Genetics and metabolomics of elite athletes: Genome-wide association study and Metabolomics profiling of elite athletes

Fatima Abdullah AL-Khelaifi
Student Number: 15104948

A thesis submitted to the University of London for the Research degree

of Doctor of Philosophy

July 2020

Division of Medicine

University College London

بسم الله الرحمن الرحيم

اللَّهُمَّ انْفَعْنِي بِمَا عَلَّمْتَنِي، وَعَلِّمْنِي مَا يَنْفَعُنِي، وَزِدْنِي عِلْمًا

Allaahum-ma anfa'ni bima 'al-lamtani, wa 'al-limni ma yanfa'uni, wa zidni 'ilma

‘Oh Allah, benefit me with what you have taught me and teach me what benefits me, and grant me increase in knowledge.’

Declaration

I, Fatima Abdulla AL-Khelaifi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, and contributions by other researchers are acknowledged.

Signature: 

Date: 6 July 2020

Abstract

AIM: The outstanding performance of elite athletes is a product of a complex interaction between genetic and environmental factors. The aims of this study was to compare differences in genetic and metabolic profiles among different classes of elite athletes and to identify genetically-influenced metabolic profiles (metabotypes) underlying these differences.

METHODS: Genome-wide association study (GWAS) was conducted in 1259 elite athlete samples using Drug core BeadChip arrays, followed by non-targeted metabolomics of 692 serum samples. Genotype distribution, differences in metabolic levels and genetically-influenced metabotypes were compared between high and moderate endurance and power sports as well as among sports with different cardiovascular demands (CVD).

RESULTS: Out of 341385 SNPs, two novel associations are reported for endurance status including rs56330321 in *ATP2B2* (p=1.47E-7) and rs2635438 in *SYNE1* (p=2.54E-7). A meta-analysis confirmed the association of rs56330321 and rs2635438 with endurance athlete status at GWAS level of significance. Metabolomics analysis of 740 metabolites was performed in in 191 (discovery cohort) and 500 (replication cohort) elite athletes. These studies revealed changes in various metabolites involved in steroid biosynthesis, fatty acid oxidation, oxidative stress response, xenobiotics and various mediators of cell signaling among different groups of endurance, power and CVD athletes. By combining GWAS with metabolomics profiling data (mGWAS), 19 common

variant metabolic quantitative trait loci (mQTLs) were identified, of which 5 were novel. When focusing on metabolites associated with endurance, power and CVD, 4 common variant mQTLs were found, of which one novel mQTL linking 4-androsten-3alpha,17alpha-diol monosulfate and *SULT2A1* involved in steroid sulfation was identified in association with endurance.

CONCLUSIONS: GWAS, metabolomics and mGWAS of elite athletes identified novel markers associated with elite athletic performance with a potential application in biomarker discovery in relation to elite athletic performance.

Impact Statement

The performance of elite athletes is a multifactorial trait, with input from both environmental (exercise and diet) and genetic factors, the combined effect of which can cause changes in blood metabolic profile. There is ample evidence suggesting genetic and environmental factors influence over several phenotypic traits related to physical performance and training responses. The genetic predisposition is believed to result from minor/small? contributions of multiple genetic variants. However, genetic studies of athletic performance have been mostly underpowered because of small sample size due to the difficulty in obtaining samples from elite athletes. Additionally, most of the previous studies lacked a clear phenotype.

In this study, we carried out a sufficiently powered genome-wide association study (GWAS) to identify common genetic variants (single nucleotide polymorphisms or SNPs) associated with elite athletic performance in distinct athletic groups (moderate vs high endurance, power, and cardiovascular demand). We also compared the blood metabolic profiles among these groups to identify the potential metabolic pathways underlying differences in their performance. Finally, we identified known and novel SNPs likely to influence the levels of metabolites differentiating the athletic groups as well as those differentiating elite athletes from the general population.

GWAS has revealed a number of SNPs associated with endurance, power and cardiovascular demand in elite athletes. Metabolomics has identified a number of metabolites differentiating these groups, including biomarkers of steroid biosynthesis, fatty acid oxidation, glutathione cycle, energy metabolites, and various membrane lipids. When focusing on xenobiotics, specific metabolites exhibited different levels in various

sports groups including those that potentially prolong exercise tolerance, carry a nootropic effect, exert a potent anti-oxidant effect, or originate from drugs used treating different types of injuries. Finally, combining GWAS and metabolomics data revealed known and novel genetically-influenced metabolites (metabotypes) with an association with performance, including those related to steroid biosynthesis and energy storage.

This project provides evidence of genetic predisposition to elite athletic performance that is most evident when considering genetically-determined metabolites with direct relevance to steroid biosynthesis and energy utilization.

The emerging novel data could lead to the development of a panel of SNPs that can influence levels of specific metabolites associated with elite athletic performance, aiming for more informed identification and management of sports talents in Qatar and potentially worldwide.

Acknowledgement

First of all, I would like to express gratitude to Allah for giving me the opportunity, ambition and the strength to pursue my PhD study.

I ‘am thankful to Qatar National Research Fund for funding this project (Grant no. NPRP 7 - 272 - 1 – 041) as well as thanks to Anti-doping Laboratory Qatar (Prof. Mohammed Al-Maadeed, Dr. Vidya Mohamed Ali and Miss. Noor Al-Motawa) and Weill Cornell Medical College-Qatar (Prof. Karsten Suhre) for collaborating with all the required platforms and the facilities that needed to accomplish the targets of this project.

Special thanks to Dr. Mohamed Aghar Elrayess (My Anti-Doping Lab Qatar supervisor and Lead PI of this grant) for believing in me and allowing me to assist him in his projects which have been a life changing experience with full of intellectual challenges. I would also like to thank Prof. David Abraham, and Prof. Aroon Hingorani (my University College London supervisors) for their continuous support. Guidance and encouragement that I received from my supervisors which helped me achieving the aims of this PhD project.

Thanks to Dr. Costas Georgakopoulos, Dr. Francesco Donati, Dr. Francesco Botrè and Mr. Najib Adnan Dbes, for their time and energy in recruiting athletes’ blood and serum samples from their respective laboratories and transferring them safely to my research lab.

I also thank Dr. Mark Kristiansen, Gaganjit Kaur and Yasmin Panchbhaya (UCL genomics team at University College London) for providing iSCAN training and logistic support.

I would like to thank Dr. Noha Abdel Rahman Yousri, Dr. Ilhame Diboun and Dr. Omar Albagha for their contribution and valuable input during data analysis. And a special thanks to Dr. Noha Abdel Rahman Yousri for her time and dedication that she provided for my bioinformatics training.

I would also like to thank Dr. Ahmetov Ildus and his team to discuss their study results on Russian cohort and invaluable data sharing for meta-analysis.

And many thanks to all my colleagues at Anti-doping Laboratory Qatar for their motivation, wishes, and for being the best companions ever.

I am grateful to my mother (Mrs. Meher Unissa) and my brother (Pilot. Khalifa AL-Khelaifi) for being by my side whenever I needed them and for their support; without their support, and approval I would never had explore the world. Correspondingly, I would also like to thank all my family members (Khamis AL-Khelaifi, Rizwana Unissa and Anjum Unissa), cousins and relatives for their prayers and wishes that they always do for my success and bright future.

Last but not the least, I would like to express my gratitude toward the appreciations I received from my previous instructors: Dr. Hamda Al-Naimi, Prof. Samir Mohamed Jaoua,

Dr. S. Muralitharan and all my teachers, as well as my special friend Dr. Asmaa Raees who build my confidence and inspired me to work in the field as research scientist.

I would like to dedicate this thesis to my late grandmother “Abeda Begum” who used to call me doctor as I was studying science. Thank you Nanyami (rest in peace) for giving me the most prosperous name “Dactarni Amma” which has been the stimulant throughout my career encouraged me to work harder and go beyond my comfort zone to make myself deserving for such title.

Publications

1. **Al-Khelaifi F**, Diboun I, Donati F, Botrè F, Alsayrafi M, Georgakopoulos C, Suhre K, Yousri NA, Elrayess MA. A pilot study comparing the metabolic profiles of elite-level athletes from different sporting disciplines. Sports Med Open. 2018 Jan 5;4(1):2.

Abstract

Background: The outstanding performance of an elite athlete might be associated with changes in their blood metabolic profile. The aims of this study were to compare the blood metabolic profiles between moderate- and high-power and endurance elite athletes and to identify the potential metabolic pathways underlying these differences.

Methods: Metabolic profiling of serum samples from 191 elite athletes from different sports disciplines (121 high- and 70 moderate-endurance athletes, including 44 high- and 144 moderate-power athletes), who participated in national or international sports events and tested negative for doping abuse at anti-doping laboratories, was performed using non-targeted metabolomics-based mass spectroscopy combined with ultrahigh-performance liquid chromatography. Multivariate analysis was conducted using orthogonal partial least squares discriminant analysis. Differences in metabolic levels between high- and moderate-power and endurance sports were assessed by univariate linear models.

Results: Out of 743 analyzed metabolites, gamma-glutamyl amino acids were significantly reduced in both high-power and high-endurance athletes compared to moderate counterparts, indicating active glutathione cycle. High-endurance athletes exhibited significant increases in the levels of several sex hormone steroids involved in testosterone and progesterone synthesis, but decreases in diacylglycerols and ecosanoids. High-power athletes had increased levels of phospholipids and xanthine metabolites compared to moderate-power counterparts.

Conclusions: This pilot data provides evidence that high-power and high-endurance athletes exhibit a distinct metabolic profile that reflects steroid biosynthesis, fatty acid metabolism, oxidative stress, and energy-related metabolites. Replication studies are warranted to confirm differences in the metabolic profiles associated with athletes' elite performance in independent data sets, aiming ultimately for deeper understanding of the underlying biochemical processes that could be utilized as biomarkers with potential therapeutic implications.

2. **Al-Khelaifi F**, Diboun I, Donati F, Botrè F, Alsayrafi M, Georgakopoulos C, Suhre K, Yousri, NA, Elrayess MA. Profiling the prevalence of xenobiotics in elite athletes: Relevance to supplements consumption. Journal of the International Society of Sports Nutrition. J Int Soc Sports Nutr. 2018 Sep 27;15(1):48.

Abstract

Background: Supplements are widely used among elite athletes to maintain health and improve performance. Despite multiple studies investigating use of dietary supplements by athletes, a comprehensive profiling of serum supplement metabolites in elite athletes is still lacking. This study aims to analyze the presence of various xenobiotics in serum samples from elite athletes of different sports, focusing on metabolites that potentially originate from nutritional supplements.

Methods: Profiling of xenobiotics in serum samples from 478 elite athletes from different sports (football, athletics, cycling, rugby, swimming, boxing and rowing) was performed using non-targeted metabolomics-based mass spectroscopy combined with ultrahigh-performance liquid chromatography. Multivariate analysis was performed using orthogonal partial least squares discriminant analysis. Differences in metabolic levels among different sport groups were identified by univariate linear models.

Results: Out of the 102 detected xenobiotics, 21 were significantly different among sport groups including metabolites that potentially prolong exercise tolerance (caffeic acid), carry a nootropic effect (2-pyrrolidinone), exert a potent anti-oxidant effect (eugenol, ferulic acid 4 sulfate, thioproline, retinol), or originate from drugs for different types of injuries (ectoine, quinate). Using Gaussian graphical modelling, a metabolic network that links various sport group-associated xenobiotics was constructed to further understand their metabolic pathways.

Conclusions: This pilot data provides evidence that athletes from different sports exhibit a distinct xenobiotic profile that may reflect their drug/supplement use, diet and exposure to various chemicals. Because of limitation in the study design, replication studies are warranted to confirm results in independent data sets, aiming ultimately for better assessment of dietary supplement use by athletes.

3. **Al-Khelaifi F**, Donati F, Botrè F, Latiff A, Abraham D, Hingorani A, Georgakopoulos C, Suhre K, Yousri NA, Elrayess MA. Metabolic profiling of elite athletes with different cardiovascular demand. Scand J Med Sci Sports, 2019. 29(7): p. 933-943.

Abstract:

Intensive exercise of elite athletes can lead to physiological alterations in the cardiovascular system in response to increased stroke volume and blood pressure, known collectively as cardiovascular demand (CD). This study aimed to compare metabolic differences in elite athletes with high vs low/moderate CD and to reveal their underlying metabolic pathways as potential biomarker signatures for assessing health, performance, and recovery of elite athletes. Metabolic profiling of serum samples from 495 elite athletes from different sport disciplines (118 high CD and 377 low/moderate CD athletes) was conducted using non-targeted metabolomics-based mass spectroscopy combined with ultra-high-performance liquid chromatography. Results show that DAGs containing arachidonic were enriched in high CD together with branched-chain amino acids, plasminogens, phosphatidylcholines, and phosphatidylethanolamines, potentially indicating increased risk of cardiovascular disease in the high CD group. Gamma-glutamyl amino acids and glutathione metabolism were increased in low/moderate CD group, suggesting more efficient oxidative stress scavenging mechanisms than the high CD group. This first most comprehensive metabolic profiling of elite athletes provides an evidence that athletes with different CD show a unique metabolic signature that reflects energy generation and oxidative stress and potentially places the high CD group at a higher risk of cardiovascular disease. Further studies are warranted for confirmation and validation of findings in other sport groups in light of potential confounders related to limited available information about participants.

4. **Fatima Al-Khelaifi**, David Abraham, Ilhame Diboun, Mohamed A Elrayess. Book Chapter Twenty-Three - Proteomics and metabolomics research in exercise and sport, in Sports, Exercise, and Nutritional Genomics, D. Barh and I.I. Ahmetov, Editors. 2019, Academic Press. p. 539-566.

Abstract

Proteomics and metabolomics research offers a quantitative measurement of metabolic profiles associated with exercise and identifies metabolic signatures of athletes from different sporting disciplines. Molecular biomarkers include peptides, proteins, amino acids, carbohydrates, lipids, and xenobiotics. Collectively, these metabolites capture the molecular pathways underlying whole body physiology including nutrition, energy generation, oxidative stress scavenging mechanisms, and hormonal balance. This chapter aims to review the recent literature investigating potential proteomic and metabolomic biomarkers associated with exercise, endurance, and power sports and discuss their functional relevance in relation to athletes' performance, training, recovery from injury and overall health.

5. **Fatima Al-Khelaifi**, Ilhame Diboun, Francesco Donati, Francesco Botrè, David Abraham, Aroon Hingorani, Omar Albagha, costas Georgakopoulos, Karsten Suhre, Noha Yousri, Mohamed A. Elrayess. "Metabolic GWAS of elite athletes reveals novel genetically-influenced metabolites associated with athletic performance". Scientific Reports. 2019;9(1):19889.

Abstract:

Genetic research of elite athletic performance has been hindered by the complex phenotype and the relatively small effect size of the identified genetic variants. The aims of this study were to identify genetic predisposition to elite athletic performance by investigating genetically-influenced metabolites that discriminate elite athletes from non-elite athletes and to identify those associated with endurance sports. By conducting a genome wide association study with high-resolution metabolomics profiling in 490 elite athletes, common variant metabolic quantitative trait loci (mQTLs) were identified and compared with previously identified mQTLs in non-elite athletes. Among the identified mQTLs, those associated with endurance metabolites were determined. Two novel genetic loci in FOLH1 and VNN1 are reported in association with N-acetyl-aspartyl-glutamate and Linoleoyl ethanolamide, respectively. When focusing on endurance metabolites, one novel mQTL linking androstenediol (3alpha, 17alpha) monosulfate and SULT2A1 was identified. Potential interactions between the novel identified mQTLs and exercise are highlighted. This is the first report of common variant mQTLs linked to elite athletic performance and endurance sports with potential applications in biomarker discovery in elite athletic candidates, non-conventional anti-doping analytical approaches and therapeutic strategies.

6. Ildus I. Ahmetov, Ekaterina A. Semenova, Eri Miyamoto-Mikami, Egor B. Akimov, **Fatima Al-Khelaifi**, Haruka Murakami, Hirofumi Zempo, Elena S. Kostryukova, Nikolay A. Kulemin, Andrey K. Larin, Oleg V. Borisov, Motohiko Miyachi, Daniil V. Popov, Eugenia A. Boulygina, Mizuki Takaragawa, Hiroshi Kumagai, Hisashi Naito, Vladimir P. Pushkarev, Dmitry A. Dyatlov, Eugene V. Lekontsev, Yuliya E. Pushkareva, Liliya B. Andryushchenko, Mohamed A. Elrayess, Edward V. Generozov, Noriyuki Fuku. "The association of HFE gene H63D polymorphism with endurance athlete status and aerobic capacity: novel findings and a meta-analysis". European journal of applied physiology. 2020.

Abstract

Purpose Iron is an important component of the oxygen-binding proteins and may be critical to optimal athletic performance. Previous studies have suggested that the G allele of C/G rare variant (rs1799945), which causes H63D amino acid replacement, in the *HFE* is associated with elevated iron indexes and may give some advantage in endurance-oriented sports. The aim of the present study was to investigate the association between the *HFE* H63D polymorphism and elite endurance athlete status in Japanese and Russian populations, aerobic capacity and to perform a meta-analysis using current findings and three previous studies.

Methods The study involved 315 international-level endurance athletes (255 Russian and 60 Japanese) and 809 healthy controls (405 Russian and 404 Japanese). Genotyping was performed using micro-array analysis or by PCR. VO_{2max} in 46 male Russian endurance athletes was determined using gas analysis system.

Results The frequency of the iron-increasing CG/GG genotypes was significantly higher in Russian (38.0 vs 24.9%; OR 1.85, $P = 0.0003$) and Japanese (13.3 vs 5.0%; OR 2.95, $P = 0.011$) endurance athletes compared to ethnically matched controls. The meta-analysis using five cohorts (two French, Japanese, Spanish, and Russian; 586 athletes and 1416 controls) showed significant prevalence of the CG/GG genotypes in endurance athletes compared to controls (OR 1.96, 95% CI 1.58–2.45; $P = 1.7 \times 10^{-9}$). Furthermore, the *HFE* G allele was associated with high VO_{2max} in male athletes [CC: 61.8 (6.1), CG/GG: 66.3 (7.8) ml/min/kg; $P = 0.036$].

Conclusions We have shown that the *HFE* H63D polymorphism is strongly associated with elite endurance athlete status, regardless ethnicities and aerobic capacity in Russian athletes.

7. **Fatima Al-Khelaifi**, Noha Yusri, Omar Albagha, Ekaterina A. Semenova, Elena S. Kostryukova, Nickolay A. Kulemin, Oleg V. Borisov, Andrey K. Larin, Edward V. Generozov, Francesco Donati, Francesco Botrè, Costas Georgakopoulos, Karsten Suhre, Ildus I. Ahmetov, Mohamed A Elrayess. "Genome-Wide Association Study Reveals a Novel Association Between MYBPC3 Gene Polymorphism, Endurance Athlete Status, Aerobic Capacity and Steroid Metabolism". Front. Genet. June 2020. 11:595. doi: 10.3389/fgene.2020.00595

Background: The genetic predisposition to elite athletic performance has been a controversial subject due to the underpowered studies and the small effect size of identified genetic variants. The aims of this study were to investigate the association of common single-nucleotide polymorphisms (SNPs) with endurance athlete status in a large cohort of elite European athletes using GWAS approach, followed by replication studies in Russian and Japanese elite athletes and functional validation using metabolomics analysis.

Results: The association of 476,728 SNPs of Illumina DrugCore Gene chip and endurance athlete status was investigated in 796 European international-level athletes (645 males, 151 females) by comparing allelic frequencies between athletes specialized in sports with high (n = 662) and low/moderate (n = 134) aerobic component. Replication of results was performed by comparing the frequencies of the most significant SNPs between 242 and 168 elite Russian high and low/moderate aerobic athletes, respectively, and between 60 elite Japanese endurance athletes and 406 controls. A meta-analysis has identified rs1052373 (GG homozygotes) in Myosin Binding Protein (*MYBPC3*; implicated in cardiac hypertrophic myopathy) gene to be associated with endurance athlete status ($P = 1.43 \times 10^{-8}$, odd ratio 2.2). Homozygotes carriers of rs1052373 G allele in Russian athletes had significantly greater VO_{2max} than carriers of the AA + AG (P = 0.005). Subsequent metabolomics analysis revealed several amino acids and lipids associated with rs1052373 G allele (1.82×10^{-05}) including the testosterone precursor androstenediol (3beta,17beta) disulfate.

Conclusions: This is the first report of genome-wide significant SNP and related metabolites associated with elite athlete status. Further investigations of the functional relevance of the identified SNPs and metabolites in relation to enhanced athletic performance are warranted.

Conference papers:

1. **Fatima Al-Khelaifi**, Francesco Donati, Francesco Botre, Costas Georgakopoulos, Karsten Suhre and Noha A Yousri, Mohamed A Elrayess. Genomics and Metabolomics of elite athletes. Seventh Annual Anti-Doping Lab Junior Symposium, Dec 5, 2019.

2. **Fatima Al-Khelaifi,** Francesco Donati, Francesco Botrè, Mohammed Al Maadheed, Costas Georgakopoulos, Mohamed A Elrayess. Poster "Metabolomics In Anti-Doping Research". Annual Workshop of Antidoping labs, Feb 2019, Cologne, Germany.

3. **Fatima Al-Khelaifi,** Ilhame Diboun, Francesco Donati, Francesco Botrè, Omar Albagha, Costas Georgakopoulos, Karsten Suhre, Noha A Yousri, Mohamed A Elrayess. Poster "Genomics and Metabolomics of elite athletes". Sixth Annual Anti-Doping Lab Junior Symposium, Dec 11, 2018.

4. **Fatima Al-Khelaifi,** Ilhame Diboun, Francesco Donati, Francesco Botre, Costas Georgakopoulos, Karsten Suhre, Noha A Yousri and Mohamed A Elrayess. Poster "A pilot study comparing the genomics and metabolic profiles of elite-level athletes". Qatar Foundation Annual Research Conference (ARC'18), March 19-20, 2018 Qatar.

5. **Fatima Al-Khelaifi**, Francesco Donati, Francesco Botre, Costas Georgakopoulos, Karsten Suhre and Noha A Yousri, Mohamed A Elrayess. Genomics and Metabolomics of elite athletes. Fifith Annual Anti-Doping Lab Junior Symposium, Dec 19, 2017.

6. **Fatima Al-Khelaifi**, Ilhame Diboun, Francesco Donati, Francesco Botre, Costas Georgakopoulos, Karsten Suhre, Noha A Yousri and Mohamed A Elrayess. Poster "Metabolomics of elite athletes reveals distinct signatures associated with power and endurance sports class". 8th International Conference and Exhibition on Metabolomics & Systems Biology May 08-10, 2017 Singapore. ***Received the Best Poster Award*** that has been attributed in recognition of research paper quality, novelty and significance. DOI: 10.4172/2153-0769-C1-037.

7. **Fatima Al-Khelaifi,** Francesco Donati, Francesco Botre, Costas Georgakopoulos, Karsten Suhre and Noha A Yousri, Mohamed A Elrayess. Genetic and metabolic profiling of elite athletes. Fourth Annual Anti-Doping Lab Junior Symposium, Dec 5, 2016.

8. **Fatima Al-Khelaifi**, Francesco Donati, Francesco Botre, Costas Georgakopoulos, Karsten Suhre and Noha A Yousri, Mohamed A Elrayess. Genetically influenced metabotypes associated with elite athletic performance. Third Annual Anti-Doping Lab Junior Symposium, Dec 7, 2015.

Table of Content

List of Abbreviations

(°C)	Degree Celsius
(µL)	Microliter
(12,13-DHOME)	12,13-Dihydroxyoctadec-9-Enoic Acid
(ACE)	Angiotensin Converting Enzyme Gene
(ACSL1)	Acyl Coenzyme A Synthetase Long-Chain 1 Gene
(ACTN3)	A-Actinin-3
(ADLQ)	Anti-Doping Laboratories In Qatar
(ADME)	Drug Absorption, Distribution, Metabolism And Excretion
(Ala)	Alanine
(ATP)	Adenosine Triphosphate
(Cat#)	Catalogue Number
(Cit)	Citric Acid
(CKM)	Muscle Creatine Kinase
(Cr)	Creatine
(CrP)	Creatine Phosphate
(Cys)	Cysteine
(DAGs)	Diacylglycerols
(DNA)	Deoxyribonucleic Acid
(EDTA)	Ethylenediaminetetraacetic Acid
(ESI)	Electrospray Ionization
(F)	Females
(FA)	Formic Acid
(FA-carnitine)	Fatty Acid Metabolism(Acyl Carnitine)
(FDR)	False Discovery Rate
(FMSI)	Federazione Medico Sportivo Italiana
(GABA)	gamma-Aminobutyric acid
(GC/TOF-MS)	Gas Chromatography-Time Of Fly/Mass Spectrum
(Gln)	Glutamine
(Glu)	Glutamic Acid
(GWAS)	Genome Wide Association Study

(HESI-II)	Heated Electrospray Ionization
(HIF1A)	Hypoxia-Inducible Factor-1alpha
(HILIC)	Hydrophilic-Interaction Chromatography
(hr)	Hours
(HTS)	High-Throughput Screening (HTS) Array
(HWE)	Hardy-Weinberg Equilibrium
(I/D)	Insertion/Deletion
(IL-6)	Interleukin 6
(IPA)	Ingenuity Pathway Analysis
(KEGG)	Kyoto Encyclopedia Of Genes And Genomes
(Lac)	Lactate
(LD)	Linkage Disequilibrium
(LIMS)	Laboratory Information Management System
(Lin)	Linoleic Acid
(M)	Males
(MAF)	Minor Allele Frequency
(Me-G)	B-D-Methylglucopyranoside
(mGWAS)	Genetically-Influenced Metabotype
(min)	Minutes
(ml)	Milliliter
(mol/L)	Moles Per Liter
(MSTN)	Myostatin
(MVC%)	Maximal Voluntary Contraction Percentage
(NFIA-AS2)	Nuclear Factor I A Antisense RNA 2
(ng)	Nanogram
(nm)	Nanometre
(OD)	Optical Density
(Ole)	Oleic Acid
(OPLS-DA)	Orthogonal Partial Least Square Discriminant Analysis
(Pal)	Palmitic Acid
(PC)	PCA Components

(PCA)	Principal Component Analysis
(PFPA)	Perfluoropentanoic Acid
(pH)	Potential Of Hydrogen
(PPARs)	Peroxisome Proliferator-Activated Receptors
(ppm)	Parts-Per-Million
(p-value)	Probability Value
(Pyr)	Pyroglutamic Acid
(QC)	Quality Control
(RBC)	Red Blood Cell
(RBFOX1)	RNA Binding Protein, Fox-1 Homolog 1
(RI)	Retention Time/Index
(RP)/UPLC-MS/MS	Reverse Phase /Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy
(rpm)	Rounds Per Minute
(RSD)	Relative Standard Deviation
(SNPs)	Single Nucleotide Polymorphisms
(T2DM)	Type 2 Diabetes Mellitus
(TCA) cycle	Tricarboxylic Acid or Krebs Cycle
(TSHR)	Thyroid Stimulating Hormone Receptor
(UPLC)	Ultra-Performance Liquid Chromatography
(UPLC-MS/MS)	Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy
(Val)	Valine
(VO_2max%)	Maximal Oxygen Uptake Percentage
(xg)	Times Gravity

List of Tables:

Table of Figures:

CHAPTER 1
INTRODUCTION

1 (Chapter 1) Introduction:

A historical debate on whether nature or nurture plays a more important role in mammalian development has been going on for centuries. Many researchers proposed that environmental factors such as physical activity, nutrition, experience, education, social relationships and culture play a more significant role in shaping individuals, while others argued that key determinants of features are carried in their DNA [1]. In his “origin of species” book, Darwin denied direct role of environment in the modification of heredity (genes), stressing on the role of natural selection as the driver of evolution [2, 3]. However, it has now become evident that phenotypes in general are determined by complex interactions between genes and the surrounding environmental factors that collectively account for the final phenotype [4]. Gene-environment interplay is comprehensive and complex. The superior performance of elite athletes, for instance, has been historically considered an outcome of a special talent shaped by intensive training. The talent is now believed to be a product of additive genetic components predisposing the athlete to endurance/power trainability under the control of strong environmental cues including exercise and nutrition. In this model, the inherited capability together with ability to respond to training are the keys to success in sport [5]. For example, initial selection for participation in organized competitive sports begins at the age of 6-7 years for children who show certain sports talents (genetic predisposition). As children grow from adolescence to young adulthood following intense exercise in the field of their interest, they benefit from regular training that brings changes in their body shape and results in increases in muscle strength and muscle power, predisposing them to becoming young elite athletes [6]. Hence, athletes may reach the elite-level only when

they are born with innate ability [7] as activation and expression of the innate ability (dormant genes) in the athletes in response to exercise training (environmental factor) constitute the fundamental factors for achieving elite-level of performance [8]. This fact can be used in describing the impact of nature (genes), nurture (environment) and their complex interaction in the context of physical performance of elite athletes.

1.1 Elite athletes: Definition and classification

1.1.1 Definition of elite athletes

The consistency and precision in performing remarkable skills with increased force, joint stability, and minimizing the risk of musculoskeletal injuries are essential characteristics in defining elite athletes and differentiating them from less or non-elite individuals [6, 9]. Some studies considered elite athletes as professional athletes who have accumulated deliberate practice for two to ten years [10, 11]. Other studies indicated elite athletes as professional performers who compete in national or international sport events [12-14] . Accordingly, consented elite athlete's samples (males and females) who participated in national or international sports events and tested negative for doping substances were used in this study.

1.1.2 Classification of elite athletes

Athletes have been classified into two broad categories according to the type and intensity of their training: dynamic (isotonic) and static (isometric) [15-17] (Figure 1). The dynamic component reflects alterations in the muscle length as a result of regular contractions generating a limited, but long-lasting, intramuscular power. Those alterations are typical of endurance for sports like long distance running, cycling or triathlons. On the other hand, static exercise causes a higher, but short-lasting, intramuscular power with limited

alterations in length of the muscle. This type of exercise is typical in power sports like weightlifting, jumping, sprinting and throwing. However, most of sports includes both static and dynamic components at varying degrees. For example, Marathon running, which has limited static but have high dynamic components, whereas water skiing has a greater static but have lower dynamic component. Rowing, on the other hand, requires both intense static and dynamic components [16-18].

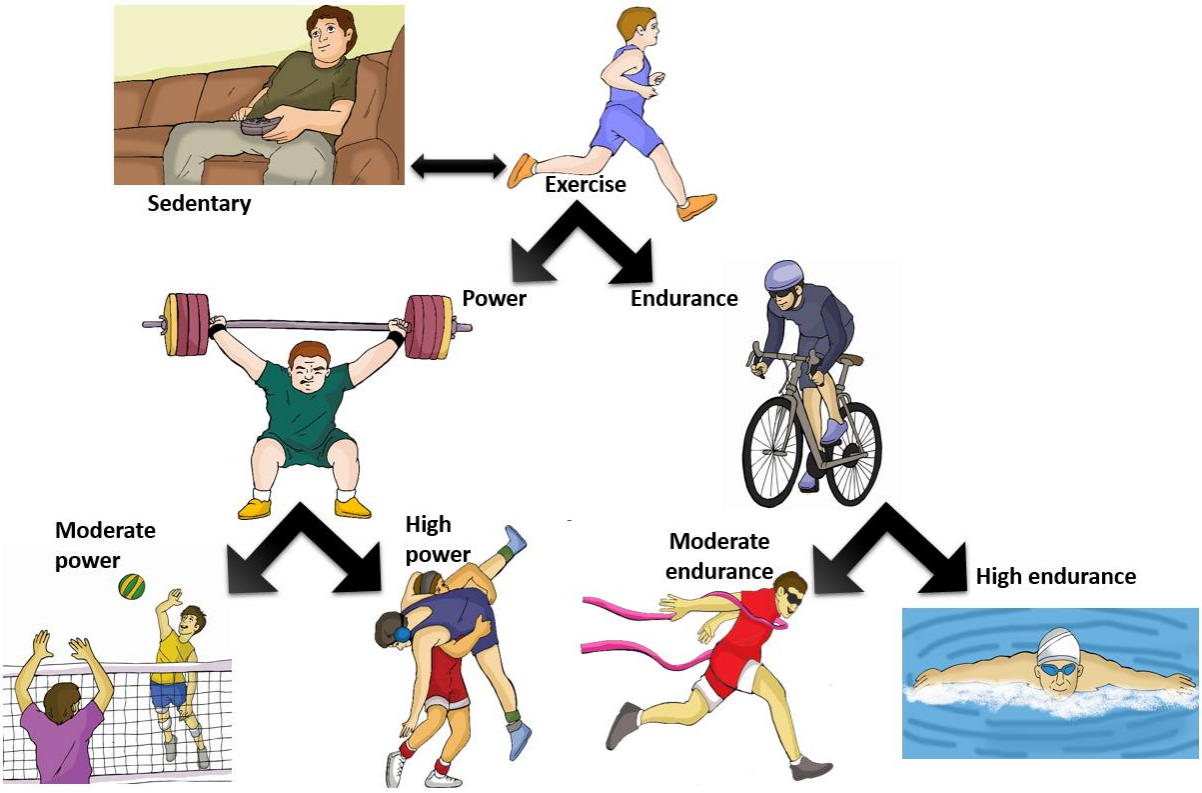


Figure 1. Classification of elite athletes into power sports (moderate and high) and endurance sports (moderate and high) [19].

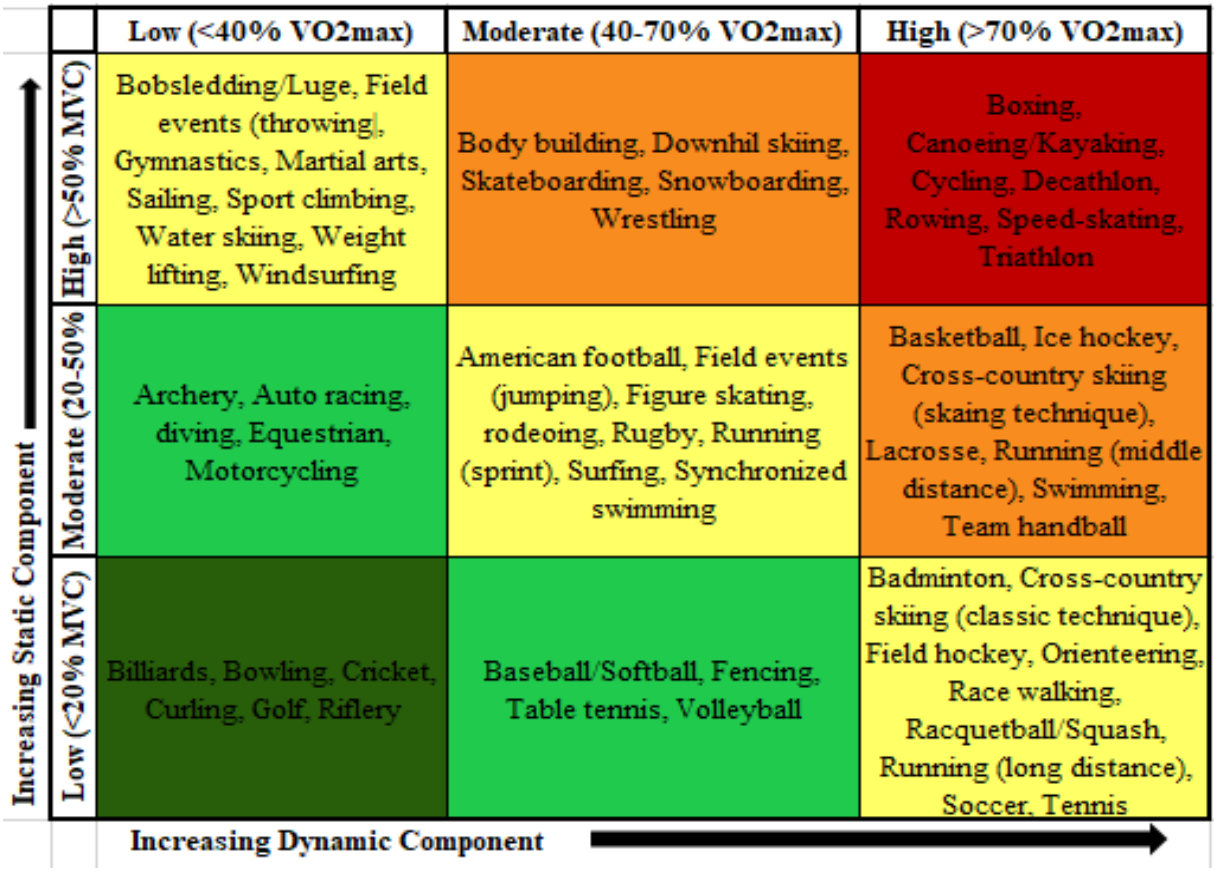
Muscle contraction can be high and prolonged to match the metabolic needs of the exercising body while regulating the blood pressure to ensure adequate perfusion pressure to all organs [20]. Dynamic exercise can be additionally classified depending

on the percentage of maximal oxygen uptake (VO_2 max%) attained with maximum cardiac output. Static exercise can also be classified depending on percentage of maximal voluntary contraction (MVC%) attained with elevated blood pressure (Table.1). Very intense dynamic exercise (such as long-distance running) as well as constant static exercise (such as gymnastics) may be grouped in one high-intensity exercise group based on their similar cardiovascular demand (CVD) that includes high cardiac output and blood pressure. Each sport can be further classified into low, medium, or high intensity dynamic or static exercise, with subcategories including (high static, high dynamic), (moderate static, moderate dynamic), (low static, low dynamic) and so on [16].

Increased aerobic exercise causes greater cardiac output whereas the increased anaerobic exercise results in higher load of blood pressure. Consequently, total CVD includes cardiac output and blood pressure combined and is utilized to divide sports into five additional classes ranging from low static/dynamic (examples are golf and riflery) to high static/dynamic (examples are boxing and cycling) [16]. Several studies examining the association between CVD and physical activity have indicated that exercise plays a role as antidepressant as well as protection from cardiovascular disease [21-23]. A recent study by Hamer and his team investigated the association between physical activity and sub-types of cardiovascular disease. The participants were categorized into four groups: Inactive (not exercising); insufficient activity (performing some moderate - high physical activity), sufficient activity (performing physical activity with intensities: moderate (150 min/week) or high (75 min/week); and high activity group (performing 300 min/week high intensity physical activity). That information was obtained from the participants by interview and assessed by a validated questionnaire. Their data showed an opposite

association between physical activity and type of cardiovascular disease called haemorrhage stroke, suggesting preventions at moderate and risk can be increased with high intensities; in contrast high intensity physical activity was not associated with increased risk of other type of cardiovascular disease called ischemic stroke [21]. However, they concluded that physical activity regardless to its intensities prevent cardiovascular disease as well as there is strong association between aerobic exercise and protection from cardiovascular disease, but they also mentioned the fact that they had a very small proportion of the cohort and lacked complete data on their detailed socioeconomic, sedentary behavior, and diet that might have influenced their results [21].

Table 1. Classification of sports based on peak dynamic (VO_2max) and static (MVC) components achieved during competition, adopted from previously published work [16].



Dynamic exercise causes a significant increase in cardiac output and systolic blood pressure but not diastolic or mean arterial blood pressure. Static exercise, on the other hand, triggers lower cardiac output, but significant elevation in systolic, diastolic, and mean arterial pressure. Therefore, dynamic exercise principally triggers a volume load on the left ventricle, whereas static exercise triggers a pressure load [16]. In Table1, the highest total CVD (cardiac output & blood pressure) are shown in red whereas the lowest are shown in green. Light green, yellow, and orange reflect low moderate, moderate, and high moderate total CVD [16].

Dynamic and static classes of exercise are complementary to another categorization based on muscle metabolism, known as aerobic and anaerobic exercise. The latter categorization describes the source of the required energy for muscle metabolism, which represents a vital element for successful performance. During dynamic exercise, energy is largely obtained through oxidative phosphorylation that requires oxygen for the synthesis of ATP via stimulation of components of the mitochondrial respiratory chain [24]. Under static conditions, the energy for muscle contraction is mostly obtained without using oxygen through anaerobic glycolysis where ATP is produced by substrate-level phosphorylation reactions [25]. Since carbohydrates are the prime energy source during exercise, their limited supplies must be preserved by using fat as a supplementary source of energy. The efficiency of the muscles of elite endurance athletes in utilizing fat as source of energy is greater than that of the power athletes. Such ability can be further enhanced by endurance training [5].

1.2 Factors influencing athletic performance

As introduced, elite athletic performance is a multifactorial trait that is influenced by multiple intrinsic (genetic, physical, physiological and psychological), extrinsic factors (training, nutrition, socioeconomical and general health conditions) and various levels of interactions among these factors [26]. The genetic predisposition was suggested by early hereditary studies, case control studies and more recently genome wide association studies (GWAS), whereas the environmental influence is believed to include exercise, diet and other environmental factors such as socioeconomical status, air quality and emotional status as described in more details.

1.2.1 Environmental factors influencing athletic performance:

Human athletic ability is largely influenced by environmental factors and can be learned and earned through extended and intense practice that interacts at variable degrees with genetic predisposition. Altitude training, for example, is known to enhance sport performance [27]. Indeed, Kenyan runners, who have won most distance running events in the past two decades, are a vibrant example of gene-environment interactions with their yet to be determined unique genetic predisposition in addition to their strict running exercise at high altitude since childhood [28, 29]. Therefore, the genetic makeup of living cells influences the responses to environmental conditions by whole-body functional organs, whereas the expression of genetic variation on a trait depends on the environmental context [30]. Exercise evokes several signaling pathways that strongly modify skeletal muscle contractile properties and myofiber metabolism [31]. These metabolic changes will subsequently alter glucose uptake from the bloodstream to meet the requirement of huge amount of energy demand during exercise [32]. It has been proven that genetic variance plays an important role in determining blood glucose levels and genetic factors may explain half of the variance in impaired fasting glucose and fasting plasma glucose [33].

Other essential environmental issues influencing athletic performance include weather condition, temperature, and air pollution that could potentially impact running capacity and adaptation to sprint conditions [34, 35].

The socio-economic status of the athletes is another example of environmental factors influencing athletic training and performance [36]. Athletes of developed countries undergo strict diet and training programs supervised by professionals who are expert in maximizing the outcome of the genetic predisposition. Elite athletes from developing

countries, on the other hand, may have less opportunity for professional supervision because of limited resources and since sports development does not occupy a top priority in their national budgets [37, 38].

The athlete's emotional status can also affect physical performance substantially by modulating the sympathetic drive resulting in changing energy burst and cardiac output that can further control perceived exertion, pacing and performance [39]. Furthermore, the emotional profiles of elite athletes were shown to predict their potential risk of injury [40]. It has been postulated that training poses stress on an athlete, shifting their physical and psychological well-being along a range of performing better and scoring higher, leading to acute fatigue, overreaching, and ultimately overtraining syndrome [41]. On the other hand, studies have reported that athletes with greater sense of mindfulness were more likely to exhibit better scores in emotion control, setting goals and positive satisfaction, together reflecting positively on their performance [42]. Measurement of athletes' status of emotion and self-satisfaction as well as identification of other environmental factors affecting their own perception have revealed the heterogeneity of motivating factors among athletes. Studies identified that personalized approaches are needed in every athlete's case for encouraging and coordinating by the support system and leadership [43].

1.2.1.1 Exercise

Several studies found that there is a positive effect and a linear relationship between physical performance and exercise protocols [44]. For example, physical inactivity, sedentary behavior and low cardiorespiratory fitness are strong risk factors for the development of chronic diseases, morbidity with resulting mortality [45, 46]. Whereas,

healthcare professionals are promoting physical activity and encouraging adolescents and young adults to exercise with new frontiers broached in elite sport that are increasingly being translated to benefit patients, and vice versa [45, 47]. Exercise improves skeletal muscle energy status and causes higher muscle force outputs for longer periods of time [17]. Overall effect of exercise on performance in terms of intensity and duration of training has been investigated. Continuous training protocols are performed at moderate or high intensities in a single bout [48, 49]. High-intensity interval training protocol is repeated bouts of high-intensity exercise with transition from maximal lactate steady state to complete rest interspersed with recovery periods of low-intensity exercise [17]. Studies tested continuous training versus high-intensity interval training protocol. Previous studies found that both are equally effective at improving exercise performance and reducing the risk of cardiovascular disease [50, 51]. It was shown that high-intensity interval training increases VO_2max and maximal exercise capacity, whereas continuous training is more effective at increasing muscle oxidative capacity [52]. Another study showed that exercise at high intensity, with shorter duration, caused a greater activation of significant regulatory pathways controlling skeletal muscle gene expression when compared to exercise at low intensity with a longer duration [53]. Another study showed that long high intensity interval training generated higher acute increases in rate perceived exertion (RPE: measurement of global perceived effort intensity), heart rate (HR) and maximal oxygen consumption than short high intensity interval training and moderate intensity continuous training, suggesting increase in cardiovascular demand on the cardiorespiratory system [48]. The study also showed

similarity between short high intensity interval training and moderate intensity continuous training in physiological and perceptual responses [48].

A recent study tested duration of recovery periods separating serial bouts of small sided football games on RPE, HR, deoxygenated hemoglobin (HHb), and time motion descriptors (speed and distance) during the sessions [54]. The study concluded that the tested recovery duration ranging from 30s to 120s between serial bouts had a similar effect on trained footballers as the experienced players. The study suggested the coaches to consider recovery duration in the training plan to provide an efficient training session that maintains performance levels [54].

1.2.1.2 Nutrition

Sufficient nutritional intake is fundamental for achieving optimal athletic performance [55], protecting athletes from health issues related to nutrients deficiency and enhancing post-exercise recovery. For example; carbohydrate intake during exercise maintains high levels of carbohydrate oxidation and prevents hypoglycemia [56]. Similarly, a balanced diet of iron is essential for endurance athletes as they have a higher tendency to develop both iron depletion and deficiency, which sooner or later can cause anemia and lead to reduction in training capabilities [57]. Dietary supplements are widely used by athletes in different sport disciplines as part of their regular training or competition routine to fulfill the requirement of having a balanced nutritional diet [58]. Other potential benefits of supplements include the ability to sustain a higher training load without suffering illness or injury, enhance responsiveness to the training stimulus and improve competition performance [59, 60]. Supplements contain essential fatty acids, amino acids, vitamins, minerals, and various "ergogenic" compounds, consumed in the form of protein shakes,

sports drinks, and fortified foods [58, 60, 61]. The regulation of supplements is different in different countries as there is no global market agreement on categorization of these products. For example, some known “dietary” supplements are defined as natural health products and marketed as a food product. Other products are treated as a therapeutic goods, prescription medicine, or controlled substances [61].

Several studies investigated supplement use by elite athletes based on interviews and surveys to determine the influence of age and gender on consuming these supplements and the opinions regarding nutritional supplement education as well as awareness of anti-doping regulations. A study based on questionnaires filled by Canadian high-performance athletes (314 males, 268 females with a mean age of 19.96 +/- 3.91) and reviewed by health professionals (dietitians, sport medicine physician, pharmacist) revealed no age or gender difference in consuming supplements. The study found that 52.7% of family/friends, 44.3% of teammates, and 40.7% of coaches were major source of information for athletes consuming the supplements without approaching an expert in nutrition. Furthermore, 83.5% of athletes indicated that health maintenance and nutritional deficiency were their prime reason for consuming supplements [62]. Another investigation on nutritional supplements performed on 1620 Norwegian elite athletes showed increased use of nutritional supplements in male elite athletes. Their coaches were the main source of information for used supplements [63]. Information acquired from doping control forms of 361 Danish elite athletes reported 100% use of at least one nutritional supplement [64]. A survey conducted on use of dietary supplements in 164 elite German athletes indicated 80% use of one or more supplements such as multi-vitamins, minerals and energy drinks

[65]. Among the supplements used by athletes, carbohydrates ranked top in the list of consumed supplements. However, only some indicated use of amino acids, creatine, or other ergogenic aids such as "blood doping" (administration of packed red blood cells) and pharmacologic agents, such as androgenic steroid supplements [65, 66]. However, some studies have revealed that most athletes were not well informed regarding the benefits and risks of supplements, due to the limited available scientific evidence [58].

A study on population health and lifestyle was performed using 15465 participants (45.2 % were men and the overall mean age was 49.8 years) through collecting data using self-completion postal questionnaire. The study showed that 35.5 % of the participants were taking dietary supplements, but participants had no knowledge of the potential risks associated with these supplements [67]. Another study reported that most of the athletes considered club coaches as the best source of information on supplements and among them only 23% were aware of the side effects of the supplements, mainly via searches on the internet [68]. Consistently, most of the studies revealed significant influence of coaches' beliefs and actions regarding performance-enhancing drugs as well as their role as anti-doping agents, on athletes' supplement consumption [69].

Since elite-level athletes are often reluctant to participate in supplement testing trials because of heavy training load and busy competition timetables, a true assessment on the effectiveness of supplements still to be investigated. Hence, a retrospective approach that investigates the existence of certain supplements/metabolites in elite athletes could offer a deeper view of potentially consumed supplements and their anticipated effects on general health and performance.

1.2.1.3 Other environmental factors

Elite athletes' performance is also affected by other environmental factors including air pollution, weather conditions and socioeconomical factors. A research into the effects of long-term exposure to air pollution on athletic performance described that pollutants like nitrogen and sulphur oxides, as well as ammonium ion, organic aerosols, particulate matter and ozone are all of concern [70]. Particularly during outdoor sports trainings and competition events, athletes are receiving higher exposure via inhalation of pollutants because of the increased ventilation during exercise, which may lead to decreased lung function, increased exacerbations of asthma, decreased diffusion capacity, pulmonary hypertension, cardiovascular effects and an overall decrease performance [70]. A number of measures can be taken into account to reduce exposure to pollutants. Athletes living and practicing in urban cities close to major roadways should consider levels of particulate matter and ozone prior to start exercise, and need to make lifestyle alterations to avoid the deleterious effects of pollutants inhalation [71].

Outdoor sports participants may also be affected by the environmental thermal stress. They might be exposed to warm/hot ambient conditions (temperature >39 °C) leading to exertional hyperthermia, or they may get hypothermia, while practicing in cold/cool-windy environments (temperature <35 °C). Thus, the chances of athletes facing health risks increases which in turns lowers their physical activities and performance [72]. A study showed that heat exposure increases muscle temperature that may benefit initial-sprint bouts performance via an improved muscle contractility [73]. However, overheat exposure leads those muscle-temperature-related benefits to be overridden by the intensified

cardiovascular and metabolic strain and decreasing voluntary muscle activation associated with greater performance decrements [73].

Studies also considered social and economic features of players as one of the environmental factors that distinguish and promote elite athletes. For instance, a study conducted on the Greek women's national volleyball team provided useful information concerning the impact of profession, financial status, relatives and social environment on the performance of female athletes with high potential [74]. The study showed that most of the female athletes were second-born children in their families with support to participate in athletic activities [74]. Their relatives and social environment had an emotional connection with sports and showed moral support towards their participation in sports [74]. Greek women volleyball athletes were financially supported by the sports ministry as their sole occupation [74]. Another study reported that athletes were more influenced by their personal family, teammates, and lack of support from relatives, than coaches who were more concerned about technical and institutional aspects and medical support [75]. The study also reported financial support as the least important factor in terms of athlete's path to sports success [75]. The most influential factor according to both athletes and coaches was dedication, whereas the factor that adversely affected performance was exposure to injuries [75].

1.3 Genetic predisposition to elite athletic performance:

The Human Genome Project has completed over a decade and half ago [76]. Successively, International HapMap Consortium, 1000 genomes and the encode [77] projects indicated that over 99% of human DNA sequences is similar. The 1% variation in human genome at several sites in an individual's DNA and their unique combinations

are likely to have the potential to predispose an individual towards elite sporting ability [78]. The identification of these variants is crucial to understand the superior performance of elite athletes and has been a subject of study for many years [79, 80].

1.3.1 Heritability of super-physical performance

Family-based studies have indicated that the propensity to engage in physical activity, particularly in the form of organized athletics, may be rooted to some extent in our genes. Heritability estimates the amount of variation in a trait that can be accounted for by variation in genes. Twin heritability studies provided the first clues of genetic predisposition to physical performance-related phenotypes. A study of 37051 twins from 7 European countries revealed 48-71 % heritability of participating in leisure-time exercise depending on the country, suggesting that genetics play a significant role in athletic performance independent of environmental effects [81]. The first genome-wide linkage scan for athletic status was performed in 700 British female dizygotic twin pairs and reported a heritability of athletic status of 66% (explained by additive genetic factors), with the remaining variance explained by environmental factors [82]. Data from the HERITAGE family study of 473 sedentary Caucasian adults from 99 families suggested that the heritability of changes in VO_2max with 20 weeks exercise training was ~47%. This GWAS revealed a set of 21 single nucleotide polymorphisms (SNPs), out of 300000 studied, that accounted for the observed variance in VO_2max trainability [83]. Among these, rs6552828 SNP in acyl coenzyme A synthetase long-chain 1 gene (*ACSL1*) exhibited the most significant association with VO_2max, accounting for 6% of the training response of VO_2max. This study, despite the relatively small sample size, has highlighted that performance is genetically influenced and that a large number of SNPs must be

studied to reveal genotype–phenotype interaction instead of single candidate genes. The HERITAGE family study group applied a molecular classifier that predicts the training responsiveness of VO_2max in non-athletes [84].

In addition to VO_2max, multiple studies have reported that muscle strength and mass are influenced by genetic factors. Twin and family studies estimated heritability for muscle strength and muscle mass between 31-78%, with significant alterations between muscle groups and lengths in addition to contraction velocities [85]. Studies focusing on the heritability of explosive power, an essential predictor of sprinting ability, have revealed 74% heritability of maximal power (5 seconds interval) and 84% heritability of total power (30 seconds interval) in 32 Caucasian male twin pairs sharing the same environmental backgrounds [86]. Therefore, it is becoming widely accepted that physical performance is indeed heritable.

1.3.2 Case-control studies

Multiple case-control studies have revealed a number of genetic variants associated with elite athletic performance. Early work identified a significant association of an insertion/deletion (I/D) polymorphism in the angiotensin converting enzyme gene (*ACE*) and endurance trainability, where the insertion genotypes (I and II) were predominant in skilled British high-altitude climbers compared to healthy non-athletic counterparts [87]. This work was followed by the GENEATHLETE study that presented a robust phenotype for elite athletic status by using world-class group of Caucasian athletes VO_2max of at least 75 ml per kg per min. This study identified association of alpha-actinin-3 (*ACTN3*) R577X as well as the Pro582Ser polymorphism of the hypoxia-inducible factor-1alpha

(*HIF1A*) gene with elite endurance athletes, but not variants in myostatin (*MSTN*) or muscle creatine kinase (*CKM*) [88-90]. Subsequent case-control studies have identified over 200 SNPs shown to be associated with physical performance, among which at least 20 SNPs were suggested to influence elite-athletic phenotype [91]. Among these, *ACTN3* remains the only gene that shows a significant and consistent association across multiple studies of elite-athletic power. The expression of the skeletal muscle protein *ACTN3* is limited to fast twitch (type II) muscle fibers as it stabilizes muscle contraction, hence provides a higher force ability than type I fibers [92]. *ACTN3* R577X polymorphism (prevalence of 10% in Africans and up to 50% in Caucasians) provides the most consistent association with reduced sprint capacity, and possible improved endurance performance in humans with the *ACTN3* XX genotype [92, 93] with more than a billion people worldwide cannot express alpha-actinin-3 in their skeletal muscle fibers (homozygous for the R577X null-allele) [94]. The functional relevance of this SNP is clearly confirmed in *ACTN3* knockout mouse model. In this model, (*ACTN3* deficient) has a reduced anaerobic glycolytic activity and increased aerobic oxidative activity [95, 96]. ACTN3 knockout mice also show greater resistance to fatigue, smaller muscle mass and reduced diameter of fast (IIB) twitch muscle fibers and strength than wild-type mice [94, 97]. The phenotypes of the *ACTN3* knockout mouse mimic the gene association studies performed in humans and provide a plausible explanation for the phenotype seen in *ACTN3* XX individuals [93]. However, the *ACTN3* R577X polymorphism was shown an account of only 2.5% of the variance seen in athletic performance [98]. Other polymorphisms associated with elite-athletic power less consistently including *ACE* (rs4646994), angiotensinogen (rs699), skeletal adenosine monophosphate deaminase

(rs17602729), interleukin-6 (rs1800795), endothelial nitric oxide synthase 3 (rs2070744), peroxisome proliferator activated receptor-alpha (rs1799983) and mitochondrial uncoupling protein 2 (rs660339) [99]. Ethnic differences were also evident as some studies associating *ACE* and *ACTN3* with endurance and sprint performance in Caucasians were not replicated in Africans, possibly due to small sample sizes, using inappropriate controls or a mere reflection of a true genotype-ethnicity interaction [100].

Another study performed by Dr. Jamshidi and her team focused on inherited cardiac function defects associated with fatty acid oxidation in response to exercise, hypertension, or obesity [101]. The study analyzed 144 young British Army males and 1148 men and women who participated in the echocardiographic sub-study of the Third Monitoring Trends and Determinants in Cardiovascular Disease Augsburg study (MONICA) [101]. The results indicated that G/C polymorphism in *PPARα* gene was significantly associated with left ventricular growth in response to exercise (P=0.009) [101]. Equally, the C allele homozygotes had significantly higher left ventricular mass, which was greater still in hypertensive subjects, and a higher prevalence of left ventricular hypertrophy (LVH) in the Third MONICA Augsburg study [101]. The study concluded that peroxisome proliferator–activated receptor α (*PPARα*) regulates genes responsible for myocardial fatty acid oxidation and is downregulated during cardiac hypertrophy, associated with glucose utilization instead of fatty acid [101].

Such preliminary studies have played an important role to identify the heritability of performance, however, a candidate gene approach of athletic performance has been hindered by a small sample size and poorly defined phenotypes used by various studies that involved multiple sports disciplines without a clear categorization. This included

measuring heritability of physical activity, a vague phenotype in identical twins and underpowered case-control studies, comparing limited number of genetic variants in athletes versus non-athletes. One issue that these studies have in common is their assumption that, controls have little or no genetic contribution to athleticism. However, the fact is that they contain a number of genetically predisposed individuals who are simply inactive and therefore not meeting their athletic potential [102].

1.3.3 Genome-Wide Association Studies

Decades of research in sport genomics have been shaped by the development in molecular biology techniques and the enhanced statistical and bioinformatics analyses and their ability to identify genetic markers linked to physical performance. Among them, GWAS have been proven a promising approach that can rapidly scan numerous genetic markers across complete sets of DNA in a large sample size to find genetic association with a physical phenotype.

For instance, one GWAS evaluated four phenotypes related to electrocardiogram (ECG): heart rate (RR interval), PR interval, QRS duration, and QT interval in large consortia of 30,000 samples [103]. Using meta-analyzed genome-wide association results, the study identified seven novel locus-trait associations, of which six were successfully replicated [103]. Among their findings, a variant in *ATP2A2* gene was significantly associated with QRS duration that is known for measurement of cardiac depolarization, which causes the ventricular muscle to contract, resulting in pulsatile blood flow [104]. The study results identified *ATP2A2* as a candidate gene in the region that encodes a sarcoplasmic reticulum Ca2+-ATPase (SERCA) [103], which is involved in calcium transport in the human heart and under regulation of phospholamban [103, 105]. This GWAS strategy,

unlike candidate gene studies, is not based on the previous knowledge of the trait that identifies candidate loci contributing to the trait of interest but follows a hypothesis-free approach.

Accordingly, the high degree of similarity in the genetic profile across athletes and the relatively few individuals with low or high numbers of favorable genetic loci indicate a low level of heterogeneity in muscular power-related traits and/or VO_2max endurance-related phenotypes in the population. This low heterogeneity provides insights into the challenges of using genetic information in predicting athletic potential [106, 107]. Therefore, GWAS were utilized to reveal genetic predisposition of elite athletic performance by using a large number of genetic markers covering the entire genome in order to maximize the chance to reveal multiple genetic sites in various athletes' cohorts. For instance, Ahmetov I and his research group have examined the association between 1,140,419 SNPs and relative maximal oxygen consumption rate (VO_2max) in both males and females of eighty international level endurance athletes from Russia. The data indicated associations of rs1572312 in *NFIA-AS2*, rs7144481 in *TSHR* and rs7191721 in *RBFOX1* with aerobic exercise as well as their endurance level. To verify their data, investigators carried out a case/control study comparing the three most significant SNPs between their endurance athletes and controls (Russian power athletes and Russian and European controls). The results indicated that the frequency of the C allele of rs1572312 was significantly higher in elite endurance athletes (95.5%) than non-elite endurance Russian athletes (88.8%) [108].

GWAS can also be applied to predict risk of sports injury in athletes. Several studies in the context of sports genetics have identified a number of DNA polymorphisms associated

with anterior cruciate ligament tear, Achilles tendon injury, low bone mineral density and stress fracture, osteoarthritis, vitamin/mineral deficiencies, and sickle cell trait [109]. Herbert and his team studied the importance of physical activity on bone mineral density, and identified genetic associations considering gene–environment interactions such as sensitivity to mechanical loading from physical activity. Their results showed significantly higher total body bone mineral density in 84 weight-bearing athletes compared to 80 controls. A subsequent study of 99 elite academy footballers by the same group indicated significant associations between SNPs in *SOST*, *P2RX7* and *TNFRSF11* and bone phenotypes, although no genotype–training interactions were observed after a 12-week period of increased football training volume [110]. Mitchell and his team performed exploratory analyses in a discovery cohort of 918 athletes and a replication cohort of 486 athletes of European descent. They calculated the genetic risk score to investigate interactions between individual variants and proportions of high impact physical activity. A significant interaction between total physical activity and one variant (rs2887571) influencing bone mineral content was observed. However, this exploratory finding did not reach statistical significance in the cross-sectional replication cohort as variants only partly (≈6%) explained the variance in adult bone mineral density. Additionally, the contribution of other genetic variants influencing the pediatric skeleton response to physical activity exposure could not be ruled out [111].

It is believed that studying elite athletes' gene–environment interactions could provide a greater understanding of biological underpinnings of physical activity and subsequently changes in exercise and rehabilitation programs with reference to genetic backgrounds could have significant impact on sporting and public health sectors [110].

1.3.4 Limitations of genetic association studies

Overall, when it comes to physical performance, the association between genotype and phenotype is complex. For simple phenotypes such as height, identified SNPs through GWAS could explain up to 45% of its variance [112]. However, complex phenotypes, such as athletic performance, rely on the association analysis of thousands of genes that only explain a small percentage of the phenotype due to their small effect size [102, 113]. Such studies require a large sample size as well as a simpler intermediate phenotype to fully understand the genotype association and contribution of inheritance in elite performance. Until this day, none of the GWAS associations in physical performance has reached the genome-wide significance level of 5x E-8 [106]. This is because genetic talent is a result of a number of genetic variants and a complex combination of several environmental conditions influencing patterns of expressed genes [114]. In addition to the small effect size of identified SNPs, the number of participants represents the most significant limitation of genetic association studies. There becomes a need to include thousands of subjects to reach a sufficient statistical power to identify genetic variants that truly predispose to the phenotype of interest. Whereas this may be possible for some common diseases such cardiovascular disease, the number of elite athletes is limited for a given ethnicity and sport type, making and conducting such studies are more challenging. In order to increase sample size, a recent meta-analysis by the international consortium (GAMES) was conducted to compare elite endurance athletes and ethnicity-matched controls in a case-control study was designed by combining GENATHLETE and Japanese endurance runners. The meta-analysis tested a panel of 45 SNPs that were identified as promising markers by early studies on the GWAS basis of sports performance. The study that analyzed 1520 endurance athletes and 2760 controls

revealed only one statistically significant SNP (rs558129) at N-acetylgalactosaminyltransferase-like 6 (*GALNTL6*) locus, exhibiting the same direction of association with endurance, although it did not reach GWAS level of significance. The study concluded that there is no evidence of association of genomic signature that differentiates endurance status in world class athletes from controls. The study also highlighted that the sample size constitutes the main limitation which is hard to tackle because of limited number of elite endurance, the majority of whom are reluctant to take part in genetic studies [115]. Differences in allele frequency and association with performance among different ethnicities represent an additional limitation of genetic association studies, making it imperative to verify SNP associations in different ethnicities before candidates are named. A third limitation is heterogeneity of phenotype as some investigation reports data from athletes of mixed (power and endurance) sports with different cardiovascular demands instead of clearly defining athletes at either end of the spectrum such high power or high endurance ranked by their increased cardiovascular demand. This would offer a better understanding of variants associated with either power or endurance elite athletic status. Finally, many of these identified associations between SNPs and athletic performance were unique to athletes who were exposed to specific environmental factors including technique, kinematics, motivation, pain tolerance, adding to the complex trait. Such factors are very difficult to reproduce.

Functional validation of identified SNPs is required to understand the underlying mechanisms of SNP association with the phenotype such as gene expression studies. However, such studies would require sample collection from various tissues such as

skeletal muscles or myocardium which is not feasible in elite athletes. Such approaches are possible in animal models and possibly in non-athletes [116].

Overall, genomics studies are limited in predicting the complex phenotypes because genetic information is mostly static and does not account for the influences of the dynamic environmental factors [117]. There are difficulties in determining the effect of GWAS-discovered SNPs on nearby or distant genes as well as in determining the effect of the environmental factors including lifestyle, diet, emotional and psychological factors and physical activity on the development of several phenotypes. Therefore, it is more appropriate to consider metabolites as an alternative phenotype that may provide a direct functional link and consequently a greater effect size of genetic variants. Metabolomics have been recently used to explain the outcome of a variety “inborn errors of metabolism” in conjunction with the environment [118]. These “inborn errors of metabolism” would be reflected by differences in the concentrations of metabolites in body fluids and could provide insight into the “missing heritability” explained as “genetically influenced metabotype” (GIM) [119].

1.4 Metabolomics

Metabolomics is the measurement of endogenous metabolites in biological fluids, providing a “snapshot” analysis of metabolic pathways, gene function, enzyme activity, and the physiological effects [120]. Previous studies focused on the identification of metabolites and measurement of metabolite concentrations for the discovery of clinically relevant biomarkers and potential therapeutic targets [121]. For example, a study considered cardiovascular disfunction as the leading cause of mortality globally and investigated heart disease-associated metabolic derangements [122]. This study

indicated that carnitine palmitoyltransferase I inhibitor-dependent suppression of free fatty acids oxidation, and glucagon-like peptide-1-dependent enhancement of glucose oxidation are significantly associated with stimulation of myocardial function and constitute potential targets for therapeutic efficacy and cardiac recovery in heart disease [122].

Another study considered the potential of metabolomics serving as prognostic tools to perceive the development of cardiovascular disease [123]. The study assessed the metabolic difference associated with obesity and/or type 2 diabetes (T2D) and their contribution in disturbing function of body organs with several deleterious changes. These included inflammation of the adipose tissue, insulin resistance of skeletal muscle, nonalcoholic fatty-liver disease, dysfunction of beta cells, and vascular and cardio-myopathies [123]. Associated with these changes are metabolic derangements that can lead to increased acyl carnitines, lactate, and branched-chain amino acids (BCAAs), which collectively contribute to an increased risk for cardiovascular disease in these individuals [123] (Figure 2). The study suggested that identification of circulating metabolites that correlate with risk of disease would aid in the treatment and overall management of CVD in obese and/or T2D patients [123].

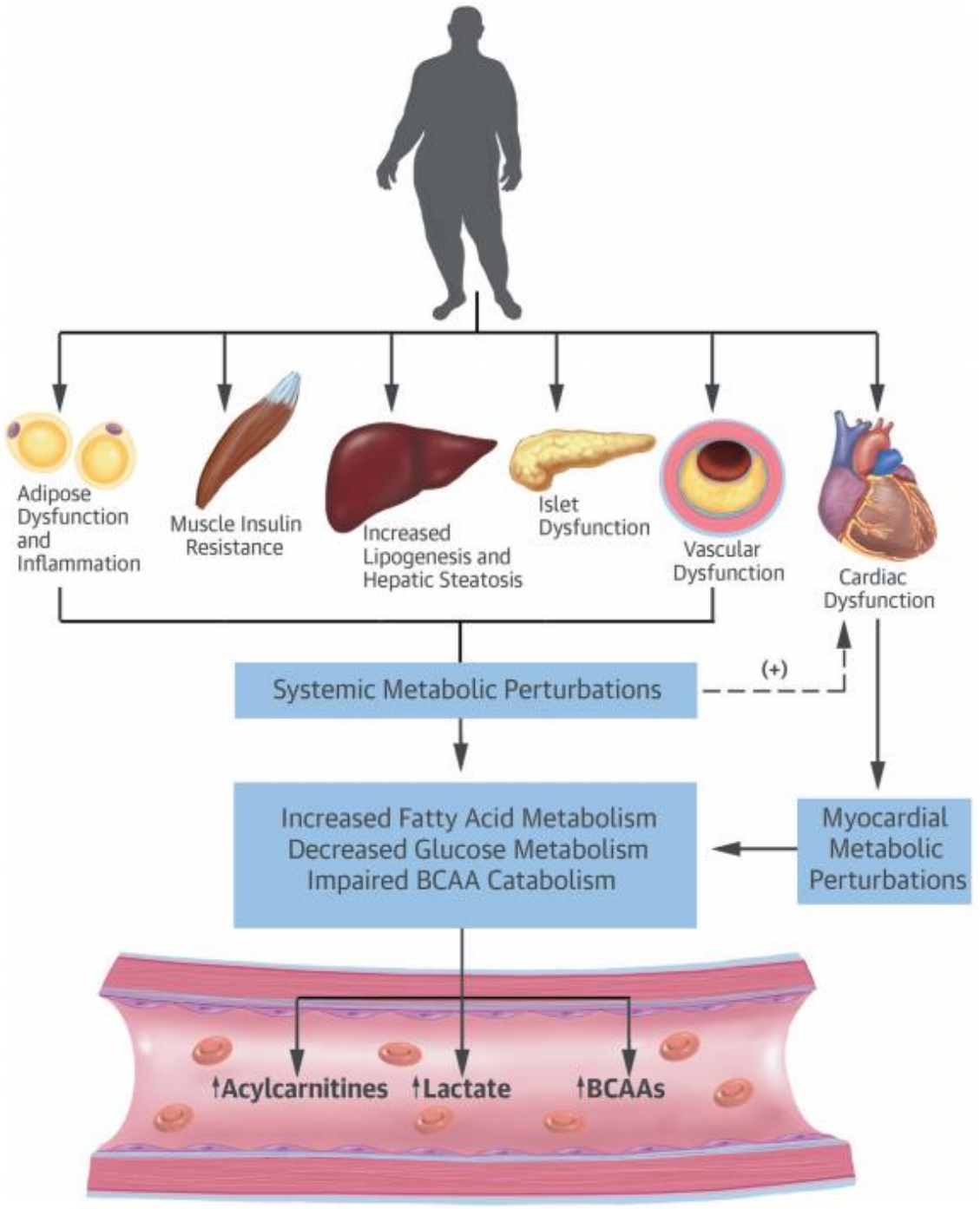


Figure 2. Blood-based metabolomic profiles of individuals with obesity and type 2 diabetes often lead to worsening of CVD system and development of cardiovascular disease. Abbreviations are CVD - cardiovascular demand and BCAAs - branched-chain. Figure adopted from previously published work [123].

The excessive training of professional athletes is associated with major alterations in their metabolic profiles that depends largely on the type and duration of their training regimen as discussed above. Exercise has putatively beneficial effects on blood pressure, blood lipids, and glucose homeostasis. Several biomarkers are commonly used to evaluate the physical status of elite athlete during their exercise [124, 125]. However, these methods were not sufficient enough to evaluate accurately the physiological alterations between endurance athletes and controls or differentiate pre versus post exercise markers [126]. Hence, there is a need for a more accurate metabolic profiling to reveal true physiological alterations in response to exercise.

Metabolomics presents a quantitative method for profiling of metabolites that change with physical activity in professional athletes, aiming at identification of biomarkers for performance, fatigue, and sports-associated disorders [126, 127]. These metabolic alterations are characteristic of glucose, lipid, amino acid and energy metabolism [126, 128]. For instance, metabolic profiling of exercising athletes showed elevation in plasma lactate [129] and breakdown products of adenine [130], typical of anaerobic respiration and ATP recycling. Other investigations of the metabolic impact of exercise indicated increased intermediates of the Krebs cycle, indicating aerobic energy generation in skeletal muscles [131]. High load training was also implicated in causing alterations in the concentrations of amino acids such as uptake of glutamate by the skeletal muscle, causing generation of alanine to induce ammonia metabolism [132], with associated alterations in plasma levels of these metabolites [133, 134]. Increase in serum concentration of sex steroid hormones was also shown in endurance athletes undergoing high exercise load [135].

Evaluation of these alterations presents invaluable assessments of athletes' physical status and response to exercise that would assist in planning their future exercise programs, avoiding potential disorders associated with intensive training and improving their general performance. The identified metabolites can also serve as potential biomarkers for professional athletes or may exhibit functional relevance as hormones [136, 137].

Metabolomics also offers a comprehensive method for identifying metabolic alterations associated with various dietary and environmental factors [138]. Approaches based on non-targeted/targeted metabolomics have substantially aided in the screening of hundreds of metabolites marking different metabolic signals [139]. Although the direction for comprehensive metabolic profiling is recommended to ensure all health aspects are accurately evaluated [140], it remains underutilized [141].

1.5 Genetically-influenced Metabotypes

Hundreds of GWAS have been processed for the identification of common genetic polymorphisms influencing common human traits. As a result, several genetic variants have been associated with numerous phenotypes by the selection of variants with an associated p value below a certain threshold in each individual SNP has a small overall effect [142]. Previous studies have shown that GWAS could explain up to 45% of some phenotypes such as height as discussed earlier, though for complex ones such as cardiovascular risk factors and coronary heart disease, it may not be able to explain more than 10–12% of the total trait variance [143, 144]. However, when taking metabolites into account, some genetic variants could explain up to 50% of their respective metabolic profile [145]. Several studies considered that the integration of personalized genetic and

metabolomics data would improve the success of precision medicine by enhancing the prediction of increased risk of disease and the therapeutic options [117]. For instance, the study conducted by Suhre and his team reported a comprehensive analysis of genotype-dependent metabolic phenotypes using a GWAS with non-targeted metabolomics to provide functional insights for many disease-related associations, including CVD and T2D [119, 146]. The study used several published CVD associated SNPs that were not statistically very robust but were considered since the identification of the biochemical function of the associated metabolic traits would explain the link to heart disease [146]. The results indicated several significant genome-wide association studies with metabolomics (mGWAS) associations. Among them was kallikrein B plasma (Fletcher factor) 1 (*KLKB1*) gene associated with the bradykinin pathway that controls blood pressure, as well as *ABO* (encoding ABO blood group) in association with transferase A (α-1-3-N-acetylgalactosaminyltransferase) and transferase B (α-1-3-galactosyltransferase) and *ALPL* (alkaline phosphatase) in association with fibrinogen A-α phosphorylation (FAaP). The data indicated that these associations could serve as biomarkers for acute myocardial infarction, and the combined additive genetic effect could predict the CVD risk [146]. Considering mGWAS results associated with T2D, the study reported ten associations included Carbamoyl-phosphate synthase 1 (*CPS1*) associated with glycine, as well as genetic variant in Fatty acid desaturase 1 (*FADS1*) associated with lysophosphatidylcholine acyl (lysoPC a C18:2). The latter is a genetic risk factor for hyperglycemia, preferentially associating with abnormal cholesterol and triglyceride levels [147].

Consequently, studies suggested that mGWAS results provide useful clues for understanding the causality of disease progression when studying the functional role of metabolic pathways in combination with genetic evidence. Figure 3 summarizes the importance of metabolites as an intermediate phenotype to examine function of the genomic allele.

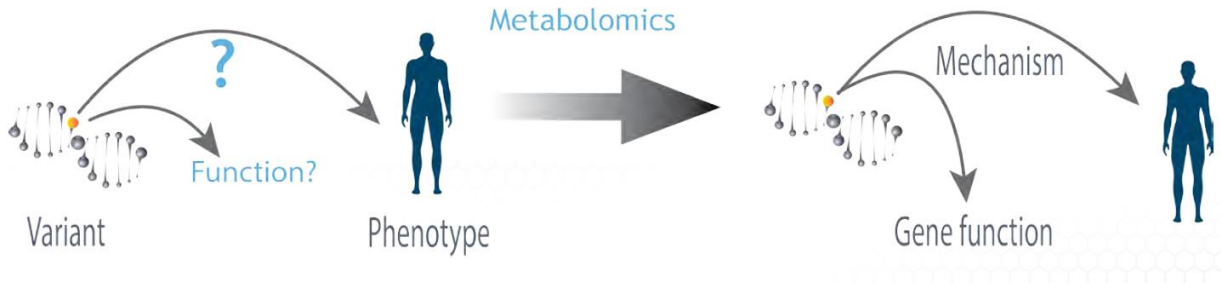


Figure 3. Metabolomics is a genome sentinel and metabolites are highly informative intermediate phenotypes between genes and their functional presence (Adopted from [148]).

Therefore, analyzing mGWAS and gathering knowledge of the heritable part of the genetic variation in metabolism could explain a significant part of complex phenotypes by providing intermediate phenotypes that collectively explain these phenotypes [146, 149]. mGWAS provides measurable hypotheses for functional genomics and metabolomics and identify novel gene functions and metabolite identities, allowing for more comprehensive and system-based downstream analyses [150]. Figure 4 summarizes the purpose of “OMICS” by identifying alleles that are associated with intermediate phonotypes that contribute collectively to the final disease phenotype.

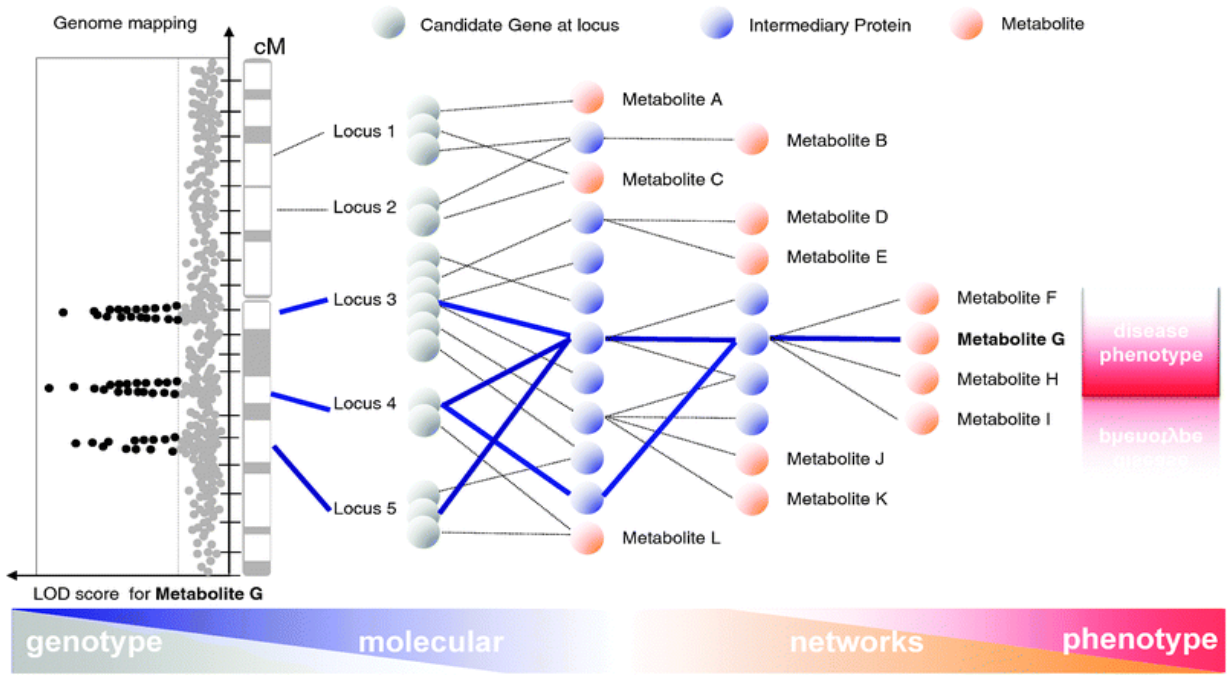


Figure 4. Genetically influenced metabotypes. Genome mapping will identify a number of SNPs at various loci linked to specific genes that are associated with an intermediate protein. These proteins mark certain metabolites that together are associated with risk of disease. This, joint "OMICS" approach is necessary to provide an intermediate phenotype for genetic variation as has been the aim of this project (Adopted from [151]).

There is ample evidence suggesting genetic influence of multiple genetic variants with small effect size over several phenotypic traits related to physical performance [152]. The identification of these variants is crucial to understand the superior performance of elite athletes and has been a subject of study for many years [79, 80, 107]. Limited evidence of genetically-influenced metabolites in the context of physical exercise exists. Pilot studies conducted by Karoly and his team investigated 200 sedentary individuals to determine the correlations between 14 SNPs and responses to a 30min session of aerobic exercise. Participants were strictly instructed what to eat, drink and when to

exercise. Blood samples were collected immediately before exercise, 10 min and 30 min into exercise to measure lactate concentration and catecholamine levels (epinephrine and norepinephrine) that were considered as physiological phenotypes in the analysis. Their results indicated that *SLIT2* SNP rs1379659 and the *FAM5C* SNP rs1935881 were related to changes in norepinephrine. *OPRM1* SNP rs1799971 was also associated with norepinephrine and lactate changes during exercise. However, the study concluded that the physiological responses to aerobic exercise involve a complex interaction of metabolic functions, thus these genes and SNPs are likely to explain only a small portion of the variability in individual differences in response to aerobic exercise [153, 154].

In conclusion, several studies used unbiased genome-wide approaches in the form of case-control study that have uncovered many new loci [83, 91]. However, for more reliable functional associations when analyzing genomic regions, further studies showed interest in metabolomics. The advancement in metabolomics tools including bioinformatics technologies has offered a unique opportunity to complement enormous sample sizes and massive genomics data with intermediate phenotypes as metabolic profiles. Several studies showed that identified metabolites exhibited direct functional associations with genetic variants and provided greater effect sizes [150, 155]. Previous mGWAS findings [146, 150, 156-164] have revealed hundreds of metabolomics quantitative trait loci (mQTLs) in the general population [145, 162-166]. The identification of mGWAS in athletes who experience unique environmental conditions including special diet and intensive exercise may provide invaluable tools for biomarker discovery in relation to exercise and performance. This unique approach could provide better informed

selection of athletic candidates and crucial information needed for optimal balance between training and recovery for every athlete [167]. The integration of genomics and metabolomics technologies has also allowed a more comprehensive coverage of the metabolic pathways involved in complex physiological and pathological processes [168, 169].

In this study, genetics, metabolomics, and genetically-influenced metabolites profiling of consented elite athletes collected at anti-doping labs during the study are investigated and reported.

1.6 Hypothesis and Aims of this project

Combining 'omic' approaches will facilitate the discovery of the metabolic and genetic determinants that underlie elite athlete performance within selective athletic disciplines

1.6.1 Overall objectives:

This study aims to determine the genetic predisposition of elite athletic performance in endurance and power athletes with different cardiovascular demand and identify intermediate phenotypes (metabolic profiling) that could explain some of the identified genetic associations.

1.6.2 Specific aims:

1. To develop a sample bank comprising of serum and DNA from cohorts of elite athletes collected at the ADLQ.

2. To explore genetic association by performing comprehensive GWAS analysis on DNA derived from cohorts of elite athletes from defined disciplines and performance activities based on endurance, power and cardiovascular demand.

3. Investigating global sera metabolic profiles in cohorts of elite athletes using the Metabolon platform.

4. Employing complex bioinformatics and statistical analysis to interrogate GWAS and Metabolomics data and defined genetically-influenced metabotypes

CHAPTER 2
MATERIALS AND METHODS

2 (Chapter 2) Materials and Methods:

2.1 Ethical Approval

Ethical approval from the Institutional Research Board of Anti-Doping Lab Qatar (F2014000009) was obtained in line with the World Medical Association Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects.

2.2 Study participants and Power calculation

Consented elite athlete's samples from various sport groups who took part in national or international sport competitions and tested negative for doping abuse at anti-doping laboratories in Qatar (ADLQ) and Italy (FMSI) were used in this study. Based on article investigating *ACTN3* polymorphism by comparison between elite swimmers and runners [170], using *ACTN3* R allele frequency in long distance (42%) vs. short distance (61.8%) elite runners, a power calculation indicates 102 participants per group to detect significant difference at alpha 0.05 and 80% power. However, based on R allele frequency in elite long distance (57%) vs low distance (62.5%) swimmers, a power calculation indicates the need for 1247 per group. Since athletes recruited in our study were a mixed population of runners, swimmers, footballers, athletics etc.., we believe that our study was probably underpowered to detect significant differences in genotypes between high (n=639) and low (n=114) endurance. However, because of limited number of available samples through the anti-doping laboratories), we have used an intermediate phenotype (metabolites) that provided a larger effect size, increasing the chance to pick up significant differences with our sample numbers.

Samples were received in ethylenediaminetetraacetic acid (EDTA) anti-coagulated tubes and/or serum-separator tubes (SSt) on dry ice. Serum samples were instantly aliquoted

into 2ml tubes and stored at -80ºC until use. Blood samples were stored at –20ºC until use. . All the other information of participants were not available, including the time of recruitment (pre or post exercise), measurements of weight, body mass index, fat mass, muscle mass, bone mass and basal metabolic rate due to the strict anonymization process undertaken by anti-doping laboratories. Table 2 summarizes number of samples collected to date and type of analysis performed.

Athletes were grouped into classes based on the endurance, power and cardiovascular demand phenotypes-associated with their sport types using previously published criteria as introduced earlier [16]. Tables 3, 4, 5, and 7 further list the number of athletes and their information (genders and sport type) used in subsequent bioinformatics analyses applied for each class.

Table 2. Total number of recruited samples and type of performed analyses throughout the study.

	Blood only	Blood + Serum	Serum only	Total
Recruited	884	512	191	1587
Processed (excluding dropouts)	758	501	191	1450
GWAS	1259		-	1259
Metabolomics	-		692	692
mGWAS	490		-	490

2.3 Genome Wide Association analysis

Blood from EDTA tubes and clotted blood from Serum separating tubes was collected from one thousand two hundred and fifty nine (1259) consented elite athletes. Table 3 summarizes the distribution of elite athletes' samples that were used in GWAS into endurance, power and cardiovascular demand (colored) groups following published criteria [16].

Table 3. Distribution of all genotyped samples into three phenotypic groups: endurance (columns: moderate and high), power (rows: moderate and high) and cardiovascular demand (white and grey: low and high) based on their sports discipline.

GWAS		Endurance	
		Moderate (<70% VO2max)	**High (>70% VO2max)**
Power	**High (>50% MVC)**	Wrestling (14M/3F), Skate boarding (2M), Athletic-Throw (3M), Judo (3M), Weight-Lifting (43M/13F), Taekwondo (3M/1F), Martial arts/ Jiu-jitsu (2M)	Boxing (9M/17F), Heptathlon (1M), Kayaking (2F), Rowing (9M/8F), Biathlon (2M/1F), Mordent Pentathlon (1F), Cycling (196M/55F), Triathlon (10M/10F)
	Moderate (20-50% MVC)	Sprint (4M), Athletic-Jump (1M/1F), Rugby (18M), Aquatics (11M/8F), Athletics (128M/90F)	Handball (30M/8F), Athletics middle distance (6M/2F), Skiing Cross Country (3M/1F), Hockey (10M/1F), Basketball (8M/1F), Swimming (34M/22F)
	Low (<20% MVC)	Baseball (2M), Volleyball (6M), Table-Tennis (13M/1F)	Athletic-Long-Distance-Marathon (67M/26F), Athletics-Ultra-Running (1M), Tennis (5M/3F), Soccer (315M), Badminton (1M/1F), Football (20M/3F)

2.3.1 DNA extraction

Extraction of genomic DNA was performed at anti-doping laboratory Qatar, using different Qiagen kits recommended for the type of blood samples according to the manufacturer's instructions. All DNA extraction methods involved proteinase K-mediated protein denaturation, followed by several steps of ethanol DNA washing, hydration and precipitation. Additional steps were performed depending on type of blood used in extraction. Accordingly, this section is divided into three sub-sections: whole blood samples (EDTA-tubes), clotted blood samples (SSt-tubes) and fully automatable DNA extraction via QIAcube. Isolated DNA was loaded on gel electrophoresis to check for DNA integrity and quantified with Nano-Drop (Thermo Scientific, cat. # ND-2000) and Qubit Fluorometer (Invitrogen, cat. # Q32866) to ensure that all DNAs exhibit the required

quantity and quality to achieve the study objectives. When needed, increased concentration of extracted DNA was achieved by vacuum evaporation. DNA samples were stored frozen at -20ºC. The detailed protocol is described as follows.

2.3.1.1 DNA extraction from whole blood samples collected in EDTA tubes

DNeasy Blood & Tissue kit from Qiagen (Cat# 69506) were used for isolation of genomic DNA from EDTA-anti-coagulated whole blood (n= 554) samples. Sample tubes (n=24) were thawed at a time of extraction and 200µL of each sample was transferred to a new 1.5ml tube. Lysate buffer with proteinase K was added and incubated for 1hr at 65ºC with 900rpm shaking. After centrifugation at 6000xg for 1min, the supernatant containing the DNA was transferred to a clean 1.5ml tube and mixed with absolute ethanol. After vortexing vigorously, the mix was transferred to filter tube, and the DNA was precipitated with nuclease-free water after two washes with washing buffers that were provided by the kit. The concentration and the quality of DNA was measured using the Nanodrop (Thermo Scientific, cat. # ND-2000) and the DNA was stored at -20ºC. Figure 5 summarizes workflow of DNA extraction using DNeasy Blood & Tissue kit.

Figure 5. workflow of DNA extraction

2.3.1.2 Clotted blood samples obtained from Serum Separation tubes

Due to interest in running GWAS and Metabolomics on the same samples, blood collected in serum-separator vacutainer (for growth factor anti-doping tests) were utilized, where serum was separated without using anti-coagulant, therefore clotting took place in the remaining blood at the bottom of the tube. Serum was collected in 2ml tubes and stored at -80^{O}C for metabolomics analysis at Metabolon, Durham, NC, USA. The clotted blood was used for genomic DNA isolation which was performed at anti-doping lab Qatar. Previously devised method was utilized to efficiently remove clotted blood (n= 501) from the serum-separation gel without being contaminated by the separation gel [171]. This method was performed after collecting serum by inverting the capped serum-separator tube free of serum within a 50ml tube and spinning at 2000xg for 5min. The clot was then transferred to a clean 15ml tube (Figure 5) and mixed with 10ml Red Blood Cell (RBC) lysing buffer (Sigma cat# R7757) and vigorously vortexed and incubated at room temperature for 40min, then centrifuged at 2000xg for 5min. If the resultant pellet was still thick and dark in color, this step was repeated. Supernatant was discarded and 200µL from the pellet was transferred to 1.5ml tube. Figure 6 summarizes steps of collecting clotted blood from the serum separation vacutainer tube into 15ml tube.

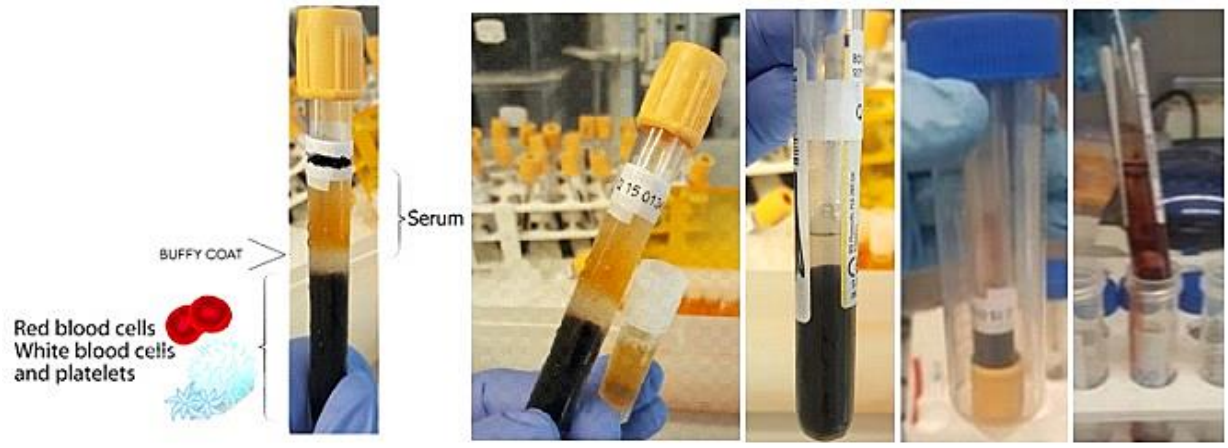


Figure 6. Collecting clotted blood from the serum separation tube. Clotted blood was used for DNA extraction whereas serum from the same tube was used for metabolomics analysis.

This was followed by DNA isolation using QIAmp DNA blood mini kit from Qiagen cat# 51106. Lysate buffer with proteinase K was added and incubated at 56^{o}C for 2 hours with shaking 1200rpm. NaCl (5mol/L) was then added and centrifuged at 6000xg for 1min. Supernatant was transferred to a new tube and mixed with absolute ethanol. The mix was vortexed and transferred to filter column then centrifuged at 6000xg for 1 minutes [172]. The column was then washed twice with washing buffers provided by the kit and centrifuged at 20,000xg for 3min to dry the filter and finally re-suspended DNA in nuclease-free water (Life Technologies, Paisley, UK). The concentration and the quality of DNA was measured using the Nanodrop (Thermo Scientific, cat. # ND-2000) and the DNA was stored at -20^{o}C.

2.3.1.3 *Fully automatable DNA extraction via QIAcube*

Some of the blood samples (n=204) were placed into QIAcube (Qiagen, cat. # 9001293) that is fully automated to increase standardization for purification of high-quality genomic DNA using QIAamp DNA Blood Mini QIAcube Kit Cat # 51126. Procedures comprises the same four steps as the manual procedure (lysing, binding, washing, and eluting) (Figure. 4), but enabling seamless integration of automated, low-throughput of laboratory workflow (Figure 6). QIAcube is carried out by placing samples to their predefined positions followed by loading buffers that all are provided in the kit with rotor adapters (preloaded with spin columns and elution tubes) in a standard microcentrifuge, on a vacuum manifold, as it is prescribed in QIAcube protocol. The procedures are designed to ensure that there is no sample-to-sample cross-contamination and to allow safe handling of potentially infectious samples. The concentration and the quality of DNA was measured using the Nanodrop and the DNA was stored at -20^{o}C. Figure 7 workflow of DNA extraction using fully automated method via QIAcube

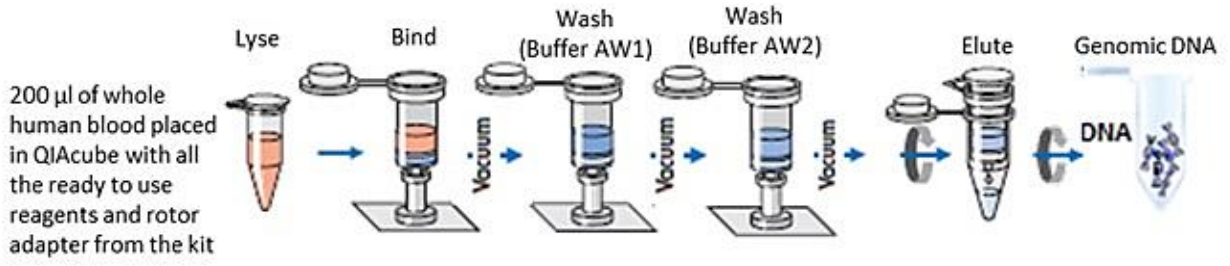


Figure 7. Workflow of DNA extraction using QIAcube

2.3.1.4 *DNA quantity and quality assessment:*

DNA was visualized by gel electrophoresis method. 2% agarose (Sigma-Aldrich cat. # A9539-10G) gel was prepared with 2% tris-acetate EDTA (Sigma-Aldrich cat. # T9650-

1L) and 0.5 µg/mL of ethidium bromide (Invitrogen cat. #15585011). 10µL of isolated DNA was mixed with 2µl of loading dye (Thermo Scientific cat. # R0611) and loaded on the gel along with 10µl of 1kb DNA ladder (Promega cat. # G5711) to be assessed under UV light. The gel was run for 45 minutes to 1 hour at a voltage of 90 and was then examined using UV benchtop transilluminators (Thomas Scientific cat. # 6284E17). The quantity and quality of extracted DNA was assessed before going further to make sure that enough DNA (≥50ng/µL) was obtained for genotyping, otherwise, additional steps of increasing DNA concentration were performed by reducing the elution volume to 10µl by evaporation by vacuum (Eppendorf™ Concentrator Plus Complete, cat. # 5305000568) under 40ºC for 2hours. DNA was quantified using Nanodrop (Thermo Scientific, cat. # ND-2000) at optical density (OD) of 260 nm and quality was assessed at ratio of 260/280 ODs. Further assessment was carried out using Qubit dsDNA High Sensitivity Assay Kit, Invitrogen cat# Q32854 that quantifies double stranded DNA accurately according to manufacturer's instructions (Figure 8). Qubit working solution was prepared at 1:200 dilution of reagent with buffer. Qubit standards and DNA samples were mixed with Qubit working solution in thin-walled polypropylene tubes (Invitrogen cat. # Q32856) and incubated at room temperature in dark for 20min then measured. DNA concentration of 50ng/µl and a 260/280 ratio >1.7 was considered acceptable for genotyping.

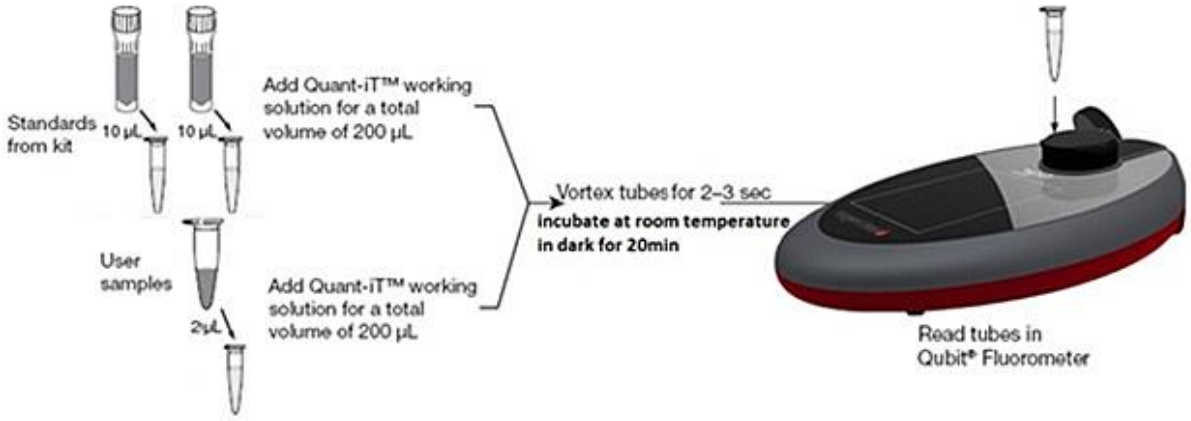


Figure 8. DNA quantification using Qubit

2.3.2 Whole Genome Genotyping Arrays

Among multiple SNP chips options available for GWAS, Illumina Drug Core array-24 BeadChips was chosen for genotyping of 476728 SNPs in athletes of different type of sport. This array represents a new product leading in translational genomics and computational chemical biology with 240,000 highly-informative genome-wide tag SNPs and a novel 200,000 custom marker set designed to support studies of drug target validation and treatment response. It supports genotyping of SNPs selected according to sets of criteria like: 1- genes involved in drug absorption, distribution, metabolism and excretion (ADME), 2- exome content coverage of genes encoding proteins closely related to targets of approved small molecule and biotherapeutic drugs or binding drug-like compounds, and 3- other useful content, including all SNPs associated at genome-wide significance with any human trait marking the X and Y chromosomes and mitochondrial DNA, and for sample fingerprinting (common SNPs represented on major genome-wide array products from both Illumina and Affymetrix). These SNPs are expected to be

involved in regulating the same essential metabolic pathways that regulate the magnitude of physical performance such as *ACTN3*, *ACE*, *IL-6* and *PPAR*.

The assay required 4 µl of the DNA sample (200 ng) as input with a concentration of at least 50 ng/µl. Exact concentrations of DNA in each sample were measured using a Qubit Fluorometer (Invitrogen, cat. # Q32866). All further procedures were performed manually at anti-doping lab Qatar following the instructions of Infinium HD Assay (Figure 9). 96 samples were processed at a time. First day of applying the genotyping assay, 4µl of obtained DNA was mixed with Illumina amplification reagents and incubated overnight at 37^{o}C in hybridization oven. On the second day, enzymatic reagents were used to fragment the amplified DNA then precipitated by centrifugation. Subsequently, the pellet was resuspended and loaded in the beadchip (twenty-four samples per array) for hybridisation to the oligonucleotide probes by overnight incubation at 48^{o}C in the hybridization oven. On the third day, a washing step was carried out to remove the unhybridised DNA. Beadchips underwent enzymatic two-colour single-base extension step in which differentially labelled nucleotides were added to the oligonucleotide probes corresponding to nucleotide in the hybridised DNA samples, followed by fluorescent staining. Lastly, after coating, the beadchips were imaged in iScan system (Illumine, cat. # SY-101-1001) (Figure 10) and the output files and recorded intensity were imported into Illumina GenomeStudio software (Figure 11) to extract the genotype information. Almost all samples passed the QC with genotyping call rate greater than 98%.

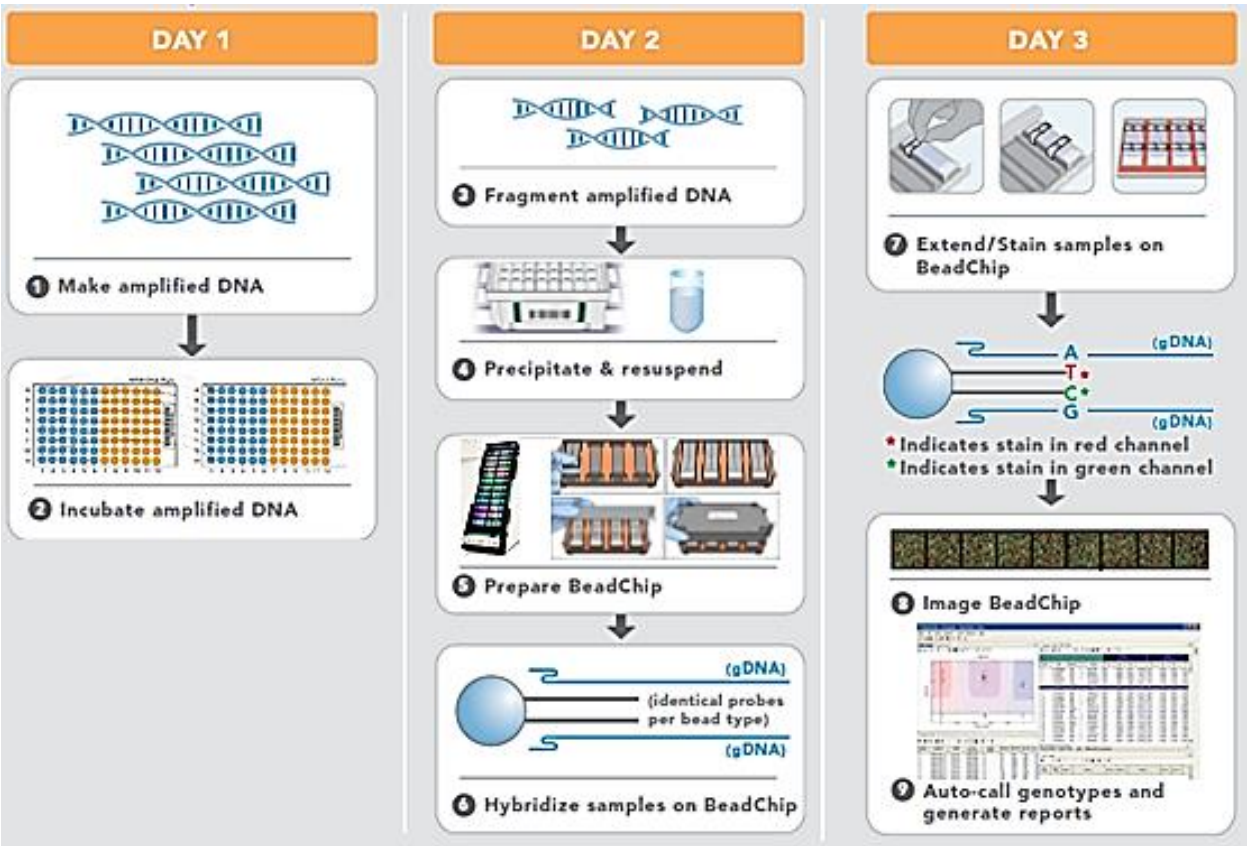


Figure 9. Workflow for genome-wide SNP genotyping using the Illumina Infinium assay.

Figure is reproduced from Illumina.com

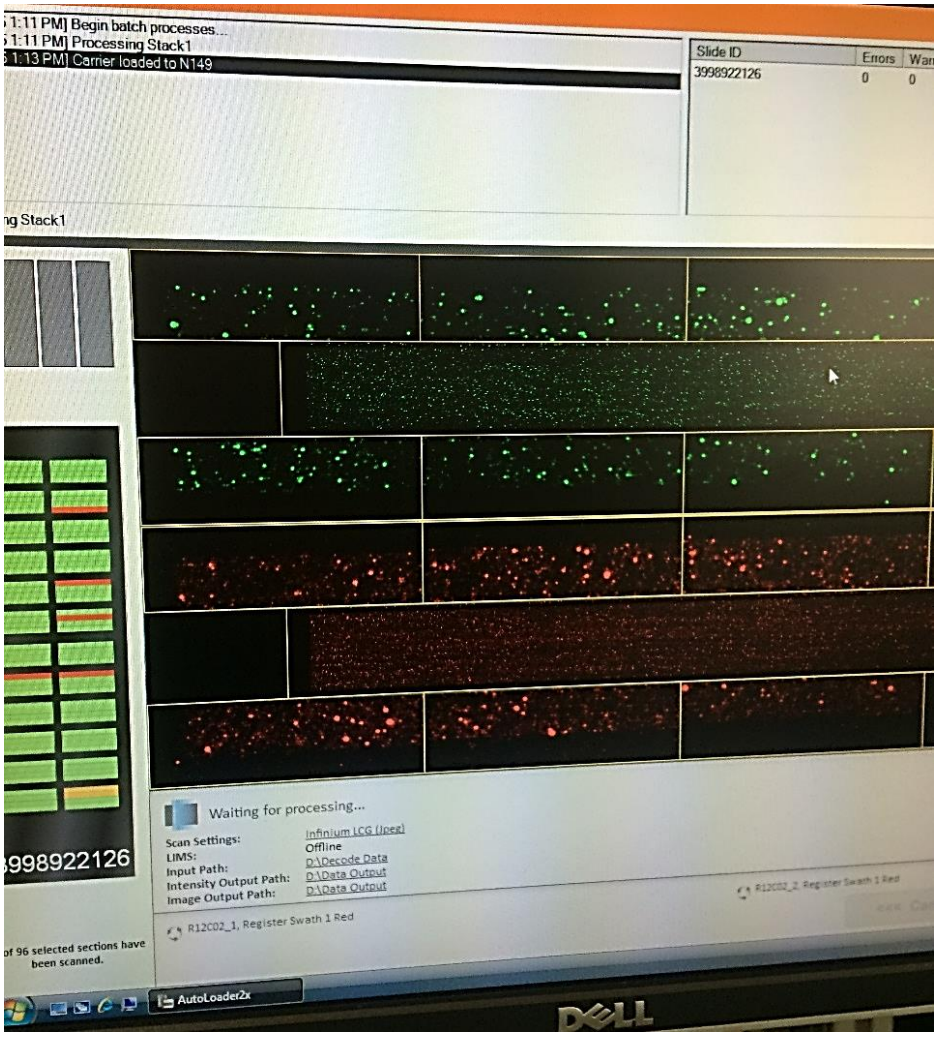


Figure 10. A screen capture of beadchips imaging in iScan system (Illumine, cat. # SY-101-1001). The preview information bar is located at the left side of the iScan Control Software Scan screen displays the section of the strip that currently being scanned. Green indicates the successful data reading. Yellow indicates the conditional iScan reading. Red indicates unsuccessful data reading. If 1 or more sections of a BeadChip did not successfully scan, Rescan can be performed by clicking on the iScan Control Software Review screen to rescan the BeadChip. The iScan Control Software only rescans the sections that are not successfully scanned.

2.3.3 Data Extraction and SNP Identification:

Illumine iScan (Illumine, cat. # SY-101-1001) hardware and software were used for raw data extraction, peak-identification and QC process. Illumina SNP array files including raw signal intensity files, cluster files and the sample sheet were loaded into GenomeStudio. The sample call rate and SNPs signal intensity were checked in GenomeStudio as shown in figure 11. Then, the exported data were selected from GenomeStudio options to save genotype data for each sample into ped/map format. SNPs were identified by statistical analyses as described later.

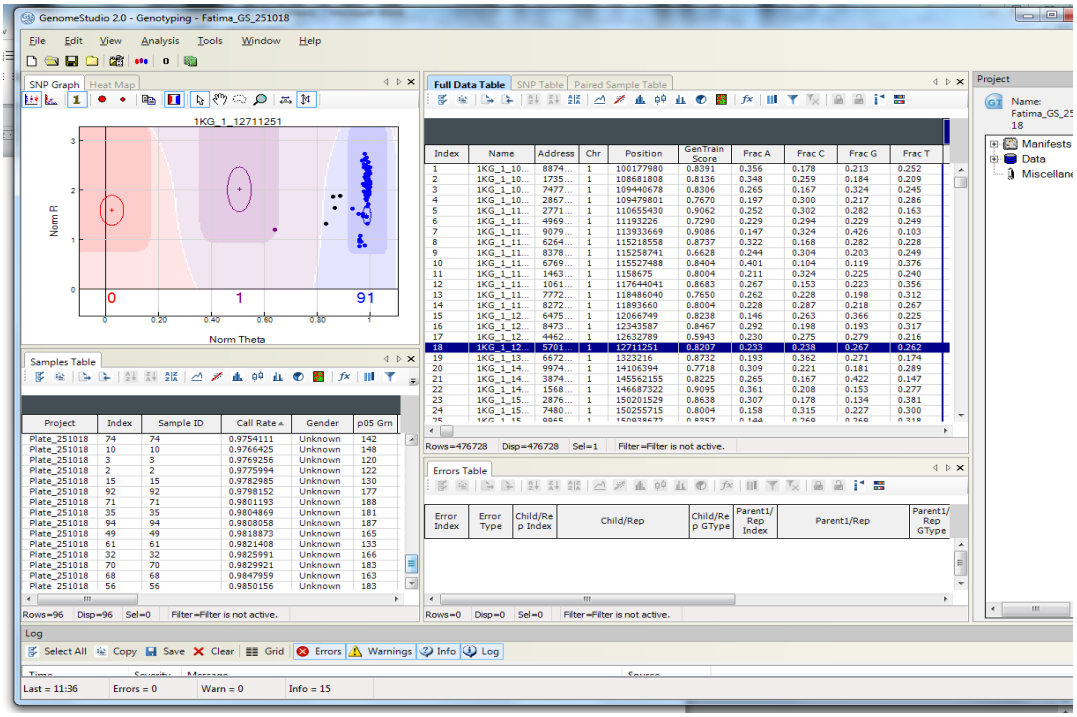


Figure 11. Shows a screenshot of signal intensities in GenomeStudio. The image indicates genoplot with samples falling in each of three genotype clusters based on their alleles (red points are AA, purple points are AB, blue points are BB). Black points in the genoplot shown in the right panel, represented the samples that dramatic drop in intensity, which may signify a homozygous deletion in those samples.

2.3.4 Statistical analysis of genomics data

Statistical analysis was performed with Plink using the SNP array genotyping data after filtering out SNPs with genotype call rate < 98%, Hardy Weinberg p value < 10^{-6} and minor allele frequency (MAF) < 0.01. The commonly used 1% cut off for minor allele frequency for GWAS studies was adopted taking into account that athletic performance is a complex, but not rare, phenotype among athletes, therefore common polymorphisms are expected to be identified, especially by combining the metabolic profile as an intermediate phenotype. Additionally, lowering the cut off will lead to increasing the penalty of multiple testing correction. Furthermore, at the most extreme level, if all but one variant cluster together, it is difficult to assess whether the only variant is truly a different genotype, or whether it is a missed call [173]. Number of removed samples and SNPs are indicated in the results chapter. Principle component analysis (PCA) was used for uncovering any hidden segregation of samples based on maximizing the variance. Association analysis was performed with PLINK v1.90b4.6 64-bit using plink function (Figure 12). Plink code incorporated sports ranked according to the lowest endurance, power or cardiovascular demands (ranked 1) and the highest (ranked 2). A categorical variable with two levels (moderate & high) is sensible since the two classes of either sport phenotype feature a heterogeneous distribution of the other phenotype's classes as described in the study design section (Table 3). With both analyses, covariates including gender, and PCA components 1, 2, 3 & 4 were included in the model. Despite not picking gender effect in PCA, gender was included as a confounder in statistical analyses because the differences in physical performance between males and females are well established. This is related to hormonal differences (higher estrogen levels in women and higher levels of

testosterone in men). Testosterone enables men to develop larger skeletal muscles as well as larger hearts [174, 175]. Therefore, sex is considered a major factor influencing physical performances. A stringent Bonferroni level of significance was calculated and used to define significant associations. For a selected set of SNPs (see tables 8, 10 and 12 in the results chapter), genotype distribution among groups of athletes was reported.

For demonstration of most significant SNPs, Manhattan and quantile-quantil (QQ) plots were generated using R (version 3.3.1). The QQ plot demonstrates a plot of the observed vs the expected test statistics of the P-values, thus is used to examine the genomic inflation [176]. The observed P-values should be in large deviations from the expected distribution in the lower tail of the QQ plot, whereas the deviation in the upper tail indicates a possible association [176].

A web-based software application called ingenuity pathway analysis (IPA) QIAGEN, cat. # 830011 was also performed to gain further insight into the functional relevance of identified SNPs. IPA searches for targeted information on genes and enables analysis, integration, and understanding of data from gene expression [177]. Data analysis and search capabilities help in identification of candidate biomarkers in the context of larger biological or chemical systems associated with deeper understanding of metabolomics, as well as discoveries about disease processes [177].

```
falkhelaifi@ubuntuADL-Plink:~/plink$ ./plink --adjust  --bfile Eur_strict --ci 0.95  --covar
covar_file.txt  --covar-name sex,PC1,PC2,PC3,PC4  --hide-covar  --logistic  --out Eur_Strict_
Endurance_PC1_4_logistic_adjust.assoc.logistic   --pheno pheno_clean1.txt  --pheno-name Endur
ance
PLINK v1.90b3y 64-bit (4 Nov 2015)         https://www.cog-genomics.org/plink2
(C) 2005-2015 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to Eur_Strict_Endurance_PC1_4_logistic_adjust.assoc.logistic.log.
Options in effect:
  --adjust
  --bfile Eur_strict
  --ci 0.95
  --covar covar_file.txt
  --covar-name sex,PC1,PC2,PC3,PC4
  --hide-covar
  --logistic
  --out Eur_Strict_Endurance_PC1_4_logistic_adjust.assoc.logistic
  --pheno pheno_clean1.txt
  --pheno-name Endurance

Note: --hide-covar flag deprecated.  Use e.g. '--linear hide-covar'.
16048 MB RAM detected; reserving 8024 MB for main workspace.
341385 variants loaded from .bim file.
753 people (595 males, 158 females) loaded from .fam.
753 phenotype values present after --pheno.
Using 1 thread (no multithreaded calculations invoked.
--covar: 5 covariates loaded.
Before main variant filters, 753 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Warning: 23742 het. haploid genotypes present (see
Eur_Strict_Endurance_PC1_4_logistic_adjust.assoc.logistic.hh ); many commands
treat these as missing.
Warning: Nonmissing nonmale Y chromosome genotype(s) present; many commands
treat these as missing.
Total genotyping rate is 0.997459.
341385 variants and 753 people pass filters and QC.
Among remaining phenotypes, 639 are cases and 114 are controls.
Writing logistic model association results to
Eur_Strict_Endurance_PC1_4_logistic_adjust.assoc.logistic.assoc.logistic ...
63%^[[done.
--adjust: Genomic inflation est. lambda (based on median chisq) = 1.15833.
--adjust values (323696 variants) written to
Eur_Strict_Endurance_PC1_4_logistic_adjust.assoc.logistic.assoc.logistic.adjusted
```

Figure 12. A screenshot of plink processing logistic regression analysis on large data set of 753 elite athletes' samples to analyze the variants in genetic models by performing basic statistical tool.

2.3.5 Validation of significant GWAS findings in a second cohort

Validation of results was performed by comparing the frequencies of the most significant SNPs (with P <E-5 - E-8) in 219 elite Russian athletes (120 middle-distance athletes, 56 long-distance athletes, 43 sprinters) and 173 Russian controls. All athletes were Olympic team members (International level) who have tested negative for doping substances. DNA extraction, genotyping and GWAS analysis of this replication study cohort was

performed by research scientist at Department of Molecular Biology and Genetics, Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia. To perform the meta-analysis, the Cochrane Review Manager (RevMan) version 5.3 was used. Random and fixed effect models were applied. The heterogeneity degree between the studies was assessed with the I^2 statistics.

2.4 Metabolomics

Metabolon, Inc., is a commercial supplier and a leader in metabolomics-driven biomarker discovery, including the identification and relative quantification, and quality-assurance components of the process [178]. Aliquot of each serum sample were shipped frozen in dry ice to Metabolon, Durham, NC, USA to perform metabolites profiling using established protocols as detailed below. Despite outsourcing the metabolomics analysis at Metabolon, Durham, towards the end of PhD, Metabolon established a sister-lab at ADLQ where I was personally involved in processing and analysis of samples belonging to other projects. This opportunity has gained me a practical experience in metabolomics analysis that was added to my analytical expertise in liquid chromatography used in ADLQ's routine doping tests:

2.4.1 Samples transportation:

Firstly, 191 serum samples were shipped to Metabolon as the discovery group (Table 4.a) followed by 501 serum samples as the replication group (Table 4.b). The overall characteristics of the study participants are provided in Table 4, and they were divided into different classes for statistical analysis as shown in Figure 13.

Table 4. Categorization of metabolomics study participants. Elite athletes were distributed in various categories: 4.a Discovery cohort and 4.b replication cohort, based on sport type as described previously [16]. These categories were based on the VO_2Max, MVC and CVD associated with their respective sports types. The number and gender (M for males and F for females) of participants in each group are also indicated.

(4.a) Discovery		Endurance	
		Moderate (40-70% VO2max)	**High (>70% VO2max)**
Power	**High (>50% MVC)**		Boxing (12M), Triathlon (4M), Rowing (2M), Cycling (21M/3F), Canoe Kayak (1F)
	Moderate (20-50% MVC)	Athletics (5M/5F), Rugby (41M), Canoe 200m Sprint (4M/1F), Motorcycle Racing / Road (5M)	Skiing Cross Country (1M), Basketball (4M), Swimming (9M)
	Low (<20% MVC)	Baseball (1M), Volleyball (4M/4F)	Tennis (1F), Football (62M), Long distance 3000m or greater (1M),

(4.b) Replication		Endurance	
		Moderate (40-70% VO2max)	**High (>70% VO2max)**
Power	**High (>50% MVC)**	Wrestling (3M), Judo (3M)	Boxing (1M/16F), Heptathlon (1M), Rowing (6M/7F), Cycling (31M/4F)
	Moderate (20-50% MVC)	Athletics (15M/22F), Rugby (16M), Triple Jump (1M)	Athletics 200-800m (4M), Hockey (1F), Skiing Cross Country (1M), Basketball (3M), Swimming (22M/16F)
	Low (<20% MVC)	Baseball (2M), Volleyball (1M)	Tennis (1M/1F), Soccer (315M), Athletics 1500-3000m (3M)

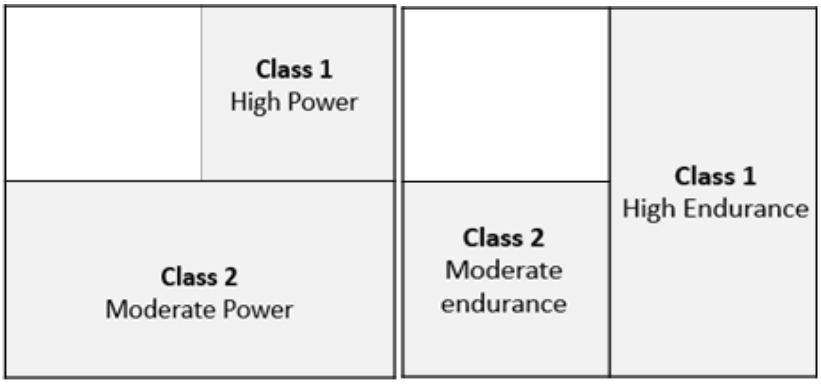


Figure 13. Classification of sport groups based on endurance only regardless of power (right) and similarly for power only ignoring endurance (left); statistical analysis these classes were used in the analysis.

2.4.2 Sample Preparation:

Following Metabolon protocols previously published by us [167], serum samples were identified then kept at -80ºC until needed. Samples were processed via automated MicroLab STAR® system (Hamilton). A number of recovery standards were included before the first extraction step for quality control. Sample were mixed with methanol and vigorously shaked for 2 min (Glen Mills GenoGrinder 2000) then centrifugated to eliminate protein via small molecules trapped in the precipitated protein matrix, as well as to regain chemically-diverse metabolites. The extracted mix was then divided into 5 fractions (2 fractions for 2 independent reverse phase (RP)/UPLC-MS/MS analyses with positive ion mode electrospray ionization (ESI), 1 for RP/UPLC-MS/MS analysis with negative ion mode ESI, 1 fraction for HILIC/UPLC-MS/MS analysis with negative ion mode ESI, and 1 fraction for backup. Then samples were briefly placed on a TurboVap® (Zymark) to eliminate organic solvent. The sample extracts were stored in nitrogen overnight until further preparation for analysis.

2.4.3 Quality Control (QC):

A number of controls were analyzed together with the tested samples. These included a pooled matrix sample produced from small volume of each tested sample (or a pool of human plasma that was previously characterized) that was used as a technical replicate during the run; extracted water samples used as blank controls; and a cocktail of quality control standards selected carefully not to interfere with the endogenous compounds. These controls were 'spiked' into every tested sample, allowing monitoring of instrument performance and aiding in chromatographic alignment. Instrument variability was assessed by computing the median relative standard deviation (RSD) for the standards that were added to each tested sample before injection into the mass spectrometers. Overall process variability was calculated by computing the median RSD for all endogenous metabolites (i.e., the non instrument ones) present in 100% of the pooled matrix samples. Tested samples were randomized throughout the platform run with quality control samples put evenly among the injections.

2.4.4 Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS):

All methods used a Waters-ACQUITY ultra performance liquid-chromatography (UPLC) and a ThermoScientific Q Exactive high-resolution accurate mass spectrometer combined with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The extracts from samples were dried then reconstituted in solvents corresponding to each of the four methods (positive early, positive late, negative and polar). Each reconstitution solvent contained a number of standards at specific concentrations to ensure consistency of injection and

chromatography. One aliquot was analyzed using acidic positive ion conditions for early positive compounds, optimized chromatographically for more hydrophilic compounds. In this “positive early” method, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analyzed using acidic positive ion conditions, however it was optimized chromatographically for more hydrophobic compounds. In this “positive late” method, the extract was gradient eluted from the same mentioned C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 µm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8 (polar compounds). The MS analysis alternated between MS and data-dependent MS^n scans using dynamic exclusion. The scan range varied slighted between methods but covered 70-1000 m/z. Raw data files are archived and extracted as described below.

2.4.5 Data Extraction and Compound Identification:

Using Metabolon’s hardware and software, raw data was extracted, peak-identified and QC processed. Metabolon’s systems are connected to a web-service platform utilizing Microsoft’s .NET technologies, which run on high-performance application servers and fiber-channel storage arrays in clusters to provide active failover and load-

balancing. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (*m/z)*, and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum. While there may be similarities between these molecules based on one of these factors, the use of all three data points can be utilized to distinguish and differentiate biochemicals. More than 3300 commercially available purified standard compounds have been acquired and registered into Laboratory Information Management System (LIMS) for analysis on all platforms for determination of their analytical characteristics. Metabolon data analysts use proprietary visualization and interpretation software to confirm the consistency of peak identification among the various samples. Library matches for each compound were checked for each sample and corrected if necessary.

2.4.6 Statistical analysis of metabolomics:

2.4.6.1 Multivariate analysis:

Batch correction of the samples was performed by Metabolon. Extracted metabolomics data were log-transformed to achieve normal distribution. Principle component analysis (PCA), an unsupervised multivariate test, was performed to get a global data view by

revealing the natural separation of samples based on maximizing the variance. PCA component's loading values were used to view the linear combination of the metabolites levels. Orthogonal partial least square discriminant analysis (OPLS-DA), a supervised multivariate regression test, was performed to differentiate components that best discriminate between predefined classes of samples whilst analyzing orthogonal components which do not differentiate between these classes. In this study, OPLS-DA was used to compare moderate versus high classes of endurance and power separately. Both PCA and OPLS-DA were run using SIMCA 14 with the default metabolite-wise sample missingness threshold of 50%. It is important to note that the purpose of the multivariate analysis performed in this study, including PCA and OPLS-DA, was purely for visualization purposes to provide an upfront global view of the data. Statistical rigor was later achieved by means of a linear regression model.

2.4.6.2 Univariate regression and enrichment analysis:

Association analysis were run using the R statistical package (version 2.14, www.r-project.org/). With general linear regression models, covariates included hemolysis levels (determined visually by Metabolon), gender, and first two PCA components that were calculated from SIMCA as explained above. Specifying power and endurance regression model, the incorporating power and endurance as a categorical variable with two levels (moderate & high) was used. Incorporating both power and endurance classification in the same model makes it possible to examine the effect of endurance class whilst correcting for power class and vice versa. This is sensible since these two classes of either sport phenotype feature a heterogeneous distribution of the other phenotype's classes as described in the study design section. A stringent Bonferroni level of

significance of p < = 0.05/743 = 6.72 E-5 was used to identify association. False discovery rate (FDR) was also performed for multiple testing correction at 5%. Function enrichment analysis was performed using the one tailed Wilcoxon sum of the ranks test. The purpose of the test is to identify sub-pathway categories (part of Metabolon annotation), whose metabolite members occupy higher ranks than can be accounted for by chance, in the list of metabolites ranked by decreasing significance of association with power/endurance sport phenotype. After exclusion of categories with less than 4 metabolites, 53 categories remained. These were tested for enrichment following association tests with endurance and power. Multiple testing correction for function enrichment analysis was also performed using FDR. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were utilized to gain further insight into the biochemistry of identified metabolites.

2.4.6.3 Statistical meta-analysis of metabolomics data

A meta-analysis was utilized to identify metabolites equally influenced by endurance, and power level in both metabolomics datasets in the current study [167, 179]. A similar linear regression model as explained above was run using R statistical package (version 2.14, www.r-project.org/) to assess association between metabolites and (moderate versus high) endurance and power classes in 2nd cohort. The model also corrected for the following possible confounders: hemolysis levels (determined visually by Metabolon), gender, and metabolites PCs as explained above. Multiple testing was Bonferroni corrected. Initially, both metabolomics data were compared and listed 691 metabolites that are similar in both data were identified. Then, values of beta, and SE.beta from the regression analysis of individual datasets were collected. Meta-analysis was performed

using “metafor” function in R (version 3.3.1). The p-values from the meta-analysis were corrected for multiple testing based on FDR correction (https://tools.carbocation.com/FDR).

Gaussian graphical modeling (GGM) was used to identify correlated metabolites based on partial correlation coefficient [180], leading potentially to unbiased reconstruction of metabolic reactions as previously reported [181]. GGM pairwise correlation were calculated using “ggm.estimate.pcor” function in R (version 3.3.1). then used Cytoscape software for visualizing complex networks and analyze human-curated pathway datasets such as KEGG with attribute data of metabolic pathways [182].

Another visualization tool called Heatmaps were generated using R (version 3.3.1) for metabolomics results. Color intensities in a heatmap image is representing the relative richness of metabolites detected in each sample [183]. In this study heatmaps are representing significant metabolic differences between low/moderate versus high endurance, power and CVD group of athletes.

2.5 mGWAS analysis

Profiling of serum metabolites and genotyping of 490 elite athletes was conducted as explained previously. Table 5 summarizes classification of 490 athletes samples used in mGWAS analysis.

Table 5. Distribution of 490 samples used in mGWAS into three phenotypes: endurance (columns), power (rows) and cardiovascular demand (colored) based on their sports discipline.

mGWAS		Endurance	
		Moderate (40-70% VO2max)	**High (>70% VO2max)**
Power	**High (>50% MVC)**	Wrestling (3M), Judo (3M)	Boxing (1M/16F), Heptathlon (1M), Rowing (6M/7F), Cycling (31M/4F)
	Moderate (20-50% MVC)	Athletics (15M/22F), Rugby (16M), Triple Jump (1M)	Athletics 200-800m (4M), Hockey (1F), Skiing Cross Country (1M), Basketball (3M), Swimming (22M/16F)
	Low (<20% MVC)	Baseball (2M), Volleyball (1M)	Tennis (1M/1F), Soccer (315M), Athletics 1500-3000m (3M)

2.5.1 Metabolomics and genomics data extraction

Illumina iScan system (Illumina, cat. # SY-101-1001) was used for genotyping raw data extraction and QC process. SNP identification was obtained from GenomeStudio that was explained previously (see section 1.1.3 Data Extraction and SNP Identification).

Metabolites data were received from Metabolon after their analysts performed the confirmation step related to the consistency of peak identification among the various samples as explained previously (see section 1.2.5 Data Extraction and Compound Identification).

2.5.2 Statistical analysis of combined metabolomics and genomics data

Statistical analysis for genotyping data of 490 samples was performed using Plink as explained previously in GWAS analysis. SNP exclusion QC filters were adopted: genotype call rate < 98% (130526 SNPs were excluded), MAF < 0.01 (70210 SNPs were

excluded) and Hardy Weinberg p value < 10-6 (976 SNPs were excluded), resulting in 275016 SNPs (Bonferroni significance= [0.05/(275016x751)] = 2.4E-10 was used for the analysis.

Following metabolomics of 490 samples, 751 metabolites were taken into account following the previously described QC filters in the metabolomics analysis. Metabolites data were log scaled and z-score normalized. Outliers were detected and replaced by missing values. Metabolites with more than 50% missing values were excluded.

Associations between SNPs and metabolite levels were computed using lm function in R (version 3.3.1) while correcting for gender, hemolysis and population stratification based on top two PCs from genotype data using plink version 1.9. An additive inheritance model was used (SNPs were coded as 0,1,2 according to their genotype group).

An average inflation factor was calculated for mGWAS metabolites using chi-squared statistic for the null markers divided by the expected median value of the chi-squared statistic (chi^2=beta*beta/(sebeta*sebeta) and lambda=median(chi^2)/0.45), providing the "inflation factor," lambda [184]. If observed average inflation factor is less than or equal to 1, no adjustment is required, but if is above 1.1 then method for correcting for population stratification should be considered [176, 184]. Principal-components analysis can control inflation as covariates in a regression model [184], hence principal-components were included in mGWAS regression model as explained above. However, studies indicated that phenotype-associated markers can cause a slight increase in the inflation factor [176].

Percent of explained variance (r^2) was calculated with formula: $r^2 = X^2/(N - 2 + X^2)$, where N is the number of samples and X^2 = (Beta/standard error of the beta)^2, values obtained from the regression analysis [165]. Percent of explained variance used to measure proportion of how well the variation of metabolite explains the variation of corresponding gene in the identified mQTLs in elite athletes.

To display significant SNPs, manhattan and box plots were generated using R (version 3.3.1). The boxplots are quick graphical examination of resulted mQTLs that comparing distributions of samples between SNP alleles associated with metabolites and genes.

Regional association plots were used to show the strength of mGWAS calculated SNP-metabolite associations. A sentinel SNP (lead SNP) or sentinel metabolite (lead metabolite) show the most significant SNP-metabolite association within the locus. Thus, mQTL locus is defined as 500Kb region around the sentinel SNP and is named according to the gene present in the locus. Regional association plots were generated using SNIPA (grch37-1kgpp3v5, eur; http://snipa.helmholtz-muenchen.de/snipa/).

2.5.3 mQTLS associated with Endurance, power and CVD

To determine mGWAS associated with endurance, power and CVD sports, a list of significant results from GWAS and metabolomics were considered. Significant metabolites associated with endurance, power and CVD metabolites were identified within the list of mQTLs from the mGWAS analysis. The following Bonferroni p-values were used to report a significant association:

- Endurance: [0.05 / (104x275016)] =1.7E-9 (104 was obtained from the meta-analysis)

- Power: [0.05 / (207x275016)] =8.9E-10 (207 was obtained from the 1st cohort only)
- CVD: [0.05 / (112x275016)] =1.6E-9 (112 was obtained from second cohort only)

Box plots were generated using R (version 3.3.1) to display distributions of samples among SNP alleles associated with metabolites and genes. Regional association plots were used to show the strength of mGWAS calculated for significant SNP-metabolite associations. Regional association plots were generated as previously described using SNIPA (grch37-1kgpp3v5, eur; http://snipa.helmholtz-muenchen.de/snipa/).

2.5.4 mQTLS associated with GWAS significant SNPs Endurance, power and CVD

GWAS significant SNPs associated with endurance were recognized from the mGWAS analysis to perform functional validation of the genes by identifying metabolites that could explain the functional relevance of the association with endurance athletes. Regional association plots were used to show the strength of GWAS calculated SNP-Gene associations. Thus, SNP locus is defined as 500Kb region around the significant SNP and is named according to the gene present in the locus. Regional association plots were generated using SNIPA (grch37-1kgpp3v5, eur; http://snipa.helmholtz-muenchen.de/snipa/). Pathway enrichment analyses were carried out using Chi square tests (https://www.mathsisfun.com/data/chi-square-calculator.html) to identify pathways with enriched metabolites ranked by p-value from the linear regression model. Boxplots were generated using R (version 3.3.1) to display distributions of samples between SNP alleles associated with metabolites and genes.

CHAPTER 3
Results: Genome Wide Association study (GWAS)

3 (Chapter 3) Results: Genome Wide Association study

3.1 Introduction

With the development of high-density SNP chip-based arrays, GWAS present a powerful tool for detecting associations between genetic variants and a phenotype of interest in a group of individuals [185]. It offers a valuable first insight into humans' trait validation by considering biological and statistical associations [186]. GWAS presents the advantages of dealing with population stratification, and effectively adjusting for common variants to avoid spurious results of no association [187]. When considering genetic associations with physical performance related phenotypes, limited progress has been published due to small size of elite atheltes cohorts and complixity of phenotypes. Most published work has adopted traditional candidate-gene approach [113, 188]. Classical twin and family genetic studies have suggested that VO_2max is up to 50% inherited [85, 189]. Case-control studies have revealed a number of SNPs associated with various aspects of elite athletic performance (Table 6). GWAS in athletes versus non-athletes have uncovered many new loci in association with VO_2max [83, 91] and elite endurance performance [108]. A more recent review of genetic predisposition to elite athletic performance has highlighted 93 endurance variants and 62 power variants [107]. Table 6 provides a summary of SNPs associated with physical performance identified from case-control and GWAS analysis. A large meta-analysis of 1520 endurance athletes and 2760 controls has revealed no evidence of association of a common genetic variation with alite athlete's status in world class athletes [115]. Therefore, the genetic predisposition to endurance or power traits remains unclear, largely due to the relatively underpowered elite athletes' cohorts.

To investigate the influence of genes on human physical performance, GWAS need to be carried out with large sample size of elite athletes using high number of gene variants that may explain differences in physical capabilities and training-induced effects between subjects. Therefore, GWAS represent a productive way that could explain genetic predisposition of athletic performance in relation to various sports-related phenotypes.

Table 6. A summary of identified DNA polymorphisms associated with physical performance in more than one study in different ethnicities, although some were not reproduced in at least one study, adopted from [190].

	GENE	SNP	SPORT DESCIPLINE	FUNCTIONAL RELEVANCE	REFERENCES
1	Angiotensin I-converting enzyme	*ACE* 287 bp Alu sequence insertion fragment	Endurance: Elite mountaineers, 5000 meter distance runners, rowers, cyclists, cross-country skiers, basketball, triathletes, and handball players	Increase synthesis of the vasodilator kinins that elevates oxygen flow in the blood	[87], [191], [192] [193-209]
2	Adrenergic receptor	*ADRA2A* 6.7-kb, *ADRB1* 49Gly, *ADRB2* 16Arg, and *ADRB3* 64Arg Alleles	Elite endurance athletes	Modulate the physiological impacts of norepinephrine/ epinephrine, therefore affecting cardiovascular responses	[210-220]
3	Angiotensin II type 2 receptor	*AGTR2* rs11091046 C Allele	Endurance athletes	Increase growth and differentiation of slow-twitch fibers in skeletal muscle with increased in oxygen consumption	[221]

4	Aquaporins	*AQP1* rs1049305 C Allele	Elite endurance athletes: Marathon runners and triathletes	Regulate osmotic reabsorption by promoting water transfer from the blood into the muscle.	[222, 223]
5	Adenosine monophosphate deaminase 1	*AMPD1* Gln12 Allele	Endurance athletes: Cyclists, runners and rowers	Prematurely terminate translation of catalyst and deaminate adenosine monophosphate to inosine monophosphate in skeletal muscle.	[224-228]
6	Bradykinin B2 receptor	*BDKRB2* _9 and rs1799722 T Alleles	Endurance athletes: 200, 400–3000, and 5000 m runners, and Triathletes	Increase efficiency of muscular contraction by endothelium-dependent vasodilator.	[229-231], [216]
7	Protein phosphatase 3	Calcineurin/NFAT-Related Genetic Markers (*NFATC4* Gly160, *PPP3CA* rs3804358 C, *PPP3CB* rs3763679 C, and *PPP3R1* 5I Alleles)	Elite endurance athlete: Runners	Regulate skeletal muscle differentiation, hypertrophyand fiber-type by Calcineurin–NFAT signaling pathway that leads to different cardiac and skeletal muscle phenotypes.	[232-238], [239]
8	The muscle isoform of creatine kinase	*CKM* rs8111989 A Allele	Elite endurance athlete: Marathon runners and cyclists	Increase aerobic performance and reduce fatigability after long-term physical activity	[240-244]
9	Collagens	Collagen-Related Genetic Markers (*COL5A1* rs12722 T, *COL5A1* rs71746744 *AGGG*, and *COL6A1* rs35796750 T Alleles)	Elite endurance athletes: Ultramarathon runners and triathletes	Associated with energy storage, inflexibility of muscles and reduction in the need for muscle-stabilizing activity	[245-251]
10	Endothelial PAS domain protein 1	*EPAS1* rs1867785 G and rs11689011 T Alleles	Endurance athletes: Cyclists, swimmers (100–800 m), middle-	Regulate hypoxia-inducible transcription factor, catecholamine and	[252, 253]

			distance runners, triathletes, and rowers	mitochondrial homeostasis, also associated with regulation of cardiac output and erythropoietin.	
11	GA-binding protein β subunit 1	*GABPB1* rs12594956 A, rs8031031 T, and rs7181866 G Alleles	Elite endurance athletes: Runners, rowers and sprinters	Stimulate mitochondrial biogenesis	[254-257]
12	Heterotrimeric guanine nucleotide-binding proteins (G proteins)	*GNB3* rs5443 T Allele	Elite endurance athletes: Long-distance runners and sprinters	Increase oxygen consumption	[258-262]
13	Hemochromatosis	*HFE* 63Asp Allele	Elite endurance athletes: cyclists, and runners	Regulate iron absorption by increasing the interaction of the transferrin receptor with transferrin.	[263-265]
14	Hypoxia-inducible factor-1α	*HIF1A* Pro582 Allele	Elite endurance athletes	Increase VO_2max through aerobic exercise training at age 55 to 65 years	[90, 266-268] [239]
15	IL-15 receptor α	*IL15RA* rs2228059 A	Elite endurance athlete: Cyclists (A allele), triathletes and rowers (C allele)	Increase whole muscle baseline and cortical bone volumes	[269]
16	Potassium inwardly rectifying channel, subfamily J, member 11	*KCNJ11* Glu23 Allele	Endurance-oriented athletes, and marathon runners	Increase VO_2max and maximal minute ventilation	[270]
17	Monocarboxylate transporters	*MCT1* (*SLC16A1*) rs1049434 A Allele	endurance-oriented athletes and rowers	Associated with lactate production by white muscle fibers	[271-275]

18	mtDNA Markers	mtDNA Markers	Elite endurance athletes: Sprinters, cyclists and runners	Regulate energy metabolism through 36 molecules of ATP per glucose molecule in contrast to the two ATP molecules produced by glycolysis.	[276, 277] [276, 278-283]
19	Nuclear factor I A	*NFIA-AS2* rs157231 C Allele	Endurance athletes	Increase VO_2max by activating erythropoiesis that enhances haemoglobin, reticulocytes and erythrocytes production.	[284], [108],
20	Endothelial nitric oxide synthase	*NOS3* Glu298, 164-bp, 4B and rs2070744 T Alleles	Elite endurance athletes: triathletes, rowers, and underwater finswimmers	Regulate NO generation in blood vessels and associated with vascular function as vasodilator	[285],[286, 287], [231]
21	Peroxisome proliferator-activated receptor α	*PPARA* rs4253778 G Allele	Endurance-oriented athletes, sprinters and rowers	Modulate lipid, glucose, and energy homeostasis and regulate body weight and vascular inflammation.	[101, 288-291], [292]
22	Peroxisome proliferator-activated receptor δ	*PPARD* rs2016520 C and rs1053049 T Alleles	Endurance athletes	PPARδ agonist GW1516 is in the WADA prohibition list because it increases the exercise tolerance and significantly linked with an increased muscle glucose uptake.	[293-297]
23	Peroxisome proliferator-activated receptor γ (PPARγ) coactivator 1α	*PPARGC1A* Gly482 and rs4697425 A Alleles	Endurance-oriented athletes, and rowers	Decrease expression of PPARGC1A in mitochondrial biogenesis, fatty acid oxidation, glucose utilization, thermogenesis, angiogenesis, and muscle fiber-type conversion toward slow-twitch type I fibers.	[298-302] [292], [239],[283]

24	PPARγ coactivator 1 β	*PPARGC1B* 203Pro and 292Ser Alleles	Elite endurance athletes	Increase insulin-stimulated glucose metabolism, and protect muscle against an age-related decline in PGC1β expression	[303-305], [239]
25	RNA-binding protein, fox-1 homolog	*RBFOX1* rs7191721 G Allele	Middle and short endurance	Associated with tissue-specific alternative splicing in heart, muscle, and neuronal tissues	[306], [108]
26	Solute carrier family 2 (Mediated glucose transporter), member 4	*SLC2A4* rs5418 A Allele	Long-distance runners	Mediate glucose metabolism in the body by limiting muscle glucose uptake under most conditions	[307]
27	TFAMprotein expression	*TFAM* 12Thr Allele	Elite endurance athletes	Modulate mtDNA transcription, replication, and maintenance.	[308] [239] [283]
28	Thyroid stimulating hormone receptor	*TSHR* rs7144481 C Allele	Long endurance athletes	Regulate thyrothropin impacts on angiogenesis by cAMP mammalian target of rapamycin signaling.	[108]
29	Uncoupling proteins 2,	*UCP2* 55Val Allele	Elite endurance athletes and rowers (55Val allele). power-oriented athletes (Ala55 allele)	Enhance metabolic efficiency by increasing body mass index, physical activity, and energy generation with higher VO_2max	[309, 310] [311] [239]
30	Uncoupling proteins 3	*UCP3* rs1800849 T Allele	Elite endurance athletes and rowers	Linked to markers of energy metabolism and aerobic capacity	[312-314] [239]

31	vascular endothelial growth factor	*VEGFA* rs2010963 C Allele	Elite endurance athletes and rowers	Regulate VEGF protein expression in human myoblasts and increase oxygen uptake before and after aerobic exercise training,	[239, 315]
32	VEGF receptor 2	*VEGFR2* 472Gln Allele	Elite endurance athletes, all-round speed skaters and rower	Regulate the full spectrum of VEGF angiogenic responses by kinase insert domain receptor in aerobic exercise	[316, 317]
33	Y-chromosome haplogroups E*, E3*, and K*(xP)	Y-Chromosomal Haplogroups	Endurance running	Specific haplogroups of the Y-chromosome have been associated in the Ethiopian athletes.	[318]
34	Circulating angiotensin I-converting enzyme	*ACE* D Allele	Power-oriented athletes, short-and-middle-distance swimmers and sprinters	Increase muscle volume and strength by increasing fast-twitch muscle fibers	[211, 319-327], [191], [205], [328], [195], [193], [311]
35	α-actininin-3	*ACTN3* Arg577 Allele	Power-oriented athletes, artistic gymnasts, sprinters, swimmers, short-distance skaters and soccer	Restricted to fast muscle fibers responsible for generating force at high velocity.	[88, 319, 329-349], [277], [195], [311], [216], [321]
36	Angiotensinogen	*AGT* 235Thr Allele	Power athletes	Modulate vascular resistance and sodium homeostasis, consequently regulate blood pressure.	[350-352], [209]
37	Adenosine monophosphate deaminase	*AMPD1* Gln12 Allele	Power oriented athletes: short-distance runners, short-distance	Reduces skeletal muscle AMPD activities	[353-355], [224]

			swimmers, sprinter, boxing, wrestling, speed skating (500–1500 m), and weightlifters		
38	cAMP-responsive element modulator	*CREM* rs1531550 A Allele	Elite sprinters	A bZIP transcription factor that binds to the cAMP-responsive element found in many viral and cellular promoters.	[356]
39	Dystrophin	*DMD* rs939787 T Allele	Strength/power athletes	Dystrophin RNA is part of dystrophin–glycoprotein complex associated with the inner surface of muscle fibers, generating a large set of protein isoforms.	[357]
40	5,10-methylenetetrahydrofolate reductase	Folate Pathway Genetic Markers (*MTHFR* rs1801131 C, *MTR* rs1805087 G, and *MTRR* rs1801394 G Alleles)	Power/strength athlete	DNA hypomethylation leads to myogenic differentiation, and increases muscle mass, and energy metabolism.	[358-360]
41	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 13 protein	*GALNT13* rs10196189 G Allele	Elite sprinters	Initiate O-linked glycosylation of mucins by the initial transfer of N-acetylgalactosamine with an alphalinkage to a serine or threonine residue and thus catalyzes the initial reaction in O-linked oligosaccharide biosynthesis.	[356]
42	HIF-1α	*HIF1A* 582Ser Allele	Power-orientated athletes, all-round speed skaters, weightlifters, and wrestlers	Increase HIF-1α protein stability and transcriptional activity by converting proline to serine in the amino	[361-364]

				acid sequence of the protein, which improve glucose metabolism	
43	Insulin-like growth factor 1	*IGF1* rs35767 T and *IGF1R* rs1464430 C Alleles	Power athletes	Increase IGF-1 impacts in skeletal muscle through transmembrane receptor	[365, 366]
44	Interleukin-1 receptor antagonist	*IL1RN**2 Allele	Volleyball, soccer, rugby, triathlon, basketball, martial arts, track-and-field sports, running, handball, and swimming	Regulate skeletal muscle inflammatory and repair reactions during and after exercise and increase adaptation to high-intensity exercise.	[367, 368]
45	interleukin-6	*IL-6* rs1800795 G Allele	Power-oriented athletes	Modulate glucose homeostasis during exercise and mediate hypertrophic muscle growth	[369-373]
46	endothelial nitric oxide synthase gene	*NOS3* rs2070744 T Allele	Power-oriented athletes: jumpers, throwers, and sprinters	Improve muscle contraction during knee-extensor exercise and increase tolerance to high-intensity exercise in humans	[374], [362], [285], [311]
47	Peroxisome proliferator-activated receptor α	*PPARA* rs4253778 C Allele	Power-oriented athletes and football players	In response to training, LV mass increases with the increase in hypertrophic effect that influences cardiac and skeletal muscle substrate consumption	[101, 288, 289], [375-377]
48	PPARγ	*PPARG* 12Ala Allele	Power-oriented athletes	Reduce receptor activity and increase insulin sensitivity and skeletal muscle glucose uptake with better cross-sectional area of muscle fibers	[362, 378-381]

49	Manganese superoxide dismutase	*SOD2* Ala16 Allele	Power athlete: 100–200 m sprinters and long jumpers	Convert anion superoxide of the mitochondria into hydrogen peroxide and oxygen. Consequently after racing, creatine kinase is increased, suggesting lower muscle damage.	[273, 382-385]
50	Vitamin D receptor	*VDR* rs10735810 T Allele	Medium-high-level male soccer	Impact on bone and skeletal muscle biology through binding of vitamin D and inhibiting parathyroid hormone production that sustain normocalcemia.	[386-388]

The aim of this chapter was to identify genetic predisposition into elite athletic performance by genotyping a large cohort of elite athletes who belong to different types of sports. Genotyping was performed using a unique SNP chip that covers various enzymes and metabolic pathways with relevance to physical performance. The study findings were validated in an independent replication cohort of elite Russian athletes and matched controls.

3.2 Material and method

3.2.1 Classification of study participants for Statistical analysis of GWAS data

Athletes were classified according to their VO_2max, MVC and cardiovascular demand as shown previously in chapter 2 Table 3, following Mitchell's previously published sports classification criteria [16]. Table 5 further lists the number of participants based on various analyses as per sport type in each class/group and their genders.

3.2.2 Data Extraction and SNP Identification:

SNP array genotyping was performed in 1259 athletes' samples using a new customized 24 BeadChips array. The Illumina Drug Core array-24 BeadChips developed by Illumina in collaboration with experts in translational genomics and computational chemical biology. This array contains 240,000 highly-informative genome-wide tag SNPs and a novel 200,000 custom marker set designed to support studies of metabolomics, drug target validation, and treatment response.

The genotypes were extracted from the raw intensities using GenomeStudio (Illumina). The genotyping data (476728 SNPs from each sample) were processed to quality control

steps using Plink v1.90b4.6 64-bit to identify candidate gene regions through regression analyses.

3.2.3 Statistical analysis of genomics data

Following genotyping using Illumina's DrugCore SNP array, PCA was conducted to investigate whether genotypes were influenced by individual sport types (Figure 14) or gender (Figure 15). PCA confirmed no evidence of such influence as genotypes did not cluster over any of the identified PCs. Further PC analyses were applied on the study groups using cardiovascular demand of their respective sport rankings (Figure 16) [16]. The latter included low/medium (2), medium (3), high/medium (4) and high (5) as shown in Figure 15. Rank 2 and 3 were then grouped as moderate CVD vs. rank 4 and 5 as high CVD. Study groups were used in analysis including low/moderate vs. high groups of CVD, endurance and power as shown in Figure 17.

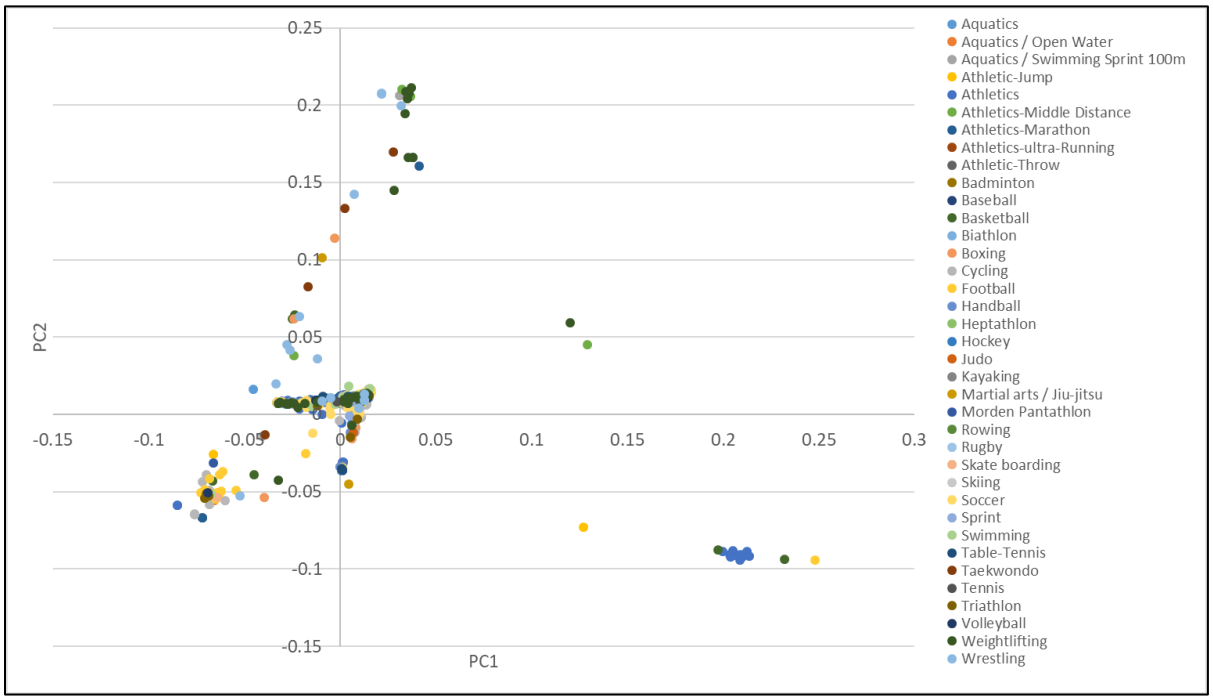


Figure 14. PCA shows no difference in genotype distribution among sports types.

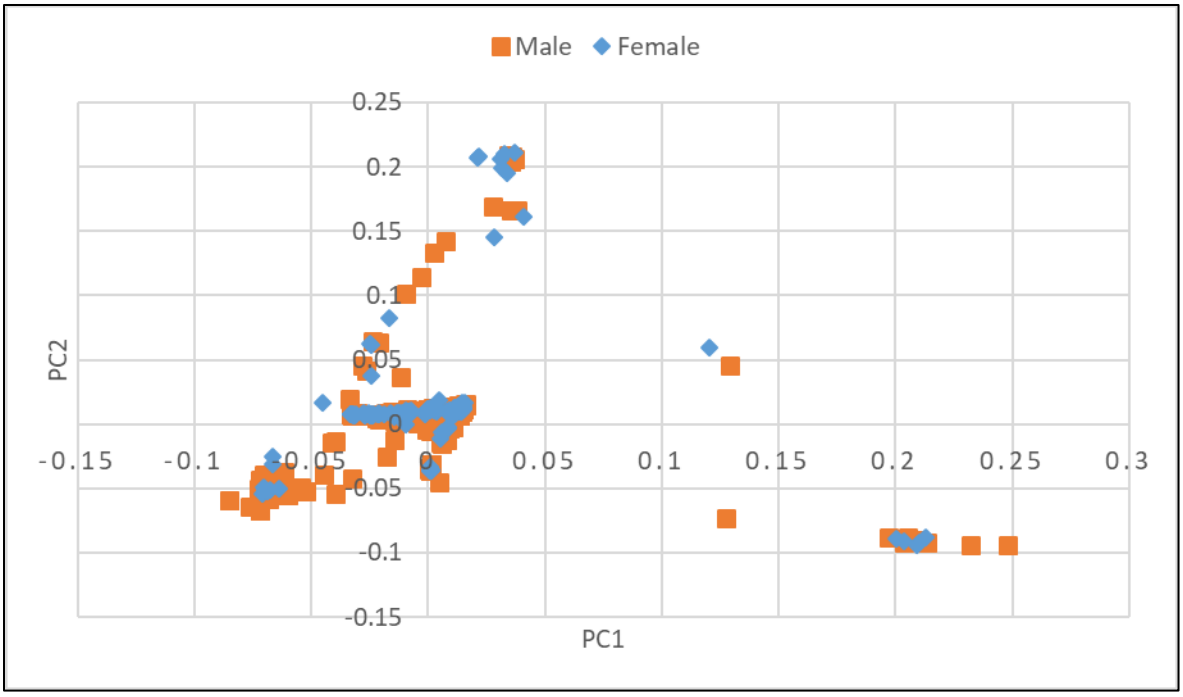


Figure 15. PCA shows no difference in genotype distribution between gender.

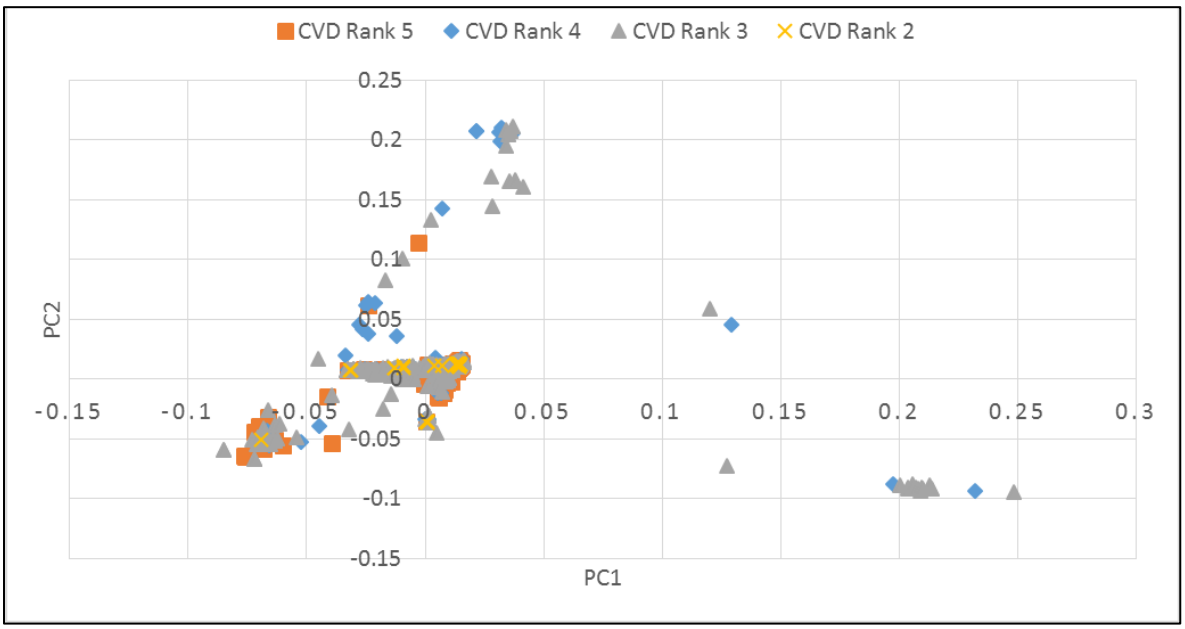


Figure 16. PCA shows no difference in genotype distribution among sport ranks according to cardiovascular demand groups.

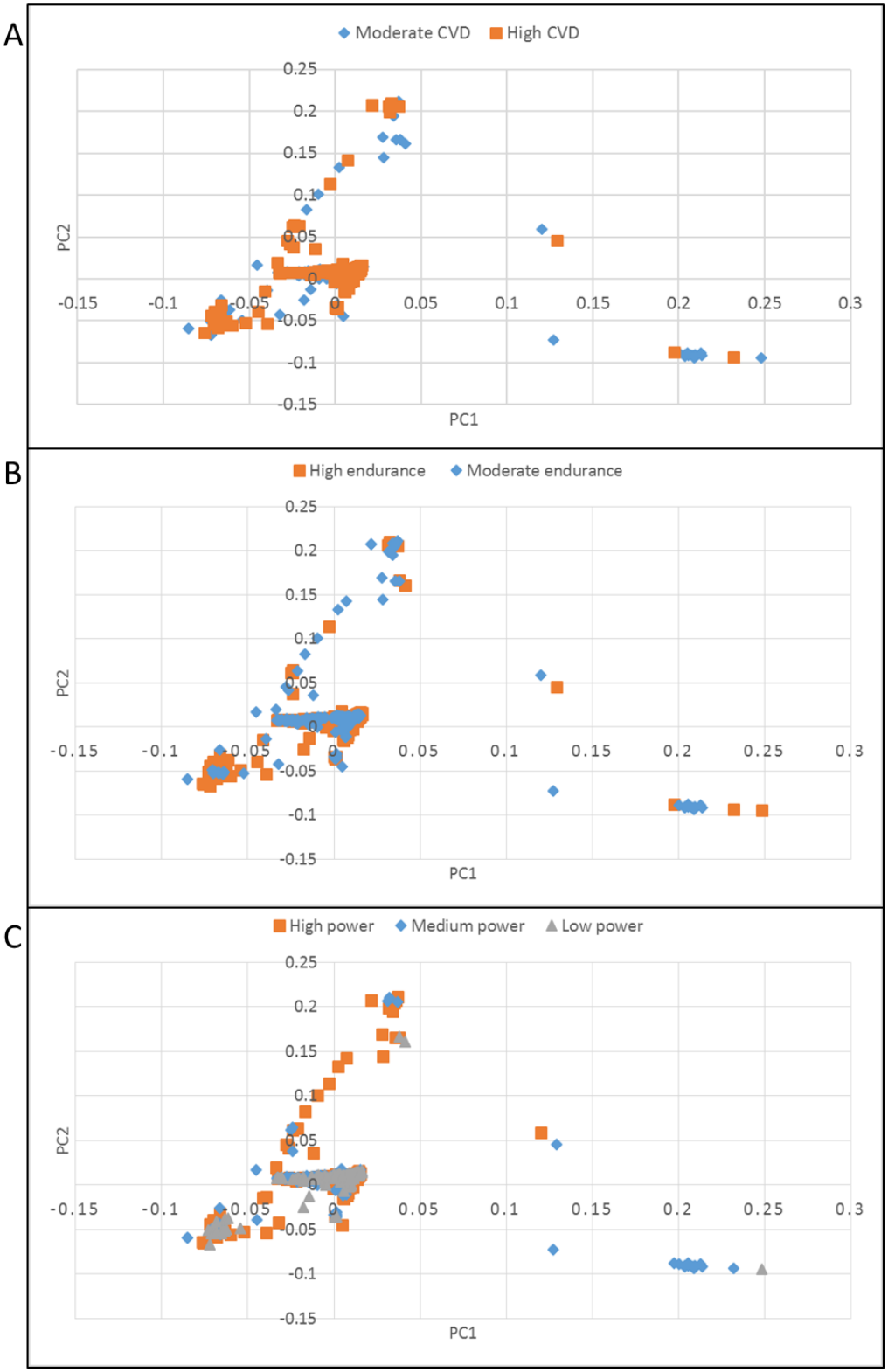


Figure 17. PCA shows no difference in genotype distribution among high vs. moderate: A. CVD, B. endurance, and C. power athletic groups.

Some clusters of samples were observed in the PCA, which probably reflect ethnic diversity. To predict ethnicities of athletes' samples, genotyped data were merged with 1000 Genome project data. PCA were conducted for merged data and then ethnicities of athletes were predicted. Figure 18 reveals clusters of ethnicities, and figure 19 indicates that the majority of the samples were from European athletes.

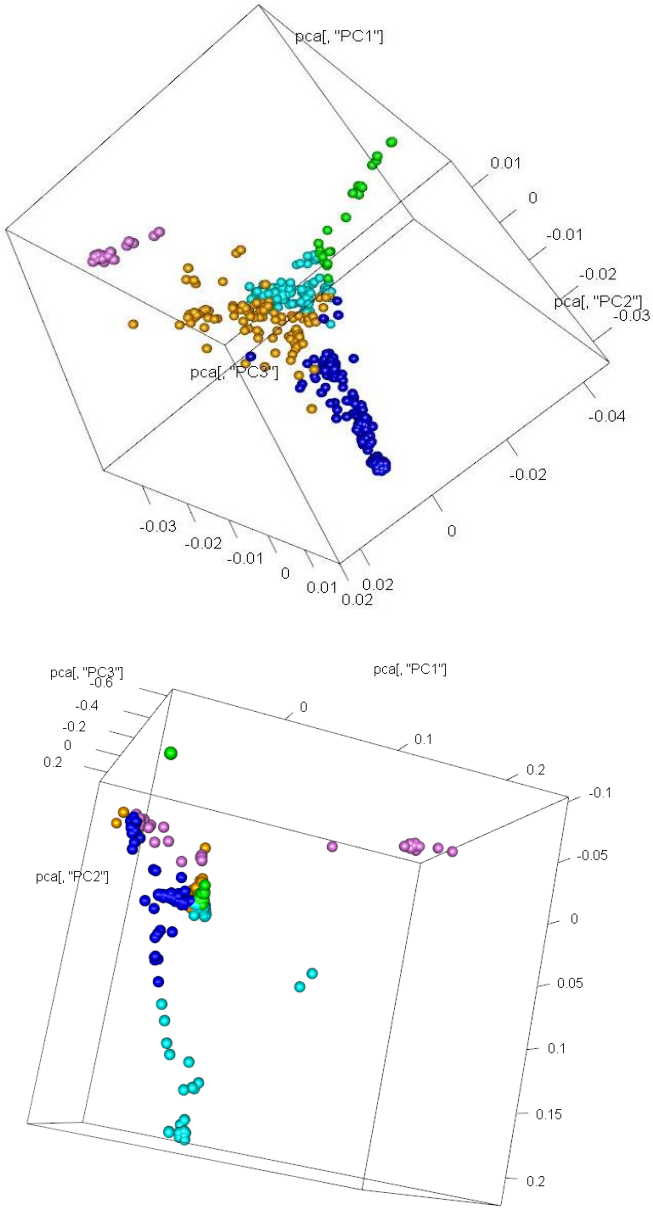


Figure 18. PCA revealed clusters of samples reflecting different ethnicities: American, orange; African, blue; South Asian, green; East Asian, violet, European, cyan.

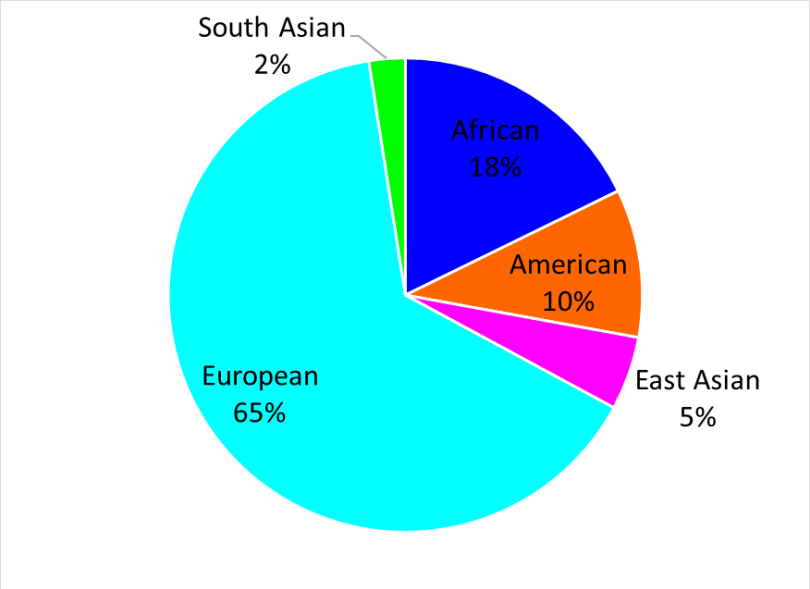


Figure 19. Pie chart showing the percentages of samples collected from different ethnicities as per mapping against PCA generated clusters in 1000 genome project.

Since PCA analysis showed that there is population stratification in genetic data that requires the execution of several quality checks and careful conductance of statistical analyses to avoid spurious associations due to several potential sources of confounding (e.g., ethnic stratification) that lead to false positive associations and/or mask true associations, because allele frequencies can differ between subpopulations [389]. Accordingly, we applied the rigorous quality control (QC) procedures on genotype data prior to conducting GWAS, including the use of appropriate methods to take into account ethnic heterogeneity [389]. Since most athletes included in the study cohort were of European origin, genetic association studies were performed on 753 European athletes' samples to minimize the effect of ethnicities. Table 7 summarizes the distribution of European elite athletes' samples into endurance, power and cardiovascular demand (colored) groups, following published criteria [16]. Figure 20 reveals clusters of European

athletes, but not pertaining to any of the study groups discussed above, suggesting no significant influence of ethnicity on sports classes used in this analysis.

Table 7. Classification of GWAS study participants. Elite athletes are distributed in defined categories, as described previously [16]. These categories were based on VO_2Max, MVC, and CVD associated with their respective sports types. The number of participants and gender (M for males and F for females) in each class are mentioned.

GWAS		Endurance	
		Moderate (<70% VO2max)	**High (>70% VO2max)**
Power	**High (>50% MVC)**	Weight-Lifting (8M/7F),Judo (2M),Wrestling (4M),Skate boarding (2M)	Kayaking (1F),Morden Pantathlon (1F),Rowing (9M/8F),Biathlon (2M/1F),Boxing (3M/12F),Cycling (149M/48F),Triathlon (9M/7F)
	Moderate (20-50% MVC)	Athletic-Jump (1F),Rugby (15M),Aquatics (8M/5F),Athletics (31M/29F)	Handball (15M/3F),Athletics middle distance (1M),Basketball (2M),Hockey (1F),Skiing Cross Country (3M/1F),Swimming (24M/17F)
	Low (<20% MVC)	Baseball (2M),Volleyball (2M),Table-Tennis (7M)	Athletic-Long-Distance-Marathon (35M/12F),Tennis (2M/3F),Soccer (243M),Athletics-Ultra-Running (1F),Football (16M/1F)

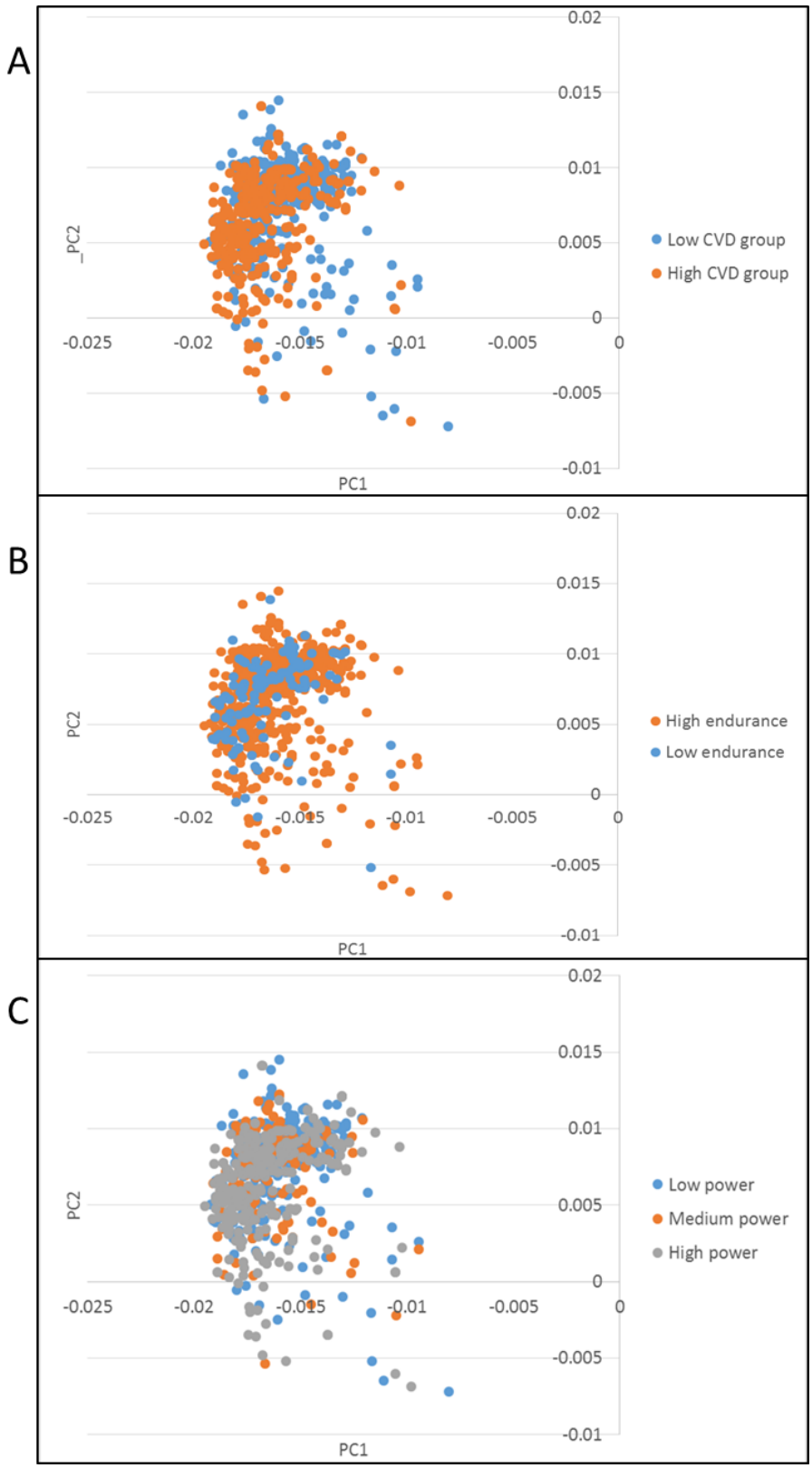


Figure 20. PCA shows no difference in genotype distribution between high vs. moderate A. CVD, B. endurance, and C. power athletic groups.

The top principal components from PCA were used in the following association analysis for correcting for any remaining population stratification. Genome wide association study was performed using Plink v1.9 Quality control measures were applied to the genotype data set to exclude samples with low genotype call rate (<95%) or excess heterozygosity. SNPs with genotype call rate < 98%, minor allele frequency < 1%, or data deviating from Hardy-Weinberg equilibrium (P < 1E-6) were excluded. Accordingly, out of 476728 SNPs, only 341385 SNPs were analyzed. A stringent Bonferroni level of significance of p < = 0.05/341385 = 1.46E-7 was used to define significant associations. The analysis was performed using linear or logistic regression models adjusting for sex and the first four top principal components. Resulted minor allele frequencies (MAF) of our study groups were compared against non-athlete group (1000 Genome project) and Trans-Omics for Precision Medicine (TOPMed) projects that contributed to this initiative by integrating whole-genome sequencing (WGS) with their clinical data that were obtained from: https://www.ncbi.nlm.nih.gov/snp/?term=rs56330321. Ingenuity pathway analysis (IPA), Qiagen, cat. # 830011) was also reported to look for functional relevance of identified SNPs.

3.2.4 Replication analysis of genomics data

The Russian athletes' study involved 43 sprinters, 120 middle-distance athletes, 56 long-distance athletes (95 females, age 21.9 (3.5) years, 124 males, age 22.1 (4.2) years). Sprinters included 8 (100-400m) runners, 5 sprint cyclers, 10 (500-1000m) speed skaters / short trackers, 19 (50-100m) swimmers, 1 (200m) kayaker. Middle-distance athletes comprised 59 rowers, 10 (0.8-1.5km) runners, 7 middle-distance cyclers, 21 middle-distance kayakers / canoers, 15 (1.5-3.0km) speed skaters, 8 (200-400m) swimmers.

Long-distance athletes included 3 (3-10km) runners, 1 marathon runner, 14 biathletes, 12 cross-country skiers, 14 (0,8-25km) swimmers, 6 triathletes, 6 race walkers. All athletes were Olympic team members (International level) who have tested negative for doping substances.

Russian controls were 173 (126 males and 47 females) unrelated individuals with no competitive sporting experience (all Eastern European descent Caucasians). The protocol was approved by the Ethics Committee of the Federal Research and Clinical Center of Physical-chemical Medicine of the Federal Medical and Biological Agency of Russia. Written informed consents were obtained from participants. The study complied with the guidelines of the Declaration of Helsinki and ethical standards in sport and exercise science research. The experimental procedures were performed according to the set of guiding principles for results reporting typical of genetic association studies as defined by the Strengthening the REporting of Genetic Association studies (STREGA) Statement. DNA extraction was performed using a commercial kit according to the manufacturer's instructions (Technoclon, Russia). HumanOmni1-Quad BeadChips (Illumina Inc, USA) were utilized for genotyping of 1,140,419 SNPs in athletes and controls. Genotyping was performed according to the instructions of the Infinium HD Assay. The analysis of Russian cohort data was performed using linear regression models incorporating sports grouped by training (i.e. sports with high vs. low/moderate aerobic component). To perform the meta-analysis, the Cochrane Review Manager (RevMan) version 5.3 was used. Random and fixed effect models were applied.

3.3 Results

3.3.1 Comparing high endurance performers against low/moderate endurance performers.

Genotype distribution of SNPs was compared between high endurance compared to moderate endurance group. Several variants were found associated with high endurance (Table 8) by logistic model regression analysis after correcting for gender, and population stratification using PC1, PC2, PC3, and PC4. One novel SNP (rs56330321 in ATPase Plasma Membrane Ca2+ Transporting 2 gene, *ATP2B2*) reached Bonferroni level of significance (p=1.47E-7) and another novel SNP (rs2635438 in Spectrin Repeat Containing Nuclear Envelope Protein 1 gene, *SYNE1*) reached FDR level of significance at 5% level of significance (p=2.54E-7). Validation of significant endurance associated variants was performed by comparing the frequencies of the most significant SNPs (with P < E-5 - E-8) in 219 elite Russian athletes and 173 Russian controls. The rs56330321 A allele was under-represented in Russian middle-distance athletes (n=120) compared to 173 Russian controls (0.8 vs 3.8%; OR=0.2036; SE=0.6435; P=0.013). The rs2635438 G allele was under-represented in 56 elite Russian long-distance athletes compared to 43 elite Russian sprinters (3.6 vs 8.1%; OR=0.132; SE=0.9004; P=0.024). A subsequent Meta-analysis has confirmed the association of rs56330321 and rs2635438 with endurance athlete status at GWAS level of significance (5.13E-09 and 1.91E-08, respectively). Table 8 shows top 10 SNPs with their odd ratios (OR), location according to function genome variation server (GVS), gene name and MAF in high and low endurance groups. MAF in non-elite athletes from 1000 Genome project were used as a reference. Manhattan and quartile-quartile (QQ) plot of GWAS hits associated with endurance are shown in Figure 21.

Table 8. Top GWAS SNPs associated with Endurance (n=753, High= 639, Low= 114). Abbreviations: single nucleotide polymorphism (SNP), chromosome (Chr), Position (BP), Reference Allele (A), number of samples (N), odds ratio (OR), standard error (SE), probability value (Pval), function genome variation server (GVS), and minor allele frequencies (MAF).

SNP	Chr.	BP	A	N	OR	SE	Pval	FDR	Significant using an FDR of 0.05?	Function GVS	Nearest Gene	MAF-High Endurance N=639	MAF-Low Endurance N=114	MAF-nonathletes
rs56330321	3	10684812	A	751	0.2501	0.26	1.15E-07	0.037	Yes	intron	*ATP2B2*	0.038	0.12	A=0.055
rs2635438	6	152506470	G	753	0.1636	0.35	2.54E-07	0.041	Yes	intron	*SYNE1*	0.015	0.08	C=0.06
rs225902	14	30459451	A	748	0.41	0.19	2.48E-06	0.268	No	intergenic	none	0.14	0.24	A=0.18
rs146654270	9	123798492	A	749	0.1621	0.39	3.38E-06	0.274	No	intron	*C5*	0.013	0.057	A=0.01655
rs1969772	1	59205102	A	752	0.4928	0.15	4.23E-06	0.274	No	intergenic	none	0.29	0.44	A=0.26
rs77471963	15	69245158	G	746	0.3671	0.29	4.38E-06	0.236	No	intron	*NOX5*	0.08	0.16	G=0.068
rs10011584	4	127213903	T	750	0.4572	0.17	5.41E-06	0.250	No	intergenic	none	0.15	0.26	T=0.13
rs142155779	7	126252753	C	751	0.2247	0.33	5.47E-06	0.221	No	intron	*GRM8*	0.021	0.082	C=0.015
rs7584904	2	106131036	T	750	0.4047	0.20	6.30E-06	0.227	No	intergenic	none	0.09	0.18	T=0.14
rs7599151	2	59308101	A	753	0.5097	0.15	7.75E-06	0.251	No	intergenic	none	0.36	0.46	A=0.33

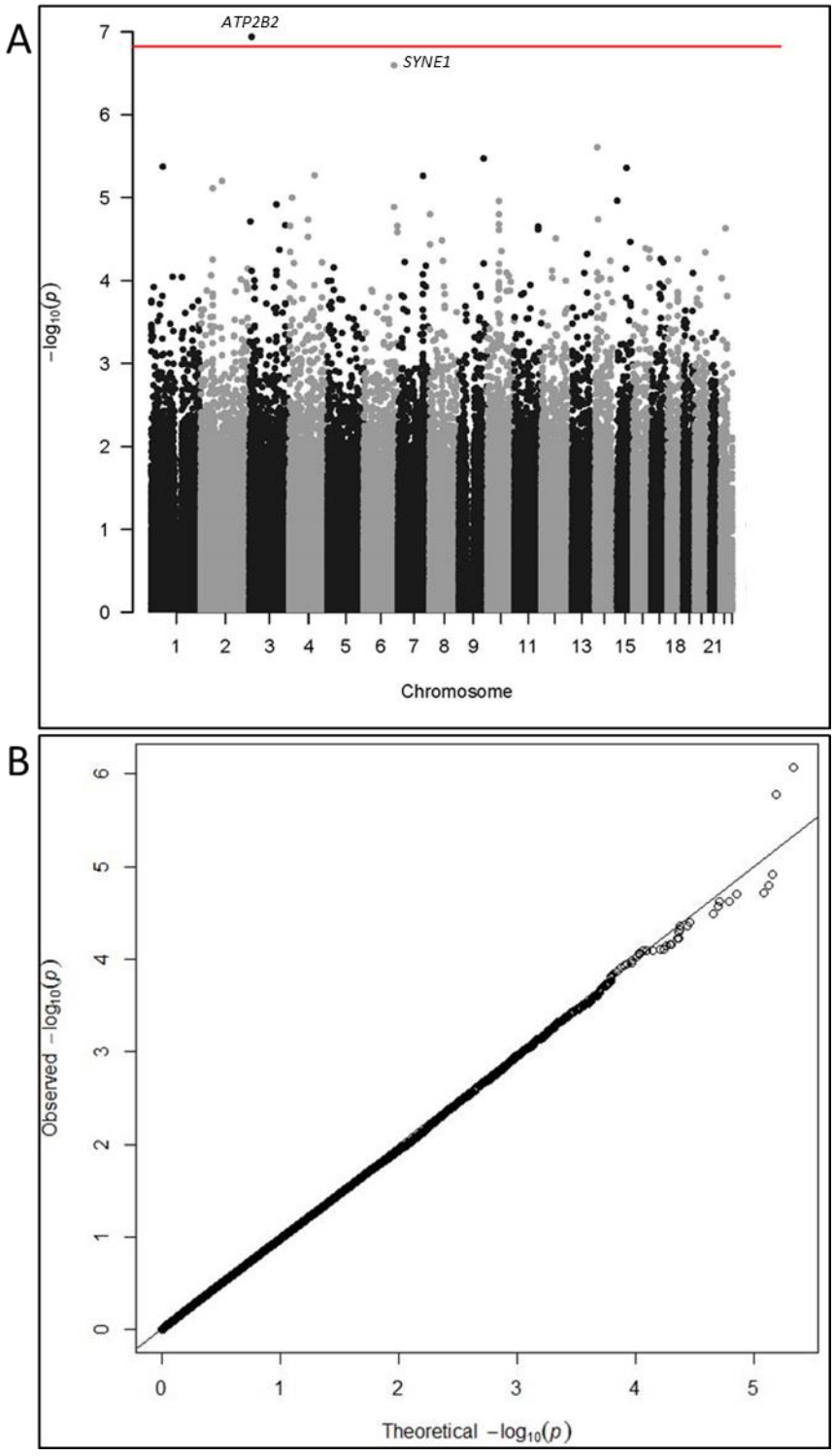


Figure 21. Manhattan (A) and quantile-quantile (B) plots illustrating GWAS results in association with endurance (red line indicates the Bonferroni level of significance, P=1.46E-7).

Pathway analysis using Ingenuity Pathway Analysis software (IPA) predicted top five affected physiological system development and functions (Table 9). Nervous system development occupies the top in the list. The model predicted abnormal morphology of neurons by variants in the following genes: *ATP2B2*, *BRAF*, *EPHB1*, *GRID2*, *NTRK3*, *SYNE1*, *THRB* and *TPH2* (p value range 2.8E-3 – 5.8E-7). The functional relevance of this association remains to be investigated. Also, SNPs in twelve genes associated with endurance are found to be related to cardiovascular system development. Of these genes, *ABCC4*, *BRAF*, *NTRK3*, *PIK3CA*, *PRKD1*, *SYNE1*, and *VAV3* are associated with cardiac enlargement and cardiac dysfunction (p=0.000016) with direct relevance to increased CVD and subsequently endurance.

Table 9. IPA analysis of top SNPs associated with high endurance athletes' performance.

Physiological System Development and Function

Name	p-value range	#SNPs
Nervous System Development and Function	2.85E-03 - 5.78E-07	20
Tissue Development	2.85E-03 - 5.78E-07	20
Cardiovascular System Development and Function	2.85E-03 - 1.25E-06	12
Organ Development	2.85E-03 - 1.25E-06	14
Organismal Development	2.85E-03 - 4.56E-06	21

Cardiotoxicity

Name	p-value range	#SNPs
Cardiac Enlargement	2.22E-01 - 1.61E-04	7
Congenital Heart Anomaly	1.83E-01 - 1.43E-03	3
Cardiac Damage	2.85E-03 - 2.85E-03	1
Cardiac Arrythmia	2.40E-02 - 1.25E-02	2
Tachycardia	2.26E-02 - 1.25E-02	2

3.3.2 Comparing high power performers against low/moderate power performers.

Genotype distribution of SNPs was compared between high power against the moderate low/moderate groups. Several SNPs associated with high power athletic performance at p-values $<10^{-5}$ were identified (Table 10) by linear regression model analysis after correcting for gender, and population stratification using PC1, PC2, PC3 and PC4, but none reached Bonferroni significance (1.47E-7). Table 10 shows top 10 SNPs with their odd ratios (OR), location according to function genome variation server (GVS), gene name and MAF in low/moderate and high power groups. MAF in non-elite athletes from 1000 genome project were used as a reference. Manhattan and quartile-quartile (QQ) plots of GWAS hits associated with endurance are shown in Figure 22.

Table 10. Top GWAS SNPs associated with Power (n=753, High= 273, Medium=156, Low= 324). Abbreviations: single nucleotide polymorphism (SNP), chromosome (Chr), Position (BP), Reference Allele (A), number of samples (N), odds ratio (OR), standard error (SE), probability value (Pval), function genome variation server (GVS), and minor allele frequencies (MAF).

SNP	Chr.	BP	A	N	OR	SE	P val	FDR	Function GVS	Nearest Gene	MAF-High Power N=273	MAF-Low Power N=324	MAF-nonathletes
rs2073307	14	69342250	C	739	0.21	0.045	3.08E-06	0.64	intron	*ACTN1*	0.46	0.44	C=0.47
rs6732989	2	75778286	C	753	0.275	0.06	4.53E-06	0.77	intron	*FAM176A*	0.2	0.12	C=0.12
rs1430779	2	67878108	T	753	-0.216	0.047	6.08E-06	0.86	intergenic	none	0.23	0.29	T=0.27
rs2072633	6	31919578	T	753	0.2154	0.047	6.44E-06	0.88	intron	*CFB*	0.41	0.27	A=0.42
rs876549	15	52602736	A	753	1.098	0.242	6.57E-06	0.88	utr-3	*MYO5A*	0.02	0.001	T=0.09
rs135230	22	49513663	G	753	0.2223	0.049	7.21E-06	0.91	intergenic	none	0.32	0.23	G=0.32
rs2106247	16	23177942	T	753	-0.211	0.047	7.95E-06	0.93	intergenic	none	0.23	0.31	A=0.30
rs2603154	4	69363530	G	752	0.2278	0.051	9.47E-06	0.96	near-gene-3	*TMPRSS11E*	0.31	0.22	G=0.27
rs3802753	11	26682387	T	753	0.238	0.054	1.07E-05	0.97	utr-3	*ANO3*	0.24	0.15	T=0.16
rs2808770	9	1.17E+08	C	748	-0.198	0.045	1.45E-05	0.99	missense	*COL27A1*	0.25	0.33	C=0.23
rs864687	7	1.05E+08	C	751	0.3949	0.091	1.65E-05	0.99	intron	*RINT1*	0.09	0.04	C=0.07

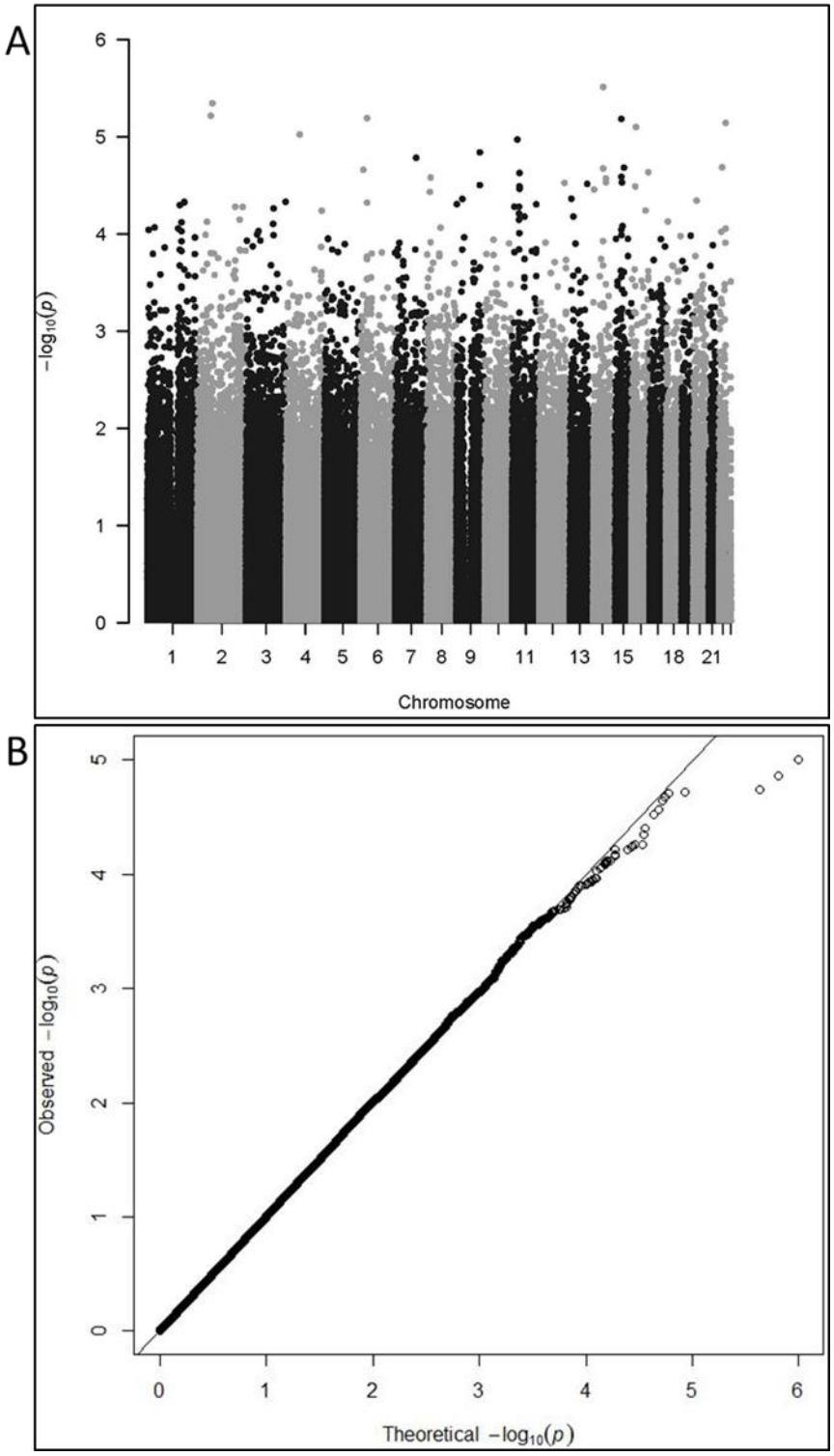


Figure 22. Manhattan (A) and quantile-quantile (B) plots illustrating GWAS results in association with power. No Bonferroni level of significance found.

Pathway analysis using IPA was also utilized to predict top five affected physiological system development and functions (Table 11). Nervous system and cardiovascular system development also occupied the top two functions in the list as predicted by variants associated with high power athletes (p value range 4.9E-2 – 4.54E-4).

Table 11. IPA analysis of top SNPs associated with high power athletes' performance.

Physiological System Development and Function

Name	p-value range	#SNPs
Nervous System Development and Function	4.94E-02 - 4.54E-04	9
Cardiovascular System Development and Function	3.97E-02 - 1.76E-03	6
Connective Tissue Development and Function	4.14E-02 - 1.76E-03	7
Hair and Skin Development and Function	4.14E-02 - 1.76E-03	3
Organ Morphology	4.48E-02 - 1.76E-03	8

3.3.3 Comparing high performers with high cardiovascular demand against low/moderate performers with lower cardiovascular demand.

Genotype distribution of SNPs was compared between high CVD (grey cells in Table 7) against low/moderate CVD group. SNPs associated with high CVD athletic performance at p-values <E-5 were identified (Table 12) by logistic regression analysis after correcting for gender, and population stratification using PC1, PC2, and PC3, but none reached Bonferroni significance (1.47E-7). Table 12 shows top 10 SNPs with their odd ratios (OR), location according to function genome variation server (GVS), gene name and MAF in high and low CVD groups. MAF in non-elite athletes from 1000 genome project were used as a reference. Manhattan and quartile-quartile (QQ) plots of GWAS hits associated with CVD are shown in Figure 23.

Table 12. Top GWAS SNPs associated with CVD (n=753, High= 335, Moderate= 418). Abbreviations: single nucleotide polymorphism (SNP), chromosome (Chr), Position (BP), Reference Allele (A), number of samples (N), odds ratio (OR), standard error (SE), probability value (Pval), function genome variation server (GVS), and minor allele frequencies (MAF).

SNP	Chr.	BP	A	N	OR	SE	Pval	FDR	Function GVS	Nearest Gene	MAF-High CVD N=335	MAF-Low CVD N=418	MAF-nonathletes
rs8032767	15	66244634	G	752	2.238	0.15	1.77E-07	0.056	intron	*MEGF11*	0.2	0.12	G=0.10
rs2171302	10	3815292	T	753	0.4317	0.18	2.34E-06	0.533	intergenic	*KLF6*	0.1	0.15	T=0.15
rs6476151	9	29730581	C	753	0.5833	0.11	2.42E-06	0.545	intergenic	none	0.34	0.45	C=0.41
rs2808661	1	159558258	A	753	1.927	0.14	4.50E-06	0.769	coding-synonymous	*APCS*	0.24	0.15	A=0.15
rs2073307	14	69342250	C	739	1.722	0.12	8.26E-06	0.932	intron	*ACTN1*	0.46	0.43	C=0.47
rs8040847	15	85447805	T	753	1.705	0.12	1.13E-05	0.975	intron	*SLC28A1*	0.39	0.29	T=0.37
rs1889055	9	29539919	C	753	0.6176	0.11	1.72E-05	0.996	intergenic	none	0.39	0.49	G=0.37
rs560764	9	29592836	C	753	1.628	0.11	1.78E-05	0.997	intergenic	none	0.46	0.42	T=0.45
rs11071854	15	66275951	G	753	1.761	0.13	1.81E-05	0.997	intron	*MEGF11*	0.28	0.19	G=0.17
rs2290272	15	85447431	T	753	1.653	0.12	2.07E-05	0.999	missense	*SLC28A1*	0.4	0.3	A=0.40

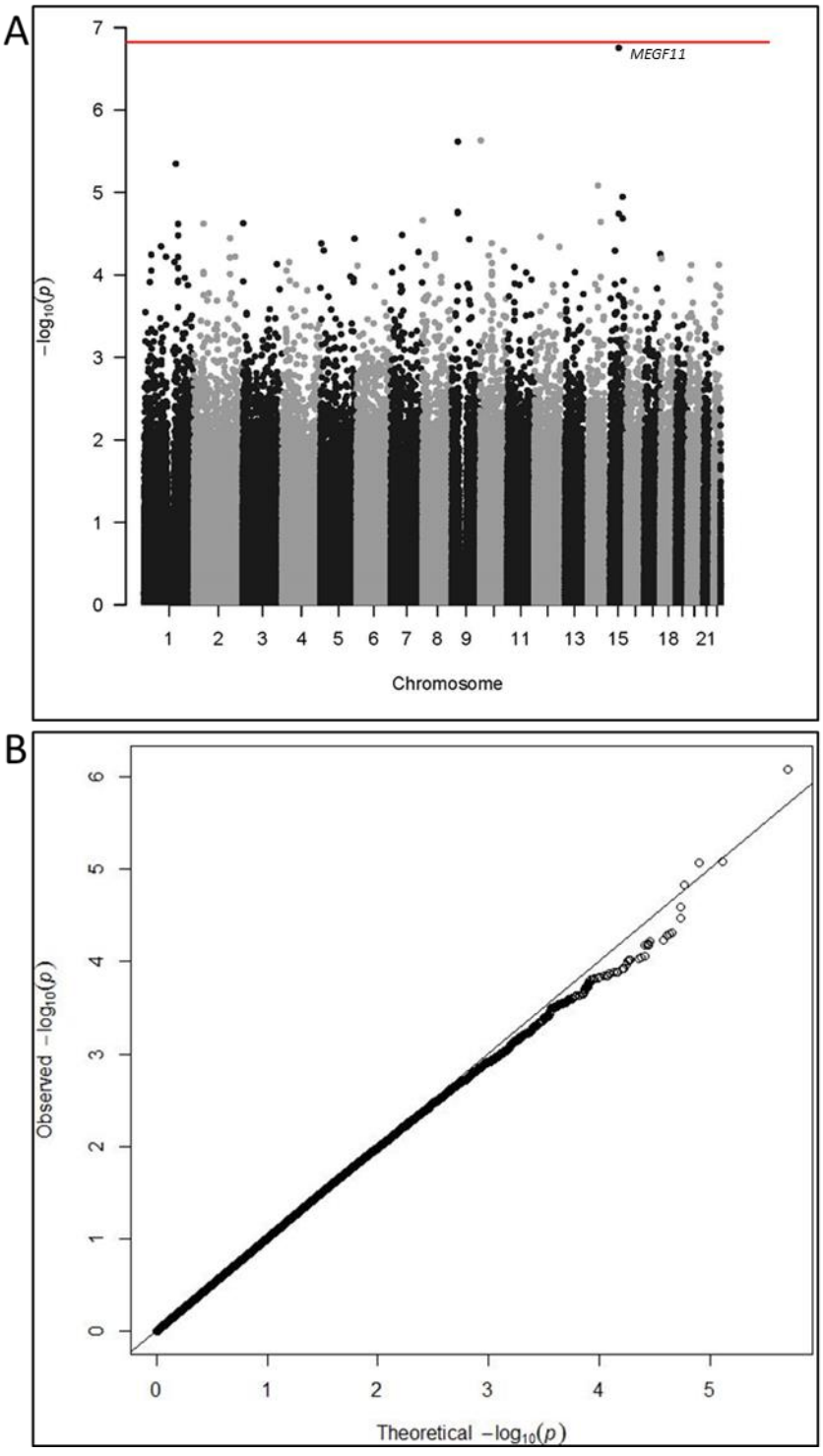


Figure 23. Manhattan (A) and quantile-quantile (B) plots illustrating GWAS results in association with CVD (genome wide line=-log10(5e-08), Bonferroni line (red) =-log10(1.5E-7)

Similarly, IPA was used for functional enrichment analysis to investigate the physiological system development and functions (Table 13). Nervous system and organ development occupied the top two functions in the list as predicted by variants associated with increased cardiovascular demand-linked to increased athletic performance (p value range 4.1E-2 – 9.5E-4).

Table 13. IPA pathway analysis of top SNPs associated with increased cardiovascular demand-linked to increased athletic performance.

Physiological System Development and Function

Name	p-value range	#SNPs
Nervous System Development and Function	4.91E-02 - 9.48E-04	3
Organ Morphology	4.11E-02 - 9.48E-04	3
Organismal Development	4.54E-02 - 9.48E-04	5
Tissue Development	4.27E-02 - 9.48E-04	5
Connective Tissue Development and Function	4.54E-02 - 1.89E-03	3

3.4 Discussion

3.4.1 Genetic associations

Genetic predisposition into cardiorespiratory fitness and response to exercise training has been previously described [189, 390-394]. Since endurance sports are characterized by increased cardiorespiratory capacity, genetic predisposition into elite endurance performance is also expected to be genetically influenced [152]. However, genetic studies of elite athletic endurance showed inconsistent results [113, 152, 190, 395]. Additionally, a large meta-analysis has revealed no evidence of genetic predisposition to endurance sports, potentially due to the small size of individual studies and the complexity of

endurance as a phenotype [115]. Taken previous study designs and findings into account when designing the current project, we decided to adopt a design that compares low vs high levels of performance (power, endurance and cardiovascular demand) athletes, instead of the classical case-control design in order to avoid the potential issue of controls carrying the genetic predisposition but not pursuing athleticism as a career. However, when we looked for an independent cohort to validate our findings, the only available ones were case-control cohorts (such as the Russian elite athletes), hence they were used for validation. Our GWAS results have revealed two novel SNPs (rs56330321 and rs2635438) associated with endurance at Bonferroni and FDR level of significance, respectively. Validation of the results in an independent cohort of elite Russian athletes and controls has confirmed the association of rs56330321 and rs2635438 with endurance athlete status. Subsequent meta-analysis of the two cohorts has shown for the first time that both SNPs were associated with endurance in elite endurance athletes at GWAS level of significance. The two novel SNPs (rs56330321 and rs2635438) are located within genes *ATP2B2* and *SYNE1*, respectively. Although these two genes have not been previously implicated directly in physical performance, their potential roles in cell signaling and cytoskeletal structure of skeletal muscle cells were previously established [396, 397].

The top endurance GWAS significant SNP (rs56330321) is located within the intron of *ATP2B2*. This gene codes for the plasma membrane Ca2+ ATPase 2 (PMCA2), a member of the P-type primary ion transport ATPases. These ATPases play a crucial role in calcium homeostasis as they remove bivalent calcium ions from the cell against high gradients [396]. PMCA2 is mostly expressed in the inner ear, the cerebellum and the mammary gland with an established role in hearing and balance in mice [398] and

humans [399]. The expression of different isoforms and splice variants is highly regulated following the physiological demand of the cell [400]. The association between PMCA2 and physical performance has not been previously described. The under representation of the A allele in the high endurance athletes in the discovery and replication cohorts, compared to their moderate endurance/non-athletes controls, may suggest that carrying the A allele is disadvantageous for high endurance athletes.

The second endurance GWAS significant SNP (rs2635438) is located within the intron of *SYNE1*. This gene codes a spectrin-repeat containing protein expressed in skeletal and smooth muscle, as well as peripheral blood lymphocytes. The protein localizes to the nuclear membrane. Mutations in this gene have been associated with autosomal recessive spinocerebellar ataxia 8, Emery-Dreifuss muscular dystrophy type 4, dominant muscular dystrophy and Emery-Dreifuss muscular dystrophy-like [401-405]. Both discovery and replication cohorts have shown that the G allele is under represented in high endurance athletes compared to moderate endurance, suggesting that carriers of the G allele may have lower endurance ability, perhaps through replacement of healthy muscle tissue by fibrosis and fatty infiltration described in recessive arthrogryposis families carrying mutations in *SYNE1* gene [405].

Subsequent analyses of GWAS data in association with power and CVD did not show Bonferroni significant hits. It is well established that sports with highly intensive training can lead to an increased cardiac mass and tissue remodeling [406-408]. Training causes expansion of the right and LV chambers, with increased wall thickness. Extreme changes in wall thickness are usually associated with rowing, cross-country skiing, cycling, and swimming [407]. Hence, genetic predisposition into larger cardiac mass may give athletes

the advantage of increased cardiac output, but may have increased their cardiovascular risk [408].

Our data identified one borderline significant SNP (rs8032767) associated with high performers in CVD group, but it has no apparent functional relevance to physical activity. SNP (rs8032767) is associated with gene *MEGF11* (Multiple EGF Like Domains 11) that has a critical function in regulating mosaic signals by two retinal interneuron subtypes, ensuring similarity in all parts of the visual field to access a full set of processing elements [409]. Functional association between this gene and cardiovascular demand phenotype need to be further investigated to confirm the genetic association.

Further investigations of the functional relevance of the identified SNPs and associated metabolites in relation to enhanced athletic performance are needed to confirm these associations and identify their functional relevance. Functional characterization of genetic variants is often required to move from statistical association to causal variants and genes, especially in the non-coding genome. Computational methods are often used to predict the regulatory effect of non-coding variants on the basis of functional annotations. Target genes can be identified or confirmed using chromatin immunoprecipitation and chromosome conformation capture methods, and experimentally validated using cell-based systems and model organisms [410]. For *ATP2B2* SNP, expression studies using reporter gene assays investigating calcium signaling in wild type and mutant *ATP2B2* SNP (rs56330321) will reveal the potential effect of the SNP on calcium homeostasis [411]. Similarly, for SYNE1 gene variant, a multi-isomeric modular protein which forms a linking network between organelles and the actin cytoskeleton, experiments comparing the connection between the nuclear lamina and the cytoskeleton between wild type and

rs2635438 mutants will reveal whether SYNE1 SNP (rs2635438) destabilizes the actin cytoskeletal structure of the cells. Further studies on its role in cardiomyocytes contraction would provide additional evidence of the functional role of this SNP in relation to athletic performance [412].

3.4.2 Genomics study limitations

Firstly, categorization of elite athletes into groups based on sport disciplines characterized by variable peak dynamic (VO_2max) component reached during competition may not be sufficient as it is not based on actual measurement of VO_2max. This limitation is caused by restricted access to athletes' information as per study ethics.

Secondly, common genetic variations exhibit relatively low effect sizes, especially in the context of complex phenotype such as physical performance [413]. Despite the fact, GWAS had a major contribution towards the identification of genetic variants associated with several common complex phenotypes. The modest effect size of these variants limits their utility in functional prediction.

Thirdly, the added complexity of sport phenotype (sport group) is one of the features that makes the search for common variants difficult even when using VO_2max, MCV and CVD-based study designs. Such complexity arises from several factors including the position of the player as, for example, we can't differentiate between a goal keeper and a outfield player in the samples collected from football team. Great difference in athletes' contribution and involvement in sport performances makes it difficult to separate the heterogenic phenotypes, as the fact that endurance, power and CVD are not monogenic. In several cases, causal variants were found within key associated genes, and their gene-environment interactions were determining the effect of some risk variants; such as CVD

associated variant *MEGF11*. Such complexity may explain why there was no clear co-segregation or association evidence for the variant identified in this study. Therefore, further investigations are necessary in order to obtain valid conclusions about the role of the identified candidate variants in relation to elite athlete physical performance.

In addition, the functional relevance of many of these variants to physical activity is unclear since most of them are mapped to intergenic or intronic regions [413]. Therefore, additional follow-up studies are necessary to fully understand the genes and mechanisms deriving the effect on study groups, as in the case of the *MEGF11* and *ATP2B2* genes.

Conclusion: This study was carried out with the largest GWAS analysis of elite European athletes to date using a unique SNP microarray that is enriched with genes involved in different metabolic pathways with direct influence on various physiological pathways characteristic of elite athletes. This study reports the first GWAS significant SNPs associated with endurance in elite athletes in genes with no previous association with physical performance. Further investigations of the functional relevance of the identified SNPs and associated metabolites in relation to enhanced athletic performance are needed to confirm these associations and identify their functional relevance. In chapter 5, metabolites associated with top endurance SNPs were further investigated as a step toward unravelling their functional validation.

CHAPTER 4
Results: Metabolomics

4 (Chapter 4) Results: Metabolomics

4.1 Introduction

The professional athlete's body adapts to the type of physical activity and elevates the capacity of physiological systems to carry out the increased workloads and to enhance performance. For instance, a study examining the association between hormone fluctuations and human behavior focused on the hypothalamic-pituitary-adrenal (HPA) axis in different exercise groups [414]. The study revealed that both moderate and high intensity of exercise could significantly increase HPA axis function, leading to increased cortisol concentration. The study concluded that salivary cortisol measurement could be a reliable biomarker of physical performance [414, 415]. Therefore metabolites profiling using the quantitative measurement of metabolic responses to that elevation in various systems could provide a potential tool for a deeper insight into athlete's physiological state, and explain the fluctuations induced by endurance/power physical exercise [127]. More importantly, it could provide straightforward phenotypes for genotype association. One example is a study aimed at examining the alterations in the metabolic biochemical compositions of professional athletes (rowers) compared with non-athlete (control) subjects. In this study, the authors monitored the endogenous metabolomics status of rowers during a training program and identified sports-associated variation in metabolic phenotype in professional athletes. They used gas chromatography/time of flight–mass spectrometry (GC/TOF-MS)-based metabolomics. Multivariate statistical analysis, such as partial least squares projection to latent structures and discriminant analysis (PLS-DA) was carried out using SIMCA-P 11 software to visualize clustering of study groups and their metabolites (Figure 24). Significant differences in metabolic profiles were observed

with mean difference that was considered significant at 0.05 or 0.01 level (Table 14) compared between the professional athletes and control subjects [124, 126, 128, 416-418].

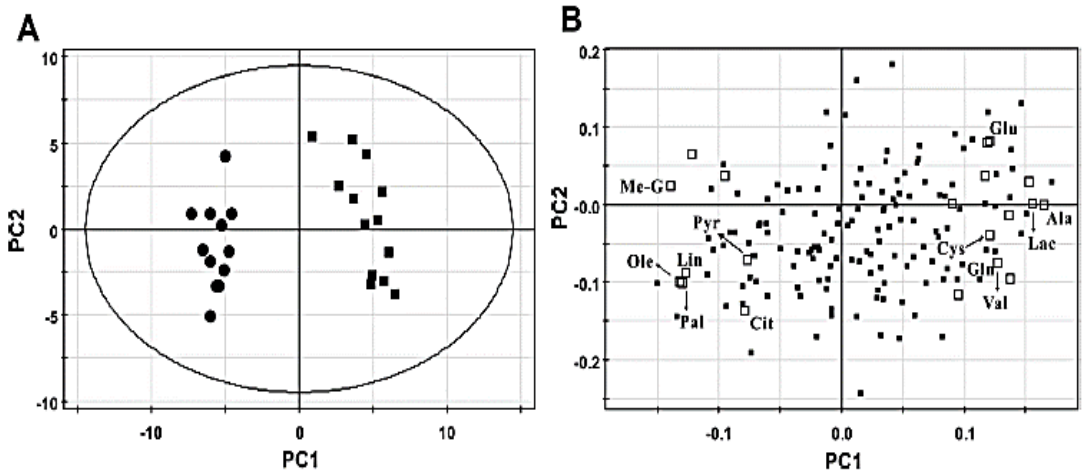


Figure 24. Significant metabolomics difference observed between ● control subjects and ■ rowers, adopted from [126]. A: PLS-DA score plots (PC1/PC2) of control subjects (left) and the professional athletes (right) indicating a significant separation in measured metabolites of studied groups. B: PLS-DA loading plots exhibited variables positively correlated with score plots. Data showed open squares "□" endogenous metabolite (Ala, alanine; Cit, citric acid; Cys, cysteine; Gln, glutamine; Glu, glutamic acid; Lac, lactate; Lin, linoleic acid; Me-G, β-D-methylglucopyranoside; Ole, oleic acid; Pal, palmitic acid; Pyr, pyroglutamic acid; and Val, valine) were statistically significant and found responsible for the discrimination of the two groups. Closed squares "▪" indicate non-significant metaoblites.

Table 14. A list of endogenous metabolites that were found statistically different between control subjects and rowers. Adopted from [126]

Related Metabolic Pathway	Metabolites	Control Subjects	Rowers
Glucose metabolism	Alanine	122.8±12.8	359.3±21.4
	Lactate	1,467.5±134.5	3,406.2±212.1
	β-D-Methylglucopyranoside	8.14e6±2.42e6	4.69e6±1.49e6
Oxidative stress	Cysteine	312.5±26.0	487.7±30.5
	Glutamic acid	32.6±3.7	51.0±2.7
Energy metabolism	Citric acid	57.0±6.9	40.3±2.9
Lipid metabolism	Palmitic acid	260.4±22.3	150.0±11.9
	Linoleic acid	185.2±21.7	83.0±10.7
	Oleic acid	242.3±27.6	93.5±16.3
Amino acids	Valine	209.9±44.8	467.7±35.0
	Glutamine	369.3±26.0	533.4±43.7

The exhaustive physical exercise is associated with myocardial adaptations such as electrical, structural and functional changes of athletes' heart that improve athletic performance [419, 420]. The nature and intensity of an athlete's sport determines the metabolic changes in systemic blood flow, whereas the cardiovascular function is influenced by the duration of exercise over time [421]. For instance, during aerobic workout, athlete's cardiovascular system reaches up to maximum cardiac output as per the need of the exercising body organs to receive as much oxygen for regulating the mean arterial pressure [422, 423].

Despite of various physiological studies measuring blood pressure and cardiac output of different groups of athletes [424], or studies collecting information about their consumed dietary supplements [140], a more comprehensive measurement of athletes' metabolic markers in response to exercise and nutrition remains limited [141].

For example, no study has adopted a retrospective approach to measure metabolites that potentially originate from drugs/supplements consumption in elite athletes or analyzed the metabolic signature of different groups of elite athletes who belong to different sports.

Such comprehensive testing of blood metabolic biomarkers may offer valuable insight into the pathophysiological and functional alterations underlying athletes' long-term performance and health considering changes in metabolites associated with their nutrition, CVD, MVC and VO_2max.

The aim of this chapter was to identify metabolites associated with endurance and power in discovery and replication cohorts. This chapter also aims to investigate metabolites associated with CVD and analyze xenobiotic profiling in serum samples from elite athletes of different sport disciplines.

4.2 Material and method

4.2.1 Classification of study participants for metabolomics analysis

Metabolomics analysis of serum samples from elite athletes was performed in discovery (n=191) and replication cohorts (n=501) as described in the general methods, section 2.4. Categorization of participants' sport groups for both cohorts into low/moderate vs. high endurance, power and CVD groups is shown in Tables 15 and 16.

Table 15. Classification of pilot metabolomics study of 191 participants. Elite athletes are distributed in various categories is based on sports type as described previously [16]. The number of participants and gender (M for males and F for females) in each group are mentioned.

(15) Discovery		Endurance	
		Moderate (40-70% VO2max)	**High (>70% VO2max)**
Power	**High (>50% MVC)**		Boxing (12M), Triathlon (4M), Rowing (2M), Cycling (21M/3F), Canoe Kayak (1F)
	Moderate (20-50% MVC)	Athletics (5M/5F), Rugby (41M), Canoe 200m Sprint (4M/1F), Motorcycle Racing / Road (5M)	Skiing Cross Country (1M), Basketball (4M), Swimming (9M)
	Low (<20% MVC)	Baseball (1M), Volleyball (4M/4F)	Tennis (1F), Football (62M), Long distance 3000m or greater (1M),

Table 16. Classification of 2nd metabolomics study of 490 participants. Elite athletes are distributed in various categories is based on sports type as described previously [16]. The number of participants and gender (M for males and F for females) in each group are mentioned.

(16) Replication		Endurance	
		Moderate (40-70% VO2max)	**High (>70% VO2max)**
Power	**High (>50% MVC)**	Wrestling (3M), Judo (3M)	Boxing (1M/16F), Heptathlon (1M), Rowing (6M/7F), Cycling (31M/4F)
	Moderate (20-50% MVC)	Athletics (15M/22F), Rugby (16M), Triple Jump (1M)	Athletics 200-800m (4M), Hockey (1F), Skiing Cross Country (1M), Basketball (3M), Swimming (22M/16F)
	Low (<20% MVC)	Baseball (2M), Volleyball (1M)	Tennis (1M/1F), Soccer (315M), Athletics 1500-3000m (3M)

4.2.2 Metabolomics profiling

Serum samples from 1st cohort (191 samples) and 2nd cohort (501 samples) were sent to Metabolon, Durham, NC, USA for metabolomics profiling that was performed using established protocols as described in the general methods section 2.4.1.

4.3 Results

4.3.1 Multivariate analysis of athlete metabolomics data

As noted above, PCA and OPLS-DA were run using SIMCA 14; the default metabolite-wise sample missingness threshold of 50% was performed. PCA is a hypothesis-free multivariate approach that can reduce the dimensionality of the dataset while preserving the original relationships inherent within the dataset including any population stratification. OPLS-DA can exhibit sets of metabolites that best differentiate the predefined classes of sports.

4.3.1.1 Principle components analysis (PCA) 1st cohort

PCA revealed no clusters of samples influenced by sport types, gender, or study groups (Figure 25A). First two principle components (PC1 and PC2) together captured 25% of the variance in the data (Figure 25). Figure 25A is scatter plot that discovered two clusters of samples along PC1, which were not explained by sport types, study groups, or gender. Figure 25B is a loading plot that showed heme and hemoglobin metabolites at the extreme positive end of PC1, suggesting that PC1 is influenced by hemolysis. A t-test performed to compare the hemolysis extent between the two clusters of samples discovered a significant difference ($p= 0.01$) between the two groups. These results reveal that PC1 captured presence of hemolysis in the samples. Likewise, arachidonate phospholipid metabolites were found enriched at the positive end of PC1 in contrast to eicosanoids enrichment at the negative end. These two metabolites are biochemically linked as substrate/product, but the link to hemolysis remains unclear. The loading plot in Figure 25C shows amino acids that feed into tricarboxylic acid cycle (TCA) and TCA cycle energy metabolites clustering at the positive end of PC2, suggesting a potential metabolic signature of exercise. A significant positive correlation has been revealed between previously identified variations in metabolites pre/post endurance exercise [128], listed in Table 17, and PC2 loading values for the same metabolites was identified ($R=0.6$, $p=0.005$). Since the loading values drive the corresponding PC score values, PC2 may represent a pre/post exercise metabolic signature. Focusing at the negative end of PC2, there is enrichment of dipeptides that suggest an opposing anabolic effect. Consequently, in summary, PCA results have been significantly beneficial in detecting possible confounders (hemolysis and pre/post exercise) that were used as covariates to be corrected for in subsequent analyses.

Table 17. Comparison of variations in previously identified metabolites pre/post exercise [416] and their PC2 loading values acquired in this study.

Metabolites	Changes in concentration 1 hour post exercise as reported in [416]	PC2 loading values from this study
Malate	69	0.07
citric/Isocitrate	44	0.052
Aconitic acid	25	0.06
Fumarate	76	0.07
Lactate	81	0.05
Pyruvate	47	0.06
Succinate	20	0.06
Methionine	10	0.04
Alanine	21	0.058
Cystathione	-38	0.03
Glutamine	12	0.02
Ornithine	-19	0.036
Hipurate	-12	0.01
Allantoin	-33	0.01
Uridine	36	0.04
Uric acid	8	0.03
Nicocianidine	53	0.021
Pantothenate	18	0.06

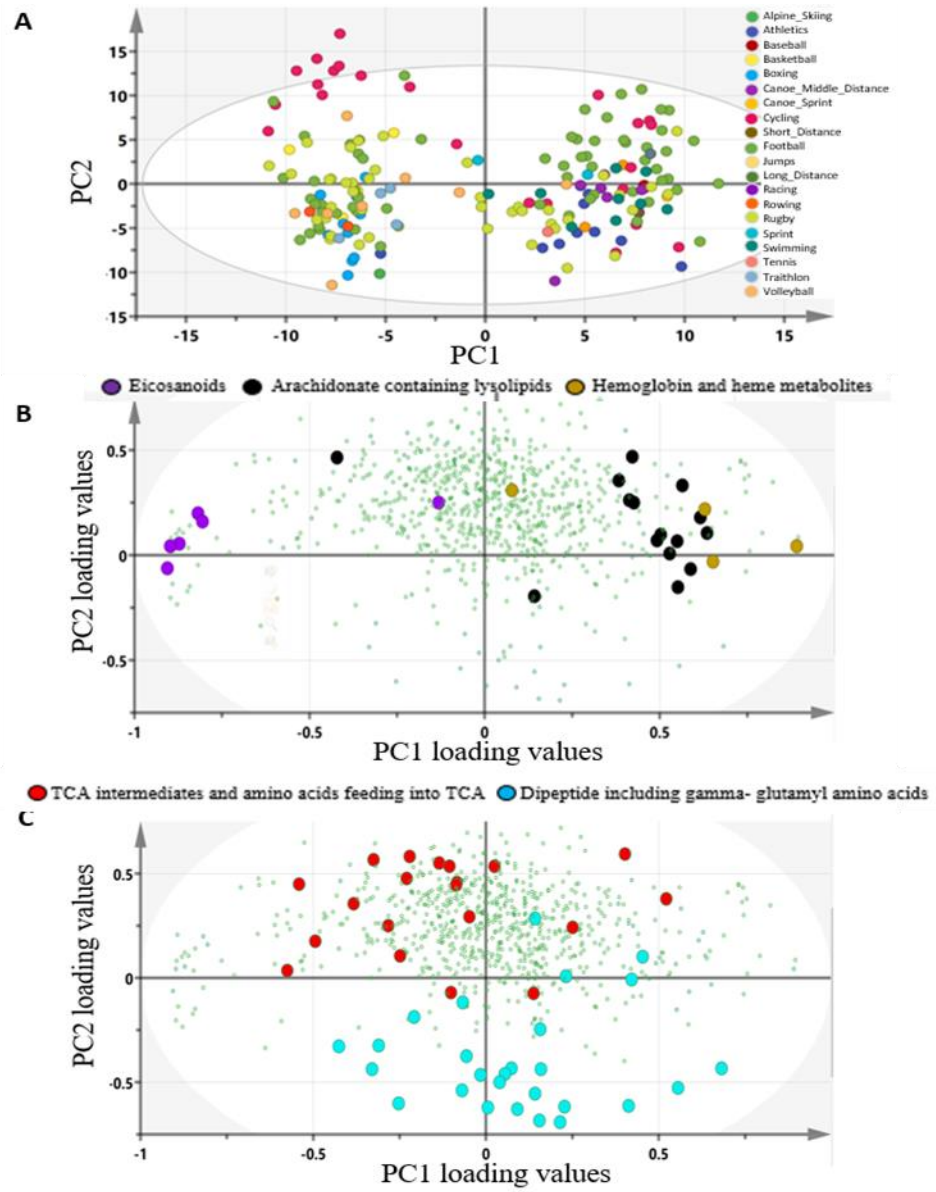


Figure 25. PCA analysis of metabolomics data from 191 elite athletes (Discovery cohort). (A) A score plot of PC1 and PC2 signifying clusters of samples into two groups according to PC1 without any influence of sport type. (B, C) Loading plots reveals that (B) the hemoglobin/heme metabolites suggest a hemolysis presentation for PC1, whilst (C) the TCA energy metabolite suggest an energy generating process for PC2 which may be associated with exercise.

4.3.1.2 OPLS-DA Endurance Group

An OPLS-DA analysis compared moderate vs. high endurance. Figure 26A indicated 66.7% class-discriminatory component of the alteration in the data due to endurance level (R-squared-Y = 0.66, Q-squared = 0.45). Figure 26B reveals the corresponding loading values, indicating a decrease in gamma-glutamyl amino acids and diacyl glycerols as well as increase in monohydroxy fatty acids and steroids with higher endurance levels.

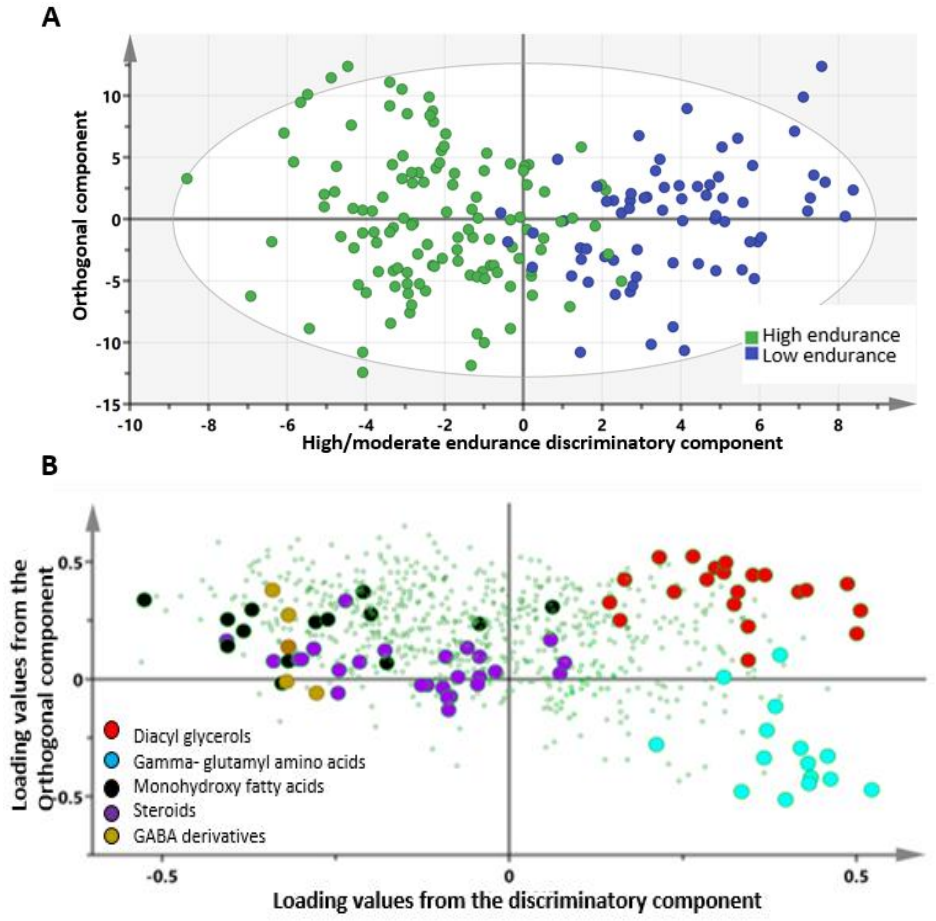


Figure 26. OPLS-DA model comparing high vs. moderate endurance groups of discovery cohort. (A) A score plot showing the orthogonal component (y-axis) versus class-discriminatory component (x-axis). (B) The corresponding loading plot discovered a positive association between monohydroxy-fatty acids and steroids with endurance and a negative association with gamma-glutamyl amino acids and diacyl-glycerols.

4.3.1.3 OPLS-DA Power Group

OPLS-DA also discovered a clear separation between high vs. moderate power. Figure 27A reveals the significant predictive component that is explaining 88% of the variation in the power (R-squared-Y =0.88, Q-squared = 0.52), together with four orthogonal components. The loading plot on Figure 27B suggests a decrease in gamma glutamyl amino acids and steroids as opposed to an increase in lyso lipids, phospholipids and xanthine metabolites with increased power.

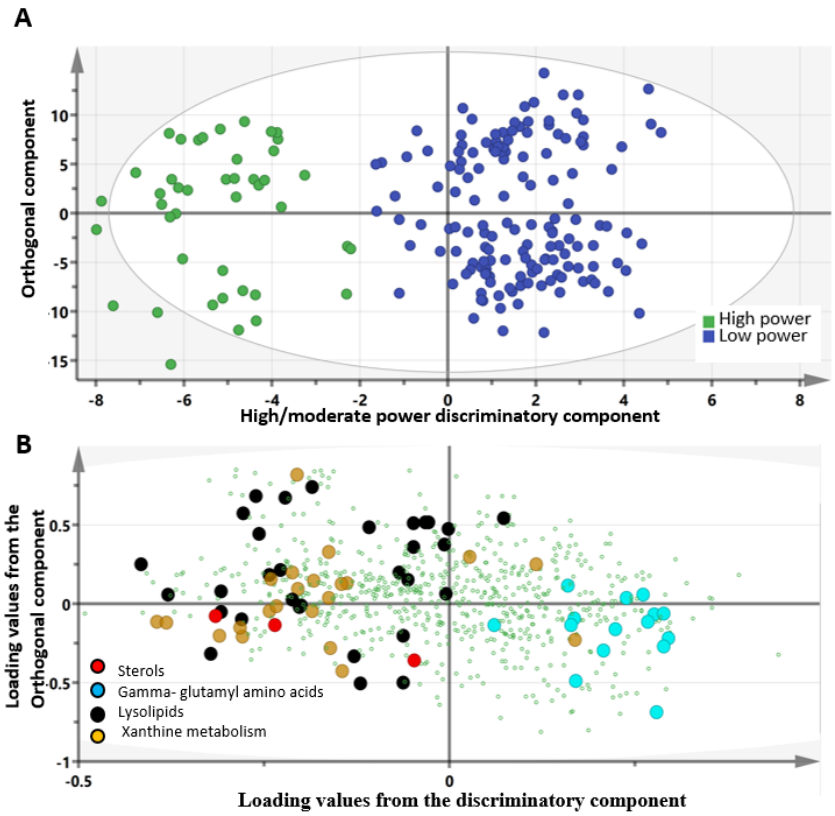


Figure 27. OPLS-DA model of high vs. moderate power groups of elite athletes. A score plot showing the orthogonal component (y-axis) versus the class-discriminatory component (x-axis). (B) The corresponding loading plot showing a positive association between power and lipids, and xanthine metabolites and a negative association with sterols and gamma-glutamyl amino acids.

4.3.2 Univariate association tests and function enrichment analysis

Metabolomics data were used in the association analysis only after log-transformation and z-score normalization. Linear regression models were run using the R statistical package (version 2.14, www.r-project.org/). Incorporating power and endurance as a categorical variable with two levels (moderate & high) was used in the model as explained above.

4.3.2.1 Endurance-associated metabolites in elite 1st cohort of 191 athletes

A linear regression model was used to identify the significant metabolite associated with high vs. moderate endurance after correcting for covariates: gender, hemolysis levels, PC1, PC2 and power. Thirty-nine metabolites were found associated with endurance at a Bonferroni level of significance ($p <= 0.05/743 = 6.72$ E-5) and Table 18 lists their associated pathways. More metabolites that are associated at FDR level of significance with endurance are shown in Table 33 in the appendix section.

Table 18. List of metabolites significantly associated with high endurance athletes (Bonferroni significance).

Metabolite	Subpathway	Change	Bonferroni p value
1-stearoyl-GPC (18:0)	Lysolipid	-0.15595	1.72E-05
Vanillylmandelate (VMA)	Phenylalanine and Tyrosine Metabolism	0.415133	2.29E-05
21-hydroxypregnenolone disulfate	Steroid	0.365863	0.000107398
Palmitoyl-linoleoyl-glycerol (16:0/18:2) [2]*	Diacylglycerol	-0.46764	0.000130998
Tartronate (hydroxymalonate)		0.290077	0.000657114
Palmitoyl-linoleoyl-glycerol (16:0/18:2) [1]*	Diacylglycerol	-0.42202	0.00090176
1-palmitoleoyl-GPC (16:1)*	Lysolipid	-0.22642	0.001172265
Cortisone	Steroid	0.395892	0.001489996
Citrate	TCA Cycle	0.200056	0.001784274
Succinimide	Polyamine metabolism	0.279317	0.002636335

stearoylcarnitine (C18)	Fatty Acid Metabolism(Acyl Carnitine)	-0.28394	0.002953686
Trans-4-hydroxyproline	Urea cycle; Arginine and Proline Metabolism	-0.27783	0.00295413
4-guanidinobutanoate	Polyamine metabolism	-0.44969	0.003796483
Dihomo-linoleoylcarnitine (C20:2)*	Fatty Acid Metabolism(Acyl Carnitine)	-0.33166	0.005028391
1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*	Plasmalogen	0.145176	0.005178692
1-palmitoyl-GPC (16:0)	Lysolipid	-0.11595	0.005429078
Linoleoyl-linoleoyl-glycerol (18:2/18:2) [1]*	Diacylglycerol	-0.54301	0.005827373
Gamma-glutamylglutamate	Gamma-glutamyl Amino Acid	-0.42069	0.006242208
Pregnanediol-3-glucuronide	Steroid	0.44061	0.006441558
Palmitoyl-arachidonoyl-glycerol (16:0/20:4) [2]*	Diacylglycerol	-0.47247	0.008366458
1-palmitoyl-2-stearoyl-GPC (16:0/18:0)	Phospholipid Metabolism	-0.1648	0.009358338
Cortisol	Steroid	0.471022	0.009967366
Linoleoyl-linolenoyl-glycerol (18:2/18:3) [2]*	Diacylglycerol	-0.53635	0.012030273
Homoarginine	Urea cycle; Arginine and Proline Metabolism	-0.22816	0.013313047
Palmitoleoyl-linoleoyl-glycerol (16:1/18:2) [1]*	Diacylglycerol	-0.42989	0.015554355
Lactosyl-N-palmitoyl-sphingosine (d18:1/16:0)	Sphingolipid Metabolism	0.131658	0.017917489
3-hydroxydecanoate	Fatty Acid, Monohydroxy	0.346756	0.018411909
Pregnenolone sulfate	Steroid	0.332031	0.01854452
Pregnenolone steroid monosulfate*	Steroid	0.292548	0.024089561
Leukotriene B4	Eicosanoid	-0.84063	0.027085708
Vanillactate	Phenylalanine and Tyrosine Metabolism	0.214757	0.028124765
12-HETE	Eicosanoid	-0.63302	0.028449419
Acetylcarnitine (C2)	Fatty Acid Metabolism(Acyl Carnitine)	0.337317	0.033107027
N1-methyladenosine	Purine Metabolism, Adenine containing	0.121048	0.036870759

Isovalerate	Leucine, Isoleucine and Valine Metabolism	-0.52129	0.039358891
5-hydroxylysine	Lysine Metabolism	-0.39575	0.040606024
1,3,7-trimethylurate	Xanthine Metabolism	0.671617	0.045828468
Fructose	Fructose, Mannose and Galactose Metabolism	0.391699	0.053677595

Enrichment analysis indicated an over-representation of monohydroxy fatty acids, eicosanoids, gamma-glutamyl amino acids and diacylglycerols (FDR-corrected p-value 0.04, 0.017, 0.005 and 0.000122 respectively) amongst metabolites associated with endurance, regardless of change direction. Noticeably, steroids had a nominal significant p-value of 0.05, but lost significance after FDR-based multiple testing. Of interest, this data is similar to the identified metabolic effects discussed by the OPLS-DA model shown earlier (Fig. 12.B).

The results concerning steroids are absolutely remarkable and will be explained further in the discussion. Additionally, there are six Bonferroni significant steroids listed in Table 18, and seven more steroid species at FDR significant at alpha=0.05. These are androstenediol (3alpha, 17alpha) monsulfate (FDR p-value=0.04), pregnen-dioldisulfate (FDR p-value=0.035), 5alpha-pregnan-3beta,20alpha-diol monosulfate (FDR p-value=0.029), androstenediol (3beta,17beta) disulfate (FDR p-value=0.025), 5alpha-pregnan-3beta,20beta-diol monosulfate (FDR p-value=0.02), 5alpha-pregnan-3beta,20alpha-diol disulfate (FDR p-value=0.01), and etiocholanolone glucuronide (FDR p-value=0.003). All these significant steroid metabolites were investigated further using KEGG to highlight their biochemical inter-relationships in Steroid Biosynthesis Pathway (Fig. 28).

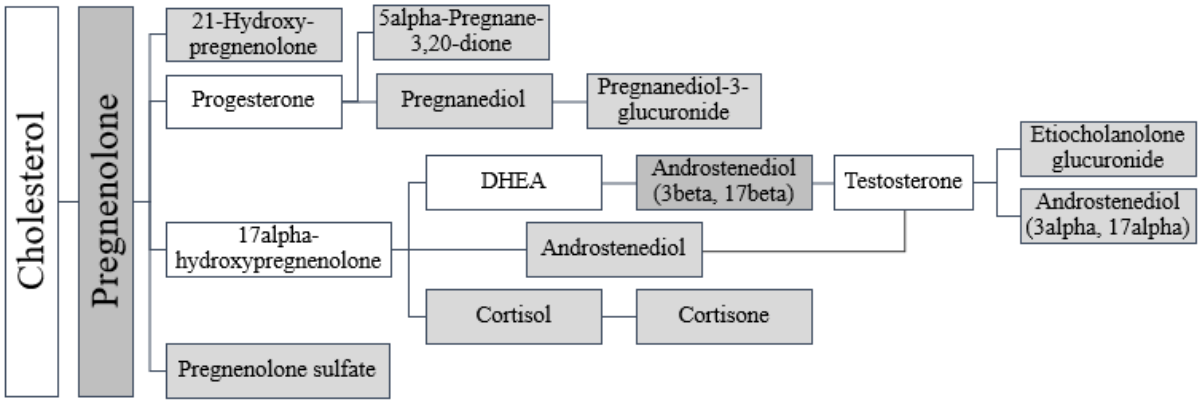


Figure 28. A schematic diagram reviewing steroid metabolites and their biochemical relationships. Shaded boxes indicate metabolite significantly associated with high endurance based on our data. This steroid hormone biosynthesis pathway is based on reference pathway (map00140) from the Kyoto Encyclopedia of Genes and Genomes (KEGG).

Significant correlations are identified between steroids metabolites, including the precursor cholesterol, are shown in Table 19, suggesting high endurance athletes show highly activate sex steroid biosynthesis pathway.

Table 19. Pearson's Correlations between steroid biosynthesis-related metabolites. Significantly associated metabolites' p values are highlighted (*<0.05, **<0.01, ***<0.001).

Metabolite																			
	cholesterol	R	1.000																
		p value																	
		N	191																
	17alpha-hydroxypregnenolone 3-sulfate	R	0.027	1.000															
		p value	0.738																
		N	152	152															
	21-hydroxypregnenolone disulfate	R	.174*	.265**	1.000														
		p value	0.016	0.001															
		N	191	152	191														
	5alpha-pregnenolone-3beta,20alpha-diol disulfate	R	0.056	.322**	.529**	1.000													
		p value	0.438	0.000	0.000														
		N	191	152	191	191													
	5alpha-pregnenolone-3beta,20alpha-diol monosulfate (2)	R	0.089	.219**	.474**	.750**	1.000												
		p value	0.222	0.007	0.000	0.000													
		N	191	152	191	191	191												
	5alpha-ppregnenolone-3beta,20beta-diol monosulfate (1)	R	-0.025	.210**	.403**	.719**	.785**	1.000											
		p value	0.736	0.010	0.000	0.000	0.000												
		N	191	152	191	191	191	191											
	androstenediol (3alpha, 17alpha) monsulfate (2)	R	.149*	.490**	.388**	.194**	.374**	.249**	1.000										
		p value	0.039	0.000	0.000	0.007	0.000	0.001											
		N	191	152	191	191	191	191	191										
	androstenediol (3beta,17beta) disulfate (2)	R	.231**	.213**	.719**	.489**	.377**	.367**	.495**	1.000									
		p value	0.001	0.008	0.000	0.000	0.000	0.000	0.000										
		N	191	152	191	191	191	191	191	191									
	cortisol	R	.285**	.337**	.191**	-0.023	0.034	-0.095	.414**	0.024	1.000								
		p value	0.000	0.000	0.008	0.755	0.639	0.191	0.000	0.739									
		N	191	152	191	191	191	191	191	191	191								
	cortisone	R	.293**	.298**	.351**	.149*	.238**	0.097	.509**	.264**	.586**	1.000							
		p value	0.000	0.000	0.000	0.040	0.001	0.181	0.000	0.000	0.000								
		N	191	152	191	191	191	191	191	191	191	191							
	dehydroisoandrosterone sulfate (DHEA-S)	R	0.057	.293**	.592**	.375**	.412**	.410**	.409**	.675**	-0.046	.284**	1.000						
		p value	0.431	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.523	0.000							
		N	191	152	191	191	191	191	191	191	191	191	191						
	etiocholanolone glucuronide	R	.262**	0.129	.246**	0.085	.214**	0.118	.445**	.334**	.409**	.378**	.265**	1.000					
		p value	0.000	0.112	0.001	0.245	0.003	0.104	0.000	0.000	0.000	0.000	0.000						
		N	191	152	191	191	191	191	191	191	191	191	191	191					
	pregnenolone steroid monosulfate*	R	0.091	.287**	.783**	.410**	.488**	.451**	.453**	.605**	.160*	.388**	.731**	.220**	1.000				
		p value	0.212	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.027	0.000	0.000	0.002					
		N	191	152	191	191	191	191	191	191	191	191	191	191	191				
	pregnanediol-3-glucuronide	R	.258**	0.136	.458**	.511**	.643**	.534**	.282**	.389**	.191**	.257**	.303**	.530**	.437**	1.000			
		p value	0.000	0.094	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.000	0.000				
		N	191	152	191	191	191	191	191	191	191	191	191	191	191	191			
	pregnenolone-diol disulfate*	R	0.016	.289**	.675**	.640**	.310**	.431**	0.109	.658**	-0.129	0.084	.593**	0.024	.671**	.265**	1.000		
		p value	0.830	0.000	0.000	0.000	0.000	0.000	0.134	0.000	0.076	0.248	0.000	0.742	0.000	0.000			
		N	191	152	191	191	191	191	191	191	191	191	191	191	191	191	191		
	pregnenolone sulfate	R	0.106	.641**	.715**	.428**	.438**	.371**	.532**	.476**	.422**	.454**	.612**	.225**	.793**	.389**	.556**	1.000	
		p value	0.146	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000		
		N	191	152	191	191	191	191	191	191	191	191	191	191	191	191	191	191	
	testosterone sulfate	R	-0.067	-0.183	-0.021	-0.055	0.002	-0.067	0.022	0.117	-0.033	-0.067	-0.077	0.146	-0.030	0.209	-0.016	-0.060	1.000
		p value	0.557	0.150	0.852	0.629	0.983	0.556	0.844	0.302	0.772	0.552	0.497	0.195	0.790	0.063	0.888	0.595	
		N	80	63	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80

In addition to steroid biosynthesis-related metabolites, enrichment analysis of endurance associated metabolites also discovered further individual metabolic signatures that are remarkable. Amongst these are 2-pyrrolidinone itself (FDR p value=0.03), byproducts of GABA cyclic lactam 2-pyrrolidinone including acisoga or N-(3-acetamidopropyl) pyrrolidin-2-one (FDR p value=0.004) and succinimide (FDR p value= 0.0002462) as well as GABA derivative 4-guanidinobutanoate (FDR p value=0.000292). There were significant correlations between 2-pyrrolidinone and its derivatives including 4-guanidinobutanoate (R=-0.146, p=0.04), succinimide (R=0.15, p=0.04), and guanidinosuccinate (R=-0.186, p=0.01), confirming these results as seen in OPLS-DA (Figure 26B) and also signifying the presence of this drug and its derivatives in high endurance athletes.

Other interesting findings include increase in citrate (a Bonferroni significant) together with increase in 2-methylcitrate (an FDR significant) in high endurance athletes. Other associations include sphingolipids, phospholipids and acyl carnitines among others (Table 33 in section appendix).

Endurance associated FDR significant metabolites with p values less than 0.01 were projected on the heatmap as shown in Figure 29. The heatmap summarizes a snapshot of the actual intensities of significant metabolites after correcting for covariates in the linear regression model described above. Samples were arranged within their respective sports classes (moderate endurance vs. high endurance) considering their sport types.

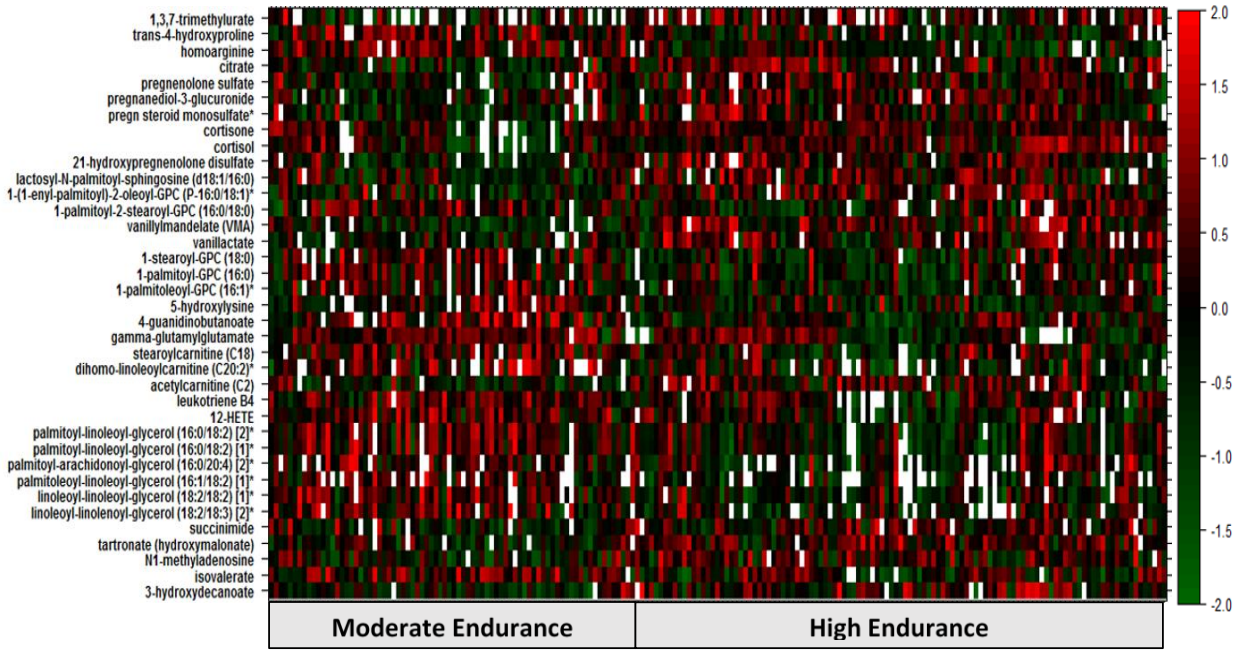


Figure 29. Heatmap of high endurance associated significant metabolites (y axis). Samples were grouped into high vs. moderate endurance groups (x axis). The color code represents scaled values of metabolites after correction of covariates.

4.3.2.2 Power-associated metabolites in elite 1st cohort of 191 athletes

A similar regression model was performed with the categorical variable 'power' becoming the variable of interest.

A linear regression model was used to identify the significant metabolite associated with moderate/low vs. high power after correcting for covariates: gender, hemolysis levels, PC1, PC2 and endurance. Thirty-three metabolites were found associated with power at a Bonferroni level of significance ($p < = 0.05/743 = 6.72$ E-5) and Table 20 listed their associated pathways. More metabolites that are associated at FDR level of significance with power are shown in Table 34 in the appendix.

Table 20. List of metabolites significantly associated with high power athletes (Bonferroni significance).

Metabolite	Subpathway	Change	Bonferroni p value
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	Phospholipid Metabolism	0.577623	5.92E-11
1-palmitoyl-2-oleoyl-GPI (16:0/18:1)*	Phospholipid Metabolism	0.42177	1.10E-07
imidazole lactate	Histidine Metabolism	0.447699	1.88E-06
1-stearoyl-2-oleoyl-GPC (18:0/18:1)	Phospholipid Metabolism	0.279019	4.51E-06
1-linolenoyl-GPC (18:3)*	Lysolipid	0.414819	1.10E-05
1-linoleoyl-2-linolenoyl-GPC (18:2/18:3)*	Phospholipid Metabolism	0.537975	1.11E-05
1-palmitoyl-2-linoleoyl-GPI (16:0/18:2)	Phospholipid Metabolism	0.447877	5.88E-05
1-palmitoyl-GPI (16:0)	Lysolipid	0.438221	0.000101

Indolelactate	Tryptophan Metabolism	0.30948	0.000178
3-methylxanthine	Xanthine Metabolism	0.788924	0.00021
1,2-dilinoleoyl-GPC (18:2/18:2)	Phospholipid Metabolism	0.324133	0.000225
1-lignoceroyl-GPC (24:0)	Lysolipid	0.321129	0.000287
1-palmitoyl-2-stearoyl-GPC (16:0/18:0)	Phospholipid Metabolism	0.222855	0.000322
N-acetylcarnosine	Dipeptide Derivative	-0.33185	0.000873
1-stearoyl-2-oleoyl-GPI (18:0/18:1)*	Phospholipid Metabolism	0.346165	0.001026
N-acetylmethionine	Methionine, Cysteine, SAM and Taurine Metabolism	-0.58119	0.001445
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)	Phospholipid Metabolism	0.153562	0.002983
argininate*	Urea cycle; Arginine and Proline Metabolism	0.422405	0.003294
7-methylxanthine	Xanthine Metabolism	0.648043	0.004023
Homoarginine	Urea cycle; Arginine and Proline Metabolism	-0.27429	0.006606
gamma-glutamylvaline	Gamma-glutamyl Amino Acid	-0.3052	0.008009
sphingosine 1-phosphate	Sphingolipid Metabolism	-0.20846	0.008168
phenyllactate (PLA)	Phenylalanine and Tyrosine Metabolism	0.306398	0.009708
arabitol/xylitol	Pentose Metabolism	0.23942	0.015147
1-palmitoleoyl-GPC (16:1)*	Lysolipid	0.229408	0.017685
methionine sulfone	Methionine, Cysteine, SAM and Taurine Metabolism	0.308995	0.02004
Guanidinoacetate	Creatine Metabolism	-0.22401	0.035446
1-stearoyl-2-linoleoyl-GPI (18:0/18:2)	Phospholipid Metabolism	0.261839	0.036305
sphingomyelin (d18:2/14:0, d18:1/14:1)*	Sphingolipid Metabolism	0.216635	0.036711
4-cholesten-3-one	Sterol	0.242711	0.037246
1-palmitoyl-GPG (16:0)*	Lysolipid	0.309379	0.040079
Cholate	Primary Bile Acid Metabolism	1.182236	0.041373
1-palmitoyl-GPE (16:0)	Lysolipid	0.230631	0.049265
1-stearoyl-2-linoleoyl-GPC (18:0/18:2)*	Phospholipid Metabolism	0.118022	0.052877

Enrichment analysis discovered an over-representation of sterols (p=0.005), gamma-glutamyl amino acids (p=0.000846), lysolipids (p=0.00042) and phospholipids (p=0.00042) among the most significantly correlated metabolites associated with power. Other significantly altered metabolites in high vs. moderate power classes included 3-methylxanthine, cholate, guanidinoacetate, N-acetylcarnosine, imidazole lactate, and indolelactate (Table 20).

The FDR significant metabolites in association with power included a decrease in creatinine (estimate=-0.1, p=0.002) and an increase in creatine (estimate=0.6, p=0.001) although this result was not confirmed with Bonferroni significance. FDR significant metabolites with p values of less than 0.01 were projected on the heatmap in Figure 30. The heatmap summarizes a snapshot of the actual intensities of significant metabolites associated with high power athletes after correcting for covariates in the linear regression model described above. Samples were arranged within their respective sport classes (moderate power vs. high power) considering their sports type.

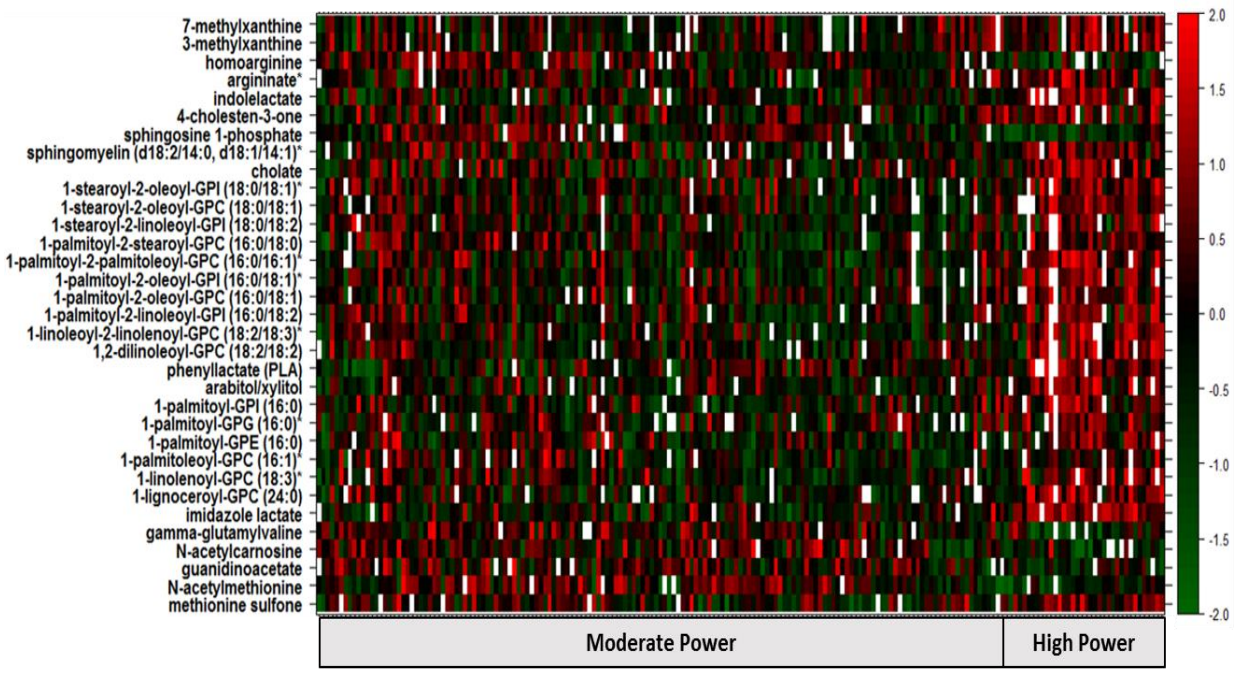


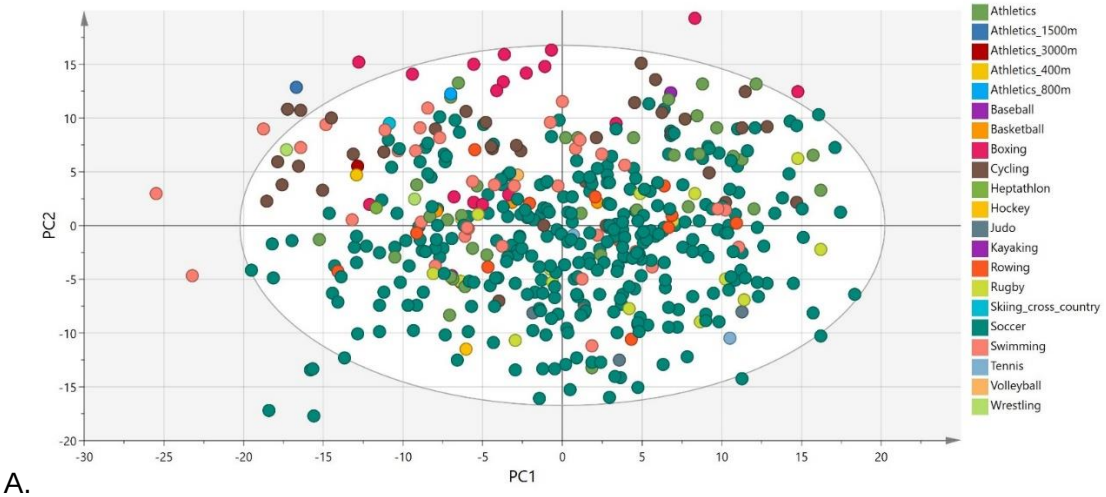
Figure 30. Heatmap of high power associated significant metabolites (y axis). Samples were grouped into moderate vs. high power groups (x axis). The color code represents scaled values of metabolites after correction of covariates.

Confirmation of metabolites associated with predefined sport classes in a 2nd cohort of 490 elite athletes

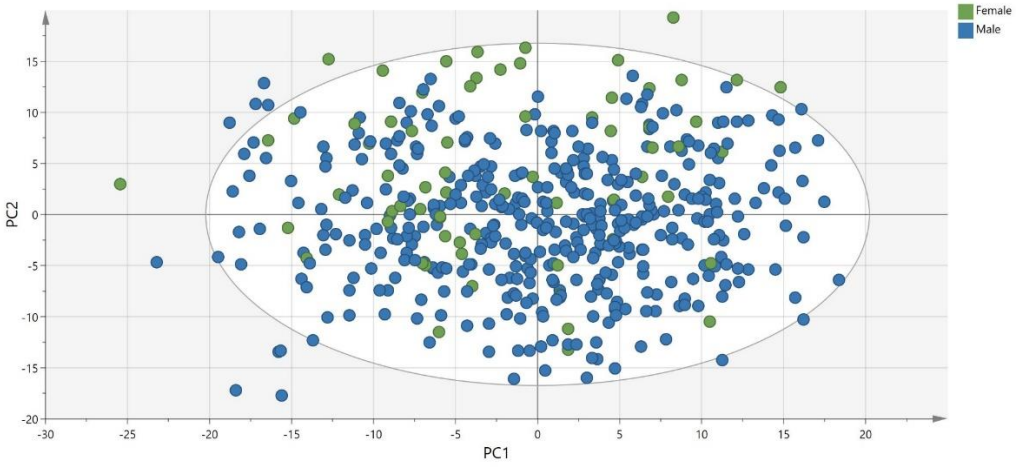
4.3.3 Multivariate analysis of athlete metabolomics data cohort 2

4.3.3.1 PCA 2nd cohort

Non-targeted metabolomics was applied to determine the metabolic signatures of 490 elite athletes. PCA revealed no clusters of samples influenced by sport types (Figure 31A), gender (Figure 31B), or study groups (Figure 32).

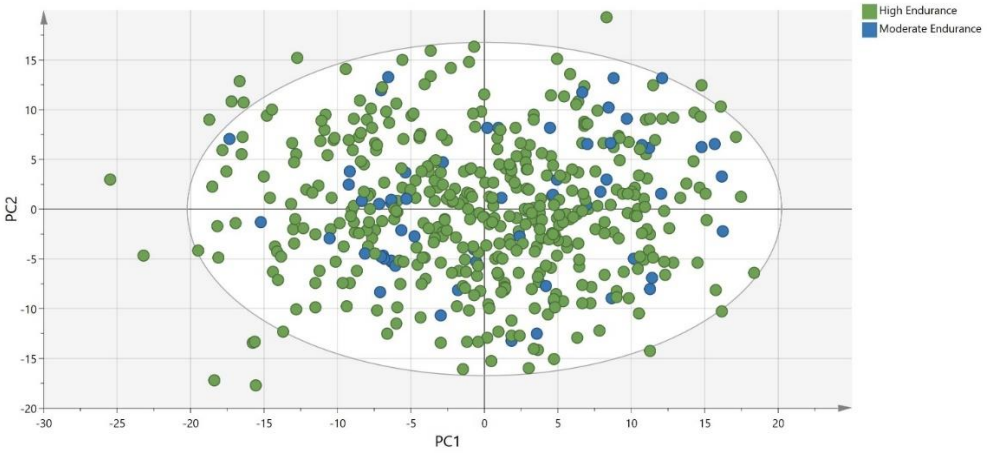


A.



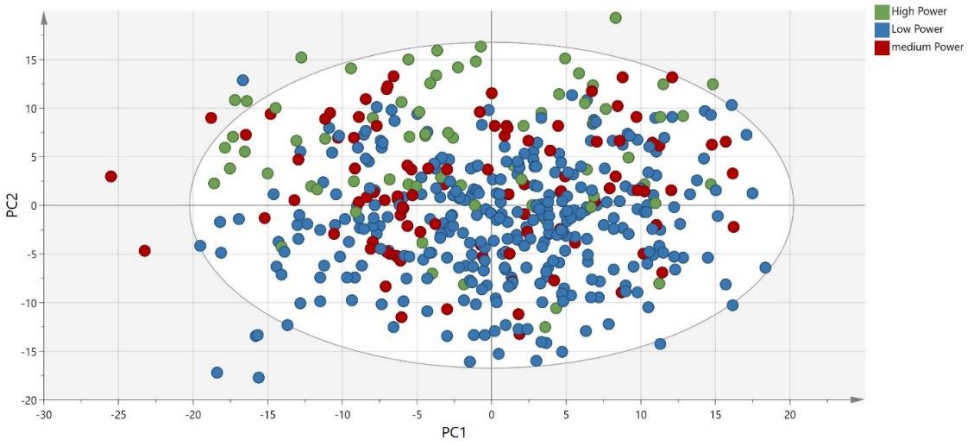
B.

Figure 31. A score plot of PC1 and PC2 signifying no clusters of samples influenced by (A) sport type, of (B) gender.



A.

B.



C.

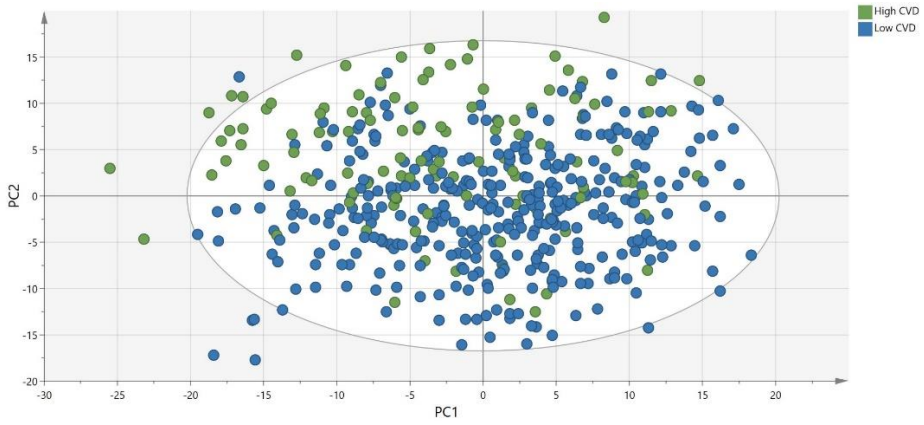


Figure 32. PCA analysis of 2nd cohort athlete metabolomics data. (A) A score plot of first two PCA components (PC1/PC2) indicating no population stratification of samples appeared due to difference in grouping (A) high vs. moderate endurance, (B) high, medium and low power, and (C) high vs. low cardiovascular demand.

4.3.3.2 OPLS-DA Endurance Group replication for cohort 2

OPLS-DA identified a number of metabolites that best distinguish predefined high vs. moderate endurance classes of samples. Figure 33A revealed 44% class-discriminatory component accounting for variation in the data due to endurance level (R-squared-Y = 0.44, Q-squared = 0.19). Two orthogonal components that do not differentiate the classes were also identified. The corresponding loading score, shown in Figure 33B, indicates a number of metabolites associated with higher endurance levels.

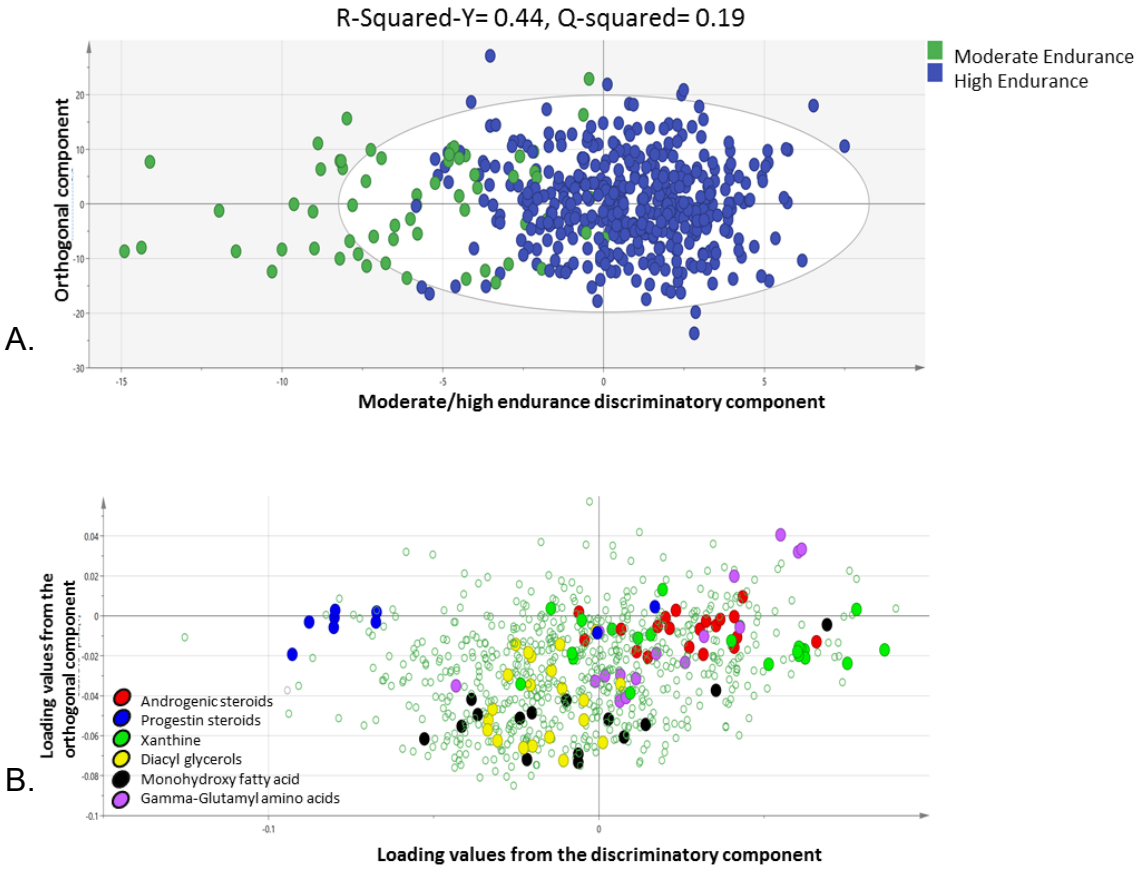


Figure 33. OPLS-DA model comparing high versus moderate endurance classes of elite athletes. (A) A score plot showing the orthogonal component (y-axis) vs. class-discriminatory component (x-axis). (B) The corresponding loading plot showing a clustering of gamma-glutamyl amino acids and androgenic steroids at the high end of endurance opposed by progestin steroids clustering at the negative end.

4.3.3.3 OPLS-DA Power Group for replication cohort 2

OPLS-DA identified number of metabolites that best distinguish predefined high vs. moderate power groups of samples. Figure 34A revealed 65% class-discriminatory component accounting for the variation in the data due to power level (R-squared-Y = 0.65, Q-squared = 0.55). Two orthogonal components that do not differentiate the

classes were also identified. The corresponding loading score, shown in Figure 34B, indicates a number of metabolites associated with higher power levels.

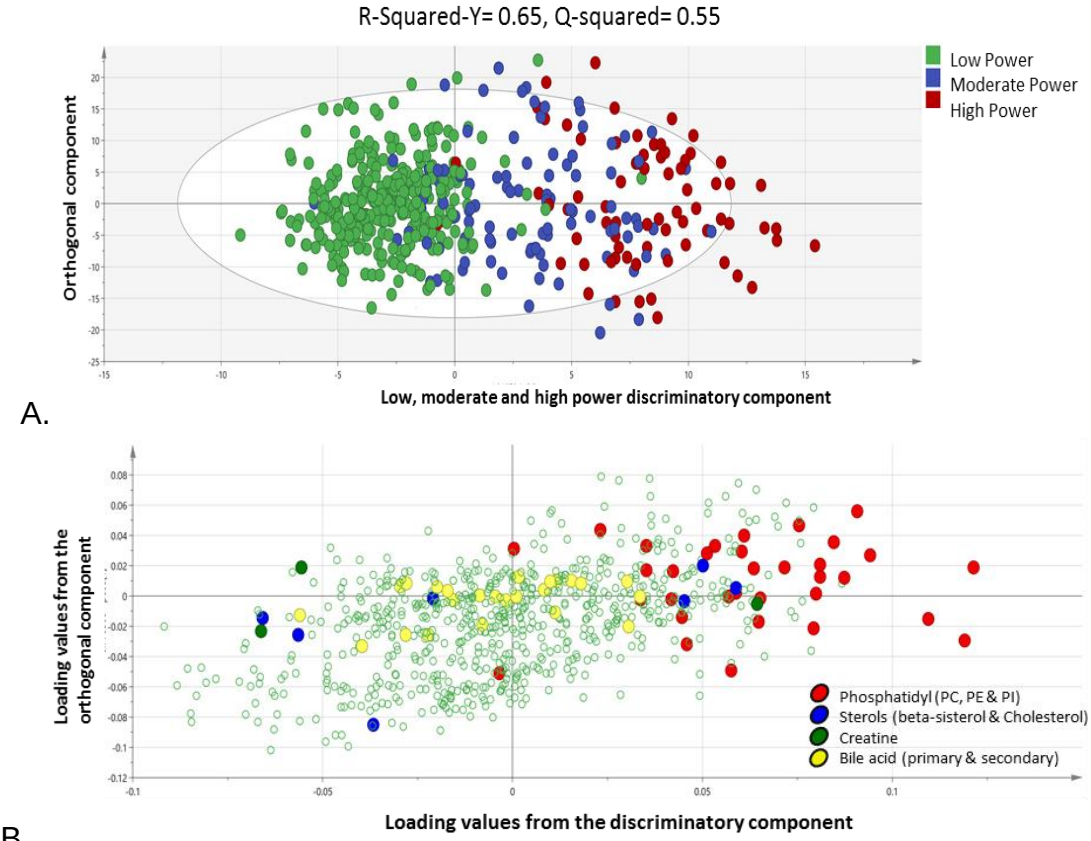


Figure 34. OPLS-DA model of high versus moderate power classes of elite athletes. (A) A score plot showing the orthogonal component (y-axis) and the class-discriminatory component (x-axis). (B) The corresponding loading plot showing a clustering of phosphatidyl (PC, PE and PI) at the high end of power as opposed to enrichment of gamma-glutamyl amino acids at the moderate and low end of power.

4.3.4 Univariate association tests and meta-analysis

A linear regression model was used to identify the significance of metabolite-associations in 2nd cohort of the athletes and their associated pathways listed below.

4.3.4.1 Endurance-associated metabolites in a 2nd cohort 490 elite athletes

Table 21 lists common metabolites-associated with endurance between the two cohorts. Among the metabolites identified, leucine, isoleucine and valine metabolites, phenylalanine and tyrosine metabolites, steroid, TCA cycle citrate and xanthine metabolites were increased with endurance. Gamma-glutamyl amino acid, fatty acid acyl carnitines, diacylglycerols, and glutamate metabolites were decreased with endurance.

Table 21. Common metabolites associated with endurance in previously reported cohort 1[167] and cohort 2.

No.	Metabolite	Sub-pathway	Cohort 1			Cohort 2		
			Fold Change	Nominal p value	FDR p value	Fold Change	Nominal p value	FDR p value
1	1-methylhistidine	Histidine Metabolism	-0.17	0.0001942	0.003005	-0.21	0.0268745	0.1708256
2	1-methylxanthine	Xanthine Metabolism	0.65	0.0001632	0.0026446	0.38	0.0370919	0.2025018
3	21-hydroxypregnenolone disulfate	Steroid	0.37	1.38E-07	3.27E-05	0.22	0.0425672	0.2208807
4	2-aminoadipate	Lysine Metabolism	-0.28	0.0010199	0.0093355	-0.32	0.0034942	0.0576194
5	2-aminobutyrate	Methionine, Cysteine, SAM and Taurine Metabolism	-0.15	0.0195432	0.0710495	-0.22	0.0019498	0.0384249
6	2-hydroxy-3-methylvalerate	Leucine, Isoleucine and Valine Metabolism	0.17	0.0084141	0.0382818	0.20	0.008995	0.0984927
7	3-(4-hydroxyphenyl)lactate	Phenylalanine and Tyrosine Metabolism	0.19	0.0016729	0.0131465	0.35	1.26E-05	0.0022683
8	5-hydroxylysine	Lysine Metabolism	-0.40	5.22E-05	0.0011279	-0.72	0.0462509	0.2292681
9	alpha-hydroxycaproate	Fatty Acid, Monohydroxy	0.23	0.0093646	0.0411619	0.40	0.0017179	0.0365281
10	alpha-hydroxyisovalerate	Leucine, Isoleucine and Valine Metabolism	0.14	0.0420618	0.1203092	0.32	2.27E-05	0.0030514
11	androstenediol (3alpha, 17alpha) monosulfate (2)	Steroid	0.21	0.0100732	0.0435385	0.22	0.0314248	0.1788115
12	citrate	TCA Cycle	0.20	2.29E-06	0.0001983	0.16	0.0063134	0.0809722
13	cortisol	Steroid	0.47	1.28E-05	0.0004531	0.44	0.0001974	0.009382
14	cortisone	Steroid	0.40	1.92E-06	0.0001863	0.26	0.0016599	0.0365281
15	dihomo-linoleoylcarnitine (C20:2)*	Fatty Acid Metabolism(Acyl Carnitine)	-0.33	6.46E-06	0.000339	-0.23	0.0075783	0.0913924
16	etiocholanolone glucuronide	Steroid	0.30	0.0002032	0.0030403	0.22	0.0427989	0.2208807

17	gamma-glutamylisoleucine*	Gamma-glutamyl Amino Acid	-0.20	0.0061989	0.0299547	-0.23	0.0257817	0.1692941
18	gamma-glutamyltryptophan	Gamma-glutamyl Amino Acid	-0.17	0.0020108	0.014714	-0.31	9.91E-05	0.006676
19	gamma-glutamyltyrosine	Gamma-glutamyl Amino Acid	-0.16	0.0024371	0.0160682	-0.18	0.0077223	0.0917588
20	glutamate	Glutamate Metabolism	-0.13	0.001245	0.010644	-0.09	0.0450395	0.2260367
21	linolenoylcarnitine (C18:3)*	Fatty Acid Metabolism(Acyl Carnitine)	-0.29	0.005164	0.0263192	-0.26	0.0175486	0.1350404
22	linoleoylcarnitine (C18:2)*	Fatty Acid Metabolism(Acyl Carnitine)	-0.30	0.0003551	0.0041861	-0.25	0.0014793	0.0351288
23	linoleoyl-linoleoyl-glycerol (18:2/18:2) [1]*	Diacylglycerol	-0.54	7.49E-06	0.000339	-0.34	0.0044899	0.0682119
24	N1-methyladenosine	Purine Metabolism, Adenine containing	0.12	4.74E-05	0.0010844	0.09	0.0104108	0.1028217
25	oleoyl-linoleoyl-glycerol (18:1/18:2) [1]	Diacylglycerol	-0.27	0.0003619	0.004202	-0.28	0.0001125	0.0069903
26	oleoyl-linoleoyl-glycerol (18:1/18:2) [2]	Diacylglycerol	-0.30	0.0001279	0.0022114	-0.28	0.0001409	0.0081319
27	phenyllactate (PLA)	Phenylalanine and Tyrosine Metabolism	0.19	0.0017918	0.0137857	0.34	3.86E-05	0.0041175
28	pyrraline	Food Component/Plant	0.35	0.0030817	0.0188784	0.51	0.0004289	0.015474
29	sphingosine 1-phosphate	Sphingolipid Metabolism	-0.16	0.0001031	0.0019098	-0.08	0.0491289	0.2379558
30	trans-4-hydroxyproline	Urea cycle; Arginine and Proline Metabolism	-0.28	3.80E-06	0.0002462	-0.22	0.0005916	0.019121
31	vanillactate	Phenylalanine and Tyrosine Metabolism	0.21	3.62E-05	0.000889	0.20	0.0141498	0.1216283
32	vanillylmandelate (VMA)	Phenylalanine and Tyrosine Metabolism	0.42	2.95E-08	1.15E-05	0.23	0.0321285	0.1815373

Among confirmed hits, elevation of pregnenolone and androgenic steroids indicate active steroid biosynthesis pathway in high endurance athletes. Reduced diacylglycerols and acyl carnitines and increased monohydroxy fatty acids suggest active fatty acid oxidation for energy generation in the high endurance group. Reduction in glutathione metabolism

and gamma glutamyl amino acids suggests active oxidative scavenging mechanisms in moderate endurance group.

4.3.4.2 Meta-analysis of endurance-associated metabolites of elite athletes

A meta-analysis shows a list of metabolites identified in both cohorts in association with endurance sports (Table 22).

Table 22. List of metabolites identified by meta-analysis in discovery and replication cohorts in association with endurance sports.

Confirmed endurance metabolites	Effect size (mean difference)	SE	p value	FDR
N6,N6,N6-trimethyllysine	-0.42	0.07	8.29E-10	5.95E-07
cortisol	0.53	0.09	9.40E-09	3.37E-06
3-(4-hydroxyphenyl)lactate	0.36	0.07	5.47E-08	1.31E-05
vanillylmandelate (VMA)	0.44	0.08	7.67E-08	1.38E-05
phenyllactate (PLA)	0.36	0.07	2.05E-07	2.79E-05
cortisone	0.37	0.07	2.33E-07	2.79E-05
oleoyl-linoleoyl-glycerol (18:1/18:2) [2]	-0.34	0.07	2.92E-07	2.99E-05
oleoyl-linoleoyl-glycerol (18:1/18:2) [1]	-0.33	0.07	4.10E-07	3.28E-05
linoleoyl-linoleoyl-glycerol (18:2/18:2) [1]*	-0.48	0.09	4.11E-07	3.28E-05
trans-4-hydroxyproline	-0.30	0.06	5.78E-07	4.15E-05
gamma-glutamyltryptophan	-0.35	0.07	8.33E-07	5.44E-05
21-hydroxypregnenolone disulfate	0.42	0.09	9.70E-07	5.80E-05
imidazole lactate	0.32	0.06	1.15E-06	6.33E-05
alpha-hydroxyisovalerate	0.32	0.07	2.11E-06	0.000108272
pyrraline	0.49	0.11	3.46E-06	0.000165834
5-hydroxylysine	-0.65	0.14	3.98E-06	0.000178779
dihomo-linoleoylcarnitine (C20:2)*	-0.35	0.08	4.94E-06	0.00020862
linoleoylcarnitine (C18:2)*	-0.32	0.07	5.72E-06	0.000228087
indolelactate	0.25	0.06	8.18E-06	0.000309153
citrate	0.24	0.05	9.83E-06	0.000352781
linoleoyl-linolenoyl-glycerol (18:2/18:3) [2]*	-0.46	0.11	1.15E-05	0.000394535
glutamine	0.22	0.05	1.39E-05	0.000452449
2-aminoadipate	-0.39	0.09	1.52E-05	0.000474453
vanillactate	0.29	0.07	1.84E-05	0.000550283
1-methylxanthine	0.53	0.12	2.18E-05	0.000626386
alpha-hydroxycaproate	0.44	0.11	4.23E-05	0.001167417
leukotriene B4	-0.37	0.09	5.99E-05	0.001593257
palmitoyl-linoleoyl-glycerol (16:0/18:2) [2]*	-0.30	0.08	6.88E-05	0.001762982
1-methylhistidine	-0.31	0.08	9.61E-05	0.002379542
1,3,7-trimethylurate	0.52	0.13	0.00010533	0.002520897
etiocholanolone glucuronide	0.34	0.09	0.000118208	0.002737861
retinol (Vitamin A)	0.24	0.06	0.000134816	0.00302494
2-aminobutyrate	-0.24	0.06	0.000143623	0.003124897
N-(2-furoyl)glycine	0.50	0.13	0.000151907	0.003207922
pregnenolone sulfate	0.34	0.09	0.000169508	0.003351897
isovalerate	-0.32	0.09	0.000172222	0.003351897
quinate	0.53	0.14	0.00017273	0.003351897
glycochenodeoxycholate	-0.42	0.11	0.000184442	0.003484981
oleoylcarnitine (C18:1)	-0.24	0.06	0.00019533	0.003596078
gamma-glutamyltyrosine	-0.22	0.06	0.000255918	0.004593735
2-hydroxy-3-methylvalerate	0.24	0.07	0.000379078	0.006539206
linolenoylcarnitine (C18:3)*	-0.33	0.09	0.000382516	0.006539206
choline phosphate	-0.26	0.07	0.000527465	0.008807444
N1-methyladenosine	0.11	0.03	0.000582633	0.009507508

gamma-glutamylglutamate	-0.24	0.07	0.000607793	0.009614412
gamma-glutamylisoleucine*	-0.29	0.09	0.000615965	0.009614412
palmitoleoyl-linoleoyl-glycerol (16:1/18:2) [1]*	-0.41	0.12	0.000705891	0.010783618
3-methyladipate	0.34	0.10	0.000772111	0.011549491
5-acetylamino-6-amino-3-methyluracil	0.44	0.13	0.000823188	0.01206222
glycoursodeoxycholate	-0.38	0.12	0.000890413	0.012786331
pro-hydroxy-pro	-0.22	0.07	0.000939544	0.013227306
1,3-dimethylurate	0.40	0.12	0.001041754	0.01394916
12-HETE	-0.31	0.09	0.00104228	0.01394916
1-methylurate	0.40	0.12	0.001049101	0.01394916
behenoyl dihydrosphingomyelin (d18:0/22:0)*	-0.24	0.07	0.001071388	0.013986487
O-sulfo-L-tyrosine	0.21	0.06	0.001096605	0.014060046
alpha-tocopherol	0.38	0.12	0.001122833	0.014133875
androstenediol (3alpha, 17alpha) monosulfate (2)	0.27	0.08	0.001141734	0.014133875
maleate	0.23	0.07	0.001255341	0.015276866
bilirubin (Z,Z)	0.41	0.13	0.001314843	0.015486434
methionine sulfoxide	-0.25	0.08	0.0013157	0.015486434
gamma-glutamylphenylalanine	-0.21	0.07	0.001400255	0.016215862
heme	0.57	0.18	0.001474442	0.016615589
eicosenoylcarnitine (C20:1)*	-0.23	0.07	0.001481055	0.016615589
acetylcarnitine (C2)	0.23	0.07	0.001624224	0.017941428
glycerol	0.26	0.08	0.001699844	0.018492247
fructose	0.27	0.09	0.001825278	0.019560446
2,3-dihydroxy-2-methylbutyrate	0.30	0.10	0.002012051	0.021244895
chenodeoxycholate	-0.34	0.11	0.002109893	0.021955119
palmitoyl-linoleoyl-glycerol (16:0/18:2) [1]*	-0.24	0.08	0.002286723	0.023455248
4-guanidinobutanoate	-0.24	0.08	0.002421021	0.024459721
oleoyl-oleoyl-glycerol (18:1/18:1) [2]*	-0.24	0.08	0.002470579	0.024459721
sphingosine 1-phosphate	-0.12	0.04	0.002487542	0.024459721
gamma-glutamyl-epsilon-lysine	-0.21	0.07	0.002548374	0.024459721
retinal	-0.23	0.08	0.002554985	0.024459721
perfluorooctanesulfonic acid (PFOS)	0.25	0.08	0.002638042	0.024922557
sphingomyelin (d18:1/14:0, d16:1/16:0)*	0.11	0.04	0.002718038	0.025344826
arachidonoylcarnitine (C20:4)	-0.26	0.09	0.002815979	0.025921448
oleoyl-oleoyl-glycerol (18:1/18:1) [1]*	-0.23	0.08	0.002878101	0.026157929
sphingomyelin (d18:0/18:0, d19:0/17:0)*	-0.20	0.07	0.002943426	0.026417249
glutamate	-0.12	0.04	0.003273165	0.029013977
gamma-glutamylleucine	-0.22	0.08	0.003342624	0.029141469
4-cholesten-3-one	-0.19	0.07	0.003368721	0.029141469
allantoin	-0.24	0.08	0.003539135	0.030251175
theophylline	0.35	0.12	0.003942121	0.032936022
O-methylcatechol sulfate	0.33	0.12	0.003973132	0.032936022
gamma-glutamylthreonine	-0.19	0.07	0.00403623	0.032936022
indolepropionate	0.33	0.12	0.004036727	0.032936022
catechol sulfate	0.31	0.11	0.004153603	0.033067881
16a-hydroxy DHEA 3-sulfate	0.29	0.10	0.004188185	0.033067881
3-acetylphenol sulfate	0.37	0.13	0.004202051	0.033067881
androstenediol (3beta,17beta) disulfate (2)	0.21	0.08	0.00423711	0.033067881
1-stearoyl-2-arachidonoyl-GPI (18:0/20:4)	-0.12	0.04	0.00432257	0.033372103
N-acetylmethionine sulfoxide	-0.27	0.09	0.004666187	0.035641729
3-aminoisobutyrate	-0.20	0.07	0.004883327	0.036675872
N-behenoyl-sphingadienine (d18:2/22:0)*	0.17	0.06	0.004910577	0.036675872
pantothenate	0.20	0.07	0.004954818	0.036675872
4-HDoHE	-0.23	0.08	0.005074441	0.037099619
gamma-glutamylvaline	-0.20	0.07	0.005115407	0.037099619
sucrose	0.35	0.13	0.005192876	0.037284848
alanine	0.09	0.03	0.005679218	0.040373054
palmitoleoylcarnitine (C16:1)*	-0.23	0.08	0.005920206	0.041673609
isovalerylcarnitine (C5)	-0.22	0.08	0.006852712	0.047542467
urate	0.13	0.05	0.006886374	0.047542467

4.3.4.3 Power-associated metabolites in a 2nd cohort of 490 elite athletes

Table 23 lists common metabolites-associated with power group between the two cohorts. Among the hit metabolites: phenylalanine and tyrosine metabolism, histidine metabolism, secondary bile acid metabolism, steroid, fatty acid metabolism (acyl Carnitine), dipeptide derivative, gamma-glutamyl amino acid, urea cycle, arginine and proline metabolism, sphingolipid metabolism, and creatine metabolism were decreased with power. Whereas, lysolipid, phospholipid metabolism, plasmalogen, sphingolipid metabolism, diacylglycerol, histidine metabolism, benzoate metabolism and primary bile acid metabolism were increased with power.

Table 23. Common metabolites associated with power in discovery and replication cohorts 1.

No.	Metabolite	Sub-pathway	Cohort DS1			Cohort DS2		
			Fold Change	Nominal	FDR	Fold Change	Nominal	FDR
1	1-(1-enyl-stearoyl)-2-linoleoyl-GPE (P-18:0/18:2)*	Plasmalogen	0.2	0.004498	0.02904	0.11	0.043037989	0.199898039
2	1,2-dilinoleoyl-GPC (18:2/18:2)	Phospholipid Metabolism	0.32	0.000000289	0.0000205	0.13	0.003501268	0.038096611
3	1,2-dipalmitoyl-GPC (16:0/16:0)	Phospholipid Metabolism	0.13	0.000127	0.002295	0.06	0.024974045	0.139373821
4	1,2-dipalmitoyl-GPE (16:0/16:0)*	Phospholipid Metabolism	0.36	0.000284	0.003874	0.16	0.043399688	0.199898039
5	1-linolenoyl-GPC (18:3)*	Lysolipid	0.41	1.41E-08	0.00000185	0.32	4.99E-07	6.80E-05
6	1-linoleoyl-2-arachidonoyl-GPC (18:2/20:4n6)*	Phospholipid Metabolism	0.18	0.006016	0.036002	0.13	0.000878327	0.017961791
7	1-linoleoyl-GPE (18:2)*	Lysolipid	0.24	0.003574	0.024606	0.19	0.001619221	0.025471595
8	1-linoleoyl-GPI (18:2)*	Lysolipid	0.24	0.000903	0.008675	0.16	0.034493041	0.17009489
9	1-oleoyl-GPC (18:1)	Lysolipid	0.11	0.002504	0.018072	0.07	0.008436737	0.071888032
10	1-palmitoleoyl-GPC (16:1)*	Lysolipid	0.23	0.0000227	0.000707	0.27	2.59E-10	7.07E-08
11	1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*	Phospholipid Metabolism	0.38	0.0000769	0.001616	0.17	0.002757743	0.031938263
12	1-palmitoyl-2-linoleoyl-GPC (16:0/18:2)	Phospholipid Metabolism	0.12	0.0000761	0.001616	0.05	0.025046396	0.139373821

13	1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	Phospholipid Metabolism	0.33	0.000713	0.0075	0.25	0.000624157	0.013435799
14	1-palmitoyl-2-linoleoyl-GPI (16:0/18:2)	Phospholipid Metabolism	0.45	7.55E-08	0.00000839	0.23	0.000963554	0.018766359
15	1-palmitoyl-2-oleoyl-GPC (16:0/18:1)	Phospholipid Metabolism	0.15	0.00000383	0.000175	0.12	6.76E-06	0.000502435
16	1-palmitoyl-2-oleoyl-GPE (16:0/18:1)	Phospholipid Metabolism	0.34	0.000166	0.002804	0.25	0.000401827	0.010016899
17	1-palmitoyl-2-oleoyl-GPI (16:0/18:1)*	Phospholipid Metabolism	0.42	1.42E-10	5.52E-08	0.27	1.07E-06	9.74E-05
18	1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	Phospholipid Metabolism	0.58	7.61E-14	5.92E-11	0.48	8.55E-14	6.99E-11
19	1-palmitoyl-2-stearoyl-GPC (16:0/18:0)	Phospholipid Metabolism	0.22	0.000000414	0.0000248	0.09	0.002401021	0.030688054
20	1-palmitoyl-GPE (16:0)	Lysolipid	0.23	0.0000633	0.001493	0.12	0.008780804	0.072552502
21	1-palmitoyl-GPI (16:0)	Lysolipid	0.44	0.00000013	0.0000127	0.27	0.001666624	0.025722613
22	1-stearoyl-2-oleoyl-GPC (18:0/18:1)	Phospholipid Metabolism	0.28	5.80E-09	0.00000113	0.15	6.33E-05	0.002466895
23	1-stearoyl-2-oleoyl-GPI (18:0/18:1)*	Phospholipid Metabolism	0.35	0.00000132	0.0000684	0.16	0.007861258	0.068409673
24	1-stearoyl-GPC (18:0)	Lysolipid	0.09	0.003811	0.025766	0.05	0.041916352	0.197055035
25	2,3-dihydroxy-2-methylbutyrate	Leucine, Isoleucine and Valine Metabolism	0.31	0.001082	0.009659	0.27	0.022860301	0.134530403
26	2-aminobutyrate	Methionine, Cysteine, SAM and Taurine Metabolism	-0.24	0.001529	0.013073	-0.12	0.034518034	0.17009489
27	2-palmitoleoyl-GPC (16:1)*	Lysolipid	0.22	0.001009	0.009407	0.44	9.05E-08	1.48E-05
28	3-(4-hydroxyphenyl)lactate	Phenylalanine and Tyrosine Metabolism	0.25	0.000258	0.003718	0.2	0.001701722	0.025777941
29	3,7-dimethylurate	Xanthine Metabolism	0.44	0.000739	0.007671	0.26	0.041016332	0.193938493
30	3beta,7alpha-dihydroxy-5-cholestenoate	Sterol	-0.25	0.000096	0.001822	-0.22	0.001082166	0.020317242
31	3-methoxycatechol sulfate (1)	Benzoate Metabolism	0.65	0.00088	0.008561	0.33	0.03625756	0.176539785
32	3-methoxytyrosine	Phenylalanine and Tyrosine Metabolism	-0.13	0.003229	0.022427	-0.12	0.002235459	0.03030887

33	3-methylhistidine	Histidine Metabolism	-0.52	0.001887	0.014983	-0.57	0.001233883	0.021474811
34	3-methylxanthine	Xanthine Metabolism	0.79	0.00000027	0.0000205	0.31	0.013594715	0.098411296
35	5,6-dihydrothymine	Pyrimidine Metabolism, Thymine containing	-0.19	0.000171	0.002838	-0.11	0.026487848	0.145618436
36	5alpha-pregnan-3beta,20alpha-diol disulfate	Steroid	-0.45	0.000177	0.002864	-0.35	0.001386678	0.022988455
37	7-methylxanthine	Xanthine Metabolism	0.65	0.00000517	0.000212	0.37	0.002471226	0.031095215
38	argininate*	Urea cycle; Arginine and Proline Metabolism	0.42	0.00000423	0.000183	0.15	0.030046587	0.156548462
39	campesterol	Sterol	0.31	0.0000948	0.001822	0.26	0.006405801	0.058642181
40	cholate	Primary Bile Acid Metabolism	1.18	0.0000532	0.001293	0.61	0.010428988	0.082652704
41	creatinine	Creatine Metabolism	-0.1	0.000283	0.003874	-0.05	0.01512514	0.104103638
42	decanoylcarnitine (C10)	Fatty Acid Metabolism(Acyl Carnitine)	-0.35	0.006458	0.037776	-0.44	4.62E-05	0.001888923
43	gamma-glutamylisoleucine*	Gamma-glutamyl Amino Acid	-0.3	0.000274	0.003874	-0.26	0.001786051	0.026563453
44	gamma-glutamylleucine	Gamma-glutamyl Amino Acid	-0.32	0.000131	0.002313	-0.2	0.006452074	0.058642181
45	gamma-glutamylvaline	Gamma-glutamyl Amino Acid	-0.31	0.0000103	0.000371	-0.16	0.014932113	0.104103638
46	guanidinosuccinate	Guanidino and Acetamido Metabolism	0.38	0.000442	0.00537	0.19	0.04893162	0.215429515
47	hexanoylcarnitine (C6)	Fatty Acid Metabolism(Acyl Carnitine)	-0.23	0.008253	0.045864	-0.29	0.000207211	0.006277735
48	homoarginine	Urea cycle; Arginine and Proline Metabolism	-0.27	0.00000849	0.00033	-0.18	0.002945385	0.033462844
49	imidazole lactate	Histidine Metabolism	0.45	2.41E-09	0.000000626	0.3	6.24E-07	7.29E-05
50	indolelactate	Tryptophan Metabolism	0.31	0.000000228	0.0000198	0.25	9.92E-07	9.74E-05
51	isoleucine	Leucine, Isoleucine and Valine Metabolism	-0.13	0.007439	0.042557	-0.11	0.002622946	0.031095215
52	N-acetylcarnosine	Dipeptide Derivative	-0.33	0.00000112	0.0000624	-0.3	4.56E-05	0.001888923
53	N-acetylthreonine	Glycine, Serine and	-0.13	0.000659	0.00715	-0.17	0.000290296	0.008188356

		Threonine Metabolism						
54	octanoylcarnitine (C8)	Fatty Acid Metabolism(Acyl Carnitine)	-0.31	0.008981	0.048188	-0.38	0.000134062	0.004388485
55	palmitoyl-arachidonoyl-glycerol (16:0/20:4) [2]*	Diacylglycerol	0.41	0.000856	0.008432	0.2	0.024353165	0.138339508
56	phenyllactate (PLA)	Phenylalanine and Tyrosine Metabolism	0.31	0.0000125	0.000422	0.25	0.000172735	0.005434511
57	pyrraline	Food Component/Plant	0.41	0.003075	0.021555	0.37	0.001405162	0.022988455
58	sphingomyelin (d18:2/14:0, d18:1/14:1)*	Sphingolipid Metabolism	0.22	0.0000472	0.001242	0.2	3.02E-05	0.001544014
59	sphingosine 1-phosphate	Sphingolipid Metabolism	-0.21	0.0000105	0.000371	-0.08	0.01840533	0.115812
60	taurocholenate sulfate	Secondary Bile Acid Metabolism	-0.49	0.002074	0.015974	-0.47	2.48E-05	0.001433649
61	thyroxine	Phenylalanine and Tyrosine Metabolism	-0.11	0.002828	0.020186	-0.12	0.002094122	0.029033752
62	tyramine O-sulfate	Phenylalanine and Tyrosine Metabolism	-0.75	0.001618	0.013392	-0.47	0.012859685	0.095009806

4.3.4.4 Meta-analysis of power-associated metabolites in a 2nd cohort of 490 elite athletes

A meta-analysis shows a list of metabolites identified in both cohorts in association with power sports (Table 24).

Table 24. List of metabolites identified by meta-analysis in current and previously published cohort in association with power sports.

Confirmed Power metabolites	Effect size (mean difference)	SE	p value	FDR
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	0.53	0.05	1.07E-28	7.81E-26
1-palmitoyl-2-oleoyl-GPI (16:0/18:1)*	0.34	0.04	2.33E-16	8.50E-14
imidazole lactate	0.36	0.05	3.02E-15	7.35E-13
1-palmitoleoyl-GPC (16:1)*	0.25	0.03	8.35E-15	1.52E-12
1-linolenoyl-GPC (18:3)*	0.36	0.05	8.48E-15	1.24E-12
indolelactate	0.27	0.04	3.74E-13	4.55E-11
1-stearoyl-2-oleoyl-GPC (18:0/18:1)	0.20	0.03	3.41E-12	3.56E-10
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)	0.13	0.02	6.52E-11	5.95E-09
N-acetylcarnosine	-0.32	0.05	8.40E-11	6.81E-09
1-palmitoyl-2-linoleoyl-GPI (16:0/18:2)	0.33	0.05	5.96E-10	4.35E-08
1-palmitoyl-GPI (16:0)	0.36	0.06	6.88E-10	4.57E-08
2-palmitoleoyl-GPC (16:1)*	0.31	0.05	1.49E-09	9.06E-08

sphingomyelin (d18:2/14:0, d18:1/14:1)*	0.20	0.03	3.28E-09	1.84E-07
phenyllactate (PLA)	0.28	0.05	5.24E-09	2.73E-07
1,2-dilinoleoyl-GPC (18:2/18:2)	0.20	0.04	3.2E-08	1.56E-06
1-palmitoyl-2-stearoyl-GPC (16:0/18:0)	0.13	0.02	3.93E-08	1.79E-06
7-methylxanthine	0.50	0.09	6.65E-08	2.86E-06
3-methylxanthine	0.51	0.10	8.44E-08	3.42E-06
homoarginine	-0.23	0.04	8.9E-08	3.42E-06
taurocholenate sulfate	-0.48	0.09	1.3E-07	4.75E-06
1-stearoyl-2-oleoyl-GPI (18:0/18:1)*	0.24	0.04	1.35E-07	4.69E-06
1-palmitoyl-2-oleoyl-GPE (16:0/18:1)	0.28	0.05	2.32E-07	7.70E-06
3beta,7alpha-dihydroxy-5-cholestenoate	-0.23	0.05	2.57E-07	8.16E-06
dihomo-linolenate (20:3n3 or n6)	0.23	0.05	6.11E-07	1.86E-05
N-acetylthreonine	-0.14	0.03	6.14E-07	1.79E-05
5alpha-pregnan-3beta,20alpha-diol disulfate	-0.39	0.08	7.18E-07	2.02E-05
decanoylcarnitine (C10)	-0.40	0.08	8.64E-07	2.34E-05
gamma-glutamylvaline	-0.23	0.05	8.69E-07	2.27E-05
3-(4-hydroxyphenyl)lactate	0.22	0.05	1.23E-06	3.10E-05
gamma-glutamylisoleucine*	-0.28	0.06	1.24E-06	3.02E-05
1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	0.28	0.06	1.37E-06	3.23E-05
campesterol	0.29	0.06	1.41E-06	3.22E-05
taurolithocholate 3-sulfate	-0.57	0.12	1.52E-06	3.36E-05
myristoleoylcarnitine (C14:1)*	-0.40	0.08	2.21E-06	4.75E-05
octanoylcarnitine (C8)	-0.35	0.08	3.38E-06	7.05E-05
1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*	0.22	0.05	3.54E-06	7.18E-05
cholate	0.85	0.18	3.75E-06	7.40E-05
gamma-glutamylleucine	-0.25	0.05	3.78E-06	7.26E-05
sphingosine 1-phosphate	-0.13	0.03	3.93E-06	7.36E-05
argininate*	0.25	0.05	3.98E-06	7.26E-05
1-palmitoyl-GPE (16:0)	0.16	0.03	4.31E-06	7.67E-05
hexanoylcarnitine (C6)	-0.27	0.06	5.03E-06	8.74E-05
3-methylhistidine	-0.54	0.12	6.08E-06	0.000103
pyrraline	0.39	0.09	1.13E-05	0.000187
N6,N6,N6-trimethyllysine	-0.22	0.05	1.24E-05	0.000201
ceramide (d18:1/20:0, d16:1/22:0, d20:1/18:0)*	0.15	0.03	1.28E-05	0.000203
1-lignoceroyl-GPC (24:0)	0.14	0.03	1.52E-05	0.000236
thyroxine	-0.12	0.03	1.52E-05	0.000231
1-linoleoyl-GPE (18:2)*	0.21	0.05	1.66E-05	0.000247
1-linoleoyl-2-arachidonoyl-GPC (18:2/20:4n6)*	0.14	0.03	1.78E-05	0.00026
3-methoxytyrosine	-0.12	0.03	1.91E-05	0.000273
5,6-dihydrothymine	-0.15	0.03	1.91E-05	0.000268
cortisol	0.32	0.08	2.67E-05	0.000368
creatinine	-0.07	0.02	2.76E-05	0.000373
1,2-dipalmitoyl-GPC (16:0/16:0)	0.08	0.02	3.03E-05	0.000402
1-palmitoyl-2-linoleoyl-GPC (16:0/18:2)	0.07	0.02	3.06E-05	0.000399
beta-sitosterol	0.23	0.06	5.33E-05	0.000683
isoleucine	-0.12	0.03	5.45E-05	0.000686
2,3-dihydroxy-2-methylbutyrate	0.30	0.07	5.73E-05	0.000709
myristoyl dihydrosphingomyelin (d18:0/14:0)*	0.13	0.03	0.000072	0.000876
1-oleoyl-GPC (18:1)	0.08	0.02	0.000074	0.000886
tyramine O-sulfate	-0.58	0.15	7.64E-05	0.0009
inosine	0.58	0.15	8.46E-05	0.00098
valine	-0.10	0.03	8.75E-05	0.000998
1-linoleoyl-GPI (18:2)*	0.20	0.05	0.000093	0.001044
1,2-dipalmitoyl-GPE (16:0/16:0)*	0.24	0.06	9.75E-05	0.001078
arabitol/xylitol	0.15	0.04	9.91E-05	0.00108
guanidinosuccinate	0.28	0.07	0.000104	0.001114
3,7-dimethylurate	0.35	0.09	0.00011	0.00116
cis-4-decenoylcarnitine (C10:1)	-0.28	0.07	0.000111	0.001154
heptanoate (7:0)	0.21	0.05	0.000122	0.00125
palmitoyl-arachidonoyl-glycerol (16:0/20:4) [2]*	0.28	0.07	0.000124	0.001255
1-oleoyl-GPI (18:1)*	0.23	0.06	0.000125	0.001253
12,13-DiHOME	0.27	0.07	0.000132	0.001304
laurate (12:0)	0.32	0.08	0.00014	0.001359
1-palmitoyl-2-arachidonoyl-GPC (16:0/20:4n6)	0.09	0.02	0.000148	0.001426
1-palmitoyl-GPG (16:0)*	0.19	0.05	0.000156	0.001481
1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4)*	0.19	0.05	0.000157	0.001472

3-methoxycatechol sulfate (1)	0.46	0.12	0.000161	0.001491
1-stearoyl-2-linoleoyl-GPI (18:0/18:2)	0.14	0.04	0.00018	0.00164
4-cholesten-3-one	0.16	0.04	0.000219	0.001978
pantothenate	0.18	0.05	0.000236	0.0021
glycerophosphoethanolamine	0.11	0.03	0.000245	0.002159
stearoyl sphingomyelin (d18:1/18:0)	-0.09	0.02	0.000256	0.002229
docosapentaenoate (n6 DPA; 22:5n6)	0.18	0.05	0.000259	0.002221
2-aminobutyrate	-0.16	0.04	0.000281	0.002381
glycolithocholate sulfate*	-0.40	0.11	0.000312	0.002621
oleoylcarnitine (C18:1)	-0.17	0.05	0.000347	0.002878
1-stearoyl-GPE (18:0)	0.11	0.03	0.000356	0.002922
2-aminoadipate	-0.23	0.06	0.00037	0.003003
2-stearoyl-GPE (18:0)*	0.15	0.04	0.000419	0.003358
kynurenine	0.09	0.03	0.000427	0.003388
laurylcarnitine (C12)	-0.27	0.08	0.00044	0.003457
4-hydroxychlorothalonil	0.20	0.06	0.000516	0.004007
methionine sulfone	0.15	0.04	0.000525	0.004034
1-palmitoleoylglycerol (16:1)*	0.22	0.06	0.000526	0.003998
1-stearoyl-GPC (18:0)	0.07	0.02	0.000566	0.004258
2-methylbutyrylcarnitine (C5)	-0.17	0.05	0.000579	0.004312
1-palmitoleoyl-GPA (16:1)*	0.33	0.10	0.000583	0.004298
theobromine	0.39	0.11	0.000603	0.004399
suberate (octanedioate)	0.18	0.05	0.000604	0.004363
oleoyl-oleoyl-glycerol (18:1/18:1) [1]*	-0.20	0.06	0.000607	0.004346
lactosyl-N-palmitoyl-sphingosine (d18:1/16:0)	-0.08	0.02	0.000641	0.004544
1-linoleoyl-GPG (18:2)*	0.16	0.05	0.000668	0.004688
1-linoleoyl-GPC (18:2)	0.08	0.02	0.000686	0.004768
gamma-glutamyl-alpha-lysine	-0.18	0.05	0.000711	0.004899
5alpha-pregnan-3beta,20alpha-diol monosulfate (2)	-0.26	0.08	0.000778	0.005311
sphingomyelin (d18:2/18:1)*	-0.09	0.03	0.000779	0.005266
creatine	0.26	0.08	0.000779	0.00522
1-(1-enyl-stearoyl)-2-linoleoyl-GPE (P-18:0/18:2)*	0.14	0.04	0.000789	0.005239
N-behenoyl-sphingadienine (d18:2/22:0)*	0.13	0.04	0.000816	0.005368
1-stearoyl-2-linoleoyl-GPC (18:0/18:2)*	0.06	0.02	0.000849	0.005532
guanosine	0.32	0.10	0.000859	0.005547
oleoyl-linoleoyl-glycerol (18:1/18:2) [1]	-0.16	0.05	0.000887	0.005679
1-linoleoyl-2-linolenoyl-GPC (18:2/18:3)*	0.20	0.06	0.000952	0.006045
palmitoleoyl-arachidonoyl-glycerol (16:1/20:4) [2]*	0.32	0.10	0.000965	0.006071
3-hydroxyhexanoate	-0.17	0.05	0.000979	0.006106
N-acetylglucosamine/N-acetylgalactosamine	-0.11	0.03	0.001137	0.007034
5-methyluridine (ribothymidine)	-0.09	0.03	0.001153	0.007072
palmitoyl-arachidonoyl-glycerol (16:0/20:4) [1]*	0.30	0.09	0.001185	0.007209
S-methylcysteine	-0.20	0.06	0.001204	0.007265
leucine	-0.09	0.03	0.001245	0.007451
C-glycosyltryptophan	-0.09	0.03	0.001334	0.007916
hypoxanthine	0.19	0.06	0.001364	0.008032
ursodeoxycholate	0.44	0.14	0.001433	0.00837
1-myristoylglycerol (14:0)	0.25	0.08	0.001451	0.008407
4-vinylguaiacol sulfate	-0.51	0.16	0.001451	0.008343
1,5-anhydroglucitol (1,5-AG)	-0.10	0.03	0.001519	0.008666
5-(galactosylhydroxy)-L-lysine	-0.14	0.04	0.001525	0.008627
3-methylglutaconate	0.15	0.05	0.00154	0.008648
3beta-hydroxy-5-cholestenoate	-0.13	0.04	0.001541	0.008588
N6-carboxymethyllysine	0.20	0.06	0.001575	0.008708
N-acetylisoleucine	-0.16	0.05	0.001586	0.008707
1-oleoyl-2-linoleoyl-GPE (18:1/18:2)*	0.22	0.07	0.001706	0.009295
propionylglycine	0.19	0.06	0.001787	0.009665
cytidine	-0.16	0.05	0.001794	0.009632
sphingomyelin (d18:1/14:0, d16:1/16:0)*	0.08	0.02	0.001866	0.00994
13-HODE + 9-HODE	0.31	0.10	0.002202	0.011647
sphingomyelin (d18:1/25:0, d19:0/24:1, d20:1/23:0, d19:1/24:0)*	-0.12	0.04	0.002231	0.011716
indoleacetylglutamine	0.27	0.09	0.002275	0.011865
pyroglutamine*	-0.18	0.06	0.002389	0.012369
glutarylcarnitine (C5-DC)	-0.14	0.05	0.002487	0.012784
isovalerylcarnitine (C5)	-0.18	0.06	0.002487	0.012697

epiandrosterone sulfate	-0.20	0.07	0.00261	0.013229
eicosenoylcarnitine (C20:1)*	-0.15	0.05	0.00272	0.013692
N-stearoyltaurine	0.17	0.06	0.002852	0.014259
2-hydroxyglutarate	0.14	0.05	0.003087	0.015329
gamma-glutamylthreonine	-0.14	0.05	0.003114	0.01536
arachidonate (20:4n6)	0.12	0.04	0.003116	0.015268
adenine	0.19	0.06	0.003121	0.015189
1-(1-enyl-palmitoyl)-2-oleoyl-GPE (P-16:0/18:1)*	0.10	0.03	0.003136	0.015159
gamma-glutamyltryptophan	-0.13	0.04	0.003181	0.015276
2-hydroxybutyrate/2-hydroxyisobutyrate	0.17	0.06	0.003182	0.015182
guanidinoacetate	-0.12	0.04	0.003207	0.015201
oleoyl-linoleoyl-glycerol (18:1/18:2) [2]	-0.14	0.05	0.003285	0.015472
glycochenodeoxycholate sulfate	-0.27	0.09	0.003382	0.015824
3b-hydroxy-5-cholenoic acid	-0.24	0.08	0.00357	0.0166
hyocholate	0.23	0.08	0.003648	0.016855
1-(1-enyl-stearoyl)-2-oleoyl-GPE (P-18:0/18:1)	0.13	0.05	0.003675	0.016871
N-stearoyl-sphingosine (d18:1/18:0)*	0.10	0.04	0.003805	0.017361
2-palmitoyl-GPC (16:0)*	0.09	0.03	0.00383	0.017364
pregnen-diol disulfate*	-0.18	0.06	0.003882	0.017494
4-guanidinobutanoate	0.19	0.07	0.004118	0.018444
N-methylproline	-0.41	0.15	0.004502	0.020041
caproate (6:0)	0.17	0.06	0.004673	0.020673
1-stearoyl-2-oleoyl-GPE (18:0/18:1)	0.17	0.06	0.004823	0.021208
N-palmitoyl-sphingadienine (d18:2/16:0)*	-0.10	0.03	0.004883	0.021345
stearoyl-arachidonoyl-glycerol (18:0/20:4) [1]*	0.12	0.04	0.004899	0.021288
N-acetylmethionine	-0.20	0.07	0.005282	0.022816
vanillic alcohol sulfate	0.56	0.20	0.005595	0.024024
dodecanedioate	0.25	0.09	0.005898	0.025177
1-palmitoyl-GPC (16:0)	0.05	0.02	0.00611	0.025934
5-HETE	0.20	0.07	0.006155	0.02597
N-palmitoyl-sphinganine (d18:0/16:0)	0.13	0.05	0.006258	0.026254
1-stearoyl-GPG (18:0)	0.17	0.06	0.006328	0.026395
choline phosphate	0.13	0.05	0.006473	0.02685
1-stearoyl-GPI (18:0)	0.15	0.06	0.00658	0.027136
9,10-DiHOME	0.21	0.08	0.006724	0.027576
palmitoleoylcarnitine (C16:1)*	-0.17	0.06	0.006744	0.027502
5alpha-pregnan-3(alpha or beta),20beta-diol disulfate	-0.29	0.11	0.006906	0.028008
gamma-glutamylglutamate	-0.15	0.06	0.006996	0.028216
chiro-inositol	-0.52	0.19	0.007196	0.028862
umbelliferone sulfate	-0.40	0.15	0.007977	0.031821
2-ethylphenylsulfate	-0.65	0.24	0.008004	0.031755
azelate (nonanedioate)	0.16	0.06	0.008871	0.035004
N6-carbamoylthreonyladenosine	-0.07	0.03	0.009037	0.035468
glycocholenate sulfate*	-0.16	0.06	0.009172	0.035806
beta-citrylglutamate	-0.14	0.05	0.009266	0.035981
1-arachidonoyl-GPE (20:4n6)*	0.08	0.03	0.009418	0.036375
urea	0.08	0.03	0.009454	0.036323
orotate	-0.12	0.05	0.009555	0.036519
gamma-glutamylhistidine	-0.12	0.05	0.010189	0.03874
1,2,3-benzenetriol sulfate (2)	0.41	0.16	0.010359	0.039182
6-oxopiperidine-2-carboxylate	-0.16	0.06	0.011127	0.041871
beta-hydroxyisovalerate	-0.16	0.06	0.01129	0.042266
asparagine	-0.07	0.03	0.011303	0.042099
3-hydroxyisobutyrate	-0.15	0.06	0.011539	0.04276
1-methylhistidine	-0.11	0.04	0.01156	0.042621
1-oleoyl-GPE (18:1)	0.13	0.05	0.01157	0.042443
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)*	0.07	0.03	0.011777	0.042987
L-urobilin	-0.41	0.17	0.012096	0.04393
retinal	0.14	0.06	0.012339	0.044591
histidylalanine	-0.18	0.07	0.012734	0.045791
cholesterol	0.07	0.03	0.012872	0.046061
stachydrine	-0.34	0.14	0.012937	0.046067
4-acetylphenol sulfate	0.25	0.10	0.013149	0.046596
16-hydroxypalmitate	0.14	0.05	0.0134	0.047257

4.3.4.5 CVD-associated metabolites in elite athletes in a 2nd cohort of 495 elite athletes

4.3.4.6 OPLS-DA in the 2nd cohort for CVD Group

An OPLS-DA was performed by comparing high vs. moderate CVD classes. The one class-discriminatory component model showed 71% of the variation in the data due to increased CVD (R-squared-Y = 0.71, Q-squared = 0.52) (Figure 35A). The diagnostic performance calculated by the leave-one-out analysis indicated sensitivity 97.9% and specificity 87.8%. Figure 35B identified the corresponding loading scores representing key metabolites that are responsible for the clear separation between high vs. moderate CVD levels. These metabolites are diacylglycerols, monohydroxy fatty acids, gamma glutamyl amino acids, leucine, isoleucine and valine metabolites, PC and PE. Consequently, OPLS confirmed linear regression model results shown in Table 25.

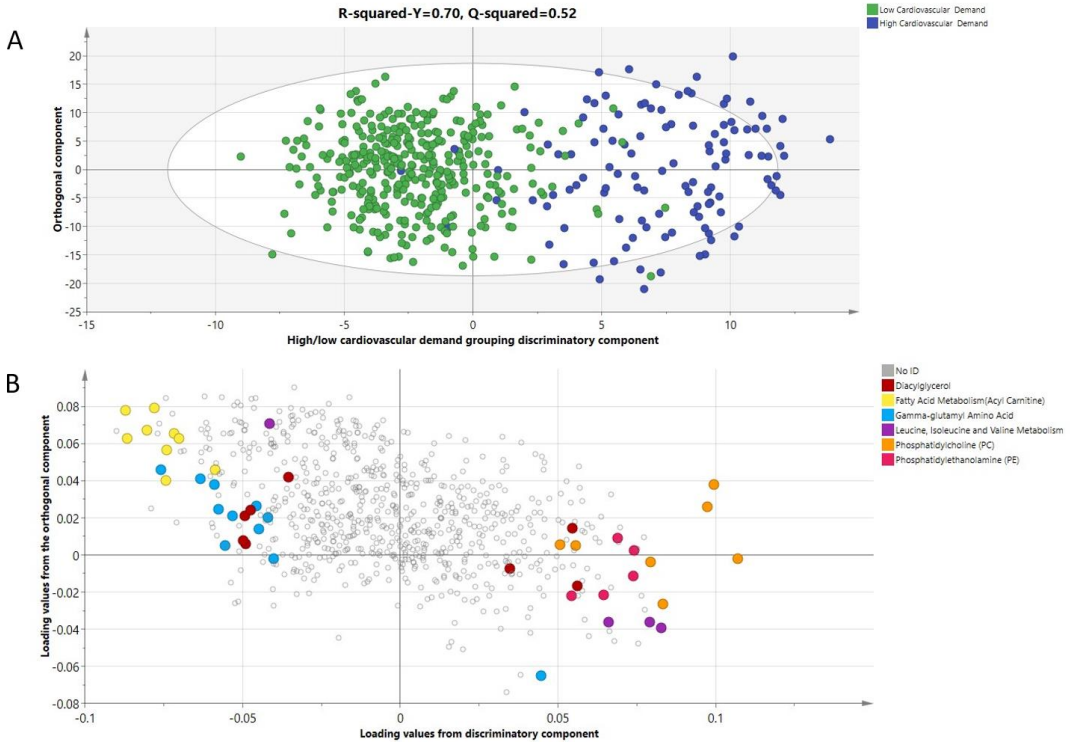


Figure 35. OPLS-DA model comparing high versus moderate CVD levels of elite athletes. **A.** A score plot indicating the orthogonal component (y-axis) and the class-discriminatory

component (x-axis). **B.** The matching loading plot indicating a cluster of gamma-glutamyl amino acids and diacylglycerols fatty acids (acyl carnitines) at the moderate end of CVD against a cluster of PC and PE at the high end.

4.3.4.7 CVD-associated metabolites in 495 elite athletes

A model incorporating CVD as a categorical variable was performed in 495 elite athletes (377 moderate versus 118 high CVD and CVD athletes). Linear regression model was run after correcting for covariates including hemolysis levels, gender, and PCA components obtained from genomic data as explained in the section 3.2.3 Statistical analysis of genomics data. Analyses included 751 known metabolites among which 112 were found significantly associated with high CVD group ($p<6.6 \times E-5$), including 2 cofactors and vitamins, 3 carbohydrates, 5 nucleotides, 8 xenobiotics, 12 peptides, 25 amino acids, and 57 lipids (Table 25). In order to validate using sports discipline as a CVD grouping criterion, a group of metabolites was utilized as surrogate markers of increased VO_2max as shown previously [425]. Figure 36 is box plot of the detected 7 metabolic markers exhibited significant differences between low/moderate and high CVD groups, suggesting that sport discipline was a reasonable criterion to dichotomize participants.

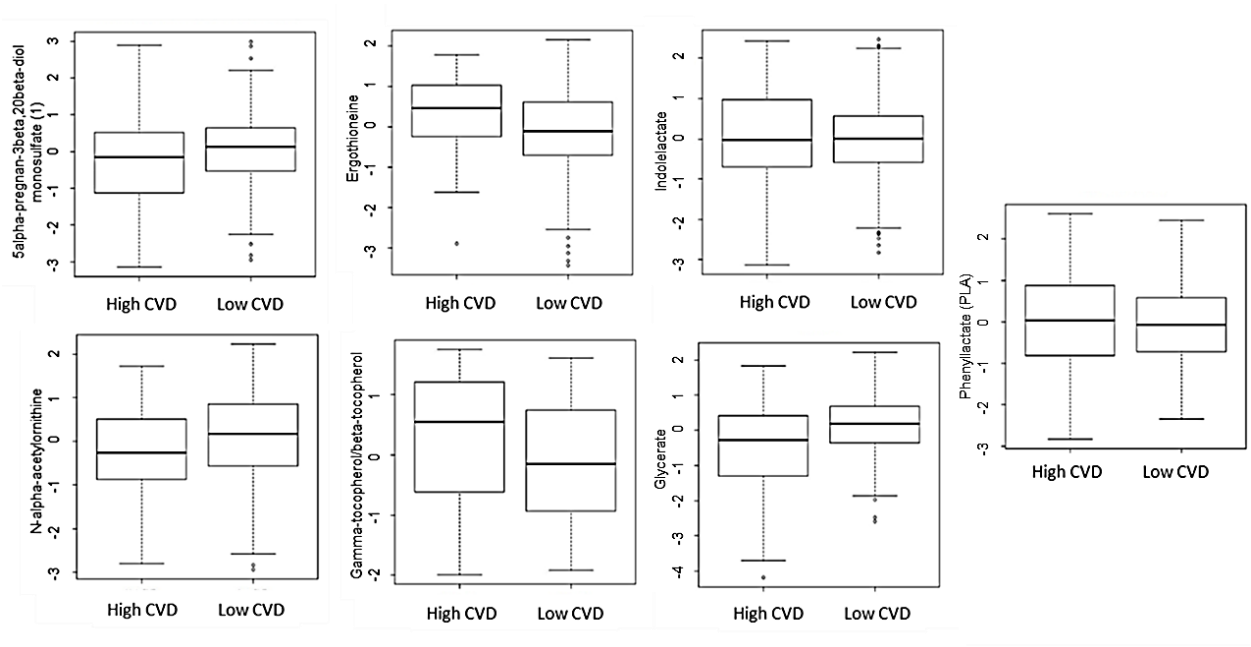


Figure 36. Boxplot describing surrogate metabolic markers of increased VO_2max that showed significant differences between low/moderate and high CVD groups.

Table 25. Metabolites that differentiate moderate vs high CVD athletes.

Metabolite	Subpathway	Beta	p value
5-oxoproline	Glutathione Metabolism	-0.90	2.19E-16
3-methyl-2-oxobutyrate	Leucine, Isoleucine and Valine Metabolism	0.91	8.42E-16
4-methyl-2-oxopentanoate	Leucine, Isoleucine and Valine Metabolism	0.91	1.27E-15
1-myristoyl-2-arachidonoyl-GPC (14:0/20:4)*	Phosphatidylcholine (PC)	0.89	2.27E-15
retinol (Vitamin A)	Vitamin A Metabolism	0.82	1.09E-13
glutamate	Glutamate Metabolism	-0.83	1.47E-13
methionine sulfoxide	Methionine, Cysteine, SAM and Taurine Metabolism	-0.83	1.60E-13
N-acetylmethionine sulfoxide	Methionine, Cysteine, SAM and Taurine Metabolism	-0.84	2.73E-13
14-HDoHE/17-HDoHE	Fatty Acid, Monohydroxy	-0.82	4.35E-13
gamma-glutamylglutamate	Gamma-glutamyl Amino Acid	-0.81	4.90E-13
imidazole lactate	Histidine Metabolism	0.83	6.22E-13
1-linoleoyl-2-arachidonoyl-GPC (18:2/20:4n6)*	Phosphatidylcholine (PC)	0.81	1.88E-12
N-acetylglucosamine/N-acetylgalactosamine	Aminosugar Metabolism	-0.79	3.40E-12
3-methyl-2-oxovalerate	Leucine, Isoleucine and Valine Metabolism	0.80	6.96E-12

aspartate	Alanine and Aspartate Metabolism	-0.77	1.14E-11
gamma-glutamylphenylalanine	Gamma-glutamyl Amino Acid	-0.75	2.70E-11
2-oxoarginine*	Urea cycle; Arginine and Proline Metabolism	0.74	7.27E-11
13-HODE + 9-HODE	Fatty Acid, Monohydroxy	-0.72	1.89E-10
carotene diol (1)	Food Component/Plant	0.70	4.51E-10
1-myristoyl-2-palmitoyl-GPC (14:0/16:0)	Phosphatidylcholine (PC)	0.70	4.62E-10
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	Phosphatidylcholine (PC)	0.67	1.11E-09
carotene diol (2)	Food Component/Plant	0.71	1.34E-09
choline	Phospholipid Metabolism	-0.69	1.77E-09
oleoylcarnitine (C18:1)	Fatty Acid Metabolism(Acyl Carnitine)	-0.67	1.86E-09
1-linolenoyl-GPC (18:3)*	Lysophospholipid	0.66	6.38E-09
N-acetylneuraminate	Aminosugar Metabolism	-0.66	9.82E-09
ximenoylcarnitine (C26:1)*	Fatty Acid Metabolism(Acyl Carnitine)	-0.64	1.21E-08
myristoleoylcarnitine (C14:1)*	Fatty Acid Metabolism(Acyl Carnitine)	-0.63	1.35E-08
glycerate	Glycolysis, Gluconeogenesis, and Pyruvate Metabolism	-0.64	1.78E-08
serine	Glycine, Serine and Threonine Metabolism	-0.65	1.94E-08
4-HDoHE	Fatty Acid, Monohydroxy	-0.66	2.84E-08
1-(1-enyl-palmitoyl)-GPC (P-16:0)*	Lysoplasmalogen	-0.63	3.16E-08
gamma-glutamylvaline	Gamma-glutamyl Amino Acid	-0.62	4.41E-08
gamma-glutamyltyrosine	Gamma-glutamyl Amino Acid	-0.62	4.75E-08
serotonin	Tryptophan Metabolism	0.61	5.29E-08
cortisol	Corticosteroids	0.63	5.99E-08
ornithine	Urea cycle; Arginine and Proline Metabolism	-0.61	6.34E-08
palmitoleoylcarnitine (C16:1)*	Fatty Acid Metabolism(Acyl Carnitine)	-0.61	6.50E-08
1-(1-enyl-palmitoyl)-2-oleoyl-GPE (P-16:0/18:1)*	Plasmalogen	0.62	8.17E-08
1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4)*	Phosphatidylethanolamine (PE)	0.60	9.38E-08
1-palmitoyl-GPA (16:0)	Lysophospholipid	-0.62	1.05E-07
4-cholesten-3-one	Sterol	-0.58	1.73E-07
1-arachidonoyl-GPA (20:4)	Lysophospholipid	-0.61	1.98E-07
1-oleoyl-GPA (18:1)	Lysophospholipid	-0.60	2.37E-07
sphingosine 1-phosphate	Sphingolipid Metabolism	-0.60	2.50E-07
4-hydroxychlorothalonil	Chemical	0.60	2.51E-07

isovalerate	Leucine, Isoleucine and Valine Metabolism	-0.59	2.63E-07
eicosenoylcarnitine (C20:1)*	Fatty Acid Metabolism(Acyl Carnitine)	-0.59	3.23E-07
1-palmitoleoyl-2-linolenoyl-GPC (16:1/18:3)*	Phosphatidylcholine (PC)	0.77	3.27E-07
gamma-glutamyltryptophan	Gamma-glutamyl Amino Acid	-0.60	3.95E-07
gamma-glutamylisoleucine*	Gamma-glutamyl Amino Acid	-0.58	4.47E-07
N-formylmethionine	Methionine, Cysteine, SAM and Taurine Metabolism	0.59	5.90E-07
lactosyl-N-palmitoyl-sphingosine (d18:1/16:0)	Ceramides	-0.56	8.50E-07
leukotriene B4	Eicosanoid	-0.58	9.75E-07
N1-methyladenosine	Purine Metabolism, Adenine containing	0.56	9.81E-07
1-palmitoyl-2-arachidonoyl-GPC (16:0/20:4n6)	Phosphatidylcholine (PC)	0.56	9.99E-07
N-acetylthreonine	Glycine, Serine and Threonine Metabolism	-0.55	1.24E-06
N6,N6,N6-trimethyllysine	Lysine Metabolism	-0.55	1.25E-06
glutamine	Glutamate Metabolism	0.56	1.80E-06
ergothioneine	Food Component/Plant	0.55	1.96E-06
gamma-glutamylleucine	Gamma-glutamyl Amino Acid	-0.54	2.50E-06
palmitoyl-oleoyl-glycerol (16:0/18:1) [2]*	Diacylglycerol	-0.56	2.62E-06
carnitine	Carnitine Metabolism	0.54	2.68E-06
allantoin	Purine Metabolism, (Hypo)Xanthine/Inosine containing	-0.52	3.05E-06
creatine	Creatine Metabolism	0.54	3.07E-06
linoleoyl-arachidonoyl-glycerol (18:2/20:4) [1]*	Diacylglycerol	0.55	3.73E-06
gamma-glutamylthreonine	Gamma-glutamyl Amino Acid	-0.53	4.05E-06
5-HETE	Eicosanoid	-0.55	4.21E-06
adenine	Purine Metabolism, Adenine containing	0.52	4.49E-06
orotate	Pyrimidine Metabolism, Orotate containing	-0.53	4.99E-06
1-(1-enyl-stearoyl)-2-linoleoyl-GPE (P-18:0/18:2)*	Plasmalogen	0.53	5.27E-06
N-acetylserine	Glycine, Serine and Threonine Metabolism	-0.52	5.70E-06
leukotriene B5	Eicosanoid	-0.55	5.71E-06
laurylcarnitine (C12)	Fatty Acid Metabolism(Acyl Carnitine)	-0.51	7.16E-06
proline	Urea cycle; Arginine and Proline Metabolism	-0.50	8.23E-06
1-stearoyl-2-arachidonoyl-GPE (18:0/20:4)	Phosphatidylethanolamine (PE)	0.51	9.06E-06
gamma-glutamyl-epsilon-lysine	Gamma-glutamyl Amino Acid	-0.50	9.10E-06
glycosyl-N-(2-hydroxynervonoyl)-sphingosine (d18:1/24:1(2OH))*	Ceramides	-0.52	9.16E-06
oleoyl-oleoyl-glycerol (18:1/18:1) [1]*	Diacylglycerol	-0.52	9.38E-06

1-methylimidazoleacetate	Histidine Metabolism	-0.52	9.57E-06
linoleoylcarnitine (C18:2)*	Fatty Acid Metabolism(Acyl Carnitine)	-0.51	9.83E-06
cysteine sulfinic acid	Methionine, Cysteine, SAM and Taurine Metabolism	-0.50	9.83E-06
N-oleoyltaurine	Endocannabinoid	-0.53	1.03E-05
1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	Phosphatidylethanolamine (PE)	0.50	1.07E-05
palmitoyl-arachidonoyl-glycerol (16:0/20:4) [2]*	Diacylglycerol	0.57	1.17E-05
alpha-tocopherol	Tocopherol Metabolism	0.49	1.32E-05
oleoyl-linoleoyl-glycerol (18:1/18:2) [2]	Diacylglycerol	-0.50	1.38E-05
caproate (6:0)	Medium Chain Fatty Acid	-0.49	1.39E-05
1-(1-enyl-oleoyl)-GPE (P-18:1)*	Lysoplasmalogen	-0.50	1.50E-05
1-palmitoyl-2-oleoyl-GPE (16:0/18:1)	Phosphatidylethanolamine (PE)	0.49	1.60E-05
decanoylcarnitine (C10)	Fatty Acid Metabolism(Acyl Carnitine)	-0.49	1.67E-05
eugenol sulfate	Food Component/Plant	-0.53	1.70E-05
oleoyl-oleoyl-glycerol (18:1/18:1) [2]*	Diacylglycerol	-0.50	1.72E-05
1-linoleoyl-GPI (18:2)*	Lysophospholipid	0.49	1.87E-05
adipoylcarnitine (C6-DC)	Fatty Acid Metabolism(Acyl Carnitine)	-0.48	2.25E-05
beta-citrylglutamate	Glutamate Metabolism	-0.48	3.09E-05
N-formylphenylalanine	Tyrosine Metabolism	-0.50	3.14E-05
sphinganine-1-phosphate	Sphingolipid Metabolism	-0.48	3.17E-05
3,7-dimethylurate	Xanthine Metabolism	0.52	3.28E-05
1,2-dilinoleoyl-GPC (18:2/18:2)	Phosphatidylcholine (PC)	0.48	3.40E-05
N-acetylcarnosine	Dipeptide Derivative	-0.42	3.52E-05
gamma-glutamylhistidine	Gamma-glutamyl Amino Acid	-0.47	3.55E-05
gamma-glutamylglutamine	Gamma-glutamyl Amino Acid	0.46	3.65E-05
1-palmitoyl-2-docosahexaenoyl-GPE (16:0/22:6)*	Phosphatidylethanolamine (PE)	0.47	3.77E-05
palmitoleoyl-arachidonoyl-glycerol (16:1/20:4) [2]*	Diacylglycerol	0.57	3.82E-05
orotidine	Pyrimidine Metabolism, Orotate containing	-0.43	4.18E-05
O-methylcatechol sulfate	Benzoate Metabolism	-0.46	5.09E-05
5alpha-pregnan-3beta,20beta-diol monosulfate (1)	Progestin Steroids	-0.48	5.60E-05
oleoyl-linoleoyl-glycerol (18:1/18:2) [1]	Diacylglycerol	-0.47	5.66E-05
3-methylxanthine	Xanthine Metabolism	0.49	5.76E-05
12-HETE	Eicosanoid	-0.47	6.10E-05
1-(1-enyl-palmitoyl)-2-arachidonoyl-GPE (P-16:0/20:4)*	Plasmalogen	0.46	6.36E-05

The thirty-two CVD associated metabolites corresponding to different sub-pathways that significantly differentiating moderate versus high CVD groups were projected on the

heatmap in Figure 37. The heatmap summarizes a snapshot of the actual intensities of significant metabolites associated with high CVD group of athletes after correcting for covariates in the linear regression model described above. Samples on *y*-axis were ordered by sports group (high CVD, low/moderate CVD). The color code denotes *z*-scaled values of metabolites after correction of confounders (red represents an increase in high CVD, green represents a decrease in high CVD). Whereas the moderate groups of CVD showed similar intensities of metabolites, the latter high CVD group was visibly different intensities of metabolites, signifying presence of athletes with the moderate CVD group exhibiting a similar metabolic profile to the high CVD group than their own low/moderate CVD group. Athletes who belong to moderate CVD group are mostly athletics who share some of the features of the high CVD group including high VO_2max. Since we do not have the detailed description of these athletes, we predict that the moderate CVD athletics belong to high endurance sports.

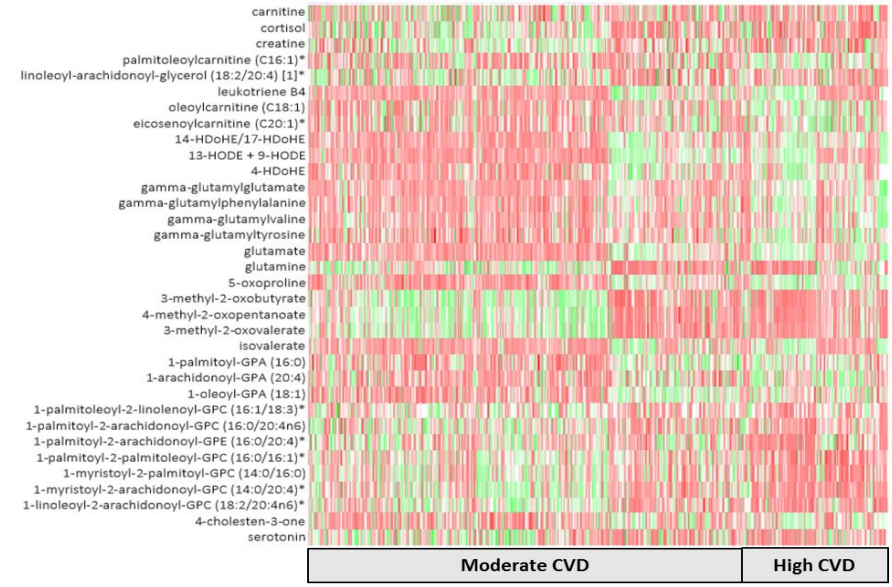
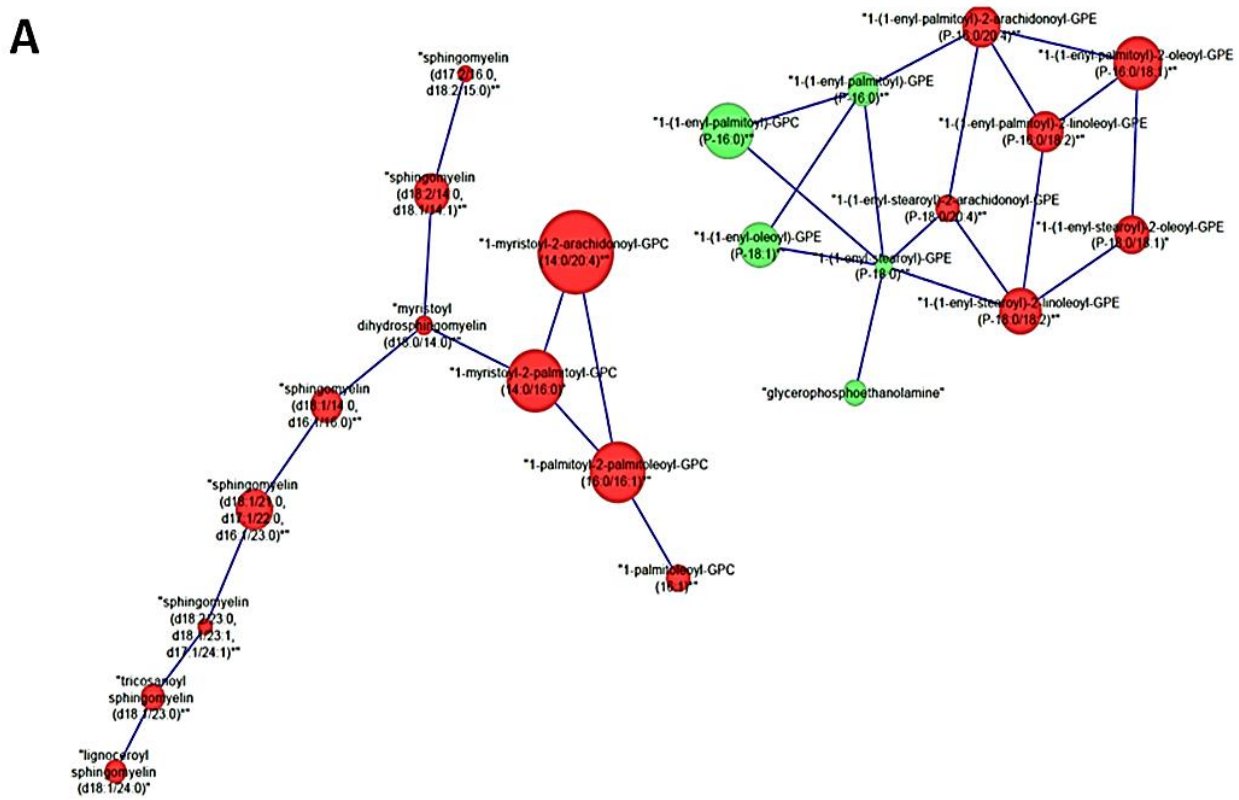
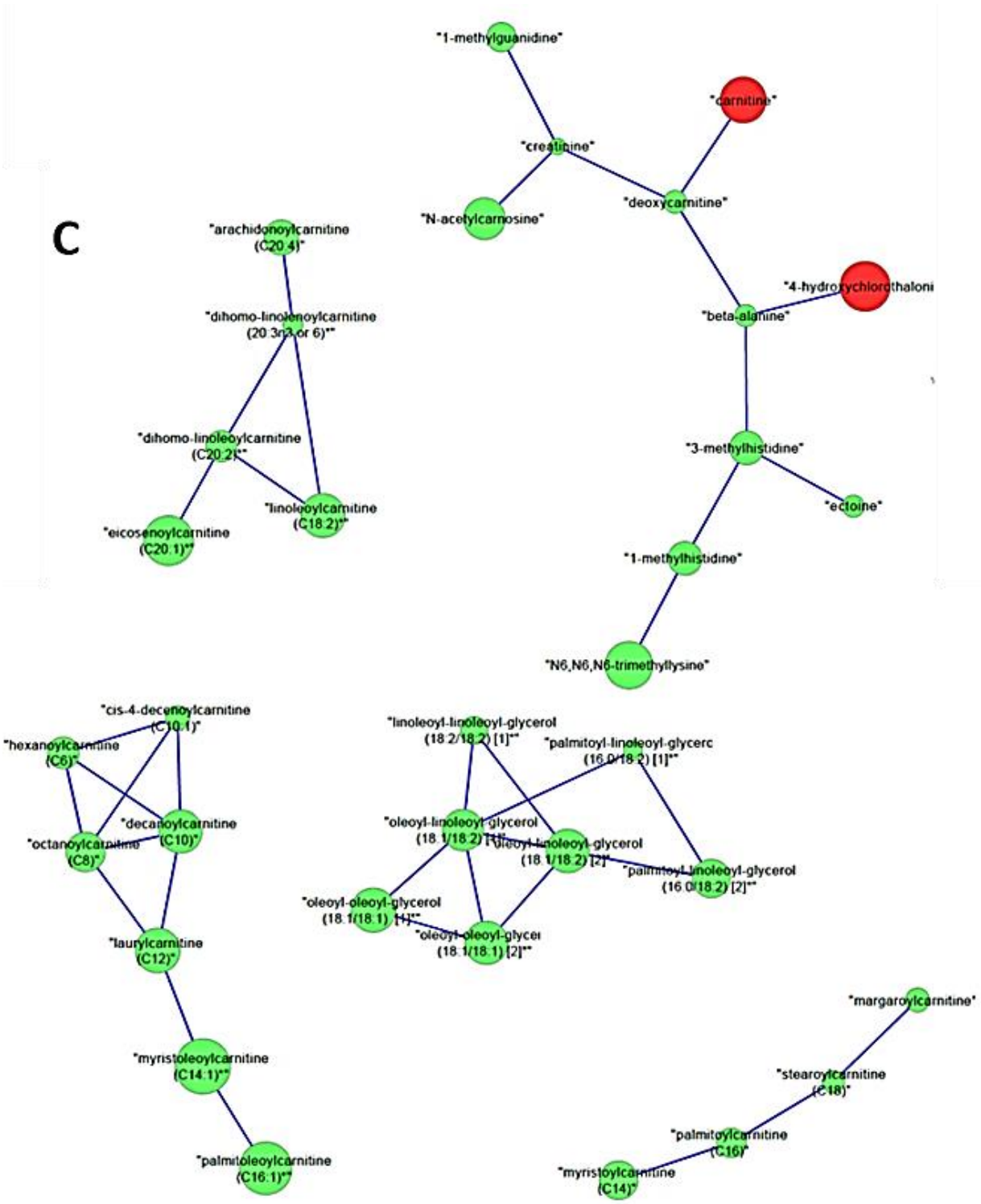


Figure 37. Heatmap of high CVD associated with 32 significant metabolites (y axis). Samples were grouped into moderate vs. high CVD groups type (x axis). The color code represents scaled values of metabolites after correction of covariates.

4.3.4.8 Six GGM networks capture pathways that change in relation to CVD

All 751 metabolites were projected into GGM pairwise correlation calculation using R (version 3.3.1). GGM networks were constructed using Cytoscape software, resulting in 60 subnetworks containing more than two metabolites with an overall 604 metabolites (nodes) connected by 600 edges. The identified subnetworks were filtered for metabolites associated with CVD at (p less than 0.05), resulting in 11 subnetworks containing more than two Bonferroni significant metabolites. The major metabolic pathways were captured by six subnetworks that were selected for discussion in relation to CVD (Figure 38), including phosphatidyls (cholines and ethanol amines) (Figure 38A), eicosanoids (Figure 38B), carnitine metabolism (Figure 38C), gamma glutamyl amino acids and their link to glutamate (Figure 38D), cortisol metabolism (Figure 38E), energy metabolites including creatine and TCA (Figure 38F).


A
"sphingomyelin (d17:2/16:0, d18:2/15:0)**"
"sphingomyelin (d18:2/14:0, d18:1/14:1)**"
"1-myristoyl-2-arachidonoyl-GPC (14:0/20:4)**"
"myristoyl dihydrosphingomyelin (d18:0/14:0)**"
"1-myristoyl-2-palmitoyl-GPC (14:0/16:0)*"
"sphingomyelin (d18:1/14:0, d16:1/16:0)**"
"1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)**"
"sphingomyelin (d18:1/23:0, d17:1/24:0, d16:1/25:0)**"
"1-palmitoleoyl-GPC (16:1)**"
"sphingomyelin (d18:2/23:0, d18:1/23:1, d17:1/24:1)**"
"tricosanoyl sphingomyelin (d18:1/23:0)**"
"lignoceroyl sphingomyelin (d18:1/24:0)*"
"1-(1-enyl-palmitoyl)-2-arachidonoyl-GPE (P-16:0/20:4)**"
"1-(1-enyl-palmitoyl)-2-oleoyl-GPE (P-16:0/18:1)**"
"1-(1-enyl-palmitoyl)-GPE (P-16:0)**"
"1-(1-enyl-palmitoyl)-GPC (P-16:0)**"
"1-(1-enyl-palmitoyl)-2-linoleoyl-GPE (P-16:0/18:2)**"
"1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4)**"
"1-(1-enyl-stearoyl)-2-oleoyl-GPE (P-18:0/18:1)*"
"1-(1-enyl-oleoyl)-GPE (P-18:1)**"
"1-(1-enyl-stearoyl)-GPE (P-18:0)**"
"1-(1-enyl-stearoyl)-2-linoleoyl-GPE (P-18:0/18:2)**"
"glycerophosphoethanolamine"
B
"14-HDoHE/17-HDoHE"
"leukotriene B4"
"5-HETE"
"4-HDoHE"
"5-HEPE"
"leukotriene B5"


C
"1-methylguanidine"
"carnitine"
"creatinine"
"N-acetylcarnosine"
"deoxycarnitine"
"arachidonoylcarnitine (C20:4)"
"4-hydroxychlorothaloni
"dihomo-linolenoylcarnitine (20:3n3 or 6)"
"beta-alanine"
"dihomo-linoleoylcarnitine (C20:2)"
"3-methylhistidine"
"linoleoylcarnitine (C18:2)"
"ectoine"
"eicosenoylcarnitine (C20:1)"
"1-methylhistidine"
"N6,N6,N6-trimethyllysine"
"cis-4-decenoylcarnitine (C10:1)"
"hexanoylcarnitine (C6)"
"linoleoyl-linoleoyl-glycerol (18:2/18:2) [1]"
"palmitoyl-linoleoyl-glycerc (16:0/18:2) [1]"
"decanoylcarnitine (C10)"
"oleoyl-linoleoyl-glycerol (18:1/18:2) [1]"
"octanoylcarnitine (C8)"
"oleoyl-linoleoyl-glycerol (18:1/18:2) [2]"
"palmitoyl-linoleoyl-glycerol (16:0/18:2) [2]"
"oleoyl-oleoyl-glycerol (18:1/18:1) [1]"
"laurylcarnitine (C12)"
"oleoyl-oleoyl-glycer (18:1/18:1) [2]"
"margaroylcarnitine"
"myristoleoylcarnitine (C14:1)"
"stearoylcarnitine (C18)"
"palmitoylcarnitine (C16)"
"palmitoleoylcarnitine (C16:1)"
"myristoylcarnitine (C14)"

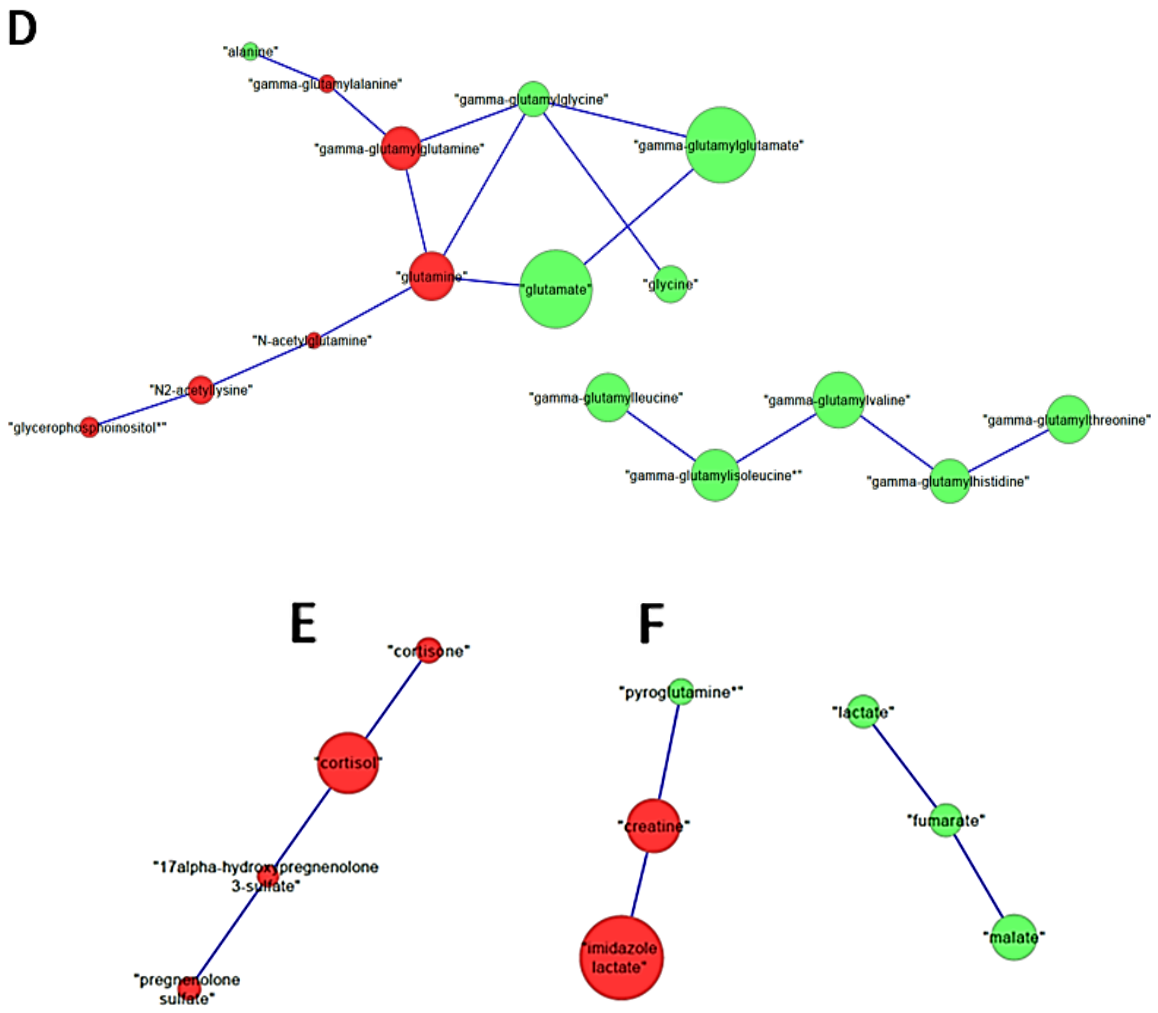


Figure 38. Six GGM subnetworks reveal metabolic networks that significantly differentiated between high versus moderate CVD groups of athletes. Alterations are signified by nodes with sizes proportional to – (log p value) (more significant metabolite association with CVD group represented by larger nodes). **A.** Phosphatidyls including cholines and ethanol amines, **B.** Eicosanoids, **C.** Carnitine metabolism, **D.** Gamma glutamyl amino acids and their link to glutamate, **E.** Cortisol metabolism, **F.** Energy metabolites including tricarboxilic acid cycle and creatine. Green shows that low/moderate CVD has higher metabolite levels and red color shows that high CVD has higher metabolite levels.

4.3.5 Metabolomics profiling of xenobiotics in elite athletes

Despite multiple studies investigating the impact of specific dietary supplements on health and performance of elite athletes, no study has focused on products of dietary consumption and environmental exposure in elite athletes by adopting a comprehensive serum xenobiotic profiling. This study aims to identify differences in various xenobiotics in serum samples from elite athletes of different sport disciplines, focusing on metabolites that certainly originate from supplements/drugs.

4.3.5.1 Multivariate analysis of athlete metabolomics data

Non-targeted metabolomics data determined 102 studied xenobiotic signatures in 478 elite athletes belonged to six classes (bacterial and fungal metabolites, tobacco metabolites, xanthine metabolites, benzoate metabolites, food components, and chemicals and drugs). Figure 39 summarizes classes and prevalence of xenobiotics found in elite athletes.

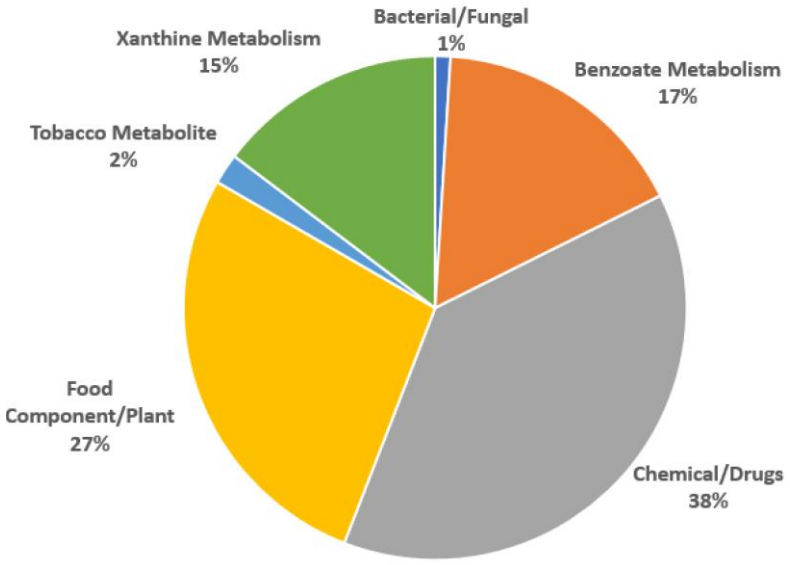


Figure 39. A pie chart summarizing classes and prevalence of xenobiotics identified in elite athletes included in this study.

An OPLS-DA analysis was performed using seven studied sport disciplines for comparing xenobiotics. OPLS results revealed one class-discriminatory component separating boxers and football players from the rest of the sport groups (Figure 40A). For easier visualization, OPLS-DA was repeated by combining football players and boxers in one group and the rest in group two (Figure 40B). With the second unique discriminatory component (x-axis, Figure 40B) model, the differences between sport groups accounted 40% and of the variation among xenobiotics was 30%. The corresponding loading score, shown in Figure 40C, suggested an increase in 2,3-dihydroxyisovalerate, 4-allylphenol.sulfate, 3.methylcatechol sulfate, 3-hydroxypyridine sulfate, 4-vinylguaiacol.sulfate, catechol sulfate, methyl.glucopyranoside, caffeic acid sulfate, ectoine, tartronate .hydroxymalonate, ferulic acid 4 sulfate, N-2.furoyl glycine, 3hippurate, O-methylcatechol sulfate, quinate, in football players and boxers, while showing an increase in 1,3.7-trimethylurate, tartronate (hydroxymalonate), thioprolinein the other groups (athletics, cyclist, rowers, rugby players and swimmers).

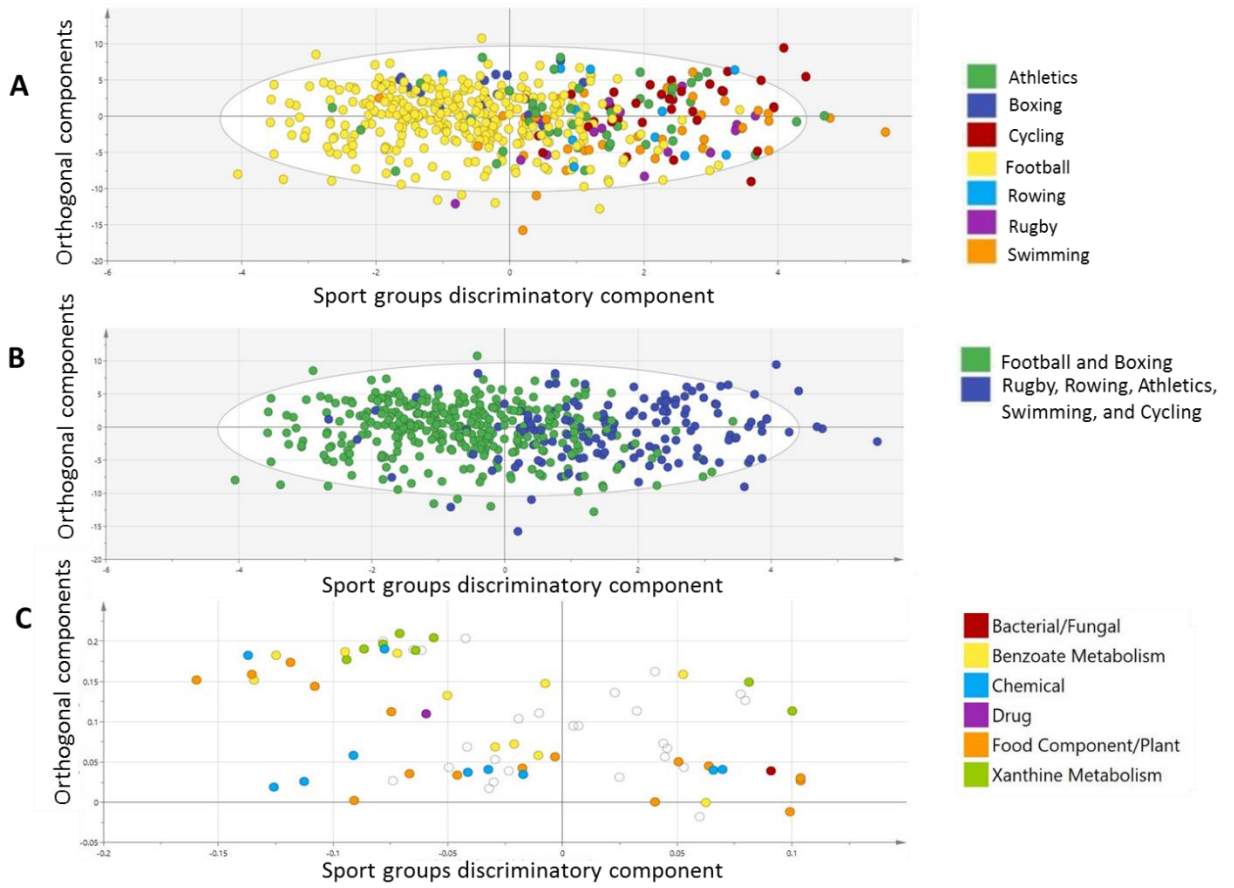


Figure 40. OPLS-DA model comparing elite athletes of different sport disciplines (cycling, athletics, rugby, boxing, rowing, swimming and football) (A) A score plot showing the orthogonal component (y-axis) vs. class-discriminatory component (x-axis) between all sport groups. (B) An updated score plot from 2nd OPLS-DA model featuring group one: football and boxing, group two: cycling, rugby, athletics, rowing, and swimming). (C) The corresponding loading plot revealing clusters of xenobiotics at opposite sides of group one or group either ends of the discriminatory component along the x-axis.

4.3.5.2 Univariate association tests

A linear regression model was used to identify the significant metabolite associated with studied sport groups: athletics, swimming, cycling, rugby, rowing, boxing and football after correcting for covariates including gender, and PCA components obtained from genomic

data as explained in the section 3.2.3 Statistical analysis of genomics data predicted ethnicities. Twenty one metabolites were significantly differentiating between studied sport groups including xanthine metabolites (catechol sulfate, 1,3,7-trimethylurate caffeic acid sulfate, and O-methylcatechol.sulfate), chemicals (2-pyrrolidinone, thioproline, 3-hydroxypyridine.sulfate, ectoine), food and plant products (ferulic acid 4-sulfate, eugenol sulfate, methyl glucopyranoside (alpha/beta), quinate, 2-furoylglycine, retinol, 4-vinylguaiacol.sulfate, stachydrine, 2,3-dihydroxyisovalerate, tartronate (hydroxymalonate), 4-allylphenol.sulfate, and benzoate metabolites (3-methylcatechol sulfate, 4-hydroxyhippurate, hippurate). Table 26 lists top significantly different metabolites (FDR p value less than 0.05) between different studied sport groups, their fold change and level of significance. After correcting for covariates, levels of significantly different metabolites are visually plotted in Figure 41.

Table 26. Metabolites differentiating between studied sport groups (FDR significance, p ≤0.05).

Metabolite	Contrast	Fold change	Nominal p-value	FDR p-value
Catechol sulfate	Football_Swimming	0.9	3.04E-09	4.86E-06
O-methylcatechol.sulfate	Football_Swimming	1.1	3.35E-08	1.78E-05
Quinate	Football_Swimming	1.8	2.89E-08	1.78E-05
2-pyrrolidinone	Boxing_Rugby	0.5	8.13E-08	3.24E-05
2-furoyl.glycine	Cycling_Football	-1.2	6.26E-07	0.0002
2-pyrrolidinone	Boxing_Cycling	0.4	8.35E-07	0.0002
2-pyrrolidinone	Athletics_Rugby	0.3	3.99E-06	0.0009
Thioproline	Boxing_Rowing	1.2	7.42E-06	0.002
1,3,7-trimethylurate	Athletics_Football	-1.0	1.79E-05	0.003
Tartronate (hydroxymalonate)	Boxing_Rowing	-0.7	1.80E-05	0.003
2-pyrrolidinone	Football_Rugby	0.3	1.83E-05	0.003

Thioproline	Boxing_Cycling	1.0	2.96E-05	0.004
Thioproline	Athletics_Boxing	-0.8	3.50E-05	0.004
Ferulic acid 4-sulfate	Cycling_Football	-0.9	3.87E-05	0.004
Tartronate (hydroxymalonate)	Boxing_Rugby	-0.7	7.85E-05	0.007
Tartronate (hydroxymalonate)	Boxing_Swimming	-0.5	6.76E-05	0.007
3-hydroxypyridine sulfate	Cycling_Football	-1.3	7.90E-05	0.007
4-vinylguaiacol sulfate	Cycling_Football	-1.2	8.00E-05	0.007
Ectoine	Football_Swimming	0.9	8.32E-05	0.007
Ferulic acid 4-sulfate	Football_Swimming	0.9	7.10E-05	0.007
2-pyrrolidinone	Athletics_Cycling	0.2	9.09E-05	0.007
2-pyrrolidinone	Boxing_Rowing	0.4	9.30E-05	0.007
3-methyl catechol sulfate	Football_Swimming	1.1	0.0001	0.007
2,3-dihydroxyisovalerate	Boxing_Football	-1.6	0.0001	0.008
Hippurate	Football_Swimming	0.8	0.0001	0.009
Eugenol.sulfate	Athletics_Swimming	1.3	0.0002	0.009
Retinol	Athletics_Boxing	-1.50	0.0002	0.010
2,3-dihydroxyisovalerate	Athletics_Boxing	1.5	0.0002	0.010
Caffeic acid sulfate	Cycling_Football	-1.1	0.0002	0.010
Tartronate (hydroxymalonate)	Boxing_Cycling	-0.5	0.0002	0.010
3-hydroxypyridine sulfate	Football_Swimming	1.3	0.0002	0.011
Tartronate (hydroxymalonate)	Boxing_Football	-0.5	0.0002	0.013
Eugenol sulfate	Football_Swimming	1.1	0.0003	0.017
Stachydrine	Athletics_Swimming	1.1	0.0004	0.018
Ectoine	Cycling_Football	-0.8	0.0004	0.020
Stachydrine	Athletics_Cycling	1.1	0.0005	0.023
4-hydroxyhippurate	Athletics_Swimming	0.7	0.0005	0.023
Thioproline	Boxing_Football	0.8	0.0007	0.028
Methyl glucopyranoside	Athletics_Boxing	1.0	0.0008	0.031
4-allylphenol sulfate	Athletics_Boxing	1.2	0.0012	0.048

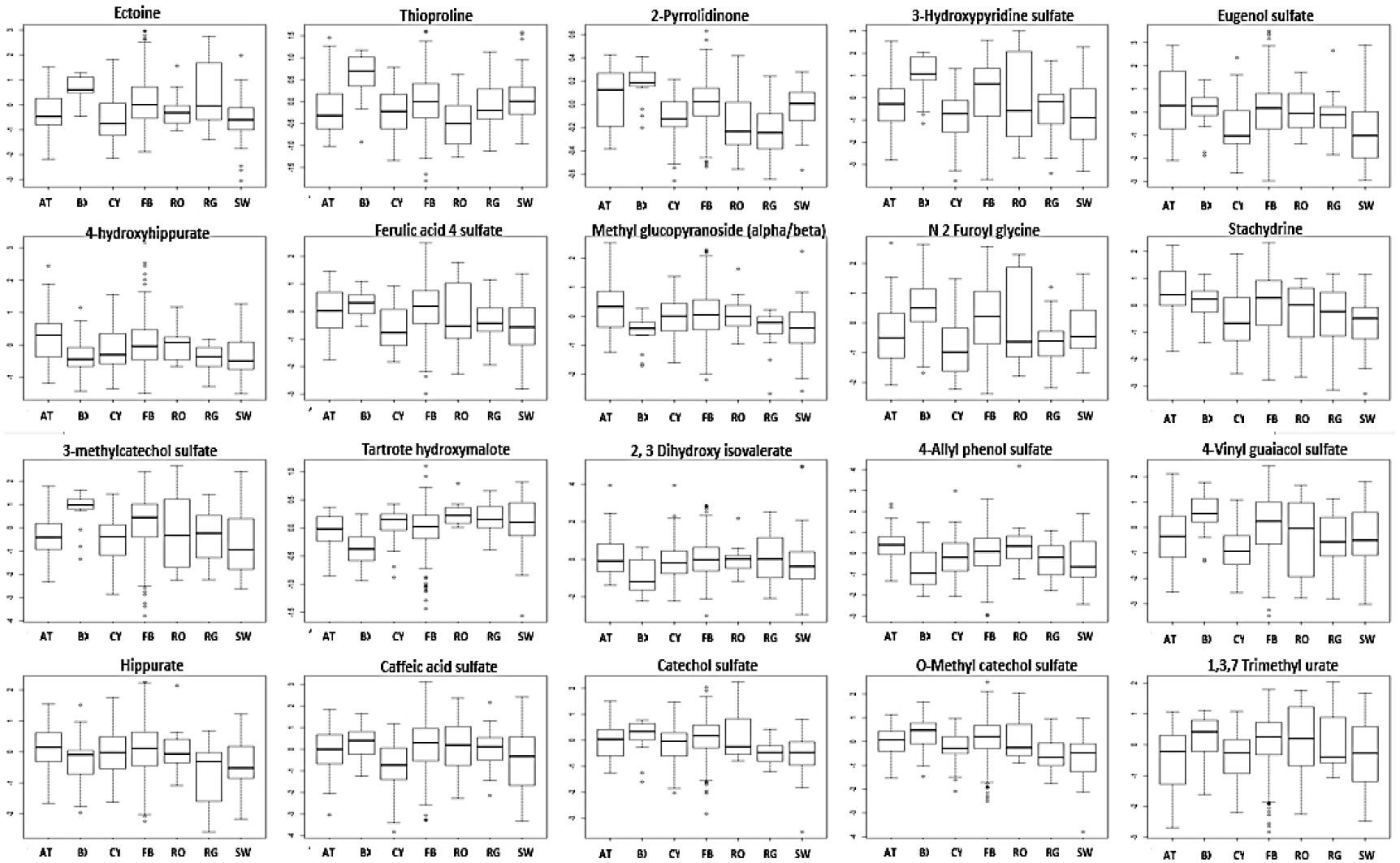

Ectoine
Thioproline
2-Pyrrolidinone
3-Hydroxypyridine sulfate
Eugenol sulfate
4-hydroxyhippurate
Ferulic acid 4 sulfate
Methyl glucopyranoside (alpha/beta)
N 2 Furoyl glycine
Stachydrine
3-methylcatechol sulfate
Tartrote hydroxymalote
2, 3 Dihydroxy isovalerate
4-Allyl phenol sulfate
4-Vinyl guaiacol sulfate
Hippurate
Caffeic acid sulfate
Catechol sulfate
O-Methyl catechol sulfate
1,3,7 Trimethyl urate
AT
B)
CY
FB
RO
RG
SW

Figure 41. Box plots visualizing levels of significantly different metabolites between seven studied groups (SW: Swimming, RG: Rugby, RO: Rowing, CY: Cycling, AT: Athletics, BX: Boxing, FB: Football). These levels are from a repeated linear regression model corrected for covariates, mean-shifted, and scaled since they show the residuals that omits the sport group while featuring only covariates.

All seventy-two xenobiotics with no missing values were used for GGM sub-network construction with size of nodes indicating the level of significance, therefore creating a map that potentially reveals biochemical connections between different sport groups-associated networks containing more than or equal to two metabolites (Figure 42). These included two xanthine sub-networks showing partial correlations between one benzoate sub-network including 5 benzoate metabolites (shown in yellow in Figure 42), 11 xanthine metabolites (shown in blue in Figure 42), and various smaller food components sub-networks that discovered direct partial correlations between ferulic acid 4-sulfate and 4-vinylguaiacol sulfate and between stachydrine and methyl glucopyranoside (alpha/beta) (shown in red in Figure 42).

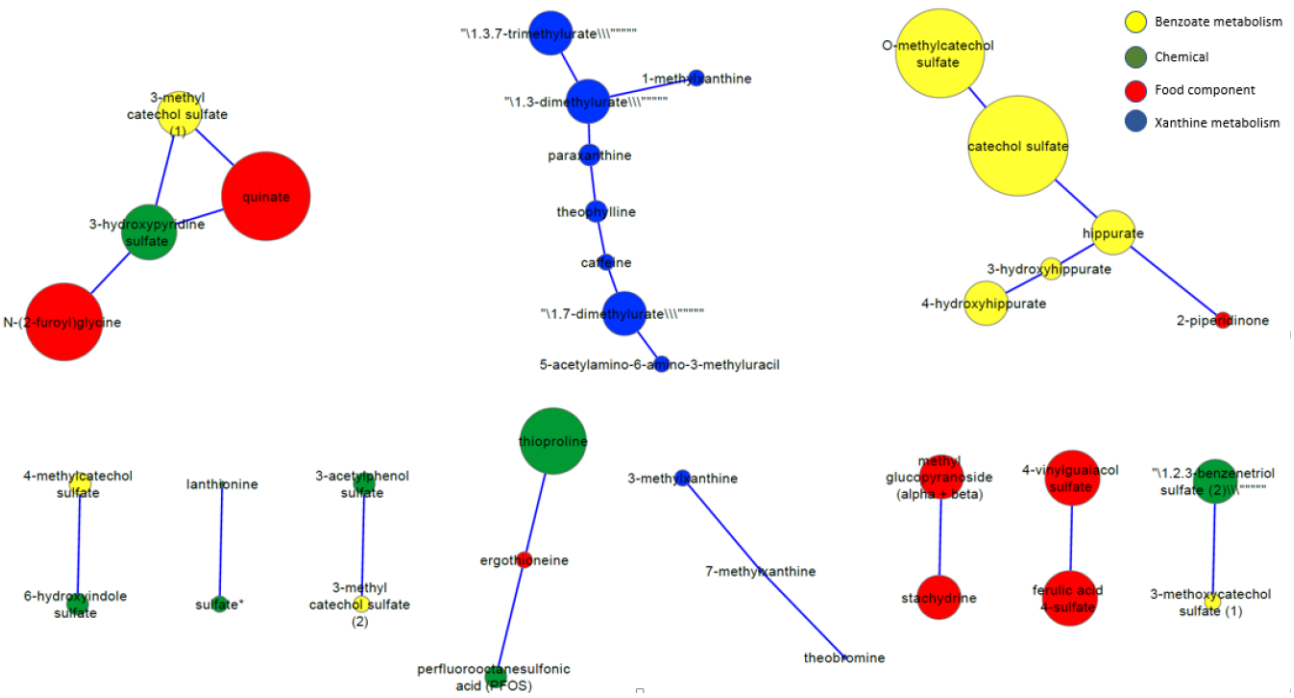


Figure 42. Xenobiotics GGM sub-networks that varied significantly between studied sport groups and between Football and Swimming. Significance are represented by nodes with sizes proportional to – (log p value) (bigger nodes reveal more significant association with particular sport group). Colors correspond to classes of metabolites (Xanthine metabolites in blue, benzoate metabolites in yellow, food components in red, and chemicals in green).

4.4 Discussion

Metabolic profiling of athletes' serum samples in response to exercise has recently discovered unique metabolic biomarkers associated with different intensities and durations of exercise [128, 426]. However, metabolomics of elite athletes from different sport classes remains to be investigated. Particularly, the metabolic pathways of power and endurance elite athletes should shed light on the molecular mechanisms underlying variations with functional significance or those that can be used as potential signatures for their respective sport class. In this study, metabolomics association analysis was utilized to identify the unique serum metabolic signature of elite athletes who competed in national or international sports events and successfully passed the anti-doping tests. Following rules and regulations of WADA (The World Anti-Doping Code), athletes' samples were provided with very limited information that included sport type and gender only. Even with the restricted information about possible confounding factors influencing their metabolic profiling, the emerging data discovered significant alterations in metabolite levels between moderate versus high endurance, power and CVD levels of sport classes. Incorporating PCA components for discovery cohort in the linear regression model has conceivably corrected for expected covariates including pre-post exercise and hemolysis effects to identify common as well as differentiating/distinct metabolic mechanisms underlying endurance and power. Bonferroni level of significance was calculated and checked. Fold change reflects beta values. They are similar to correlation coefficient and range from -1 to +1 depending on the direction of the relationship. The closer the value is to 1 or -1, the stronger the relationship. Low beta therefore reflect low correlation with the

phenotype [427]. Second cohort was used as a replication cohort to confirm the list of metabolites associated with endurance and power group of athletes.

4.4.1 Metabolites associations

4.4.1.1 Metabolites associated with endurance

Exercise can cause alterations in sex steroid hormone levels in the serum of athletes as well as non-athletes [135, 428], including concentrations of cortisol and testosterone [429, 430]. Interestingly, this study's results showed elevated levels of several metabolites involved in sex steroid hormone biosynthesis associated with high endurance athletes. Some of the significant metabolites were conjugated with one or more sulfate group(s) which renders them inactive. However, these can be reactivated through the activity of enzyme steroid sulfatase [431]. The list of elevated steroids (Fig. 28) involved pregnenolone that mediates biosynthesis of progesterone and corticosteroids, 21-hydroxypregnenolone disulfate that mediates biosynthesis of corticoids (cortisol and cortisone), corticosteroids, various metabolites of progesterone (5alpha-pregnan-3beta,20alpha-diol, 5alpha-pregnan-3beta,20beta-diol, pregnanediol), testosterone precursor (androstenediol (3beta,17beta)) and testosterone metabolites (androstenediol (3alpha, 17alpha), etiocholanoloneglucuronide). The increase in cortisol-related metabolites in response to sustained aerobic exercise correlated positively with intensity of exercise measured by oxygen uptake [432]. However, exercise-induced changes in sex steroid hormone levels lasts only for 1–3 hours which are considered usually short lived [135]. The typical exercise regiments that are usually followed by elite endurance athletes may have accounted for this maintained systemic increase. Sex steroid hormones play a crucial role in regulation of redox homeostasis, as well as maintain

protein synthesis and glucose metabolism in the muscle [433-435]. Among significant results is pregnen-dioldisulfate that is known for its role as neurosteroid that alters neuronal excitability and also works as a potent negative allosteric modulator of the GABAA receptor [436] and pregnenolone sulfate that acts as a potent negative allosteric modulator of the GABAA receptor and a weak positive allosteric modulator of the N-methyl-D-aspartate (NMDA) receptor [437]. The stimulatory effects of steroids on neuronal excitability, muscle mass and energy generation may have accounted partially to the superior performance associated with high endurance elite athletes. The fact that athletes included in this study all tested negative in anti-doping tests, alterations in steroids levels may reflect either enrichment in endogenous anabolic steroids biosynthesis due to genetic predisposition, physiological adaptation to exercise and/or increased dietary intake. Therefore, further genetic association studies are required to reveal the potential genetic variants underlying elevated activity of enzymes involved in steroid biosynthesis. Remarkably, in addition to increased neurosteroids, findings of this study suggested increased elevated levels of a number of GABA byproducts including 2-pyrrolidinone, the cyclic lactate form of GABA [438], its derivatives (N-(3-acetamidopropyl)pyrrolidin-2-one), 4-guanidinobutanoate, and succinimide, possibly contributing to GABA-mediated muscle growth in response to exercise [439].

Other findings of high endurance associated metabolic alterations involved elevated acylated carnitine and decreased FA-carnitine and DAGs. Changes in these metabolites may suggest enhanced hydrolysis of DAGs, shuttling of FA intracellularly followed by fatty acid oxidation and energy generation [440]. Lipids and fatty acids are ideal substrates for exercising muscle and the results of this study suggest a bigger beta oxidation of fatty

acids in athletes in response to higher endurance exercises. Hence, lipolysis is highly active in high endurance athletes compared to moderate endurance athletes during physical activity.

Focusing on the increased levels of acylated carnitine may indicate prevention from cellular damage and recovery from exercise stress since carnitine can reduce post-exercise plasma lactate [441]. Isocitrate and citrate were also increased in high endurance elite athletes, indicating enhanced aerobic energy generation through TCA cycle.

4.4.1.2 Metabolites associated with power athletes

Alterations in guanidinoacetate, creatine, and creatinine were significantly associated with power athletes. Maintaining the previously reported balance of creatine metabolism can be seen here, as creatine increased in the high power athletes, its breakdown product (creatinine) and precursor (guanidinoacetate) were both significantly reduced [442]. Creatine phosphate and creatine play crucial roles in the transmission and storage of phosphate-bound energy. Alterations in creatine homeostasis in high power athletes may suggest more creatine phosphate storage in muscles that adapted during exercise and will constitute a primary source for high energy to reload ATP in the first few seconds of intense activity 7-methylxanthine (adenine breakdown products) and 3-methylxanthine were other examples of increased energy-related metabolites elevated in high power athletes, possibly indicating heightened utilization of fuel substrates in several metabolic pathways [443]. Xanthine supplementation allows high power athletes to exercise at a higher power output for longer durations [444].

Furthermore, N-acetylcarnosine was significantly decreased in high power athletes. This metabolite is known as oxidative stress scavenger in muscles where it scavenges the lipid peroxidation over the imidazolium groups that stabilize adducts formed at the primary amino group [445].

Various byproducts of phosphatidates were elevated with high power athletes, possibly indicating alterations in cellular membranes dynamics in response to oxidative stress [446]. Among those, 12,13-DHOME was earlier revealed to accumulate in response to stress-induced membrane dynamics. This long-chain fatty acid inhibits osteogenesis and enhances adipogenesis due to its role as a proliferator-activated receptor (*PPAR*) gamma 2 ligand [447]. Another elevated metabolite likely to be a product of muscle contraction during hypoxia is inositol phospholipids [448].

4.4.1.3 Global stress response in both high power and endurance athletes

Intensive exercise causes elevation in free radical generation in active skeletal muscle resulting in the formation of oxidized lipids [446]. Overall in both endurance and power athletes, there was a clear stress metabolic response. Alterations in gamma-glutamyl amino acids, associated with increased cysteine-glutathione disulfide (fold change 0.24, nominal p value of 0.03), between high and moderate performance athletes may reflect activation of gamma-glutamyl cycle that plays a significant role in the glutathione mediated radical detoxification during oxidative stress [449]. The cycle involves synthesis and degradation of glutathione by transferring gamma-glutamyl functional groups from glutathione to an amino acid, leaving the cysteine product to preserve intracellular homeostasis of oxidative stress [450, 451]. Decrease in serum levels of gamma-glutamyl-amino acids in high performance athletes (both high endurance and high power) may

suggest increased glutathione synthesis. The enrichment of glutathione in the blood stream marks elevated oxidative stress and reactive oxygen species scavenging activity. These metabolic alterations detected in high endurance and high-power elite athletes may indicate several cellular adaptations to prolonged exercise-induced oxidative stress. These may involve modulation of energy utilization, muscle mass and deployment of stress-scavenging mechanisms. Further research should investigate whether elite athletes may have developed more efficient mechanisms to counteract exercise-induced stress.

4.4.1.4 Metabolites associated with CVD

Cardiovascular system of elite athletes is also influenced by the intensive exercise causing certain physiological adaptations leading to increased stroke volume and blood pressure in response to enhanced performance. Metabolomics association analysis with different CVD levels of athletes could provide vital information about their systemic metabolic alterations with effect on their performance and health [421]. In this study, comparison of metabolic biomarkers was performed between athletes who belong to a high CVD group versus those who belong to a low/moderate CVD group. This is the first attempt to analyze over 750 metabolites in such a big cohort of elite athletes (n=495) with linear regression model incorporating different levels of CVD. Association analysis revealed 112 Bonferroni significant metabolites associated with CVD and OPLS-DA confirmed association of metabolites into two groups: 70 metabolites that showed higher levels in low/moderate CVD group of athletes and 40 metabolites that showed higher levels in high CVD group of athletes. The findings of this study identified differences in oxidative stress scavenging mechanisms, energy utilization, and membrane dynamics

between the two studied CVD groups. Metabolic alterations in high CVD group may indicate elevated cardiovascular risk, hypothetically due to elevated blood pressure and exercise-induced left ventricular hypertrophy [452].

By focusing on higher levels of 40 metabolites that associated with high CVD group of athletes, energy-related metabolites such adenine, the building block of ATP used in cellular metabolism as the main source of energy, were apparent. Creatine was also elevated in the high CVD group. It plays an important role in ATP recycling mainly in muscles through donation of phosphate groups to adenosine diphosphate (ADP). Creatine also plays an important role as a pH buffer in skeletal muscle tissues [453], consequently increased creatine is consistent with the intensive training associated with high CVD levels. Remarkably, levels of imidazole lactate were positively associated with levels of creatine as indicated in the GGM subnetworks in Figure 38F, possibly indicating enrichment of lactate in exercising muscle. Lactate is also used in energy generation via its oxidation to pyruvate in well-oxygenated muscle and heart cells, followed by re-entry into TCA cycle [454]. Glutamine was also elevated in high CVD group. It is mostly synthesized in the muscle tissue, accounting for 90% of all synthesized glutamine, where it can serve as a source of cellular energy next to glucose [455]. Carnitine was also increased in the high CVD group. Carnitine plays an important role as a transporter of long-chain fatty acids into the mitochondria to be oxidized and produce energy in exercising skeletal muscles [456]. However, since there was a negative correlation between carnitine associated with several fatty acids and free carnitine levels (Figure 38C), fatty acid oxidation was lower in high CVD group of athletes compared to their low/moderate counterpart.

Furthermore, there were specific signaling molecules elevated in high CVD group. Among those three different diacylglycerols containing arachidonic acid (C20:4) were increased with high CVD group of athletes including palmitoleoyl-arachidonoyl-glycerol, linoleoyl-arachidonoyl-glycerol, and palmitoyl-arachidonoyl-glycerol. Arachidonic acid retention is highly active in skeletal muscle, accounting for 10-20% of the phospholipid fatty acid content on average [457]. In addition to its crucial role as a second messenger involved in regulating various signaling enzymes from the phospholipase C family (PLC-γ, PLC-δ, and PKC-α/β/γ), arachidonic acid plays a critical role as a primary inflammatory intermediate [457].

Another metabolite found to be elevated in high CVD group of athletes is the plasmologens. This group of metabolites constitute up to 20% of the total phospholipid mass in humans and greater than or equal to 50% of the ethanol amines fraction in the brain, heart, neutrophils and eosinophils [458]. They play important roles as modulators of membrane dynamics and as signaling molecules, mediating various signaling processes, providing unique structural properties, and guarding membrane lipids from oxidation [458]. GGM subnetworks revealed positive correlations between various plasmologens, GPEs and GPCs (Figure 38A), indicating biochemical relationships between these various phosphatides.

Also, cortisol was elevated in high CVD. It is related to glucocorticoid type of hormones that are secreted during stress and low blood glucose. It plays significant roles in stimulating gluconeogenesis to elevate blood sugar [459], immunomodulation and metabolism of carbohydrates, fat, and protein [460]. Cortisol is also important to keep normal blood pressure and extra high levels of cortisol can cause hypertension [461].

Therefore, the elevated cortisol in high CVD group of athletes could indicate their low blood glucose in response to increased stress triggers increase in their blood pressure. Remarkably, GGM subnetwork analysis (Figure 38E) also showed a positive association among cortisol and other steroids such as pregnenolone, an intermediate in the biosynthesis of most of the steroid hormones, indicating elevated steroid biosynthesis revealed earlier in endurance athletes [167].

Likewise, an increase in branched chain amino acids (BCAAs, valine, leucine, and isoleucine) metabolites was revealed in the high CVD group of athletes as demonstrated by increased levels of 4-methyl-2-oxopentanoate, 3-methyl-2-oxovalerate, and 3-methyl-2-oxobutyrate. Exercise promotes energy expenditure and activates oxidation of BCAAs that act as substrates to TCA intermediates and gluconeogenesis [462]. Similarly, leucine acts as a regulator of intracellular signaling pathway, increasing muscle-protein synthesis in vivo [463].

Even if exercise is associated with enhancement of body function, it can also be risky for heart function and can cause diseases such as myocardial infarction, arrhythmias, aortic dissection and sudden cardiac arrest [464]. Increases in some metabolites associated with high CVD like PC, PE, and BCAAs may also indicate an elevated risk of cardiovascular disease in high CVD groups of athletes. Whereas PC were previously shown to be correlated with elevated cardiovascular mortality independent of traditional risk factors [465], PE (precursors of PC) have been revealed earlier to be amongst the strongest predictive lipid species for risk of cardiovascular disease [466]. BCAAs were also previously associated with increased risk of cardiovascular disease [467]. Whether these associations are actual signatures of elevated risk of cardiovascular disease in high

CVD group of athletes or just an indication of differences in dietary requirement and exercise compared to low/moderate CVD group of athletes remains to be investigated. Nevertheless, it is possible to accept that high CVD group of athletes could manifest indicators of higher risk of cardiovascular disease due to their intensive physical training that is linked to functional, electrical, and structural myocardial adaptations [419, 420], initiating left ventricular hypertrophy and causing high blood pressure [452].

The 70 elevated metabolites associated with low/moderate CVD group of athletes include metabolites indicative of sterol biosynthesis, energy generation, and oxidative stress. The increase in several metabolites like acylated carnitines, DAGs, and fatty acid-carnitines suggests improved hydrolysis of diacylglycerols, followed by transfer of fatty acids inside the cells for oxidation and energy production [440]. These alterations indicate that low/moderate CVD group of athletes adapt by increasing fatty acids beta oxidation for energy generation and by exhibiting a greater potential to activate lipolysis during exercise than high CVD group of athletes. Moreover, enrichment of acylated carnitine in low/moderate CVD group of athletes may suggest a better exercise recovery, because carnitine can lower prolonged exhaustion by lowering plasma lactate [468]. Remarkably, four DAGs containing oleic acid, C18:1 (arachidonic acid precursor) were elevated in low/moderate CVD group (Table 25). Oleic acid has a hypotensive effect that partially explains the increase of this metabolite in athletes of this group that is characterized by lower blood pressure [469]. In addition, eicosanoids (products of arachidonic acid) were also elevated in low/moderate CVD group including Leukotriene B4 & B5 and 5 & 12 HETE and were positively associated as showed in Figure 38B. Eicosanoids play a key

role in several pathological and physiological processes including platelet aggregation and immunomodulation [470, 471].

Similarly, 5-oxoproline, a metabolite of glutathione cycle, beta-citrylglutamate, glutamate, and gamma glutamyl amino acids were higher in low/moderate CVD group of athletes indicating improved anti-oxidative stress scavenging mechanism [472]. The GGM subnetworks revealed a strong positive association among various gamma glutamyl amino acids, gamma glutamylglutamate and glutamate (Figure 38D), confirming their biochemical and functional relationship.

4.4.1.5 Metabolites associated with athletes' supplement consumption

Athletes' ideal dietary requirements are dictated by their training regimen and sport-related energy demand as well as their own metabolic needs. Maintaining athletes' optimal dietary requirement is important because inadequate nutrition could compromise their physical performance and lower their ability to exercise and recovery from injury. Such comprehensive nutritional need differs between various sport groups although the most challenging finding when reviewing the literature is the shortage of such data [473]. As a compensation for nutritional deficiencies, athletes are frequently taking supplements that boost their nutritional consumption and fulfil their optimal energy demand. Still, the potential adverse effects of supplements and their effectiveness remain speculative. Information concerning athletes' supplement consumption is somewhat scarce and mostly relays on surveys and interviews. In this study, profiling of supplement consumption in elite athletes was attempted by identifying xenobiotics in their serum samples received from Anti-Doping Laboratories. Various xenobiotics that possibly originated from food products, supplements, drugs, and other chemical contaminants

were found significantly different between sport disciplines. Even if the potential effects of supplements and their exact sources remain questionable, this study provides a snapshot of xenobiotics existing in different sport disciplines that may be beneficial or infelicitous to athletes.

Xenobiotics increased in athletics: Higher levels of both stachydrine and eugenolsulfate were found in Athletics group of athletes. These have possibly originated from supplements and/or food products. Stachydrine is contained in various citrus fruits and several supplements that claim to provide soothing effects promoting calm and relief from anxiety. It plays an osmoprotective role in the kidneys and in animal models it exerted anti-oxidative stress and anti-inflammatory effects [474]. There was a partial correlation showed in GGM sub-networks between methyl glucopyranoside (alpha/beta) and stachydrine, also increased in athletics, indicating a similar source, possibly orange juice as previously shown [475]. Eugenol is also a strong anti-oxidant found in many spices, herbs, and plants especially in clove but is also present in some supplements promoting reduction of risk of gingivitis and heart disease and providing blood purification [476]. 2, 3-Dihydroxy-isovalerate is also increased in athletics, and has been known as a substrate of dihydroxyacid dehydratase that is sensitive to nitric oxide [477]. 4-hydroxyhippurate is also elevated in athletics, and it is known as a microbial end-product derived from polyphenol metabolism by the microflora in the intestine.

Xenobiotics increased in football: Higher levels of several xenobiotics were found associated with footballers. Those possibly originated from supplements and/or food products. These involved caffeic acid (3,4-dihydroxycinnamic acid) that can be consumed as a supplement as well as found naturally in many plants such as sage, thyme,

spearmint, star anise, cinnamon, black chokeberry, coffee and tea. Caffeic acid-treated animals had enhanced hepatic oxidation, higher exercise tolerance, and lower blood lactate [478]. The derivative of caffeic acid (caffeic acid phenethyl ester) was shown previously to protect against hyperthermal stress induced by prolonged exercise [479]. Caffeic acid has also showed an anti-oxidant effect in both vivo and vitro [480, 481]. Ferulic acid 4-sulfate was also high among footballers, and it is known as a strong ubiquitous plant anti-oxidant found in high concentration in oranges, peanuts, apples, wheat, and rice [482] and in oral supplement form. Ferulic acid is a direct product of caffeic acid in plants and can be generated from tyrosine and phenylalanine metabolism. Other compounds with higher concentrations in footballers involved drugs like ectoine, eye drops or a nasal spray, frequently used in allergic rhinoconjunctivitis and rhinitis treatment for relief of sneezing and nose block [483], possibly as a result of continuous exposure to grass. Quinine, is another xenobiotic that is also a drug and found increased in footballers. It is an alkaloid synthesized in plants and is the active component of quinate that is used for treating muscle cramps between footballers [484]. A benzoate metabolite, hippurate, was also increased between footballers. Hippurate is rich in fruits and whole grains and has been found associated with lower risk of metabolic syndrome [485]. From the GGM sub-networks, hippurate concentrations were partially associated with other benzoate metabolites including O-methyl catechol sulfate and catechol sulfate, both found in higher levels in football. Other xenobiotics increased in footballers included 4-vinylguaiacol sulfate, an additive flavor in beer, also revealed in GGM sub-network in partial correlation with ferulic acid. 2-furoyl.glycine was elevated in footballers, and also found partially

associated with quinate. A minor metabolite of caffeine called 1,3,7-trimethylurate, was also increased in football.

Xenobiotics increased in boxing: In this small study of female boxers (n=17), association analysis has revealed three significantly increased xenobiotics. One of the compounds was retinol (vitamin A), a dietary supplement that has strong anti-oxidant effects and is contained in anti-aging creams [486]. 2.pyrrolidinone was also elevated in boxers. It constitutes the simplest γ-lactam with nootropic effects, proving efficient as an antiepileptic agent and enhancing neuroprotection after stroke [487]. The existence of this drug in boxers may indicate a prophylactic treatment for multiple head injury. Another xenobiotic increased in boxers was thioproline. It is an intracellular sulfhydril antioxidant and free radical scavenger that improves immune functions. Results from in vivo studies in mice revealed that thioproline provides an anorexic effect correlated with a decreased oxidative damage and better neurological function as well as greater survival [488].

Athletes from different sport disciplines may intake different nutritional supplements according to their nature of physical performances and the obligatory outcomes from the nutritional supplements, aiming eventually for improved recovery from exercise and injury. Our findings provide evidence that athletes with different sports types show a different xenobiotic profile that may indicate their exposure to some chemicals, food, and drug/supplement. These metabolites possibly provide an anti-oxidant effect, a nootropic effect, and enhance prolong exercise tolerance, or originate from drugs for different types of treatments. GGM showed metabolic networks that associated several xenobiotics correlated with different sport groups, providing additional evidence of functional relationship between these xenobiotics. To validate these metabolic differences,

replication studies are required to verify the associations with different sport disciplines in independent data sets, targeting for monitoring of athletes' exposure to several environmental chemicals and assessment of their supplement use that might be beneficial or cause harm to their health and career.

4.4.1.6 Metabolomics study limitations

Since athletes' blood samples were collected at different sites and at multiple times (IN and OUT of competition), several confounding factors were unavoidable. For instance, batch effect was inevitable, and might have lowered the associations between metabolite levels and sports groups. This batch effect may have impacted several critical pre-analytical features that can significantly influence the metabolic profiling of samples including the blood collection procedure, transportation conditions, such as period to reach anti-doping laboratories, sample handling and storage [489]. Despite these confounders, clear signatures were revealed after correcting for potential covariates. Furthermore, the lack of information about studied group of athletes such as their body mass index, origin and age was a penalty this study has endured to access large number of elite athletes. However, since most of national and international sports competitions include athletes at young age as well as the wide range of sports involved in the study analysis may have decreased the effect of other possible confounders.

Ambiguity of sports classifications based on earlier published work was another issue as differences exist between different members of the same team, for instance, football midfielders and goal keepers. Strikers, defenders, and mid-fielders differ in their CVD levels due to their intensity of exercise and style of playing. For example, playing with change of direction is known to elicit higher levels of CVD and causes greater metabolic

changes compared to running in a straight line with the same mean speed [490, 491]. Another confounder in this study is the number of participants in classified group of athletes. Some classes were underrepresented and other overrepresented. Finally, differences in nutritional consumption between moderate and high endurance and power elite athletes, including several other ergogenic and supplements, may have significant effect on their metabolic profile [492]. Such alterations are difficult to take into account as they differ between different sport groups and participants and are not generally revealed. Taking all these restrictions and limits on board, data of this study require further validation and confirmation in other sports group of athletes.

Conclusion: In summary, our pilot study of 191 elite athletes identified a number of metabolites associated with endurance and power [167]. By carrying out metabolomics in a second cohort of 490 elite athlete we confirmed a number of these metabolites, followed by a meta-analysis of the two cohorts. These metabolic changes seen in high performance elite athletes may reflect various cellular adaptations to prolonged exercise-induced oxidative stress. These may include modulation of energy utilization, muscle mass and deployment of stress-scavenging mechanisms as previously suggested [167]. CVD metabolites data also indicate an elevated risk of heart disease in the high CVD group of athletes compared to low/moderate CVD group of athletes. Metabolomics findings present a snapshot summarizing differences among classified sports groups of elite athletes with metabolic differences. Replication studies are desired to validate these findings in independent data sets, aiming for discovery of metabolic signatures for assessing performance, recovery, and health of elite athletes. Such metabolic signatures could be used as screening tools for young athletic candidates by revealing their extreme

trainability potential associated that could predict their performance. Metabolic profiling in elite athletes could also help guiding training programs and avoiding potential disorders that can be caused by intensive training.

CHAPTER 5
Results: mGWAS

5 (Chapter 5) Results: mGWAS analysis

5.1 Introduction

Following completion of GWAS and Metabolomics, assessment of genetically influenced metabotypes (GIM) has been expedient, where metabolites served as intermediate phenotypes with a greater effect size [145, 165]. The coupling of genetics to metabolomics (mGWAS) has been proven a powerful tool for clinical translation [493-495]. It has been used to reveal molecular mechanisms that link human genetic variation to inter-individual differences in metabolite concentrations [494, 496]. Utilizing mGWAS approach [146, 150, 156-164] in elite athletes may enlighten the environmental exposures affecting athletes and provide information related to athlete's health and training requirements. Indeed, theses robust and quantitative information can explain environmental conditions, special diet as well as type and intensity of exercise that best fit to athletes' innate capabilities [167]. mGWAS results could serve as important tools in biomarker discovery associated with exercise that can be used in selection of athletic candidates. It could also help in developing anti-doping analytical strategies that include gene doping of specific performance enhancing metabolites.

The aims of this chapter are (1) to discover novel genetically influenced metabolites in elite athletes by fine-mapping loci to putative functional variants at or near sentinel SNPs (a sentinel SNP or sentinel metabolite refers to a lead SNP or a lead metabolite), (2) to discover novel variant loci associated with significant metabolites identified from metabolomics associated with high endurance, power and CVD group of athletes, and to

perform functional validation of significant endurance-related SNPs by identifying associated metabolites from mGWAS results.

5.2 Material and method

5.2.1 Sample classification, metabolomics and genomics data extraction

mGWAS analysis was conducted in only four hundred and ninety (490) elite athletes who were included in both genetic and metabolic studies. Categorization of participants' groups (Low/moderate vs. high endurance, power and CVD) is shown in Table 27.

Table 27. Classification of 490 study participants. Distribution of elite athletes in various categories based on trainability as described previously [16]. The number and gender (M for males and F for females) of participants in each group are indicated.

mGWAS		Endurance	
		Moderate (40-70% VO2max)	**High (>70% VO2max)**
Power	**High (>50% MVC)**	Wrestling (3M), Judo (3M)	Boxing (1M/16F), Heptathlon (1M), Rowing (6M/7F), Cycling (31M/4F)
	Moderate (20-50% MVC)	Athletics (15M/22F), Rugby (16M), Triple Jump (1M)	Athletics 200-800m (4M), Hockey (1F), Skiing Cross Country (1M), Basketball (3M), Swimming (22M/16F)
	Low (<20% MVC)	Baseball (2M), Volleyball (1M)	Tennis (1M/1F), Soccer (315M), Athletics 1500-3000m (3M)

As discussed in chapter 3, section 1.2.3, SNP array genotyping was performed using Illumina Drug Core array-24 BeadChips. Corresponding metabolomics data were collected from Metabolon, Durham, NC, USA as detailed in chapter 4 section 1.2.2.

5.2.2 Statistical analysis of combined metabolomics and genomics data

The genotyping data (476728 SNPs from each sample) were filtered using Plink v1.9 as follows: SNPs with genotype call rate < 98% were excluded (130526 SNPs). SNPs with minor allele frequency (MAF) < 0.01 were excluded (70210 SNPs). SNPs exhibiting Hardy Weinberg p value < E-6 were excluded (976 SNPs). Similarly, metabolites with more than 50% missingness were excluded (87 metabolites). After quality control filtering, only 275016 SNPs and 751 metabolites were used in mGWAS analysis, providing a 2.4E-10 Bonferroni level of significance ($p \leq 0.05/(275016 \times 751)$). Linear regression model associations between SNPs and metabolite were performed using "lm" function in R (version 3.3.1) while correcting for gender, hemolysis, and genotypic population stratification (using two genomic PCA components that were calculated with plink version 1.9).

An average inflation factor was calculated for mGWAS metabolites using chi2=beta*beta/(sebeta*sebeta) and lambda=median(chi^2)/0.45) [184]. Percent of explained variance (r^2) was calculated with formula: $r^2 = X^2/(N - 2 + X^2)$, where N is the number of samples and X^2 = (Beta/standard error of the beta)^2, values obtained from the regression analysis [165]. Manhattan and boxplots were generated using R (version 3.3.1). Regional association plots were produced using SNIPA (http://snipa.helmholtz-muenchen.de/snipa/). Pathway enrichment analyses were carried out using Chi square tests (https://www.mathsisfun.com/data/chi-square-calculator.html) as explained in chapter 2 section 1.3.4.

5.3 Results

5.3.1 Loci and sentinel SNPs

By combining genomics and metabolomics analyses, 145 significant associations (Bonferroni $p \leq 2.4\,E{-}10$) were identified, with an average inflation factor for mGWAS metabolites of 1.07 (0.96 - 1.19). Table 28 lists 145 mGWAS association results.

Table 28. Metabolite quantitative trait loci (mQTL) associated with elite athletes (Bonferroni significance $p \leq 2.4\,E{-}10$).

SNP	Gene	Beta	P.Value	Metabolite
rs7072216	*PYROXD2*	-0.96	1.23E-59	N-methylpipecolate
rs1061115	*HPS1*	-0.89	4.05E-50	N-methylpipecolate
rs1881245	*NAT8*	0.87	5.76E-39	N-acetyl-1-methylhistidine*
rs1881245	*NAT8*	0.85	2.98E-37	N-acetylasparagine
rs7924303	*PYROXD2*	-0.82	5.72E-37	N-methylpipecolate
rs2066938	*UNC119B*	0.82	6.16E-37	ethylmalonate
rs1881245	*NAT8*	0.83	9.31E-36	N-acetylarginine
rs6759452	*NAT8*	0.93	2.23E-34	N-acetyl-1-methylhistidine*
rs1881245	*NAT8*	-0.77	4.12E-32	N-delta-acetylornithine
rs6584202	*PYROXD2*	-0.82	6.69E-32	N-methylpipecolate
rs12763326	*PYROXD2*	-0.81	1.85E-31	N-methylpipecolate
rs6759452	*NAT8*	0.89	3.96E-31	N-acetylarginine
rs45446698	*CYP3A7, ZKSCAN5*	-1.92	4.62E-31	androsterone sulfate
rs2421668	*NAT8*	0.68	9.91E-29	N-acetylasparagine
rs12472502	*NAT8*	0.67	1.44E-28	N-acetylasparagine
rs7586361	*NAT8*	0.72	2.54E-28	N-acetylasparagine
rs10431384	*MLEC*	0.67	9.75E-28	ethylmalonate
rs6759452	*NAT8*	0.83	6.22E-27	N-acetylasparagine
rs12995433	*NAT8*	0.66	1.28E-26	N-acetylasparagine
rs4852977	*NAT8*	0.67	1.46E-26	N-acetylarginine
rs12472502	*NAT8*	0.65	1.94E-26	N-acetylarginine
rs6759452	*NAT8*	-0.81	2.97E-26	N-delta-acetylornithine
rs2421668	*NAT8*	0.66	5.50E-26	N-acetylarginine
rs11126414	*NAT8*	0.66	6.35E-26	N-acetylarginine
rs2421668	*NAT8*	-0.64	1.09E-25	N-delta-acetylornithine

rs6429759	*AGMAT*	0.74	2.98E-25	beta-guanidinopropanoate
rs6546824	*NAT8*	0.69	3.93E-25	N-acetylasparagine
rs12995433	*NAT8*	-0.63	5.49E-25	N-delta-acetylornithine
rs4767935	*SPPL3*	0.65	6.44E-25	ethylmalonate
rs12472502	*NAT8*	-0.62	1.33E-24	N-delta-acetylornithine
rs7586361	*NAT8*	-0.66	1.70E-24	N-delta-acetylornithine
rs45446698	*CYP3A7*	-1.68	2.50E-24	epiandrosterone sulfate
rs4852977	*NAT8*	0.63	7.36E-24	N-acetylasparagine
rs1881245	*NAT8*	0.74	7.59E-24	N2-acetyllysine
rs45446698	*CYP3A7*	-1.76	9.66E-24	5alpha-androstan-3alpha,17beta-diol monosulfate (1)
rs11126416	*NAT8*	0.64	1.01E-23	N-acetylarginine
rs12995433	*NAT8*	0.63	1.48E-23	N-acetylarginine
rs9204	*ACADS*	0.63	3.02E-23	ethylmalonate
rs11126414	*NAT8*	0.62	3.25E-23	N-acetylasparagine
rs7586361	*NAT8*	0.66	3.50E-23	N-acetylarginine
rs2421668	*NAT8*	0.62	5.21E-23	N-acetyl-1-methylhistidine*
rs12472502	*NAT8*	0.61	1.02E-22	N-acetyl-1-methylhistidine*
rs1052571	*CASP9*	-0.70	1.52E-22	beta-guanidinopropanoate
rs1052576	*CASP9*	0.70	1.88E-22	beta-guanidinopropanoate
rs1881245	*NAT8*	0.77	3.54E-22	N-acetylcitrulline
rs6546824	*NAT8*	0.65	3.62E-22	N-acetylarginine
rs6546824	*NAT8*	-0.63	5.67E-22	N-delta-acetylornithine
rs6546824	*NAT8*	0.65	5.83E-22	N-acetyl-1-methylhistidine*
rs7586361	*NAT8*	0.65	6.31E-22	N-acetyl-1-methylhistidine*
rs11126416	*NAT8*	0.61	9.33E-22	N-acetylasparagine
rs10927796	*DNAJC16*	-0.69	1.19E-21	beta-guanidinopropanoate
rs2421668	*NAT8*	0.66	1.22E-21	N2-acetyllysine
rs10492987	*DDI2*	0.66	2.34E-21	beta-guanidinopropanoate
rs12472502	*NAT8*	0.65	4.08E-21	N2-acetyllysine
rs12995433	*NAT8*	0.59	8.54E-21	N-acetyl-1-methylhistidine*
rs6759452	*NAT8*	0.79	1.30E-20	N2-acetyllysine
rs6546824	*NAT8*	0.66	4.72E-19	N2-acetyllysine
rs4852977	*NAT8*	0.56	4.32E-18	N-acetyl-1-methylhistidine*
rs7586361	*NAT8*	0.64	4.46E-18	N2-acetyllysine
rs4852977	*NAT8*	-0.54	7.27E-18	N-delta-acetylornithine
rs11126416	*NAT8*	0.56	9.21E-18	N-acetyl-1-methylhistidine*
rs6759452	*NAT8*	0.78	2.26E-17	N-acetylcitrulline
rs4852977	*NAT8*	0.60	2.76E-17	N2-acetyllysine
rs1881245	*NAT8*	0.57	3.74E-17	N-acetylglutamine
rs11568825	*CYP3A7*	-1.88	3.91E-17	androsterone sulfate

rs6759452	*NAT8*	0.65	4.12E-17	N-acetylglutamine
rs11126414	*NAT8*	-0.53	4.78E-17	N-delta-acetylornithine
rs1495741	none	0.74	5.48E-17	5-acetylamino-6-formylamino-3-methyluracil
rs80218044	*CYP3A5,ZSCAN25*	-1.77	5.57E-17	androsterone sulfate
rs7586361	*NAT8*	0.68	8.65E-17	N-acetylcitrulline
rs12995433	*NAT8*	0.58	1.77E-16	N2-acetyllysine
rs2421668	*NAT8*	0.51	1.88E-16	N-acetylglutamine
rs1495743	none	0.73	3.08E-16	5-acetylamino-6-formylamino-3-methyluracil
rs11126414	*NAT8*	0.52	3.43E-16	N-acetyl-1-methylhistidine*
rs11126416	*NAT8*	-0.51	8.19E-16	N-delta-acetylornithine
rs2421668	*NAT8*	0.63	1.45E-15	N-acetylcitrulline
rs11126414	*NAT8*	0.56	1.65E-15	N2-acetyllysine
rs6546824	*NAT8*	0.66	1.94E-15	N-acetylcitrulline
rs12472502	*NAT8*	0.48	2.20E-15	N-acetylglutamine
rs1495747	*NAT2*	0.68	2.24E-15	5-acetylamino-6-formylamino-3-methyluracil
rs1495741	none	0.56	3.09E-15	N-acetylputrescine
rs12472502	*NAT8*	0.62	3.77E-15	N-acetylcitrulline
rs11126416	*NAT8*	0.56	6.37E-15	N2-acetyllysine
rs12130370	*CELA2B*	-0.61	7.23E-15	beta-guanidinopropanoate
rs80218044	*CYP3A5,ZSCAN25*	-1.82	1.06E-14	epiandrosterone sulfate
rs1961456	*NAT2*	0.66	2.15E-14	5-acetylamino-6-formylamino-3-methyluracil
rs12995433	*NAT8*	0.61	2.35E-14	N-acetylcitrulline
rs6685648	*CASP9*	-0.66	2.59E-14	beta-guanidinopropanoate
rs10168416	*UGT1A10*	0.50	2.67E-14	biliverdin
rs1495747	*NAT2*	0.52	4.16E-14	N-acetylputrescine
rs2070959	*UGT1A10*	0.49	4.56E-14	biliverdin
rs7586361	*NAT8*	0.50	5.33E-14	N-acetylglutamine
rs1495743	none	0.54	6.28E-14	N-acetylputrescine
rs12210538	*SLC22A16*	-0.54	8.67E-14	dihomo-linolenoylcarnitine (20:3n3 or 6)*
rs1105880	*UGT1A10*	0.47	1.00E-13	biliverdin
rs6546824	*NAT8*	0.49	1.05E-13	N-acetylglutamine
rs174547	*FADS1*	-0.48	1.18E-13	1-arachidonoyl-GPC (20:4n6)*
rs76657705	*CYP3A43*	-1.29	1.20E-13	androsterone sulfate
rs174550	*FADS1*	-0.48	1.34E-13	1-arachidonoyl-GPC (20:4n6)*
rs12995433	*NAT8*	0.46	1.97E-13	N-acetylglutamine
rs174546	*FADS1*	-0.48	2.40E-13	1-arachidonoyl-GPC (20:4n6)*
rs45446698	*CYP3A7*	-1.20	2.44E-13	androstenediol (3alpha, 17alpha) monosulfate (3)
rs45446698	*CYP3A7*	1.38	2.80E-13	andro steroid monosulfate C19H28O6S (1)*
rs45446698	*CYP3A7*	1.54	3.07E-13	16a-hydroxy DHEA 3-sulfate
rs3798793	*VNN1*	0.46	3.15E-13	linoleoyl ethanolamide

rs174536	*MYRF*	-0.48	4.63E-13	1-arachidonoyl-GPC (20:4n6)*
rs75859219	*SLC22A16, SLC22A24*	0.96	5.04E-13	etiocholanolone glucuronide
rs4363657	*SLCO1B1*	0.69	7.18E-13	glycochenodeoxycholate glucuronide (1)
rs11045879	*SLCO1B1*	0.70	7.69E-13	glycochenodeoxycholate glucuronide (1)
rs12988520	*UGT1A10*	-0.44	9.53E-13	biliverdin
rs680379	*SPLTLC3*	0.49	1.20E-12	sphingomyelin (d18:1/25:0, d19:0/24:1, d20:1/23:0, d19:1/24:0)*
rs6930032	*VNN1*	-0.50	1.21E-12	linoleoyl ethanolamide
rs1961456	*NAT2*	0.49	1.29E-12	N-acetylputrescine
rs887829	*UGT1A10*	0.46	2.55E-12	biliverdin
rs11568825	*CYP3A7*	-1.55	2.82E-12	epiandrosterone sulfate
rs9431	*SPPL3*	-0.42	3.55E-12	ethylmalonate
rs35307342	*TMPRSS11E*	0.47	6.48E-12	5alpha-androstan-3alpha,17beta-diol monosulfate (1)
rs3733402	*KLKB1*	0.40	6.80E-12	leucylglycine
rs174574	*FADS2*	-0.43	9.99E-12	1-arachidonoyl-GPC (20:4n6)*
rs174547	*FADS1*	-0.44	1.31E-11	1-stearoyl-2-arachidonoyl-GPC (18:0/20:4)
rs4661636	*CASP9*	0.55	1.44E-11	beta-guanidinopropanoate
rs80218044	*CYP3A5, ZSCAN25*	-1.76	1.51E-11	5alpha-androstan-3alpha,17beta-diol monosulfate (1)
rs174550	*FADS1*	-0.44	1.54E-11	1-stearoyl-2-arachidonoyl-GPC (18:0/20:4)
rs174546	*FADS1*	-0.44	1.75E-11	1-stearoyl-2-arachidonoyl-GPC (18:0/20:4)
rs11613331	*SLC6A13*	0.39	1.93E-11	deoxycarnitine
rs4241816	*KLKB1*	0.39	2.02E-11	leucylglycine
rs4852977	*NAT8*	0.42	2.09E-11	N-acetylglutamine
rs174574	*FADS2*	-0.43	2.15E-11	1-stearoyl-2-arachidonoyl-GPC (18:0/20:4)
rs55729124	*FOLH1*	-0.95	2.17E-11	N-acetyl-aspartyl-glutamate (NAAG)
rs61886483	*FOLH1*	-0.94	2.25E-11	N-acetyl-aspartyl-glutamate (NAAG)
rs2741045	*UGT1A10*	0.46	2.25E-11	biliverdin
rs11568825	*CYP3A7*	-1.53	3.31E-11	5alpha-androstan-3alpha,17beta-diol monosulfate (1)
rs1912826	*KLKB1*	0.39	4.43E-11	leucylglycine
rs10774563	none	0.41	4.64E-11	ethylmalonate
rs7258249	*CERS4*	0.40	7.17E-11	sphingomyelin (d18:1/20:1, d18:2/20:0)*
rs61886481	*FOLH1*	-0.98	7.70E-11	N-acetyl-aspartyl-glutamate (NAAG)
rs28899170	*UGT1A10*	0.43	1.03E-10	biliverdin
rs3820071	*CELA2B*	-0.62	1.31E-10	beta-guanidinopropanoate
rs4294999	*UGT1A10*	0.40	1.40E-10	biliverdin
rs17101394	*SGPP1*	0.52	1.52E-10	ceramide (d16:1/24:1, d18:1/22:1)*
rs174536	*MYRF*	-0.42	1.61E-10	1-stearoyl-2-arachidonoyl-GPC (18:0/20:4)
rs174583	*FADS2*	-0.41	1.65E-10	1-arachidonoyl-GPC (20:4n6)*
rs174601	*FADS2*	-0.40	1.81E-10	1-arachidonoyl-GPC (20:4n6)*
rs174583	*FADS2*	-0.41	1.95E-10	1-stearoyl-2-arachidonoyl-GPC (18:0/20:4)
rs3903703	none	0.44	2.38E-10	sphingomyelin (d18:1/25:0, d19:0/24:1, d20:1/23:0, d19:1/24:0)*

Association results were divided into gene loci with each locus having sentinel SNP and sentinel metabolite defined according to the SNP-metabolite association with the highest significant p-value. These SNP-metabolite associations define the metabolite quantitative trait loci (mQTLs). Table 29 lists 19 mQTLs resulted from gene-metabolite associations found by analyzing sentinel SNPs. In case of locus 5, the sentinel SNP rs3733402 with highest P-value 6.80E-12 could not be found in SNIPA regional association plotting. Therefore, another SNP rs4241816 with the second highest P-value 2.02E-11 had been considered in the regional association. Both were associated with same gene *KLKB1*.

Table 29. Nineteen unique locus-metabolite mGWAS pairs identified in 490 elite athletes. The novel gene/metabolite associations appear in red and the known gene/metabolite associations, but with novel SNPs, appear in blue. Abbreviations; SNP is single nucleotide polymorphism, MAF is minor allele frequency, SE is standard error, and r^2 is percent of explained variance. Highlighted rows indicate novel significant mGWAS. Biochemical Name* indicates compounds that have not been confirmed using reference standards, but Metabolon is confident in their identities based on exact mass and fragmentation pattern.

Locus by gene	Gene	SNP	MAF	Metabolite	P value	Beta	SE. Beta	r2 (%)	Function GVS	Comment	Reported r2 (%) [27] [25]
1	*AGMAT*	rs6429759	0.48	Beta-guanidinopropanoate	2.98E-25	0.74	0.07	28.54	Intron	Reported	2.3, NA
2	*NAT8*	rs1881245	0.27	N-acetyl-1-methylhistidine*	5.76E-39	0.87	0.06	30.92	Intron	Reported	1.8, 26.6
3	*UGT1A10*	rs10168416	0.31	Biliverdin	2.67E-14	0.5	0.06	11.68	Intron	Reported	0.5, NA
4	*TMPRSS11E*	rs35307342	0.36	5alpha-androstan-3alpha,17beta-diol monosulfate (1)	6.48E-12	0.47	0.07	9.9	Intron	Reported	1.2, 21.4
5	*KLKB1*	rs3733402	0.48	Leucylglycine	6.80E-12	0.4	0.06	9.68	Missense	Reported	8.8, NA
6	*SLC22A16*	rs12210538	0.24	Dihomo-linolenoylcarnitine (20:3n3 or 6)*	8.67E-14	-0.54	0.07	11.24	Missense	Reported	3.8, NA
7	*VNN1*	rs3798793	0.42	Linoleoyl ethanolamide	3.15E-13	0.46	0.06	10.82	Intron	Novel gene/metabolite association	NA
8	*CYP3A7*	rs11568825	0.01	Androsterone sulfate	3.91E-17	-1.88	0.21	14.11	Upstream-gene	Reported gene/metabolite	NA

				Epiandrosterone sulfate	2.82E-12	-1.55	0.22	9.95		association but not with this SNP	NA
				5alpha-androstan-3alpha,17beta-diol monosulfate (1)	3.31E-11	-1.53	0.23	9.2			NA
	CYP3A7	rs45446698	0.03	Androsterone sulfate	4.62E-31	-1.92	0.15	24.82	Upstream-gene	Reported	0.5
9	*NAT2*	rs1495741	0.27	5-acetylamino-6-formylamino-3-methyluracil	5.48E-17	0.74	0.08	21.44	Intergenic	Reported	4.9
10	*PYROXD2*	rs7072216	0.35	N-methylpipecolate	1.23E-59	-0.96	0.05	43.68	Intron	Reported	NA, 18.3
11	*FOLH1*	rs55729124	0.06	N-acetyl-aspartyl-glutamate (NAAG)	2.17E-11	-0.95	0.14	9.35	Intron	Novel gene/metabolite association	NA
12	*FADS1*	rs174547	0.3	1-arachidonoyl-GPC (20:4n6)*	1.18E-13	-0.48	0.06	11.04	Intron	Reported	10.2, NA
13	*SLC22A10, SLC22A24*	rs75859219	0.06	Etiocholanolone glucuronide	5.04E-13	0.96	0.13	10.73	Upstream-gene	Reported	4.8, NA
14	*SLC6A13*	rs11613331	0.46	Deoxycarnitine	1.93E-11	0.39	0.06	9.15	Intron	Reported	5.8, nA
15	*SLCO1B1*	rs4363657	0.14	Glycochenodeoxycholate glucuronide (1)	7.18E-13	0.69	0.09	10.53	Intron	Reported	0.9, 18.5
16	*UNC119B*	rs2066938	0.29	Ehylmalonate	6.16E-37	0.82	0.06	29.79	3-prime-UTR	Reported	1.5, NA
17	*SGPP1*	rs17101394	0.17	Ceramide (d16:1/24:1, d18:1/22:1)*	1.52E-10	0.52	0.08	8.59	Intergenic	Reported SNP association but with different metabolites	2.4, NA
18	*CERS4*	rs7258249	0.46	Sphingomyelin (d18:1/20:1, d18:2/20:0)*	7.17E-11	0.4	0.06	8.6	Upstream-gene	Reported	2.6, NA
19	*SPTLC3*	rs680379	0.35	Sphingomyelin (d18:1/25:0, d19:0/24:1, d20:1/23:0, d19:1/24:0)*	1.20E-12	0.49	0.07	10.24	Intergenic	Reported	1.5, NA

5.3.2 mQTLS associated with athletes vs. non-athletes

Genetic loci were investigated for known expression quantitative trait loci (eQTLs), mQTLs and functional associations using several databases including:

- SNIPA http://snipa.helmholtzmuenchen.de/snipa/
- PhenoScanner V2 A database of human genotype-phenotype associations http://www.phenoscanner.medschl.cam.ac.uk/
- GTEx portal (version 2.1, Build #201) www.gtexportal.org
- OMIM www.omim.org
- Overview of Bravo variant server resources https://bravo.sph.umich.edu/freeze3a/hg19/
- GnomAD http://gnomad.broadinstitute.org/

By defining the identities of the sentinel SNPs by their genes, these associations collapsed into 19 independent loci (Table 29, Figure 43). The percent of variance explained by these SNPs ranges from the highest value of 43.68% (N-methylpipecolate with rs7072216 in *PYROXD2* locus) to the lowest value of 8.59% (Ceramide-d16:1/24:1-d18:1/22:1 in *SGPP1* locus) with an average of 16.09% (Figure 44). The 19 independent loci replicated 15 previously reported loci [145, 166] (Table 29). The remaining four mGWAS loci represented novel associations between specific SNPs and metabolites. Two of these included novel gene/metabolite associations, namely rs55729124 in folate hydrolase 1 (*FOLH1*) in association with N-acetyl-aspartyl-glutamate (NAAG) and rs3798793 in vascular non-inflammatory molecule 1 (*VNN1*) in association with linoleoyl ethanolamide (Table 29, Figures 43 and 44). Other novel mGWAS involved known gene/metabolite associations of existing loci with novel SNP. These included the third novel mGWAS rs11568825 SNP within the cytochrome P450 family 3 subfamily A member 7 (*CYP3A7*) gene, exhibiting significant association with three different metabolites (androsterone sulfate, epiandrosterone sulfate and 5 alpha-androstan-3alpha, 17 beta-diol monosulfate 1). The fourth novel mGWAS represented association between rs17101394 in sphingosine-1-phosphate phosphatase 1 (*SGPP1*) gene and Cermamide. For the 4 novel mGWAS loci, there were clear genotype-dependent effects on levels of associated metabolites as shown in boxplots in Figure 45.

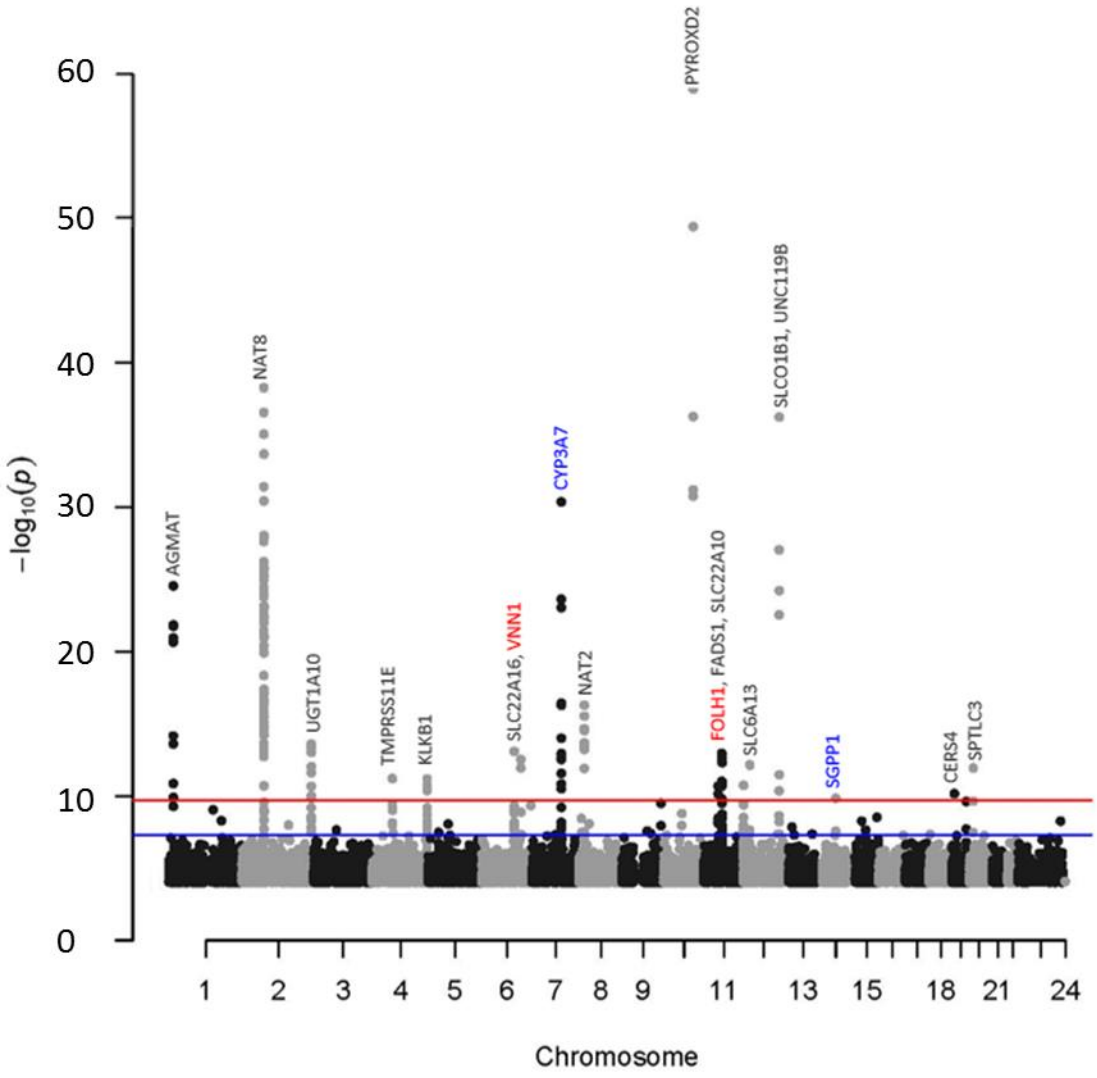


Figure 43. Manhattan plot for the discovered mGWAS loci. The red line indicates the Bonferroni threshold (2.4 E-10) and the blue line indicates the genome wide significance threshold (5 E-8). The newly identified replicated loci are typed in black, and the non-replicated novel loci are in red. The known existing loci with novel SNP is typed in blue.

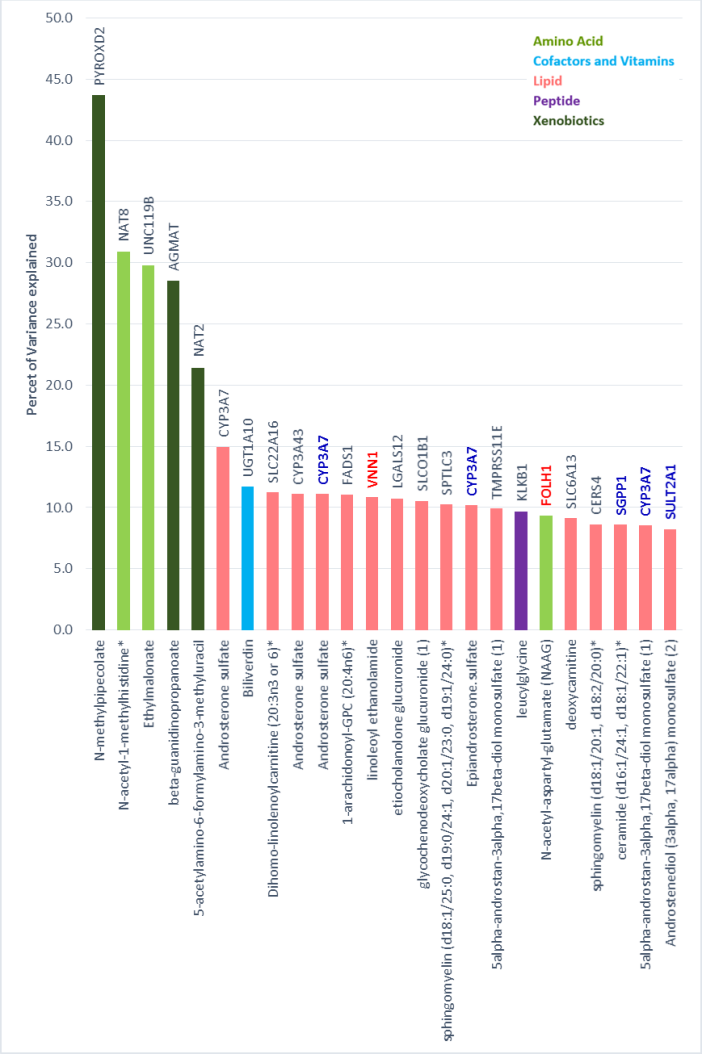


Figure 44. Percent of explained variance of metabolites by the corresponding SNP in the identified mGWAS loci in elite athletes. The height of a column bar reflects the percent of variance explained for each locus. Loci genes are indicated above the column bar and corresponding metabolite name on the X-axis. The newly identified/novel replicated loci are typed in black, and the novel mGWAS loci appear in red. The known identified replicated loci, but with novel SNP or metabolite are typed in blue. Bars are colored according to Metabolon specified pathway for the metabolites associated with the locus. Biochemical Name* indicates compounds that have not been officially confirmed based on a standard, but Metabolon is confident in their identities.

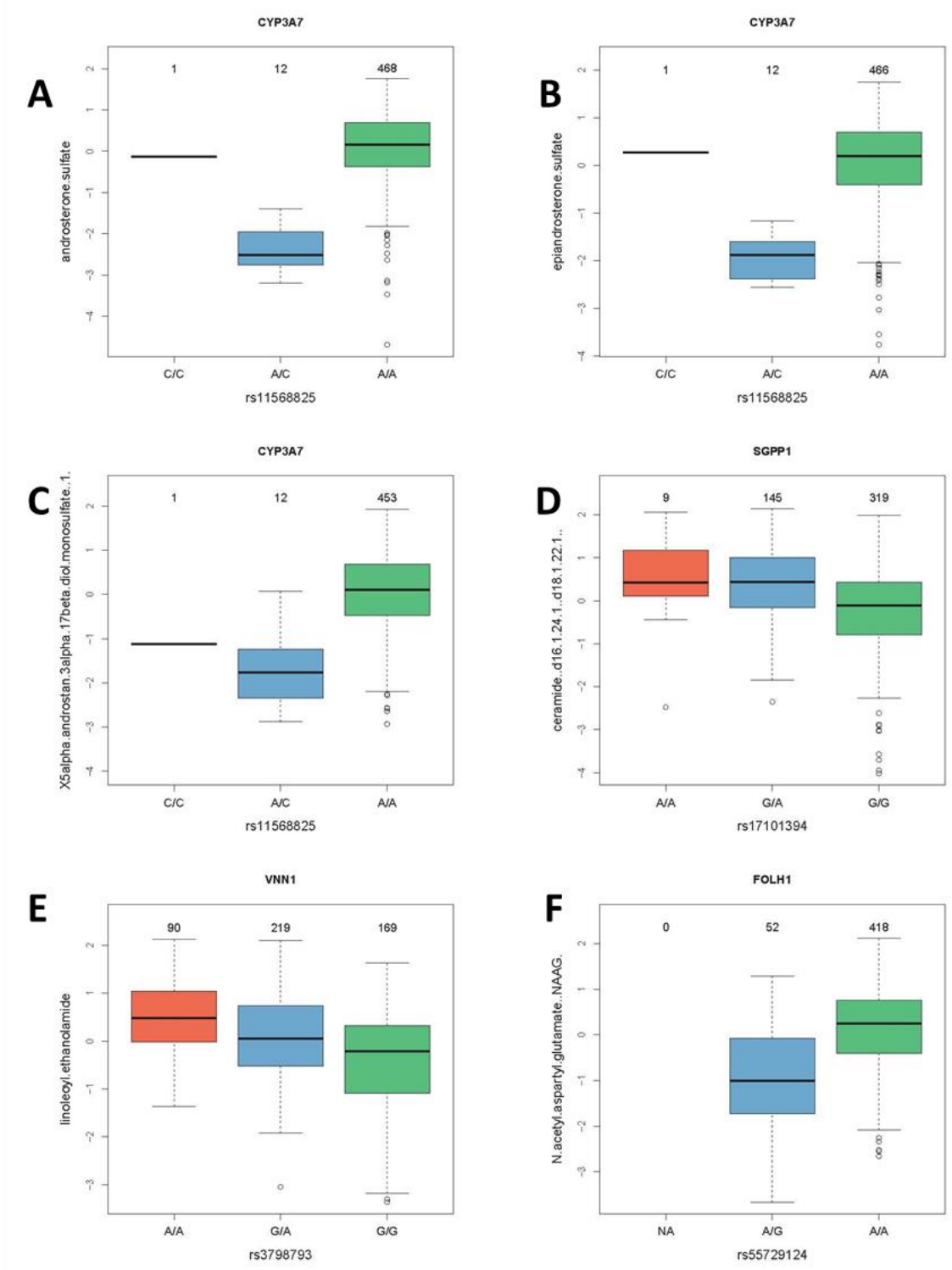


Figure 45. Boxplots of levels of metabolites by genotype for novel loci. Boxplots for the loci *CYP3A7* (A-C), *SGPP1* (D), *VNN1* (E), and *FOLH1* (F) indicating the metabolite level and the number of samples for each genotype group.

Regional association plots for the novel loci *VNN1* and *FOLH1* are shown in Figure 46. The intronic SNPs within *VNN1* (rs3798793, Figure 5a) and *FOLH1* (rs55729124, Figure 5b) loci show the strongest association (−log10 (p-value)) with linoleoyl ethanolamide and N-acetyl-aspartyl-glutamate (NAAG), respectively. The colors correspond to different linkage disequilibrium (LD) thresholds, where LD is computed between the sentinel SNP (lowest p-value, colored in blue) and all SNPs.

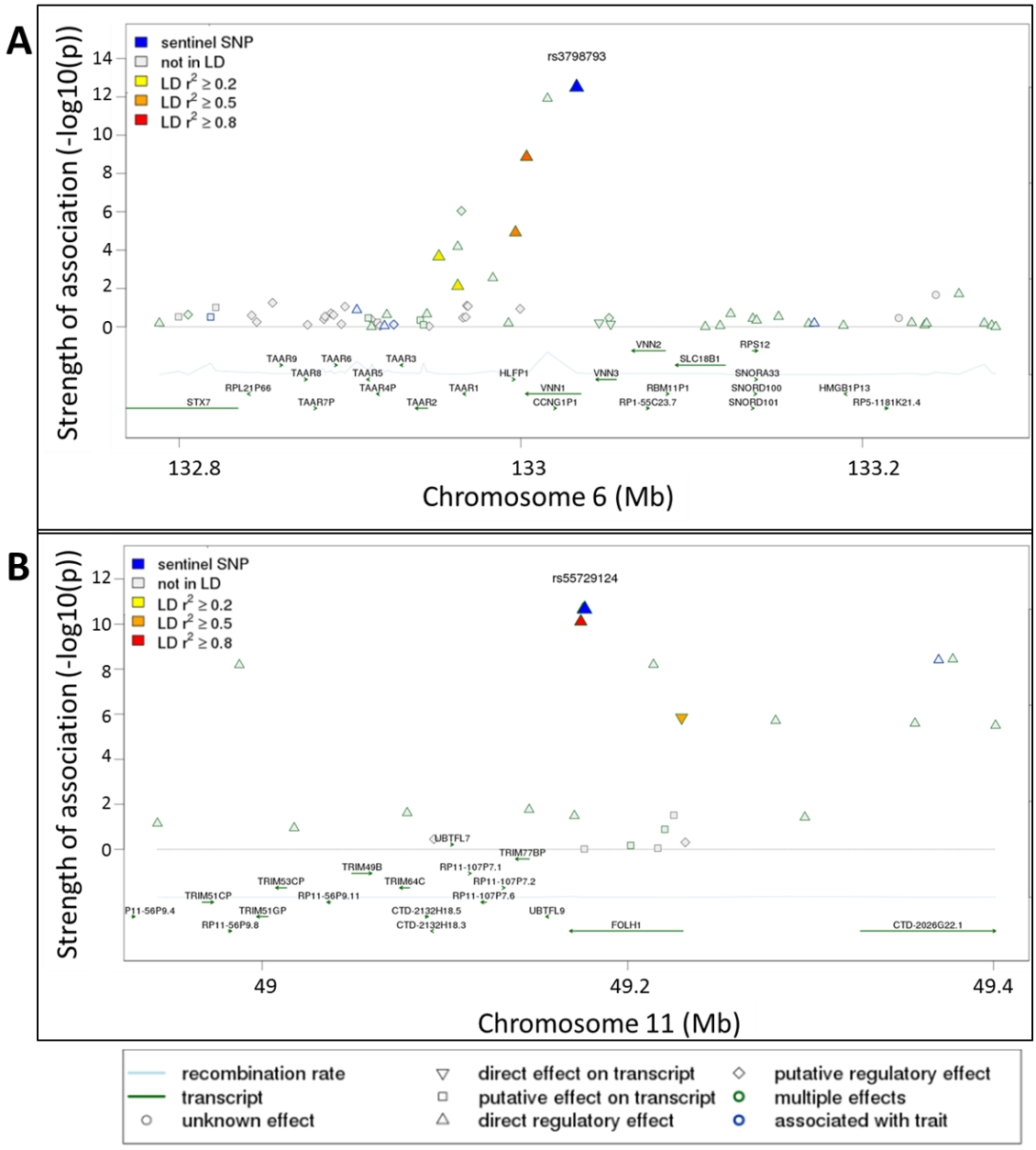


Figure 46. Regional association plots for the two new loci (*VNN1* and *FOLH1*). *VNN1* (A) and *FOLH1* (B) loci intron SNPs showing the strength of the association (−log10 (p-value)) for linoleoyl ethanolamide and N-acetyl-aspartyl-glutamate (NAAG), respectively, on the Y-axis and the genes on the X-axis. The colors resemble the difference in LD thresholds, where LD is calculated around 500Kb region between the SNPs and sentinel SNP (colored in blue). Shapes and colors of the symbols correspond to their functionality as described in the legend.

5.3.3 mQTLs associated with significant metabolites of endurance, power and CVD

Various metabolites were associated with endurance, power or CVD. In order to identify mQTLs differentiating these different groups, the corresponding SNP associations were identified from all identified mQTLs in elite athletes according to their calculated Bonferroni level of significance as follows:

- Endurance mQTLs are p <0.05/104*275016 = 1.7E-9
- Power mQTLs are p <0.05/207*275016 = 8.78E-10
- CVD mQTLs are p <0.05/112*275016 = 1.6E-9

Four endurance significant associations are shown in table 30 including one novel mGWAS association between rs10426201 in *SULT2A1* gene and androstenediol (3alpha, 17alpha) monosulfate (2). Although the latter association was reported before, it did not reach statistical significance [166]. One power and CVD significant association is shown in table 30 between rs174574 SNP in *FADS2* gene and 1-palmitoyl-2-linoleoyl-GPE (16:0/18:2), which was previously reported in non-athletes.

Table 30. Unique locus-metabolite pairs associated with endurance, power and CVD sports in comparison with previous reports [166]. Abbreviations; SNP is single nucleotide polymorphism, N is number if samples, and SE standard error.

Identified										Published [165]		Endurance		Power		Power and CVD
Gene	SNP	functionGVS	N	Beta	SE.Beta	P.Value	Metabolite	SUB_PATHWAY	Metabolite	P. Value						
SLC22A16	rs12210538	missense	457	-0.47	0.07	5.19E-10	dihomo-linoleoylcarnitine (C20:2)*	Fatty Acid Metabolism(Acyl Carnitine)	linoleoylcarnitine	1.58E-24						
SLC22A24	rs75859219	upstream-gene	462	0.96	0.13	5.04E-13	etiocholanolone glucuronide	Androgenic Steroids	etiocholanolone glucuronide	9.13E-38						
SULT2A1	rs10426201	intron	470	0.52	0.08	2.47E-10	androstenediol (3alpha, 17alpha) monosulfate (2)	Androgenic Steroids	4-androsten-3alpha,17alpha-diol monosulfate (2)	2.70E-07						
CYP3A7	rs45446698	upstream-gene	468	1.54	0.2	3.07E-13	16a-hydroxy DHEA 3-sulfate	Androgenic Steroids	16a-hydroxy DHEA 3-sulfate	2.07E-47						
CYP3A7	rs45446698	upstream-gene	473	-1.68	0.16	2.50E-24	epiandrosterone sulfate	Androgenic Steroids	epiandrosterone sulfate	3.43E-100						
CYP3A5	rs80218044	intron	473	-1.82	0.23	1.06E-14	epiandrosterone sulfate	Androgenic Steroids	epiandrosterone sulfate	4.00E-35						
SPTLC3	rs680379	intergenic	470	0.49	0.07	1.20E-12	sphingomyelin (d18:1/25:0, d19:0/24:1, d20:1/23:0, d19:1/24:0)*	Sphingolipid Metabolism	sphingomyelin (d18:1/21:0, d17:1/22:0, d16:1/23:0)	7.91E-09						
CYP3A7	rs11568825	upstream-gene	468	-1.55	0.22	2.82E-12	epiandrosterone sulfate	Androgenic Steroids	Reported gene/metabolite association but not with this SNP							
none	rs3903703	intergenic	470	0.44	0.07	2.38E-10	sphingomyelin (d18:1/25:0, d19:0/24:1, d20:1/23:0, d19:1/24:0)*	Sphingolipid Metabolism	sphingomyelin (d18:1/21:0, d17:1/22:0, d16:1/23:0)	2.28E-08						
FADS2	rs174574	intron	471	0.397855	0.062	3.36E-10	1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	Phosphatidylethanolamine (PE)	1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	6.78E-34						

For the novel endurance mGWAS locus in *SULT2A1* gene, there was a clear genotype-dependent effect on levels of associated metabolite as shown in Figure 47. The regional association plot indicates that the intronic SNP (rs10426201) in *SULT2A1* gene shows the strongest association (−log10 (p-value)) with androstenediol (3alpha, 17alpha) monosulfate (2) (Figure 48). The colors correspond to different LD thresholds, where LD is computed between the sentinel SNP (lowest p-value, colored in blue) and all SNPs.

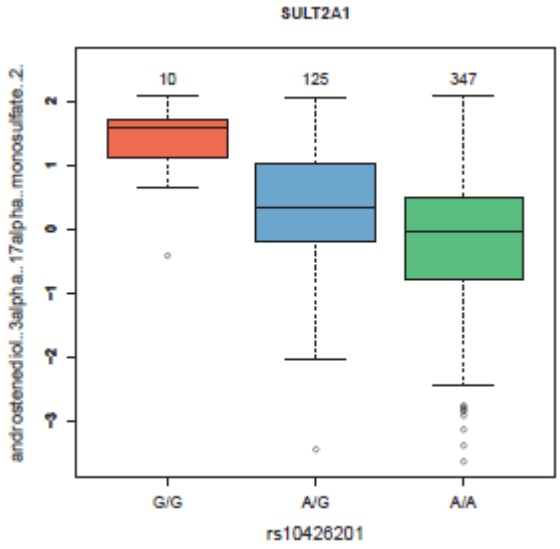


Figure 47. Boxplot for metabolite-locus pair associated with endurance, indicating the metabolite level and the number of samples for each genotype group

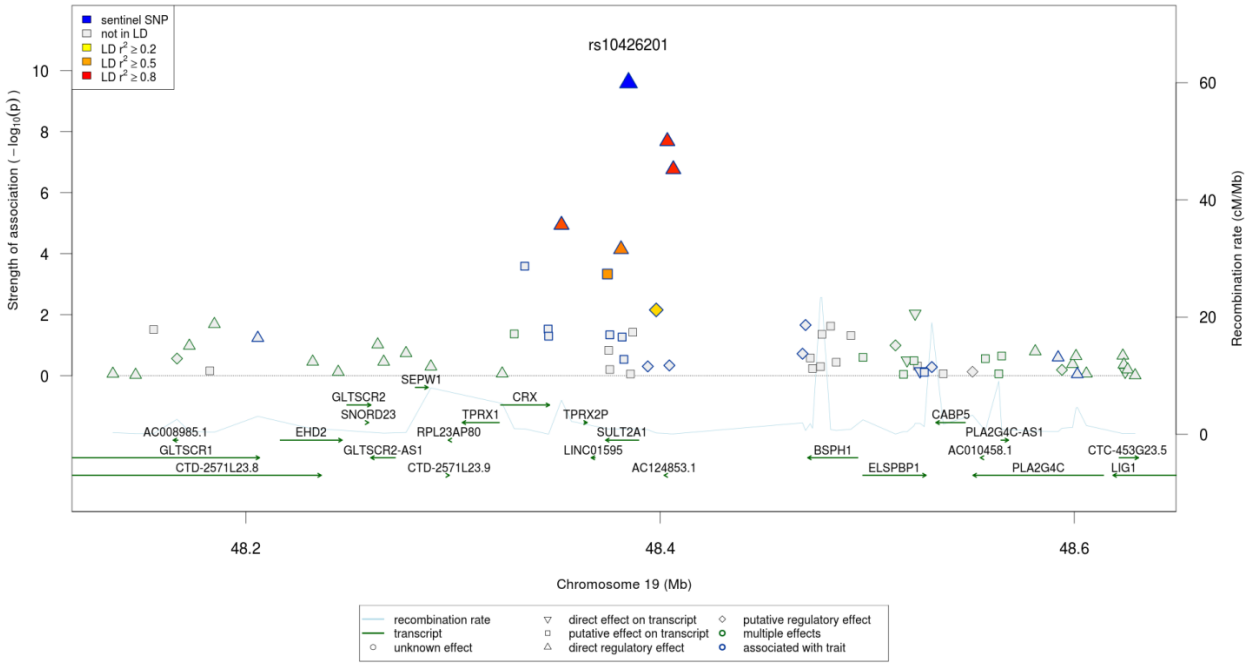


Figure 48. Regional association plots for the novel locus *SULT2A1*

5.3.4 Metabolites associated with GWAS significant SNPs of endurance

To validate the potential functionality of the identified GWAS SNPs, mGWAS results of 751 metabolites in a subset of the discovery cohort (n=490) and enriched metabolic pathways associated with the rs56330321 and rs2635438 were determined (Table 31). Among the metabolic pathways associated with rs56330321, ceramides, fatty acid (Acyl Carnitine), polyamine and creatine metabolites were significantly altered by rs56330321 genotype (Table 31, Figure 50). Whereas, gamma-glutamyl amino acid and glutamate metabolic pathways were significantly changed with rs2635438 (Table 31, Figure 51). Regional association plots for the novel loci *ATP2B2* and *SYNE1* are shown in Figure 49. The intronic SNPs within (rs56330321, Figure 49A) and (rs2635438, Figure 49B) show the strongest association (−log10 (p-value)) with loci *ATP2B2* and *SYNE1*, respectively.

The colors correspond to different linkage disequilibrium (LD) thresholds, where LD is computed between the significant SNP (lowest p-value, colored in blue) and all SNPs.

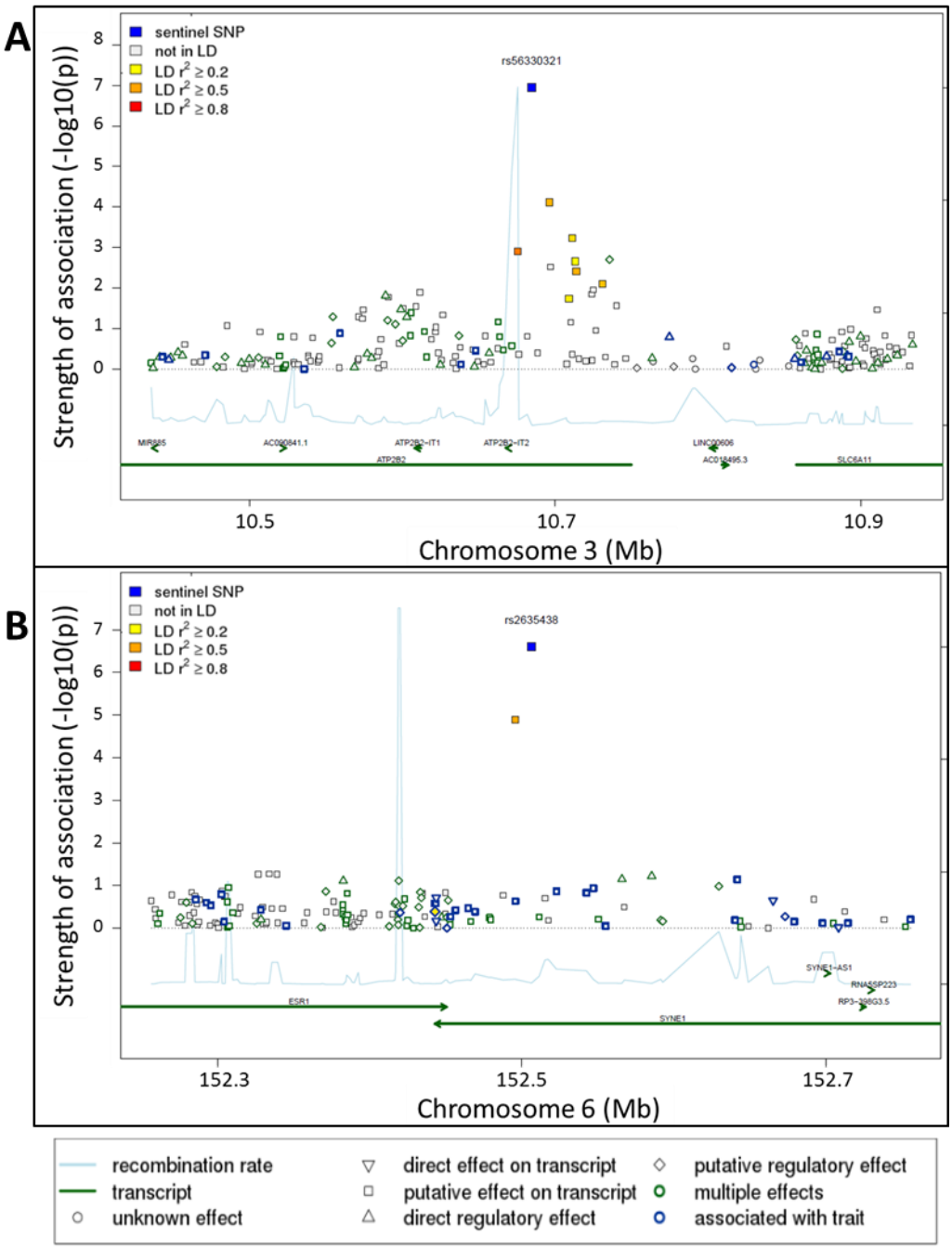
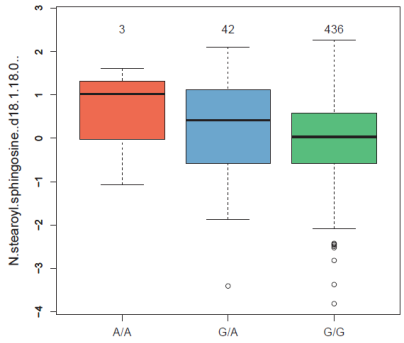


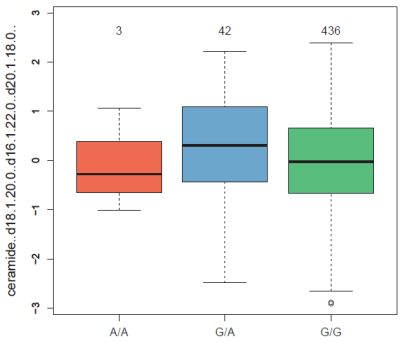
Figure 49. Regional association plots for the two new loci (*ATP2B2* and *SYNE1*). *ATP2B2* (A) and *SYNE1* (B) loci intron SNPs showing the strength of the association (−log10 (p-value)) on the Y-axis and the genes on the X-axis. The colors resemble the difference in

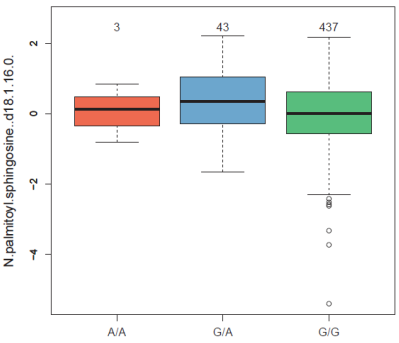
LD thresholds, where LD is calculated around 500Kb region between the SNPs and sentinel SNP (colored in blue). Shapes and colors of the symbols correspond to their functionality as described in the legend.

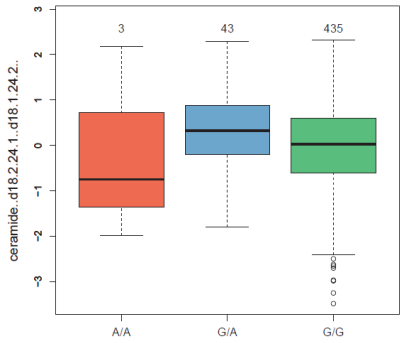
Table 31. Metabolites that belong to the significantly enriched phospholipids pathway Top metabolites associated with significant SNPs.

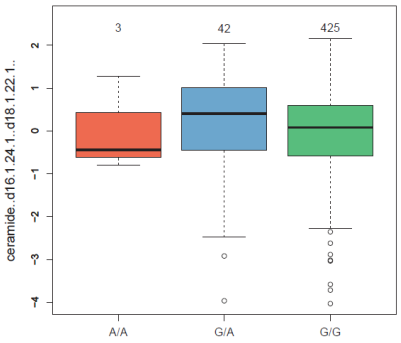
SNP	Metabolite	Beta	SE Beta	P value	Super-pathways	Enriched Sup-pathways	Enrichment p value
rs56330321	N-Stearoyl-Sphingosine (D18:1/18:0)*	0.33	0.14	0.01	Lipid	Ceramides	0.04
	Ceramide (D18:1/20:0, D16:1/22:0, D20:1/18:0)*	0.31	0.14	0.03			
	N-Palmitoyl-Sphingosine (D18:1/16:0)	0.29	0.13	0.03			
	Ceramide (D18:2/24:1, D18:1/24:2)*	0.28	0.14	0.04			
	Ceramide (D16:1/24:1, D18:1/22:1)*	0.26	0.13	0.05			
	Adrenoylcarnitine (C22:4)*	0.5	0.17	0		Fatty Acid Metabolism (Acyl Carnitine)	0.04
	Dihomo-Linolenoylcarnitine (20:3N3 Or 6)*	0.37	0.14	0.01			
	3-Hydroxybutyrylcarnitine (2)	0.36	0.14	0.01			
	3-Hydroxybutyrylcarnitine (1)	0.36	0.14	0.01			
	Palmitoylcarnitine (C16)	0.3	0.13	0.02			
	Acetylcarnitine (C2)	0.3	0.14	0.03			
	Oleoylcarnitine (C18:1)	0.28	0.13	0.04			
	Cerotoylcarnitine (C26)*	0.27	0.13	0.04			
	4-Acetamidobutanoate	0.38	0.14	0.01	Amino Acid	Polyamine Metabolism	0.03
	Acisoga	-0.46	0.2	0.02			
	N-Acetylputrescine	0.3	0.14	0.03			
	Creatine	0.32	0.14	0.02		Creatine Metabolism	0.03
	Creatinine	0.25	0.13	0.05			
rs2635438	Gamma-Glutamyl-2-Aminobutyrate	-0.58	0.21	0.006	Peptide	Gamma-glutamyl Amino Acid	0.00002
	Gamma-Glutamylhistidine	-0.5	0.19	0.01			
	Gamma-Glutamylglutamine	-0.53	0.21	0.012			
	Gamma-Glutamylalanine	-0.54	0.22	0.012			
	Gamma-Glutamylthreonine	-0.52	0.21	0.015			
	Gamma-Glutamylmethionine	-0.44	0.22	0.042			
	Gamma-Glutamyl-Alpha-Lysine	-0.42	0.21	0.045			
	N-Acetyl-Aspartyl-Glutamate (NAAG)	-0.68	0.22	0.002	Amino Acid	Glutamate Metabolism	0.0002
	Glutamate	0.43	0.18	0.021			
	Glutamine	-0.47	0.21	0.026			
	Beta-Citrylglutamate	0.45	0.22	0.041			
	Carboxyethyl-GABA	0.41	0.21	0.047			

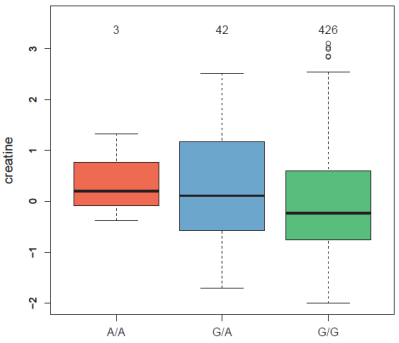

N.stearoyl.sphingosine..d18.1.18.0..
3
42
436
A/A
G/A
G/G

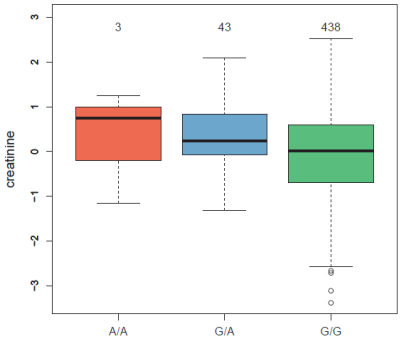

ceramide..d18.1.20.0..d16.1.22.0..d20.1.18.0..
3
42
436
A/A
G/A
G/G

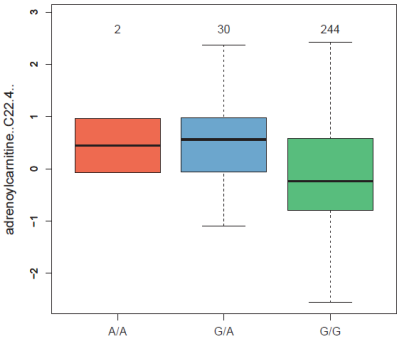

N.palmitoyl.sphingosine..d18.1.16.0.
3
43
437
A/A
G/A
G/G

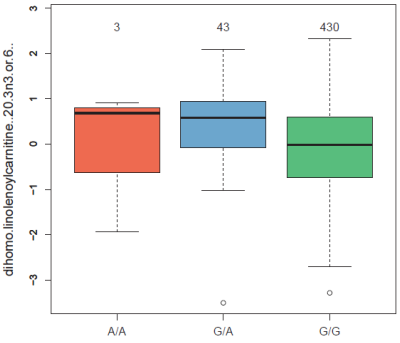

ceramide..d18.2.24.1..d18.1.24.2..
3
43
435
A/A
G/A
G/G

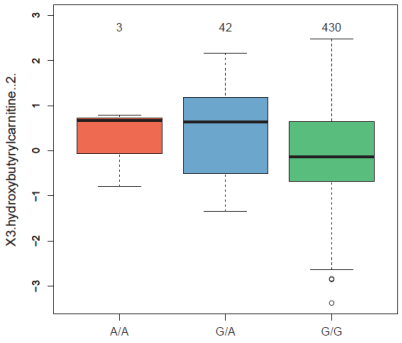

ceramide..d16.1.24.1..d18.1.22.1..
3
42
425
A/A
G/A
G/G

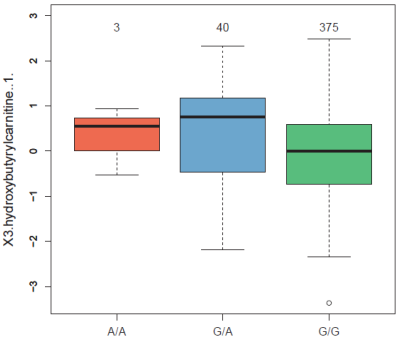

creatine
3
42
426
A/A
G/A
G/G

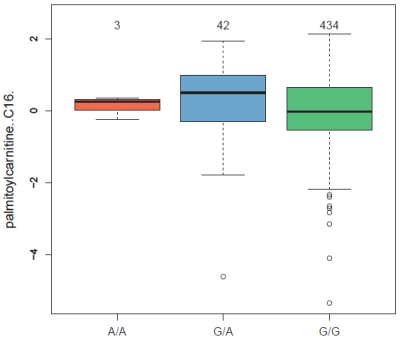

creatinine
3
43
438
A/A
G/A
G/G


adrenoylcarnitine..C22.4..
2
30
244
A/A
G/A
G/G


dihomo.linolenoylcarnitine..20.3n3.or.6..
3
43
430
A/A
G/A
G/G


X3.hydroxybutyrylcarnitine..2.
3
42
430
A/A
G/A
G/G


X3.hydroxybutyrylcarnitine..1.
3
40
375
A/A
G/A
G/G


palmitoylcarnitine..C16.
3
42
434
A/A
G/A
G/G

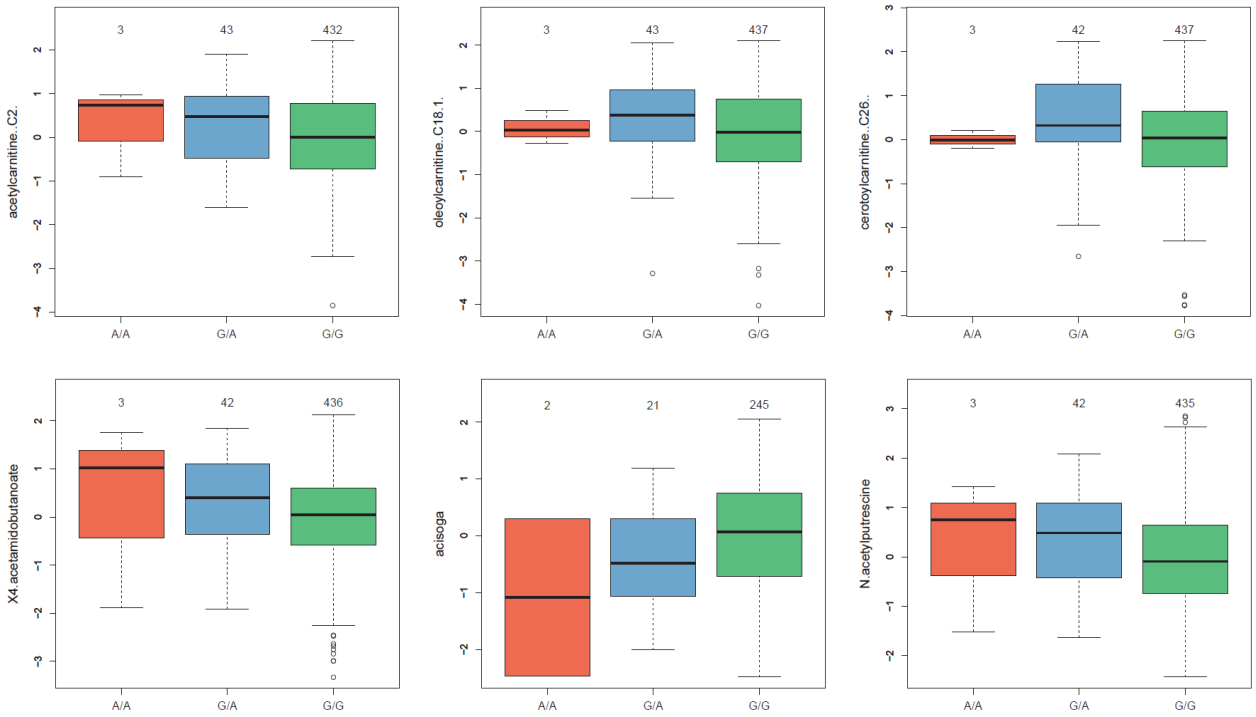


Figure 50. Boxplots representing levels of metabolites affiliated with enriched metabolic pathways in association with the *ATP2B2* (rs56330321) genotype groups.

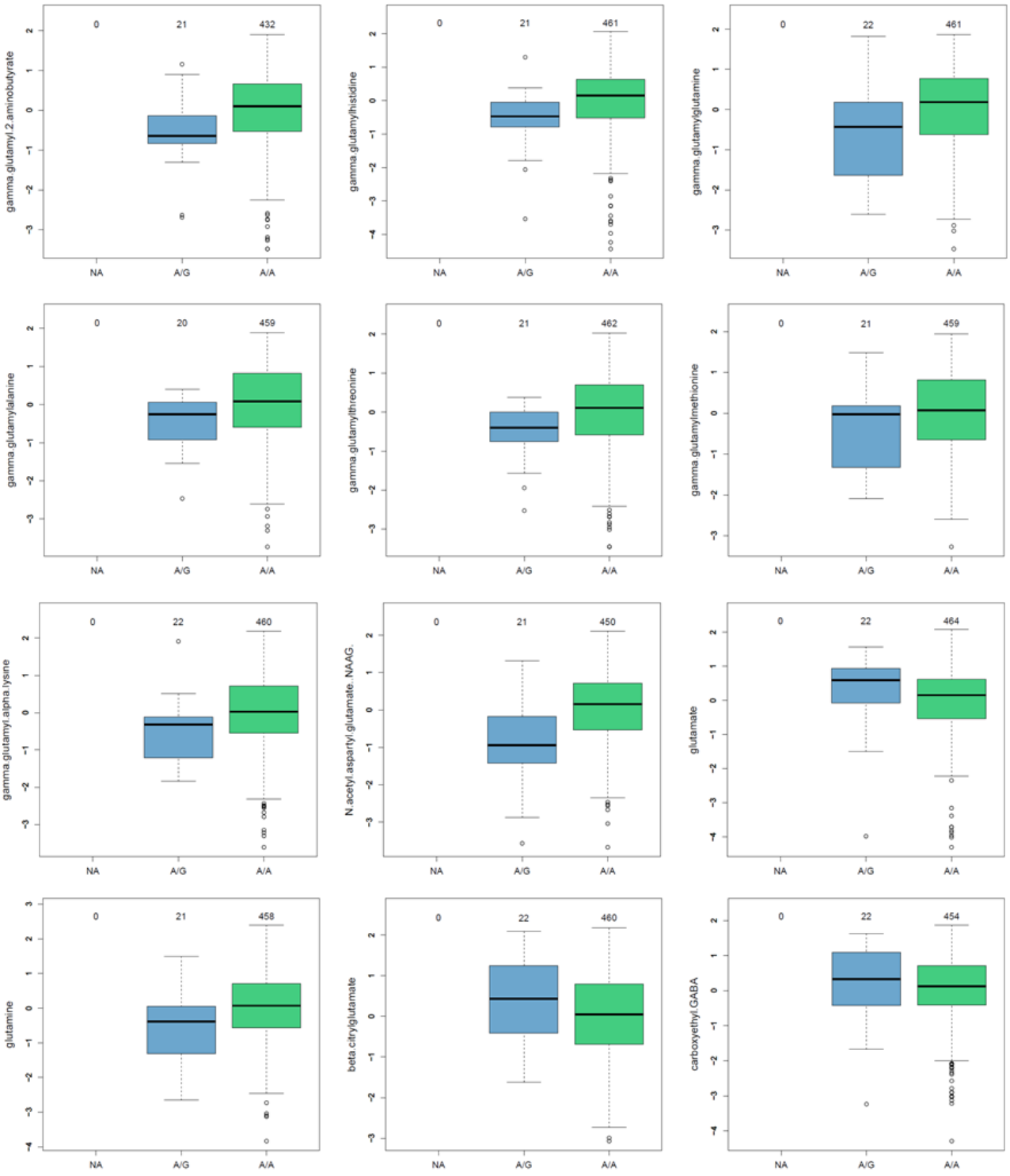


Figure 51. Boxplots representing levels of metabolites affiliated with enriched metabolic pathways in association with the *SYNE*1 (rs2635438) genotype groups.

5.4 Discussion

5.4.1 Common variant loci influence metabolites (mQTLs) in elite athletes

Our genotyping data revealed several SNPs associated with endurance, power and CVD, but none has reached the GWAS level of significance. Except for the association of *ATP2B2* SNP with endurance, the small sample size and effect size of the genetic variants as well as the complexity of the phenotype may have reduced the power of the study to confirm previously suggested associations or identify novel ones. Therefore, a more precise phenotype (metabolites) was sought to obtain larger effect size and a better chance for detection. Additionally, validation in a replication cohort was also pursued.

Combining genotyping and metabolomics analyses, genetically-influenced metabolites were firstly assessed between elite athletes cohort and published data from non-elite athletes [145, 165, 166], and secondly within the elite athletes cohort between moderate and high endurance, power, and CVD groups. Both analyzes revealed novel mGWAS associations with significant effect size (between 8%-14%, Table 29 and Figures 44 and 45), clear genotype-dependent effect (Figures 46 and 48) and evidence of multiple SNP associations within the same genomic region (Figures 47 and 50).

The mGWAS results between elite and non-elite athletes revealed 4 novel mGWAS associations. The first of which is a negative association between rs55729124 in Folate Hydrolase 1 (*FOLH1*) and NAAG levels. *FOLH1* hydrolyzes NAAG to NAA and glutamate [497]. The intronic SNP rs55729124 may therefore be associated with enhanced *FOLH1* activity leading to the breakdown of NAAG and accumulation of NAA and glutamate. NAA is a nervous system specific metabolite found predominantly in cell bodies of neurons. Aerobic fitness was reported to enhance NAA levels, leading to improved cognitive

enhancement [498]. The identification of this novel mGWAS in elite athletes may suggest augmentation of *FOLH1* activity in elite athletes with exercise, resulting in higher NAA levels compared to non-elite athletes in other published studies (Table 29). The second mGWAS identified in our athletic cohort is a positive association between rs3798793 in vascular non-inflammatory molecule 1 (*VNN1*) in association with linoleoyl ethanolamide. *VNN1* protein possess pantetheinase activity that may play a role in oxidative-stress response. The endocannabinoid linoleoyl ethanolamide has a role as fatty acid amide hydrolase inhibitor as it inhibits arachidonoylethanolamide amidohydrolase. The link between *VNN1* and linoleoyl ethanolamide is not clear yet. Exercise, however, has been shown to increase serum concentrations of endocannabinoids [499], thus the identification of this novel mGWAS may be due to exercise interaction (Table 29). The third mGWAS involved association of SNP (rs45446698) in the Cytochrome P450 Family 3 Subfamily A Member 7 (*CYP3A7*) with lower serum sulfated steroids [500]. This is the first report of a negative association between the novel SNP rs11568825 in *CYP3A7* with 5alpha- androstan-3alpha, 17beta-diol monosulfate (1), although association of other SNPs within the same gene with same metabolites were previously reported [166]. The linkage disequilibrium between the novel SNP rs11568825 and the previously published SNP rs45446698 is r2 >0.4 that are part of a polymorphism in the promoter of *CYP3A7* [501]. Cytochrome P450 enzymes are important for the metabolism of many endogenous compounds including various steroids [502]. We have indicated previously that 5alpha-androstan-3alpha, 17beta-diol monosulfate is increased in endurance sports, potentially providing evidence of environmental interaction with endurance exercise [167] (Table 29). The fourth novel mGWAS was a positive association between rs17101394 in

Sphingosine-1-Phosphate Phosphatase 1 (*SGPP1*) in association with Ceramide (d16:1/24:1, d18:1/22:1). Although the association of the same SNP with multiple different metabolites was previously reported. These metabolites included various spingolipids such as palmitoyl dihydrosphingomyelin (d18:0/16:0), sphingomyelin (d18:1/14:0, d16:1/16:0), sphingomyelin (d18:1/15:0, d16:1/17:0), sphingomyelin (d18:1/20:0, d16:1/22:0), and sphingomyelin (d18:1/21:0, d17:1/22:0, d16:1/23:0) [166, 503] as well as X-08402, and X-10510 [145] that are also related to sphingolipid pathway [160]. *SGPP1* catalyzes the degradation of Sphingosine-1-phosphate (S1P), a bioactive sphingolipid metabolite that regulates diverse biologic processes, via salvage and recycling of sphingosine into long-chain ceramides [504]. Acute prolonged exercise was shown previously to influence ceramide metabolism in human skeletal muscle [505], perhaps explaining identification of this mGWAS in our elite athlete cohort (Table 32). In addition to novel mQTLs identified in elite athletes, we have confirmed 16 previously published mQTLS, two of which exhibited greater effect sizes in our elite atheltes compared to the ones reported in non-elite athletes [165], including rs7072216 in *PYROXD2* in association with N-methylpipecolate (effect size 43% vs 31%) and rs1881245 in *NAT8* in association with N-acetyl-1-methylhistidine (effect size 30% vs 26.6%). The functional relevance of these elevated effect sizes remains to be investigated.

Table 32. Novel elite athletes-associated mQTLSs reflecting gene/environment (exercise) interaction

SNP	Gene	Metabolite	Functional relationship between gene and metabolite	Interaction with athletic performance (exercise)
rs55729124	*FOLH1*	N-Acetylaspartylglutamic acid (NAAG)	Gene encodes an enzyme that directly cleaves NAAG into NAA + Glutamate	Aerobic fitness was reported to enhance NAA levels, leading to increased cognitive enhancement [498]
rs3798793	*VNN1*	Linoleoyl ethanolamide	Gene encodes a membrane protein that participates in hematopoietic cell trafficking	Exercise increases serum concentrations of endocannabinoids including linoleoyl ethanolamide [499]
rs17101394	*SGPP1*	Ceramide	Gene encodes enzyme that directly mediates recycling of sphinogsine into cermides	Acute prolonged exercise was shown to influence ceramide metabolism in human skeletal muscle [505]

5.4.2 mQTLS associated with endurance, power and CVD athletes

When focusing on identified endurance, power and CVD associated metabolites, four mQTLs were found significant for endurance and only one mQTL was significant for both power and CVD. Association between rs174574 SNP in *FADS2* gene and 1-palmitoyl-2-linoleoyl-GPE (16:0/18:2) with power and CVD level were not novel as it was previously reported in non-athletes [166]. A positive association between rs10426201 in Sulfotransferase Family 2A Member 1 (*SULT2A1*) in association with androstenediol (3alpha, 17alpha) monosulfate (2) was novel for endurance athletes. *SULT2A1* catalyzes the sulfation of steroids, a process that is fundamental for their function. Following biosynthesis, hydrophobic steroids become sulfated to accelerate their circulatory shuttling to target tissues. The expression of anion transporting polypeptides on target cells enables their uptake. Subsequently, intracellular sulfatases activate them by hydrolyzing the steroid sulfate esters [506]. The genetic predisposition of steroid sulfation

in elite high endurance athletes may therefore explain active steroid biosynthesis in this group and could potentially contribute to their elite physical performance.

5.4.3 Metabolites associated with endurance SNPs

GWAS results showed only two significant SNPs associated with endurance. Metabolomics data were used for functional validation of this association. The rs56330321 A allele found in mGWAS results was associated with higher levels of several ceramides, fatty acidsacyl carnitine, polyamines, except for acisoga (N-(3-acetamidopropyl)pyrrolidin-2-one), and creatine/creatinine (Figure 51). Ceramides tend to accumulate in skeletal muscles and promote insulin resistance. Chronic endurance exercise lowers muscle ceramides and promotes the insulin-sensitivity in exercising muscle [507]. Since the A allele is associated with higher ceramides levels, it could be potentially compromising the beneficial effect of exercise in carriers on improving insulin sensitivity [508]. The A allele is also associated with higher levels of fatty acids acyl carnitines, a hallmark of active fatty acid oxidation. During endurance exercise, fatty acids oxidation increases, sparing glycogen and delaying muscle fatigue [509]. Despite the beneficial effect of fatty acid oxidation in endurance athletes, the elevated fatty acid acyl carnitines in A allele carriers may represent a compensatory mechanism to counteract ceramides-induced impairment of fatty acid oxidation [510]. The A allele was also associated with higher polyamine accumulation, except for acisoga. The increase in polyamines concentration in exercising skeletal muscle after physical exercise reflects oxidative processes related to muscle adaptation to exercise [511]. The elevated polyamines in A allele carriers may therefore reflect higher oxidative mechanisms, also suggested by the increased acyl carnitines, in response to endurance exercise. The

elevated creatine/creatinine levels in A allele carriers may suggest worse renal functions compared to GG individuals, perhaps contributing to their lower prevalence in high endurance athletes [512]. The direct link between *ATP2B2* (rs56330321) and levels of these metabolites is yet to be determined.

The second GWAS significant SNP (rs2635438) G allele is associated with lower gamma-glutamyl amino acids and glutamine but higher glutamate metabolism. Glutamine has various ergogenic benefits including increased muscle strength and promotes better recovery [513]. The lower levels of glutamine in G allele carriers may partially explain their lower prevalence in endurance athletes. The direct link between *SYNE1* (rs2635438) and levels of these and other metabolites remains to be determined.

The data described in this chapter reveal for the first-time evidence of genetically-influenced metabolites associated with elite athletic status in general and endurance sports in particular. Uncovering these novel associations in elite athletes, but not in the general population, could reflect a gene-environment (intensive exercise) interaction that augments the effect size of these genetic variants. Among the novel identified mGWAS, SNPs associated with enhanced endogenous steroids activity may play an important role in elite athletic performance, especially among endurance athletes. The utilization of these mGWAS as biomarkers for selecting athletic candidates with a greater potential to becoming elite endurance athletes is warranted and should be further validated.

CHAPTER 6
DISCUSSION

6 (Chapter 6) Discussion:

This thesis describes an investigation of the superior physical performance of elite athletes as a complex phenotype that results from interactions between genetic and environmental factors (exercise and diet). At the start of this work, samples were collected from the anti-doping laboratory in Qatar and Italy after they tested negative for doping abuse, following strict lab procedures of using anonymous clean athletes' samples in research. Sport's type and gender were the only information available with the samples because of the strict anonymization process followed by anti-doping laboratories according to the world anti-doping laboratories agency (WADA). Without any access to actual measurement of VO_2max, MVC and cardiovascular demand, athletes were classified into high vs moderate endurance, power and CVD groups according to their known sport type depending on previously published sports classification criteria [16].

The first aim of this study was to examine the association of the identified polymorphisms with functional relevance to classified study groups. The second aim of this study was to investigate the metabolomic fluctuations and mechanisms through which those changes exhibit associations with the study groups.

6.1.1 Genome Wide Association study

There have been several investigations over genetic contribution to the susceptibility to sport performance. Most of the findings in the field of sports genetics came from studying case-control and monogenic forms of results. The identification and understanding of the genetic contributors alone to different sport disciplines has proven to be more challenging. The discovery GWAS analysis in our cohort followed by a replication study in Russian elite athletes has revealed two significant SNPs associated with endurance.

GWAS identified two significant SNPs in *ATP2B2* gene and *SYN2A* gene. Whereas the first gene is known to be involved in calcium intracellular signaling, the second gene codes for nucleo-cytoskeletal structural protein. Neither of these two genes was reported before in association with physical activity, however their critical functions in cell signaling and muscle contraction may explain these associations [396, 397].

Despite identification of two novel SNPs in association with endurance, their effect size remains relatively small because of small sample size and the complexity of the phenotype. Therefore, an intermediate phenotype with direct functional relevance to genetic variants was sought. To achieve this goal, metabolomics profiling was performed and compared among different sport disciplines. Metabolites were then used as intermediate phenotypes for identifying genetically influenced metabotypes associated with elite athletes.

6.1.2 Metabolomics

Metabolomics offer a comprehensive approach for detecting metabolic changes in response to environmental factors, dietary, and lifestyle. Indeed, profiling metabolic changes in elite athletes from various sport disciplines has given a deeper insight into the biological state and physical adaptation in response to their respective sports. Metabolomics analysis of discovery and replication cohorts of elite athletes revealed changes in various metabolites involved in steroid biosynthesis, fatty acid oxidation, oxidative stress response, xenobiotics and various mediators of cell signaling among different groups of endurance, power and CVD athletes. These changes reflect various cellular adaptations to prolonged exercise-induced oxidative stress and increased energy demand. Using metabolomics data as potential biomarkers requires further validation.

Once validated it could help in managing athletes' future training programs and in avoiding potential disorders related to excessive training as well as improving their performance in general.

Since supplement use by athletes is of interest, and because most information related to supplement use are obtained from surveys, we have decided to use our data to provide the evidence from metabolomics analysis of supplement use in elite athletes. This was achieved by focusing on profiling xenobiotics that likely came from supplements present in different groups of athletes. Our data indicated presence of various types of supplements that can benefit the athlete, but others may harm the athlete's health, and reputation; in case of an anti-doping rule violation [60].

6.1.3 Metabolomics GWAS study

In order to link GWAS and metabolomics results, mGWAS analysis was conducted. The analysis revealed novel mQTLs found in elite athletes when compared to non-elite athlete [145, 165, 166]. The analysis also revealed novel mQTL in association with endurance. The functional relevance of these associations remains suggestive despite the direct functional relationship between gene function and metabolite levels, especially the ones that were shown to respond to exercise. Final functional validation of these SNPs is important for their potential use as biomarkers for selecting athletic candidates and enhance performance and recovery from sports injuries.

CHAPTER 7
CONCLUSION

7 (Chapter 7) Conclusion:

This thesis reports the identification of genetic predisposition into elite athletic performance that has long been considered as a controversial matter due to the small effect size of genes and the vague phenotype of elite performance. By increasing the number of participants and focusing on intermediate phenotypes with direct functional relationships with genetic variants, the study results provide the first evidence of genetically-influenced metabolites associated with elite athletic performance. The identified novel genetic variants are associated with metabolites that can be influenced by exercise and potentially enhance elite athletic performance by increasing the cognitive function, energy utilization and endogenous steroids activity of elite athletes. Uncovering these novel associations in elite athletes, but not in the general population, could reflect a gene-environment (intensive exercise) interaction that augments the effect size of these genetic variants. The identified genetic variants and associated metabolites, once validated in independent cohorts, could potentially be utilized in biomarker discovery in elite athletic candidates, non-conventional anti-doping analytical approaches and therapeutic strategies. Future research should therefore investigate how an athletes genotype and metabolic profile may contribute to training outcomes and adaptations. A major application to prospective athletes or sport clubs is providing personalized training programs to the individual athletes. Introducing genetic and metabolic testing to these programs will allow athlete specific training regimes to improve athletic, skill, and match performance.

Future Plan:

Following the completion of this project, a number of studies are planned to confirm and further validate our findings. These include the following studies:

1. Validation of GWAS and mGWAS data in larger cohorts of athletes from different sport groups and a better stratification based on defined age/weight/cardiac/respiratory function or some type of functional analyses of the identified SNPs and mQTLs?
2. Validation of metabolites and genetically influenced metabotypes as potential biomarkers for athletic talent identification and development using athletic candidates recruited at Aspire Academy, Doha, Qatar. The overall aims of the study are to determine the natural variation of known metabolites in young athletic candidates from different sports disciplines and to validate/identify metabolic biomarkers for talent development. If successful, this can be developed into an anti-doping testing protocol following all required validation steps. This will include the following steps:
    - Assessment of the natural range of metabolites, their specificity and stability over period of times (relevance to doping research initiated recently by WADA).
    - Validate our panel of metabolic biomarkers for talent identification and selection. This panel could help in the selection process of athletic talents in Qatar and potentially worldwide.
    - Establish a panel of metabolic biomarkers for talent development. This panel will reflect best adaptation to training allowing optimal training programs for young promising athletic candidates and measures of their physical status. Furthermore, this could open new avenues for exercise prescription in the general population and target the approach for metabolic diseases

- Verification of the predictive power of the established panel of selected metabolites of elite athletes in relation to young candidates' performance. Such panel could be further developed based on the new findings into a screening panel for elite power or endurance athletes from young age. It would also help directing future training programs, preventing potential disorders associated with excessive exercise as well as improving their overall performance
- Profiling xenobiotics including supplements and their products with reference to information collected from candidates and their coaches.
- Detection and determination of extreme metabolic outliers in relation to their genetic profile following repeated metabolomics analyses. Such information could be used by anti-doping labs to avoid false positives.

References:

1. in *Genes, Behavior, and the Social Environment: Moving Beyond the Nature/Nurture Debate*, L.M. Hernandez and D.G. Blazer, Editors. 2006: Washington (DC).
2. Darwin, C., *On the origin of spicies.* John Murray, UK., 1859.
3. Singh, R.S., *Darwin's legacy II: why biology is not physics, or why it has taken a century to see the dependence of genes on the environment.* Genome, 2015. **58**(1): p. 55-62.
4. Hunter, P., *Extended phenotype redux. How far can the reach of genes extend in manipulating the environment of an organism?* EMBO Rep, 2009. **10**(3): p. 212-5.
5. Maughan, R.J.G., Michael, *The Biochemical Basis of Sports Performance* OUP Oxford.
6. Armstrong, N. and A.M. McManus, *Physiology of elite young male athletes.* Med Sport Sci, 2011. **56**: p. 1-22.
7. Georgiades, E., et al., *Why nature prevails over nurture in the making of the elite athlete.* BMC Genomics, 2017. **18**(Suppl 8): p. 835.
8. KA, E., *Deliberate practice and the modifiability of body and mind: Toward a science of the structure and acquisition of expert and elite performance.* International Journal of Sport Psychology, 2007. **38**(1): p. 4-34.
9. Swann, C., A. Moran, and D. Piggott, *Defining elite athletes: Issues in the study of expert performance in sport psychology.* Psychology of Sport & Exercise, 2015. **16**(1): p. 3-14.
10. Welch, A.S. and M. Tschampl, *Something to Shout About: A Simple, Quick Performance Enhancement Technique Improved Strength in Both Experts and Novices.* Journal of Applied Sport Psychology, 2012. **24**(4): p. 418-428.
11. Gladwell, M., *Outliers : the story of success / Malcolm Gladwell*. 2009, London: London : Penguin.
12. Macarthur, D.G. and K.N. North, *Genes and human elite athletic performance.* Hum Genet, 2005. **116**(5): p. 331-9.
13. Grant, M. and P. Schempp, *Analysis and Description of Olympic Gold Medalists' Competition-Day Routines.* The Sport Psychologist, 2013. **27**(2): p. 156.
14. Voss, M.W., et al., *Are expert athletes 'expert' in the cognitive laboratory? A meta-analytic review of cognition and sport expertise.* Applied Cognitive Psychology, 2010. **24**(6): p. 812-826.
15. Asmussen, E., *Similarities and dissimilarities between static and dynamic exercise.* Circ Res, 1981. **48**(6 Pt 2): p. I3-10.
16. Mitchell, J.H., et al., *Task Force 8: classification of sports.* J Am Coll Cardiol, 2005. **45**(8): p. 1364-7.
17. Laursen, P.B., *Training for intense exercise performance: high-intensity or high-volume training?* Scand J Med Sci Sports, 2010. **20 Suppl 2**: p. 1-10.
18. Thein, L.A., *Environmental conditions affecting the athlete.* J Orthop Sports Phys Ther, 1995. **21**(3): p. 158-71.
19. Al-Khelaifi, F., et al., *Chapter Twenty-Three - Proteomics and metabolomics research in exercise and sport*, in *Sports, Exercise, and Nutritional Genomics*, D. Barh and I.I. Ahmetov, Editors. 2019, Academic Press. p. 539-566.
20. Joyner, M.J. and D.P. Casey, *Regulation of increased blood flow (hyperemia) to muscles during exercise: a hierarchy of competing physiological needs.* Physiol Rev, 2015. **95**(2): p. 549-601.
21. Hamer, M., G. O'Donovan, and E. Stamatakis, *Association between physical activity and sub-types of cardiovascular disease death causes in a general population cohort.* Eur J Epidemiol, 2019. **34**(5): p. 483-487.

22. Hamer, M., et al., *Examining associations between physical activity and cardiovascular mortality using negative control outcomes.* Int J Epidemiol, 2018.
23. Hamer, M., *Psychosocial stress and cardiovascular disease risk: the role of physical activity.* Psychosom Med, 2012. **74**(9): p. 896-903.
24. Mole, P.A., L.B. Oscai, and J.O. Holloszy, *Adaptation of muscle to exercise. Increase in levels of palmityl Coa synthetase, carnitine palmityltransferase, and palmityl Coa dehydrogenase, and in the capacity to oxidize fatty acids.* J Clin Invest, 1971. **50**(11): p. 2323-30.
25. Spencer, M.R. and P.B. Gastin, *Energy system contribution during 200- to 1500-m running in highly trained athletes.* Med Sci Sports Exerc, 2001. **33**(1): p. 157-62.
26. Guilherme, J.P.L.F., et al., *Genetics and sport performance: current challenges and directions to the future.* Revista Brasileira de Educação Física e Esporte, 2014. **28**(1): p. 177-193.
27. Lundby, C. and P. Robach, *Does 'altitude training' increase exercise performance in elite athletes?* Exp Physiol, 2016. **101**(7): p. 783-8.
28. Saltin, B., et al., *Aerobic exercise capacity at sea level and at altitude in Kenyan boys, junior and senior runners compared with Scandinavian runners.* Scand J Med Sci Sports, 1995. **5**(4): p. 209-21.
29. Wilber, R.L. and Y.P. Pitsiladis, *Kenyan and Ethiopian distance runners: what makes them so good?* Int J Sports Physiol Perform, 2012. **7**(2): p. 92-102.
30. Taylor, D.L., et al., *Interactions between genetic variation and cellular environment in skeletal muscle gene expression.* PLoS One, 2018. **13**(4): p. e0195788.
31. Bassel-Duby, R. and E.N. Olson, *Signaling pathways in skeletal muscle remodeling.* Annu Rev Biochem, 2006. **75**: p. 19-37.
32. Ferraro, E., et al., *Exercise-induced skeletal muscle remodeling and metabolic adaptation: redox signaling and role of autophagy.* Antioxid Redox Signal, 2014. **21**(1): p. 154-76.
33. Wang, G., et al., *Tracking blood glucose and predicting prediabetes in Chinese children and adolescents: a prospective twin study.* PLoS One, 2011. **6**(12): p. e28573.
34. Shephard, R.J., *Athletic performance and urban air pollution.* Can Med Assoc J, 1984. **131**(2): p. 105-9.
35. El Helou, N., et al., *Impact of environmental parameters on marathon running performance.* PLoS One, 2012. **7**(5): p. e37407.
36. Garland, T., Jr., et al., *The biological control of voluntary exercise, spontaneous physical activity and daily energy expenditure in relation to obesity: human and rodent perspectives.* J Exp Biol, 2011. **214**(Pt 2): p. 206-29.
37. Eime, R.M., et al., *Participation in sport and physical activity: associations with socio-economic status and geographical remoteness.* BMC Public Health, 2015. **15**: p. 434.
38. Eime, R.M., et al., *Socioecological factors potentially associated with participation in physical activity and sport: A longitudinal study of adolescent girls.* J Sci Med Sport, 2015. **18**(6): p. 684-90.
39. Wagstaff, C.R., *Emotion regulation and sport performance.* J Sport Exerc Psychol, 2014. **36**(4): p. 401-12.
40. Devonport, T.J., A.M. Lane, and Y.L. Hanin, *Emotional States of athletes prior to performance-induced injury.* J Sports Sci Med, 2005. **4**(4): p. 382-94.
41. Saw, A.E., L.C. Main, and P.B. Gastin, *Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a systematic review.* Br J Sports Med, 2016. **50**(5): p. 281-91.
42. Rothlin, P., et al., *Psychological skills training and a mindfulness-based intervention to enhance functional athletic performance: design of a randomized controlled trial using ambulatory assessment.* BMC Psychol, 2016. **4**(1): p. 39.

43. Saw, A.E., L.C. Main, and P.B. Gastin, *Monitoring athletes through self-report: factors influencing implementation.* J Sports Sci Med, 2015. **14**(1): p. 137-46.
44. Saucedo Marquez, C.M., et al., *High-intensity interval training evokes larger serum BDNF levels compared with intense continuous exercise.* J Appl Physiol (1985), 2015. **119**(12): p. 1363-73.
45. Kumar, B., R. Robinson, and S. Till, *Physical activity and health in adolescence.* Clin Med (Lond), 2015. **15**(3): p. 267-72.
46. Weiler, R., et al., *Sedentary behaviour among elite professional footballers: health and performance implications.* BMJ Open Sport Exerc Med, 2015. **1**(1).
47. Kumar, B., C. Hornby, and K. Joshi, *Sport and exercise medicine.* Medicine, 2018. **46**(3): p. 196-201.
48. Naves, J.P.A., et al., *Cardiorespiratory and perceptual responses of two interval training and a continuous training protocol in healthy young men.* Eur J Sport Sci, 2019. **19**(5): p. 653-660.
49. Cochran, A.J., et al., *Intermittent and continuous high-intensity exercise training induce similar acute but different chronic muscle adaptations.* Exp Physiol, 2014. **99**(5): p. 782-91.
50. Gibala, M.J., et al., *Short-term sprint interval versus traditional endurance training: similar initial adaptations in human skeletal muscle and exercise performance.* J Physiol, 2006. **575**(Pt 3): p. 901-11.
51. Iellamo, F., et al., *Matched dose interval and continuous exercise training induce similar cardiorespiratory and metabolic adaptations in patients with heart failure.* Int J Cardiol, 2013. **167**(6): p. 2561-5.
52. Gorostiaga, E.M., et al., *Uniqueness of interval and continuous training at the same maintained exercise intensity.* Eur J Appl Physiol Occup Physiol, 1991. **63**(2): p. 101-7.
53. Egan, B., et al., *Exercise intensity-dependent regulation of peroxisome proliferator-activated receptor coactivator-1 mRNA abundance is associated with differential activation of upstream signalling kinases in human skeletal muscle.* J Physiol, 2010. **588**(Pt 10): p. 1779-90.
54. McLean, S., et al., *The Effect of Recovery Duration on Vastus Lateralis Oxygenation, Heart Rate, Perceived Exertion and Time Motion Descriptors during Small Sided Football Games.* PLoS One, 2016. **11**(2): p. e0150201.
55. Alaunyte, I., V. Stojceska, and A. Plunkett, *Iron and the female athlete: a review of dietary treatment methods for improving iron status and exercise performance.* J Int Soc Sports Nutr, 2015. **12**: p. 38.
56. Beck, K.L., et al., *Role of nutrition in performance enhancement and postexercise recovery.* Open Access J Sports Med, 2015. **6**: p. 259-67.
57. Naclerio, F., et al., *Effects of Supplementation with Beef or Whey Protein Versus Carbohydrate in Master Triathletes.* J Am Coll Nutr, 2017. **36**(8): p. 593-601.
58. Maughan, R.J., et al., *The use of dietary supplements by athletes.* J Sports Sci, 2007. **25 Suppl 1**: p. S103-13.
59. Knapik, J.J., et al., *Prevalence of Dietary Supplement Use by Athletes: Systematic Review and Meta-Analysis.* Sports Med, 2016. **46**(1): p. 103-23.
60. Maughan, R.J., *Infographic: Helping athletes make decisions on dietary supplement use.* Br J Sports Med, 2018. **52**(23): p. 1492-1493.
61. Dwyer, J.T., P.M. Coates, and M.J. Smith, *Dietary Supplements: Regulatory Challenges and Research Resources.* Nutrients, 2018. **10**(1).
62. Erdman, K.A., et al., *Dietary supplementation of high-performance Canadian athletes by age and gender.* Clin J Sport Med, 2007. **17**(6): p. 458-64.
63. Sundgot-Borgen, J., B. Berglund, and M.K. Torstveit, *Nutritional supplements in Norwegian elite athletes--impact of international ranking and advisors.* Scand J Med Sci Sports, 2003. **13**(2): p. 138-44.

64. Solheim, S.A., et al., *Use of nutritional supplements by Danish elite athletes and fitness customers.* Scand J Med Sci Sports, 2017. **27**(8): p. 801-808.
65. Braun, H., et al., *Dietary supplement use among elite young German athletes.* Int J Sport Nutr Exerc Metab, 2009. **19**(1): p. 97-109.
66. Silver, M.D., *Use of ergogenic aids by athletes.* J Am Acad Orthop Surg, 2001. **9**(1): p. 61-70.
67. Harrison, R.A., et al., *Are those in need taking dietary supplements? A survey of 21 923 adults.* Br J Nutr, 2004. **91**(4): p. 617-23.
68. Maryam Ekramzadeh, a.S.T., *The Prevalence of Sport Dietary Supplements Consumption among Male Athlete Students in Shiraz University of Medical Sciences, Shiraz, Iran.* International Journal of Nutrition Sciences, 2017. **2.4**: p. 218-23. Web.
69. Sullivan, P. and P. Razavi, *Are Athletes' Doping-Related Attitudes Predicted by Their Perceptions of Coaches' Confrontation Efficacy?* Subst Use Misuse, 2017. **52**(8): p. 1098-1103.
70. Rundell, K.W., *Effect of air pollution on athlete health and performance.* Br J Sports Med, 2012. **46**(6): p. 407-12.
71. Cutrufello, P.T., J.M. Smoliga, and K.W. Rundell, *Small things make a big difference: particulate matter and exercise.* Sports Med, 2012. **42**(12): p. 1041-58.
72. Brocherie, F., O. Girard, and G. Millet, *Emerging Environmental and Weather Challenges in Outdoor Sports.* Climate, 2015. **3**(3): p. 492-521.
73. Girard, O., F. Brocherie, and D.J. Bishop, *Sprint performance under heat stress: A review.* Scand J Med Sci Sports, 2015. **25 Suppl 1**: p. 79-89.
74. Papadopoulou, S., A. Lailoglou, and A. Fachantidou, *The Socioeconomic Profile of the Greek Women's National Volleyball Team.* Women in Sport & Physical Activity Journal, 2008. **17**(2): p. 50-61.
75. Ruiz-Tendero, G. and J.J.S. Martín, *Psycho-Social Factors Determining Success in High-Performance Triathlon: Compared Perception in the Coach-Athlete Pair.* Perceptual and Motor Skills, 2012. **115**(3): p. 865-880.
76. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
77. Genomes Project, C., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
78. Moran, C.N. and Y.P. Pitsiladis, *Tour de France Champions born or made: where do we take the genetics of performance?* J Sports Sci, 2017. **35**(14): p. 1411-1419.
79. Eynon, N., et al., *Genes and elite athletes: a roadmap for future research.* J Physiol, 2011. **589**(Pt 13): p. 3063-70.
80. Ahmetov, II and V.A. Rogozkin, *Genes, athlete status and training -- An overview.* Med Sport Sci, 2009. **54**: p. 43-71.
81. Stubbe, J.H., et al., *Genetic influences on exercise participation in 37,051 twin pairs from seven countries.* PLoS One, 2006. **1**: p. e22.
82. De Moor, M.H., et al., *Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs.* Twin Res Hum Genet, 2007. **10**(6): p. 812-20.
83. Bouchard, C., et al., *Genomic predictors of the maximal O(2) uptake response to standardized exercise training programs.* J Appl Physiol (1985), 2011. **110**(5): p. 1160-70.
84. Timmons, J.A., et al., *Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans.* J Appl Physiol (1985), 2010. **108**(6): p. 1487-96.
85. Peeters, M.W., et al., *Genetics and sports: an overview of the pre-molecular biology era.* Med Sport Sci, 2009. **54**: p. 28-42.
86. Calvo, M., et al., *Heritability of explosive power and anaerobic capacity in humans.* Eur J Appl Physiol, 2002. **86**(3): p. 218-25.

87. Montgomery, H.E., et al., *Human gene for physical performance.* Nature, 1998. **393**(6682): p. 221-2.
88. Doring, F.E., 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): p. 1355-9.
89. Doring, F., 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): p. 841-5.
90. Doring, F., et al., *A common haplotype and the Pro582Ser polymorphism of the hypoxia-inducible factor-1alpha (HIF1A) gene in elite endurance athletes.* J Appl Physiol (1985), 2010. **108**(6): p. 1497-500.
91. Rankinen, T., et al., *Advances in exercise, fitness, and performance genomics.* Med Sci Sports Exerc, 2010. **42**(5): p. 835-46.
92. Mills, M., et al., *Differential expression of the actin-binding proteins, alpha-actinin-2 and -3, in different species: implications for the evolution of functional redundancy.* Hum Mol Genet, 2001. **10**(13): p. 1335-46.
93. Berman, Y. and K.N. North, *A gene for speed: the emerging role of alpha-actinin-3 in muscle metabolism.* Physiology (Bethesda), 2010. **25**(4): p. 250-9.
94. MacArthur, D.G. and K.N. North, *ACTN3: A genetic influence on muscle function and athletic performance.* Exerc Sport Sci Rev, 2007. **35**(1): p. 30-4.
95. MacArthur, D.G., et al., *Loss of ACTN3 gene function alters mouse muscle metabolism and shows evidence of positive selection in humans.* Nat Genet, 2007. **39**(10): p. 1261-5.
96. MacArthur, D.G. and K.N. North, *A gene for speed? The evolution and function of alpha-actinin-3.* Bioessays, 2004. **26**(7): p. 786-95.
97. MacArthur, D.G., et al., *An Actn3 knockout mouse provides mechanistic insights into the association between alpha-actinin-3 deficiency and human athletic performance.* Hum Mol Genet, 2008. **17**(8): p. 1076-86.
98. Moran, C.N., et al., *Association analysis of the ACTN3 R577X polymorphism and complex quantitative body composition and performance phenotypes in adolescent Greeks.* Eur J Hum Genet, 2007. **15**(1): p. 88-93.
99. Eynon, N., et al., *Genes for elite power and sprint performance: ACTN3 leads the way.* Sports Med, 2013. **43**(9): p. 803-17.
100. Tucker, R., J. Santos-Concejero, and M. Collins, *The genetic basis for elite running performance.* Br J Sports Med, 2013. **47**(9): p. 545-9.
101. Jamshidi, Y., et al., *Peroxisome proliferator--activated receptor alpha gene regulates left ventricular growth in response to exercise and hypertension.* Circulation, 2002. **105**(8): p. 950-5.
102. Tucker, R. and M. Collins, *What makes champions? A review of the relative contribution of genes and training to sporting success.* Br J Sports Med, 2012. **46**(8): p. 555-61.
103. van Setten, J., et al., *Genome-wide association meta-analysis of 30,000 samples identifies seven novel loci for quantitative ECG traits.* Eur J Hum Genet, 2019. **27**(6): p. 952-962.
104. van der Harst, P., et al., *52 Genetic Loci Influencing Myocardial Mass.* J Am Coll Cardiol, 2016. **68**(13): p. 1435-1448.
105. Kranias, E.G. and R.J. Hajjar, *Modulation of cardiac contractility by the phospholamban/SERCA2a regulatome.* Circ Res, 2012. **110**(12): p. 1646-60.
106. Roth, S.M., et al., *Advances in exercise, fitness, and performance genomics in 2011.* Med Sci Sports Exerc, 2012. **44**(5): p. 809-17.
107. Ahmetov, II, et al., *Genes and Athletic Performance: An Update.* Med Sport Sci, 2016. **61**: p. 41-54.

108. Ahmetov, I., et al., *Genome-wide association study identifies three novel genetic markers associated with elite endurance performance.* Biol Sport, 2015. **32**(1): p. 3-9.
109. Goodlin, G.T., et al., *Applying personal genetic data to injury risk assessment in athletes.* PLoS One, 2014. **10**(4): p. e0122676.
110. Herbert, A.J., et al., *The interactions of physical activity, exercise and genetics and their associations with bone mineral density: implications for injury risk in elite athletes.* Eur J Appl Physiol, 2019. **119**(1): p. 29-47.
111. Mitchell, J.A., et al., *Physical Activity Benefits the Skeleton of Children Genetically Predisposed to Lower Bone Density in Adulthood.* J Bone Miner Res, 2016. **31**(8): p. 1504-12.
112. Yang, J., et al., *Common SNPs explain a large proportion of the heritability for human height.* Nat Genet, 2010. **42**(7): p. 565-9.
113. Pitsiladis, Y.P., et al., *Athlome Project Consortium: a concerted effort to discover genomic and other "omic" markers of athletic performance.* Physiol Genomics, 2016. **48**(3): p. 183-90.
114. Jacob, Y., et al., *The Potential Role of Genetic Markers in Talent Identification and Athlete Assessment in Elite Sport.* Sports (Basel), 2018. **6**(3).
115. Rankinen, T., et al., *No Evidence of a Common DNA Variant Profile Specific to World Class Endurance Athletes.* PLoS One, 2016. **11**(1): p. e0147330.
116. Timmons, J.A., *Variability in training-induced skeletal muscle adaptation.* J Appl Physiol (1985), 2011. **110**(3): p. 846-53.
117. Contrepois, K., L. Liang, and M. Snyder, *Can Metabolic Profiles Be Used as a Phenotypic Readout of the Genome to Enhance Precision Medicine?* Clin Chem, 2016. **62**(5): p. 676-8.
118. Dharuri, H., et al., *Genetics of the human metabolome, what is next?* Biochim Biophys Acta, 2014. **1842**(10): p. 1923-1931.
119. Suhre, K. and C. Gieger, *Genetic variation in metabolic phenotypes: study designs and applications.* Nat Rev Genet, 2012. **13**(11): p. 759-69.
120. Serkova, N.J., T.J. Standiford, and K.A. Stringer, *The emerging field of quantitative blood metabolomics for biomarker discovery in critical illnesses.* Am J Respir Crit Care Med, 2011. **184**(6): p. 647-55.
121. Stringer, K.A., et al., *Metabolic consequences of sepsis-induced acute lung injury revealed by plasma (1)H-nuclear magnetic resonance quantitative metabolomics and computational analysis.* Am J Physiol Lung Cell Mol Physiol, 2011. **300**(1): p. L4-L11.
122. Wang, Z.V., D.L. Li, and J.A. Hill, *Heart failure and loss of metabolic control.* J Cardiovasc Pharmacol, 2014. **63**(4): p. 302-13.
123. Ussher, J.R., et al., *The Emerging Role of Metabolomics in the Diagnosis and Prognosis of Cardiovascular Disease.* J Am Coll Cardiol, 2016. **68**(25): p. 2850-2870.
124. Lac, G. and F. Maso, *Biological markers for the follow-up of athletes throughout the training season.* Pathol Biol (Paris), 2004. **52**(1): p. 43-9.
125. Rietjens, G.J., et al., *Physiological, biochemical and psychological markers of strenuous training-induced fatigue.* Int J Sports Med, 2005. **26**(1): p. 16-26.
126. Yan, B., et al., *Metabolomic investigation into variation of endogenous metabolites in professional athletes subject to strength-endurance training.* J Appl Physiol (1985), 2009. **106**(2): p. 531-8.
127. Nicholson, J.K., et al., *Metabonomics: a platform for studying drug toxicity and gene function.* Nat Rev Drug Discov, 2002. **1**(2): p. 153-61.
128. Lewis, G.D., et al., *Metabolic signatures of exercise in human plasma.* Sci Transl Med, 2010. **2**(33): p. 33ra37.
129. Goodwin, M.L., et al., *Blood lactate measurements and analysis during exercise: a guide for clinicians.* J Diabetes Sci Technol, 2007. **1**(4): p. 558-69.

130. Dudzinska, W., et al., *Adenine, guanine and pyridine nucleotides in blood during physical exercise and restitution in healthy subjects.* Eur J Appl Physiol, 2010. **110**(6): p. 1155-62.
131. Howarth, K.R., et al., *Effect of endurance training on muscle TCA cycle metabolism during exercise in humans.* J Appl Physiol (1985), 2004. **97**(2): p. 579-84.
132. Henriksson, J., *Effect of exercise on amino acid concentrations in skeletal muscle and plasma.* J Exp Biol, 1991. **160**: p. 149-65.
133. Sahlin, K., A. Katz, and S. Broberg, *Tricarboxylic acid cycle intermediates in human muscle during prolonged exercise.* Am J Physiol, 1990. **259**(5 Pt 1): p. C834-41.
134. Eriksson, L.S., et al., *Ammonia metabolism during exercise in man.* Clin Physiol, 1985. **5**(4): p. 325-36.
135. Sato, K., et al., *Responses of sex steroid hormones to different intensities of exercise in endurance athletes.* Exp Physiol, 2016. **101**(1): p. 168-75.
136. Sapieha, P., et al., *The succinate receptor GPR91 in neurons has a major role in retinal angiogenesis.* Nat Med, 2008. **14**(10): p. 1067-76.
137. He, W., et al., *Citric acid cycle intermediates as ligands for orphan G-protein-coupled receptors.* Nature, 2004. **429**(6988): p. 188-93.
138. Zivkovic, A.M. and J.B. German, *Metabolomics for assessment of nutritional status.* Curr Opin Clin Nutr Metab Care, 2009. **12**(5): p. 501-7.
139. Nieman, D.C. and S.H. Mitmesser, *Potential Impact of Nutrition on Immune System Recovery from Heavy Exertion: A Metabolomics Perspective.* Nutrients, 2017. **9**(5).
140. German, J.B., M.A. Roberts, and S.M. Watkins, *Personal metabolomics as a next generation nutritional assessment.* J Nutr, 2003. **133**(12): p. 4260-6.
141. Gibney, M.J., et al., *Metabolomics in human nutrition: opportunities and challenges.* Am J Clin Nutr, 2005. **82**(3): p. 497-503.
142. Ehret, G.B., et al., *A multi-SNP locus-association method reveals a substantial fraction of the missing heritability.* Am J Hum Genet, 2012. **91**(5): p. 863-71.
143. Ehret, G.B., *The Contribution of the Framingham Heart Study to Gene Identification for Cardiovascular Risk Factors and Coronary Heart Disease.* Glob Heart, 2013. **8**(1): p. 59-65.
144. Visscher, P.M., et al., *Five years of GWAS discovery.* Am J Hum Genet, 2012. **90**(1): p. 7-24.
145. Shin, S.Y., et al., *An atlas of genetic influences on human blood metabolites.* Nat Genet, 2014. **46**(6): p. 543-550.
146. Suhre, K., et al., *Human metabolic individuality in biomedical and pharmaceutical research.* Nature, 2011. **477**(7362): p. 54-60.
147. Lee, H.-S., et al., *Identification of putative biomarkers for type 2 diabetes using metabolomics in the Korea Association REsource (KARE) cohort.* Metabolomics, 2016. **12**(12): p. 1-12.
148. Beebe, K. and A.D. Kennedy, *Sharpening Precision Medicine by a Thorough Interrogation of Metabolic Individuality.* Comput Struct Biotechnol J, 2016. **14**: p. 97-105.
149. Petersen, A.K., et al., *On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies.* BMC Bioinformatics, 2012. **13**: p. 120.
150. Kastenmuller, G., et al., *Genetics of human metabolism: an update.* Hum Mol Genet, 2015. **24**(R1): p. R93-R101.
151. Dumas, M.E., *Metabolome 2.0: quantitative genetics and network biology of metabolic phenotypes.* Mol Biosyst, 2012. **8**(10): p. 2494-502.
152. Guth, L.M. and S.M. Roth, *Genetic influence on athletic performance.* Curr Opin Pediatr, 2013. **25**(6): p. 653-8.
153. Bouchard, C., et al., *Genetic effect in resting and exercise metabolic rates.* Metabolism, 1989. **38**(4): p. 364-70.

154. Karoly, H.C., et al., *Genetic Influences on Physiological and Subjective Responses to an Aerobic Exercise Session among Sedentary Adults.* J Cancer Epidemiol, 2012. **2012**: p. 540563.
155. Tanaka, M., G. Wang, and Y.P. Pitsiladis, *Advancing sports and exercise genomics: moving from hypothesis-driven single study approaches to large multi-omics collaborative science.* Physiol Genomics, 2016. **48**(3): p. 173-4.
156. Gieger, C., et al., *Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum.* PLoS Genet, 2008. **4**(11): p. e1000282.
157. Illig, T., et al., *A genome-wide perspective of genetic variation in human metabolism.* Nat Genet, 2010. **42**(2): p. 137-41.
158. Nicholson, G., et al., *A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection.* PLoS Genet, 2011. **7**(9): p. e1002270.
159. Draisma, H.H.M., et al., *Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels.* Nat Commun, 2015. **6**: p. 7208.
160. Yet, I., et al., *Genetic Influences on Metabolite Levels: A Comparison across Metabolomic Platforms.* PLoS One, 2016. **11**(4): p. e0153672.
161. Kettunen, J., et al., *Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA.* Nat Commun, 2016. **7**: p. 11122.
162. Yazdani, A., et al., *Identification of Rare Variants in Metabolites of the Carnitine Pathway by Whole Genome Sequencing Analysis.* Genet Epidemiol, 2016. **40**(6): p. 486-91.
163. Guo, L., et al., *Plasma metabolomic profiles enhance precision medicine for volunteers of normal health.* Proc Natl Acad Sci U S A, 2015. **112**(35): p. E4901-10.
164. Demirkan, A., et al., *Insight in genome-wide association of metabolite quantitative traits by exome sequence analyses.* PLoS Genet, 2015. **11**(1): p. e1004835.
165. Yousri, N.A., et al., *Whole-exome sequencing identifies common and rare variant metabolic QTLs in a Middle Eastern population.* Nat Commun, 2018. **9**(1): p. 333.
166. Long, T., et al., *Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites.* Nat Genet, 2017. **49**(4): p. 568-578.
167. Al-Khelaifi, F., et al., *A pilot study comparing the metabolic profiles of elite-level athletes from different sporting disciplines.* Sports Med Open, 2018. **4**(1): p. 2.
168. Tyers, M. and M. Mann, *From genomics to proteomics.* Nature, 2003. **422**(6928): p. 193-7.
169. Misra, B.B., et al., *Integrated Omics: Tools, Advances, and Future Approaches.* J Mol Endocrinol, 2018.
170. Ben-Zaken, S., et al., *ACTN3 Polymorphism: Comparison Between Elite Swimmers and Runners.* Sports Med - Open, 2015. **1**(1): p. 1-8.
171. Se Fum Wong, S., et al., *A simple method for DNA isolation from clotted blood extricated rapidly from serum separator tubes.* Clin Chem, 2007. **53**(3): p. 522-4.
172. Salazar, L.A., et al., *Optimized procedure for DNA isolation from fresh and cryopreserved clotted human blood useful in clinical molecular testing.* Clin Chem, 1998. **44**(8 Pt 1): p. 1748-50.
173. Coleman, J.R., et al., *Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray.* Brief Funct Genomics, 2016. **15**(4): p. 298-304.
174. Handelsman, J.D., L.A. Hirschberg, and L.S. Bermon, *Circulating Testosterone as the Hormonal Basis of Sex Differences in Athletic Performance.* Endocrine Reviews, 2018. **39**(5): p. 803-829.
175. Haizlip, K.M., B.C. Harrison, and L.A. Leinwand, *Sex-based differences in skeletal muscle kinetics and fiber-type composition.* Physiology (Bethesda), 2015. **30**(1): p. 30-9.
176. Melum, E., A. Franke, and T.H. Karlsen, *Genome-wide association studies--a summary for the clinical gastroenterologist.* World J Gastroenterol, 2009. **15**(43): p. 5377-96.
177. Kramer, A., et al., *Causal analysis approaches in Ingenuity Pathway Analysis.* Bioinformatics, 2014. **30**(4): p. 523-30.

178. Yousri, N.A., et al., *Long term conservation of human metabolic phenotypes and link to heritability.* Metabolomics, 2014. **10**(5): p. 1005-1017.
179. Al-Khelaifi, F., et al., *Metabolic profiling of elite athletes with different cardiovascular demand.* Scand J Med Sci Sports, 2019. **29**(7): p. 933-943.
180. Yousri, N.A., et al., *Large Scale Metabolic Profiling identifies Novel Steroids linked to Rheumatoid Arthritis.* Sci Rep, 2017. **7**(1): p. 9137.
181. Krumsiek, J., et al., *Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data.* BMC Syst Biol, 2011. **5**: p. 21.
182. Nishida, K., et al., *KEGGscape: a Cytoscape app for pathway data integration.* F1000Res, 2014. **3**: p. 144.
183. Benton, P.H., et al., *An Interactive Cluster Heat Map to Visualize and Explore Multidimensional Metabolomic Data.* Metabolomics, 2015. **11**(4): p. 1029-1034.
184. Hinrichs, A.L., E.K. Larkin, and B.K. Suarez, *Population stratification and patterns of linkage disequilibrium.* Genet Epidemiol, 2009. **33 Suppl 1**: p. S88-92.
185. Grimm, D.G., et al., *easyGWAS: A Cloud-Based Platform for Comparing the Results of Genome-Wide Association Studies.* Plant Cell, 2017. **29**(1): p. 5-19.
186. Korte, A. and A. Farlow, *The advantages and limitations of trait analysis with GWAS: a review.* Plant Methods, 2013. **9**: p. 29.
187. Hakonarson, H. and S.F. Grant, *Planning a genome-wide association study: points to consider.* Ann Med, 2011. **43**(6): p. 451-60.
188. Wang, G., *Genetic studies of elite athlete status*. 2013, ProQuest Dissertations Publishing.
189. Bouchard, C., et al., *Familial resemblance for VO2max in the sedentary state: the HERITAGE family study.* Med Sci Sports Exerc, 1998. **30**(2): p. 252-8.
190. Ahmetov, II and O.N. Fedotovskaya, *Current Progress in Sports Genomics.* Adv Clin Chem, 2015. **70**: p. 247-314.
191. Myerson, S., et al., *Human angiotensin I-converting enzyme gene and endurance performance.* J Appl Physiol (1985), 1999. **87**(4): p. 1313-6.
192. Gayagay, G., et al., *Elite endurance athletes and the ACE I allele--the role of genes in athletic performance.* Hum Genet, 1998. **103**(1): p. 48-50.
193. Amir, O., et al., *The ACE deletion allele is associated with Israeli elite endurance athletes.* Exp Physiol, 2007. **92**(5): p. 881-6.
194. Muniesa, C.A., et al., *World-class performance in lightweight rowing: is it genetically influenced? A comparison with cyclists, runners and non-athletes.* Br J Sports Med, 2010. **44**(12): p. 898-901.
195. Gineviciene, V., et al., *Genetic variation of the human ACE and ACTN3 genes and their association with functional muscle properties in Lithuanian elite athletes.* Medicina (Kaunas), 2011. **47**(5): p. 284-90.
196. Orysiak, J., et al., *The association between ace gene variation and aerobic capacity in winter endurance disciplines.* Biol Sport, 2013. **30**(4): p. 249-53.
197. Tobina, T., et al., *Association between the angiotensin I-converting enzyme gene insertion/deletion polymorphism and endurance running speed in Japanese runners.* J Physiol Sci, 2010. **60**(5): p. 325-30.
198. Ash, G.I., et al., *No association between ACE gene variation and endurance athlete status in Ethiopians.* Med Sci Sports Exerc, 2011. **43**(4): p. 590-7.
199. Shenoy, S., et al., *Association of Angiotensin Converting Enzyme gene Polymorphism and Indian Army Triathletes Performance.* Asian J Sports Med, 2010. **1**(3): p. 143-50.
200. Min, S.K., et al., *Is there a gender difference between ACE gene and race distance?* Appl Physiol Nutr Metab, 2009. **34**(5): p. 926-32.

201. Cieszczyk, P., et al., *The angiotensin converting enzyme gene I/D polymorphism in Polish rowers.* Int J Sports Med, 2009. **30**(8): p. 624-7.
202. Tsianos, G., et al., *The ACE gene insertion/deletion polymorphism and elite endurance swimming.* Eur J Appl Physiol, 2004. **92**(3): p. 360-2.
203. Turgut, G., et al., *The angiotensin converting enzyme I/D polymorphism in Turkish athletes and sedentary controls.* Acta Medica (Hradec Kralove), 2004. **47**(2): p. 133-6.
204. Scanavini, D., et al., *Increased frequency of the homozygous II ACE genotype in Italian Olympic endurance athletes.* Eur J Hum Genet, 2002. **10**(10): p. 576-7.
205. Nazarov, I.B., et al., *The angiotensin converting enzyme I/D polymorphism in Russian athletes.* Eur J Hum Genet, 2001. **9**(10): p. 797-801.
206. Hruskovicova, H., et al., *The angiotensin converting enzyme I/D polymorphism in long distance runners.* J Sports Med Phys Fitness, 2006. **46**(3): p. 509-13.
207. Lucia, A., et al., *Is there an association between ACE and CKMM polymorphisms and cycling performance status during 3-week races?* Int J Sports Med, 2005. **26**(6): p. 442-7.
208. Collins, M., et al., *The ACE gene and endurance performance during the South African Ironman Triathlons.* Med Sci Sports Exerc, 2004. **36**(8): p. 1314-20.
209. Alvarez, R., et al., *Genetic variation in the renin-angiotensin system and athletic performance.* Eur J Appl Physiol, 2000. **82**(1-2): p. 117-20.
210. Santiago, C., et al., *Trp64Arg polymorphism in ADRB3 gene is associated with elite endurance performance.* Br J Sports Med, 2011. **45**(2): p. 147-9.
211. Kim, S.M., et al., *Distribution of the Trp64Arg polymorphism in the ss 3-adrenergic receptor gene in athletes and its influence on cardiovascular function.* Kardiol Pol, 2010. **68**(8): p. 920-6.
212. Umekawa, T., et al., *Trp64Arg mutation of beta3-adrenoceptor gene deteriorates lipolysis induced by beta3-adrenoceptor agonist in human omental adipocytes.* Diabetes, 1999. **48**(1): p. 117-20.
213. Walston, J., et al., *Time of onset of non-insulin-dependent diabetes mellitus and genetic variation in the beta 3-adrenergic-receptor gene.* N Engl J Med, 1995. **333**(6): p. 343-7.
214. Sawczuk, M., et al., *Association of the ADRB2 Gly16Arg and Glu27Gln polymorphisms with athlete status.* J Sports Sci, 2013. **31**(14): p. 1535-44.
215. Wagoner, L.E., et al., *Polymorphisms of the beta(2)-adrenergic receptor determine exercise capacity in patients with heart failure.* Circ Res, 2000. **86**(8): p. 834-40.
216. Tsianos, G.I., et al., *Associations of polymorphisms of eight muscle- or metabolism-related genes with performance in Mount Olympus marathon runners.* J Appl Physiol (1985), 2010. **108**(3): p. 567-74.
217. Wolfarth, B., et al., *Association between a beta2-adrenergic receptor polymorphism and elite endurance performance.* Metabolism, 2007. **56**(12): p. 1649-51.
218. Snyder, E.M., et al., *Genotype related differences in beta2 adrenergic receptor density and cardiac function.* Med Sci Sports Exerc, 2006. **38**(5): p. 882-6.
219. Kochanska-Dziurowicz, A.A., et al., *CATECHOLAMINES AND beta2-ADRENOCEPTOR GENE EXPRESSION BEFORE AND AFTER MAXIMAL INCREMENTAL CYCLE TEST IN YOUNG ICE HOCKEY PLAYERS: RELATION TO WORK PERFORMED.* Biol Sport, 2013. **30**(2): p. 85-90.
220. Wolfarth, B., et al., *A polymorphism in the alpha2a-adrenoceptor gene and endurance athlete status.* Med Sci Sports Exerc, 2000. **32**(10): p. 1709-12.
221. Mustafina, L.J., et al., *AGTR2 gene polymorphism is associated with muscle fibre composition, athletic status and aerobic performance.* Exp Physiol, 2014. **99**(8): p. 1042-52.
222. Saunders, C.J., et al., *A variant within the AQP1 3'-untranslated region is associated with running performance, but not weight changes, during an Ironman Triathlon.* J Sports Sci, 2015. **33**(13): p. 1342-8.

223. Rivera, M.A., et al., *AQP-1 association with body fluid loss in 10-km runners.* Int J Sports Med, 2011. **32**(3): p. 229-33.
224. Gineviciene, V., et al., *AMPD1 rs17602729 is associated with physical performance of sprint and power in elite Lithuanian athletes.* BMC Genet, 2014. **15**: p. 58.
225. Cieszczyk, P., et al., *Is the C34T polymorphism of the AMPD1 gene associated with athlete performance in rowing?* Int J Sports Med, 2011. **32**(12): p. 987-91.
226. Rubio, J.C., et al., *Frequency of the C34T mutation of the AMPD1 gene in world-class endurance athletes: does this mutation impair performance?* J Appl Physiol (1985), 2005. **98**(6): p. 2108-12.
227. Thomaes, T., et al., *A genetic predisposition score for muscular endophenotypes predicts the increase in aerobic power after training: the CAREGENE study.* BMC Genet, 2011. **12**: p. 84.
228. Rico-Sanz, J., et al., *Associations between cardiorespiratory responses to exercise and the C34T AMPD1 gene polymorphism in the HERITAGE Family Study.* Physiol Genomics, 2003. **14**(2): p. 161-6.
229. Sawczuk, M., et al., *The -9/+9 polymorphism of the bradykinin receptor Beta 2 gene and athlete status: a study involving two European cohorts.* Hum Biol, 2013. **85**(5): p. 741-56.
230. Eynon, N., et al., *Is there an interaction between BDKRB2 -9/+9 and GNB3 C825T polymorphisms and elite athletic performance?* Scand J Med Sci Sports, 2011. **21**(6): p. e242-6.
231. Saunders, C.J., et al., *The bradykinin beta 2 receptor (BDKRB2) and endothelial nitric oxide synthase 3 (NOS3) genes and endurance performance during Ironman Triathlons.* Hum Mol Genet, 2006. **15**(6): p. 979-87.
232. Poirier, O., et al., *Polymorphisms of genes of the cardiac calcineurin pathway and cardiac hypertrophy.* Eur J Hum Genet, 2003. **11**(9): p. 659-64.
233. Akhmetov, II, et al., *[The influence of calcineurin gene polymorphism on morphofunctional characteristics of cardiovascular system of athletes].* Ross Fiziol Zh Im I M Sechenova, 2008. **94**(8): p. 915-22.
234. Tang, W., et al., *Identification of a novel 5-base pair deletion in calcineurin B (PPP3R1) promoter region and its association with left ventricular hypertrophy.* Am Heart J, 2005. **150**(4): p. 845-51.
235. He, Z.H., et al., *Are calcineurin genes associated with athletic status? A function, replication study.* Med Sci Sports Exerc, 2011. **43**(8): p. 1433-40.
236. He, Z.H., et al., *Polymorphisms in the calcineurin genes are associated with the training responsiveness of cardiac phenotypes in Chinese young adults.* Eur J Appl Physiol, 2010. **110**(4): p. 761-7.
237. He, Z.H., et al., *Are calcineurin genes associated with endurance phenotype traits?* Eur J Appl Physiol, 2010. **109**(3): p. 359-69.
238. Sakuma, K. and A. Yamaguchi, *The functional role of calcineurin in hypertrophy, regeneration, and disorders of skeletal muscle.* J Biomed Biotechnol, 2010. **2010**: p. 721219.
239. Ahmetov, II, et al., *The combined impact of metabolic gene polymorphisms on elite endurance athlete status and related phenotypes.* Hum Genet, 2009. **126**(6): p. 751-61.
240. Rivera, M.A., et al., *Muscle-specific creatine kinase gene polymorphisms in elite endurance athletes and sedentary controls.* Med Sci Sports Exerc, 1997. **29**(11): p. 1444-7.
241. Zhou, D.Q., et al., *Muscle-specific creatine kinase gene polymorphism and running economy responses to an 18-week 5000-m training programme.* Br J Sports Med, 2006. **40**(12): p. 988-91.
242. Rivera, M.A., et al., *Linkage between a muscle-specific CK gene marker and VO2max in the HERITAGE Family Study.* Med Sci Sports Exerc, 1999. **31**(5): p. 698-701.
243. Rivera, M.A., et al., *Muscle-specific creatine kinase gene polymorphism and VO2max in the HERITAGE Family Study.* Med Sci Sports Exerc, 1997. **29**(10): p. 1311-7.
244. van Deursen, J., et al., *Skeletal muscles of mice deficient in muscle creatine kinase lack burst activity.* Cell, 1993. **74**(4): p. 621-31.

245. O'Connell, K., M. Posthumus, and M. Collins, *COL6A1 gene and Ironman triathlon performance.* Int J Sports Med, 2011. **32**(11): p. 896-901.
246. Bonaldo, P., et al., *Collagen VI deficiency induces early onset myopathy in the mouse: an animal model for Bethlem myopathy.* Hum Mol Genet, 1998. **7**(13): p. 2135-40.
247. Abrahams, S., M. Posthumus, and M. Collins, *A polymorphism in a functional region of the COL5A1 gene: association with ultraendurance-running performance and joint range of motion.* Int J Sports Physiol Perform, 2014. **9**(3): p. 583-90.
248. Brown, J.C., et al., *The COL5A1 gene, ultra-marathon running performance, and range of motion.* Int J Sports Physiol Perform, 2011. **6**(4): p. 485-96.
249. Posthumus, M., M.P. Schwellnus, and M. Collins, *The COL5A1 gene: a novel marker of endurance running performance.* Med Sci Sports Exerc, 2011. **43**(4): p. 584-9.
250. Collins, M., et al., *The COL5A1 genotype is associated with range of motion measurements.* Scand J Med Sci Sports, 2009. **19**(6): p. 803-10.
251. Brown, J.C., et al., *Range of motion measurements diverge with increasing age for COL5A1 genotypes.* Scand J Med Sci Sports, 2011. **21**(6): p. e266-72.
252. Voisin, S., et al., *EPAS1 gene variants are associated with sprint/power athletic performance in two cohorts of European athletes.* BMC Genomics, 2014. **15**: p. 382.
253. Henderson, J., et al., *The EPAS1 gene influences the aerobic-anaerobic contribution in elite endurance athletes.* Hum Genet, 2005. **118**(3-4): p. 416-23.
254. Eynon, N., 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): p. 135-9.
255. Eynon, N., et al., *Interaction between SNPs in the NRF2 gene and elite endurance performance.* Physiol Genomics, 2010. **41**(1): p. 78-81.
256. Eynon, N., et al., *NRF2 intron 3 A/G polymorphism is associated with endurance athletes' status.* J Appl Physiol (1985), 2009. **107**(1): p. 76-9.
257. He, Z., et al., *NRF2 genotype improves endurance capacity in response to training.* Int J Sports Med, 2007. **28**(9): p. 717-21.
258. Sawczuk, M., et al., *Is gnb3 c825t polymorphism associated with elite status of polish athletes?* Biol Sport, 2014. **31**(1): p. 21-5.
259. Ruiz, J.R., et al., *GNB3 C825T Polymorphism and elite athletic status: A replication study with two ethnic groups.* Int J Sports Med, 2011. **32**(2): p. 151-3.
260. Eynon, N., et al., *The guanine nucleotide binding protein beta polypeptide 3 gene C825T polymorphism is associated with elite endurance athletes.* Exp Physiol, 2009. **94**(3): p. 344-9.
261. Faruque, M.U., et al., *Association of GNB3 C825T polymorphism with peak oxygen consumption.* Int J Sports Med, 2009. **30**(5): p. 315-9.
262. Siffert, W., et al., *Association of a human G-protein beta3 subunit variant with hypertension.* Nat Genet, 1998. **18**(1): p. 45-8.
263. Chicharro, J.L., et al., *Mutations in the hereditary haemochromatosis gene HFE in professional endurance athletes.* Br J Sports Med, 2004. **38**(4): p. 418-21.
264. Deugnier, Y., et al., *Increased body iron stores in elite road cyclists.* Med Sci Sports Exerc, 2002. **34**(5): p. 876-80.
265. Burt, M.J., et al., *The significance of haemochromatosis gene mutations in the general population: implications for screening.* Gut, 1998. **43**(6): p. 830-6.
266. Prior, S.J., et al., *Sequence variation in hypoxia-inducible factor 1alpha (HIF1A): association with maximal oxygen consumption.* Physiol Genomics, 2003. **15**(1): p. 20-6.
267. Lunde, I.G., et al., *Hypoxia inducible factor 1 links fast-patterned muscle activity and fast muscle phenotype in rats.* J Physiol, 2011. **589**(Pt 6): p. 1443-54.

268. Mason, S.D., et al., *Loss of skeletal muscle HIF-1alpha results in altered exercise endurance.* PLoS Biol, 2004. **2**(10): p. e288.
269. Pistilli, E.E., et al., *Loss of IL-15 receptor alpha alters the endurance, fatigability, and metabolic characteristics of mouse fast skeletal muscles.* J Clin Invest, 2011. **121**(8): p. 3120-32.
270. Yi, Y., et al., *Association between KCNJ11 E23K genotype and cardiovascular and glucose metabolism phenotypes in older men and women.* Exp Physiol, 2008. **93**(1): p. 95-103.
271. Ben-Zaken, S., et al., *Differences in MCT1 A1470T polymorphism prevalence between runners and swimmers.* Scand J Med Sci Sports, 2015. **25**(3): p. 365-71.
272. Sawczuk, M., et al., *MCT1 A1470T: a novel polymorphism for sprint performance?* J Sci Med Sport, 2015. **18**(1): p. 114-8.
273. Fedotovskaya, O.N., et al., *A common polymorphism of the MCT1 gene and athletic performance.* Int J Sports Physiol Perform, 2014. **9**(1): p. 173-80.
274. Cupeiro, R., et al., *MCT1 genetic polymorphism influence in high intensity circuit training: a pilot study.* J Sci Med Sport, 2010. **13**(5): p. 526-30.
275. Merezhinskaya, N., et al., *Mutations in MCT1 cDNA in patients with symptomatic deficiency in lactate transport.* Muscle Nerve, 2000. **23**(1): p. 90-7.
276. Kim, K.C., H.I. Cho, and W. Kim, *MtDNA haplogroups and elite Korean athlete status.* Int J Sports Med, 2012. **33**(1): p. 76-80.
277. Niemi, A.K. and K. Majamaa, *Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes.* Eur J Hum Genet, 2005. **13**(8): p. 965-9.
278. Mikami, E., et al., *Comprehensive analysis of common and rare mitochondrial DNA variants in elite Japanese athletes: a case-control study.* J Hum Genet, 2013. **58**(12): p. 780-7.
279. Mikami, E., et al., *Polymorphisms in the control region of mitochondrial DNA associated with elite Japanese athlete status.* Scand J Med Sci Sports, 2013. **23**(5): p. 593-9.
280. Mikami, E., et al., *Mitochondrial haplogroups associated with elite Japanese athlete status.* Br J Sports Med, 2011. **45**(15): p. 1179-83.
281. Scott, R.A., et al., *Mitochondrial haplogroups associated with elite Kenyan athlete status.* Med Sci Sports Exerc, 2009. **41**(1): p. 123-8.
282. Castro, M.G., et al., *Mitochondrial haplogroup T is negatively associated with the status of elite endurance athlete.* Mitochondrion, 2007. **7**(5): p. 354-7.
283. Maruszak, A., et al., *Mitochondrial DNA variation is associated with elite athletic status in the Polish population.* Scand J Med Sci Sports, 2014. **24**(2): p. 311-8.
284. Starnes, L.M., et al., *NFI-A directs the fate of hematopoietic progenitors to the erythroid or granulocytic lineage and controls beta-globin and G-CSF receptor expression.* Blood, 2009. **114**(9): p. 1753-63.
285. Gomez-Gallego, F., et al., *The -786 T/C polymorphism of the NOS3 gene is associated with elite performance in power sports.* Eur J Appl Physiol, 2009. **107**(5): p. 565-9.
286. Ahsan, A., et al., *Simultaneous selection of the wild-type genotypes of the G894T and 4B/ 4A polymorphisms of NOS3 associate with high-altitude adaptation.* Ann Hum Genet, 2005. **69**(Pt 3): p. 260-7.
287. Wolfarth, B., et al., *Endothelial nitric oxide synthase gene polymorphism and elite endurance athlete status: the Genathlete study.* Scand J Med Sci Sports, 2008. **18**(4): p. 485-90.
288. Eynon, N., et al., *Do PPARGC1A and PPARalpha polymorphisms influence sprint or endurance phenotypes?* Scand J Med Sci Sports, 2010. **20**(1): p. e145-50.
289. Ahmetov, II, et al., *PPARalpha gene variation and physical performance in Russian athletes.* Eur J Appl Physiol, 2006. **97**(1): p. 103-8.
290. Tural, E., et al., *PPAR-alpha and PPARGC1A gene variants have strong effects on aerobic performance of Turkish elite endurance athletes.* Mol Biol Rep, 2014. **41**(9): p. 5799-804.

291. Maciejewska, A., M. Sawczuk, and P. Cieszczyk, *Variation in the PPARalpha gene in Polish rowers.* J Sci Med Sport, 2011. **14**(1): p. 58-64.
292. Akhmetov, II, et al., *[Association of regulatory genes polymorphisms with aerobic and anaerobic performance of athletes].* Ross Fiziol Zh Im I M Sechenova, 2007. **93**(8): p. 837-43.
293. Hautala, A.J., et al., *Peroxisome proliferator-activated receptor-delta polymorphisms are associated with physical performance and plasma lipids: the HERITAGE Family Study.* Am J Physiol Heart Circ Physiol, 2007. **292**(5): p. H2498-505.
294. Eynon, N., et al., *Is there an interaction between PPARD T294C and PPARGC1A Gly482Ser polymorphisms and human endurance performance?* Exp Physiol, 2009. **94**(11): p. 1147-52.
295. Vanttinen, M., et al., *Single nucleotide polymorphisms in the peroxisome proliferator-activated receptor delta gene are associated with skeletal muscle glucose uptake.* Diabetes, 2005. **54**(12): p. 3587-91.
296. Skogsberg, J., et al., *Evidence that peroxisome proliferator-activated receptor delta influences cholesterol metabolism in men.* Arterioscler Thromb Vasc Biol, 2003. **23**(4): p. 637-43.
297. Wang, Y.X., et al., *Regulation of muscle fiber type and running endurance by PPARdelta.* PLoS Biol, 2004. **2**(10): p. e294.
298. He, Z.H., et al., *PGC-related gene variants and elite endurance athletic status in a Chinese cohort: a functional study.* Scand J Med Sci Sports, 2015. **25**(2): p. 184-95.
299. Maciejewska, A., et al., *The PPARGC1A gene Gly482Ser in Polish and Russian athletes.* J Sports Sci, 2012. **30**(1): p. 101-13.
300. Lucia, A., et al., *PPARGC1A genotype (Gly482Ser) predicts exceptional endurance capacity in European men.* J Appl Physiol (1985), 2005. **99**(1): p. 344-8.
301. Stefan, N., et al., *Genetic variations in PPARD and PPARGC1A determine mitochondrial function and change in aerobic physical fitness and insulin sensitivity during lifestyle intervention.* J Clin Endocrinol Metab, 2007. **92**(5): p. 1827-33.
302. Ling, C., et al., *Multiple environmental and genetic factors influence skeletal muscle PGC-1alpha and PGC-1beta gene expression in twins.* J Clin Invest, 2004. **114**(10): p. 1518-26.
303. Ling, C., et al., *Impact of the peroxisome proliferator activated receptor-gamma coactivator-1beta (PGC-1beta) Ala203Pro polymorphism on in vivo metabolism, PGC-1beta expression and fibre type composition in human skeletal muscle.* Diabetologia, 2007. **50**(8): p. 1615-20.
304. Olsson, A.H., et al., *The expression of myosin heavy chain (MHC) genes in human skeletal muscle is related to metabolic characteristics involved in the pathogenesis of type 2 diabetes.* Mol Genet Metab, 2011. **103**(3): p. 275-81.
305. Arany, Z., et al., *The transcriptional coactivator PGC-1beta drives the formation of oxidative type IIX fibers in skeletal muscle.* Cell Metab, 2007. **5**(1): p. 35-46.
306. Fogel, B.L., et al., *RBFOX1 regulates both splicing and transcriptional networks in human neuronal development.* Hum Mol Genet, 2012. **21**(19): p. 4171-86.
307. Xia, X., et al., *A functional promoter polymorphism of SLC2A4 is associated with aerobic endurance in a Chinese population.* Eur J Sport Sci, 2014. **14**(1): p. 53-9.
308. Norrbom, J., et al., *Training response of mitochondrial transcription factors in human skeletal muscle.* Acta Physiol (Oxf), 2010. **198**(1): p. 71-9.
309. Astrup, A., et al., *Impact of the v/v 55 polymorphism of the uncoupling protein 2 gene on 24-h energy expenditure and substrate oxidation.* Int J Obes Relat Metab Disord, 1999. **23**(10): p. 1030-4.
310. Buemann, B., et al., *The association between the val/ala-55 polymorphism of the uncoupling protein 2 gene and exercise efficiency.* Int J Obes Relat Metab Disord, 2001. **25**(4): p. 467-71.
311. Sessa, F., et al., *Gene polymorphisms and sport attitude in Italian athletes.* Genet Test Mol Biomarkers, 2011. **15**(4): p. 285-90.

312. Hudson, D.E., et al., *The -55 C/T polymorphism within the UCP3 gene and performance during the South African Ironman Triathlon.* Int J Sports Med, 2004. **25**(6): p. 427-32.
313. Schrauwen, P., et al., *A novel polymorphism in the proximal UCP3 promoter region: effect on skeletal muscle UCP3 mRNA expression and obesity in male non-diabetic Pima Indians.* Int J Obes Relat Metab Disord, 1999. **23**(12): p. 1242-5.
314. Schrauwen, P. and M. Hesselink, *UCP2 and UCP3 in muscle controlling body metabolism.* J Exp Biol, 2002. **205**(Pt 15): p. 2275-85.
315. Prior, S.J., et al., *DNA sequence variation in the promoter region of the VEGF gene impacts VEGF gene expression and maximal oxygen consumption.* Am J Physiol Heart Circ Physiol, 2006. **290**(5): p. H1848-55.
316. Ahmetov, II, et al., *Association of the VEGFR2 gene His472Gln polymorphism with endurance-related phenotypes.* Eur J Appl Physiol, 2009. **107**(1): p. 95-103.
317. Wang, Y., et al., *Polymorphisms of KDR gene are associated with coronary heart disease.* J Am Coll Cardiol, 2007. **50**(8): p. 760-7.
318. Moran, C.N., et al., *Y chromosome haplogroups of elite Ethiopian endurance runners.* Hum Genet, 2004. **115**(6): p. 492-7.
319. Scott, R.A., et al., *ACTN3 and ACE genotypes in elite Jamaican and US sprinters.* Med Sci Sports Exerc, 2010. **42**(1): p. 107-12.
320. Shahmoradi, S., A. Ahmadalipour, and M. Salehi, *Evaluation of ACE gene I/D polymorphism in Iranian elite athletes.* Adv Biomed Res, 2014. **3**: p. 207.
321. Wang, G., et al., *Association analysis of ACE and ACTN3 in elite Caucasian and East Asian swimmers.* Med Sci Sports Exerc, 2013. **45**(5): p. 892-900.
322. Boraita, A., et al., *Cardiovascular adaptation, functional capacity and Angiotensin-converting enzyme I/D polymorphism in elite athletes.* Rev Esp Cardiol, 2010. **63**(7): p. 810-9.
323. Costa, A.M., et al., *Association between ACE D allele and elite short distance swimming.* Eur J Appl Physiol, 2009. **106**(6): p. 785-90.
324. Woods, D., et al., *Elite swimmers and the D allele of the ACE I/D polymorphism.* Hum Genet, 2001. **108**(3): p. 230-2.
325. Zhang, B., et al., *The I allele of the angiotensin-converting enzyme gene is associated with an increased percentage of slow-twitch type I fibers in human skeletal muscle.* Clin Genet, 2003. **63**(2): p. 139-44.
326. Charbonneau, D.E., et al., *ACE genotype and the muscle hypertrophic and strength responses to strength training.* Med Sci Sports Exerc, 2008. **40**(4): p. 677-83.
327. Williams, A.G., et al., *Circulating angiotensin converting enzyme activity is correlated with muscle strength.* Med Sci Sports Exerc, 2005. **37**(6): p. 944-8.
328. Papadimitriou, I.D., et al., *The ACE I/D polymorphism in elite Greek track and field athletes.* J Sports Med Phys Fitness, 2009. **49**(4): p. 459-63.
329. Grealy, R., et al., *The genetics of endurance: frequency of the ACTN3 R577X variant in Ironman World Championship athletes.* J Sci Med Sport, 2013. **16**(4): p. 365-71.
330. Lucia, A., et al., *ACTN3 genotype in professional endurance cyclists.* Int J Sports Med, 2006. **27**(11): p. 880-4.
331. Saunders, C.J., et al., *No association of the ACTN3 gene R577X polymorphism with endurance performance in Ironman Triathlons.* Ann Hum Genet, 2007. **71**(Pt 6): p. 777-81.
332. Paparini, A., et al., *ACTN3 genotyping by real-time PCR in the Italian population and athletes.* Med Sci Sports Exerc, 2007. **39**(5): p. 810-5.
333. Shang, X., et al., *Association between the ACTN3 R577X polymorphism and female endurance athletes in China.* Int J Sports Med, 2010. **31**(12): p. 913-6.

334. Ahmetov, II, et al., *The ACTN3 R577X polymorphism in Russian endurance athletes.* Br J Sports Med, 2010. **44**(9): p. 649-52.
335. Ahmetov, II, A.E. Donnikov, and D.Y. Trofimov, *Actn3 genotype is associated with testosterone levels of athletes.* Biol Sport, 2014. **31**(2): p. 105-8.
336. Ahmetov, II, O.L. Vinogradova, and A.G. Williams, *Gene polymorphisms and fiber-type composition of human skeletal muscle.* Int J Sport Nutr Exerc Metab, 2012. **22**(4): p. 292-303.
337. Vincent, B., et al., *ACTN3 (R577X) genotype is associated with fiber type distribution.* Physiol Genomics, 2007. **32**(1): p. 58-63.
338. Yang, N., et al., *The ACTN3 R577X polymorphism in East and West African athletes.* Med Sci Sports Exerc, 2007. **39**(11): p. 1985-8.
339. Mikami, E., et al., *ACTN3 R577X genotype is associated with sprinting in elite Japanese athletes.* Int J Sports Med, 2014. **35**(2): p. 172-7.
340. Cieszczyk, P., et al., *Association of the ACTN3 R577X Polymorphism in Polish Power-Orientated Athletes.* J Hum Kinet, 2011. **28**: p. 55-61.
341. Ahmetov, II, et al., *The dependence of preferred competitive racing distance on muscle fibre type composition and ACTN3 genotype in speed skaters.* Exp Physiol, 2011. **96**(12): p. 1302-10.
342. Eynon, N., et al., *ACTN3 R577X polymorphism and Israeli top-level athletes.* Int J Sports Med, 2009. **30**(9): p. 695-8.
343. Chiu, L.L., et al., *ACTN3 genotype and swimming performance in Taiwan.* Int J Sports Med, 2011. **32**(6): p. 476-80.
344. Massidda, M., G. Vona, and C.M. Calo, *Association between the ACTN3 R577X polymorphism and artistic gymnastic performance in Italy.* Genet Test Mol Biomarkers, 2009. **13**(3): p. 377-80.
345. Druzhevskaya, A.M., et al., *Association of the ACTN3 R577X polymorphism with power athlete status in Russians.* Eur J Appl Physiol, 2008. **103**(6): p. 631-4.
346. Roth, S.M., et al., *The ACTN3 R577X nonsense allele is under-represented in elite-level strength athletes.* Eur J Hum Genet, 2008. **16**(3): p. 391-4.
347. Santiago, C., et al., *ACTN3 genotype in professional soccer players.* Br J Sports Med, 2008. **42**(1): p. 71-3.
348. Papadimitriou, I.D., et al., *The ACTN3 gene in elite Greek track and field athletes.* Int J Sports Med, 2008. **29**(4): p. 352-5.
349. Yang, N., et al., *ACTN3 genotype is associated with human elite athletic performance.* Am J Hum Genet, 2003. **73**(3): p. 627-31.
350. Zarebska, A., et al., *Association of rs699 (M235T) polymorphism in the AGT gene with power but not endurance athlete status.* J Strength Cond Res, 2013. **27**(10): p. 2898-903.
351. Gomez-Gallego, F., et al., *The C allele of the AGT Met235Thr polymorphism is associated with power sports performance.* Appl Physiol Nutr Metab, 2009. **34**(6): p. 1108-11.
352. Alves, G.B., et al., *Influence of angiotensinogen and angiotensin-converting enzyme polymorphisms on cardiac hypertrophy and improvement on maximal aerobic capacity caused by exercise training.* Eur J Cardiovasc Prev Rehabil, 2009. **16**(4): p. 487-92.
353. Cieszczyk, P., et al., *Distribution of the AMPD1 C34T polymorphism in Polish power-oriented athletes.* J Sports Sci, 2012. **30**(1): p. 31-5.
354. Fedotovskaya, O.N., A.A. Danilova, and Akhmetov, II, *Effect of AMPD1 gene polymorphism on muscle activity in humans.* Bull Exp Biol Med, 2013. **154**(4): p. 489-91.
355. Fischer, H., et al., *AMP deaminase deficiency is associated with lower sprint cycling performance in healthy subjects.* J Appl Physiol (1985), 2007. **103**(1): p. 315-22.
356. G. Wang, S.P., E. Miyamoto-Mikami, N. Fuku, M. Tanaka, M. Miyachi, and Y.C.C. H. Murakami, B.D. Mitchell, K.G. Austin, Y.P. Pitsiladis, *GWAS of elite Jamaican, African American and Japanese sprint athletes.* Med. Sci. Sports Exerc, 2014. **46 (5S)** p. 596–598.

357. V.A. Naumov, I.I.A., A.K. Larin, E.V. Generozov, N.A. Kulemin, E.A. Ospanova, A.V. Pavlenko, E.S. Kostryukova, D.G. Alexeev, V.M. Govorun,, *Genome-wide association analysis identifies a locus on DMD (dystrophin) gene for power athlete status in Russians.* Eur. J. Hum. Genet. , 2014. **22 (Supp. 1)** p. 502.
358. Zarebska, A., et al., *Association of the MTHFR 1298A>C (rs1801131) polymorphism with speed and strength sports in Russian and Polish athletes.* J Sports Sci, 2014. **32**(4): p. 375-82.
359. Terruzzi, I., et al., *Genetic polymorphisms of the enzymes involved in DNA methylation and synthesis in elite athletes.* Physiol Genomics, 2011. **43**(16): p. 965-73.
360. Barres, R., et al., *Acute exercise remodels promoter methylation in human skeletal muscle.* Cell Metab, 2012. **15**(3): p. 405-11.
361. Eynon, N., et al., *Is the interaction between HIF1A P582S and ACTN3 R577X determinant for power/sprint performance?* Metabolism, 2010. **59**(6): p. 861-5.
362. Drozdovska, S.B., et al., *The association of gene polymorphisms with athlete status in ukrainians.* Biol Sport, 2013. **30**(3): p. 163-7.
363. Gabbasov, R.T., et al., *The HIF1A gene Pro582Ser polymorphism in Russian strength athletes.* J Strength Cond Res, 2013. **27**(8): p. 2055-8.
364. Ahmetov, II, et al., *Effect of HIF1A gene polymorphism on human muscle performance.* Bull Exp Biol Med, 2008. **146**(3): p. 351-3.
365. Ben-Zaken, S., et al., *IGF-I receptor 275124A>C (rs1464430) polymorphism and athletic performance.* J Sci Med Sport, 2015. **18**(3): p. 323-7.
366. Ben-Zaken, S., et al., *Can IGF-I polymorphism affect power and endurance athletic performance?* Growth Horm IGF Res, 2013. **23**(5): p. 175-8.
367. Cauci, S., et al., *Variable number of tandem repeat polymorphisms of the interleukin-1 receptor antagonist gene IL-1RN: a novel association with the athlete status.* BMC Med Genet, 2010. **11**: p. 29.
368. Pedersen, B.K., *Special feature for the Olympics: effects of exercise on the immune system: exercise and cytokines.* Immunol Cell Biol, 2000. **78**(5): p. 532-5.
369. Eynon, N., et al., *Is the -174 C/G polymorphism of the IL6 gene associated with elite power performance? A replication study with two different Caucasian cohorts.* Exp Physiol, 2011. **96**(2): p. 156-62.
370. Eider, J., et al., *Association of the 174 G/C polymorphism of the IL6 gene in Polish power-orientated athletes.* J Sports Med Phys Fitness, 2013. **53**(1): p. 88-92.
371. Ruiz, J.R., et al., *The -174 G/C polymorphism of the IL6 gene is associated with elite power performance.* J Sci Med Sport, 2010. **13**(5): p. 549-53.
372. Serrano, A.L., et al., *Interleukin-6 is an essential regulator of satellite cell-mediated skeletal muscle hypertrophy.* Cell Metab, 2008. **7**(1): p. 33-44.
373. Febbraio, M.A. and B.K. Pedersen, *Contraction-induced myokine production and release: is skeletal muscle an endocrine organ?* Exerc Sport Sci Rev, 2005. **33**(3): p. 114-9.
374. Bailey, S.J., et al., *Dietary nitrate supplementation enhances muscle contractile efficiency during knee-extensor exercise in humans.* J Appl Physiol (1985), 2010. **109**(1): p. 135-48.
375. Broos, S., et al., *Is PPARalpha intron 7 G/C polymorphism associated with muscle strength characteristics in nonathletic young men?* Scand J Med Sci Sports, 2013. **23**(4): p. 494-500.
376. Egorova, E.S., et al., *The polygenic profile of Russian football players.* J Sports Sci, 2014. **32**(13): p. 1286-93.
377. Ahmetov, II, et al., *The association of ACE, ACTN3 and PPARA gene variants with strength phenotypes in middle school-age children.* J Physiol Sci, 2013. **63**(1): p. 79-85.
378. Ahmetov, II, et al., *PPARG Gene polymorphism and locomotor activity in humans.* Bull Exp Biol Med, 2008. **146**(5): p. 630-2.

379. Vanttinen, M., et al., *The effect of the Ala12 allele of the peroxisome proliferator-activated receptor-gamma2 gene on skeletal muscle glucose uptake depends on obesity: a positron emission tomography study.* J Clin Endocrinol Metab, 2005. **90**(7): p. 4249-54.
380. Deeb, S.S., et al., *A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity.* Nat Genet, 1998. **20**(3): p. 284-7.
381. Maciejewska-Karlowska, A., et al., *Association between the Pro12Ala polymorphism of the peroxisome proliferator-activated receptor gamma gene and strength athlete status.* PLoS One, 2013. **8**(6): p. e67172.
382. Ben-Zaken, S., et al., *Increased prevalence of MnSOD genetic polymorphism in endurance and power athletes.* Free Radic Res, 2013. **47**(12): p. 1002-8.
383. Akimoto, A.K., et al., *Evaluation of gene polymorphisms in exercise-induced oxidative stress and damage.* Free Radic Res, 2010. **44**(3): p. 322-31.
384. Sutton, A., et al., *The Ala16Val genetic dimorphism modulates the import of human manganese superoxide dismutase into rat liver mitochondria.* Pharmacogenetics, 2003. **13**(3): p. 145-57.
385. Shimoda-Matsubayashi, S., et al., *Structural dimorphism in the mitochondrial targeting sequence in the human manganese superoxide dismutase gene. A predictive evidence for conformational change to influence mitochondrial transport and a study of allelic association in Parkinson's disease.* Biochem Biophys Res Commun, 1996. **226**(2): p. 561-5.
386. Micheli, M.L., et al., *Angiotensin-converting enzyme/vitamin D receptor gene polymorphisms and bioelectrical impedance analysis in predicting athletic performances of Italian young soccer players.* J Strength Cond Res, 2011. **25**(8): p. 2084-91.
387. Diogenes, M.E., et al., *Vitamin D receptor gene FokI polymorphisms influence bone mass in adolescent football (soccer) players.* Eur J Appl Physiol, 2010. **108**(1): p. 31-8.
388. Hopkinson, N.S., et al., *Vitamin D receptor genotypes influence quadriceps strength in chronic obstructive pulmonary disease.* Am J Clin Nutr, 2008. **87**(2): p. 385-90.
389. Marees, A.T., et al., *A tutorial on conducting genome-wide association studies: Quality control and statistical analysis.* Int J Methods Psychiatr Res, 2018. **27**(2): p. e1608.
390. Lortie, G., et al., *Familial similarity in aerobic power.* Hum Biol, 1982. **54**(4): p. 801-12.
391. Bouchard, C., et al., *The response to exercise with constant energy intake in identical twins.* Obes Res, 1994. **2**(5): p. 400-10.
392. Hamel, P., et al., *Heredity and muscle adaptation to endurance training.* Med Sci Sports Exerc, 1986. **18**(6): p. 690-6.
393. Prud'homme, D., et al., *Sensitivity of maximal aerobic power to training is genotype-dependent.* Med Sci Sports Exerc, 1984. **16**(5): p. 489-93.
394. Bouchard, C., et al., *Familial aggregation of VO(2max) response to exercise training: results from the HERITAGE Family Study.* J Appl Physiol (1985), 1999. **87**(3): p. 1003-8.
395. Wang, G., et al., *The Future of Genomic Research in Athletic Performance and Adaptation to Training.* Med Sport Sci, 2016. **61**: p. 55-67.
396. Brini, M., et al., *The plasma membrane calcium pump in health and disease.* FEBS J, 2013. **280**(21): p. 5385-97.
397. Zhang, Q., et al., *Nesprin-2 is a multi-isomeric protein that binds lamin and emerin at the nuclear envelope and forms a subcellular network in skeletal muscle.* J Cell Sci, 2005. **118**(Pt 4): p. 673-87.
398. Street, V.A., et al., *Mutations in a plasma membrane Ca2+-ATPase gene cause deafness in deafwaddler mice.* Nat Genet, 1998. **19**(4): p. 390-4.
399. Smits, J.J., et al., *De novo and inherited loss-of-function variants of ATP2B2 are associated with rapidly progressive hearing impairment.* Hum Genet, 2019. **138**(1): p. 61-72.

400. Strehler, E.E. and D.A. Zacharias, *Role of alternative splicing in generating isoform diversity among plasma membrane calcium pumps.* Physiol Rev, 2001. **81**(1): p. 21-50.
401. Yoshinaga, T., et al., *A novel frameshift mutation of SYNE1 in a Japanese family with autosomal recessive cerebellar ataxia type 8.* Hum Genome Var, 2017. **4**: p. 17052.
402. Sandra, M., et al., *Emery-Dreifuss muscular dystrophy type 4: A new SYNE1 mutation associated with hypertrophic cardiomyopathy masked by a perinatal distress-related spastic diplegia.* Clin Case Rep, 2019. **7**(5): p. 1078-1082.
403. Izumi, Y., et al., *Cerebellar ataxia with SYNE1 mutation accompanying motor neuron disease.* Neurology, 2013. **80**(6): p. 600-1.
404. Fanin, M., et al., *Dominant muscular dystrophy with a novel SYNE1 gene mutation.* Muscle Nerve, 2015. **51**(1): p. 145-7.
405. Baumann, M., et al., *Homozygous SYNE1 mutation causes congenital onset of muscular weakness with distal arthrogryposis: a genotype-phenotype correlation.* Eur J Hum Genet, 2017. **25**(2): p. 262-266.
406. Earing, M., *The athlete and heart disease: diagnosis, evaluation, and management.* Pediatr Cardiol, 1999. **20**(4): p. 247.
407. Maron, B.J., *Sudden death in young athletes.* N Engl J Med, 2003. **349**(11): p. 1064-75.
408. Lavie, C.J., et al., *Exercise and the heart: risks, benefits, and recommendations for providing exercise prescriptions.* Ochsner J, 2001. **3**(4): p. 207-13.
409. Kay, J.N., M.W. Chu, and J.R. Sanes, *MEGF10 and MEGF11 mediate homotypic interactions required for mosaic spacing of retinal neurons.* Nature, 2012. **483**(7390): p. 465-9.
410. Tam, V., et al., *Benefits and limitations of genome-wide association studies.* Nat Rev Genet, 2019. **20**(8): p. 467-484.
411. Strehler, E.E., *Plasma membrane calcium ATPases as novel candidates for therapeutic agent development.* Journal of pharmacy & pharmaceutical sciences : a publication of the Canadian Society for Pharmaceutical Sciences, Societe canadienne des sciences pharmaceutiques, 2013. **16**(2): p. 190.
412. Haskell, G.T., et al., *Whole Exome Sequencing Identifies Truncating Variants in Nuclear Envelope Genes in Patients With Cardiovascular Disease.* Circ Cardiovasc Genet, 2017. **10**(3).
413. McCarthy, M.I. and J.N. Hirschhorn, *Genome-wide association studies: potential next steps on a genetic journey.* Hum Mol Genet, 2008. **17**(R2): p. R156-65.
414. Hamer, M., *Psychoneuroendocrinology and Physical Activity*. 2012, Oxford University Press.
415. Bowden Davies, K.A., et al., *Physical Activity and Sedentary Time: Association with Metabolic Health and Liver Fat.* Med Sci Sports Exerc, 2019. **51**(6): p. 1169-1177.
416. Heaney, L.M., K. Deighton, and T. Suzuki, *Non-targeted metabolomics in sport and exercise science.* J Sports Sci, 2019. **37**(9): p. 959-967.
417. Pitsiladis, Y.P. and R.J. Maughan, *The effects of exercise and diet manipulation on the capacity to perform prolonged exercise in the heat and in the cold in trained humans.* J Physiol, 1999. **517 (Pt 3)**: p. 919-30.
418. Duft, R.G., et al., *Metabolomics and Exercise: possibilities and perspectives.* Motriz: Revista de Educação Física, 2017. **23**(2).
419. Chevalier, L., et al., *Athlete's heart patterns in elite rugby players: effects of training specificities.* Arch Cardiovasc Dis, 2013. **106**(2): p. 72-8.
420. Barbier, J., et al., *Sports-specific features of athlete's heart and their relation to echocardiographic parameters.* Herz, 2006. **31**(6): p. 531-43.
421. Gonzalez-Alonso, J., *Human thermoregulation and the cardiovascular system.* Exp Physiol, 2012. **97**(3): p. 340-6.

422. Rowell, L.B., *Human cardiovascular control*. 1993, New York: Oxford University Press. xv, 500 : ill. ; 24 cm.
423. Laughlin, M.H., *Cardiovascular response to exercise.* Am J Physiol, 1999. **277**(6 Pt 2): p. S244-59.
424. Fagard, R., *Athlete's heart.* Heart, 2003. **89**(12): p. 1455-61.
425. Lustgarten, M.S., et al., *Identification of serum analytes and metabolites associated with aerobic capacity.* Eur J Appl Physiol, 2013. **113**(5): p. 1311-20.
426. Heaney, L.M., K. Deighton, and T. Suzuki, *Non-targeted metabolomics in sport and exercise science.* J Sports Sci, 2017: p. 1-9.
427. Schranner, D., et al., *Metabolite Concentration Changes in Humans After a Bout of Exercise: a Systematic Review of Exercise Metabolomics Studies.* Sports Med Open, 2020. **6**(1): p. 11.
428. Smith, A.A., et al., *Dihydrotestosterone is elevated following sprint exercise in healthy young men.* J Appl Physiol (1985), 2013. **114**(10): p. 1435-40.
429. Budde, H., et al., *The cortisol response to exercise in young adults.* Front Behav Neurosci, 2015. **9**: p. 13.
430. Budde, H., et al., *Hormonal responses to physical and cognitive stress in a school setting.* Neurosci Lett, 2010. **474**(3): p. 131-4.
431. Fietz, D., et al., *Membrane transporters for sulfated steroids in the human testis--cellular localization, expression pattern and functional analysis.* PLoS One, 2013. **8**(5): p. e62638.
432. Fuqua, J.S. and A.D. Rogol, *Neuroendocrine alterations in the exercising human: implications for energy homeostasis.* Metabolism, 2013. **62**(7): p. 911-21.
433. Jacob, M.H., et al., *Age-related effects of DHEA on peripheral markers of oxidative stress.* Cell Biochem Funct, 2010. **28**(1): p. 52-7.
434. Ferrando, A.A., et al., *Testosterone administration to older men improves muscle function: molecular and physiological mechanisms.* Am J Physiol Endocrinol Metab, 2002. **282**(3): p. E601-7.
435. Sato, K., et al., *Testosterone and DHEA activate the glucose metabolism-related signaling pathway in skeletal muscle.* Am J Physiol Endocrinol Metab, 2008. **294**(5): p. E961-8.
436. Belelli, D. and K.W. Gee, *5 alpha-pregnan-3 alpha,20 alpha-diol behaves like a partial agonist in the modulation of GABA-stimulated chloride ion uptake by synaptoneurosomes.* Eur J Pharmacol, 1989. **167**(1): p. 173-6.
437. Harteneck, C., *Pregnenolone sulfate: from steroid metabolite to TRP channel ligand.* Molecules, 2013. **18**(10): p. 12012-28.
438. Fasolato, C., et al., *Effect of 2-pyrrolidone on the concentration of GABA in rat tissues.* Pharmacology, 1988. **36**(4): p. 258-64.
439. Powers, M.E., et al., *Growth hormone isoform responses to GABA ingestion at rest and after exercise.* Med Sci Sports Exerc, 2008. **40**(1): p. 104-10.
440. Hoppel, C.L., *Carnitine and carnitine palmitoyltransferase in fatty acid oxidation and ketosis.* Fed Proc, 1982. **41**(12): p. 2853-7.
441. Karlic, H. and A. Lohninger, *Supplementation of L-carnitine in athletes: does it make sense?* Nutrition, 2004. **20**(7-8): p. 709-15.
442. Walker, J.B., *Creatine: biosynthesis, regulation, and function.* Adv Enzymol Relat Areas Mol Biol, 1979. **50**: p. 177-242.
443. Tullson, P.C. and R.L. Terjung, *Adenine nucleotide metabolism in contracting skeletal muscle.* Exerc Sport Sci Rev, 1991. **19**: p. 507-37.
444. Graham, T.E., *Caffeine and exercise: metabolism, endurance and performance.* Sports Med, 2001. **31**(11): p. 785-807.
445. Boldyrev, A. and H. Abe, *Metabolic transformation of neuropeptide carnosine modifies its biological activity.* Cell Mol Neurobiol, 1999. **19**(1): p. 163-75.

446. Powers, S.K. and M.J. Jackson, *Exercise-induced oxidative stress: cellular mechanisms and impact on muscle force production.* Physiol Rev, 2008. **88**(4): p. 1243-76.
447. Lecka-Czernik, B., et al., *Divergent effects of selective peroxisome proliferator-activated receptor-gamma 2 ligands on adipocyte versus osteoblast differentiation.* Endocrinology, 2002. **143**(6): p. 2376-84.
448. Coburn, R.F., C. Baron, and M.T. Papadopoulos, *Phosphoinositide metabolism and metabolism-contraction coupling in rabbit aorta.* Am J Physiol, 1988. **255**(6 Pt 2): p. H1476-83.
449. Circu, M.L. and T.Y. Aw, *Glutathione and modulation of cell apoptosis.* Biochim Biophys Acta, 2012. **1823**(10): p. 1767-77.
450. Lu, S.C., *Regulation of glutathione synthesis.* Mol Aspects Med, 2009. **30**(1-2): p. 42-59.
451. Yokoyama, H., *[Gamma glutamyl transpeptidase (gammaGTP) in the era of metabolic syndrome].* Nihon Arukoru Yakubutsu Igakkai Zasshi, 2007. **42**(3): p. 110-24.
452. Lovic, D., et al., *Left ventricular hypertrophy in athletes and hypertensive patients.* J Clin Hypertens (Greenwich), 2017. **19**(4): p. 413-417.
453. Wallimann, T., M. Tokarska-Schlattner, and U. Schlattner, *The creatine kinase system and pleiotropic effects of creatine.* Amino Acids, 2011. **40**(5): p. 1271-96.
454. Cruz, R.S., et al., *Intracellular shuttle: the lactate aerobic metabolism.* ScientificWorldJournal, 2012. **2012**: p. 420984.
455. Aledo, J.C., *Glutamine breakdown in rapidly dividing cells: waste or investment?* Bioessays, 2004. **26**(7): p. 778-85.
456. Roepstorff, C., et al., *Regulation of plasma long-chain fatty acid oxidation in relation to uptake in human skeletal muscle during exercise.* Am J Physiol Endocrinol Metab, 2004. **287**(4): p. E696-705.
457. Korotkova, M. and I.E. Lundberg, *The skeletal muscle arachidonic acid cascade in health and inflammatory disease.* Nat Rev Rheumatol, 2014. **10**(5): p. 295-303.
458. Braverman, N.E. and A.B. Moser, *Functions of plasmalogen lipids in health and disease.* Biochim Biophys Acta, 2012. **1822**(9): p. 1442-52.
459. Khani, S. and J.A. Tayek, *Cortisol increases gluconeogenesis in humans: its role in the metabolic syndrome.* Clin Sci (Lond), 2001. **101**(6): p. 739-47.
460. Christiansen, J.J., et al., *Effects of cortisol on carbohydrate, lipid, and protein metabolism: studies of acute cortisol withdrawal in adrenocortical failure.* J Clin Endocrinol Metab, 2007. **92**(9): p. 3553-9.
461. Kelly, J.J., et al., *Cortisol and hypertension.* Clin Exp Pharmacol Physiol Suppl, 1998. **25**: p. S51-6.
462. Shimomura, Y., et al., *Exercise promotes BCAA catabolism: effects of BCAA supplementation on skeletal muscle during exercise.* J Nutr, 2004. **134**(6 Suppl): p. 1583S-1587S.
463. Wolfe, R.R., *Branched-chain amino acids and muscle protein synthesis in humans: myth or reality?* J Int Soc Sports Nutr, 2017. **14**: p. 30.
464. Lawless, C.E., et al., *Sports and exercise cardiology in the United States: cardiovascular specialists as members of the athlete healthcare team.* J Am Coll Cardiol, 2014. **63**(15): p. 1461-72.
465. Zheng, Y., et al., *Dietary phosphatidylcholine and risk of all-cause and cardiovascular-specific mortality among US women and men.* Am J Clin Nutr, 2016. **104**(1): p. 173-80.
466. Stegemann, C., et al., *Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study.* Circulation, 2014. **129**(18): p. 1821-31.
467. Tobias, D.K., et al., *Circulating Branched-Chain Amino Acids and Incident Cardiovascular Disease in a Prospective Cohort of US Women.* Circ Genom Precis Med, 2018. **11**(4): p. e002157.
468. Orer, G.E. and N.A. Guzel, *The effects of acute L-carnitine supplementation on endurance performance of athletes.* J Strength Cond Res, 2014. **28**(2): p. 514-9.

469. Teres, S., et al., *Oleic acid content is responsible for the reduction in blood pressure induced by olive oil.* Proc Natl Acad Sci U S A, 2008. **105**(37): p. 13811-6.
470. Lone, A.M. and K. Tasken, *Proinflammatory and immunoregulatory roles of eicosanoids in T cells.* Front Immunol, 2013. **4**: p. 130.
471. Dennis, E.A. and P.C. Norris, *Eicosanoid storm in infection and inflammation.* Nat Rev Immunol, 2015. **15**(8): p. 511-23.
472. Liu, Y., et al., *Emerging regulatory paradigms in glutathione metabolism.* Adv Cancer Res, 2014. **122**: p. 69-101.
473. Grandjean, A.C., *Diets of elite athletes: has the discipline of sports nutrition made an impact?* J Nutr, 1997. **127**(5 Suppl): p. 874S-877S.
474. Zhang, J., et al., *Stachydrine ameliorates carbon tetrachloride-induced hepatic fibrosis by inhibiting inflammation, oxidative stress and regulating MMPs/TIMPs system in rats*. Biomed Pharmacother, 2018. **97**: p. 1586-1594.
475. Rangel-Huerta, O.D., et al., *A serum metabolomics-driven approach predicts orange juice consumption and its impact on oxidative stress and inflammation in subjects from the BIONAOS study.* Mol Nutr Food Res, 2017. **61**(2).
476. Lahlou, S., et al., *Cardiovascular effects of eugenol, a phenolic compound present in many plant essential oils, in normotensive rats.* J Cardiovasc Pharmacol, 2004. **43**(2): p. 250-7.
477. Hyduke, D.R., et al., *Integrated network analysis identifies nitric oxide response networks and dihydroxyacid dehydratase as a crucial target in Escherichia coli.* Proc Natl Acad Sci U S A, 2007. **104**(20): p. 8484-9.
478. Novaes, R.D., et al., *3,4-Dihydroxycinnamic acid attenuates the fatigue and improves exercise tolerance in rats.* Biosci Biotechnol Biochem, 2012. **76**(5): p. 1025-7.
479. Chen, Y.J., et al., *Caffeic acid phenethyl ester, an antioxidant from propolis, protects peripheral blood mononuclear cells of competitive cyclists against hyperthermal stress.* J Food Sci, 2009. **74**(6): p. H162-7.
480. Olthof, M.R., P.C. Hollman, and M.B. Katan, *Chlorogenic acid and caffeic acid are absorbed in humans.* J Nutr, 2001. **131**(1): p. 66-71.
481. Jung, U.J., et al., *Antihyperglycemic and antioxidant properties of caffeic acid in db/db mice.* J Pharmacol Exp Ther, 2006. **318**(2): p. 476-83.
482. Graf, E., *Antioxidant potential of ferulic acid.* Free Radic Biol Med, 1992. **13**(4): p. 435-48.
483. Eichel, A., et al., *Meta-analysis of the efficacy of ectoine nasal spray in patients with allergic rhinoconjunctivitis.* J Allergy (Cairo), 2014. **2014**: p. 292545.
484. Diener, H.C., et al., *Effectiveness of quinine in treating muscle cramps: a double-blind, placebo-controlled, parallel-group, multicentre trial.* Int J Clin Pract, 2002. **56**(4): p. 243-6.
485. Pallister, T., et al., *Hippurate as a metabolomic marker of gut microbiome diversity: Modulation by diet and relationship to metabolic syndrome.* Sci Rep, 2017. **7**(1): p. 13670.
486. Kafi, R., et al., *Improvement of naturally aged skin with vitamin A (retinol).* Arch Dermatol, 2007. **143**(5): p. 606-12.
487. Shorvon, S., *Pyrrolidone derivatives.* Lancet, 2001. **358**(9296): p. 1885-92.
488. Navarro, A., et al., *Dietary thioproline decreases spontaneous food intake and increases survival and neurological function in mice.* Antioxid Redox Signal, 2007. **9**(1): p. 131-41.
489. Yin, P., R. Lehmann, and G. Xu, *Effects of pre-analytical processes on blood samples used in metabolomics studies.* Anal Bioanal Chem, 2015. **407**(17): p. 4879-92.
490. Buchheit, M., et al., *Assessing running economy during field running with changes of direction: application to 20 m shuttle runs.* Int J Sports Physiol Perform, 2011. **6**(3): p. 380-95.

491. Buchheit, M. and P.B. Laursen, *High-intensity interval training, solutions to the programming puzzle. Part II: anaerobic energy, neuromuscular load and practical applications.* Sports Med, 2013. **43**(10): p. 927-54.
492. Williams, M.H., *Dietary supplements and sports performance: introduction and vitamins.* J Int Soc Sports Nutr, 2004. **1**: p. 1-6.
493. Rattray, N.J.W., et al., *Beyond genomics: understanding exposotypes through metabolomics.* Hum Genomics, 2018. **12**(1): p. 4.
494. Kottgen, A., et al., *Genome-Wide Association Studies of Metabolite Concentrations (mGWAS): Relevance for Nephrology.* Semin Nephrol, 2018. **38**(2): p. 151-174.
495. Adamski, J., *Genome-wide association studies with metabolomics.* Genome Med, 2012. **4**(4): p. 34.
496. Suhre, K., J. Raffler, and G. Kastenmuller, *Biochemical insights from population studies with genetics and metabolomics.* Arch Biochem Biophys, 2016. **589**: p. 168-76.
497. Tiffany, C.W. and B.S. Slusher, *Measurement of glutamate carboxypeptidase II (NAALADase) enzyme activity by the hydrolysis of [(3)H]-N-acetylaspartylglutamate (NAAG).* Curr Protoc Pharmacol, 2002. **Chapter 3**: p. Unit3 10.
498. Erickson, K.I., et al., *Beyond vascularization: aerobic fitness is associated with N-acetylaspartate and working memory.* Brain Behav, 2012. **2**(1): p. 32-41.
499. Dietrich, A. and W.F. McDaniel, *Endocannabinoids and exercise.* Br J Sports Med, 2004. **38**(5): p. 536-41.
500. Smit, P., et al., *A common polymorphism in the CYP3A7 gene is associated with a nearly 50% reduction in serum dehydroepiandrosterone sulfate levels.* J Clin Endocrinol Metab, 2005. **90**(9): p. 5313-6.
501. Ruth, K.S., et al., *Genome-wide association study with 1000 genomes imputation identifies signals for nine sex hormone-related phenotypes.* Eur J Hum Genet, 2016. **24**(2): p. 284-90.
502. Zhang, Y.Y. and L. Yang, *Interactions between human cytochrome P450 enzymes and steroids: physiological and pharmacological implications.* Expert Opin Drug Metab Toxicol, 2009. **5**(6): p. 621-9.
503. Demirkan, A., et al., *Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations.* PLoS Genet, 2012. **8**(2): p. e1002490.
504. Le Stunff, H., et al., *Recycling of sphingosine is regulated by the concerted actions of sphingosine-1-phosphate phosphohydrolase 1 and sphingosine kinase 2.* J Biol Chem, 2007. **282**(47): p. 34372-80.
505. Helge, J.W., et al., *Exercise and training effects on ceramide metabolism in human skeletal muscle.* Exp Physiol, 2004. **89**(1): p. 119-27.
506. Mueller, J.W., et al., *The Regulation of Steroid Action by Sulfation and Desulfation.* Endocr Rev, 2015. **36**(5): p. 526-63.
507. Bergman, B.C., et al., *Muscle sphingolipids during rest and exercise: a C18:0 signature for insulin resistance in humans.* Diabetologia, 2016. **59**(4): p. 785-98.
508. Borghouts, L.B. and H.A. Keizer, *Exercise and insulin sensitivity: a review.* Int J Sports Med, 2000. **21**(1): p. 1-12.
509. in *Food Components to Enhance Performance: An Evaluation of Potential Performance-Enhancing Food Components for Operational Rations*, B.M. Marriott, Editor. 1994: Washington (DC).
510. Fucho, R., et al., *Ceramides and mitochondrial fatty acid oxidation in obesity.* FASEB J, 2017. **31**(4): p. 1263-1272.
511. Turchanowa, L., et al., *Influence of physical exercise on polyamine synthesis in the rat skeletal muscle.* Eur J Clin Invest, 2000. **30**(1): p. 72-8.

512. Hodgson, L.E., et al., *Acute kidney injury associated with endurance events-is it a cause for concern? A systematic review.* BMJ Open Sport Exerc Med, 2017. **3**(1): p. e000093.
513. Phillips, G.C., *Glutamine: the nonessential amino acid for performance enhancement.* Curr Sports Med Rep, 2007. **6**(4): p. 265-8.

Appendix:

Table 33. Metabolites differentiating between moderate and high endurance athletes (FDR significance).

Metabolite	Subpathway	Change	FDR p value
1-stearoyl-GPC (18:0)	Lysolipid	-0.15595	1.15E-05
vanillylmandelate (VMA)	Phenylalanine and Tyrosine Metabolism	0.41513	1.15E-05
21-hydroxypregnenolone disulfate	Steroid	0.36586	3.27E-05
palmitoyl-linoleoyl-glycerol (16:0/18:2) [2]*	Diacylglycerol	-0.46764	3.27E-05
tartronate (hydroxymalonate)	Bacterial/Fungal	0.29008	0.0001314
palmitoyl-linoleoyl-glycerol (16:0/18:2) [1]*	Diacylglycerol	-0.42202	0.0001503
1-palmitoleoyl-GPC (16:1)*	Lysolipid	-0.22642	0.0001675
cortisone	Steroid	0.39589	0.0001862
Citrate	TCA Cycle	0.20006	0.0001983
succinimide	Chemical	0.27932	0.0002462
stearoylcarnitine (C18)	Fatty Acid Metabolism(Acyl Carnitine)	-0.28394	0.0002462
trans-4-hydroxyproline	Urea cycle; Arginine and Proline Metabolism	-0.27783	0.0002462
4-guanidinobutanoate	Guanidino and Acetamido Metabolism	-0.44969	0.000292
dihomo-linoleoylcarnitine (C20:2)*	Fatty Acid Metabolism(Acyl Carnitine)	-0.33166	0.000339
1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*	Plasmalogen	0.14518	0.000339
1-palmitoyl-GPC (16:0)	Lysolipid	-0.11595	0.000339
linoleoyl-linoleoyl-glycerol (18:2/18:2) [1]*	Diacylglycerol	-0.54301	0.000339
gamma-glutamylglutamate	Gamma-glutamyl Amino Acid	-0.42069	0.000339
pregnanediol-3-glucuronide	Steroid	0.44061	0.000339
palmitoyl-arachidonoyl-glycerol (16:0/20:4) [2]*	Diacylglycerol	-0.47247	0.0004183
1-palmitoyl-2-stearoyl-GPC (16:0/18:0)	Phospholipid Metabolism	-0.16480	0.0004456
cortisol	Steroid	0.47102	0.0004531
linoleoyl-linolenoyl-glycerol (18:2/18:3) [2]*	Diacylglycerol	-0.53635	0.0005231

homoarginine	Urea cycle; Arginine and Proline Metabolism	-0.22816	0.0005547
palmitoleoyl-linoleoyl-glycerol (16:1/18:2) [1]*	Diacylglycerol	-0.42989	0.0006222
lactosyl-N-palmitoyl-sphingosine (d18:1/16:0)	Sphingolipid Metabolism	0.13166	0.0006623
3-hydroxydecanoate	Fatty Acid, Monohydroxy	0.34676	0.0006623
pregnenolone sulfate	Steroid	0.33203	0.0006623
pregn steroid monosulfate*	Steroid	0.29255	0.0008307
leukotriene B4	Eicosanoid	-0.84063	0.000889
vanillactate	Phenylalanine and Tyrosine Metabolism	0.21476	0.000889
12-HETE	Eicosanoid	-0.63302	0.000889
acetylcarnitine (C2)	Fatty Acid Metabolism(Acyl Carnitine)	0.33732	0.0010032
N1-methyladenosine	Purine Metabolism, Adenine containing	0.12105	0.0010844
isovalerate	Leucine, Isoleucine and Valine Metabolism	-0.52129	0.0011245
5-hydroxylysine	Lysine Metabolism	-0.39575	0.0011279
1,3,7-trimethylurate	Xanthine Metabolism	0.67162	0.0012386
fructose	Fructose, Mannose and Galactose Metabolism	0.39170	0.0014126
N-(2-furoyl)glycine	Food Component/Plant	0.81319	0.0015494
indolepropionate	Tryptophan Metabolism	0.49356	0.0016031
lactosyl-N-nervonoyl-sphingosine (d18:1/24:1)*	Sphingolipid Metabolism	0.17154	0.00168
sphingosine 1-phosphate	Sphingolipid Metabolism	-0.15758	0.0019098
gamma-glutamylphenylalanine	Gamma-glutamyl Amino Acid	-0.17502	0.0020209
Maleate	Fatty Acid, Dicarboxylate	0.25860	0.0021182
oleoyl-linoleoyl-glycerol (18:1/18:2) [2]	Diacylglycerol	-0.29989	0.0022114
homovanillate (HVA)	Phenylalanine and Tyrosine Metabolism	0.40224	0.0023827
2-palmitoyl-GPC (16:0)*	Lysolipid	-0.13333	0.0026446
1-methylxanthine	Xanthine Metabolism	0.64534	0.0026446
1-methylhistidine	Histidine Metabolism	-0.17145	0.003005
12-HHTrE	Eicosanoid	-0.75757	0.003005

palmitoyl-arachidonoyl-glycerol (16:0/20:4) [1]*	Diacylglycerol	-0.50107	0.003005
etiocholanolone glucuronide	Steroid	0.30363	0.0030403
1-palmitoleoylglycerol (16:1)*	Monoacylglycerol	-0.33670	0.0032212
O-sulfo-L-tyrosine	Chemical	0.16206	0.0035299
Lysine	Lysine Metabolism	-0.13641	0.0035706
arachidonoylcarnitine (C20:4)	Fatty Acid Metabolism(Acyl Carnitine)	-0.36442	0.0035706
gamma-glutamyl-epsilon-lysine	Gamma-glutamyl Amino Acid	-0.37501	0.0035777
1-myristoylglycerol (14:0)	Monoacylglycerol	-0.38981	0.0035777
beta-alanine	Pyrimidine Metabolism, Uracil containing	-0.25684	0.0035777
oleoyl ethanolamide	Endocannabinoid	0.20627	0.0035777
3-sulfo-L-alanine	Alanine and Aspartate Metabolism	-0.47391	0.003658
2-hydroxypalmitate	Fatty Acid, Monohydroxy	0.11888	0.0036606
Acisoga	Polyamine Metabolism	0.30843	0.0037833
glycosyl ceramide (d18:2/24:1, d18:1/24:2)*	Ceramides	0.15276	0.0041224
methylmalonate (MMA)	Fatty Acid Metabolism (also BCAA Metabolism)	0.27267	0.0041861
linoleoylcarnitine (C18:2)*	Fatty Acid Metabolism(Acyl Carnitine)	-0.29551	0.0041861
oleoyl-linoleoyl-glycerol (18:1/18:2) [1]	Diacylglycerol	-0.26673	0.004202
3beta-hydroxy-5-cholestenoate	Sterol	0.19787	0.0044408
stearoyl-arachidonoyl-glycerol (18:0/20:4) [2]*	Diacylglycerol	-0.28368	0.0044408
gamma-glutamyl-alpha-lysine	Gamma-glutamyl Amino Acid	-0.30540	0.0044408
3-hydroxysebacate	Fatty Acid, Monohydroxy	0.40398	0.0044746
1-linolenoylglycerol (18:3)	Monoacylglycerol	-0.30419	0.0044983
2-stearoyl-GPE (18:0)*	Lysolipid	-0.22072	0.0052777
sphingomyelin (d18:1/20:1, d18:2/20:0)*	Sphingolipid Metabolism	0.10195	0.0054039
cysteine sulfinic acid	Methionine, Cysteine, SAM and Taurine Metabolism	-0.33477	0.0064682
hydroquinone sulfate	Drug	0.41002	0.006698
ethyl glucuronide	Chemical	-2.42842	0.006698
4-HDoHE	Fatty Acid, Monohydroxy	-0.50237	0.0070239
catechol sulfate	Benzoate Metabolism	0.41217	0.0073787
3-hydroxylaurate	Fatty Acid, Monohydroxy	0.35708	0.0073787

glycerophosphoethanolamine	Phospholipid Metabolism	-0.14061	0.0076147
4-hydroxychlorothalonil	Chemical	-0.31832	0.0079474
Heme	Hemoglobin and Porphyrin Metabolism	1.15089	0.0084345
Quinate	Food Component/Plant	0.83873	0.0091191
2-aminoadipate	Lysine Metabolism	-0.27526	0.0093355
1-stearoyl-GPE (18:0)	Lysolipid	-0.15296	0.0100114
5alpha-pregnan-3beta,20alpha-diol disulfate	Steroid	0.33586	0.0102103
gamma-glutamylthreonine	Gamma-glutamyl Amino Acid	-0.20547	0.0103058
Glucose	Glycolysis, Gluconeogenesis, and Pyruvate Metabolism	0.18846	0.0103058
gamma-glutamylhistidine	Gamma-glutamyl Amino Acid	-0.23908	0.0104625
glutamate	Glutamate Metabolism	-0.12524	0.010644

Table 34. Metabolites differentiating between moderate and high power athletes (FDR significance).

Metabolite	Subpathway	Change	FDR p value
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	Phospholipid Metabolism	0.57762	5.92E-11
1-palmitoyl-2-oleoyl-GPI (16:0/18:1)*	Phospholipid Metabolism	0.42177	5.52E-08
imidazole lactate	Histidine Metabolism	0.44770	6.26E-07
1-stearoyl-2-oleoyl-GPC (18:0/18:1)	Phospholipid Metabolism	0.27902	1.13E-06
1-linolenoyl-GPC (18:3)*	Lysolipid	0.41482	1.85E-06
1-linoleoyl-2-linolenoyl-GPC (18:2/18:3)*	Phospholipid Metabolism	0.53798	1.85E-06
1-palmitoyl-2-linoleoyl-GPI (16:0/18:2)	Phospholipid Metabolism	0.44788	8.39E-06
1-palmitoyl-GPI (16:0)	Lysolipid	0.43822	1.27E-05
indolelactate	Tryptophan Metabolism	0.30948	1.98E-05
3-methylxanthine	Xanthine Metabolism	0.78892	2.05E-05
1,2-dilinoleoyl-GPC (18:2/18:2)	Phospholipid Metabolism	0.32413	2.05E-05
1-lignoceroyl-GPC (24:0)	Lysolipid	0.32113	2.39E-05
1-palmitoyl-2-stearoyl-GPC (16:0/18:0)	Phospholipid Metabolism	0.22285	2.48E-05
N-acetylcarnosine	Dipeptide Derivative	-0.33185	6.24E-05

1-stearoyl-2-oleoyl-GPI (18:0/18:1)*	Phospholipid Metabolism	0.34616	6.84E-05
N-acetylmethionine	Methionine, Cysteine, SAM and Taurine Metabolism	-0.58119	9.03E-05
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)	Phospholipid Metabolism	0.15356	0.000175446
argininate*	Urea cycle; Arginine and Proline Metabolism	0.42241	0.000182981
7-methylxanthine	Xanthine Metabolism	0.64804	0.000211714
homoarginine	Urea cycle; Arginine and Proline Metabolism	-0.27429	0.000330318
gamma-glutamylvaline	Gamma-glutamyl Amino Acid	-0.30520	0.000371267
sphingosine 1-phosphate	Sphingolipid Metabolism	-0.20846	0.000371267
phenyllactate (PLA)	Phenylalanine and Tyrosine Metabolism	0.30640	0.000422103
arabitol/xylitol	Pentose Metabolism	0.23942	0.00063113
1-palmitoleoyl-GPC (16:1)*	Lysolipid	0.22941	0.000707412
methionine sulfone	Methionine, Cysteine, SAM and Taurine Metabolism	0.30900	0.000770758
guanidinoacetate	Creatine Metabolism	-0.22401	0.001241547
1-stearoyl-2-linoleoyl-GPI (18:0/18:2)	Phospholipid Metabolism	0.26184	0.001241547
sphingomyelin (d18:2/14:0, d18:1/14:1)*	Sphingolipid Metabolism	0.21663	0.001241547
4-cholesten-3-one	Sterol	0.24271	0.001241547
1-palmitoyl-GPG (16:0)*	Lysolipid	0.30938	0.001292881
cholate	Primary Bile Acid Metabolism	1.18224	0.001292896
1-palmitoyl-GPE (16:0)	Lysolipid	0.23063	0.001492875
1-stearoyl-2-linoleoyl-GPC (18:0/18:2)*	Phospholipid Metabolism	0.11802	0.001555213
12,13-DiHOME	Fatty Acid, Dihydroxy	0.43529	0.001575022
1-palmitoyl-2-linoleoyl-GPC (16:0/18:2)	Phospholipid Metabolism	0.11566	0.001616411
1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*	Phospholipid Metabolism	0.38399	0.001616411
creatine	Creatine Metabolism	0.55544	0.001733863
gamma-glutamylhistidine	Gamma-glutamyl Amino Acid	-0.33728	0.001810227
campesterol	Sterol	0.30689	0.001822284
3beta,7alpha-dihydroxy-5-cholestenoate	Sterol	-0.25261	0.001822284
N6-carboxymethyllysine	Advanced Glycation End-product	0.38516	0.00216275
1,2-dipalmitoyl-GPC (16:0/16:0)	Phospholipid Metabolism	0.12589	0.002294911

gamma-glutamylleucine	Gamma-glutamyl Amino Acid	-0.32465	0.002313387
sulfate*	Chemical	0.14466	0.002415096
1-palmitoyl-2-oleoyl-GPE (16:0/18:1)	Phospholipid Metabolism	0.34116	0.002803794
5,6-dihydrothymine	Pyrimidine Metabolism, Thymine containing	-0.18793	0.002838364
5alpha-pregnan-3beta,20alpha-diol disulfate	Steroid	-0.45080	0.002864448
stearoyl sphingomyelin (d18:1/18:0)	Sphingolipid Metabolism	-0.14082	0.003218016
1-oleoyl-GPI (18:1)*	Lysolipid	0.30088	0.003528577
4-ethylphenylsulfate	Benzoate Metabolism	0.70216	0.003528577
13-HODE + 9-HODE	Fatty Acid, Monohydroxy	0.66970	0.003528577
arabonate/xylonate	Pentose Metabolism	0.26505	0.003717738
3-(4-hydroxyphenyl)lactate	Phenylalanine and Tyrosine Metabolism	0.25141	0.003717738
gamma-glutamylisoleucine*	Gamma-glutamyl Amino Acid	-0.30406	0.003874496
creatinine	Creatine Metabolism	-0.10295	0.003874496
1,2-dipalmitoyl-GPE (16:0/16:0)*	Phospholipid Metabolism	0.35941	0.003874496
gamma-glutamyl-alpha-lysine	Gamma-glutamyl Amino Acid	-0.36266	0.003874496
heptanoate (7:0)	Medium Chain Fatty Acid	0.24158	0.004255491
beta-sitosterol	Sterol	0.26580	0.004255491
3-methylglutaconate	Leucine, Isoleucine and Valine Metabolism	0.28198	0.004255491
4-acetylphenol sulfate	Drug	0.52182	0.004586737
kynurenine	Tryptophan Metabolism	0.13812	0.004889979
guanidinosuccinate	Guanidino and Acetamido Metabolism	0.37809	0.00537015
pyroglutamine*	Glutamate Metabolism	-0.32981	0.005666422
gamma-glutamylglutamate	Gamma-glutamyl Amino Acid	-0.37683	0.005692175
1,2-dilinoleoyl-GPE (18:2/18:2)*	Phospholipid Metabolism	0.40914	0.005692175
tryptophan betaine	Tryptophan Metabolism	0.75506	0.006032691
C-glycosyltryptophan	Tryptophan Metabolism	-0.12468	0.006859968
laurate (12:0)	Medium Chain Fatty Acid	0.41691	0.006859968
N-acetylthreonine	Glycine, Serine and Threonine Metabolism	-0.12801	0.007149839
7-ketodeoxycholate	Secondary Bile Acid Metabolism	0.66922	0.007149839

4-hydroxyphenylacetylglutamine	Acetylated Peptides	0.33818	0.007258688
1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	Phospholipid Metabolism	0.32668	0.007500475
3,7-dimethylurate	Xanthine Metabolism	0.44430	0.007670622
9,10-DiHOME	Fatty Acid, Dihydroxy	0.37944	0.007822248
1-palmitoleoyl-GPA (16:1)*	Lysolipid	0.60491	0.008276742
N-palmitoyl-sphingadienine (d18:2/16:0)*	Sphingolipid Metabolism	-0.17777	0.008431536
palmitoyl-arachidonoyl-glycerol (16:0/20:4) [2]*	Diacylglycerol	0.41426	0.008431536
3-methoxycatechol sulfate (1)	Benzoate Metabolism	0.65147	0.008560722
1-linoleoyl-GPI (18:2)*	Lysolipid	0.24058	0.008675055
glycerophosphoethanolamine	Phospholipid Metabolism	0.16015	0.00918601
2-palmitoleoyl-GPC (16:1)*	Lysolipid	0.22177	0.009406825
homocitrulline	Urea cycle; Arginine and Proline Metabolism	0.26455	0.009406825
theobromine	s Metabolism	0.58107	0.009406825
4-guanidinobutanoate	Guanidino and Acetamido Metabolism	0.36764	0.0096592
2,3-dihydroxy-2-methylbutyrate	Leucine, Isoleucine and Valine Metabolism	0.31408	0.0096592
hyocholate	Secondary Bile Acid Metabolism	0.42358	0.0096592
5-methyluridine (ribothymidine)	Pyrimidine Metabolism, Uracil containing	-0.10724	0.01221666
tartarate	Food Component/Plant	0.80659	0.012344054
2-aminobutyrate	Methionine, Cysteine, SAM and Taurine Metabolism	-0.23588	0.013072806
gamma-glutamylthreonine	Gamma-glutamyl Amino Acid	-0.23184	0.013162001
gamma-glutamylglycine	Gamma-glutamyl Amino Acid	-0.39783	0.013257666
tyramine O-sulfate	Phenylalanine and Tyrosine Metabolism	-0.74942	0.013392292
4-acetamidobutanoate	Polyamine Metabolism	0.55369	0.014175929
3-methylhistidine	Histidine Metabolism	-0.52178	0.014982769
adenine	Purine Metabolism, Adenine containing	0.28513	0.014982769
alpha-CEHC sulfate	Tocopherol Metabolism	0.51623	0.014982769

N6-succinyladenosine	Purine Metabolism, Adenine containing	0.28515	0.015243642
choline phosphate	Phospholipid Metabolism	0.18862	0.015961835
taurocholenate sulfate	Secondary Bile Acid Metabolism	-0.49022	0.015974106
vanillic alcohol sulfate	Phenylalanine and Tyrosine Metabolism	1.04294	0.016756831
palmitoyl-arachidonoyl-glycerol (16:0/20:4) [1]*	Diacylglycerol	0.47842	0.016871845
androsterone sulfate	Steroid	-0.35775	0.016871845
2-ethylphenylsulfate	Benzoate Metabolism	-0.93329	0.018071801
1-oleoyl-GPC (18:1)	Lysolipid	0.10707	0.018071801
caproate (6:0)	Medium Chain Fatty Acid	0.25455	0.018071801
5-hydroxymethyl-2-furoic acid	Phenylalanine and Tyrosine Metabolism	0.53433	0.018071801
thyroxine	Phenylalanine and Tyrosine Metabolism	-0.11465	0.020186359
bilirubin (E,E)*	Hemoglobin and Porphyrin Metabolism	-0.73304	0.020267553
pyrraline	Food Component/Plant	0.40521	0.021554639
3-methoxytyrosine	Phenylalanine and Tyrosine Metabolism	-0.13376	0.022426992
1-linoleoyl-GPE (18:2)*	Lysolipid	0.24171	0.024606025
dodecanedioate	Fatty Acid, Dicarboxylate	0.40277	0.025766032
1-stearoyl-GPC (18:0)	Lysolipid	0.09055	0.025766032
N-delta-acetylornithine	Urea cycle; Arginine and Proline Metabolism	0.28407	0.025766032
retinal	Food Component/Plant	0.27933	0.0265451
5alpha-pregnan-3beta,20alpha-diol monosulfate (2)	Steroid	-0.33846	0.0285104
N-palmitoyl-sphinganine (d18:0/16:0)	Sphingolipid Metabolism	0.18664	0.028798657
lanthionine	Chemical	0.42303	0.029039852
1-(1-enyl-stearoyl)-2-linoleoyl-GPE (P-18:0/18:2)*	Plasmalogen	0.20378	0.029039852
succinylcarnitine (C4-DC)	TCA Cycle	0.21632	0.029039852
sphingomyelin (d18:2/18:1)*	Sphingolipid Metabolism	-0.11159	0.029826152
pimeloylcarnitine/3-methyladipoylcarnitine (C7-DC)	Fatty Acid Metabolism(Acyl Carnitine)	0.31804	0.030688053

gamma-glutamylmethionine	Gamma-glutamyl Amino Acid	-0.36220	0.031640482
leucylalanine	Dipeptide	0.39203	0.032623177
1-stearoyl-GPI (18:0)	Lysolipid	0.22787	0.032623177
lignoceroylcarnitine (C24)*	Fatty Acid Metabolism(Acyl Carnitine)	0.22343	0.035483139
1-stearoyl-2-linoleoyl-GPE (18:0/18:2)*	Phospholipid Metabolism	0.26898	0.035631174
1-linoleoyl-2-arachidonoyl-GPC (18:2/20:4n6)*	Phospholipid Metabolism	0.17919	0.036002122
1-(1-enyl-stearoyl)-2-oleoyl-GPE (P-18:0/18:1)	Plasmalogen	0.19758	0.036760403
chenodeoxycholate	Primary Bile Acid Metabolism	0.42480	0.036877113
decanoylcarnitine (C10)	Fatty Acid Metabolism(Acyl Carnitine)	-0.34885	0.037775742
N-formylanthranilic acid	Tryptophan Metabolism	0.24946	0.038257071
1-oleoyl-2-linoleoyl-GPE (18:1/18:2)*	Phospholipid Metabolism	0.29884	0.041917323
isoleucine	Leucine, Isoleucine and Valine Metabolism	-0.13383	0.042556633
o-cresol sulfate	Phenylalanine and Tyrosine Metabolism	-0.27722	0.04423568
epiandrosterone sulfate	Steroid	-0.27333	0.04423568
5alpha-pregnan-3beta,20beta-diol monosulfate (1)	Steroid	-0.28840	0.04423568
hexanoylcarnitine (C6)	Fatty Acid Metabolism(Acyl Carnitine)	-0.23139	0.045863819
1-stearoyl-2-oleoyl-GPE (18:0/18:1)	Phospholipid Metabolism	0.26286	0.046313759
kynurenate	Tryptophan Metabolism	0.17134	0.046687188
urea	Urea cycle; Arginine and Proline Metabolism	0.11514	0.046687188
octanoylcarnitine (C8)	Fatty Acid Metabolism(Acyl Carnitine)	-0.31235	0.048188258
N6-carbamoylthreonyladenosine	Purine Metabolism, Adenine containing	-0.09472	0.048188258