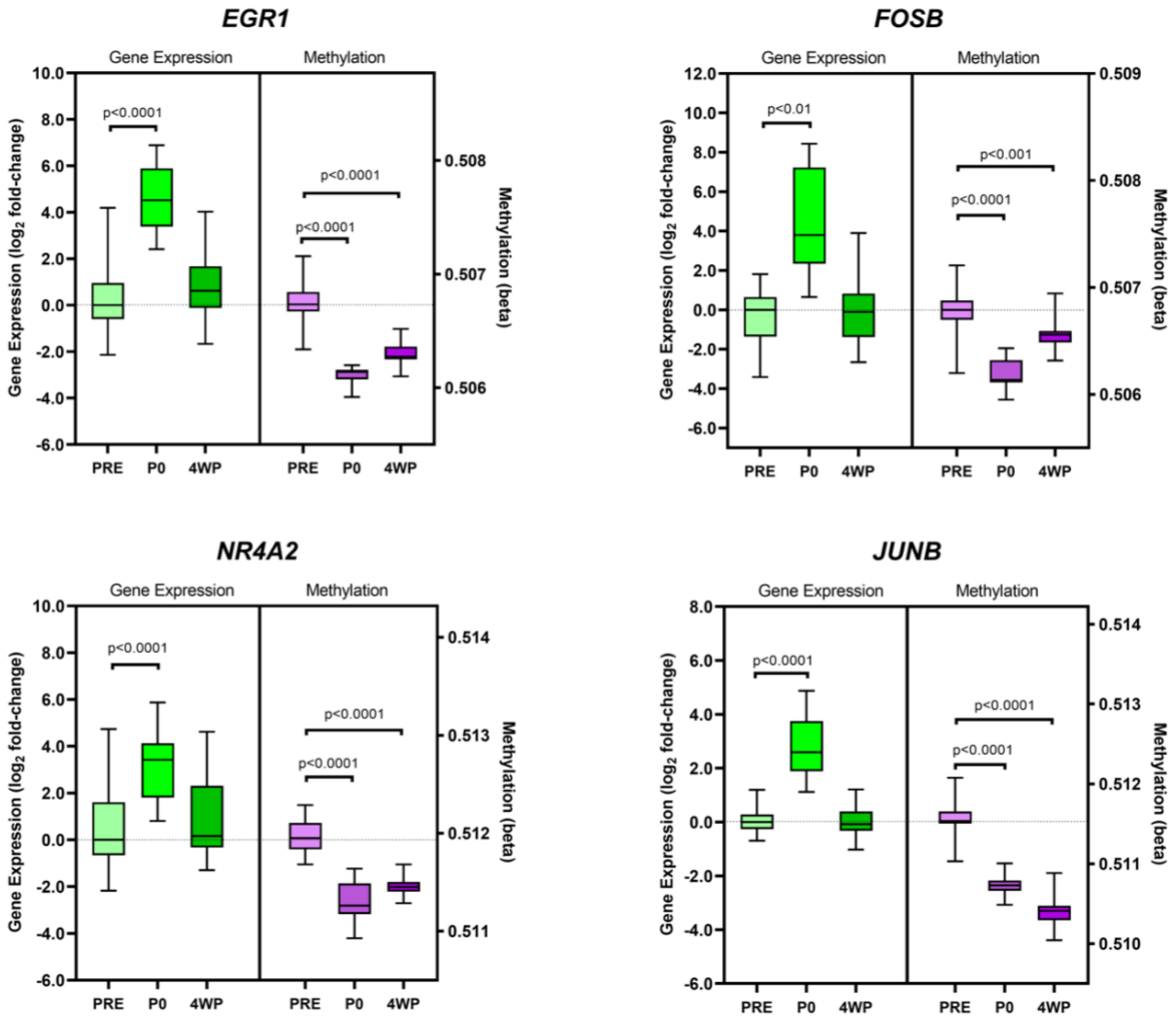


Figure 7-8: *Gene expression and DNA promoter methylation for GOIs identified for acute HIIE. Green boxplots show longitudinal gene expression changes (logFC) identified within* ***Chapter 6*** *(colour intensity corresponds to increase in time). Purple boxplots show longitudinal promoter methylation changes (beta value – colour intensity corresponds to increase in time).*

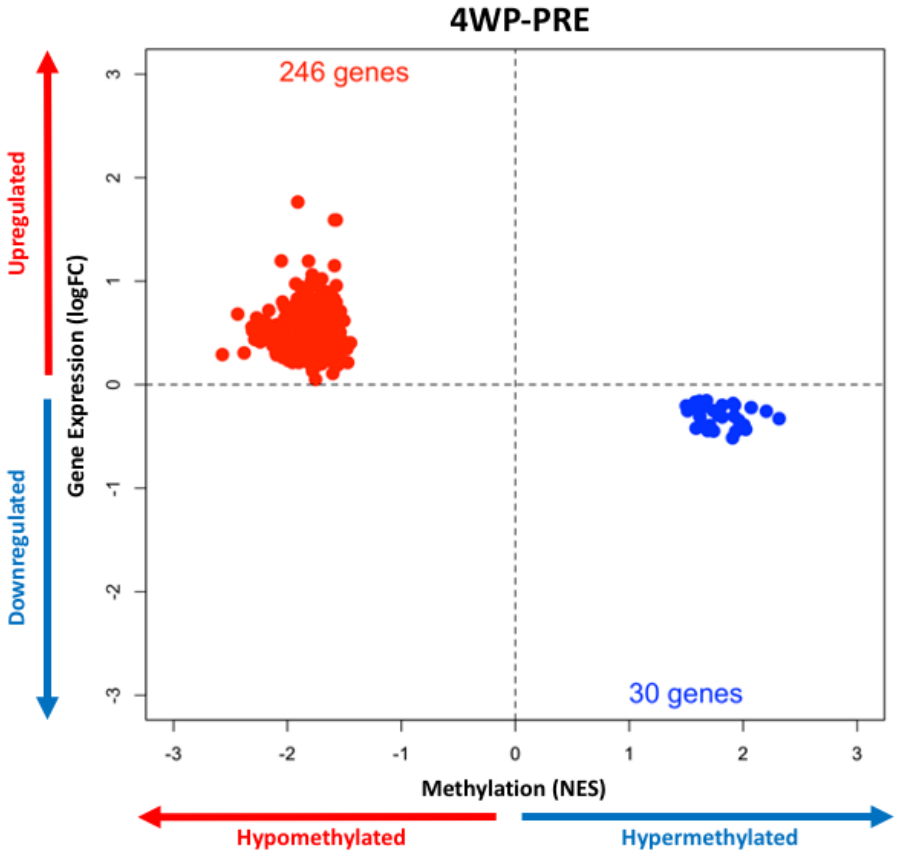
7.5.3 Epigenetically regulated DEGs following chronic HIIT (4WP)

Following integration analysis, we identified a total of 246 upregulated/hypomethylated and 30 downregulated/hypermethylated genes for chronic HIIT (4WP). Contrary to acute HIIE, the top 10 up and downregulated genes from **Section 6.4.2** were not observed in the top 10 up and down correlated genes. This was expected due to the sampling time utilised for the four-week post HIIT time point (48 hours post, at rest). In addition, only moderate correlations ($r <$ -0.6), and moderate gene expression changes (logFC< 1.5) were observed for the top genes identified from the integration analysis.

The top upregulated/hypomethylated genes included *CUGBP Elav-Like Family Member 6 (BRUNOL6),* which is an RNA binding protein that mediates mRNA splicing*; Kelch Like Family Member 6 (KLHL6),* which is involved in B-cell antigen receptor signalling*;* and *DAZ Interacting Zinc Finger Protein 1 Like (DZIP1L),* which acts as a transition zone protein for localisation of PDK1 and PDK2 (374). PDK1/2 protein levels have been shown to be elevated in endurance trained subjects (375) and, as such, these results indicate a mechanism by which promoter methylation mediated gene expression of *DZIP1L* may lead to functional change in response to chronic HIIT.

The top downregulated/hypermethylated genes were linked with calcium signalling and muscle growth and differentiation. Genes linked to calcium signalling included *Myozenin 1* (*MYOZ1*) and *Histidine Rich Calcium Binding Protein* (*HRC*), which have been associated previously with calcineurin signalling (376), and calcium sequestration/release in muscle respectively (377). Genes linked to muscle growth and differentiation included *Striated Muscle Enriched Protein Kinase* (*SPEG*) and *Phospholipid Phosphatase 7* (*PPAPDC3*), which have been associated with high responsiveness to training (378), and negative regulation of myoblast differentiation.

As such, the correlated genes were expected to contribute to ongoing biological function rather than a large-scale stress response, as per our conclusions in **Section 6.5.3**. We proceeded with over-representation analysis to determine biological processes that were likely to be influenced by the correlated gene sets.



Gene Symbol	Correlation coefficient	FDR	Promoter methylation (NES)	Gene expression (LogFC)
Upregulated/Hypomethylated				
ZC3HAV1L	-0.62	0.009	-1.74	0.95
BRUNOL6	-0.60	0.009	-1.57	1.59
CDO1	-0.55	0.017	-1.78	0.61
LOC151534	-0.55	0.017	-1.81	0.77
KLHL6	-0.55	0.017	-1.78	1.06
DZIP1L	-0.56	0.017	-1.59	1.15
LRMP	-0.54	0.020	-1.59	0.44
CCDC8	-0.53	0.020	-1.83	0.77
BMP4	-0.49	0.024	-1.84	0.28
IKBKG	-0.49	0.024	-1.70	0.45
Downregulated/Hypermethylated				
PPAPDC3	-0.46	0.025	2.02	-0.43
HRC	-0.46	0.025	1.82	-0.31
PRO1768	-0.48	0.025	1.62	-0.31
FBXO40	-0.45	0.028	1.59	-0.42
PNLDC1	-0.42	0.030	1.68	-0.15
C19orf76	-0.42	0.031	1.74	-0.25
MYOZ1	-0.41	0.033	1.92	-0.31
LOC100505795	-0.40	0.035	1.51	-0.25
SLC27A5	-0.40	0.035	1.66	-0.21
SPEG	-0.39	0.039	2.01	-0.39

Figure 7-9: *Correlation assessment of gene expression and promoter DNA methylation for chronic HIIT (4WP). Red data points show upregulated genes with accompanied promoter hypomethylation, blue data points show downregulated genes with accompanied promoter hypermethylation (r<-0.3; FDR<0.05).*

To visualise the correlation of the genes of interest (GOIs: *BRUNOL6, DZIP1L, PPAPDC3, MYOZ1*), we constructed boxplots of the gene expression (logFC) and promoter methylation (beta) changes for each (**Figure 7-10**). The gene expression for all GOIs was significantly changed for chronic HIIT, with a corresponding significant change in promoter methylation. No change to either gene expression or promoter methylation (except for *BRUNOL6*) was observed for the chronic HIIT responsive genes. These findings suggest that genes responsive to chronic HIIT are likely to not be involved in the acute response to HIIE.

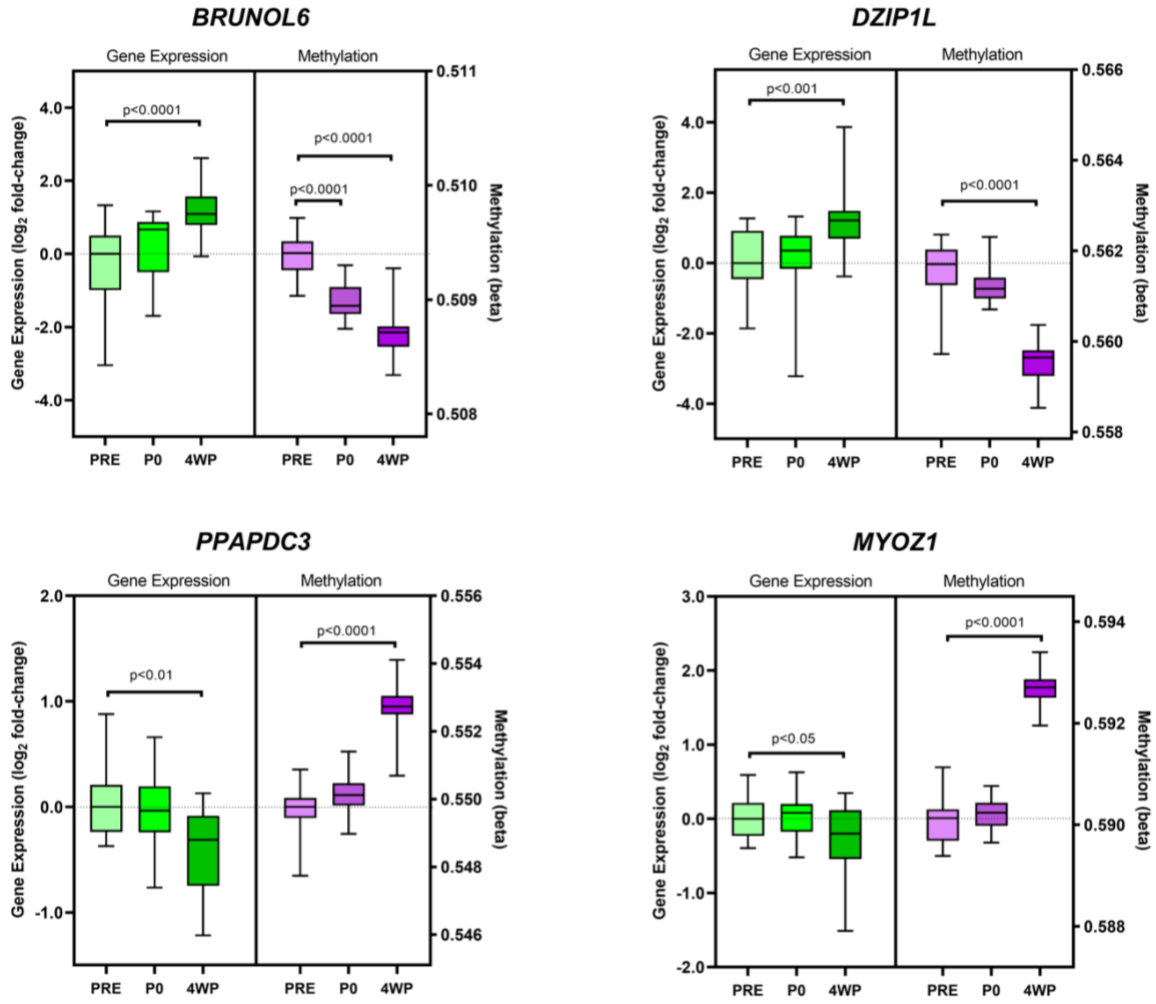


Figure 7-10: *Gene expression and DNA promoter methylation for GOIs identified in chronic HIIT.* ***Green*** *boxplots show longitudinal gene expression changes (logFC) identified within* ***Chapter 6*** *(colour intensity corresponds to increase in time). Purple boxplots show longitudinal promoter methylation changes (beta value – colour intensity corresponds to increase in time).*

7.5.4 Integrated biological processes identified for acute HIIE (P0)

Interestingly, only upregulated BPs were identified for the correlated gene set at acute HIIE (P0) (**Figure 7-7**). These findings matched the correlated gene level findings, in which all but one gene was upregulated in response to HIIE. BPs in common with the immediate transcriptome data (**Section 6.5.1**) included clusters for response to reactive oxygen species, processes involved in development, skeletal muscle tissue development, and response to peptide (hormone response).

As per **Section 2.3.3**, calcium flux in skeletal muscle immediately following exercise has been shown previously for HIIE (62). Interestingly, this response was not observed for this immediate response in the transcriptomic data (**Section 6.5.1**), indicating that these correlated mechanisms may be more subtle than individual Omic analysis, but also may be more biologically informative. We found the most significant BPs were primarily in response to stress and development processes (**Table 7-2**). *Response to cAMP* was noted as the most significant BP cluster within the ORA enrichment map for acute HIIE. Levels of cAMP have been shown to be dissociated from exercise intensity and gluconeogenesis (379), however, the response to camp levels following training has been extensively studied (380). Like calcium response, we did not observe a cAMP cluster for the singular transcriptomic data, further highlighting the significance of integrated analysis.

Of note, the correlated genes identified at this early time point may have been beginning to influence gene expression, which has been previously shown for exercise responsive genes (137). As such, the BPs identified from the integration analysis showed similarities with the BPs identified from the transcriptomic analysis for both immediate and three-hour acute HIIE time points. BPs similar to the three-hour acute HIIE time point were regulation of leukocyte differentiation, and response to hypoxia. These findings confirmed our conclusions from **Sections 6.5.1** and **6.5.2** as the BPs suggest an immediate, DNA methylation mediated, stress response in the skeletal muscle microenvironment following a single bout of HIIE.

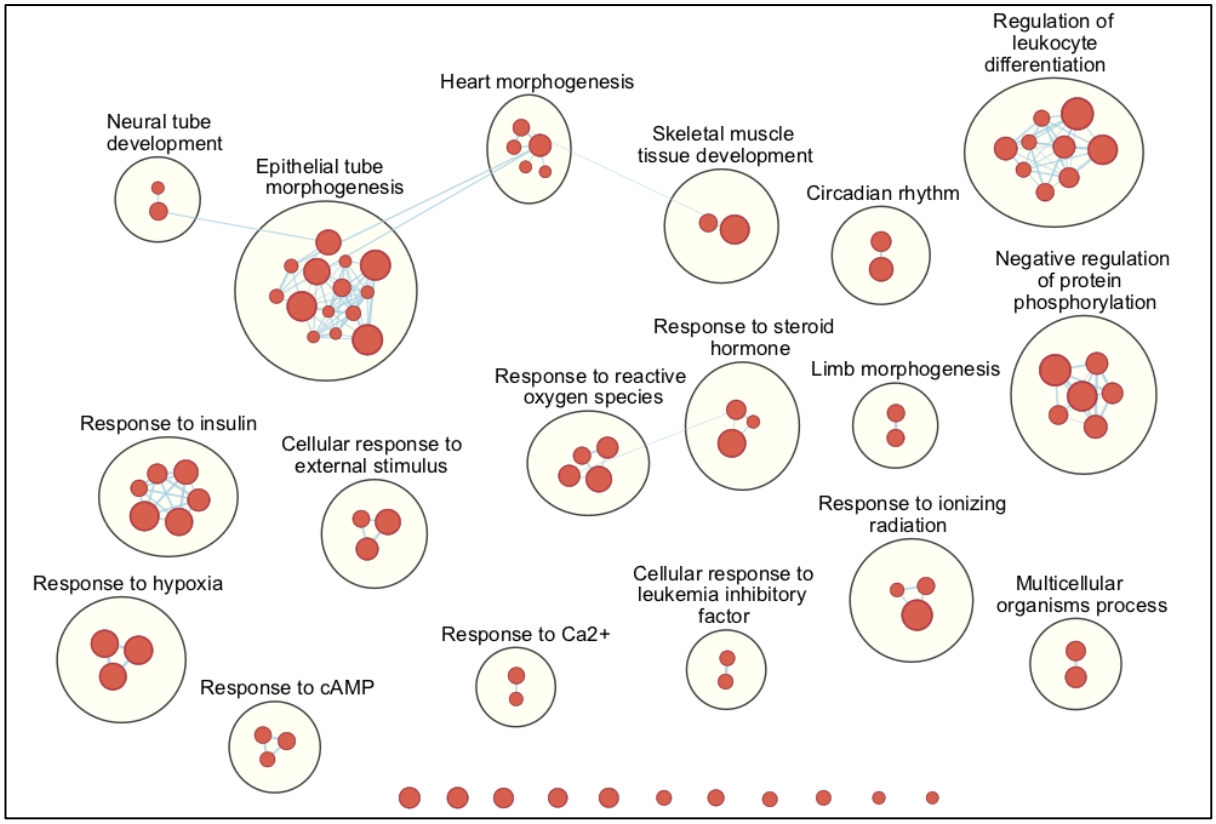


Figure 7-11: *Enrichment map showing correlated BPs for immediate acute HIIE (P0). Each node represents a BP term that is significantly differentially regulated (FDR<0.05). Jaccard overlap between nodes was set to 0.375. Size of the nodes correlate with number of genes within the BP and colours represent normalised enrichment scores (red is upregulated). Circles encompassing the molecular nodes are an auto annotation of the larger BP of the clustered nodes.*

Table 7-2: *Top 10 most significant BP clusters likely to be regulated by promoter methylation of key genes at acute HIIE (P0)*

Cluster Name	No. Nodes	FDR
Response to cAMP	3	1.23E-04
Response to ionising radiation	3	7.55E-04
Epithelial tube morphogenesis	14	2.07E-03
Response to Ca^{2+}	2	5.00E-03
Response to insulin	6	1.00E-02
Neural tube development	2	1.20E-02
Heart morphogenesis	5	1.40E-02
Negative regulation of protein phosphorylation	6	1.50E-02
Regulation of leukocyte differentiation	9	1.50E-02
Circadian rhythm	2	1.90E-02

7.5.5 Integrated biological processes identified for chronic HIIT (4WP)

The epigenetically regulated BPs at chronic HIIT (4WP) were reminiscent of the transcriptomic data for the same time point. With the majority (22/24 clusters) of gene sets being upregulated (**Figure 7-8**). This was expected as there were more upregulated/hypomethylated genes than downregulated/hypermethylated genes at this time point (as previously shown in **Figure 7-7**).

Interestingly, the two downregulated BP clusters were related to muscle contraction and muscle tissue development, which were notably absent from the sole transcriptome analysis. This indicated an epigenetic mechanism for the downregulation of muscle functional processes, and growth processes at conclusion of the training. Due to the sampling (at rest) for this time point, the downregulation of muscle functional BPs was somewhat expected as these processes would likely be enhanced during or soon after exercise.

As per **Section 6.5.3**, immune BPs were upregulated in the integrated analysis, with six BP clusters (*Regulation of leukocyte differentiation, Regulation of myeloid differentiation, Negative regulation of leukocyte differentiation, Neutrophil activation, Negative regulation of leukocyte apoptosis, and Regulation of leukocyte proliferation*) corresponding to regulation of immune cell subtypes. As previously discussed, the skeletal muscle microenvironment contains resident immune cells (18), and the elucidation of these signals from skeletal muscle myofibers is likely further complicated by immune cell extravasation into the skeletal muscle microenvironment (322).

When assessing the top 10 most significant BP clusters, we found that structural, developmental, and response to virus BPs were represented (**Table 7-3**). ECM organisation was the most significant BP cluster and was also observed for the transcriptome data at this time point. In addition, an overrepresented collagen biosynthesis BP cluster indicated that the microenvironment was being remodelled following chronic HIIT (4WP), which has been shown previously (126).

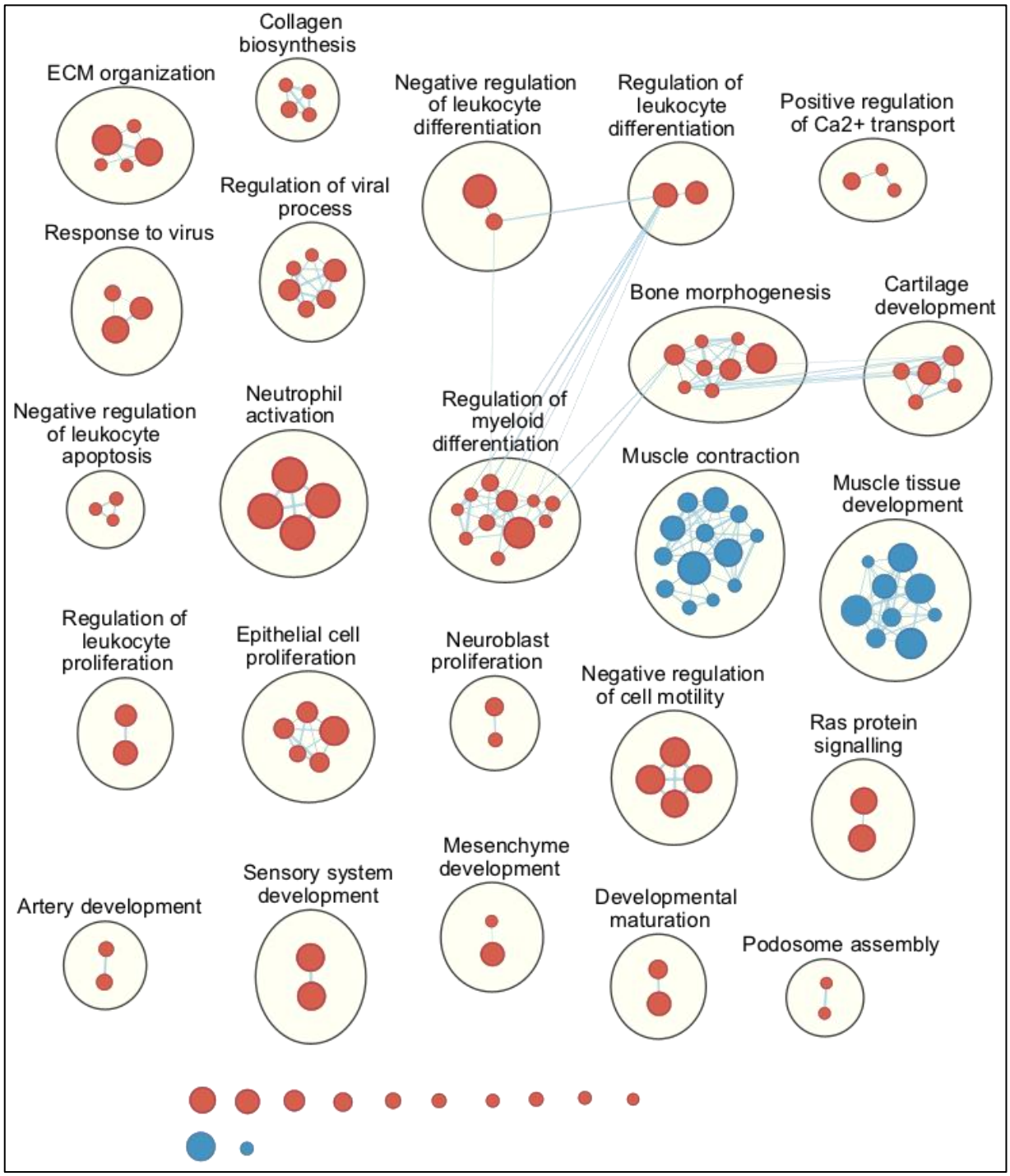


Figure 7-12: *Enrichment map showing correlated BPs for chronic HIIT (4WP). Each node represents a BP term that is significantly differentially regulated (FDR<0.05). Jaccard overlap between nodes was set to 0.375. Size of the nodes correlate with number of genes within the BP and colours represent normalised enrichment scores (red is upregulated, and blue is downregulated). Circles encompassing the molecular nodes are an auto annotation of the larger BP of the clustered nodes.*

Table 7-3: *Top 10 most significant BP clusters likely to be regulated by promoter methylation of key genes at chronic HIIT (4WP)*

Cluster Name	**No. Nodes**	**FDR**
ECM organisation	5	1.92E-06
Collagen biosynthesis	4	6.77E-05
Bone morphogenesis	8	1.25E-04
Regulation of myeloid differentiation	11	3.18E-04
Muscle tissue development	9	1.86E-03
Muscle contraction	13	1.86E-03
Cartilage development	5	4.30E-03
Podosome assembly	2	8.71E-03
Regulation of viral process	6	8.71E-03
Response to virus	3	1.00E-02

7.6 SUMMARY

In summary, multiple epigenetic mechanisms were governing possible adaptations to exercise training. Firstly, DMPs were assessed at each exercise time point when compared with the baseline (PRE). Following differential methylation analysis (delta Beta (Δβ) ±2%, FDR<0.05), we identified a total of 1,138 (746 hypermethylated and 392 hypomethylated DMPs) for acute HIIE. For chronic HIIT, 7,470 probes were differentially methylated, of which, 2,371 DMPs were hypermethylated and 5,099 hypomethylated. We observed similar findings to established studies (150), with comparable methylation changes (i.e. <10% change in beta value). As previously discussed, in **Section 6.7**, the skeletal muscle microenvironment contains 11-16 different cell types which contribute nuclear material (381-383). These studies highlighted the inherent heterogeneity of cell types within skeletal muscle microenvironments, whereby ~25% of nuclei are not myonuclei. Therefore, the use of analytical methods that can assess more subtle, yet coordinated and consistent methylation changes are likely more informative for this tissue type. Emerging analytical approaches such as *mCSEA* aim to address this gap by assessing small but coordinated changes to methylation that are likely far more informative of biological function (372). Therefore, the strict and arbitrary threshold stated previously did not confound our analysis. Groups of probes were associated to promoter sites and the resulting genes were integrated with gene expression data to better capitulate biological function for acute HIIE and chronic HIIT.

The correlated genes for acute HIIE were extremely similar to the sole transcriptome data, where transcription factors such as *JUNB, EGR1, FOS, FOSB, EGR3,* and *MYC*, were upregulated with corresponding promoter hypomethylation. The upregulation of these genes has been identified previously for HIIE, and as per **Section 6.5.1**, it was evident that the skeletal muscle microenvironment was adjusting to the stressful large-scale intervention (HIIE). When assessing GOIs, we observed a reversion of gene expression to baseline for chronic HIIT, however, the DNA promoter methylation remained significantly hypomethylated. This finding suggested a mechanism whereby promoter methylation is reduced for acute exercise responsive genes in a chronic manner for sustained response to future bouts of HIIE.

For chronic HIIT, the correlated genes seemed to suggest a more functional adaptive response involving upregulation/hypermethylation of RNA processing (*BRUNOL6*) and protein

localisation genes (*DZIP1L*), and downregulation/hypermethylation of genes involved in calcium signalling (*MYOZ1*, *HRC*) and muscle differentiation and growth (*SPEG, PPAPDC3*). Our findings for chronic HIIT GOIs suggested that genes responsive to chronic HIIT are likely to have no involvement in acute response to HIIE.

The methylation regulated BPs associated with acute HIIE strongly overlapped the BP clusters identified from the sole transcriptome analysis (*Response to reactive oxygen species, Processes involved in development, Skeletal muscle tissue development,* and *Response to peptide*) and confirmed our findings that the skeletal muscle microenvironment was adjusting to stress. We also found that some BPs were similar to BP clusters identified for the three-hour acute HIIE time point (*Regulation of leukocyte differentiation,* and *Response to hypoxia*). Firstly, these results support previous findings, in which promoter hypomethylation of exercise responsive genes leads to upregulation of these genes at three-hours post exercise (137). Secondly, our results suggest that the BPs identified for acute HIIE are regulated by promoter methylation of key transcription factor genes that have been previously under-represented in the literature (284, 384).

For chronic HIIT, both upregulated and downregulated BPs were identified. Downregulated BP clusters were associated with muscle contraction and muscle tissue development, which represented a novel finding for response to sustained HIIT. Upregulated BP clusters contained similar immune BPs to the sole transcriptome data (*Regulation of leukocyte differentiation, Regulation of myeloid differentiation, Negative regulation of leukocyte differentiation, Neutrophil activation, Negative regulation of leukocyte apoptosis, and Regulation of leukocyte proliferation*). *ECM organisation* was noted as the most significant BP cluster that was likely to be regulated by promoter methylation of chronic HIIT responsive genes. This finding was consistent with existing literature where similar biological processes were identified for prolonged (six weeks), low volume HIIT (126). For this study, we have identified that a four-week period of moderate volume HIIT is sufficient to trigger these same biological responses, and indeed demonstrated that other, novel biological processes are epigenetically regulated in acute HIIE and chronic HIIT.

Chapter 8: Summary, Future Directions, and Concluding Remarks

This thesis explored the genetics, transcriptomics, and epigenetics behind adaptations to high intensity exercise and training. Very few exercise studies have examined MultiOmics when attempting to delineate the molecular processes of exercise adaptability. Indeed, exercise studies have been consistently beset by low participant numbers and consequently low reproducibility, especially within genetic studies. In addition, there remains a large gap in the literature for molecular response to long term HIIT, with most analyses being conducted on acute HIIE responses. Because of this, only two genetic variants (*ACTN3 R577X, ACE I/D*) have been consistently replicated within exercise studies and the bulk of the literature has focussed on these within the framework of exercise training even though there remain large knowledge gaps within the field. Transcriptomics have identified many genes associated with acute endurance exercise, however, many of these findings have not been replicated, and the studies do not usually expand the analysis to gene ontology or GSEA, which is more informative of biological function. DNA methylation has been shown to be an important regulatory process during acute exercise, however, studies to date have been limited in their assessment of longer term training. As such, the projects within this thesis framework were able to contribute a large amount of novel knowledge to the field of exercise science at large due to the longitudinal nature of the Gene SMART study, and the wide variety of physiological measures taken for each participant.

8.1 PREVIOUS EXERCISE GENETIC MARKERS WITHIN THE GENE SMART POPULATION

In **Chapter 4** of this thesis, this study aimed to investigate previously implicated exercise genetic variants within the moderately trained Gene SMART population. This chapter pertains to **Aim 1**: Genotype previously discovered SNP markers in the Gene SMART participants using a MassARRAY platform.

a. Research previously discovered exercise SNP markers to develop a list of potential candidates for the Gene SMART population
b. Design a MassARRAY with the highest multiplex possible to ensure cost effective genotyping
c. Perform a quantitative association of the Gene SMART genotypes to discover significant exercise genetic variants

In conjunction, a highly trained exercise cohort was utilised from the 2008 Ironman triathlon to discern genetic variants that were associated in different exercise groups. A MassARRAY was designed to genotype 36 variants that were previously associated with either exercise cohorts, or exercise intolerant disease states such as arterial hypertension, cardiovascular disease, and diabetes. 17 genetic variants were found to be nominally associated with exercise training in either the moderately trained or highly trained cohorts. Only one variant remained associated following FDR correction. The rs1474347 A allele within the *IL6* gene was found to be associated with positive change (4.016 mL/(kg·min)) in the VO_{2peak} response phenotype. Of note, the variant was identified within the triathlon group, but was not significantly associated with time to completion measures. Further, this variant was in strong linkage with a proximal SNP within the *IL6* gene, possibly accounting for the mixed evidence of association for this gene in the exercise field. Interestingly, the associated variant was upstream of an uncharacterised long non-coding RNA, which may reveal a post-translational epigenetic process for adaptations to training. This study was not able to replicate many previous findings from the literature, indicating a strong need for continued replication cohorts and studies within the field of exercise genetics. In addition, functional genomics (i.e.

identifying genetic marks that correspond to changes in mRNA or protein expression) may be a useful avenue of conducting genetic research on exercise participants. Such studies require large cohort sizes such as within the Athlome consortium and will invariably lead to functional annotation of novel exercise genetic loci.

In conclusion, the study was able to successfully complete thesis Aim 1, including all three sub aims. A single variant within a cytokine gene was identified that passed FDR and was associated with the VO_{2peak} response phenotype.

8.2 MITOCHONDRIAL AND MITOCHONDRIAL RELATED GENETIC VARIANTS INVOLVED IN EXERCISE ADAPTATIONS

Chapter 5 discussed the mitochondrial associated genetics that correlate with different exercise response outcomes. A knowledge gap surrounding mitochondrial genetic variants associated with exercise traits has been revealed as the number of studies assessing these variants are limited. Mitochondrial variants have the capacity to effect exercise traits due to the high number of energy maintenance genes on the mitochondrial genome. This chapter pertained to **Aim 2** within the thesis framework: Identify mtDNA haplogroups and mitochondrial variants that may affect exercise response.

a. Adapt and optimise a mitochondrial genome sequencing protocol to sequence the Gene SMART participants
b. Identify mtDNA haplogroups that are associated with exercise response
c. Identify mtDNA point variants that are associated with exercise response
d. Identify nuclear encoded mitochondrial variants that are associated with exercise response

Several of the response phenotypes were found to be correlated and as such this study used a principal component analysis to build composite traits and then used these to identify genetic variants that were influencing multiple traits but would not be identified in either of the individual traits. A method for mitochondrial genome sequencing was developed and optimised for this chapter and was published. Further, the Gene SMART population was genotyped using CoreExome-24 BeadChips. As such, all the SNPs within nuclear encoded mitochondrial genes were extracted to assess both the mitochondrial genome, and genetic factors that influence mitochondrial structure, function, and interaction with the cell. A total of 28 mitochondrial and 4,325 nuclear encoded mitochondrial associated variants passed the nominal significance thresholds for the various candidate gene association tests. No mitochondrial genetic variants were associated with exercise training at a statistically meaningful level (FDR<0.05). Further, the number of specific mitochondrial DNA haplogroups within the study was equivalent to the number of participants, and as such this study was unable to discern any association between

regional isolates with exercise response. However, the replication of these findings with a larger exercise cohort may reveal significant associations between the nominal variants and the respective exercise traits. Specifically, a mechanism was proposed by which the mitochondrial genetic variants may influence training adaptations. A tRNA gene coding for the leucine amino acid was noted and may have led to the differential regulation of this amino acid and therefore may play a role in adaptation processes. Further, there were notable synonymous mtDNA variants within key genes that also coded for the leucine amino acid. Most notably, the *ATP6* gene contained a synonymous variant for leucine and the codon change resulted in a codon with far less frequent use in mitochondrial coding genes. As tRNAs are costly with respect to intracellular ATP levels, it was hypothesised that this change may result in premature energy deficiency during and immediately post exercise training and therefore lead to differential regulation of normal downstream molecular pathway changes. In the nuclear encoded genetic association analysis, nine novel variants in eight different genes were identified *(rs11061368: DIABLO, rs113400963: FAM185A, rs6062129* and *rs6121949: MTG2, rs7231304: AFG3L2, rs2041840: NDUFAF7, rs7085433: TIMM23, rs1063271: SPTLC2, rs2275273: ALDH18A1)* that were significantly associated with the VO_{2peak} and LT measures. Each of the associated variants were assessed using the UCSC genome browser to ascertain the functional consequence of each variant to the residing or proximal gene. Of note, the variant identified within the *NDUFAF7* gene was correlated with the expression of the proximal *QPCT* gene. Further, the two intronic variants within the *MTG2* gene were within regulatory regions connected to the *MTG2* transcription start site. A non-coding transcript variant within the *TIMM23* gene was associated with the time trial phenotype. This was significant biologically as the gene is a target of the *NRF2* transcriptional activator and the gene product forms a part of the mitochondrial transmembrane proteins. The *SPTLC2* gene contained a 3`UTR variant, which was assessed for functional consequence using the STarMir database. A list of pre-existing exercise gene interacting miRNA sequences were used to determine the possible effect of the variant on miRNA binding properties. All of the miRNA sequences assessed were likely to bind to the 3`UTR region and it was therefore impossible to distinguish the exact mechanism by which the variant may affect exercise response. The genetic variants within the *DIABLO* and *FAM185A* genes were not associated with any regulatory elements and therefore this study was unable to elucidate the molecular mechanisms for these genes. The novel findings within this chapter represent new avenues for adaptations to exercise training that should be replicated in larger cohorts, and then functionally assessed with proteomics. To summarise, this study was able to successfully complete most of Aim 2. Sub-aim b) was unable to be completed as

the number of haplogroups within the Gene SMART participants was approximately equivalent to the total number of participants, and therefore there were no significant haplogroups associated with any particular response phenotype. The study was able to identify novel genetic variants associated with exercise response phenotypes and composite traits. In addition, the study implicated a novel mechanism by which leucine metabolism may be contributing to lack of exercise response.

8.3 TRANSCRIPTOMICS AND EXERCISE TRAINING

We assessed the transcriptomic changes between a single bout of HIIE and prolonged HIIT in **Chapter 6** of this thesis. This chapter pertained to **Aim 3** of this thesis: Utilise targeted RNA sequencing to identify longitudinal global gene expression changes and pathways in response to training.

a. QC RNA samples prior to experimentation to address any bias in gene expression values.
b. Perform targeted RNA sequencing using the AmpliSeq Transcriptome methodology.
c. Perform differential gene expression analysis based on the results from the sequencing.
d. Perform pathway analysis from the differentially regulated genes
e. Link molecular pathways changes to previous chapters and aims

Briefly, targeted RNA sequencing was utilised on an Ion Torrent platform to sequence the transcriptomes of 55 participants over four exercise time points (PRE: prior to training, P0: immediately post single bout, P3: three hours post initial bout, and 4WP: four weeks post initial bout after repeated bouts). Current pre-processing and analysis techniques were used to ensure the transcript data was robust and applicable to previous and future exercise studies. A total of 12,862 genes were expressed across all samples. Of these, 96 genes were differentially regulated at P0, 3,939 genes at P3, and 2,812 genes at 4WP. Firstly, the top DEGs in each of the time points were examined to ascertain the key adaptation responses over time. An overabundance of transcription factors were found to be upregulated immediately post training. This was likely due to the high amount of stress that HIIE training places on the skeletal muscle microenvironment. Three hours post-exercise, there was a notable increase in the magnitude of DEGs. The transcription factors seen in the immediate time point were also upregulated at three hours but to a lesser extent. This suggested that the skeletal muscle was starting to adjust to the microenvironment and begin the transcription of genes involved in adaptation rather than immediate stress response. Confidence was maintained that the results were applicable outside of the current study as replication of the upregulation status of the *NR4A3* gene was confirmed,

recently implicated from a large-scale meta-analysis (124). At the four-week time point, there was no evidence of transcription factors being highly upregulated. Further, an unusually dense cluster of histone protein subunit genes within the top upregulated gene lists was noted. The magnitude of DEGs was back to immediate post exercise levels. It was expected that the transcriptional changes observed in the four-week time point were more respective of true adaptation to prolonged exercise and therefore the focus of this chapter was primarily on chronic HIIT (4WP) time point, although the HIIE responses are discussed to a lesser extent. Following this genic level differential analysis, GSEA was performed to determine the biological process that were key for adaptation to training. Interestingly, the level of genic differential regulation was not correlated with the level of differentially regulated pathways at each time point. There were 96 differentially regulated GOBPs at P0, 848 at P3, and 1,147 at 4WP. The differentially regulated processes were assessed using Cytoscape and the gene ontology enrichment terms were grouped into super clusters to gauge the BPs at each time point.

At the P0 time point, the differentially regulated pathways were grouped into 15 separate pathway clusters. Most notably, pathways involving nuclear transport, angiogenesis, oxidative stress response and cell cycle were upregulated following HIIE. Protein localisation was the only pathway cluster that was downregulated at this early time point. It was evident from the immediate post exercise time point that the skeletal muscle niche was adjusting to the extreme exercise bout for each participant. Response pathways such as oxidative stress response, and heat shock response were upregulated immediately following HIIE. There was also evidence for upregulation of angiogenesis and mRNA processing/transcription.

At three-hours post HIIE, all significant pathways were upregulated and there were 27 pathway clusters seen on the enrichment maps. Six of these were involved in immune processes and were strongly linked to each other, indicating a very strong requirement for immune mediate adaptation processes at early exercise time points. Interestingly, protein degradation and apoptosis pathways were upregulated, further implicating immune mediated activation and adaptation. Further, we noted that reactive oxygen species (ROS) metabolism pathways were upregulated, signalling a change from oxidative stress immediately post exercise, to metabolism of ROS and clearing of the stress inducing metabolites at three hours.

As specified previously, the four-week HIIT time point contained the largest number of differentially regulated pathways. These were stratified into 42 BP clusters and there was a mixture of up and down regulated significant pathways. The upregulated clusters at four-weeks included immune processes, ECM development, skeletal development, and angiogenesis. Further, there were many more upregulated pathways nodes supporting differing avenues for exercise adaptations. Due to the large nature of this data, the study scrutinised the enriched pathways at four weeks that pertained to the previous findings within this thesis framework. The majority of the downregulated pathways were associated with RNA processing, gene expression, and mitochondrial function, indicating a large molecular change from immediate exercise response time points (upregulations in response to oxidative stress, large scale increases in translational processes) to downregulation of these processes as the cell adapts to longer term HIIT.

This study was able to identify previous molecular processes associated with acute HIIT such as inflammation, angiogenesis, cytokine signalling, and muscle hypertrophic processes respective of fibre type specificity. In addition, this study was able to identify many novel biological processes that were differentially regulated in conjunction with the previously outlined processes in acute HIIT. Perhaps the most important finding of this study was the amount of differentially regulated biological processes that were significantly enriched following four-weeks of HIIT. Evidence of cytokine and inflammation processes remained following four-weeks however the vast majority of processes represented novel findings within the field of exercise science.

Aim 3 was successfully completed including sub-aims **a)-d)**. All sample RINs were obtained prior to transcriptome sequencing. Differentially expressed genes were found to be associated with each of the exercise time points utilised for the study. Differentially regulated biological process changes were discovered in each time point using gene set enrichment analysis in conjunction with gene ontology. Sub-aim **e)** was also completed and is discussed further within **Section 8.5** of the current chapter.

8.4 DNA METHYLATION RESPONSE TO EXERCISE TRAINING

Epigenetics represents a connection between inheritance and environmental influence on adaptation processes. DNA methylation is an epigenetic mark, and has been shown to correlate with fitness through repeated bouts of training. Within the current study, longitudinal changes in methylation were assessed in response to a single bout of HIIE and four-weeks of HIIT. The findings within **Chapter 7** correspond to **Aim 4** of this thesis:

a. Ascertain the quality of the methylation signals across arrays, samples, and time points
b. Discover differentially methylated probes and loci at each exercise time point
c. Discover differentially methylated genes at each exercise time point

The analysis for **Chapter 7** was not focussed on individual probes but rather the small and coordinated changes in promoter methylation that lead to changes in biological function. Therefore, the change in methylation threshold issues stated previously (**Section 7.6**) did not confound the analysis.

Firstly, differentially methylated probes were assessed at each exercise time point when compared with the baseline (PRE). Following differential methylation analysis (delta Beta ($\Delta\beta$) ±2%, FDR<0.05), we identified a total of 1,138 (746 hypermethylated and 392 hypomethylated differentially methylated probes (DMPs)) for acute HIIE. For chronic HIIT, 7,470 probes were differentially methylated, of which, 2,371 DMPs were hypermethylated and 5,099 hypomethylated. When considering the promoter regions (as determined by *mCSEA*), we found that TSS1500 and 5`UTR regions showed an initial drop in DMPs following HIIE (P0), but these returned to baseline for chronic HIIT (4WP). In contrast, DMP levels for TSS200 and 1st exon regions dropped immediately following HIIE (P0), and the drop was maintained in chronic HIIT (4WP).

8.5 INTERCONNECTIVITY BETWEEN THESIS CHAPTERS

True MultiOmics within this study were not able to be performed due to time constraints, including the large amount of data generated and analysed as part of this PhD's main aims. However, correlations between the functional data and the genomic results obtained within **Chapters 4** and **5** was investigated. Firstly, transcriptomic data was utilised to perform this rudimentary MultiOmics. In **Chapter 4**, a SNP proximal to the *IL6* gene was identified to be significantly associated and negatively correlated with higher VO_{2peak} threshold measurements within the study. Interestingly, there was no evidence of *IL6* expression within the study, indicating a probable repression of the gene, either due to the aforementioned SNP, or increased prevalence of other myokine mechanisms within the context of response to training. Further, the *ZFP36* gene was found to be upregulated in response to training, which is a negative regulator of *IL6* expression at the mRNA level, also representing a possible mechanism by which repression of *IL6* may contribute to adaptations to exercise in both short and long term HIIT.

In **Chapter 5** of this thesis, several genetic markers were found to be associated with exercise response. Specifically, an intronic SNP (rs2041840) within the *NDUFAF7* gene was identified as significantly associated with better responses within the VO_{2peak}, peak power, and composite trait response phenotypes. When the enriched molecular pathways pertaining to mitochondrial function were assessed, a negative enrichment and therefore downregulation of mitochondrial biological function was observed at the four-week exercise time point. Interestingly, the levels of *NDUFAF7* did not change over the exercise time points however the differences between expression based on genotype were not assessed within the current study. From the GSEA, the '*Mitochondrial gene expression'* GOBP was the most significantly enriched at the four-week time point with the largest gene-set size. Of note, the most enriched genes within the pathway at the four-week time point were associated with insulin signalling through the *MICAL-L2* gene. Briefly, the product of this gene forms a complex with the Rab13 protein, which allows for the binding of *ACTN2* in an insulin dependent manner.

Identifying correlations between epigenetics and transcriptomics remains elusive, particularly in cases with strong interventions such as exercise training. With the current

understanding of epigenetic mechanisms, the study aimed to identify genes involved in the regulation of epigenetic marks (histone methylation/DNA methylation) for exercise adaptation over different time points. The *'Macromolecule methylation'* molecular pathway was identified to be highly enriched at the four-week time point. Firstly, the well-known transcription factors *MYC, FOS*, and *BTG2* were highly upregulated following a single bout of training at both the immediate and three-hour time points. Secondly, large scale changes in gene expression three-hours post HIIE were noted even though the molecular pathway was non-significant at this time point. Further, histone protein methyltransferases were found to be highly upregulated, in conjunction with extremely downregulated levels of histone demethylase proteins. In particular, the *EZH2* gene was proposed as a central epigenetic mechanism to early stress response pathways. Interestingly, similar processes in the four-week time point were observed although different methyltransferase and demethylation genes were evident. The differentially regulated genes within this pathway at four-weeks were largely uncharacterised in exercise training. These findings represent novel mechanisms for epigenetic regulation of prolonged adaptation to training and should be explored in future studies.

Following this transcriptomic based MultiOmic approach, the study aimed to assess correlations between the epigenetic and transcriptomic data sets. We utilised *mCSEAIntegrate* to discover DNA promoter methylation that was negatively associated with gene expression (i.e. upregulated gene expression with corresponding promoter hypermethylation). We identified 122, and 276 significant (FDR<0.05) genes that were likely to be regulated via promoter methylation for acute HIIE (P0) and chronic HIIT (4WP) respectively. The correlated genes for acute HIIE were extremely similar to the sole transcriptome data, where transcription factors such as *JUNB, EGR1, FOS, FOSB, EGR3*, and *MYC*, were upregulated with corresponding promoter hypomethylation. For chronic HIIT, the correlated genes seemed to suggest a more functional adaptive response involving upregulation/hypermethylation of RNA processing (*BRUNOL6*) and protein localisation genes (*DZIP1L*), and downregulation/hypermethylation of genes involved in calcium signalling (*MYOZ1, HRC*) and muscle differentiation and growth (*SPEG, PPAPDC3*). The methylation regulated biological processes associated with acute HIIE were like BP clusters identified from the sole transcriptome analysis (response to reactive oxygen species, processes involved in development, skeletal muscle tissue development, and response to peptide). For chronic HIIT, both upregulated and downregulated BPs were identified. Upregulated BP clusters contained

similar immune BPs to the sole transcriptome data (*Regulation of leukocyte differentiation, Regulation of myeloid differentiation, Negative regulation of leukocyte differentiation, Neutrophil activation, negative regulation of leukocyte apoptosis, and Regulation of leukocyte proliferation*). Downregulated BP clusters were associated with muscle contraction and muscle tissue development, which represented a novel finding for response to sustained HIIT.

8.6 LIMITATIONS

Whilst this compilation of large-scale projects may introduce novel and exciting results to the field of exercise science, there were several limitations that may affect the scope and clarity of the findings within this thesis.

8.6.1 Sample size

Firstly, the sample size for aspects of this study was far too small to see genome-wide significance (Chapter 5) within our genetic data (i.e. n<100 compared to n>100,000). *A priori* power calculations were performed for the mitochondrial project that indicated the associated variants remained statistically significant, however the size of the cohort may have convoluted the results and induced false negatives within the data set. The study aimed to limit this factor by increasing statistical power through quantitative association, and the consideration of nominally significant variants.

8.6.2 Bias

The covariates that form the basis of the inclusion criteria for the Gene SMART study are also representative of bias that leads to limitations in the results from each project. Excluding females from the study limited the potential implications of the research to half of the population. This has become a matter of contention within exercise cohorts as the inclusion of female athletes/exercise participants is rare in the literature (151). Conversely, the current study was limited by sex to prevent skewing of the data by hormone influenced genetic and transcriptomic factors. Further, the Gene SMART inclusion criteria was limited to Caucasians and therefore only respective of one ethnicity. Whilst this largely matches the current research and opens avenues for large-scale meta-analysis, the reduced representation of other ethnicities further increases the disparity between accurate results, and results that are representative of entire populations. As such, there remains a need to carefully balance the inclusion of potential confounding covariates with the potential of representing all population groups.

8.6.3 Constraints

The work within this thesis represents Omic level findings and as such the large quantity of data was difficult to assess within the timeframe of a PhD project. Further, there does not yet exist a robust analysis methodology for MultiOmic data (385). Arguably, cancer genomics is currently at the forefront of Omic level analysis techniques (386), however this primarily extends only to mass mutation genomics and splice variant transcriptomics. Even so, each Omic dataset was assessed independently from the others and the subsequent investigation of any correlation between differentially regulated molecular pathways was assessed with the genetic and epigenetic data.

8.6.4 Skeletal muscle biology

As previously mentioned in **Section 6.7**, skeletal muscle is a unique tissue whereby many nuclei may be contributing to one cellular environment. Recent studies have proved this with single cell and single nuclei transcriptomic profiling of skeletal muscle (381-383). Firstly, they found that skeletal muscle tissue contains ~75% myonuclei, with the other ~25% consisting of endothelial cells, mesenchymal progenitors (fibro-adipogenic progenitors (FAP) cells), and resident immune cells. As such, myonuclei were likely the predominant nuclear material contributing to the gene expression and DNA methylation data within our study. Secondly, the single cell studies found that individual myonuclei within the same skeletal muscle myofiber show gene expression level heterogeneity. Therefore, these myonuclei likely show differing DNA methylation patterns also, which may have led to the observation of low beta changes within our study, and previous work (149, 150). Even so, the presence of multiple cell types makes it difficult to assess which cell type is responsible for the expression changes to exercise. For instance, the expression changes may be a result of changes in cell-type proportions, rather than response to intervention. Advances in deconvolution analysis with respect to human skeletal muscle have recently provided promise of a proxy to assess cell type changes for gene expression data from complex tissues (18).

In addition, muscle damage due to exercise (especially from overload training) may lead to recruitment of satellite cells for healing and growth. As such, this multinucleation may be enhanced by repeated bouts of training. For genomic approaches, the above should be considered when examining mitochondrial heteroplasmy.

8.7 FUTURE DIRECTIONS

Despite the large quantity of work that has formed this thesis, there were many avenues of analysis and understanding that were not feasible to assess during the timeframe of this project. As such, several possible (and strongly suggested) future directions for the large quantity of data produced are detailed in the following.

8.7.1 Replication and Validation

As stated previously and within each individual chapter of this thesis, replication cohorts within the exercise genetics field are rare. Further, specific processes are often not as significant in replication cohorts for several reasons. Chiefly, trained cohorts typically contain individuals with large variability in environmental factors such as diet, sleep, and habitual physical activity patterns. As a future direction, we suggest that the results obtained within this thesis be replicated in similar cohorts such as the ATHLOME Consortium (218, 233).

The genetic results obtained from this thesis have currently not been validated by the current gold standard: Sanger sequencing. Whilst this is becoming less necessary within genetic studies as methodologies become more robust, variants discovered during exploratory studies such as within **Chapter 5** of this thesis should be heavily scrutinised in future studies. Additionally, the transcriptomic (**Chapter 6**) and epigenetic (**Chapter 7)** changes could be validated through RT-qPCR, and bisulfite pyrosequencing, respectively. Further validation and extension of these changes using Western blotting and/or immunohistochemistry would also provide insights into associated proteomic changes.

8.7.2 Differing Avenues of Analysis

Due to the time constraints of the projects within this thesis, every avenue of potentially interesting analysis within each project was unable to be explored. Firstly, the phenotypic data analysed within this study contained several instances of participant dropout, and as such the current analysis was limited by participants that had completed rather than total numbers. This issue may be fixed through missing data PCA based imputation techniques. Missing phenotypic data imputation for the drop out samples from the Gene SMART study should be conducted using the '*MissMDA*' R/Bioconductor package prior to further Omic analysis. This will

contribute novel information to each of the projects outlined in this thesis. Further, the genomic data obtained from the CoreExome chips as outlined in **Chapter 5** may be imputed with larger genetic data sets to increase the number of genetic variants being assessed. Whilst this will likely increase the multiple testing burden, the results will reveal a greater clarity of the exact genes being associated with each response trait. These may then be functionally assessed and will deliver more promising results than less specific genomic tests.

8.7.3 Further testing

As discussed within **Section 8.6.3**, time constraints were prevalent during the projects within this thesis. As such, and partly due to the large-scale nature of the data sets generated for this thesis, it was impossible to explore every single avenue of analysis. Whilst all available samples were utilised for transcriptomic analysis, there were cost constraints for the methylation project within this thesis. As such, a representative subset of the Gene SMART study participants and time points were selected. If further funding could be garnered for the methylation aspect of this study, then the data sets and therefore MultiOmic analysis would be based on a larger cohort, and more complete data sets.

8.8 CONCLUDING REMARKS

The findings from this thesis represent novel and exciting results for the field of exercise science at large. Further, the results discussed within the framework of this thesis represent a summary of the overall findings from each of the projects. The Omic level data generated from this thesis will continue to be useful for the exercise genetics community and will assist with future meta-analyses that aim to delineate the adaptive response to training.

Chapter 9: References

1. Keith SW, Redden DT, Katzmarzyk PT, Boggiano MM, Hanlon EC, Benca RM, et al. Putative contributors to the secular increase in obesity: exploring the roads less traveled. Int J Obes (Lond). 2006;30(11):1585-94.

2. Francis K. The burden of physical inactivity & cardiovascular heart disease. Compr Ther. 1998;24(2):87-92.

3. Hegde BM. Health benefits of exercise. J Assoc Physicians India. 2003;51:297-8.

4. Kokkinos P, Myers J, Faselis C, Panagiotakos DB, Doumas M, Pittaras A, et al. Exercise capacity and mortality in older men: a 20-year follow-up study. Circulation. 2010;122(8):790-7.

5. Dishman RK, Berthoud HR, Booth FW, Cotman CW, Edgerton VR, Fleshner MR, et al. Neurobiology of exercise. 2006;14(3):345-56.

6. Hawley JA, Hargreaves M, Joyner MJ, Zierath JR. Integrative biology of exercise. Cell. 2014;159(4):738-49.

7. Dugan SA. Exercise for health and wellness at midlife and beyond: balancing benefits and risks. Phys Med Rehabil Clin N Am. 2007;18(3):555-75, xi.

8. Courneya KS, Segal RJ, McKenzie DC, Dong H, Gelmon K, Friedenreich CM, et al. Effects of exercise during adjuvant chemotherapy on breast cancer outcomes. Med Sci Sports Exerc. 2014;46(9):1744-51.

9. Wiskemann J, Huber GJBmt. Physical exercise as adjuvant therapy for patients undergoing hematopoietic stem cell transplantation. 2008;41(4):321-9.

10. Tucker JM, Welk GJ, Beyler NKJAjopm. Physical activity in US adults: compliance with the physical activity guidelines for Americans. 2011;40(4):454-61.

11. Pratt M, Epping JN, Dietz WHJPm. Putting physical activity into public health: a historical perspective from the CDC. 2009;49(4):301-2.

12. Katzmarzyk PT, Janssen I. The economic costs associated with physical inactivity and obesity in Canada: an update. Can J Appl Physiol. 2004;29(1):90-115.

13. Cadilhac DA, Cumming TB, Sheppard L, Pearce DC, Carter R, Magnus A. The economic benefits of reducing physical inactivity: an Australian example. Int J Behav Nutr Phys Act. 2011;8(1):99.

14. Morton JP, Kayani AC, McArdle A, Drust B. The exercise-induced stress response of skeletal muscle, with specific emphasis on humans. Sports Med. 2009;39(8):643-62.

15. Cermak NM, Snijders T, McKay BR, Parise G, Verdijk LB, Tarnopolsky MA, et al. Eccentric exercise increases satellite cell content in type II muscle fibers. Med Sci Sports Exerc. 2013;45(2):230-7.

16. Frontera WR, Ochala J. Skeletal muscle: a brief review of structure and function. Calcif Tissue Int. 2015;96(3):183-95.

17. Pandorf CE, Haddad F, Owerkowicz T, Carroll LP, Baldwin KM, Adams GR. Regulation of myosin heavy chain antisense long noncoding RNA in human vastus lateralis in response to exercise training. Am J Physiol Cell Physiol. 2020;318(5):C931-c42.

18. Rubenstein AB, Smith GR, Raue U, Begue G, Minchev K, Ruf-Zamojski F, et al. Single-cell transcriptional profiles in human skeletal muscle. Sci Rep. 2020;10(1):229.

19. Petrany MJ, Swoboda CO, Sun C, Chetal K, Chen X, Weirauch MT, et al. Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. Nat Commun. 2020;11(1):6374.

20. Pollard TD, Weihing RR. Actin and myosin and cell movement. CRC Crit Rev Biochem. 1974;2(1):1-65.

21. Higashi-Fujime S. Active movement of bundles of actin and myosin filaments from muscle: a simple model for cell motility. Cold Spring Harb Symp Quant Biol. 1982;46 Pt 1:69-75.

22. Zhang Y, Yu XJ, Chen WS, Gao HL, Liu KL, Shi XL, et al. Exercise training attenuates renovascular hypertension partly via RAS- ROS- glutamate pathway in the hypothalamic paraventricular nucleus. Sci Rep. 2016;6:37467.

23. Wall BT, Dirks ML, van Loon LJ. Skeletal muscle atrophy during short-term disuse: implications for age-related sarcopenia. Ageing Res Rev. 2013;12(4):898-906.

24. Psilander N, Eftestøl E, Cumming KT, Juvkam I, Ekblom MM, Sunding K, et al. Effects of training, detraining, and retraining on strength, hypertrophy, and myonuclear number in human skeletal muscle. J Appl Physiol (1985). 2019;126(6):1636-45.

25. Hannan AL, Hing W, Simas V, Climstein M, Coombes JS, Jayasinghe R, et al. High-intensity interval training versus moderate-intensity continuous training within cardiac rehabilitation: a systematic review and meta-analysis. 2018;9:1.

26. Amundsen BH, Rognmo O, Hatlen-Rebhan G, Slordahl SA. High-intensity aerobic exercise improves diastolic function in coronary artery disease. Scand Cardiovasc J. 2008;42(2):110-7.

27. Moholdt TT, Amundsen BH, Rustad LA, Wahba A, Lovo KT, Gullikstad LR, et al. Aerobic interval training versus continuous moderate exercise after coronary artery bypass surgery: a randomized study of cardiovascular effects and quality of life. Am Heart J. 2009;158(6):1031-7.

28. Rocco EA, Prado DM, Silva AG, Lazzari JM, Bortz PC, Rocco DF, et al. Effect of continuous and interval exercise training on the PETCO2 response during a graded exercise test in patients with coronary artery disease. Clinics (Sao Paulo). 2012;67(6):623-8.

29. Currie KD, Dubberley JB, McKelvie RS, MacDonald MJ. Low-volume, high-intensity interval training in patients with CAD. Med Sci Sports Exerc. 2013;45(8):1436-42.

30. Boukabous I, Marcotte-Chénard A, Amamou T, Boulay P, Brochu M, Tessier D, et al. Low-Volume High-Intensity Interval Training (HIIT) versus Moderate-Intensity Continuous Training on Body Composition, Cardiometabolic Profile and Physical Capacity in Older Women. J Aging Phys Act. 2019;27(4):879-89.

31. MacInnis MJ, Gibala MJ. Physiological adaptations to interval training and the role of exercise intensity. J Physiol. 2017;595(9):2915-30.

32. Gibala MJ, Jones AM. Physiological and performance adaptations to high-intensity interval training. Limits of human endurance. 76: Karger Publishers; 2013. p. 51-60.

33. Zuhl M, Kravitz LJFJ. HIIT vs continuous endurance training: battle of the aerobic titans. 2012.

34. Gibala M. Molecular responses to high-intensity interval exercise. Appl Physiol Nutr Metab. 2009;34(3):428-32.

35. Granata C, Jamnick NA, Bishop DJJSM. Principles of exercise prescription, and how they influence exercise-induced changes of transcription factors and other regulators of mitochondrial biogenesis. 2018;48(7):1541-59.

36. Gibala MJ, McGee SL, Garnham AP, Howlett KF, Snow RJ, Hargreaves M. Brief intense interval exercise activates AMPK and p38 MAPK signaling and increases the expression of PGC-1alpha in human skeletal muscle. J Appl Physiol (1985). 2009;106(3):929-34.

37. MacDougall JD, Hicks AL, MacDonald JR, McKelvie RS, Green HJ, Smith KM. Muscle performance and enzymatic adaptations to sprint interval training. J Appl Physiol (1985). 1998;84(6):2138-42.

38. Egan B, Zierath JR. Exercise metabolism and the molecular regulation of skeletal muscle adaptation. Cell Metab. 2013;17(2):162-84.

39. Chin ER. The role of calcium and calcium/calmodulin-dependent kinases in skeletal muscle plasticity and mitochondrial biogenesis. The Proceedings of the Nutrition Society. 2004;63(2):279-86.

40. Hoshino D, Kitaoka Y, Hatta HJTJoPF, Medicine S. High-intensity interval training enhances oxidative capacity and substrate availability in skeletal muscle. 2016;5(1):13-23.

41. Kao SC, Westfall DR, Soneson J, Gurd B, Hillman CH. Comparison of the acute effects of high-intensity interval training and continuous aerobic walking on inhibitory control. Psychophysiology. 2017;54(9):1335-45.

42. Kaspar F, Jelinek HF, Perkins S, Al-Aubaidy HA, deJong B, Butkowski E. Acute-Phase Inflammatory Response to Single-Bout HIIT and Endurance Training: A Comparative Study. Mediators Inflamm. 2016;2016:5474837.

43. Lira FS, Dos Santos T, Caldeira RS, Inoue DS, Panissa VLG, Cabral-Santos C, et al. Short-Term High- and Moderate-Intensity Training Modifies Inflammatory and Metabolic Factors in Response to Acute Exercise. Front Physiol. 2017;8:856.

44. Olney N, Wertz T, LaPorta Z, Mora A, Serbas J, Astorino TAJTJoS, et al. Comparison of acute physiological and psychological responses between moderate-intensity continuous exercise and three regimes of high-intensity interval training. 2018;32(8):2130-8.

45. Drigny J, Gremeaux V, Guiraud T, Gayda M, Juneau M, Nigam A. Long-term high-intensity interval training associated with lifestyle modifications improves QT dispersion parameters in metabolic syndrome patients. Ann Phys Rehabil Med. 2013;56(5):356-70.

46. Azuma K, Osawa Y, Tabata S, Katsukawa F, Ishida H, Oguma Y, et al. Decrease in regional body fat after long-term high-intensity interval training. 2017;6(2):103-10.

47. Li F-H, Sun L, Wu D-S, Gao H-E, Min ZJA. Proteomics-based identification of different training adaptations of aged skeletal muscle following long-term high-intensity interval and moderate-intensity continuous training in aged rats. 2019;11(12):4159-82.

48. Torma F, Gombos Z, Jokai M, Takeda M, Mimura T, Radak ZJSM, et al. High intensity interval training and molecular adaptive response of skeletal muscle. 2019;1(1):24-32.

49. Ferreira LF, Laitano O. Regulation of NADPH oxidases in skeletal muscle. Free Radic Biol Med. 2016;98:18-28.

50. Rush JW. Exercising an option to prevent age related decline of vascular BH4 and uncoupling of eNOS. J Physiol. 2009;587(Pt 15):3755.

51. Sachdev S, Davies KJ. Production, detection, and adaptive responses to free radicals in exercise. Free Radic Biol Med. 2008;44(2):215-23.

52. St-Pierre J, Drori S, Uldry M, Silvaggi JM, Rhee J, Jager S, et al. Suppression of reactive oxygen species and neurodegeneration by the PGC-1 transcriptional coactivators. Cell. 2006;127(2):397-408.

53. Bishop DJ, Granata C, Eynon N. Can we optimise the exercise training prescription to maximise improvements in mitochondria function and content? Biochimica et Biophysica Acta (BBA) - General Subjects. 2014;1840(4):1266-75.

54. Daussin FN, Zoll J, Dufour SP, Ponsot E, Lonsdorfer-Wolf E, Doutreleau S, et al. Effect of interval versus continuous training on cardiorespiratory and mitochondrial functions: relationship to aerobic performance

improvements in sedentary subjects. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology. 2008;295(1):R264-R72.

55. Wagner PD. Skeletal muscle angiogenesis. Hypoxia: Springer; 2001. p. 21-38.

56. Wagner PD. Skeletal muscle angiogenesis. A possible role for hypoxia. Adv Exp Med Biol. 2001;502:21-38.

57. Rodas G, Ventura JL, Cadefau JA, Cussó R, Parra JJEjoap. A short training programme for the rapid improvement of both aerobic and anaerobic metabolism. 2000;82(5-6):480-6.

58. Scribbans TD, Edgett BA, Vorobej K, Mitchell AS, Joanisse SD, Matusiak JB, et al. Fibre-specific responses to endurance and low volume high intensity interval training: striking similarities in acute and chronic adaptation. PLoS One. 2014;9(6):e98119.

59. Chinsomboon J, Ruas J, Gupta RK, Thom R, Shoag J, Rowe GC, et al. The transcriptional coactivator PGC-1alpha mediates exercise-induced angiogenesis in skeletal muscle. Proc Natl Acad Sci U S A. 2009;106(50):21401-6.

60. Tan R, Nederveen JP, Gillen JB, Joanisse S, Parise G, Tarnopolsky MA, et al. Skeletal muscle fiber-type-specific changes in markers of capillary and mitochondrial content after low-volume interval training in overweight women. Physiological Reports. 2018;6(5):e13597.

61. Burkewitz K, Zhang Y, Mair William B. AMPK at the Nexus of Energetics and Aging. Cell Metabolism. 2014;20(1):10-25.

62. Place N, Ivarsson N, Venckunas T, Neyroud D, Brazaitis M, Cheng AJ, et al. Ryanodine receptor fragmentation and sarcoplasmic reticulum Ca²⁺ leak after one session of high-intensity interval exercise. Proceedings of the National Academy of Sciences. 2015;112(50):15492.

63. Kelly DP, Scarpulla RC. Transcriptional regulatory circuits controlling mitochondrial biogenesis and function. Genes Dev. 2004;18(4):357-68.

64. Little JP, Safdar A, Bishop D, Tarnopolsky MA, Gibala MJ. An acute bout of high-intensity interval training increases the nuclear abundance of PGC-1alpha and activates mitochondrial biogenesis in human skeletal muscle. Am J Physiol Regul Integr Comp Physiol. 2011;300(6):R1303-10.

65. Tan R, Nederveen JP, Gillen JB, Joanisse S, Parise G, Tarnopolsky MA, et al. Skeletal muscle fiber-type-specific changes in markers of capillary and mitochondrial content after low-volume interval training in overweight women. Physiol Rep. 2018;6(5):e13597.

66. Summermatter S, Santos G, Pérez-Schindler J, Handschin CJPotNAoS. Skeletal muscle PGC-1α controls whole-body lactate homeostasis through estrogen-related receptor α-dependent activation of LDH B and repression of LDH A. 2013;110(21):8738-43.

67. Fiorenza M, Gunnarsson TP, Hostrup M, Iaia FM, Schena F, Pilegaard H, et al. Metabolic stress-dependent regulation of the mitochondrial biogenic molecular response to high-intensity exercise in human skeletal muscle. J Physiol. 2018;596(14):2823-40.

68. Burgomaster KA, Howarth KR, Phillips SM, Rakobowchuk M, Macdonald MJ, McGee SL, et al. Similar metabolic adaptations during exercise after low volume sprint interval and traditional endurance training in humans. J Physiol. 2008;586(1):151-60.

69. Perry CG, Heigenhauser GJ, Bonen A, Spriet LL. High-intensity aerobic interval training increases fat and carbohydrate metabolic capacities in human skeletal muscle. Appl Physiol Nutr Metab. 2008;33(6):1112-23.

70. Wood G, Murrell A, van der Touw T, Smart N. HIIT is not superior to MICT in altering blood lipids: a systematic review and meta-analysis. BMJ Open Sport Exerc Med. 2019;5(1):e000647.

71. Das DK, Graham ZA, Cardozo CP. Myokines in skeletal muscle physiology and metabolism: Recent advances and future perspectives. Acta Physiol (Oxf). 2020;228(2):e13367.

72. Pedersen BKJCP. Muscle as a secretory organ. 2013;3(3):1337-62.

73. King NA, Hopkins M, Caudwell P, Stubbs RJ, Blundell JE. Individual variability following 12 weeks of supervised exercise: identification and characterization of compensation for exercise-induced weight loss. Int J Obes (Lond). 2008;32(1):177-84.

74. Ross R, Goodpaster BH, Koch LG, Sarzynski MA, Kohrt WM, Johannsen NM, et al. Precision exercise medicine: understanding exercise response variability. Br J Sports Med. 2019;53(18):1141-53.

75. Sarzynski MA, Ghosh S, Bouchard C. Genomic and transcriptomic predictors of response levels to endurance exercise training. J Physiol. 2017;595(9):2931-9.

76. Bouchard C, Sarzynski MA, Rice TK, Kraus WE, Church TS, Sung YJ, et al. Genomic predictors of the maximal O(2) uptake response to standardized exercise training programs. J Appl Physiol (1985). 2011;110(5):1160-70.

77. Bray MS, Hagberg JM, Perusse L, Rankinen T, Roth SM, Wolfarth B, et al. The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. Med Sci Sports Exerc. 2009;41(1):35-73.

78. Hulten B, Thorstensson A, Sjödin B, Karlsson JJAPS. Relationship between isometric endurance and fibre types in human leg muscles. 1975;93(1):135-8.

79. Vincent B, De Bock K, Ramaekers M, Van den Eede E, Van Leemputte M, Hespel P, et al. ACTN3 (R577X) genotype is associated with fiber type distribution. 2007;32(1):58-63.

80. Eynon N, Ruiz JR, Oliveira J, Duarte JA, Birk R, Lucia A. Genes and elite athletes: a roadmap for future research. J Physiol. 2011;589(Pt 13):3063-70.

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

82. Grealy R, Herruer J, Smith CL, 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):e0145171.

83. Del Coso J, Hiam D, Houweling P, Pérez LM, Eynon N, Lucía AJEjoap. More than a 'speed gene': ACTN3 R577X genotype, trainability, muscle damage, and the risk for injuries. 2019;119(1):49-60.

84. Broos S, Malisoux L, Theisen D, Van Thienen R, Ramaekers M, Jamart C, et al. Evidence for ACTN3 as a speed gene in isolated human muscle fibers. 2016;11(3):e0150594.

85. Norman B, Esbjornsson M, Rundqvist H, Osterlund T, Glenmark B, Jansson E. ACTN3 genotype and modulation of skeletal muscle response to exercise in human subjects. J Appl Physiol (1985). 2014;116(9):1197-203.

86. Delmonico MJ, Kostek MC, Doldo NA, Hand BD, Walsh S, Conway JM, et al. Alpha-actinin-3 (ACTN3) R577X polymorphism influences knee extensor peak power response to strength training in older men and women. J Gerontol A Biol Sci Med Sci. 2007;62(2):206-12.

87. Norman B, Esbjornsson M, Rundqvist H, Osterlund T, von Walden F, Tesch PA. Strength, power, fiber types, and mRNA expression in trained men and women with different ACTN3 R577X genotypes. J Appl Physiol (1985). 2009;106(3):959-65.

88. Niemi AK, Majamaa K. Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. Eur J Hum Genet. 2005;13(8):965-9.

89. Cooper R, McFarlane-Anderson N, Bennett FI, Wilks R, Puras A, Tewksbury D, et al. ACE, angiotensinogen and obesity: a potential pathway leading to hypertension. J Hum Hypertens. 1997;11(2):107-11.

90. Collins M, Xenophontos SL, Cariolou MA, Mokone GG, Hudson DE, Anastasiades L, et al. The ACE gene and endurance performance during the South African Ironman Triathlons. Med Sci Sports Exerc. 2004;36(8):1314-20.

91. Amir O, Amir R, Yamin C, Attias E, Eynon N, Sagiv M, et al. The ACE deletion allele is associated with Israeli elite endurance athletes. 2007;92(5):881-6.

92. Pescatello LS, Kostek MA, Gordish-Dressman H, Thompson PD, Seip RL, Price TB, et al. ACE ID genotype and the muscle strength and size response to unilateral resistance training. Med Sci Sports Exerc. 2006;38(6):1074-81.

93. Rankinen T, Wolfarth B, Simoneau J-A, Maier-Lenz D, Rauramaa R, Rivera MA, et al. No association between the angiotensin-converting enzyme ID polymorphism and elite endurance athlete status. 2000;88(5):1571-5.

94. Papadimitriou ID, Lockey SJ, Voisin S, Herbert AJ, Garton F, Houweling PJ, et al. No association between ACTN3 R577X and ACE I/D polymorphisms and endurance running times in 698 Caucasian athletes. 2018;19(1):13.

95. Pulkes T, Liolitsa D, Wills AJ, Hargreaves I, Heales S, Hanna MG. Nonsense mutations in mitochondrial DNA associated with myalgia and exercise intolerance. Neurology. 2005;64(6):1091-2.

96. DiMauro S, Andreu AL. Mutations in mitochondrial DNA as a cause of exercise intolerance. Ann Med. 2001;33(7):472-6.

97. Andreu AL, Hanna MG, Reichmann H, Bruno C, Penn AS, Tanji K, et al. Exercise intolerance due to mutations in the cytochrome b gene of mitochondrial DNA. N Engl J Med. 1999;341(14):1037-44.

98. Bouchard C, Leblanc J, Cote J, Jobin J, Jobin M, Labrie A, et al. Familial resemblance in catecholamine changes to cold stress and maximal exercise. Hum Hered. 1983;33(3):170-8.

99. Simonen RL, Perusse L, Rankinen T, Rice T, Rao DC, Bouchard C. Familial aggregation of physical activity levels in the Quebec Family Study. Med Sci Sports Exerc. 2002;34(7):1137-42.

100. Simonen RL, Rankinen T, Perusse L, Rice T, Rao DC, Chagnon Y, et al. Genome-wide linkage scan for physical activity levels in the Quebec Family study. Med Sci Sports Exerc. 2003;35(8):1355-9.

101. Loos RJ, Rankinen T, Tremblay A, Perusse L, Chagnon Y, Bouchard C. Melanocortin-4 receptor gene and physical activity in the Quebec Family Study. Int J Obes (Lond). 2005;29(4):420-8.

102. Sarzynski MA, Rankinen T, Sternfeld B, Fornage M, Sidney S, Bouchard C. SNP-by-fitness and SNP-by-BMI interactions from seven candidate genes and incident hypertension after 20 years of follow-up: the CARDIA Fitness Study. J Hum Hypertens. 2011;25(8):509-18.

103. Sarzynski MA, Rankinen T, Sternfeld B, Grove ML, Fornage M, Jacobs Jr DR, et al. Association of SNPs from 17 Candidate Genes with Baseline Symptom-Limited Exercise Test Duration and Decrease in Duration over 20 Years: The CARDIA Fitness Study. 2010;3(6):531.

104. Rivera MA, Dionne FT, Simoneau JA, Perusse L, Chagnon M, Chagnon Y, et al. Muscle-specific creatine kinase gene polymorphism and VO2max in the HERITAGE Family Study. Med Sci Sports Exerc. 1997;29(10):1311-7.

105. Angelopoulos TJ, Lowndes J, Seip RL. Genetics of the adaptation to exercise. World Rev Nutr Diet. 2011;102:144-9.

106. Bouchard C, Daw EW, Rice T, Perusse L, Gagnon J, Province MA, et al. Familial resemblance for VO2max in the sedentary state: the HERITAGE family study. Med Sci Sports Exerc. 1998;30(2):252-8.

107. Bouchard C, Leon AS, Rao DC, Skinner JS, Wilmore JH, Gagnon J. The HERITAGE family study. Aims, design, and measurement protocol. Med Sci Sports Exerc. 1995;27(5):721-9.

108. Rankinen T, An P, Rice T, Sun G, Chagnon YC, Gagnon J, et al. Genomic scan for exercise blood pressure in the Health, Risk Factors, Exercise Training and Genetics (HERITAGE) Family Study. Hypertension. 2001;38(1):30-7.

109. Rankinen T, Sung YJ, Sarzynski MA, Rice TK, Rao DC, Bouchard C. Heritability of submaximal exercise heart rate response to exercise training is accounted for by nine SNPs. J Appl Physiol (1985). 2012;112(5):892-7.

110. Wilmore JH, Leon AS, Rao DC, Skinner JS, Gagnon J, Bouchard C. Genetics, response to exercise, and risk factors: the HERITAGE Family Study. World Rev Nutr Diet. 1997;81:72-83.

111. Ried JS, Jeff MJ, Chu AY, Bragg-Gresham JL, van Dongen J, Huffman JE, et al. A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape. Nat Commun. 2016;7:13357.

112. Lanouette CM, Chagnon YC, Rice T, Perusse L, Muzzin P, Giacobino JP, et al. Uncoupling protein 3 gene is associated with body composition changes with training in HERITAGE study. J Appl Physiol (1985). 2002;92(3):1111-8.

113. An P, Borecki IB, Rankinen T, Perusse L, Leon AS, Skinner JS, et al. Evidence of major genes for exercise heart rate and blood pressure at baseline and in response to 20 weeks of endurance training: the HERITAGE family study. Int J Sports Med. 2003;24(7):492-8.

114. Panagiotou OA, Ioannidis JP, Genome-Wide Significance P. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. Int J Epidemiol. 2012;41(1):273-86.

115. Williams CJ, Li Z, Harvey N, Lea RA, Gurd BJ, Bonafiglia JT, et al. Genome wide association study of response to interval and continuous exercise training: the Predict-HIIT study. J Biomed Sci. 2021;28(1):37.

116. Rankinen T, Wolfarth B, Simoneau JA, Maier-Lenz D, Rauramaa R, Rivera MA, et al. No association between the angiotensin-converting enzyme ID polymorphism and elite endurance athlete status. J Appl Physiol (1985). 2000;88(5):1571-5.

117. Doring FE, Onur S, Geisen U, Boulay MR, Perusse L, Rankinen T, et al. ACTN3 R577X and other polymorphisms are not associated with elite endurance athlete status in the Genathlete study. J Sports Sci. 2010;28(12):1355-9.

118. Doring F, Onur S, Kurbitz C, Boulay MR, Perusse L, Rankinen T, et al. Single nucleotide polymorphisms in the myostatin (MSTN) and muscle creatine kinase (CKM) genes are not associated with elite endurance performance. Scand J Med Sci Sports. 2011;21(6):841-5.

119. Wolfarth B, Rankinen T, Muhlbauer S, Ducke M, Rauramaa R, Boulay MR, et al. Endothelial nitric oxide synthase gene polymorphism and elite endurance athlete status: the Genathlete study. Scand J Med Sci Sports. 2008;18(4):485-90.

120. Li X, Wang SJ, Tan SC, Chew PL, Liu L, Wang L, et al. The A55T and K153R polymorphisms of MSTN gene are associated with the strength training-induced muscle hypertrophy among Han Chinese men. J Sports Sci. 2014;32(9):883-91.

121. Kikuchi N, Tsuchiya Y, Nakazato K, Ishii N, Ochi E. Effects of the ACTN3 R577X Genotype on the Muscular Strength and Range of Motion Before and After Eccentric Contractions of the Elbow Flexors. Int J Sports Med. 2018;39(2):148-53.

122. Sailors MH, Jackson AS, McFarlin BK, Turpin I, Ellis KJ, Foreyt JP, et al. Exposing college students to exercise: the Training Interventions and Genetics of Exercise Response (TIGER) study. J Am Coll Health. 2010;59(1):13-20.

123. Herring MP, Fernandez JR, Jackson AS, Mary SH, O'Connor DP, Dishman RK, et al. Variation In The FTO Gene Predicts Exercise Response In The Training Interventions And Genetics Of Exercise Response (TIGER) Study. Medicine and Science in Sports and Exercise. 2013;45(5):61-.

124. Pillon NJ, Gabriel BM, Dollet L, Smith JAB, Sardon Puig L, Botella J, et al. Transcriptomic profiling of skeletal muscle adaptations to exercise and inactivity. Nat Commun. 2020;11(1):470.

125. Mizuiri S, Hemmi H, Kumanomidou H, Iwamoto M, Miyagi M, Sakai K, et al. Angiotensin-converting enzyme (ACE) I/D genotype and renal ACE gene expression. Kidney Int. 2001;60(3):1124-30.

126. Miyamoto-Mikami E, Tsuji K, Horii N, Hasegawa N, Fujie S, Homma T, et al. Gene expression profile of muscle adaptation to high-intensity intermittent exercise training in young men. Sci Rep. 2018;8(1):16811.

127. Jaksik R, Iwanaszko M, Rzeszowska-Wolny J, Kimmel MJBd. Microarray experiments and factors which affect their reliability. 2015;10(1):46.

128. Kothapalli R, Yoder SJ, Mane S, Loughran TP. Microarray results: how accurate are they? BMC Bioinformatics. 2002;3(1):22.

129. Stevens JR, Herrick JS, Wolff RK, Slattery ML. Power in pairs: assessing the statistical value of paired samples in tests for differential expression. BMC Genomics. 2018;19(1):953.

130. Wolffe AP, Matzke MA. Epigenetics: regulation through repression. Science. 1999;286(5439):481-6.

131. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16(1):6-21.

132. Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature. 1993;366(6453):362-5.

133. Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. Science. 2001;293(5532):1068-70.

134. Bird A. The essentials of DNA methylation. Cell. 1992;70(1):5-8.

135. Jones PA. The DNA methylation paradox. Trends Genet. 1999;15(1):34-7.

136. Angeloni A, Bogdanovic O. Enhancer DNA methylation: implications for gene regulation. Essays Biochem. 2019;63(6):707-15.

137. Barres R, Yan J, Egan B, Treebak JT, Rasmussen M, Fritz T, et al. Acute exercise remodels promoter methylation in human skeletal muscle. Cell Metab. 2012;15(3):405-11.

138. Narkar VA. PGC1α Promoter Methylation and Nucleosome Repositioning: Insights Into Exercise and Metabolic Regulation in Skeletal Muscle. Endocrinology. 2017;158(7):2084-5.

139. Bajpeyi S, Covington JD, Taylor EM, Stewart LK, Galgani JE, Henagan TM. Skeletal Muscle PGC1alpha -1 Nucleosome Position and -260 nt DNA Methylation Determine Exercise Response and Prevent Ectopic Lipid Accumulation in Men. Endocrinology. 2017;158(7):2190-9.

140. Edgett BA, Foster WS, Hankinson PB, Simpson CA, Little JP, Graham RB, et al. Dissociation of increases in PGC-1alpha and its regulators from exercise intensity and muscle activation following acute exercise. PLoS One. 2013;8(8):e71623.

141. Barres R, Osler ME, Yan J, Rune A, Fritz T, Caidahl K, et al. Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density. Cell Metab. 2009;10(3):189-98.

142. Alibegovic AC, Sonne MP, Højbjerre L, Bork-Jensen J, Jacobsen S, Nilsson E, et al. Insulin resistance induced by physical inactivity is associated with multiple transcriptional changes in skeletal muscle in young men. 2010;299(5):E752-E63.

143. Voisin S, Eynon N, Yan X, Bishop DJ. Exercise training and DNA methylation in humans. Acta Physiol (Oxf). 2015;213(1):39-59.

144. Pilegaard H, Neufer PD. Transcriptional regulation of pyruvate dehydrogenase kinase 4 in skeletal muscle during and after exercise. The Proceedings of the Nutrition Society. 2004;63(2):221-6.

145. Wende AR, Huss JM, Schaeffer PJ, Giguere V, Kelly DP. PGC-1alpha coactivates PDK4 gene expression via the orphan nuclear receptor ERRalpha: a mechanism for transcriptional control of muscle glucose metabolism. Mol Cell Biol. 2005;25(24):10684-94.

146. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8(3):389-99.

147. Lindholm ME, Marabita F, Gomez-Cabrero D, Rundqvist H, Ekstrom TJ, Tegner J, et al. An integrative analysis reveals coordinated reprogramming of the epigenome and the transcriptome in human skeletal muscle after training. Epigenetics. 2014;9(12):1557-69.

148. Nitert MD, Dayeh T, Volkov P, Elgzyri T, Hall E, Nilsson E, et al. Impact of an exercise intervention on DNA methylation in skeletal muscle from first-degree relatives of patients with type 2 diabetes. Diabetes. 2012;61(12):3322-32.

149. Seaborne R, Strauss J, Cocks M, Shepherd S, O'Brien T, van Someren K, et al. Human skeletal muscle possesses an epigenetic memory of hypertrophy. Sci Rep 8: 1898. 2018.

150. Robinson MM, Dasari S, Konopka AR, Johnson ML, Manjunatha S, Esponda RR, et al. Enhanced Protein Translation Underlies Improved Metabolic and Physical Adaptations to Different Exercise Training Modes in Young and Old Humans. Cell Metab. 2017;25(3):581-92.

151. Halperin I, Vigotsky AD, Foster C, Pyne DB. Strengthening the Practice of Exercise and Sport-Science Research. Int J Sports Physiol Perform. 2018;13(2):127-34.

152. Yan X, Eynon N, Papadimitriou ID, Kuang J, Munson F, Tirosh O, et al. The gene SMART study: method, study design, and preliminary findings. BMC Genomics. 2017;18(Suppl 8):821.

153. Hautala AJ, Kiviniemi AM, Makikallio TH, Kinnunen H, Nissila S, Huikuri HV, et al. Individual differences in the responses to endurance and resistance training. Eur J Appl Physiol. 2006;96(5):535-42.

154. Bouchard C, Rankinen T. Individual differences in response to regular physical activity. Med Sci Sports Exerc. 2001;33(6 Suppl):S446-51; discussion S52-3.

155. Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, et al. Familial aggregation of VO(2max) response to exercise training: results from the HERITAGE Family Study. J Appl Physiol (1985). 1999;87(3):1003-8.

156. Islam H, Gurd BJ. Exercise response variability: Random error or true differences in exercise response? Exp Physiol. 2020;105(12):2022-4.

157. Bishop D, Jenkins DG, McENIERY M, Carey MFJM, sports si, exercise. Relationship between plasma lactate parameters and muscle characteristics in female cyclists. 2000;32(6):1088.

158. Bishop D, Jenkins DG, Mackinnon LTJM, sports si, exercise. The relationship between plasma lactate parameters, Wpeak and 1-h cycling performance in women. 1998;30(8):1270-5.

159. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. Journal of statistical software. 2008;25(1):1-18.

160. Harvey NR, Voisin S, Lea RA, Yan X, Benton MC, Papadimitriou ID, et al. Investigating the influence of mtDNA and nuclear encoded mitochondrial variants on high intensity interval training outcomes. Sci Rep. 2020;10(1):11089.

161. Hart SN, Therneau TM, Zhang Y, Poland GA, Kocher JP. Calculating sample size estimates for RNA sequencing data. J Comput Biol. 2013;20(12):970-8.

162. Li M, Zou D, Li Z, Gao R, Sang J, Zhang Y, et al. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. Nucleic Acids Res. 2019;47(D1):D983-D8.

163. De Feo PJN, Metabolism, Diseases C. Is high-intensity exercise better than moderate-intensity exercise for weight loss? 2013;23(11):1037-42.

164. O'Donoghue G, Blake C, Cunningham C, Lennon O, Perrotta C. What exercise prescription is optimal to improve body composition and cardiorespiratory fitness in adults living with obesity? A network meta-analysis. Obes Rev. 2021;22(2):e13137-e.

165. Boutcher SH. High-intensity intermittent exercise and fat loss. J Obes. 2011;2011:868305.

166. Voisin S, Jacques M, Lucia A, Bishop DJ, Eynon N. Statistical Considerations for Exercise Protocols Aimed at Measuring Trainability. Exerc Sport Sci Rev. 2019;47(1):37-45.

167. Issurin VB. Evidence-Based Prerequisites and Precursors of Athletic Talent: A Review. Sports Med. 2017;47(10):1993-2010.

168. den Hoed M, Brage S, Zhao JH, Westgate K, Nessa A, Ekelund U, et al. Heritability of objectively assessed daily physical activity and sedentary behavior. Am J Clin Nutr. 2013;98(5):1317-25.

169. Yan X, Dvir N, Jacques M, Cavalcante L, Papadimitriou ID, Munson F, 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):923-30.

170. Houweling PJ, Papadimitriou ID, Seto JT, Perez LM, Coso JD, North KN, 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):1774-87.

171. Bouchard C. Overcoming barriers to progress in exercise genomics. Exerc Sport Sci Rev. 2011;39(4):212-7.

172. McCoy J, Bates M, Eggett C, Siervo M, Cassidy S, Newman 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):e000632.

173. Buford TW, Roberts MD, Church TS. Toward exercise as personalized medicine. Sports Med. 2013;43(3):157-65.

174. Grealy R, Smith CL, Chen T, Hiller D, Haseler LJ, Griffiths LR. The genetics of endurance: frequency of the ACTN3 R577X variant in Ironman World Championship athletes. J Sci Med Sport. 2013;16(4):365-71.

175. Ellis JA, Ong B. The MassARRAY((R)) System for Targeted SNP Genotyping. Methods Mol Biol. 2017;1492:77-94.

176. Jurinke C, van den Boom D, Cantor CR, Koster H. Automated genotyping using the DNA MassArray technology. Methods Mol Biol. 2002;187:179-92.

177. Harvey NR, Voisin S, Dunn PJ, Sutherland H, Yan X, Jacques M, et al. Genetic variants associated with exercise performance in both moderately trained and highly trained individuals. Mol Genet Genomics. 2020;295(2):515-23.

178. Ahmetov, II, Fedotovskaya ON. Current Progress in Sports Genomics. Adv Clin Chem. 2015;70:247-314.

179. De Moor MH, Liu YJ, Boomsma DI, Li J, Hamilton JJ, Hottenga JJ, et al. Genome-wide association study of exercise behavior in Dutch and American adults. Med Sci Sports Exerc. 2009;41(10):1887-95.

180. Rankinen T, Perusse L, Rauramaa R, Rivera MA, Wolfarth B, Bouchard C. The human gene map for performance and health-related fitness phenotypes: the 2003 update. Med Sci Sports Exerc. 2004;36(9):1451-69.

181. Walsh S, Haddad CJ, Kostek MA, Angelopoulos TJ, Clarkson PM, Gordon PM, 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):66-70.

182. He ZH, Hu Y, Li YC, Gong LJ, Cieszczyk P, Maciejewska-Karlowska A, 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):184-95.

183. Garatachea N, Santiago C, Yvert T, Verde-Rello Z, Fiuza-Luces C, Santos-Lozano A, et al. Genetic variants in the PPARD-PPARGC1A-NRF-TFAM mitochondriogenesis pathway are neither associated with muscle characteristics nor physical performance in elderly. [Variaciones genéticas en la vía de la mitocondriogénesis PPARD-PPARGC1A-NRF-TFAM no están asociadas ni con características musculares ni con rendimiento físico en personas mayores]. RICYDE Revista internacional de ciencias del deporte. 2015;11(41):196-208.

184. He Z, Hu Y, Feng L, Li Y, Liu G, Xi Y, et al. NRF-1 genotypes and endurance exercise capacity in young Chinese men. Br J Sports Med. 2008;42(5):361-6.

185. Montani D, Girerd B, Gunther S, Riant F, Tournier-Lasserve E, Magy L, et al. Pulmonary arterial hypertension in familial hemiplegic migraine with ATP1A2 channelopathy. Eur Respir J. 2014;43(2):641-3.

186. Panicker V, Cluett C, Shields B, Murray A, Parnell KS, Perry JR, 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):3075-81.

187. Fuku N, Alis R, Yvert T, Zempo H, Naito H, Abe Y, et al. Muscle-Related Polymorphisms (MSTN rs1805086 and ACTN3 rs1815739) Are Not Associated with Exceptional Longevity in Japanese Centenarians. PLoS One. 2016;11(11):e0166605.

188. Chathoth S, Ismail MH, Vatte C, Cyrus C, Al Ali Z, Ahmed KA, et al. Association of Uncoupling Protein 1 (UCP1) gene polymorphism with obesity: a case-control study. BMC Med Genet. 2018;19(1):203.

189. Ramos AV, Bastos-Rodrigues L, Resende BA, Friedman E, Campanha-Versiani L, Miranda DM, et al. The contribution of FTO and UCP-1 SNPs to extreme obesity, diabetes and cardiovascular risk in Brazilian individuals. BMC Med Genet. 2012;13:101.

190. Brondani LA, Assmann TS, Duarte GC, Gross JL, Canani LH, Crispim D. 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):215-25.

191. Rankinen T, Church T, Rice T, Markward N, Leon AS, Rao DC, et al. Effect of endothelin 1 genotype on blood pressure is dependent on physical activity or fitness levels. Hypertension. 2007;50(6):1120-5.

192. Panoulas VF, Douglas KM, Smith JP, Taffe P, Stavropoulos-Kalinoglou A, Toms TE, et al. Polymorphisms of the endothelin-1 gene associate with hypertension in patients with rheumatoid arthritis. Endothelium. 2008;15(4):203-12.

193. Arora P, Garcia-Bailo B, Dastani Z, Brenner D, Villegas A, Malik S, et al. Genetic polymorphisms of innate immunity-related inflammatory pathways and their association with factors related to type 2 diabetes. BMC Med Genet. 2011;12:95.

194. Liu Y, Berthier-Schaad Y, Fallin MD, Fink NE, Tracy RP, Klag MJ, et al. IL-6 haplotypes, inflammation, and risk for cardiovascular disease in a multiethnic dialysis cohort. J Am Soc Nephrol. 2006;17(3):863-70.

195. Alves AJ, Goldhammer E, Ribeiro F, Eynon N, Ben-Zaken Cohen S, Duarte JA, et al. GNAS A-1121G variant is associated with improved diastolic dysfunction in response to exercise training in heart failure patients. Int J Sports Med. 2013;34(3):274-80.

196. Burguete-Garcia AI, Martinez-Nava GA, Valladares-Salgado A, Bermudez Morales VH, Estrada-Velasco B, Wacher N, 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):1327-34.

197. Pan Q, Delahanty LM, Jablonski KA, Knowler WC, Kahn SE, Florez JC, et al. Variation at the melanocortin 4 receptor gene and response to weight-loss interventions in the diabetes prevention program. Obesity (Silver Spring). 2013;21(9):E520-6.

198. Tejedor MT, Garcia-Sobreviela MP, Ledesma M, Arbones-Mainar JM. The apolipoprotein E polymorphism rs7412 associates with body fatness independently of plasma lipids in middle aged men. PLoS One. 2014;9(9):e108605.

199. Liang KY, Mintun MA, Fagan AM, Goate AM, Bugg JM, Holtzman DM, et al. Exercise and Alzheimer's disease biomarkers in cognitively normal older adults. Ann Neurol. 2010;68(3):311-8.

200. Deeny SP, Poeppel D, Zimmerman JB, Roth SM, Brandauer J, Witkowski S, et al. Exercise, APOE, and working memory: MEG and behavioral evidence for benefit of exercise in epsilon4 carriers. Biol Psychol. 2008;78(2):179-87.

201. Fredriksson J, Anevski D, Almgren P, Sjogren M, Lyssenko V, Carlson J, et al. Variation in GYS1 interacts with exercise and gender to predict cardiovascular mortality. PLoS One. 2007;2(3):e285.

202. Hennis PJ, O'Doherty AF, Levett DZ, Grocott MP, Montgomery HM. Genetic factors associated with exercise performance in atmospheric hypoxia. Sports Med. 2015;45(5):745-61.

203. Brown BM, Rainey-Smith SR, Castalanelli N, Gordon N, Markovic S, Sohrabi HR, 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):562-70.

204. Lin Y, Gu SJ, Wu M, Chen Q, Zhou ZY, Yu H, et al. [Association between peroxisome proliferator-activated receptors gene polymorphism and essential hypertension]. Zhonghua Liu Xing Bing Xue Za Zhi. 2012;33(6):597-601.

205. Uthurralt J, Gordish-Dressman H, Bradbury M, Tesi-Rocha C, Devaney J, Harmon B, et al. PPARalpha L162V underlies variation in serum triglycerides and subcutaneous fat volume in young males. BMC Med Genet. 2007;8:55.

206. Andrulionyte L, Kuulasmaa T, Chiasson JL, Laakso M, Group S-NS. 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):1181-6.

207. Doney AS, Fischer B, Lee SP, Morris AD, Leese G, Palmer CN. 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;3:4.

208. Dong C, Zhou H, Shen C, Yu LG, Ding Y, Zhang YH, et al. Role of peroxisome proliferator-activated receptors gene polymorphisms in type 2 diabetes and metabolic syndrome. World J Diabetes. 2015;6(4):654-61.

209. Evans D, Aberle J, Wendt D, Wolf A, Beisiegel U, Mann WA. 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):198-204.

210. Flavell DM, Ireland H, Stephens JW, Hawe E, Acharya J, Mather H, et al. Peroxisome proliferator-activated receptor alpha gene variation influences age of onset and progression of type 2 diabetes. Diabetes. 2005;54(2):582-6.

211. Feser WJ, Fingerlin TE, Strand MJ, Glueck DH. Calculating Average Power for the Benjamini-Hochberg Procedure. J Stat Theory Appl. 2009;8(3):325-52.

212. Bonetti DL, Hopkins WG. Variation in performance times of elite flat-water canoeists from race to race. Int J Sports Physiol Perform. 2010;5(2):210-7.

213. Munoz-Canoves P, Scheele C, Pedersen BK, Serrano AL. Interleukin-6 myokine signaling in skeletal muscle: a double-edged sword? FEBS J. 2013;280(17):4131-48.

214. Karanikolou A, Wang G, Pitsiladis Y. Letter to the editor: A genetic-based algorithm for personalized resistance training. Biol Sport. 2017;34(1):31-3.

215. Woods D, Hickman M, Jamshidi Y, Brull D, Vassiliou V, Jones A, et al. Elite swimmers and the D allele of the ACE I/D polymorphism. Hum Genet. 2001;108(3):230-2.

216. Galy O, Hue O, Chamari K, Boussana A, Chaouachi A, Prefaut C. Influence of performance level on exercise-induced arterial hypoxemia during prolonged and successive exercise in triathletes. Int J Sports Physiol Perform. 2008;3(4):482-500.

217. Zervou S, Whittington HJ, Ostrowski PJ, Cao F, Tyler J, Lake HA, et al. Increasing creatine kinase activity protects against hypoxia / reoxygenation injury but not against anthracycline toxicity in vitro. PLoS One. 2017;12(8):e0182994.

218. Pitsiladis YP, Tanaka M, Eynon N, Bouchard C, North KN, Williams AG, et al. Athlome Project Consortium: a concerted effort to discover genomic and other "omic" markers of athletic performance. Physiol Genomics. 2016;48(3):183-90.

219. Vellers HL, Kleeberger SR, Lightfoot JT. Inter-individual variation in adaptations to endurance and resistance exercise training: genetic approaches towards understanding a complex phenotype. Mamm Genome. 2018;29(1-2):48-62.

220. Bishop DJ, Botella J, Genders AJ, Lee MJ, Saner NJ, Kuang J, et al. High-Intensity Exercise and Mitochondrial Biogenesis: Current Controversies and Future Research Directions. Physiology (Bethesda). 2019;34(1):56-70.

221. Williams CJ, Williams MG, Eynon N, Ashton KJ, Little JP, Wisloff U, et al. Genes to predict VO2max trainability: a systematic review. BMC Genomics. 2017;18(Suppl 8):831.

222. Jacques M, Hiam D, Craig J, Barres R, Eynon N, Voisin S. Epigenetic changes in healthy human skeletal muscle following exercise- a systematic review. Epigenetics. 2019;14(7):633-48.

223. Hood DA, Memme JM, Oliveira AN, Triolo M. Maintenance of Skeletal Muscle Mitochondria in Health, Exercise, and Aging. Annu Rev Physiol. 2019;81:19-41.

224. Wyckelsma VL, Levinger I, McKenna MJ, Formosa LE, Ryan MT, Petersen AC, et al. Preservation of skeletal muscle mitochondrial content in older adults: relationship between mitochondria, fibre type and high-intensity exercise training. J Physiol. 2017;595(11):3345-59.

225. Seo DY, Lee SR, Kim N, Ko KS, Rhee BD, Han J. Age-related changes in skeletal muscle mitochondria: the role of exercise. Integr Med Res. 2016;5(3):182-6.

226. Bishop DJ, Granata C, Eynon N. Can we optimise the exercise training prescription to maximise improvements in mitochondria function and content? Biochim Biophys Acta. 2014;1840(4):1266-75.

227. Nicholls TJ, Gustafsson CM. Separating and Segregating the Human Mitochondrial Genome. Trends Biochem Sci. 2018;43(11):869-81.

228. Eynon N, Moran M, Birk R, Lucia A. The champions' mitochondria: is it genetically determined? A review on mitochondrial DNA and elite athletic performance. Physiol Genomics. 2011;43(13):789-98.

229. Shockett PE, Khanal J, Sitaula A, Oglesby C, Meachum WA, Castracane VD, et al. Plasma cell-free mitochondrial DNA declines in response to prolonged moderate aerobic exercise. Physiol Rep. 2016;4(1).

230. McCann BJ, Tuppen HA, Kusters B, Lammens M, Smeitink JA, Taylor RW, et al. A novel mitochondrial DNA m.7507A>G mutation is only pathogenic at high levels of heteroplasmy. Neuromuscul Disord. 2015;25(3):262-7.

231. Grady JP, Murphy JL, Blakely EL, Haller RG, Taylor RW, Turnbull DM, et al. Accurate measurement of mitochondrial DNA deletion level and copy number differences in human skeletal muscle. PLoS One. 2014;9(12):e114462.

232. Spendiff S, Reza M, Murphy JL, Gorman G, Blakely EL, Taylor RW, et al. Mitochondrial DNA deletions in muscle satellite cells: implications for therapies. Hum Mol Genet. 2013;22(23):4739-47.

233. Vlahovich N, Hughes DC, Griffiths LR, Wang G, Pitsiladis YP, Pigozzi F, et al. Genetic testing for exercise prescription and injury prevention: AIS-Athlome consortium-FIMS joint statement. BMC Genomics. 2017;18(Suppl 8):818.

234. Martinez-Redondo D, Marcuello A, Casajus JA, Ara I, Dahmani Y, Montoya J, et al. Human mitochondrial haplogroup H: the highest VO2max consumer--is it a paradox? Mitochondrion. 2010;10(2):102-7.

235. Marcuello A, Martinez-Redondo D, Dahmani Y, Casajus JA, Ruiz-Pesini E, Montoya J, et al. Human mitochondrial variants influence on oxygen consumption. Mitochondrion. 2009;9(1):27-30.

236. Dumoulin R, Sagnol I, Ferlin T, Bozon D, Stepien G, Mousson B. A novel gly290asp mitochondrial cytochrome b mutation linked to a complex III deficiency in progressive exercise intolerance. Mol Cell Probes. 1996;10(5):389-91.

237. Dionne FT, Turcotte L, Thibault MC, Boulay MR, Skinner JS, Bouchard C. Mitochondrial DNA sequence polymorphism, VO2max, and response to endurance training. Med Sci Sports Exerc. 1993;25(7):766-74.

238. Zhang F, Lupski JR. Non-coding genetic variants in human disease. Hum Mol Genet. 2015;24(R1):R102-10.

239. Cooper DN. Functional intronic polymorphisms: Buried treasure awaiting discovery within our genes. Hum Genomics. 2010;4(5):284-8.

240. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, et al. Library construction for next-generation sequencing: overviews and challenges. Biotechniques. 2014;56(2):61-4, 6, 8, passim.

241. Deiner K, Renshaw MA, Li Y, Olds BP, Lodge DM, Pfrender ME, et al. Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. Methods in Ecology and Evolution. 2017;8(12):1888-98.

242. De Fanti S, Vianello D, Giuliani C, Quagliariello A, Cherubini A, Sevini F, et al. Massive parallel sequencing of human whole mitochondrial genomes with Ion Torrent technology: an optimized workflow for Anthropological and Population Genetics studies. Mitochondrial DNA A DNA Mapp Seq Anal. 2017;28(6):843-50.

243. Knierim E, Lucke B, Schwarz JM, Schuelke M, Seelow D. Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. PLoS One. 2011;6(11):e28240.

244. Harvey NR, Albury CL, Stuart S, Benton MC, Eccles DA, Connell JR, et al. Ion torrent high throughput mitochondrial genome sequencing (HTMGS). PLoS One. 2019;14(11):e0224847.

245. Brandon MC, Ruiz-Pesini E, Mishmar D, Procaccio V, Lott MT, Nguyen KC, et al. MITOMASTER: a bioinformatics tool for the analysis of mitochondrial DNA sequences. Hum Mutat. 2009;30(1):1-6.

246. Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, Mishmar D, et al. An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Res. 2007;35(Database issue):D823-8.

247. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res. 2018;27(2):e1608.

248. Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. 2016;44(D1):D1251-7.

249. Falk MJ, Pierce EA, Consugar M, Xie MH, Guadalupe M, Hardy O, et al. Mitochondrial disease genetic diagnostics: optimized whole-exome analysis for all MitoCarta nuclear genes and the mitochondrial genome. Discov Med. 2012;14(79):389-99.

250. Stuart S, Benton MC, Eccles DA, Sutherland HG, Haupt LM, Lea RA, et al. Gene-centric analysis implicates nuclear encoded mitochondrial protein gene variants in migraine susceptibility. Mol Genet Genomic Med. 2017;5(2):157-63.

251. Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE, et al. A mitochondrial protein compendium elucidates complex I disease biology. Cell. 2008;134(1):112-23.

252. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336-7.

253. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford). 2017;2017.

254. Mitochondrial Solarplot [Internet]. 2017 [cited 30 Jan 2019]. Available from: https://github.com/stephenturner/solarplot

255. Zurita Rendon O, Silva Neiva L, Sasarman F, Shoubridge EA. The arginine methyltransferase NDUFAF7 is essential for complex I assembly and early vertebrate embryogenesis. Hum Mol Genet. 2014;23(19):5159-70.

256. Speedy HE, Di Bernardo MC, Sava GP, Dyer MJ, Holroyd A, Wang Y, et al. A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. Nat Genet. 2014;46(1):56-60.

257. Pourteymour S, Eckardt K, Holen T, Langleite T, Lee S, Jensen J, et al. Global mRNA sequencing of human skeletal muscle: Search for novel exercise-regulated myokines. Mol Metab. 2017;6(4):352-65.

258. Cesnekova J, Rodinova M, Hansikova H, Zeman J, Stiburek L. Loss of Mitochondrial AAA Proteases AFG3L2 and YME1L Impairs Mitochondrial Structure and Respiratory Chain Biogenesis. Int J Mol Sci. 2018;19(12).

259. Mancini C, Hoxha E, Iommarini L, Brussino A, Richter U, Montarolo F, et al. Mice harbouring a SCA28 patient mutation in AFG3L2 develop late-onset ataxia associated with enhanced mitochondrial proteotoxicity. Neurobiol Dis. 2019;124:14-28.

260. Prieto-Ruiz JA, Alis R, Garcia-Benlloch S, Saez-Atienzar S, Ventura I, Hernandez-Andreu JM, et al. Expression of the human TIMM23 and TIMM23B genes is regulated by the GABP transcription factor. Biochim Biophys Acta Gene Regul Mech. 2018;1861(2):80-94.

261. Eynon N, Ruiz JR, Bishop DJ, Santiago C, Gomez-Gallego F, Lucia A, et al. The rs12594956 polymorphism in the NRF-2 gene is associated with top-level Spanish athlete's performance status. J Sci Med Sport. 2013;16(2):135-9.

262. Eynon N, Ruiz JR, Meckel Y, Moran M, Lucia A. Mitochondrial biogenesis related endurance genotype score and sports performance in athletes. Mitochondrion. 2011;11(1):64-9.

263. Steri M, Idda ML, Whalen MB, Orru V. Genetic variants in mRNA untranslated regions. Wiley Interdiscip Rev RNA. 2018;9(4):e1474.

264. Silva GJJ, Bye A, El Azzouzi H, Wisloff U. MicroRNAs as Important Regulators of Exercise Adaptation. Prog Cardiovasc Dis. 2017;60(1):130-51.

265. Han G, Gupta SD, Gable K, Niranjanakumari S, Moitra P, Eichler F, et al. Identification of small subunits of mammalian serine palmitoyltransferase that confer distinct acyl-CoA substrate specificities. Proc Natl Acad Sci U S A. 2009;106(20):8186-91.

266. Alexaki A, Gupta SD, Majumder S, Kono M, Tuymetova G, Harmon JM, et al. Autophagy regulates sphingolipid levels in the liver. J Lipid Res. 2014;55(12):2521-31.

267. De Paepe ME, Chu S, Hall SJ, McDonnell-Clark E, Heger NE, Schorl C, et al. Intussusceptive-like angiogenesis in human fetal lung xenografts: Link with bronchopulmonary dysplasia-associated microvascular dysangiogenesis? Exp Lung Res. 2015;41(9):477-88.

268. Zhu Y, Gu X, Xu C. Mitochondrial DNA 7908-8816 region mutations in maternally inherited essential hypertensive subjects in China. BMC Med Genomics. 2018;11(1):89.

269. Moraes CT, Ciacci F, Bonilla E, Jansen C, Hirano M, Rao N, et al. Two novel pathogenic mitochondrial DNA mutations affecting organelle number and protein synthesis. Is the tRNA(Leu(UUR)) gene an etiologic hot spot? J Clin Invest. 1993;92(6):2906-15.

270. Jia W, Higgs PG. Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. Mol Biol Evol. 2008;25(2):339-51.

271. Ahmetov, II, Druzhevskaya AM, Astratenkova IV, Popov DV, Vinogradova OL, Rogozkin VA. The ACTN3 R577X polymorphism in Russian endurance athletes. Br J Sports Med. 2010;44(9):649-52.

272. Amorim CE, Acuna-Alonzo V, Salzano FM, Bortolini MC, Hunemeier T. Differing evolutionary histories of the ACTN3*R577X polymorphism among the major human geographic groups. PLoS One. 2015;10(2):e0115449.

273. Döring FE, Onur S, Geisen U, Boulay MR, Pérusse L, Rankinen T, et al. ACTN3 R577X and other polymorphisms are not associated with elite endurance athlete status in the Genathlete study. Journal of sports sciences. 2010;28(12):1355-9.

274. Druzhevskaya AM, Ahmetov, II, Astratenkova IV, Rogozkin VA. Association of the ACTN3 R577X polymorphism with power athlete status in Russians. Eur J Appl Physiol. 2008;103(6):631-4.

275. Eynon N, Banting LK, Ruiz JR, Cieszczyk P, Dyatlov DA, Maciejewska-Karlowska A, et al. ACTN3 R577X polymorphism and team-sport performance: a study involving three European cohorts. J Sci Med Sport. 2014;17(1):102-6.

276. Eynon N, Duarte JA, Oliveira J, Sagiv M, Yamin C, Meckel Y, et al. ACTN3 R577X polymorphism and Israeli top-level athletes. Int J Sports Med. 2009;30(9):695-8.

277. Eynon N, Ruiz JR, Femia P, Pushkarev VP, Cieszczyk P, Maciejewska-Karlowska A, et al. The ACTN3 R577X polymorphism across three groups of elite male European athletes. PLoS One. 2012;7(8):e43132.

278. Fattahi Z, Najmabadi H. Prevalence of ACTN3 (the athlete gene) R577X polymorphism in Iranian population. Iran Red Crescent Med J. 2012;14(10):617-22.

279. Fiuza-Luces C, Ruiz JR, Rodriguez-Romo G, Santiago C, Gomez-Gallego F, Yvert T, et al. Are 'endurance' alleles 'survival' alleles? Insights from the ACTN3 R577X polymorphism. PLoS One. 2011;6(3):e17558.

280. Bouchard C. Exercise genomics--a paradigm shift is needed: a commentary. Br J Sports Med. 2015;49(23):1492-6.

281. Llano-Diez M, Fury W, Okamoto H, Bai Y, Gromada J, Larsson L. RNA-sequencing reveals altered skeletal muscle contraction, E3 ligases, autophagy, apoptosis, and chaperone expression in patients with critical illness myopathy. Skelet Muscle. 2019;9(1):9.

282. Chariker JH, Sharp M, Ohri SS, Gomes C, Brabazon F, Harman KA, et al. RNA-seq data of soleus muscle tissue after spinal cord injury under conditions of inactivity and applied exercise. Data Brief. 2020;28:105056.

283. Halliday MR, Abeydeera D, Lundquist AJ, Petzinger GM, Jakowec MW. Intensive treadmill exercise increases expression of hypoxia-inducible factor 1alpha and its downstream transcript targets: a potential role in neuroplasticity. Neuroreport. 2019;30(9):619-27.

284. Lindholm ME, Rundqvist H. Skeletal muscle hypoxia-inducible factor-1 and exercise. Exp Physiol. 2016;101(1):28-32.

285. Sato S, Basse AL, Schonke M, Chen S, Samad M, Altintas A, et al. Time of Exercise Specifies the Impact on Muscle Metabolic Pathways and Systemic Energy Homeostasis. Cell Metab. 2019;30(1):92-110 e4.

286. Papp AC, Azad AK, Pietrzak M, Williams A, Handelman SK, Igo RP, Jr., et al. AmpliSeq transcriptome analysis of human alveolar and monocyte-derived macrophages over time in response to Mycobacterium tuberculosis infection. PLoS One. 2018;13(5):e0198221.

287. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

288. Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner: ; Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States); 2014. Medium: ED p.

289. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.

290. Andrews S. FastQC: a quality control tool for high throughput data. 0.11.5 ed2010.

291. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics. 2012;28(11):1530-2.

292. Law CW, Alhamdoosh M, Su S, Dong X, Tian L, Smyth GK, et al. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Res. 2016;5.

293. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2015;4:1521.

294. Conesa A, Nueda MJ. maSigPro: Significant Gene Expression Profile Differences in Time Course Gene Expression Data. R package version 15002017.

295. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

296. Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016;12(2):477-9.

297. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7.

298. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.

299. Kucera M, Isserlin R, Arkhangorodsky A, Bader GD. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. F1000Res. 2016;5:1717.

300. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38(Web Server issue):W214-20.

301. Claycomb WC, Lanson NA, Jr. Proto-oncogene expression in proliferating and differentiating cardiac and skeletal muscle. Biochem J. 1987;247(3):701-6.

302. Bessereau JL, Fontaine B, Changeux JP. Denervation of mouse skeletal muscle differentially affects the expression of the jun and fos proto-oncogenes. New Biol. 1990;2(4):375-83.

303. Puntschart A, Wey E, Jostarndt K, Vogt M, Wittwer M, Widmer HR, et al. Expression of fos and jun genes in human skeletal muscle after exercise. Am J Physiol. 1998;274(1):C129-37.

304. Louis E, Raue U, Yang Y, Jemiolo B, Trappe S. Time course of proteolytic, cytokine, and myostatin gene expression after acute exercise in human skeletal muscle. J Appl Physiol (1985). 2007;103(5):1744-51.

305. Allen DL, Hittel DS, McPherron AC. Expression and function of myostatin in obesity, diabetes, and exercise adaptation. Med Sci Sports Exerc. 2011;43(10):1828-35.

306. Apple FS, Hellsten Y, Clarkson PM. Early detection of skeletal muscle injury by assay of creatine kinase MM isoforms in serum after acute exercise. Clin Chem. 1988;34(6):1102-4.

307. Mason SD, Howlett RA, Kim MJ, Olfert IM, Hogan MC, McNulty W, et al. Loss of skeletal muscle HIF-1alpha results in altered exercise endurance. PLoS Biol. 2004;2(10):e288.

308. Mason S, Johnson RS. The role of HIF-1 in hypoxic response in the skeletal muscle. Adv Exp Med Biol. 2007;618:229-44.

309. Mason SD, Rundqvist H, Papandreou I, Duh R, McNulty WJ, Howlett RA, et al. HIF-1alpha in endurance training: suppression of oxidative metabolism. Am J Physiol Regul Integr Comp Physiol. 2007;293(5):R2059-69.

310. Suhr F, Brixius K, Marées Md, Bölck B, Kleinöder H, Achtzehn S, et al. Effects of short-term vibration and hypoxia during high-intensity cycling exercise on circulating levels of angiogenic regulators in humans. Journal of Applied Physiology. 2007;103(2):474-83.

311. Hoffman-Goetz L, Pedersen BK. Exercise and the immune system: a model of the stress response? Immunol Today. 1994;15(8):382-7.

312. Vider J, Lehtmaa J, Kullisaar T, Vihalemm T, Zilmer K, Kairane C, et al. Acute immune response in respect to exercise-induced oxidative stress. Pathophysiology. 2001;7(4):263-70.

313. Aoi W, Naito Y, Takanami Y, Kawai Y, Sakuma K, Ichikawa H, et al. Oxidative stress and delayed-onset muscle damage after exercise. Free Radic Biol Med. 2004;37(4):480-7.

314. Kayani AC, Morton JP, McArdle A. The exercise-induced stress response in skeletal muscle: failure during aging. Appl Physiol Nutr Metab. 2008;33(5):1033-41.

315. Gehlert S, Klinz FJ, Willkomm L, Schiffer T, Suhr F, Bloch W. Intense Resistance Exercise Promotes the Acute and Transient Nuclear Translocation of Small Ubiquitin-Related Modifier (SUMO)-1 in Human Myofibres. Int J Mol Sci. 2016;17(5).

316. Pugh JK, Faulkner SH, Turner MC, Nimmo MA. Satellite cell response to concurrent resistance exercise and high-intensity interval training in sedentary, overweight/obese, middle-aged individuals. European Journal of Applied Physiology. 2018;118(2):225-38.

317. Zhao J, Wang Y, Zhao D, Zhang L, Chen P, Xu X. Integration of metabolomics and proteomics to reveal the metabolic characteristics of high-intensity interval training. Analyst. 2020;145(20):6500-10.

318. Breen EC, Johnson EC, Wagner H, Tseng HM, Sung LA, Wagner PD. Angiogenic growth factor mRNA responses in muscle to a single bout of exercise. J Appl Physiol (1985). 1996;81(1):355-61.

319. Roca J, Gavin TP, Jordan M, Siafakas N, Wagner H, Benoit H, et al. Angiogenic growth factor mRNA responses to passive and contraction-induced hyperperfusion in skeletal muscle. J Appl Physiol (1985). 1998;85(3):1142-9.

320. Bloor CM. Angiogenesis during exercise and training. Angiogenesis. 2005;8(3):263-71.

321. Gavin TP, Drew JL, Kubik CJ, Pofahl WE, Hickner RC. Acute resistance exercise increases skeletal muscle angiogenic growth factor expression. Acta Physiol (Oxf). 2007;191(2):139-46.

322. Giordani L, He GJ, Negroni E, Sakai H, Law JYC, Siu MM, et al. High-Dimensional Single-Cell Cartography Reveals Novel Skeletal Muscle-Resident Cell Populations. Mol Cell. 2019;74(3):609-21 e6.

323. McMillan DW, Maher JL, Jacobs KA, Nash MS, Bilzon JLJ. Physiological responses to moderate intensity continuous and high-intensity interval exercise in persons with paraplegia. Spinal Cord. 2021;59(1):26-33.

324. Miguet M, Fillon A, Khammassi M, Masurier J, Julian V, Pereira B, et al. Appetite, energy intake and food reward responses to an acute High Intensity Interval Exercise in adolescents with obesity. Physiology & Behavior. 2018;195:90-7.

325. Hubal MJ, Chen TC, Thompson PD, Clarkson PM. Inflammatory gene changes associated with the repeated-bout effect. Am J Physiol Regul Integr Comp Physiol. 2008;294(5):R1628-37.

326. Beiter T, Hoene M, Prenzler F, Mooren FC, Steinacker JM, Weigert C, et al. Exercise, skeletal muscle and inflammation: ARE-binding proteins as key regulators in inflammatory and adaptive networks. Exerc Immunol Rev. 2015;21:42-57.

327. Stoecklin G, Stoeckle P, Lu M, Muehlemann O, Moroni C. Cellular mutants define a common mRNA degradation pathway targeting cytokine AU-rich elements. RNA. 2001;7(11):1578-88.

328. Hsieh PN, Sweet DR, Fan L, Jain MK. Aging and the Kruppel-like factors. Trends Cell Mol Biol. 2017;12:1-15.

329. Owens GK, Kumar MS, Wamhoff BR. Molecular regulation of vascular smooth muscle cell differentiation in development and disease. Physiol Rev. 2004;84(3):767-801.

330. Prosdocimo DA, Sabeh MK, Jain MK. Kruppel-like factors in muscle health and disease. Trends Cardiovasc Med. 2015;25(4):278-87.

331. Himeda CL, Ranish JA, Pearson RC, Crossley M, Hauschka SD. KLF3 regulates muscle-specific gene expression and synergizes with serum response factor on KLF binding sites. Mol Cell Biol. 2010;30(14):3430-43.

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

333. Matsumoto A, Mason SR, Flatscher-Bader T, Ward LC, Marsh SA, Wilce PA, et al. Effects of exercise and antioxidant supplementation on endothelial gene expression. Int J Cardiol. 2012;158(1):59-65.

334. Law RH, Zhang Q, McGowan S, Buckle AM, Silverman GA, Wong W, et al. An overview of the serpin superfamily. Genome Biol. 2006;7(5):216.

335. Eftedal I, Jørgensen A, Flatberg A. Prior eccentric exercise modifies diving-induced inflammatory gene expression in rats. 2012;26(1_supplement):1082.10-.10.

336. Izuhara K, Conway SJ, Moore BB, Matsumoto H, Holweg CT, Matthews JG, et al. Roles of Periostin in Respiratory Disorders. Am J Respir Crit Care Med. 2016;193(9):949-56.

337. Izuhara K, Nunomura S, Nanri Y, Ogawa M, Ono J, Mitamura Y, et al. Periostin in inflammation and allergy. Cell Mol Life Sci. 2017;74(23):4293-303.

338. Kudo A. Periostin in fibrillogenesis for tissue regeneration: periostin actions inside and outside the cell. Cell Mol Life Sci. 2011;68(19):3201-7.

339. Oka T, Xu J, Kaiser RA, Melendez J, Hambleton M, Sargent MA, et al. Genetic manipulation of periostin expression reveals a role in cardiac hypertrophy and ventricular remodeling. Circ Res. 2007;101(3):313-21.

340. Özdemir C, Akpulat U, Sharafi P, Yıldız Y, Onbaşılar İ, Kocaefe ÇJG. Periostin is temporally expressed as an extracellular matrix component in skeletal muscle regeneration and differentiation. 2014;553(2):130-9.

341. Porter JD, Merriam AP, Leahy P, Gong B, Khanna S. Dissection of temporal gene expression signatures of affected and spared muscle groups in dystrophin-deficient (mdx) mice. Hum Mol Genet. 2003;12(15):1813-21.

342. Bune AJ, Hayman AR, Evans MJ, Cox TM. Mice lacking tartrate-resistant acid phosphatase (Acp 5) have disordered macrophage inflammatory responses and reduced clearance of the pathogen, Staphylococcus aureus. Immunology. 2001;102(1):103-13.

343. Hayman AR, Macary P, Lehner PJ, Cox TM. Tartrate-resistant acid phosphatase (Acp 5): identification in diverse human tissues and dendritic cells. J Histochem Cytochem. 2001;49(6):675-84.

344. Kemi OJ, Fowler E, McGlynn K, Primrose D, Smirthwaite R, Wilson J. Intensity-dependence of exercise and active recovery in high-intensity interval training. J Sports Med Phys Fitness. 2019;59(12):1937-43.

345. Miyamoto-Mikami E, Tsuji K, Horii N, Hasegawa N, Fujie S, Homma T, et al. Gene expression profile of muscle adaptation to high-intensity intermittent exercise training in young men. Scientific Reports. 2018;8(1):16811.

346. Ramos-Filho D, Chicaybam G, de-Souza-Ferreira E, Guerra Martinez C, Kurtenbach E, Casimiro-Lopes G, et al. High Intensity Interval Training (HIIT) Induces Specific Changes in Respiration and Electron Leakage in the Mitochondria of Different Rat Skeletal Muscles. PLOS ONE. 2015;10(6):e0131766.

347. Gorgoglione R, Porcelli V, Santoro A, Daddabbo L, Vozza A, Monne M, et al. The human uncoupling proteins 5 and 6 (UCP5/SLC25A14 and UCP6/SLC25A30) transport sulfur oxyanions, phosphate and dicarboxylates. Biochim Biophys Acta Bioenerg. 2019;1860(9):724-33.

348. Soriano-Arroquia A, House L, Tregilgas L, Canty-Laird E, Goljanek-Whysall K. The functional consequences of age-related changes in microRNA expression in skeletal muscle. Biogerontology. 2016;17(3):641-54.

349. Tachiiri S, Katagiri T, Tsunoda T, Oya N, Hiraoka M, Nakamura Y. Analysis of gene-expression profiles after gamma irradiation of normal human fibroblasts. Int J Radiat Oncol Biol Phys. 2006;64(1):272-9.

350. Jaldin-Fincati JR, Pavarotti M, Frendo-Cumbo S, Bilan PJ, Klip A. Update on GLUT4 Vesicle Traffic: A Cornerstone of Insulin Action. Trends Endocrinol Metab. 2017;28(8):597-611.

351. Sun Y, Jaldin-Fincati J, Liu Z, Bilan PJ, Klip A. A complex of Rab13 with MICAL-L2 and alpha-actinin-4 is essential for insulin-dependent GLUT4 exocytosis. Mol Biol Cell. 2016;27(1):75-89.

352. Edwards MG, Anderson RM, Yuan M, Kendziorski CM, Weindruch R, Prolla TA. Gene expression profiling of aging reveals activation of a p53-mediated transcriptional program. BMC Genomics. 2007;8:80.

353. Wu C, Xu G, Tsai SA, Freed WJ, Lee CT. Transcriptional profiles of type 2 diabetes in human skeletal muscle reveal insulin resistance, metabolic defects, apoptosis, and molecular signatures of immune activation in response to infections. Biochem Biophys Res Commun. 2017;482(2):282-8.

354. Yuan R, Zhang X, Fang Y, Nie Y, Cai S, Chen Y, et al. mir-127-3p inhibits the proliferation of myocytes by targeting KMT5a. Biochem Biophys Res Commun. 2018;503(2):970-6.

355. Simon P, Fehrenbach E, Niess AM. Regulation of immediate early gene expression by exercise: short cuts for the adaptation of immune function. Exerc Immunol Rev. 2006;12:112-31.

356. Palmer C, Diehn M, Alizadeh AA, Brown PO. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. BMC Genomics. 2006;7(1):115.

357. Małecki J, Jakobsson ME, Ho AY, Moen A, Rustan AC, Falnes PØJJoBC. Uncovering human METTL12 as a mitochondrial methyltransferase that modulates citrate synthase activity through metabolite-sensitive lysine methylation. 2017;292(43):17950-62.

358. Brass EP, Peters MA, Hinchcliff KW, He YD, Ulrich RG. Temporal pattern of skeletal muscle gene expression following endurance exercise in Alaskan sled dogs. J Appl Physiol (1985). 2009;107(2):605-12.

359. Mikeska T, Craig JM. DNA methylation biomarkers: cancer and beyond. Genes (Basel). 2014;5(3):821-64.

360. De Smet C, Lurquin C, Lethe B, Martelange V, Boon T. DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. Mol Cell Biol. 1999;19(11):7327-35.

361. Leenen FA, Muller CP, Turner JD. DNA methylation: conducting the orchestra from exposure to phenotype? Clin Epigenetics. 2016;8:92.

362. Illingworth RS, Bird AP. CpG islands--'a rough guide'. FEBS letters. 2009;583(11):1713-20.

363. Ronn T, Ling C. Effect of exercise on DNA methylation and metabolism in human adipose tissue and skeletal muscle. Epigenomics. 2013;5(6):603-5.

364. Denham J. Exercise and epigenetic inheritance of disease risk. Acta Physiol (Oxf). 2018;222(1):e12881.

365. McGee SL, Hargreaves M. Epigenetics and Exercise. Trends Endocrinol Metab. 2019;30(9):636-45.

366. Denham J, Marques FZ, O'Brien BJ, Charchar FJ. Exercise: putting action into our epigenome. Sports Med. 2014;44(2):189-209.

367. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Research. 2016;45(4):e22-e.

368. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016;17(1):208.

369. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genom Data. 2016;9:22-4.

370. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29(2):189-96.

371. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32(2):286-8.

372. Martorell-Marugan J, Gonzalez-Rumayor V, Carmona-Saez P. mCSEA: detecting subtle differentially methylated regions. Bioinformatics. 2019;35(18):3257-62.

373. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics. 2010;11:587.

374. Lu H, Galeano MCR, Ott E, Kaeslin G, Kausalya PJ, Kramer C, et al. Mutations in DZIP1L, which encodes a ciliary-transition-zone protein, cause autosomal recessive polycystic kidney disease. Nat Genet. 2017;49(7):1025-34.

375. Gudiksen A, Bertholdt L, Stankiewicz T, Tybirk J, Plomgaard P, Bangsbo J, et al. Effects of training status on PDH regulation in human skeletal muscle during exercise. Pflugers Arch. 2017;469(12):1615-30.

376. Yaghoob Nezhad F, Verbrugge SAJ, Schönfelder M, Becker L, Hrabě de Angelis M, Wackerhage H. Genes Whose Gain or Loss-of-Function Increases Endurance Performance in Mice: A Systematic Literature Review. Front Physiol. 2019;10:262.

377. Arvanitis DA, Vafiadaki E, Sanoudou D, Kranias EG. Histidine-rich calcium binding protein: the new regulator of sarcoplasmic reticulum calcium cycling. J Mol Cell Cardiol. 2011;50(1):43-9.

378. Kusić D, Connolly J, Kainulainen H, Semenova EA, Borisov OV, Larin AK, et al. Striated muscle-specific serine/threonine-protein kinase beta segregates with high versus low responsiveness to endurance exercise training. Physiol Genomics. 2020;52(1):35-46.

379. Goldfarb AH, Bruno JF, Buckenmeyer PJ. Intensity and duration of exercise effects on skeletal muscle cAMP, phosphorylase, and glycogen. Journal of Applied Physiology. 1989;66(1):190-4.

380. Berdeaux R, Stewart R. cAMP signaling in skeletal muscle adaptation: hypertrophy, metabolism, and regeneration. American Journal of Physiology-Endocrinology and Metabolism. 2012;303(1):E1-E17.

381. Dos Santos M, Backer S, Saintpierre B, Izac B, Andrieu M, Letourneur F, et al. Single-nucleus RNA-seq and FISH identify coordinated transcriptional activity in mammalian myofibers. Nat Commun. 2020;11(1):5102.

382. Zeng W, Jiang S, Kong X, El-Ali N, Ball AR, Jr., Ma CIH, et al. Single-nucleus RNA-seq of differentiating human myoblasts reveals the extent of fate heterogeneity. Nucleic Acids Research. 2016;44(21):e158-e.

383. Santos MD, Gioftsidi S, Backer S, Machado L, Relaix F, Maire P, et al. Extraction and sequencing of single nuclei from murine skeletal muscles. STAR Protocols. 2021;2(3):100694.

384. Granata C, Jamnick NA, Bishop DJ. Principles of Exercise Prescription, and How They Influence Exercise-Induced Changes of Transcription Factors and Other Regulators of Mitochondrial Biogenesis. Sports Med. 2018;48(7):1541-59.

385. Worheide MA, Krumsiek J, Kastenmuller G, Arnold M. Multi-omics integration in biomedical research - A metabolomics-centric review. Anal Chim Acta. 2021;1141:144-62.

386. Ramos M, Schiffer L, Re A, Azhar R, Basunia A, Rodriguez C, et al. Software for the Integration of Multiomics Experiments in Bioconductor. Cancer Res. 2017;77(21):e39-e42.