DEVELOPING INNOVATIVE ANALYTICAL AND COMPUTATIONAL METHODS FOR THE STUDY OF HUMAN METABOLISM

by

Annie Joann Harwood-Stamper

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences

College of Life and Environmental Sciences

University of Birmingham

December 2021

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Covid-19 Statement

During my PhD studies, the Covid-19 pandemic restrictions caused multiple disruptions to my planned PhD projects.

Two planned human biological research chapters and one planned analytical chemistry research chapter were unable to be performed because of restrictions on sample collection for human studies and the availability of laboratory time to collect data for the planned analytical chemistry chapter. These three chapters were replaced by two alternative chapters. These alternative chapters included chapter two (Retention Time (RT) mapping) which was conducted as a computational chapter to reduce time lost because of the laboratories being closed at the University of Birmingham for six months. The research described in chapter three was also included to take advantage of samples already available from studies completed prior to the emergence of COVID-19. This rescoped research allowed for this project to progress and avoid any additional delays that had emerged from COVID-19 regulations and delays.

Chapter four was planned to be performed before Covid-19 restrictions came in to place in March 2020. The research was conducted in collaboration with the University of Portsmouth and the Defence Science and Technology Laboratory. The human study was delayed by six months and was reduced from sixteen participants to six to complete the study in time for metabolomic analysis. This is because the number of participants allowed in the experimental laboratories was reduced from four to one, and multiple extra precautions were required to keep both participants and researchers protected from possible exposure to Covid-19. The study was designed to include two arms of research, traditional acclimation over ten days and a shortened acclimation period of three days. However, due to Covid-19 only the shortened acclimation arm was performed at the University of Portsmouth.

Abstract

Currently, within untargeted metabolomics there are several limitations that are delaying the speed at which new knowledge can be gained. One of these is the analysis methods applied where only low sample volume/masses are available. Another is metabolite annotation and identification because two orthogonal data matches with matching data collected for chemical standards in-house are required for identification. Metabolomics is commonly applied within heath and exercise science and is extremely important to investigate the effects of exercise on the human body for health research, sports performance, and nutritional needs. There have been multiple studies conducted to study the differences in medium, high, and maximal intensity exercise, but none that have compared all three at once. Another area of health and exercise metabolomics that is currently limited is heat acclimation. As this is commonly applied to both military and athletes, allowing for optimal performance in hot climates, more understanding on how heat acclimation affects the human metabolome is required.

This thesis aims to develop and optimize UHPLC-MS methods to allow for a wider range of sample types to be analysed along with expanding the bottleneck area of metabolite annotation further. Aqueous reversed phase, lipidomic C18 reversed phase and HILIC assays were developed applying 1mm internal diameter (ID) columns and compared to commonly applied 2.1mm ID columns for the same assays. Following this, the 1mm ID assays were validated using porcine plasma, urine, and tear fluid samples. Results demonstrated that current 2.1mm ID assays are suitable for biofluid sample volumes typically collected in untargeted metabolomic studies. When applying more dilute samples, a 1mm ID column becomes more advantageous. Results for the porcine tear fluid demonstrated that tear fluid is a metabolite rich sample type that is applicable for future metabolomics research.

Next, retention time mapping methods were developed and cross validated to allow for different RT libraries around the world to be able to be transferred between research groups. Results concluded that RT mapping methods proved that it is possible to transfer retention times between research groups, even when alternative UHPLC-MS instrumentation, column stationary phases and mobile phases were applied. An in-house lipids retention time library was constructed from regression models for each class with the aim to predict the retention time for lipids of the same class where chemical standards are not available.

This thesis also aims to expand the knowledge gap areas on how the human body reacts to heat acclimation and different intensities of exercise. Two different biological studies were conducted to investigate how the human metabolome responds to different intensities of exercise and heat acclimation, respectively. Results showed that as the intensity of exercise increases, the body response changes as there is a larger requirement for energy and that heat acclimation influences biofluid metabolomes, with multiple pathways being perturbed across both days of condensed heat acclimation sessions.

Acknowledgements

First and foremost, I would like to extend my sincere gratitude to my supervisor Prof. Warwick Dunn for all the help and supervision he has provided throughout the period of my studies at Birmingham. I know I would not be the scientist I am today without your input.

Secondly to Dr Caroline Rowland, Thank you for all your support and guidance throughout my PhD.

To all the members of the Fourth floor, thank you for being such a great support network over the past four years. A special thanks go to Dr Judith Ngere, for the hours of laughter, motivation and countless mental health healing sessions. I know I wouldn't have made it this far without your support and friendship. Dr Lukáš Najdekr; for your support in the lab and making me feel less of a failure whenever I was struggling the instrument. Dr Rose Barnett, for always keeping me grounded and providing perspective on many issues that the last four years have bought. (Nearly) Dr Tara Bowen; for listening to my rants and allowing me to babble on whilst I problem solved out loud. Your office was truly my salvage and home away from home whilst working on campus.

Thanks, are also given to the Birmingham Metabolomics Training Centre and Phenome Centre Birmingham. I would also like to extend my thanks to the Engineering and Physical Sciences Research Council (EPSRC) and the Defence Science and Technology Laboratory (Dstl) for funding for this project.

To my family, thank you for always believing in me and for all for the love and support you have given throughout my education. Thank you for providing me with the opportunities and experiences that have made me who I am. You have always encouraged me to explore new directions in life and chase my dreams, teaching me to overcome any obstacles in my path with tenacity, hard work and wit. I hope I have made you proud with all I have achieved.

To Amber, your kindness, jokes and empathy have been appreciated more than you know. Especially throughout Covid-19 when you provided me with an online office and an open ear daily.

Finally, Scott. I don't think I will ever be able to thank you enough for all the emotional help and support that you have given me. You have been exceedingly patient, kind and understanding these past 4 years. I am sure that if not for you, I would not be in this position right now and for that I will forever be grateful.

Table of Contents

1 Introduction 1
1.1 Metabolomics 1
1.1.1 Untargeted 4
1.1.2 Targeted 5
1.1.3 Semi-Targeted 5
1.2 Data Acquisition 5
1.2.1 Ultra-High-Performance Liquid Chromatography Theory 7
1.2.2 Parameters which can influence chromatographic resolution 11
1.2.3 Column Chemistry 14
1.2.4 Electrospray Ionisation (ESI) 17
1.2.5 Mass Spectrometry 18
1.3 Data Processing 21
1.3.1 Data Pre-Processing 22
1.4 Metabolite Annotation and Identification 23
1.5 Univariate and Multivariate Analysis 24
1.5.1 Missing value imputation and data normalisation 24
1.5.2 Multivariate Analysis 25
1.5.3 Univariate analysis 25
1.6 Sample Types 28
1.6.1 Blood and Blood Fractions 28
1.6.2 Urine 28
1.6.3 Tears 29
1.6.4 Other common sample types 31
1.7 Sample Preparation 31
1.7.1 Monophasic extraction 31
1.7.2 Biphasic Extraction 32
1.8 Importance Of Metabolomics In Exercise 32

1.9 Overview and Current Limitations Conclusions 33

1.10 Objectives of the PhD 33

2 Materials And Methods 35

2.1 Materials 35

2.2 Ethics Statements 35

2.2.1 Ethics Statement For Research Featured In Development Of Microflow Ultra High-Performance Liquid Chromatography -Mass Spectrometry Metabolomic Assays For Analysis Of Mammalian Biofluids 35

2.2.2 Ethics Statement For An Investigation Into How The Type And Intensity Of Exercise Perturbs Metabolism 35

2.2.3 Ethics Statement For Research Into Heat Stress And Acclimation Effects On Human Metabolism 35

2.3 Sample Collection 36

2.3.1 Sample Collection Methods For Research Featured In Development Of Microflow Ultra High-Performance Liquid Chromatography -Mass Spectrometry Metabolomic Assays For Analysis Of Mammalian Biofluids 36

2.3.2 Sample Collection, Delivery, And Receipt Methods For An Investigation Into How The Type And Intensity Of Exercise Perturbs Metabolism 36

2.3.3 Sample Collection, Shipment And Receipt Methods for Research Into Heat Stress And Acclimation Effects On Human Metabolism 36

2.4 Sample Extraction 37

2.5 UHPLC Analysis Methods 39

2.5.1 UHPLC-MS Methods Applied For Chapters 3 And 4 For Samples Analysed By Researchers At The University Of Birmingham 40

2.6 Methods applied to research featured in Development of Microflow Ultra High-Performance Liquid Chromatography -Mass Spectrometry Metabolomic Assays for Analysis of Mammalian Biofluids 41

2.6.1 C_{18} Reversed-Phase Lipidomics (Lipids) Assay Details 41

2.6.2 C_{18} Aqueous Reversed-Phase (RPaq) Assay 41

2.6.3 Hydrophilic Interaction Liquid Chromatography (HILIC) Assay 42

2.6.4 Raw Data Processing, Metabolite Annotation 43

2.6.5 Data Analysis 43

2.7 Methods applied to In-silico extension of RT libraries for water-soluble metabolites and lipids when applying UHPLC-MS. 44

2.7.1 Retention Time Mapping Datasets – Reversed Phase 44

2.7.2 Retention Time Mapping Datasets – HILIC 44

2.7.3 UHPLC Methods Applied By Different Research Groups 44

2.7.4 University of Liverpool UHPLC-MS Analysis 45

2.7.5 Fiehn HILIC MoNA UHPLC-MS Analysis Methods 46

2.7.6 University of Florida UHPLC-MS Analysis Methods 46

2.8 Methods Applied To An Investigation Into How The Type And Intensity Of Exercise Perturbs Metabolism 46

2.8.1 Experimental Visits 46

2.8.2 Sample Extraction 48

2.8.3 UHPLC-MS Analysis 49

2.8.4 Raw Data Processing, Metabolite Annotation 50

2.9 Methods Applied To Heat Stress And Acclimation Effects On Human Metabolism 51

2.9.1 Sample Extraction 51

2.9.2 UHPLC-MS Analysis 52

2.9.3 Raw Data Processing, Metabolite Annotation 53

2.9.4 Statistical Analysis 53

2.9.5 Pathway Enrichment Analysis 53

3 Development Of Microflow Ultra High-Performance Liquid Chromatography – Mass Spectrometry Metabolomic Assays For Analysis Of Mammalian Biofluids 54

3.1 Introduction 54

3.1.1 Research Objectives 56

3.2 Materials And Methods 56

3.3 Results And Discussion 56

Method Development Using Human Plasma And Urine Samples 56

3.3.1 Comparison Of The Number Of Detected Compounds 59

3.3.2 Application Of 1.0mm ID Columns To Study Different Biofluids 68

3.3.3 Putative Metabolite Annotations For Tear Fluid 70
3.4 Conclusions 70
3.5 Future Work 72
4 In-Silico Extension Of RT Libraries For Water-Soluble Metabolites And Lipids When Applying UHPLC-MS. 73
4.1 Introduction 73
4.1.1 Research Objectives 75
4.2 Results And Discussion 76
4.2.1 Retention Time Mapping 76
4.2.2 Retention Time Mapping Conclusions 96
4.2.3 Lipidomics Retention Time Library 97
4.2.4 Lipidomics In-Silico Retention Time Library Conclusions 115
4.3 Conclusions 116
4.3.1 Retention Time Mapping 116
4.3.2 Lipid Library 116
4.4 Future work 116
4.4.1 Retention Time Mapping 116
4.4.2 Lipidomics In-Silico Retention Time Library 117
5 An Investigation Into How The Type And Intensity Of Exercise Perturbs Metabolism 118
5.1 Introduction 118
5.1.1 Research Objectives 119
5.2 Materials And Methods 120
5.2.1 Study Design 120
5.2.2 Methods 120
5.3 Results and Discussion 120
5.3.1 What Are The Metabolic Changes Over Time Related To Exercise Type? 120
5.3.2 What Are The Metabolic Changes When Comparing All Exercise Types At Each Time Point? 132
5.3.3 Are There Any Differences In Metabolic Response To Exercise When It Comes To Exercise Type Vs Sex? 136

5.4 Conclusions 141
5.4.1 What Are The Metabolic Changes Over Time Related To Exercise Type? 141
5.4.2 What Are The Metabolic Changes When Comparing All Exercises At Each Time Point? 142
5.4.3 Are There Any Differences When It Comes To Exercise Type Vs Sex? 142
5.5 Future Work 143
6 Heat Stress And Acclimation Effects On Human Metabolism 144
6.1 Introduction 144
6.1.1 Research Objectives 145
6.2 Study Design 146
6.3 Materials And Methods 150
6.4 Results and Discussion 151
6.4.1 Are There Any Metabolic Differences In Samples Collected Before And After Heat Acclimation As Demonstrated By Two Independent Heat Stress Tests? 151
6.4.2 Are There Any Metabolic Differences In Samples Collected Before And After Heat Acclimation As Demonstrated By Two Independent Hypoxic Tolerance Tests? 157
6.4.3 Research Question 2: Are There Metabolic Changes Over Time During The Heat Acclimation Sessions? 158
6.5 Conclusions 164
7 Final Conclusions 166
8 References 175
9 Appendix 192

List of Figures

Figure 1-1: Examples of lipids from each lipid class to demonstrate the difference in structure. Figure adapted and recreated from[3] 2

Figure 1-2: A comparison of the two main workflows used in targeted and untargeted metabolomics research. Targeted first analyses metabolite standards to optimise methodology and gain exact spectra before applying these methods to biological samples, resulting in the ability to quantify certain metabolite levels within the biological samples. Untargeted metabolomic methods, in comparison, starts with the biological samples and aims to give a global profile of all the possible metabolites present, only allowing for relative quantification of metabolite concentrations. Figure recreated from [25] 4

Figure 1-3: A schematic diagram of a liquid chromatography instrument. Featuring mobile phases (bottles labelled A and B), pumps for each mobile phase to deliver each mobile phase at the desired flow rate, a mixer to mix the two mobile phases together into the correct percentages for that part of the elution gradient, auto sampler where all the samples are housed and kept at a stable temperature before they are automatically injected into the system at the start of each sample gradient, column (above in grey) for separation of analytes in sample, mass spectrometer for detection and identification of analytes. 7

Figure 1-4: The Van Deemter equation in graph form to show how altering the flow rate of a method can affect chromatography resolution. 10

Figure 1-5: Van Deemter graph applied to columns using different particle sizes to show the difference in the optimum flow velocity and how flow rate affects each particle size differently. Graph x-axis is flow velocity in mm/s and the y-axis is height of theoretical plate in µm. the three most common particle sizes used in analytical chemistry are featured on the graph, with the optimum flow velocity for each being defined by µopt. Figure has been recreated from[77] 13

Figure 1-6: Chemical structures for C_{18}, Biphenyl and pentafluorophenyl (PFP) column stationary phases..... 15

Figure 1-7: A diagram demonstrating the different separation mechanisms used in HILIC separation assays. 16

Figure 1-8: A schematic representation of an electrospray ionisation source...... 17

Figure 1-9: Adapted schematic of a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer from www.thermofisher.com..... 18

Figure 1-10: Schematic diagram to explain how ions are detected using an Orbitrap mass analyser, Figure reproduced from [95] 20

Figure 1-11: Two common data processing workflows used within untargeted metabolomics. The left-hand side shows the workflow when using Compound Discoverer ™ *workflow, and the right-hand side shows the workflow when using XCMS.*..... 22

Figure 1-12: A diagram of the different layers found within tear fluid. 29
Figure 2-1: "Participants (50/50 male/female split) took part in three separate exercise visits in a randomized order, consisting of 30 min of moderate, steady-state exercise at 65% VO2peak (MICT); four 4 min high intensity intervals at 85% HRmax (HIIT); and four 30 s supramaximal sprint intervals at 200% Wmax (SIT). All exercise protocols were preceded by 5 min of seated rest and a 5 min warm-up, and then followed by 3 mins of active recovery at 50 W and 15 mins of seated passive recovery. Graph Key: The timing of blood sampling is marked by red blood droplets. VO2peak = maximum oxygen consumption; HRmax = maximum heart rate; Wmax = maximum aerobic power."[206] Figure reconstructed from[206]. 47
Figure 3-2: PCA Scores plots to show the variation between different dilutions when applying the 2.1mm ID column (left) or 1.0mm ID column (right) to lipidomic assays in negative ion mode using plasma samples 57
Figure 3-1: PCA Scores plots to show the variation between different dilutions when applying the 2.1mm ID column (left) or 1.0mm ID column (right) to lipidomic assays in negative ion mode using plasma samples 57
Figure 3-3: PCA Scores plots to show the variation between different dilutions when applying the 2.1mm ID column (left) or 1.0mm ID column (right) to reversed phase aqueous assays in negative ion mode using urine samples 58
Figure 3-4:The number of reported compounds for each dilution in each assay. All compounds reported have a FWHM <0.2 minutes, were present in ¾ of injections and RSD <30% in each separate injection. The numbers in table format can be seen in appendix Figure D-1 60
*Figure 3-5: The average number of reported compounds in human plasma and urine samples following a 1 in 4 dilution for each of three UHPLC-MS assays. Compounds which were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. 2.1mm ID column (green) and 1.0mm ID column (Blue) *** has a significance of <0.01, ** has a significance of <0.05.* 61
*Figure 3-6: The average number of reported compounds in human plasma and urine samples following a 1 in 16 dilution for each of three UHPLC-MS assays. Compounds which were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. 2.1mm ID column (green) and 1.0mm ID column (Blue) *** has a significance of <0.001, NS has no significance.* 62
Figure 3-7: The average full width half max in minutes for each sample dilution in each assay. Compounds used were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. NEG = negative ion mode, POS = positive ion mode, SD = standard deviation. 63
*Figure 3-8: The distribution of chromatographic peak widths in minutes for the 1 in 16 dilutions of each assay type for both the 2.1mm ID (blue) and 1mm ID (red) columns. Compounds which were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. *** has a significance of <0.001, NS has no significance.* 64

Figure 3-9: A summary of the mean distribution of chromatographic peak areas for all dilutions of 2.1mm ID vs. 1mm ID for all assays and sample types. Compounds used were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. NEG = negative ion mode, POS = positive ion mode, SD = standard deviation. A table displaying the information can be found in the appendix under E-1. 66

Figure 3-10: The distribution of chromatographic peak areas in the 1 in 16 dilutions of each assay type for both the 2.1mm ID (blue) and 1mm ID (red) columns. Compounds which were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. Most comparisons were not significant (NS) (P > 0.05), whilst one was significant (*) (P ≤ 0.05), one was moderately significant (**) (P≤ 0.01) and one was very significant (***) (P ≤ 0.001). 67

Figure 3-11: Overlap of porcine plasma (red), urine (blue), and tear fluid (yellow) compounds detected during analysis using HILIC assay in negative ion mode. 68

Figure 3-12: Overlap of porcine plasma (red), urine (blue), and tear fluid (yellow) compounds detected during analysis using HILIC assay in positive ion mode 69

Figure 4-1: The relationship between retention times for the University of Liverpool method 1 dataset and the University of Liverpool method 2 dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time for both datasets was measured in minutes. Line of best fit: y = 0.9639x + 1.7167, R^2 = 0.9843 77

Figure 4-2: The relationship between retention times for the University Birmingham dataset and the University of Liverpool method 1 dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time for both datasets was measured in minutes. Line of best fit: y = -0.1765x3 + 1.8838x2 - 4.0261x + 3.1709 R^2 = 0.9694 78

Figure 4-3: The relationship between retention times for the University Birmingham dataset and the University of Liverpool method 2 dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time for both datasets is recorded in minutes. Line of best fit: y = 0.0455x4 - 0.8373x3 + 5.1526x2 - 10.462x + 7.3291 R^2 = 0.9773..... 80

Figure 4-4: The relationship between retention times for the University of Liverpool (method 2) dataset and the University of Florida dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time for both datasets is recorded in minutes. y = -0.001x3 + 0.0928x2 - 0.3663x + 7.6796 R^2 = 0.853..... 81

Figure 4-5: The relationship between retention times for the University of Birmingham dataset and the University of Florida dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time for both datasets is recorded in minutes...... 82

Figure 4-6: The relationship between retention times for the University of Birmingham dataset and the University of Florida dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time for both datasets is recorded in minutes...... 84

Figure 4-7: The relationship between retention times for the University of Birmingham positive ion mode dataset and the University of Birmingham negative ion mode dataset. The retention time for both datasets is recorded in minutes. Line of best fit equation: y = 0.9144x + 0.7678 R² value = 0.8801.86
Figure 4-8: The relationship between retention times for University of Birmingham positive ion mode dataset and the University of Birmingham negative ion mode dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time both datasets is recorded in minutes. Line of best fit equation: y = 0.9914x + 0.3017 R² value= 0.9538...87
Figure 4-9: The relationship between retention times for University of Birmingham positive ion mode dataset and the University of Liverpool dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time both datasets is recorded in minutes. Line of best fit equation: y = 0.0264x3 - 0.3157x2 + 1.5621x + 2.79; R² = 0.834 88
Figure 4-10: The relationship between retention times for the University of Birmingham negative ion mode dataset and the Fiehn MoNA dataset. The retention time for both datasets is recorded in minutes. Line of Best Fit equation: y = 0.858x + 1.8163; R² = 0.6043 90
Figure 4-11: The relationship between retention times for the University of Birmingham positive ion mode dataset and the Fiehn MoNA dataset. The retention time for both datasets is recorded in minutes. Line of best fit equation: y = 1.0483x + 1.4621; R² = 0.8496.....92
Figure 4-12: The relationship between retention times for University of Liverpool dataset and the Fiehn MoNA dataset. The retention time for both datasets is recorded in minutes. Line of best fit equation: y = 0.0208x3 - 0.3181x2 + 1.9889x + 0.1989; R² = 0.7481.....93
Figure 4-13: The relationship between retention times for University of Birmingham Negative Ion Mode and the University of Liverpool dataset. The retention time for both datasets is recorded in minutes. Line of best fit equation: y = 0.5414x + 2.521; R² = 0.3547.....95
Figure 4-14: The trend in mean retention time as the number of carbons increased for all acyl carnitines with zero or one double bond present. Acyl carnitine structures with zero double bonds present are denoted by blue data points and orange data points denote acyl carnitine structures with one double bond present......98
Figure 4-15: The trend in mean retention time as the total number of carbons increased in the lysophosphocholines containing zero or one double bond in the structure......100
Figure 4-16: The trend in mean retention time as the total number of carbons increased in the ceramides containing zero. d, m, t denotes whether the ceramide base is m- mono hydroxy, d- dihydroxy, or t- trihydroxy......102
Figure 4-17: The trend in mean retention time as the total number of carbons increased in the ceramides containing one double bond that have a dihydroxyl base......103
Figure 4-18: The trend in mean retention time as the total number of carbons increased in the ceramides containing two double bonds that have a dihydroxyl base. 103

Figure 4-19: The trend in mean retention time as the number of carbons increased in the metabolites classed as Hexosyl-2-ceramides by LipidSearch ™. All metabolites contained one double bond in the structure and were dihydroxyl bases. Line of best fit equation: y = 0.1708x + 1.4776; R² = 0.9867 105
Figure 4-20: The trend in mean retention time as the number of carbons increased in the metabolites classed as Glycerophosphocholines with zero double bonds by LipidSearch ™......107
Figure 4-21: The trend in mean retention time as the number of carbons increased in the metabolites classed as Glycerophosphoethanolamines by LipidSearchTM. All metabolites contained one double bond in the structure. Line of best fit equation: y = 0.1113x + 4.1328; R² = 0.9595.....109
Figure 4-22: The trend in mean retention time as the number of carbons increased in the metabolites classed as Glycerophosphoinositols. The light blue data points represent the trend in the metabolites containing one double bond. The orange data points represent the trend in the metabolites containing two double bonds. The grey data points represent the trend in the metabolites containing three double bonds......111
Figure 4-23: The trend in mean retention time as the number of carbons increased in the metabolites classed as Triacylglycerides with zero, one, two, three, four and six double bonds in the structure. 113
Figure 4-24: The different elution zones for different classes of lipids. AcCA = Acytlcarnitines, CL = Cardiolipins, DG = Diacylglycerols, HEX2CER =Hexosyl-2-ceramides, LPC = lyso- phosphocholines, MePC = Monoetherglycerophosphocholines, PA=Phosphatidates, PC = Phosphocholines, PE = Phosphatidylethanolamines, PI=Phosphatidylinositols, PS=Phosphatidylserines, SM=Sphingomyelins, TAG=Triacylglycerides. The carbon range is denoted by CX-Y, and the double bond range is denoted by dbX-Y. 115
Figure 5-1: The purine metabolism pathway as recreated from KEGG Pathway Database[290,291] and adapted to highlight the changing metabolites, along with the fold change increases seen in the paired t-test between starting rest vs active recovery. Fold change calculations were active recovery / starting rest. 122
Figure 5-2: Normalised box plots of two different lipids that increased in relative concentration when compared at rest (0) and active recovery (1).....124
Figure 5-3: Normalised box plots of different fatty acyls that decreased in relative concentration when compared at rest (0) and passive recovery (1)......128
Figure 6-1: Summary of the hypoxic tolerance test protocol performed in temperate conditions (21°C, 50%). Participants were exposed to a hypoxic mixture of 15% O2 (FIO2 = 0.15). Venous blood samples were taken before and after hypoxic exposure. Heart rate, inspired and expired oxygen and carbon dioxide and oxygen saturation were monitored and recorded throughout the protocol. 148
Figure 6-2: Summary of the heat stress test (HST) protocol undertaken in hot conditions (40 degrees, 50% RH). Nude mass and a urine sample were recorded before and after entering the chamber. Data from the rectal thermistor, skin thermistors and heart rate were logged at 60 second intervals throughout the HST. Skin blood flow using a Doppler probe and local sweat rate from the right scapula

using a Q sweat were measured continuously. Oxygen uptake and carbon dioxide production were measured at 20, 40and 60 minutes of each exercise work rate. At the same time points, perceptual ratings of RPE, thermal sensation and thermal comfort were taken. A custom patch containing a cotton wool swab to absorb sweat were attached during the rest period and removed following the exercise test for sampling. Blood samples were taken before the heat stress teat and after exercise is complete as indicated by the red arrows. 149

Figure 6-3: Summary of the condensed heat acclimation (Cond HA) protocol undertaken by participants. Two days of heat acclimation (HA) sessions were conducted with each HA session was 75 minutes followed by a 60 minute rest period. Blood samples were collected before and after the first HA session of each day, along with after the second, fourth, sixth and eighth HA sessions (red arrow). Urine samples were collected at the start and end of each day (green arrow), whilst sweat samples (not analysed in this thesis) were collected during the first and last HA session of each day (purple arrow). 150

Figure 6-4: Number of overlapping compounds detected in plasma samples from each separate t-test for heat stress tests conducted before and after the condensed heat acclimation sessions. HST 1= heat stress test one, conducted before heat acclimation sessions. HST2 = heat stress test two, conducted after the heat acclimation sessions. 153

Figure 6-5: Number of overlapping annotated compounds that were detected in urine samples in each separate t-test conducted before and after the condensed heat acclimation sessions. HST 1= heat stress test one, conducted before heat acclimation sessions. HST2 = heat stress test two, conducted after the heat acclimation sessions. 156

List of Tables

Table 1-1: Compound classes for water soluble metabolites as defined by HMDB. Table recreated from[1] 1

Table 1-2: The advantages and disadvantages for each of the commonly used analytical techniques within the metabolomics research field. 6

Table 2-1: serial dilution calculation table with the volumes of sample and solvent required for the serial dilution created as part of the sample preparation for all analytical methods 37

Table 3-1: RSD range in each Dilution for each assay before filtering 59

Table 3-2: A summary of the average full width half max in minutes for each dilution in each assay. Compounds used were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. NEG = negative ion mode, POS = positive ion mode, SD = standard deviation....**Error! Bookmark not defined.**

Table 4-1: The total number of unique metabolites within each dataset and/or library along with the number of overlapping metabolites between datasets for the reversed phase assays from different research groups. For example, the Mayo Clinic and University of Birmingham libraries/datasets have 676 and 536 metabolites, respectively but only eight metabolites which are present in both libraries/datasets. 76

Table 4-2: The total number of unique metabolites within each dataset and/or library along with the number of overlapping metabolites between datasets for the HILIC assays from different research groups. For example, the Mayo Clinic and University of Birmingham Negative ion mode libraries/datasets have 670 and 535 metabolites, respectively but only ten metabolites which are present in both libraries/datasets. 85

Table 4-3: The mean, minimum, maximum and median differences in retention time in both seconds and minutes related to the validation test metabolites used to test the retention time mapping model based off of the relationship between University of Birmingham positive ion mode dataset and the University of Liverpool dataset. All differences were calculated in minutes. 89

Table 4-4: The mean, minimum, maximum and median differences in retention time in both seconds and minutes related to the validation test metabolites used to test the retention time mapping model based off of the relationship between University of Birmingham negative ion mode dataset and the Fiehn MoNA dataset. All differences are calculated in minutes. 91

Table 4-5: Acyl Carnitine retention time model equations and correlation coefficients for each retention time trend based on the number of double bonds within the structure. 98

Table 4-6: The mean, minimum, maximum, and median difference in retention time in minutes between the predicted RT and the mean RT detected for Acyl carnitines with zero or one double bond present in the structure. 99

Table 4-7: Lysophosphocholine retention time model equations and correlation coefficients for each retention time trend based on the number of double bonds within the structure......100
Table 4-8: The mean, minimum, maximum and median difference in retention time between the predicted RT and the mean RT detected for lysophosphocholines with zero or one double bond present in the structure......101
Table 4-9: Ceramides retention time model equations and correlation coefficients for each retention time trend based on the number of double bonds within the structure. d, m, t denote whether the ceramide base is m- mono hydroxy, d- dihydroxy, or t-trihydroxy......104
Table 4-10: The minimum, maximum, mean, and median difference in retention time between the predicted retention time and the detected retention time for ceramides with zero, one or two double bonds. d, m,t denote whether the ceramide base is m- mono hydroxy, d- dihydroxy, or t-trihydroxy.
.....104
Table 4-11: The mean, minimum, maximum, and median difference in retention time between the predicted RT and the mean RT detected for Hexosyl-2-ceramides with a dihydroxyl base and one double bond present in the structure......105
Table 4-12: Trend line equations and correlation coefficient values for the relationships between the total number of carbons and mean retention time for each set of Glycerophosphocholines with a different number of double bonds......107
Table 4-13: The minimum, maximum, mean and median differences in the retention time (RT) predicted by the models and the RT detected in minutes. All metabolites were Glycerophosphocholines with either zero, one, two, three, four, five, or six double bonds present in their structure......108
Table 4-14: The mean retention time (RT) detected, the predicted RT and the difference between the two in minutes and in seconds for the Glycerophosphoethanolamines with one double bond present in the structure......110
Table 4-15: Trend line equations and correlation coefficient values for the relationships between the total number of carbons and mean retention time for each set of Glycerophosphoinositols with a different number of double bonds.....111
Table 4-16: The minimum, maximum, mean, and median differences in the retention time (RT) predicted by the models and the RT detected in minutes. All metabolites were Glycerophosphoinositols with either one, two, or three bonds present in their structure......112
Table 4-17: Trend line equations and correlation coefficient values for the relationships between the total number of carbons and mean retention time for each set of Triacylglycerides with a different number of double bonds......113
Table 4-18: The minimum, maximum, mean, and median differences in the retention time (RT) predicted by the models and the RT detected in minutes. All metabolites were Triacylglycerides with either zero, one, two, three, four, or six double bonds present in their structure......114

Table 5-1: The fold changes for compounds that were found to be statistically significant in the one-way repeated measures ANOVA as well as in the separate t-tests from the MICT protocol. Fold changes had a statistical significance of P<0.05 and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. All changes that were not statistically significant are labelled as NSS. Table Key: OWRMA = one-way repeated measures ANOVA fold change, T1 = at rest, T2 = end of exercise, T3 = active recovery, T4 = passive recovery. 121
Table 5-2: The fold changes for lipid compounds that were found to be statistically significant in the one-way repeated measures ANOVA as well as in the t-tests for Starting Rest (T1) vs Active Recovery (T4) and Starting Rest (T1) and Passive Recovery (T5) time points from the HIIT protocol. Fold changes had a statistical significance of P<0.05 and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. All changes that were not statistically significant are labelled as NSS. 125
Table 5-3: The fold changes for compounds that were found to be statistically significant in t-tests for At Rest (T1) vs Active Recovery (T4) time points from the HIIT protocol. All fold changes had a statistical significance of P<0.05 and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. The glycerophospholipids are denoted by Class (number of carbons: number of double bonds). GPC= Glycerophosphocholines, GPA=Glycerophosphates, GPE/ GPE-NMe/ GPE-NMe2 = Glycerophosphoethanolamines, LysoGPE= Glycerophosphoethanolamines, GPS= Glycerophosphoserines. 126
Table 5-4: The fold changes for lipid compounds that were found to be statistically significant in the one-way repeated measures ANOVA as well as in the t-tests for Starting Rest (T1) vs Active Recovery (T3) and Starting Rest (T1) and Passive Recovery (T4) time points from the SIT protocol. All fold changes had a statistical significance of P<0.05 and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. All changes that were not statistically significant are labelled as NSS. The sphingolipids and glycerophospholipids are denoted by Class (number of carbons: number of double bonds). LysoGPA= LysoGlycerophosphates, LysoGPC= LysoGlycerophosphocholines, GPS= Glycerophosphoserines, SM= Sphingomyelins. .. 128
Table 5-5: The fold changes for lipid compounds that were found to be statistically significant in t-tests for Starting Rest (T1) vs Active Recovery (T4) time points from the SIT protocol. All fold changes had a statistical significance of P<0.05 and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. Fold change decreases are calculated as T4/T1. The sphingolipids, glyceroplipids and glycerophospholipids are denoted by Class (number of carbons: number of double bonds). LysoPA= LysoGlycerophosphates, LysoPC= LysoGlycerophosphocholines, LysoPE= Lyso Glycerophosphoethanolamines, PA=Glycerophosphates, PE/PE-NMe/PE-NMe2= Glycerophosphoethanolamines, PI= Glycerophosphoinositols, CL= Glycerophosphoglycerophosphoglycerols (Cardiolipins), CDP-DG=

cytidine diphosphate diacylglycerol, PS= Glycerophosphoserines, SM= Sphingomyelins, MG= Monoradylglycerolipids, TG= Triradylglycerolipids. 129
Table 5-6: The fold changes for lipid compounds that were found to be statistically significant in t-tests for Starting Rest (T1) vs Passive Recovery (T5) time points from the SIT protocol. All fold changes had a statistical significance of P<0.05 and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. Fold change decreases are calculated as T5/T1. The glyceroplipids and glycerophospholipids are denoted by Class (number of carbons: number of double bonds). LysoPE= Lyso Glycerophosphoethanolamines, GPE/GPE-NMe/GPE-NMe2= Glycerophosphoethanolamines, GPC= Glycerophosphocholines, CL= Glycerophosphoglycerophosphoglycerols (Cardiolipins), TG= Triradylglycerolipids. 131
Table 5-7: The compound that was found to be statistically significant in relative concentration when comparing all three exercise protocols across the at rest time point. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. Fold change increases are highlighted in green. Fold change decreases were calculated as SIT/MICT, MICT/HIIT and SIT/HIIT. 133
Table 5-8: Fold change values for the relative concentration of the compounds that were found to be statistically significant when comparing all three exercise protocols across the end of exercise time point. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. Fold change decreases are highlighted in orange, and any fold change increases are highlighted in green. Fold change decreases are calculated as SIT/MICT, MICT/HIIT and SIT/HIIT. The glycerophospholipids are denoted by Class (number of carbons: number of double bonds). LysoGPE= Lyso Glycerophosphoethanolamines. 133
Table 5-9: Fold change values for the relative concentration of the compounds that were found to be statistically significant when comparing all three exercise protocols across the end of active recovery time point. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. Fold change decreases are highlighted in orange, and any fold change increases are highlighted in green. Fold change decreases are calculated as SIT/MICT, MICT/HIIT and SIT/HIIT. 135
Table 5-10: Fold change values for the relative concentration of the compounds that were found to be statistically significant when comparing all three exercise protocols across the end of passive recovery time point. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. Fold change decreases are highlighted in orange, and any fold change increases are highlighted in green. Fold change decreases are calculated as SIT/MICT, MICT/HIIT and SIT/HIIT. 136
Table 5-11: The compounds that were statistically significantly changing in relative concentration between exercises at the end of exercise time point (time point 2 in MICT and time point 3 in HIIT and SIT), within the female participants in the study. Fold changes were calculated as MICT/SIT,

MICT/HIIT, and SIT/HIIT. Fold change increases are highlighted in green and fold change decreases are highlighted in orange. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. The p-value for the repeated one-way ANOVA was adjusted by applying FDR correction. 137
Table 5-12: The compounds that were statistically significantly changing in relative concentration between exercises at the end of exercise time point (time point 2 in MICT and time point 3 in HIIT and SIT), within the male participants in the study. Fold changes were calculated as MICT/SIT, MICT/HIIT, and SIT/HIIT. Fold change increases are highlighted in green and fold change decreases are highlighted in orange. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. The p-value for the repeated one-way ANOVA was adjusted by applying FDR correction. 137
Table 5-13: The compounds that were statistically significantly changing in relative concentration between exercises at the active recovery time point within the female participants in the study. Fold changes were calculated as MICT/SIT, MICT/HIIT, and SIT/HIIT. Fold change increases are highlighted in green and fold change decreases are highlighted in orange. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. The p-value for the repeated one-way ANOVA was adjusted by applying FDR correction. 138
Table 5-14: The compounds that were statistically significantly changing in relative concentration between exercises at the active recovery time point within the male participants in the study. Fold changes were calculated as MICT/SIT, MICT/HIIT, and SIT/HIIT. Fold change increases are highlighted in green and fold change decreases are highlighted in orange. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. The p-value for the repeated one-way ANOVA was adjusted by applying FDR correction. 138
Table 5-15: The compounds that were statistically significantly changing in relative concentration between exercises at the passive recovery time point within the female participants in the study. Fold changes were calculated as MICT/SIT, MICT/HIIT, and SIT/HIIT. Fold change increases are highlighted in green and fold change decreases are highlighted in orange. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. The p-value for the repeated one-way ANOVA was adjusted by applying FDR correction. 140
Table 5-16: The compounds that were statistically significantly changing in relative concentration between exercises at the passive recovery time point within the male participants in the study. Fold changes were calculated as MICT/SIT, MICT/HIIT, and SIT/HIIT. Fold change increases are highlighted in green and fold change decreases are highlighted in orange. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. The p-value for the repeated one-way ANOVA was adjusted by applying FDR correction. 141
Table 6-1: Timeline for the condensed heat acclimation trial. Key: GXT = Graded Exercise Test; HypTT = Hypoxic Tolerance Test; HST = Heat Stress Test; Cond HA = Heat Acclimation. 147

Table 6-2: One Way Repeated Measures ANOVA (OWRMA) results featuring annotated compounds with statistically significant fold changes across time points pre/post heat stress test 1 (HST1PostvsHST1Pre), heat stress test 1 Post/ heat stress test 2 post (HST1PostvsHST2Post) and pre/post heat stress test 2 HST2PostvsHST2Pre. Positive fold changes are highlighted green, negative fold changes are highlighted orange and any non-statistically significant changes are labelled as NSS. LysoPE = Lysophosphatidylethanolamine. FC= Fold Change. OWRMA= one way repeated measures ANOVA...... 152

Table 6-3: One Way Repeated Measures ANOVA (OWRMA) results for annotated lipid compounds featuring statistically significant fold changes across time points pre/post heat stress test 1 (HST1PostvsHST1Pre), heat stress test 1 Post/ heat stress test 2 post (HST1PostvsHST2Post), pre/post heat stress test 2 HST2PostvsHST2Pre and heat stress test 1 pre/ heat stress test 2 pre (HST1PrevsHST2Pre). All interactions had a statistical significance of P<0.05 and any non-significant interactions are labelled as NSS. Positive fold changes are highlighted green, negative fold changes are highlighted orange and any non-statistically significant changes are labelled as NSS. OWRMA = one-way repeated measures ANOVA. FC= fold change...... 154

Table 6-4: Repeated one-way ANOVA results, comparing fold changes across the compounds that were statistically significant at one or more time point comparison. All interactions had a statistical significance of P<0.05 and any non-significant interactions are labelled as NSS. Fold chance increases in relative concentration are filled in green and any fold change decreases in relative concentration are filled in orange. OWRMA = one way repeated measures ANOVA. 157

Table 6-5: OWRMA results for compounds that were statistically significant during the HTTs conducted before and after heat acclimation sessions. All interactions had a statistical significance of P<0.05 and any non-significant interactions are labelled as NSS. Fold chance increases in relative concentration are filled in green and any fold change decreases in relative concentration are filled in orange. OWRMA= one way repeated measures ANOVA. CHA = condensed heat acclimnation. 159

Table 6-6: Metabolite pathways that were significantly enriched with multiple metabolites detected in urine samples taken across time points during the condensed heat acclimation sessions changing in relative concentration. 160

Table 6-7: Repeated one-way ANOVA results for the compounds detected as significant in the alanine, aspartate, and glutamate metabolism pathway. All compounds in red are the compound annotations featured in the pathway, other annotations are possible annotations for that compound based on the MS1 database searches. All interactions had a statistical significance of P<0.05 and any non-significant interactions are labelled as NSS. Fold chance increases in relative concentration are filled in green and any fold change decreases in relative concentration are filled in orange...... 161

Table 6-8: OWRMA results for lipids that were statistically significant during the condensed heat acclimation sessions conducted. All interactions had a statistical significance of P<0.05 and any non-significant interactions are labelled as NSS. All fold change increases are highlighted green, whilst all

fold change decreases are highlighted orange. OWRMA = one-way repeated measures ANOVA, CHA = Condensed heat acclimation. FC= fold change. 162

List of Abbreviations

AA	Acetic Acid
AC	Alternating Current
AcCA	Acyl Carnitines
ACMG	Analytical And Clinical Metabolomics Group
ACN	Acetonitrile
ADP	Adenosine Diphosphate
AMP	Adenosine Monophosphate
ANOVA	Analysis Of Variance
APCI	Atmospheric Pressure Chemical Ionisation
API	Atmospheric Pressure Ionisation
APnAs	Diadenosine Polyphosphates
APPI	Atmospheric Pressure Photo Ionisation
aqRP	Aqueous Reversed Phase
ATP	Adenosine Triphosphate
BMI	Body Mass Index
CD	Compound Discoverer Tm
CDP-DG	Cytidine Diphosphate Diacylglycerol
CE	Capillary Electrophoresis
CE-MS	Capillary Electrophoresis-Mass Spectrometry
Cer	Ceramides
CID	Collision Induced Dissociation
CL	Glycerophosphoglycerophosphoglycerols
Cond HA	Condensed Heat Acclimation

C-trap	Curved Linear Ion Trap
DC	Direct Current
DDA	Data Dependent Acquisition
DG	Diradylglycerolipids
DMSO	Dimethyl Sulfoxide
Dstl	Defence Science And Technology Laboratory
ESI	Electrospray Ionisation
FA	Formic Acid
FDR	False Discovery Rate
FT	Fourier Transform
FWHM	Full Width At Half Maximum
GC	Gas Chromatography
GC-MS	Gas Chromatography-Mass Spectrometry
GPA	Glycerophosphates
GPC	Glycerophosphocholines
GPE	Glycerophosphoethanolamines
GPE-NMe	Glycerophosphoethanolamines
GPE-NMe2	Glycerophosphoethanolamines
GPI	Glycerophosphoinositols
GPS	Glycerophosphoserines
GXT	Graded Exercise Test
H_2O	Water
HA	Heat Acclimation
HCD	Higher Energy Collision Dissociation
HETP	Height Equivalent To A Theoretical Plate

Hex2Cer	Hexosyl-2-Ceramides
HIIT	High-Intensity Interval Training
HILIC	Hydrophilic Interaction Chromatography
HMDB	Human Metabolome Database
HPLC	High Pressure (Performance) Liquid Chromatography
HST	Heat Stress Test
HTT	Hypoxic Tolerance Test
ID	Internal Diameter
IM	Ion Mobility
IMS-MS	Ion Mobility Spectrometry - Mass Spectrometry
IPA	Isopropyl Alcohol
KNN	K-Nearest Neighbours
LC	Liquid Chromatography
LLE	Liquid-Liquid Extraction
LPC	Lysophosphocholines
LTQ	Linear Trap Quadrupole
LysoGPA	Lysoglycerophosphates
LysoGPC	Lysoglycerophosphocholines
LysoGPE	Lysoglycerophosphoethanolamines
m/z	Mass-To-Charge Ratio
MALDI	Matrix Assisted Laser Desorption Ionisation
MeOH	Methanol
MG	Monoradylglycerolipids
MICT	Moderate-Intensity Continuous Training
mins	Minutes

mm	Millimetres
MoNA	Massbank Of North America
MP	Mobile Phase
MS	Mass Spectrometry/Mass Spectrometer
MS/MS	Tandem Mass Spectrometry
MS1	Full-Scan MS
MS2	Tandem Mass Spectrometry
NCE	Normalised Collision Energy
nm	Nano Meters
NMR	Nuclear Magnetic Resonance
NP	Normal Phase
OWRMA	One-Way Repeated Measures ANOVA
PC	Principal Components
PCA	Principal Components Analysis
PCB	Phenome Centre Birmingham
PFP	Pentafluorophenyl
PLS-DA	Partial Least Squares- Discriminant Analysis
PPE	Protein Precipitation Extraction
ppm	Parts Per Million
psi	Pound-Force Per Square Inch
QC	Quality Control
QIT	Quadrupole Ion Trap
Qs	Quadrupoles
RF	Radio Frequency
RP	Reversed Phase

RSD	Relative Standard Deviation
RT	Retention Time
S/N	Signal To Noise Ratio
SCE	Stepped Collision Energy
SIT	Sprint Interval Training
SLE	Solid Liquid Extraction
SM	Sphingomyelins
SP	Stationary Phase
SPE	Solid Phase Extraction
TA	Targeted Analysis
TAG	Triacylglycerides
TG	Triradylglycerolipids
TOF	Time Of Flight
UHPLC	Ultra-High-Performance Liquid Chromatography
UHPLC-MS	Ultra-High-Performance Liquid Chromatography-Mass Spectrometry
UoB	University Of Birmingham
UoF	University Of Florida
UoL	University Of Liverpool
UoP	University Of Portsmouth
UPLC	Ultra-Performance Liquid Chromatography
UV	Ultraviolet
v/v	Volume In Volume
v/v/v	Volume In Volume In Volume

Chapter one

1 INTRODUCTION

1.1 METABOLOMICS

Metabolites are defined as low molecular weight chemicals, normally under 2000 Da, that are included across major biological roles, such as metabolism, regulation and signalling, as well as in the construction of biological structures. There are many different classes of metabolites; lipids, amino acids, carbohydrates, nucleotides and coenzymes are examples. Metabolites in this thesis are classified using HMDB[1] for water soluble metabolites and LipidMaps[2-4] for lipids. The two biggest separations in metabolites are water soluble and lipids. There are many different classes for water soluble metabolites, which can be found below in Table 1-1.

Table 1-1: Compound classes for water soluble metabolites as defined by HMDB. Table recreated from[1]

Compound Class	**Compound Class**	**Compound Class**
Minerals And Elements	Polyphenols	Sugar Phosphates
Alcohols And Polyols	Dicarboxylic Acids	Glucuronides
Keto Acids	Alkanes And Alkenes	Sugar Phosphates
Carbohydrates	Glycolipids	Biotin And Derivatives
Purines And Purine Derivatives	Hydroxy Acids	Bile Acids
Catecholamines And Derivatives	Prostanoids	Amino Acid Phosphates
Acyl Phosphates	Peptides	Quinones And Derivatives
Amino Ketones	Nucleotides	Pyridoxals And Derivatives
Retinoids	Cyclic Amines	Acyl Glycines
Pterins	Nucleosides	Ketones
Carnitines	Aromatic Acids	Inorganic Ions And Gases
Amino Acids	Amino Alcohols	Polyamines
Porphyrins	Leukotrienes	Alcohol Phosphates
Coenzyme A Derivatives	Indoles And Indole Derivatives	Aldehydes
Pyrimidines And Pyrimidine Derivatives	Tricarboxylic Acids	Cobalamin Derivatives

Lipids are a diverse set of metabolites that are less soluble/insoluble in water. Lipids can be broken down into classes, of structurally similar compounds. In the work presented in this theses, the classifications system used is determined by LipidMaps[4]. LipidMaps is a database with a large variety of information on lipids with eight different classes under which there are many subclasses. The main eight classes are: fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids and polyketides.

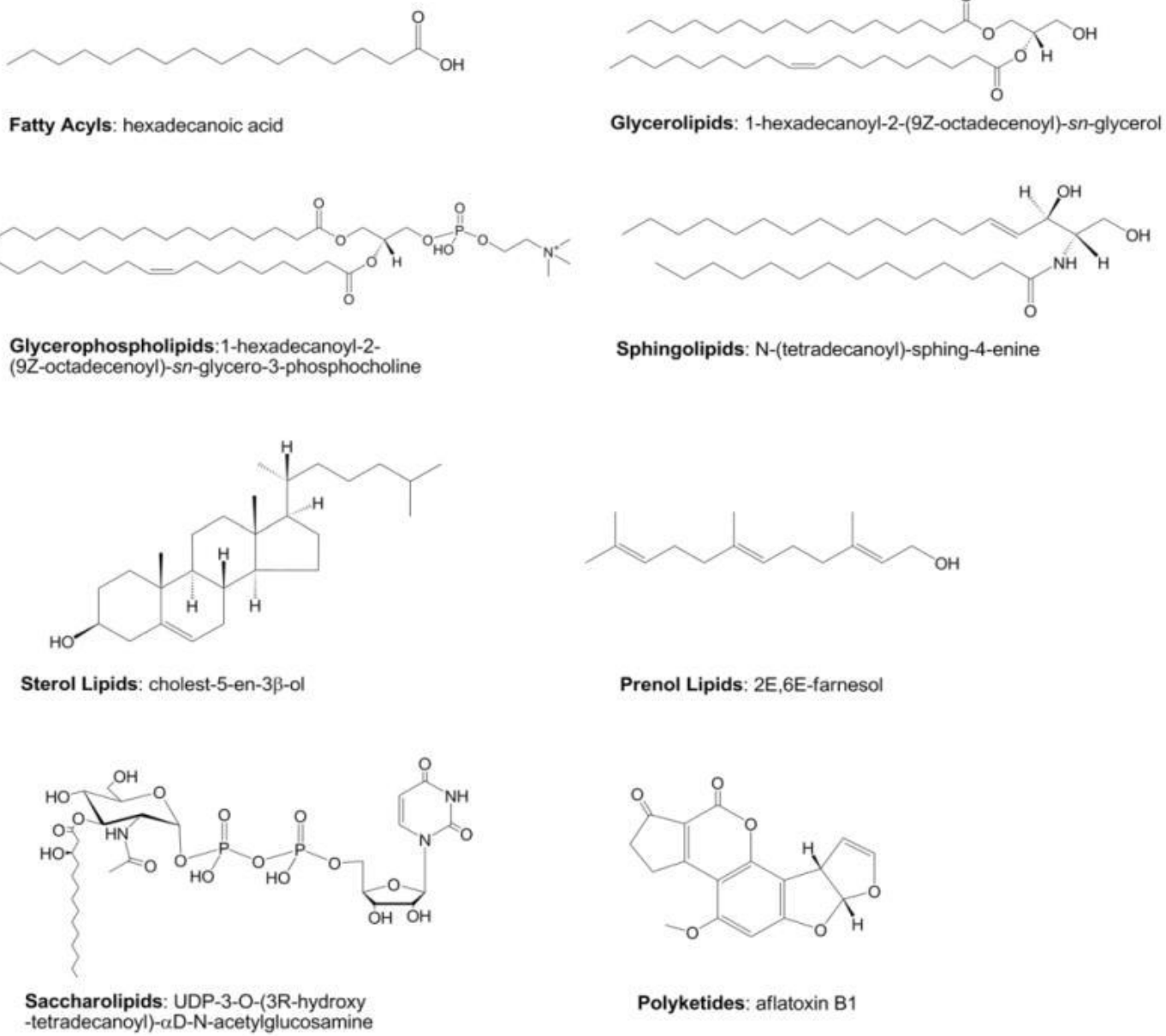


Figure 1-1: Examples of lipids from each lipid class to demonstrate the difference in structure. Figure adapted and recreated from[3]

Fatty acyls are the simplest form of lipid and form the fundamental building block for many of the more complex lipid classes. Fatty acyls are "synthesized by chain elongation of an acetyl-CoA primer with malonyl-CoA (or methylmalonyl-CoA) groups that may contain a cyclic functionality and/or are substituted with heteroatoms."[3] This lipid category includes the compound subclasses of fatty acids, acyl carnitines, fatty esters and fatty amides, fatty alcohols and fatty aldehydes.

Glycerolipids and glycerophospholipids are structurally similar in that they both have a glycerol head group, although the glycerophospholipids have a phosphate group esterified to one of the glycerol hydroxyl groups. The head group on both lipid classes are then esterically joined to fatty acids. Both classes can be mono-, or di-substituted, but it is only the glycerolipids that have the ability to become

tri-substituted. It is the tri-substituted glycerolipids that are the most well-known within the lipid class. Triacylglycerols, as they are commonly known, act as energy storage and therefore form the majority of adipose in mammalian species. Glycerolipids are also used in other bodily processes such as membrane formation, caloric storage, and crucial intracellular signalling processes[5]. Glycerophospholipids are most commonly found in cell membranes[6].

Sphingolipids are well described in the Biochemistry of Lipids, Lipoproteins and Membranes (Fifth Edition) as "…composed of a sphingoid base backbone to which a fatty acid may be attached through an amide bond and a head group at the primary hydroxyl. The head groups range from a simple hydrogen to more complex species, such as the phosphocholine moiety of sphingomyelin and the simple to complex glycans of glycosphingolipids."[4] Sphingolipids are commonly found in cell membranes but are also used in cell signalling processes.

Sterol lipids are the third lipid class that are involved in the processes found within cell membranes. This class of lipids includes the subclasses of sterols, steroids, secosteroids, steroid conjugates, and Bile acids and derivatives. Sterol lipids are well known for their four-ring structure. The most well-known sterol lipid is cholesterol, which is commonly found across all mammals[7].

Prenol lipids are also synthesised from 5-carbon precursors isopentenyl diphosphate and dimethylallyl diphosphate, similarly to sterol lipids, although their structure is very different[3,8].

Metabolite concentrations are dynamic in biological systems including the human body which makes metabolomics incredibly useful for real-time metabolic phenotypic analysis and can provide crucial information about the health or disease state of a human[9]. Metabolomics, the study of the metabolome was first derived in the 1960s[10] and 70s[11]. Although originally it was not known as metabolomics as the term "metabolome" was not around until 1998[12]. The metabolome is defined as a complete set of metabolites within a biological sample. Today, metabolomics research is used in a lot of different biological areas, including but not limited to: biotechnology[13], environment[14], plant[15], toxicology[16] and clinical research[17]. Clinical metabolomics is a translational research area because results have the potential to improve medical diagnostics and treatments for the whole population or for individuals (personalised/stratified medicine). Metabolomics has shown great potential within the medical sector [18-22] but due to it being less advanced compared with proteomics and transcriptomics, a number of studies integrate metabolomic data with data from one of the other -omic methods to allow for a more detailed and accurate phenotypical view at the systems level[23]. Using metabolomics for medical monitoring shows great potential because continuously changing metabolite levels allows the possibility that important molecular changes could be highlighted before physical changes are visible. These continuously changing metabolite levels are a great advantage but also provide a caveat in metabolomics. This is because metabolic activity in samples must be quenched incredibly quickly and frozen rapidly to stop the metabolic processes continuing and therefore later producing false results.

There are three strategies applied to perform metabolomics studies: targeted, semi-targeted and untargeted metabolomics. The three techniques differ in relation to the level of quantification, the number of metabolites detected and the aim of the study [24]. The experimental workflows for untargeted and targeted studies are illustrated in Figure 1-2.

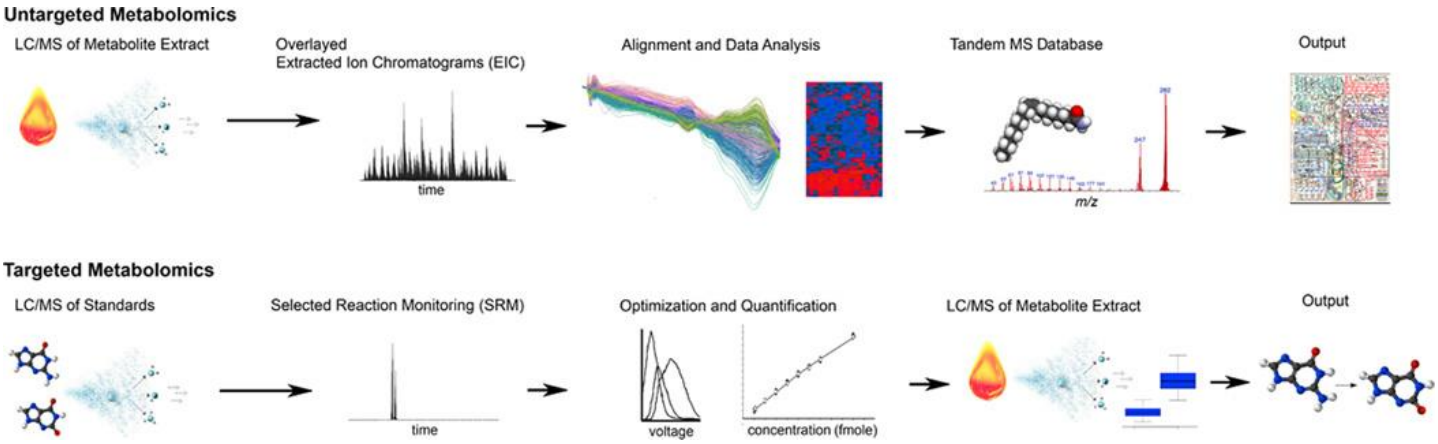


Figure 1-2: A comparison of the two main workflows used in targeted and untargeted metabolomics research. Targeted first analyses metabolite standards to optimise methodology and gain exact spectra before applying these methods to biological samples, resulting in the ability to quantify certain metabolite levels within the biological samples. Untargeted metabolomic methods, in comparison, starts with the biological samples and aims to give a global profile of all the possible metabolites present, only allowing for relative quantification of metabolite concentrations. Figure recreated from [25]

1.1.1 Untargeted

Untargeted metabolomics (also known as non-targeted analysis, non-targeted screening, metabolic phenotyping or metabonomics) is often used in "discovery" based projects [26] as it is a hypothesis-generating approach that has the aim of detecting as many metabolites as possible within a sample in a reproducible way. Often this figure is in the high hundreds. This gives a clearer global view of the metabolic network and multiple different metabolic pathways that may have been affected in each sample. This then allows for the identification of any changes in metabolite concentrations related to the general biological question. Once an untargeted approach has been completed and there are some clear putatively identified metabolites then a more targeted approach can be used to validate any hypothesis generated after the first study is complete. Untargeted metabolomics cannot be conducted with absolute quantification, but relative quantification can be completed where relative differences between peak areas are applied rather than metabolite concentrations. This is because chemical standards are required to construct a calibration curve to quantify metabolites and chemical standards

are also required to identify a metabolite. This is not possible for all metabolites because for many metabolites there is no available chemical standard.

1.1.2 Targeted

Targeted metabolomics is an approach that focuses on only a few known metabolites (approximately 5-15) in a hypothesis-driven approach and can be used for absolute quantification. This is completed using a mixture of methods including using chemical standards to construct calibration curves, isotopically labelled internal standards and rigorously validated methods to ensure highly accurate and precise results. This method is used for diseases that have known biomarkers or if there is a specific metabolic change that is expected or previously described. An example of where this method can be useful is looking at the pharmacokinetics of drug metabolism[25]. Other well-known example applications of targeted metabolomics include diabetes research [27-29] and screening inborn errors of metabolism (IEMs) in new-borns [30,31], drug development[32], biomarker validation research [33], and forensic analysis [34].

1.1.3 Semi-Targeted

Semi-targeted metabolomics is a halfway fit between untargeted and targeted metabolomics. A higher level of identification is achieved for 100s metabolites. This technique aims to identify and quantify low hundreds of metabolites that are known. This type of metabolomics research is different to untargeted metabolomics as it defines the target metabolites before data is collected. This technique also differs from targeted metabolomics because relative quantification is still applied due to either no calibration curves are applied, or one calibration curve is applied for multiple metabolites. An example of where this is used is by Biocrates, who have created several kits that can be used to quantify a range of metabolites at high throughput [35]. A different example of this when Bouatra et al. used the Absolute IDQ p180 Kit to identify metabolites and quantify 127 metabolites in urine samples[36].

1.2 DATA ACQUISITION

For metabolomics research, there are a wide range of different analytical systems that are applied. These include but are not limited to: liquid chromatography – mass spectrometry (LC-MS)[37-39]; gas chromatography – mass spectrometry (GC-MS) [40-45]; ion mobility mass spectrometry - mass spectrometry (IMS-MS) [46-49]; capillary electrophoresis-mass spectrometry (CE-MS) [50-55]; and nuclear magnetic resonance (NMR) spectroscopy [56-59]. All the hyphenated techniques use mass spectrometry in a similar way but use different techniques to separate out analytes within a sample. Gas chromatography separates analytes based on size. Ion mobility separates analytes based on the speed that they travel through a gas at when an electrical field is applied. Capillary electrophoresis uses electric fields and electroosmotic flow to separate analytes by their electrophoretic mobility. NMR applies a magnetic field to excite nuclei into their above energy state. Upon returning to their original energy state, energy

is emitted at frequencies equivalent to radio frequencies. These frequencies can then be used to produce spectra representative of a compounds structure[60]. There are multiple benefits and caveats to using each technique for untargeted metabolomics, most of which can be seen in Table 1-2.

Table 1-2: The advantages and disadvantages for each of the commonly used analytical techniques within the metabolomics research field.

Analytical Method	**Advantages**	**Disadvantages**
LC-MS	• Allows heat sensitive metabolites to be detected that would otherwise degrade if using GC	• Retention times are non-transferable between laboratories • Only relative concentrations can be achieved without standards present. • Isomers cannot be identified without MS/MS
GC-MS	• Both retention time and m/z can be used for compound annotation • Less likely to get matrix effects compared to LC	• Requires extensive sample preparation • Only relative concentrations can be achieved without standards present. • Isomers cannot be identified without MS/MS • Only volatile metabolites of low molecular weight can be detected.
IMS-MS	• Able to identify metabolite isomers[61-63]	• Only relative concentrations can be achieved without standards present.
CE-MS	• Better separation efficiency compared to LC	• Only relative concentrations can be achieved without standards present. • Lower sensitivity compared to LC
NMR	• Requires very little to no sample preparation. • Samples are not destroyed in the process	• No separation required • Lower sensitivity compared to MS based methods

	• Able to provide complete structural data and therefore annotation without the need of a standard. • Technique is quantitative as well as qualitative.	• Only quantative for concentrations upwards of 1µM[64]

Throughout this thesis, the focus will be on UHPLC-MS as it is the most commonly applied analytical technique within metabolomics[65] and was the analytical technique applied for the research presented in this thesis.

1.2.1 Ultra-High-Performance Liquid Chromatography Theory

Column liquid chromatography has been used within science for many years, since the early 1900s when Mikhail S. Tswett pioneered the technique whilst separating leaf pigments from plants[66]. Then in 1941, Martin and Synge devised the theory of using a stationary and a mobile phase to separate analytes [67], this method is still used today. In 1969, Waters produced the first HPLC instrument (ALC100 HPLC). A schematic of an LC instrument can be seen below in Figure 1-3.

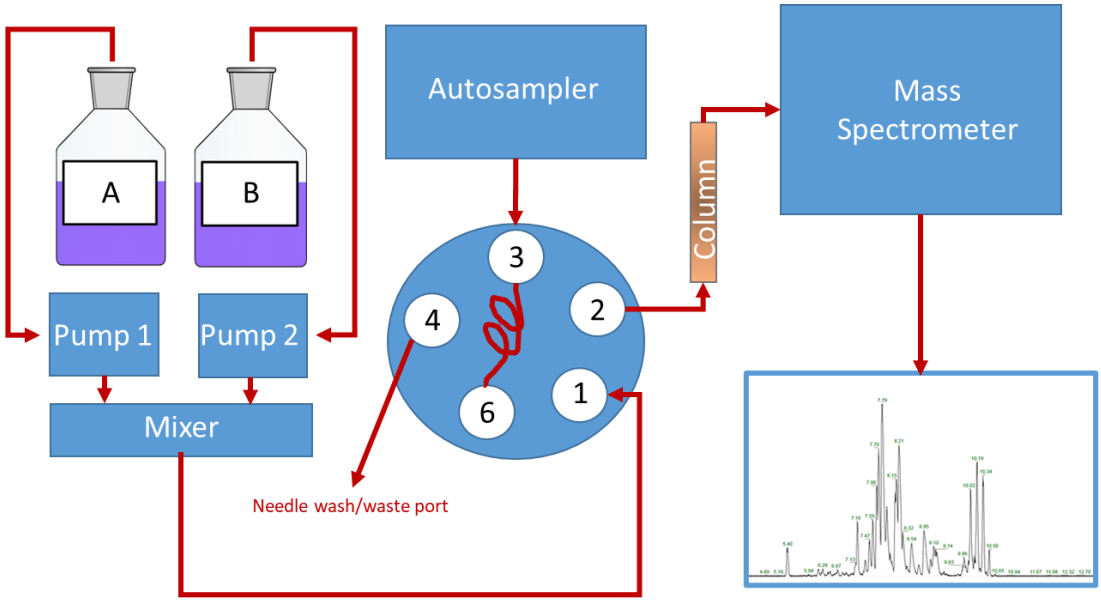


Figure 1-3: A schematic diagram of a liquid chromatography instrument. Featuring mobile phases (bottles labelled A and B), pumps for each mobile phase to deliver each mobile phase at the desired flow rate, a mixer to mix the two mobile phases together into the correct percentages for that part of the elution gradient, auto sampler where all the samples are housed and kept at a stable temperature before they are automatically injected into the system at the start of each sample gradient, column (above in grey) for separation of analytes in sample, mass spectrometer for detection and identification of analytes.

1.2.1.1 Technique theory

The basics of liquid chromatography can be simplified to an equilibrium equation.

$$A_{Mobile} \rightleftharpoons A_{Stationary} \quad \text{Eq. 1}$$

Thus, stating that a portion of the analyte will be in the mobile phase and the other portion in the stationary phase. Each analyte will have a different equilibrium point dependent on the chemical structure, thus allowing them to be separated.

There has been a lot of research in column efficiency [68-70] and how to achieve better chromatographic resolution [71-73] in order to make HPLC more efficient. Chromatographic resolution is calculated using the equation below.[74]

$$R = \frac{1}{4}\sqrt{N} \times \left(\frac{k}{k+1}\right) \times (\alpha - 1) \quad \text{Eq. 2}$$

R= resolution

N= number of theoretical plates

k = retention factor

α= selectivity

This equation can be split into 3, where the first section (red) represents the efficiency, the green section is retention factor (green) and the final section is selectivity (purple).

Retention factor is a measure of retention and can be calculated using the equation below.

$$k' = \frac{t_r - t_0}{t_0} \quad \text{Eq. 3}$$

t_r = retention time of a retained analyte (analyte of interest)

t_0 = retention time of first unretained analyte (i.e., end of dead volume)

A high k value indicates a highly retained analyte. Although, it should be noted though, that this equation only works for isocratic gradients.

Selectivity is simply thought of as the order peaks elute from a column. It can be written as a ratio of the adjusted retention time of two peaks.

$$\alpha = \frac{k'_1}{k'_2} \quad \text{Eq. 4}$$

k'_1 = adjusted retention time of analyte one

k'_2 = adjusted retention time of analyte two

It is known that if the selectivity between two peaks is equal to 1, then no changing in column efficiency will help to separate the two peaks. This is because selectivity is independent of column efficiency. Instead, it is dependent on analytes, mobile phase composition and stationary phase composition.

Column efficiency is a measure of the distribution of peaks, where narrower peaks are more efficient as they allow for more peaks to be fully separated. It is calculated by equation 5.

$$\frac{1}{4}\sqrt{N} \quad \text{Eq. 5}$$

In the above equation, N is equal to the following equation (Eq.6)

$$N = \frac{L}{H} \quad \text{Eq. 6}$$

L being the length of the column and H being the theoretical plate height. Optimising this equation means, making the theoretical plate height smaller. This is where using the Van Deemter equation (below) comes of use.

$$H = A + \frac{B}{u} + [C_s + C_m]u \quad \text{Eq. 7}$$

This equation shows how different parameters of a column can affect band broadening and separation efficiency, and how to achieve optimum flow rates for column choice.

H= (HETP) the plate height. The smaller this number, the more efficient separation will be.

A= Eddy Diffusion

B= Longitudinal Diffusion

U= Linear Velocity

C_s = contribution of the stationary phase to resistance to mass transfer

C_m= contribution of the mobile phase to resistance to mass transfer

Published in 1956 [75], the Van Deemter equation has been used closely in column efficiency for many years and can be seen in graph form in Figure 1-4 below.

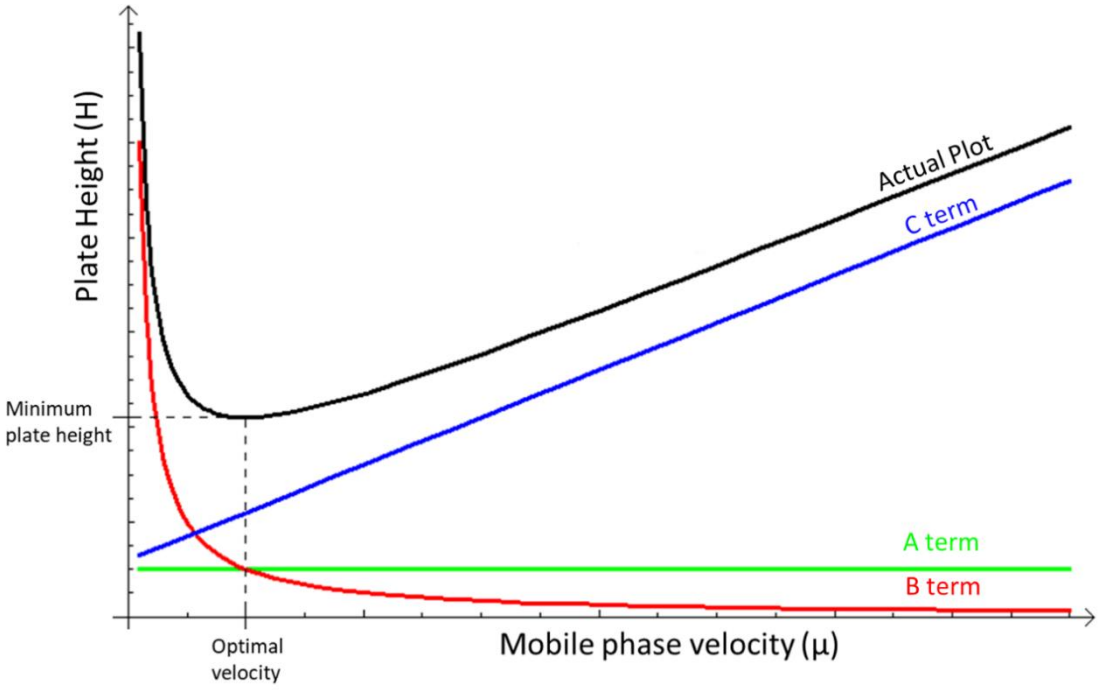


Figure 1-4: The Van Deemter equation in graph form to show how altering the flow rate of a method can affect chromatography resolution.

1.2.1.1.1 Eddy diffusion

Eddy diffusion occurs within a column when analyte molecules take multiple different routes though a column. Eddy diffusion is not affected by flow rate but is affected by the size of the particles used for packing and how well the column is packed.

1.2.1.1.2 Longitudinal Diffusion

Longitudinal diffusion is directly related to the linear velocity used. The lower the velocity selected, the more time that is given for the analyte molecules to diffuse amongst the mobile phase both directions longitudinally. This means that the starting and finishing point of one analyte being detected can be greatly increased, causing band broadening. This factor is greatly affected when the linear velocity selected is below the optimum plate height.

1.2.1.1.3 Linear Velocity

Linear velocity is crucial to stopping band broadening from occurring and plays a vital role in good separation of analytes in a sample. Selecting the optimal velocity for a method is dependent on many factors; the internal diameter (ID) of the column, the length of the column, the size of the stationary phase particles, the ID of the tubing used before and after the column, the viscosity of the mobile phase used and the temperature of the mobile phases. These are all factors as they all affect the backpressure produced. The more backpressure that is produced, the higher the linear velocity required.

1.2.1.1.4 Mass transfer

Mass transfer is broken down into two parts; C_m and C_s. C_m represents the mass transfer effects caused from the mobile phase velocity being slower at the stationary phase surface compared to the centre of the column. This means that analytes closer to the stationary phase are travelling at a different speed to those closer to the centre of column. This can cause band broadening. C_s represents the mass transfer effects from the stationary phase. This effect occurs as analytes interact with the stationary phase and diffuse in and out of the mobile phase. This can cause band broadening when the linear velocity is too high and the analytes interacting more with the stationary phase can get "left behind".

1.2.2 Parameters which can influence chromatographic resolution

When developing and/or optimising a liquid chromatography (LC) method, there are many parameters that can be altered that change how selective/sensitive the method is.

1.2.2.1 Flow Rate

Flow rate is the volume of mobile phase that is pumped through the Ultra-High-Performance Liquid Chromatography (UHPLC) system in a certain time frame, this is normally calculated in millilitres per minute. Flow rate is directly related to linear velocity; therefore, it plays an important role in obtaining good chromatographic resolution. Increasing flow rates decreases longitudinal diffusion but increases mass transfer. Back pressure is often a limiting factor when calculating the optimum flow rate, this is because pressures must be kept below the maximum operating limits for both the UHPLC column and UHPLC system. Mobile phases used must be considered when calculating optimum flow rate, this is because more viscous mobile phases will increase back pressure faster. Applying higher flow rates allows for shorter analysis times[76] but often at the expense of peak resolution.

1.2.2.2 Temperature

Changing the temperature of the mobile phase will affect the separation efficiency of the assay. This can alter by the mobile phase becoming less viscous, thus causing a drop in back pressure and therefore allowing higher flow rates to be used. Heating the mobile phase will also affect the retention of analytes in the column as the level of interactions between analytes and stationary phase will alter, as increases in temperature speed up the separation process. This is important because as temperature is a compromise between heat to allow for a higher flow rate and quicker separations for high throughput analysis, but the temperature applied cannot be too hot as the stationary phase inside will start to degrade, or the mobile phases and/or analytes might become gaseous. It is crucial to maintain a constant column temperature throughout analysis in order to not get discrepancies in resolution and retention times. There are two main ways to heat a column, block heaters and air baths. Block heaters control column temperature by being in direct contact with the column, usually by means of a metal plate, whereas air baths rely on heated air to bring the column up to equilibrium temperate. Block heaters are

much more efficient at heating than air baths, they allow the column to reach equilibrium temperate much quicker and react to temperature changes much quicker than air baths.

1.2.2.3 Column length

Column length affects the separation efficiency in that longer columns allows for a better chromatographic resolution, producing better separation between chromatographic peaks. The disadvantage to using a longer column, is that the analytical run will take longer which can be a limitation when wanting to run high-throughput studies along with the cost increase of running the instrument for longer and the larger amount of consumables required. Common column lengths include but are not limited to 50mm, 100mm, 150mm, and 200mm.

1.2.2.4 Column ID

Column internal diameter (ID) has a large effect on column separation efficiency. In general terms, the smaller the column ID, the better the chromatographic resolution though this is dependent on other parameters also. This is because there is an increase in sensitivity, which is a result of taller and sharper peaks which allows for better resolution of peaks and therefore more analytes to be separated and detected. Decreasing column ID has other benefits such as a dramatic decrease in solvent use, and a lesser requirement for large volumes of sample. One limitation is that analysis times are normally longer which means less samples can be analysed in a high throughput manner. There are many different column IDs available, but the most used in metabolomics is 2.1mm, other UHPLC IDs commonly applied within analytical chemistry include but are not limited to: 1mm and 4.6mm.

1.2.2.5 Particle size

The smaller the particle size of the stationary phase, the higher the chromatographic resolution should be. This is because in theory there should be less eddy diffusion as the size of gaps between particles is reduced. There is also more stationary phase interaction with analytes because of a larger surface area of stationary phase in the column, increasing chromatographic resolution. Due to the increase of mass transfer, flow velocity has different effects on columns of different particle size, with smaller particle sizes able to better accommodate higher flow rates (see Figure 1-5). For UHPLC columns, particles are typically sub 2µm diameter. This differs from HPLC columns which are packed with particles sized of 5µm, 3.5µm or 3µm in diameter.

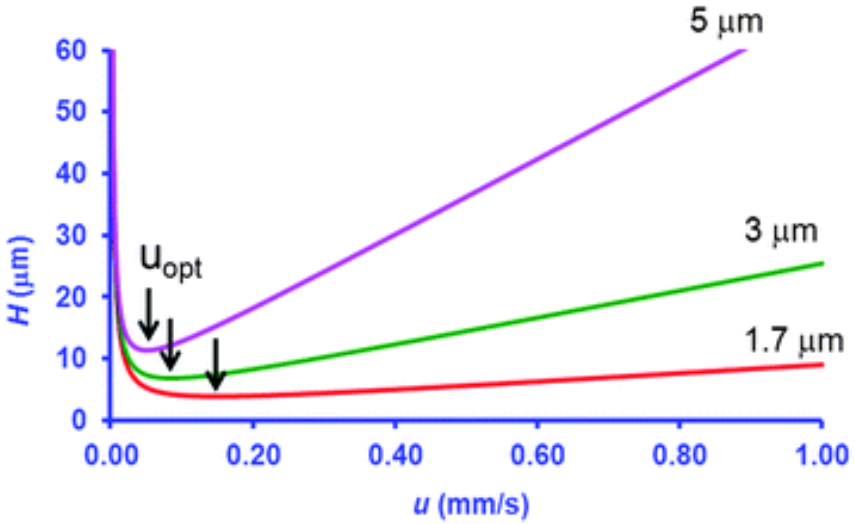


Figure 1-5: Van Deemter graph applied to columns using different particle sizes to show the difference in the optimum flow velocity and how flow rate affects each particle size differently. Graph x-axis is flow velocity in mm/s and the y-axis is height of theoretical plate in µm. the three most common particle sizes used in analytical chemistry are featured on the graph, with the optimum flow velocity for each being defined by µopt. Figure has been recreated from[77]

1.2.2.6 Pore size

Pore size is important to consider when choosing a column for analysis. It is known that smaller pore sizes allow for a greater surface area within a column, therefore, allowing for a higher number of stationary phase interactions. For small molecule interactions, pore sizes of 6 – 12 nm (60-120 Å) are generally used, with larger molecules requiring larger pore sizes for interactions to occur more successfully.

1.2.2.7 Injection volume

Injection volume is an important factor of separation efficiency. Inject too little onto the column and peak areas may not be large enough to distinguish from background noise or mobile phase peaks. Injecting too much onto a column can overload the column and separation will be compromised as peak shapes become less Gaussian and peak broadening occurs from fronting and fattened apexes. Column overloading not only occurs with injection volume but also with the amount of mass injected. If the sample being injected is too concentrated, this can also cause overloading. Mass overloading differs from volume overloading in that the peak shape will remain Gaussian, but the peaks will be broader, and therefore, retention time will shift.

1.2.2.8 Dead volume

Dead volume is the amount of liquid present in the LC system before the sample injected into the system reaches the column.

1.2.2.9 Pump pressures

Pump pressures are affected by multiple parameters within an LC system. More backpressure is applied on the pumps when tubing and columns with smaller IDs are used. Different solvents used for mobile phases also affect backpressure. For the most used solvents, the viscosity increases in this order of ACN>MeOH>H_2O>IPA. The more viscous the solvent, the more backpressure is produced at a specific flow rate and column ID and stationary phase particle size. This needs to be considered when a flow rate is being chosen for an assay. Pumps on UHPLC systems can now withstand much higher pressures than HPLC and LC systems (15,000 psi vs 6,000 psi). This allows for higher flow rates to be used, and for columns with smaller IDs and smaller particles to be used. This in turn has allowed for better separation and resolution of analytes.

1.2.3 Column Chemistry

Liquid chromatography columns were originally developed as gas chromatography columns, principles were taken from gas chromatography and transferred to liquid chromatography in the late 1950s[78]. There are two main types of liquid chromatography, reversed phase and normal phase. Normal phase chromatography uses a non-polar mobile phase (e.g., Hexane or isopropyl alcohol) that is pumped over a polar stationary phase (e.g., Silica) and reversed phase applies a polar mobile phase (e.g., water or MeOH) pumped over a non-polar stationary phase (e.g., non-polar carbon chain (C_{18}) bonded to silica).

There are two main types of columns, open tubular, or packed columns. These depict how the inner area of the column is filled. Open tubular columns have, as expected, an open inner space for uninterrupted mobile phase flow whilst the inner walls are covered with a stationary phase. Packed columns are columns that have a stationary phase that is packed tightly into the column as particles, leaving small gaps which the turbulent flow of mobile phase passes through the column. Packed columns provide a higher chromatographic resolution than open tubular columns, but open tubular columns allow for higher throughput as much larger quantities of mobile phase can be pumped through the liquid chromatography system at the same back pressure as a packed column.

To begin with, most chromatography methods started by applying manual packed columns and using normal phase chromatography. Chromatography columns work by a mobile phase passing over a stationary phase. Any analytes from a sample which are suspended in the mobile phase are then separated by their interaction with the stationary phase. The stronger the interaction between analyte and stationary phase, the longer the analyte is retained on all stationary phase particles as it passes through the column and the longer the analyte is retained on the column.

Over the years many different types of columns have been introduced with the progression of technology to help better separate different classes of analytes. Currently, UHPLC and HPLC are most conducted using packed reversed phase columns.

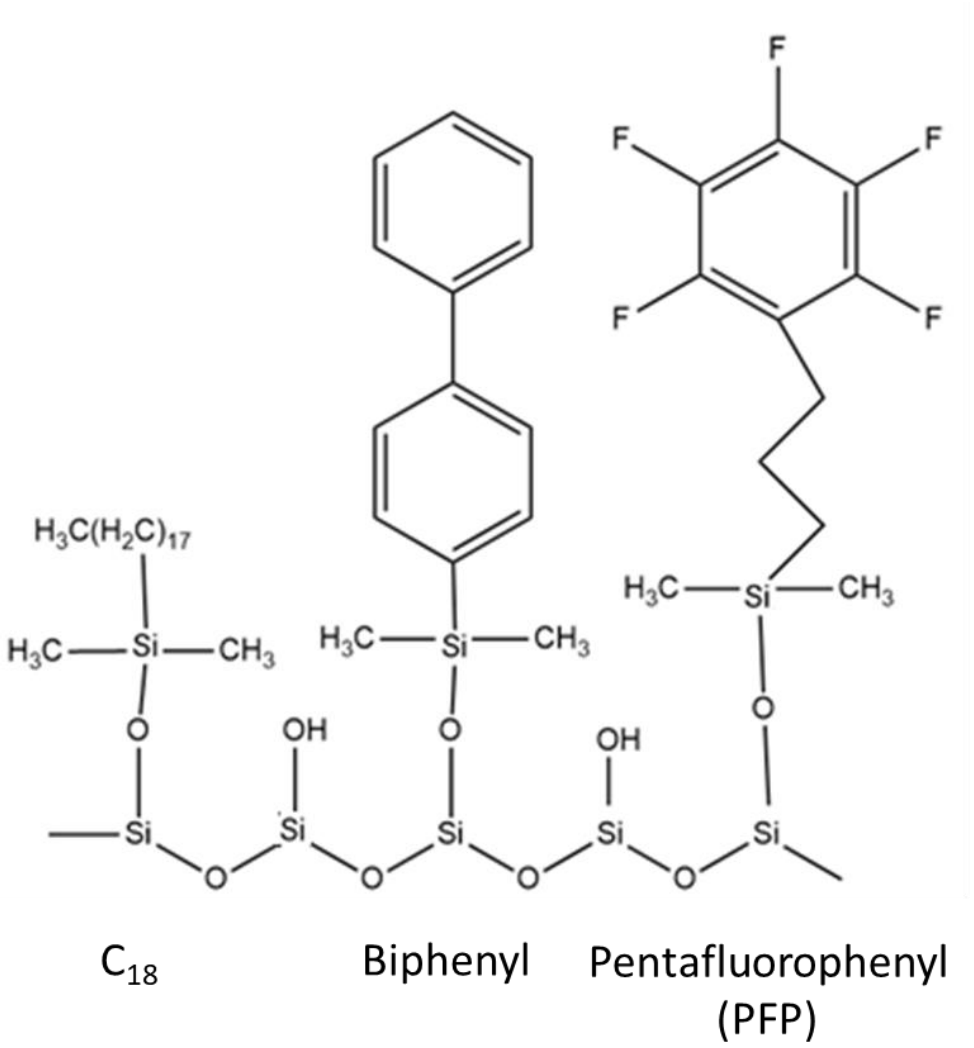


Figure 1-6: Chemical structures for C_{18}, Biphenyl and pentafluorophenyl (PFP) column stationary phases

1.2.3.1 C_{18} columns

C_{18} columns are the most popular column of choice for reversed phase chemistry. They commonly consist of an alkyl chain of 18 carbons attached to modified silica (Figure 1-6). This stationary phase is non-polar and is commonly used with polar mobile phases (e.g., acetonitrile, isopropyl alcohol (IPA), methanol and water). The main method of separation within a C_{18} column is dispersion forces such as Van der Waals or London dispersion forces. These occur when neighbouring non-polar molecules form temporary dipoles due to shifts in electron density. Using this method, lipids can be separated by both carbon chain length and level of unsaturation. C_{30} columns can also be used for lipidomics and are a better technical fit as they are better at separating isomers[79]. As they are a more established column format, C_{18} columns are used more often than C_{30} columns. They are more routinely further developed and are used more routinely for analysis[79-81].

1.2.3.2 Hydrophilic interaction chromatography (HILIC) columns

HILIC columns can be constructed from many different stationary phases, such as modified bare silica, amide bonded phase, amino bonded phase or zwitterionic bonded phase. HILIC methods require a polar stationary phase (e.g., modified bare silica) and a mix of polar and less polar solvents (e.g., water and acetonitrile). Assays using a HILIC column are aimed to retain polar analytes but can also retain some more less polar metabolites (for example polar lipids can be retained and separated based on their polar head group).

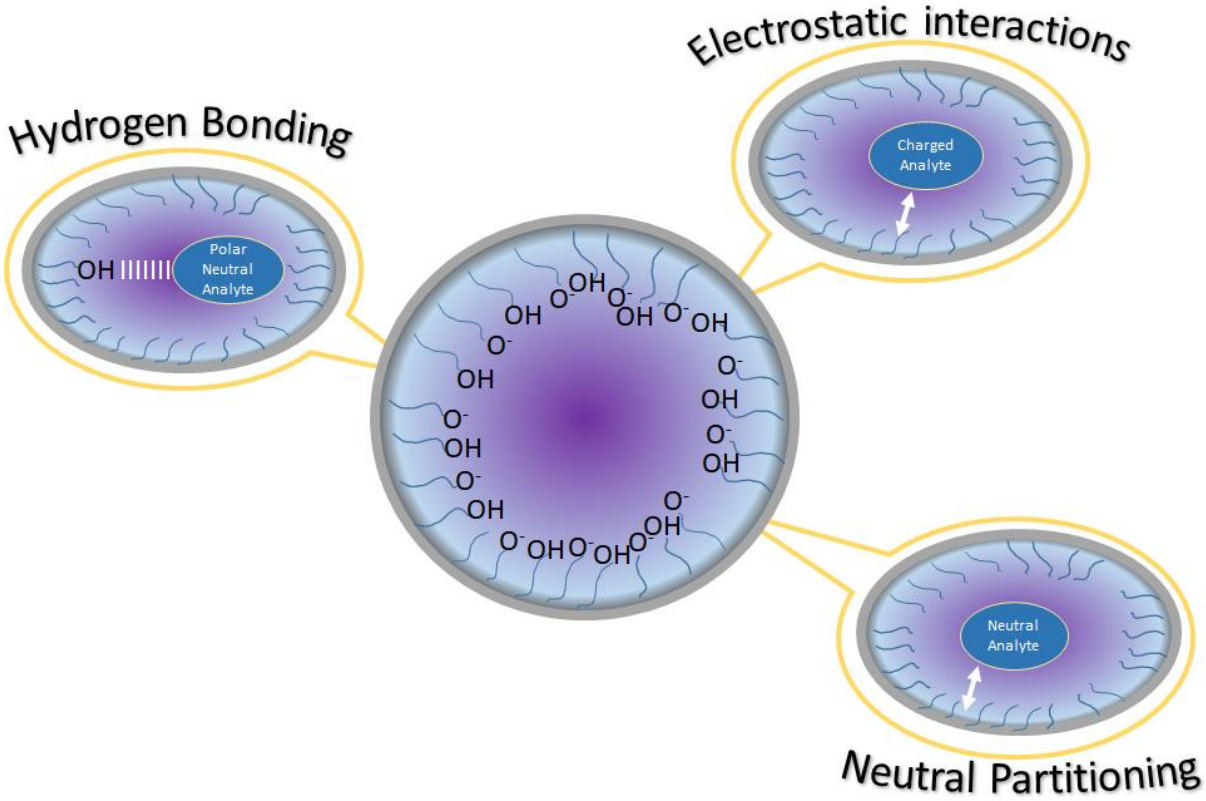


Figure 1-7: A diagram demonstrating the different separation mechanisms used in HILIC separation assays.

The water in the mobile phase interacts with the stationary phase (normally bare silica or silanol groups) to create a small aqueous-rich solvent layer, leaving an aqueous-depleted solvent area. Polar compounds are then drawn to the aqueous-rich part of the column, much like in any other normal phase liquid chromatography[82]. There are multiple different separation interactions that occur within the column (see Figure 1-7), these are hydrogen bonding, electrostatic interations and neutral partitioning. Buffers are very important in HILIC assays as they are used within the mobile phase to stabilise pH and to help retain analytes on the columns.[83]

1.2.3.3 Alternative stationary phases

C_{18} and HILIC columns are most frequently applied in metabolomics[84] but there are alternative stationary phases that when applied can be advantageous for metabolomics research. Biphenyl columns

are an example of this. They are made up of two connected phenyl rings attached to modified silica. Interactions using this column can depend on mobile phases used. If using ACN, this column type can be used for interactions similar to a C_{18} column, due to the triple bonds from ACN having pi-pi interactions with the stationary phase. If using methanol or any alkyl mobile phase, it can be used for pi-pi interactions or induced dipole moments. This is because of the double bonds around the two phenyl rings. This column type has been useful for drug analysis and allows for better separation than a C_{18} for reversed phase methods. Pentafluorophenyl (PFP) columns are another type of alternative stationary phase that are less commonly applied within analytical chemistry but have the potential to be beneficial. The chemical structure of the stationary phase (seen in Figure 1-6) allows for separation of analytes via multiple interaction methods including dipole, pi-pi, charge transfer, and ion-exchange interactions.

1.2.4 Electrospray Ionisation (ESI)

Electrospray ionisation is a soft ionisation technique that has been used in conjunction with mass spectrometry since the 1980s [85]. The technique applies a high voltage to a liquid, which creates a Taylor cone, which then produces an aerosol of charged droplets at the tip. These droplets then undergo desolvation in the high temperature of the enclosed ionisation source until the Rayleigh limit is reached, at which the droplets undergo coulombic repulsion and become a stream of ions as they enter the mass spectrometer capillary. Sheath and sweep gases are applied to aid in focusing the desolvated ions into the capillary entrance of the mass spectrometer. [86,87] See Figure 1-8 below for detail.

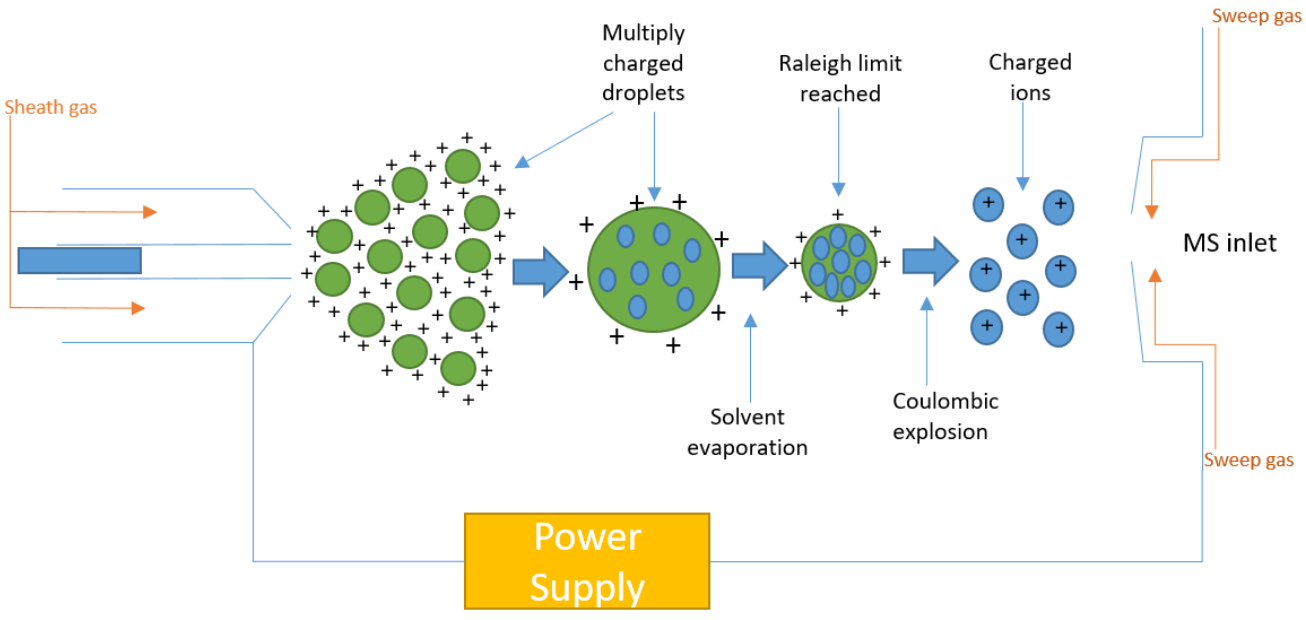


Figure 1-8: A schematic representation of an electrospray ionisation source.

ESI is not the only form of soft ionisation used in conjunction with mass spectrometry, other atmospheric pressure ionisation techniques include atmospheric pressure chemical ionization (APCI) and atmospheric pressure photoionization (APPI). When applying APCI, carrier reagent gas is ionised

by electron ionisation. The ionised carrier gas molecules the react with the gaseous analytes to form the ionisation process. [88] APPI works by ionising gaseous analytes by reacting them with photons by subjecting them to UV light. [89] All three of these methods are classed as soft ionisation techniques as they result in less fragmentation than using hard ionisation techniques that produce a lot of fragmentation, electron impact ionisation being an example of a hard ionisation technique. This is important for metabolomics since the compounds that are to be detected are considered to be small (i.e., low molecular weight) to begin with. It also means that much of the time, the molecular ion will be detected, which is helpful for structure determination and compound annotation/identification.

1.2.5 Mass Spectrometry

1.2.5.1 Theory

Mass Spectrometry is also an important analytical instrument in untargeted metabolomics as it allows the detection of the metabolites. There are many advantages to applying mass spectrometry over other analytical methods (such as NMR or flame ionisation detection) with the most notable advantages being; high sensitivity[23] and ability to detect a large number of metabolites at one time[90-92]. Mass spectrometry works by measuring the mass-to-charge ratio (*m/z*) of ions. This is then recorded by creation of a spectrum of the mass to charge ratio (*m/z*) on the x axis and the intensity of that ion detected on the y axis. It was first invented by J.J. Thomson in 1897[93] but has been rapidly developing in speed, sensitivity and selectivity ever since[94]. An example schematic of a mass spectrometer can be seen below (Figure1-9).

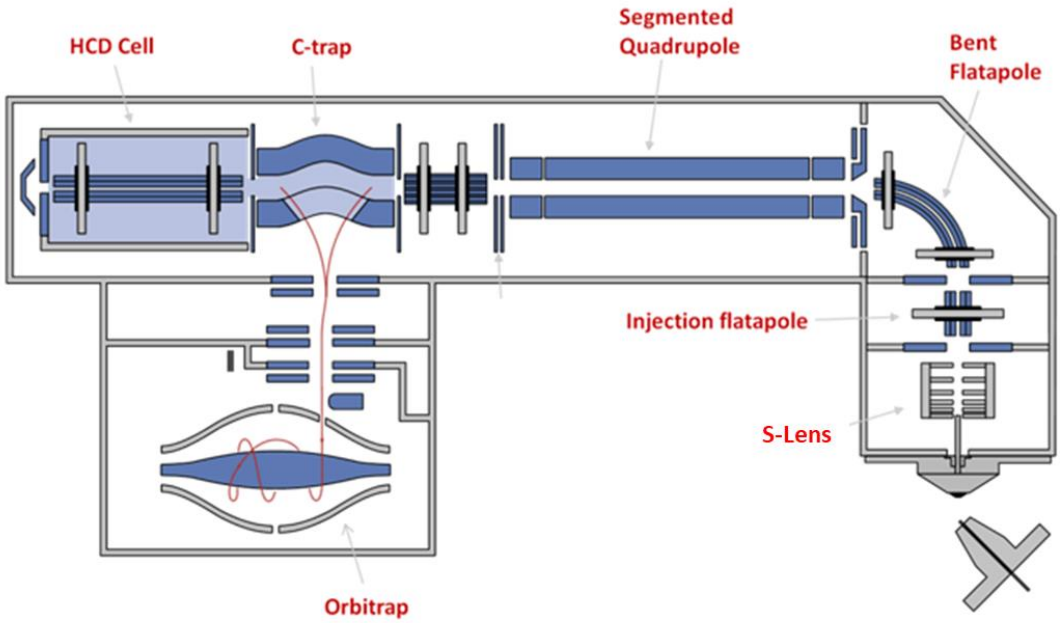


Figure 1-9: Adapted schematic of a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer from www.thermofisher.com

There are multiple different commercially available types of mass spectrometer with different types of mass analysers. These all vary greatly in how they operate; thus, this thesis will focus on the Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer as this was the mass spectrometer applied to the research presented in this thesis.

Ions first enter the mass spectrometer through the source cone, they next enter the S-lens where a radio frequency (RF) is applied to focus ions entering the mass spectrometer into a tight beam. The S-lens is constructed of series of flat electrode rings with wider beam guide holes at the start, getting smaller towards the rear of the lens. This, in conjunction with smaller gaps between rings at the start and larger gaps towards the rear, helps to capture as many ions as possible. Ions are moved and focused through the lens by odd and even electrode rings have alternating RF alternating current (AC) voltages applied, this technique forgoes the need to apply a direct current (DC) current to move the ions along the instrument. Next, the ions reach the injection flatapole and the bent flatapole. These are used to prefilter ions and remove neutrals, reducing noise whilst moving the ions towards the segmented quadrupole mass analyser. The segmented quadrupole mass analyser is constructed of four hyperbolic-profile round rods shaped into a square, the rods are separated into pairs with opposite rods being in a pair. These pairs have RF and DC voltages applied, each pair of rods has the same frequency and amplitudes applied, but opposite signs. The RF/DC ratios applied allow only select ion frequencies to have a stable flight path through the quadrupole. This increases ion transmission. Ions then move on to the C-Trap, where the ions are collected using RF waveforms before being "cooled" with N_2, which, when collided with ions absorbs their kinetic energy. After "cooling", the RF is rapidly reduced, and a DC voltage applied to release the ions and "push" them into the Orbitrap mass analyser.

Orbitrap mass analysers consist of two outer electrodes, and an inner spindle electrode (see Figure 1-10). Ions enter the Orbitrap and are "squeezed". By ramping the DC voltage on the inner electrode, the ions' radial and axial trajectory is compressed, thus decreasing the number of ions lost to the outer electrode. Ions then oscillate around and along the inner electrode. Different ions oscillate at different speeds and different distances from the inner electrode, as well as moving different distances along the inner electrode. These differences allow for detection of different *m/z* ions at higher resolution (specifically *z* direction).

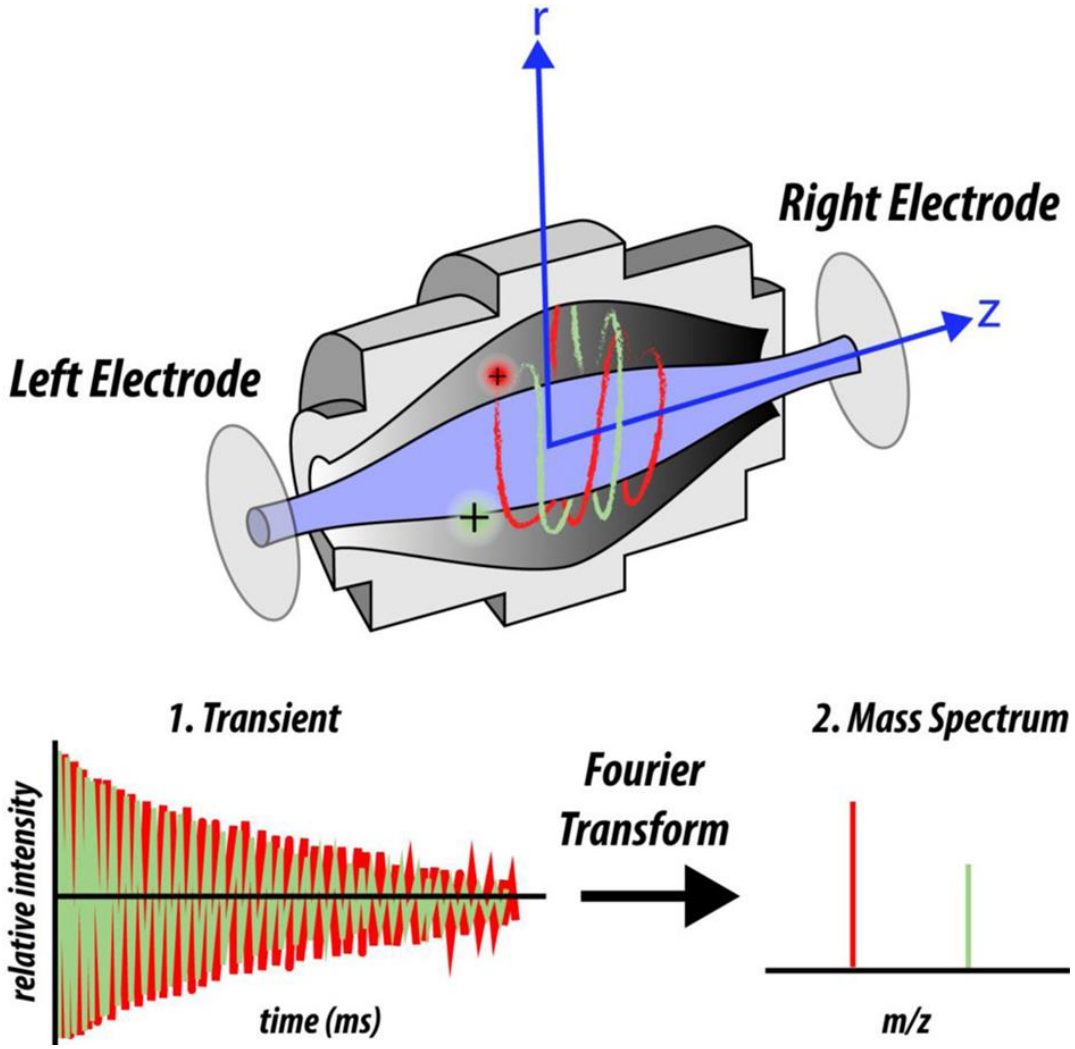


Figure 1-10: Schematic diagram to explain how ions are detected using an Orbitrap mass analyser, Figure reproduced from [95]

Mass spectrometry is useful for attaining a broad overview of what compounds may be present in a sample, but for a more detailed approach tandem mass spectrometry is applied. Within untargeted metabolomics, mass spectrometry is performed as both singular (MS1) and tandem mass spectrometry (MS/MS or MS^2) approach. MS/MS is when a mass/charge (*m/z*) ratio is recorded for each mass, and then the masses are fragmented further via a HCD cell or by using collision induced dissociation (CID), before being injected back into the Orbitrap to be separated and detected again. Precursor ions enter the HCD cell and are subjected to collisions in a beam type collision induced dissociation. The HCD cell is at a higher pressure due to the RF voltages used to trap the ions and fragments. Due to the higher pressure of the cell, the precursor ions and fragments are all subjected to higher energies and are likely to fragment further. After fragmentation is complete, the fragments are then transferred back to the C-trap and onto the Orbitrap for mass analysation and detection. HCD and CID are similar techniques and are both commonly used within MS/MS assays. CID is typically carried out using low kinetic energies. Precursor ions are accelerated to increase kinetic energy and then collided with neutral molecules. During this collision, the kinetic energy is transferred into internal energy, which in turn causes bonds

to break and fragmentation to occur. An advantage to using HCD is that the 1/3 rule of CID does not exist. The 1/3 rule of CID is that molecules that are less than M/3 in weight are not detected. This is because the electrical potential that is used to move the ions and fragments does not affect the fragments that differ that much on weight compared to the original precursor ion. This gives more detail about each different ion within the mass spectrometer and helps to identify between isomers and confirmers, whilst later allowing for better annotations of metabolites to be conducted.

1.3 DATA PROCESSING

After the data has been collected using UHPLC-MS, the data must undergo processing before it can be statistically analysed. Data analysis produces a large challenge for untargeted metabolomics studies. This is because untargeted metabolomics produces a large volume of data as there is an extremely large number of metabolite features (approx. 14,000+) which comprise information about metabolites as well as background information and/or noise. Although many metabolite features may be detected, there is not always that number of metabolites present in the sample. This is because metabolites can be detected at the same retention time with different *m/z* values. Along with these, different adducts and isomers can also be detected. This means that for each metabolite, multiple features can easily be seen. For 10,000 features, there could be 1000-2000 metabolites within the sample[96].

There are many software packages (both open and commercially available) that can be used for data processing, such as MZmine[97] or MassLynx™[98], but two of the most popular are XCMS[99] and Compound Discoverer™[100]. Below are the two workflows followed for data processing and analysis when using XCMS for data pre-processing or Compound Discoverer ™ for data pre-processing and metabolite annotation (Figure 1-11).

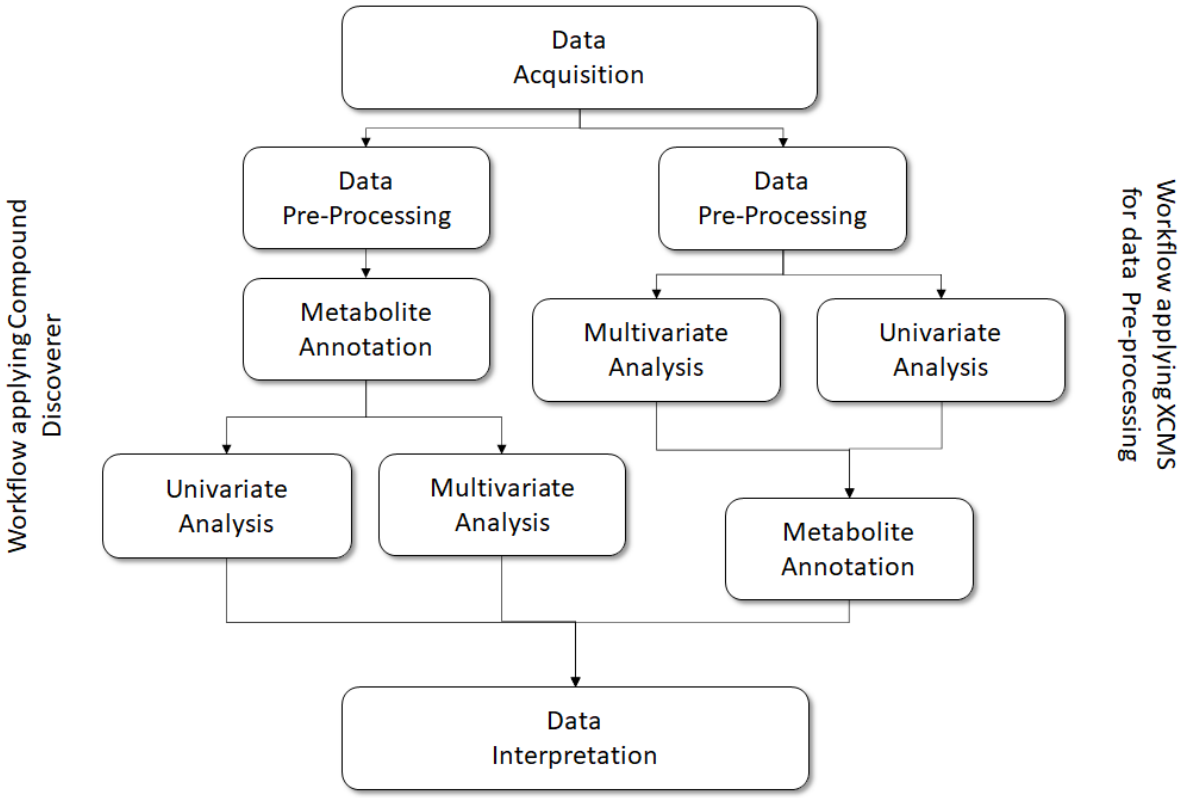


Figure 1-11: Two common data processing workflows used within untargeted metabolomics. The left-hand side shows the workflow when using Compound Discoverer ™ workflow, and the right-hand side shows the workflow when using XCMS.

1.3.1 Data Pre-Processing

Although all instrument files contain m/z, RT and intensity data, the file format is dependent on instrument manufacturer. For example, Thermo Fisher Scientific software instruments produce .RAW files. Depending on what software is being used for data processing, some pre-processing of data files is often required. Where the commercially available software was created by the same company as the instrument manufacturer, the original file format can normally be used, but for open-source software packages, files are commonly required to be converted into specific formats such as .mzml files. One software used to do this is MSConvert by ProteoWizard[101]. It should be at this point that any clear outliers with low file size should be removed from the files which will be processed. After this, data can be pre-processed. These steps include peak picking, retention time alignment and grouping of compounds before integrating all of the data for each sample into one large data matrix. Each of these steps has a set of parameters and algorithms that are followed to reveal a set of peaks that are detected across a chromatographic run. All work presented this thesis applied the Thermo Fisher Scientific closed source software package called Compound Discoverer ™. Compound Discoverer ™ allows for the raw data pre-processing steps along with metabolite annotation and integration into a data matrix to be conducted together at once, which means that less reformatting is required, saving time and potential

mistakes. Workflows can be uniquely customised to each user's needs, allowing a lot of personalisation and in-depth analysis to be performed. Compound Discoverer ™ is not the only software that allows for personalisation, as XCMS is allows for full optimization for different datasets. Opportunity for optimisation with different datasets is important because change in analytical methods brings change to m/z ranges applied along with changes to peak widths, retention times, mass accuracies etc. These changes mean that if peak picking parameters are not optimised, biological information will be lost or wrongly included the final data matrix.

1.4 METABOLITE ANNOTATION AND IDENTIFICATION

In metabolomics data analysis there are currently 4 levels of confidence when identifying metabolites within a sample[102,103]. These levels are defined as the following:

1. Identified compounds
2. Putatively annotated compounds
3. Putatively characterized compound classes
4. Unknown compounds

A compound that had no annotation would still be kept and classified under level 4. A compound classed under level 3 would be based on *m/z* only and characterised from an open database on spectral similarity. A compound annotated to confidence level 2 would require a structural match from an online spectral library (such as mzCloud) and a retention time (RT)[103]. To be identified to level 1 confidence would require two independent data matches (such as RT and *m/z* or RT and NMR spectrum) with matching data from standards. Data from standards must also be collected within the same laboratory under matching experimental conditions.

Meeting confidence levels 2 and 1 are difficult since databases used for metabolite annotation are very small and undistinguished compared to other -omic methods (such as proteomics). This is because the research area is less established, along with the fact that many metabolites are yet to be discovered. Another difficulty is that many metabolites are endogenous and/or cannot be made as standards, therefore any level 1 annotations are very advantageous and highly sought-after. One way to overcome this is to make more use of RT data. Currently RT libraries are limited when compared to mass spectral libraries and are non-transferable between instruments. Using RT data has some complications to overcome, mainly in that RT is easily altered between laboratories as methods in metabolomics are still far from standardised and vary greatly between research groups (from column or instrument manufacturer to elution gradient applied). Due to this, RT libraries are mainly composed in house for each research group and cannot be shared outside of the lab. Current research into ways to overcome this have been to try and predict retention time[104-109], this has been done with varying levels of success.

Another alternative would be to try to map retention times[110-112], this method is a lot less researched and will be further explored later in chapter three of this thesis.

Currently, identification of lipids in metabolomics is difficult. This is for multiple reasons including:

- Their complex structures, which means that with MS1 data would suggest multiple different lipid annotations for the same metabolite
- Many lipid intermediates are hard to make into analytical standards which means that matching LC-MS/MS data across for a high confidence annotation incredibly difficult

One way to overcome this would be to make an in-silico RT library and to use this to match across any lipids detected. While MS/MS in-silico libraries are not new and have been used previously [113-115], RT in-silico libraries have been investigated less so[116,117]. This is because the number of differing parameters makes RT much harder to predict.

Compound Discoverer ™ uses several steps for annotation of metabolites. Workflows predict compositions of each of the features, along with searching online databases (such as mzCloud) for MS/MS matches to the peaks detected in pre-processing. Compound Discoverer ™ also allows the user to submit a mass library to the software, which is useful for matching data to standards previously run on the same instrument with ease. Compound Discoverer ™ also predicts the compound compositions for each of the features, the software will then search online databases such as HMDB, LipidMaps and Chemspider for possible matches. After all this is completed, it will assign the most likely best hit for that compound. This is done in a priority order set by the user. If applying an open-source software, such as XCMS, metabolite annotation needs to be done afterwards using separate software. There are two main alternatives that are used within metabolomics, PutmedID, and BEAMS.

1.5 UNIVARIATE AND MULTIVARIATE ANALYSIS

A mixture of univariate and multivariate analysis is used so that as many variables are discovered as possible. This is because some variables may be dependent on others, whilst several variables may be independent. Before any statistical analysis is conducted all datasets go through quality assurance to make sure that the most reproducible results are being used for statistical analysis.

1.5.1 Missing value imputation and data normalisation

Before completing multivariate analysis, the data must first undergo missing value imputation. To do this, K-nearest neighbours algorithm (KNN) is applied. KNN is a non-parametric algorithm that classifies an object based on K number of neighbours. K is a small positive integer. Another form of missing value imputation that is commonly applied is random forest. Random forest is a useful for missing value imputation as it does not assume normality or require specification of parametric

models.[118] . After missing value imputation, data normalisation is carried out in order to create a more robust relationship between variables for multivariate analysis and regression tools.

Data normalisation makes sure the data from all samples is on a similar intensity scale to allow for better interpretation by reducing any systematic variation[119]. Probabilistic quotient normalisation (PQN) normalisation is a commonly applied normalisation method in metabolomics. This method of normalisation works by "calculating the most probable dilution factor by looking at the distribution of the quotients of the amplitudes of a test spectrum by those of a reference spectrum"[120,121].

1.5.2 Multivariate Analysis

1.5.2.1 Principal Component Analysis

Principal Component Analysis (PCA) is an unsupervised multivariate analysis technique (meaning that metadata is not included in model construction but is included afterwards to allow for data interpretation) that looks at orthogonal variation using eigenvectors and eigenvalues to determine where the most variation lies within the data. Each Principal Component (PC) is constructed with an eigenvector (the direction through the data) and an eigenvalue (the amount of variation in that direction), each subsequent principal component will feature less and less variation from the data. Several different plots from PCA can be created, but the two most used are the scores plot and the loadings plot. The scores plot can be used to see large differences in variation between samples and is often used to single out outliers within a sample class. The loadings plot shows the relationship between each variable and a PC, this is useful to show if certain variables are influencing the PC or not.

1.5.2.2 Partial Least Squares- Discriminant Analysis

Partial Least Squares- Discriminant Analysis (PLS-DA) is a supervised multivariate regression technique, meaning that sample metadata is included. It is similar to PCA in that it uses vectors to find the maximum variability between variables, although it projects the X and Y matrices (projected and observed data) onto new spaces. PLS-DA is different to PLS in that it combines PLS, and DA. DA is a technique used to define whether there are significant differences between predictor variable groups.

1.5.3 Univariate analysis

1.5.3.1 t-test

The t-test is a commonly used statistical test that defines whether there is a statistical difference between two groups. This is done by using the mean, standard deviation, and number of values from each group. There are two types of t-tests that are commonly applied: parametric and non-parametric. Mann Whitney tests (non-parametric t-tests) are applied when the two groups are independent of each other, whereas parametric t-tests are applied when the samples are related. Equations for both can be found below.

Mann Whitney $$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s(\frac{1}{n_1} + \frac{1}{n_2}))}}$$ Eq. 10

Parametric $$t = \frac{m}{s/\sqrt{n}}$$ Eq.11

$\bar{x}_1$ = mean one

$\bar{x}_2$ = mean two

S = standard deviation

m = mean

n = sample size

n_1= sample size one

n_2= sample size two

1.5.3.2 Analysis of Variance ANOVA

Analysis of variance (ANOVA) is a technique that looks at different group means to see if any are significantly different. It works similar to a T-test in that it is testing a null hypothesis (that all group means are equal). If two or more groups are significantly different, then a statistically significant result is returned and thus an alternative hypothesis would be accepted. One disadvantage to using ANOVA is that if significant results are returned, which groups are the ones causing the significant difference cannot be uniquely identified, to find this out ad hoc t tests must be conducted.

1.5.3.3 Repeated One Way Analysis of Variance ANOVA

Repeated one way analysis of variance (ANOVA) or, as it is sometimes known, within-subjects ANOVA is a multivariate analysis that is applied to compare three or more groups where the subjects are the same in all three groups to see if there are any statistically significant differences between the groups. To apply the statistical technique, there are a number of assumptions that are made. These are:

1. The data in the groups you are comparing is continuous
2. The data is normally distributed
3. Each participant is present at all the time points
4. The data is independent
5. The variance between all possible within-subject comparisons is equal.

1.5.3.4 Multiple Testing

Since there are multiple comparisons on the same data, false positives are more likely and the type 1 error rate of that statistical analysis increases. The likelihood of producing false positives can be calculated using the family wise error rate equation (FWER).

$$FWER = 1 - (1 - \alpha)^m \qquad \text{Eq. 7}$$

m is equal to the number of comparisons being performed

α is equal to the alpha level for an individual test (most commonly 0.05)

Normally when performing statistical comparisons with an alpha level of 0.05 we are accepting that 5% of statistically significant differences may be false positives. This number rises rapidly when the number of comparisons made increases, with 10 comparisons giving a FWER of 40% and 30 comparisons giving a FWER of 79%. One way to overcome this is to use a data correction method. The two most commonly used for repeated one way ANOVA are false discovery rate (FDR or Benjamini-Hochberg method)[122] and Bonferroni correction[123]. FDR is conducted by ranking all P values from smallest to largest, we then rank those P values (1-x). For the largest ranked P value, the adjusted P value is the same as the original P value. All adjusted P values below that take on the smallest of two options; the previous adjusted P value or the outcome of the below calculation.

$$adj\ P\ value = current\ P\ value \times \left(\frac{total\ number\ of\ P\ values}{P\ value\ rank}\right) \qquad \text{Eq. 8}$$

Current P value multiplied by the total number of P values divided by the p value rank. Bonferroni correction is conducted differently, instead of adjusting P values, the alpha level is corrected by using the following equation.

$$FWER = 1 - (1 - \frac{\alpha}{\mathcal{K}})^m \qquad \text{Eq. 9}$$

m is equal to the number of comparisons being performed

α is equal to the alpha level for an individual test (most commonly 0.05)

K is equal to the number of comparisons being performed

Applying this correction keeps the likelihood of false positives to ~5% and any statistically significant comparisons will be ones with a P value smaller than the newly corrected α value. One of the disadvantages to using the Bonferroni correction method is that whilst it decreases the type 1 error rate, it increases the type 2 error rate (likelihood of false negatives). Due to this, when conducting lots of comparisons, using the FDR method is much better as it avoids such a large increase in type 2 error.

1.6 SAMPLE TYPES

The field of metabolomics is fast becoming a well-researched area and the overall aim to annotate the human metabolome using multiple different sample types is progressing steadily. To do this, multiple sample types and analysis techniques have been used as no single technique alone would ever be able to quantify an entire sample metabolome, let alone the complete human metabolome.

1.6.1 Blood and Blood Fractions

Metabolomics is most commonly conducted on serum and plasma but can also be applied to whole blood. Serum has a similar composition to plasma but doesn't contain any clotting factors. Plasma is an extensively researched sample type in comparison to other biofluids. It is collected as whole blood and separated using an anticoagulant (heparin, EDTA, citrate). For metabolomics the most commonly applied anticoagulant is heparin as it does not cause matrix effects, unlike EDTA. Matrix effects occur when a component of a sample that is not one of the desired analytes being analysed affects the analytical technique [124]. An example of this is ion suppression of co-eluting compounds, where the detection of an analyte of interest is supressed due to an alternative sample component co-eluting and ionising more efficiently/at a higher concentration than the analyte of interest [125]. Citrate is never used as it is a metabolite and would therefore interfere with the assay. Plasma and serum are used frequently for samples as taking blood is a simple procedure and both are very complex biofluids with very similar metabolomes[126,127]. Currently (as of 02/08/2021) there are 3083 metabolites that have been detected and quantified in the Human Metabolome Database (HMDB) for blood. [128] The HMDB is a renowned database used by many metabolomics researchers for research on the human metabolome. It is already greatly detailed with it constantly expanding as further research is conducted and analytical instruments become more and more sensitive.

1.6.2 Urine

Urine is an excellent sample type to use for research as it is available in large quantities and is easier to obtain than plasma and serum. This is because it is a non-invasive method and therefore more likely to be volunteered than blood samples. The simple sample collection also means that it can be used simply in large scale studies, and away from clinical environments. Urine is a valuable biofluid for metabolomics as the body uses a large quantity of urine to remove toxins and waste. Urine is still chemically complex, containing a multiple of different compounds such as "urea, inorganic salts, creatinine, ammonia, organic acids" [36] and carbohydrates. It is mostly free of proteins which makes sample preparation simpler than other biofluids. These advantages make it a great sample for large cohort studies [129,130]. One limitation of using it for large cohort studies is the variability between samples produced by one subject as well as the variability of samples between subjects. High amounts of salts within samples along with sample concentration differences also pose difficulties for analysis. As concentration of the sample, varies largely on how hydrated each participant is, correction for

differences is required and can either be applied pre-or post-analysis. Pre-analysis methods include: individual dilution to a uniform creatinine concentration[131], uniform creatinine by injection volume adjustment[132], normalisation of samples to equal osmolality concentration[133]. One caveat that using the creatinine concentration as a constant is that creatinine concentrations can differ by a factor of 20-100[134]. The most common post-analysis methods include normalisation of the datasets to osmolality or creatinine concentration or to mass spectrum total signal (MSTS) [135]. At the time of writing (as of 02/08/2021), there are 1763 metabolites that have been detected and quantified in human urine in the Human Metabolome Database (HMDB) [128]. There is a dedicated database within the HMDB that focuses on just the urine [36].

1.6.3 Tears

There is an extensive range of tear types that can be found, based on how they are produced. These can be categorized into basal, reflex, and psychological tears. Psychological tears are caused by emotions and physical pain; reflex tears are stimulated from actions (yawning, irritations etc.) and basal tears are the natural secretion that covers the eye and is not stimulated by anything[136]. For metabolite composition in tear samples during human health research basal tears will be required because the metabolite levels in these tears will be closest to the natural levels[137] within the body and change diurnally rather than caused by a stimulus. Due to the nature of the research, the type of tear used cannot be interchanged as the metabolic makeup of the tears would be different. This could end up leading to false positives or false negatives in later research.

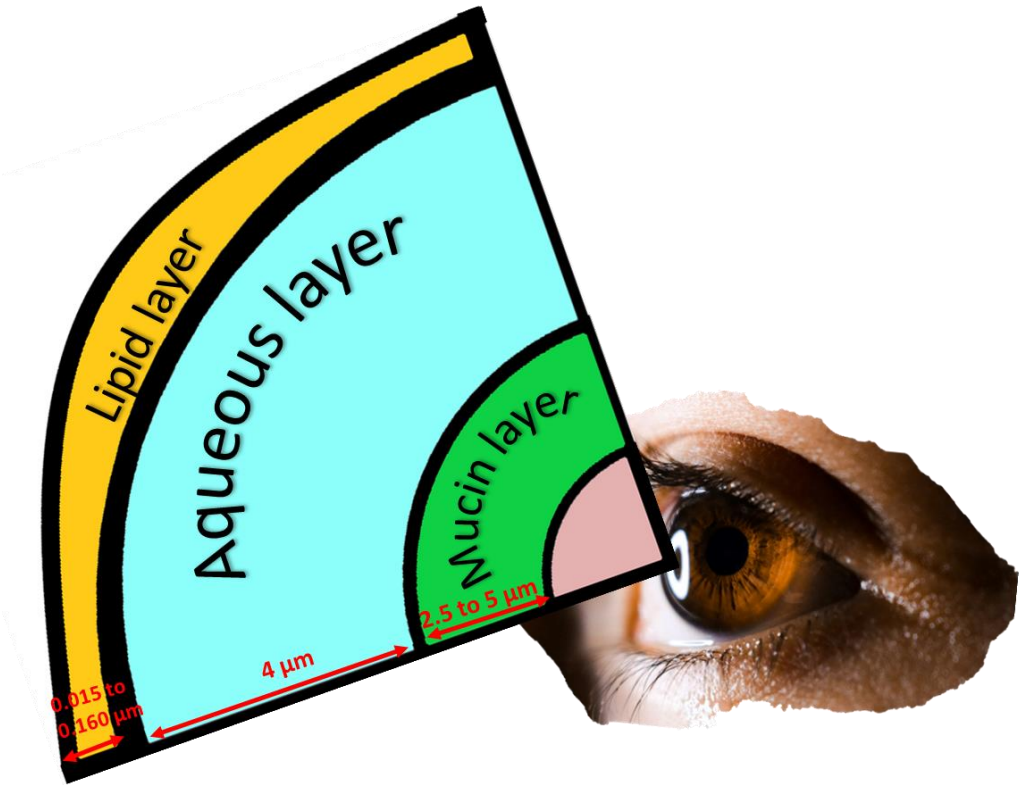


Figure 1-12: A diagram of the different layers found within tear fluid.

There is a wide variety of metabolites present in tears because they have three different layers: mucous, aqueous and lipid. This is visualised in Figure 1-12. The lipid layer is the top layer of the tear film and helps provide lubrication to the eye by stopping the other layers from evaporating too quickly [138,139]. The aqueous layer contains antibodies and proteins. The mucous layer is the bottom layer of the tear. This layer is hydrophilic and helps the aqueous layer of the tear spread across the cornea better [140]. All three layers of the tear will contain metabolites.

There has been a lot of previous research on tears but most of these studies have focused on proteomic[141-146] and lipidomic[147-149] assays, with metabolomics research on tears only becoming more common in recent years[147,150-153]. The limited metabolomics research that has been reported has shown that metabolites can be found in tears, but not all the metabolites identified have been quantified. An example of metabolites that are found in blood and expected to be seen in tears are diadenosine polyphosphates (APnAs) and vitamin A. APnAs, such as diadenosine triphosphate, are important as their biological role is to help facilitate tear secretion. Vitamin A is converted into retinol within the body and is used to restore the retina, hence why it is expected to be detected in tear samples. For future biological studies to be successful and for tear samples to become a well-researched sample type, the tear metabolome will have to be more extensively characterized as currently databases have very little information on metabolites currently detected in tears. For example, HMDB currently only has one metabolite listed as being present in tear fluid -H_2O (27/11/21). It would also be of use in research to compare the concentrations of the metabolites found in tears with another matrix, such as plasma.

Tear sampling can be conducted by applying a variety of methods.[154-159] These include: Schirmer's strip, glass capillary tube, absorbent methods (cellulose sponges, polyester fibre rod wick etc.), and eye flushing. Schirmer's strip method utilises a Schirmer's strip placed inside the lower eyelid (with or without anaesthesia) where the tear is absorbed by the strip. This method is commonly used by ophthalmology clinics as a test for dry eye disease. This method takes approximately 5 minutes. Glass capillary tubes are held against the lower eyelid with close attention not to touch the ocular surface. They are held there until the sample is collected, ~1µL per minute. All absorbent methods are conducted in a similar manner, with the absorbent material held in the lower tear meniscus for a fixed period. This method is simple and less uncomfortable than some of the other methods but can be more complex for sample extraction. Eye flushing is the least irritating and quickest of the tear collection methods, where a known volume of saline is pipetted into the eye, the patient then blinks before the solution and the tear are retracted. The disadvantage to using this method is that the tear is then diluted, and it is very difficult to calculate an accurate volume of tear fluid collected. Currently, across the different -omics techniques there has been very little investigation as to how the different sampling techniques affect sample composition, but the limited literature available shows that there is possibility that there are differences in samples collected[160,161].

1.6.4 Other common sample types

Metabolomics has an ever-growing collection of sample types applied in biomedical and clinical research that are commonly used. Some of these include tissue[162-166], saliva[167], sweat[168-170] and immortalised cell lines[171,172]. Tissue samples cover a wide range of samples with types of tissue varying from muscular (e.g. exercise related research on muscle metabolism[173]), liver[174] or kidney (can be very useful for research into tumours), placenta (research for metabolism of pregnancy), to brain and fat. Tissue samples are commonly collected via biopsies, these are very intrusive and painful, but an advantage is that they are the able to give the most accurate results on cellular metabolism as metabolism is focused on cells and tissues and not biofluids. Saliva and sweat samples are relatively new sample types in comparison to tissue samples, they are both much simpler and less intrusive to sample making them a preferable choice by participants. Saliva has been fairly well researched in comparison to a lot of other biofluids (other than urine and blood), with the human metabolome database (HMDB) having data on 885 metabolites that are detected and quantified in saliva. This is much higher than the 78 metabolites that are detected and quantified in sweat in the same database. Cell lines are a useful alternative for metabolomics research that would have otherwise required work using animals, an example of a commonly used cell line is hepaRG. HepaRG cell line is a liver tissue cell line, and so is highly suitable for metabolomics as there is a lot of metabolic activity that occurs in liver cells[175].

1.7 SAMPLE PREPARATION

To obtain as many metabolites from the sample as possible, untargeted metabolomics usually uses minimal sample preparation. Common examples of sample preparation used for untargeted metabolomics include protein precipitation (Section 1.7.1.1) and ‘dilute and shoot’ (Section 1.7.1.2). For both sample preparation methods, solvent choice is important because each solvent will influence the UHPLC separation and metabolite solubility. Polar solvents are also more suited to polar focused assay types such as aqueous reversed phase or HILIC methods. Commonly applied solvents include H_2O, ACN, and MeOH[82]. For lipidomics, non-polar solvents are more commonly applied, examples of these include IPA, ACN, and THF[176]. Using a solvent for sample preparation that is not being used in the assay can affect the resolution of the chromatography, meaning that the sample will not be separated very well across the method.

1.7.1 Monophasic extraction

1.7.1.1 Protein Precipitation Extraction (PPE)

Protein precipitation extraction is one of the main extraction methods applied to plasma and serum. This is because these sample types have a high protein concentration and proteins can significantly affect UHPLC-MS assays and reduce the metabolite information collected. PPE is primarily conducted, within metabolomics, by diluting the sample in a 1:3 (v/v) ratio (sample: solvent) using a suitable solvent

(usually acetonitrile or isopropyl alcohol or MeOH or as solvent mixtures of two or more of these solvents and sometimes with water), followed by vortex mixing and then centrifuging the mixture before using the supernatant for analysis. Solvent choice is important here, not only for the reasons above, but also because some solvents are better at precipitating protein than others[177].

1.7.1.2 "Dilute and Shoot"

Dilute and shoot is commonly used for urine metabolomics research. This is because urine normally does not include any, or very little, protein and the sample can therefore be simply diluted 1:1 (v/v) or greater with an appropriate solvent before being used for analysis. Urine samples must normally be diluted to not mass overload the UHPLC column or saturate the mass spectrometer detector for high concentration metabolites, and to balance out any differentiation on concentration of urine samples between participants.

1.7.2 Biphasic Extraction

A biphasic extraction utilises the creation of two phases to separate lipid metabolites from polar metabolites. This is useful for analysing small volume biofluids, and samples which include a physical mass (e.g., tissue or cell line samples.) The method requires two main steps; the dilution of the sample in methanol and water, later the addition of chloroform (alternatively methyl tert-butyl ether (MTBE) or dichloromethane (DCM) can be used) and water creates a phase separation (the final ratios of solvent are 2/2/1.8 methanol/chloroform/water)[178]. An advantage to using this sample preparation technique is that less sample is required for preparation of both reversed phase and polar assays, making it more appealing for sample types that are very intrusive or difficult to obtain. A disadvantage to using this method for sample preparation is that analytes that sit close to the precipitation layer in the middle may be lost in preparation in attempts to avoid the precipitated proteins. Sample may also be lost in the multiple transfers between multiple vials/Eppendorfs across the extraction steps.

1.8 IMPORTANCE OF METABOLOMICS IN EXERCISE

Metabolomics has most commonly been applied within the medical research fields to investigate different diseases (such as cancer [179], diabetes [180], obesity [181,182] etc.) and help to understand the human metabolome in disease states before physical symptoms may develop, and to help understand which metabolic pathways different diseases affect. Appling metabolic research into heath and exercise science is extremely important to fully understand how exercise affects the human metabolome. Investigating the effects of exercise on the human body is not only important for health research, but is also useful for sports performance, and nutritional needs. It also aids in the understanding of how different exercise types affect the body. There are lots of different types of studies that have been conducted when pairing metabolomics with exercise in research. These include but are not limited to: comparing metabolic profiles of athletes from a wide range of sporting disciplines[183], comparing

different elite level athletes from different disciplines[183], looking at low intensity exercise[184], looking at extended exercise[185,186] , looking at short high intensity exercise[187,188], and looking at long high intensity exercise[189]. A lot of exercise-based research focusses on set metabolic pathways such as glycolysis or β-oxidation of free fatty acids. This issue with this is that metabolomic pathways work together, and so looking at what changes are occurring globally can provide a much clearer picture of the metabolic changes taking place and the affects that the exercise is causing on the human body[190]. The most common sample types applied in exercise studies are plasma[137,191-193], urine[194-196], and sweat[168,197,198].

1.9 Overview and Current Limitations Conclusions

Overall, the application of liquid chromatography – mass spectrometry for research into untargeted metabolomics has quickly propelled the number of metabolic studies conducted and enabled detailed characterisation of metabolomes for multiple sample types with new technology being developed and optimised all the time. Further studies into multiple sample types have also allowed for the expansion of understanding about the human metabolome with knowledge on how different parts of the body are affected with age, health, and external factors and stressors. There is still a lot more knowledge to be discovered here and additional development of UHPLC-MS methods and metabolite annotation techniques will help further the knowledge within this research area. Further research will also bring the opportunity of method standardisation across research groups worldwide, something in which this research field needs in order to raise confidence levels in metabolite annotation.

1.10 Objectives of the PhD

This thesis aims to further develop and optimize UHPLC-MS methods in order to allow for a wider range of sample types to be analyzed by UHPLC-MS along with expanding the bottlenecked area of metabolite annotation, which will allow for better characterized metabolomes of a wider range of sample types, along with higher confidence annotations. Therefore, the following research chapters will aim to address the research knowledge gaps within metabolomics analytical techniques before applying them to studies within exercise-based metabolomics where there are also research gaps in the knowledge of how exercise affects the human metabolome.

Chapter 3 aims to develop a microbore UHPLC-MS method to be applied for sample types of small volumes. Current research has shown there is a lack of methods that are well suited to the growing range of sample types that are collected in small volumes/masses. Although there has been a wide range of work completed on comparing column internal diameters and sensitivity, there is only a very small number of studies that apply this to metabolomic samples, of which, have used samples that are abundantly available. Therefore, this chapter aims to bridge this research gap by developing a method

that is appropriate for samples of low volume/mass and validating this method by comparing the analysis of low volume samples to samples of high abundance applying the newly developed method.

Chapter 4 aims to help bridge the knowledge gap in metabolite annotation. Current research has highlighted the fact that metabolite annotation is a bottleneck in that so many metabolite features are detected but only a very small fraction of these are annotated, and even fewer that are annotated to a high level of confidence. Therefore, this research chapter will aim to investigate how the number of false positive annotations for lipids can be reduced by looking at the retention time patterns that occur by class type and lipid structure.

Chapter 5 involves the study of a set of participants undertaking three different exercise protocols with varying levels of intensity. Plasma samples collected at time points across each of the exercise sessions was analyzed and investigated on the differences and similarities of metabolic responses to the three different exercise protocols.

Chapter 6 involves the study of a set of participants undergoing a condensed heat acclimation trial to understand whether participants can be heat acclimatized in a shorter period than is traditionally applied. Multiple sample types were collected at several time points across the acclimations as well as in heat stress tests before and after the heat acclimation sessions. These samples were then analyzed and investigated for a better understanding of how the human metabolome changes with heat acclimatization and if the condensed heat acclimatization had any sustained effect on the metabolome.

Chapter Two

2 MATERIALS AND METHODS

2.1 MATERIALS

The following materials were purchased for all sample analysis completed at the University of Birmingham. Methanol, water, isopropyl alcohol (IPA) acetonitrile and chloroform (all HPLC grade) were purchased from VWR International (Loughborough, UK). Formic acid (HPLC grade) and ammonium formate (99%+ purity) were purchased from Sigma-Aldrich (Poole, UK).

2.2 ETHICS STATEMENTS

2.2.1 Ethics Statement For Research Featured In Development Of Microflow Ultra High-Performance Liquid Chromatography -Mass Spectrometry Metabolomic Assays For Analysis Of Mammalian Biofluids

Porcine samples were obtained from non-recovery anaesthetised animals within a pre-existing programme of animal work carried out by Dstl in accordance with the UK Animals (Scientific Procedures) Act, 1986.

2.2.2 Ethics Statement For An Investigation Into How The Type And Intensity Of Exercise Perturbs Metabolism

This study involved human participants and was therefore reviewed and approved by the University of Birmingham Ethics Committee. All participants provided their written informed consent to participate in this study. All data was pseudo anonymised to protect the identity of participants involved and I did not have the key to un-anonymise.

2.2.3 Ethics Statement For Research Into Heat Stress And Acclimation Effects On Human Metabolism

Participants in the study were all white males, had a BMI range of 18.9 – 29.2, and an age range of 20-32. Participants all signed participant consent and all data was pseudo anonymised to protect the identity of participants involved and I did not have the key to un-anonymise. Ethics application number 1017/MoDREC/19.

2.3 SAMPLE COLLECTION

2.3.1 Sample Collection Methods For Research Featured In Development Of Microflow Ultra High-Performance Liquid Chromatography - Mass Spectrometry Metabolomic Assays For Analysis Of Mammalian Biofluids

Human urine and plasma from one male donor were purchased in 5mL volumes from BioIVT (Sussex, UK). These were delivered frozen, thawed for aliquoting, and were then frozen again at -80°C before being thawed for analysis, giving a total of two freeze/thaw cycles. 2mL porcine urine samples, 1mL porcine plasma samples and ~20µL porcine basal tear fluid samples were all collected hourly for 13 hours by the Defence Science and Technology Laboratory (Porton Down, UK). These were delivered frozen (-20°C), stored at -80°C and only thawed for analysis, giving a total of 1 freeze/thaw cycle.

2.3.2 Sample Collection, Delivery, And Receipt Methods For An Investigation Into How The Type And Intensity Of Exercise Perturbs Metabolism

Multiple plasma samples collected at multiple time points in the School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham (Birmingham, UK) were centrifuged to remove any cellular material before being frozen at -80°C and transferred to School of Biosciences, University of Birmingham (Birmingham, UK). These samples were delivered frozen, thawed on ice for aliquoting and creation of pooled QCs and appropriate sample quantities and were then frozen again at -80°C before being thawed for sample preparation and analysis, giving a total of two freeze/thaw cycles per sample. This was done to ensure that the samples would not have to be at temperatures where they could quickly degrade for too long as if all thawing of large sample volumes and sample extraction and preparation was to take place on the same day, samples would have taken too long to defrost and the speed at which some would have defrosted compared to others would have led to data variance that was not due to the biological changes caused by exercise.

2.3.3 Sample Collection, Shipment And Receipt Methods for Research Into Heat Stress And Acclimation Effects On Human Metabolism

Multiple plasma and urine samples collected at multiple time points in the School of Sport, Health and Exercise science at University of Portsmouth (Portsmouth, UK) were centrifuged to remove any cellular material before being frozen at -80°C and transferred to School of Biosciences, University of

Birmingham (Birmingham, UK). These samples were delivered frozen, thawed for aliquoting and were then frozen again at -80°C before being thawed for analysis, giving a total of two freeze/thaw cycles per sample. This was done to ensure that the samples would not have to be at temperatures where they could quickly degrade for too long as if all thawing of large sample volumes and sample extraction and preparation was to take place on the same day, samples would have required a long time to defrost and the speed at which some would have defrosted compared to others would have led to data variance that was not due to the biological changes.

2.4 SAMPLE EXTRACTION

All sample extraction methods completed at the university of Birmingham followed the procedures below.

2.4.1.1 Human Plasma And Urine

Sample extraction applied three different solutions; ice-cold 50/50 (v/v) methanol/water (for the aqueous C_{18} reversed-phase assay), ice cold IPA (for the lipidomics C_{18} reversed-phase assay) and 3:1 acetonitrile/methanol (for the HILIC assay). For all plasma samples, protein precipitation was performed by mixing 50 µL of plasma with 150 µL of extraction solution followed by vortex mixing (30 seconds) and centrifugation (21,885 xg, 20 minutes, 4 °C). For all neat urine samples, dilution was performed by mixing 50 µL of urine with 150 µL of extraction solution followed by vortex mixing (30 seconds) and centrifugation (21,885 xg, 20 minutes, 4 °C).

For chapter 3 only, the supernatant from the human plasma and human urine extraction solutions were then serially diluted with extraction solvent to prepare the following dilutions: 1 in 4, 1 in 8, 1 in 16, 1 in 32, 1 in 64 and 1 in 128 (see Table 2-1 below).

Table 2-1: serial dilution calculation table with the volumes of sample and solvent required for the serial dilution created as part of the sample preparation for all analytical methods

Dilution	Sample (µL)	Solvent (µL)	Total volume (µL)
1/4	197.00	591.00	788.00
1/8	388.00	388.00	776.00
1/16	375.00	375.00	750.00
1/32	350.00	350.00	700.00
1/64	300.00	300.00	600.00
1/128	200.00	200.00	400.00

Each new dilution was created using the next highest dilution sample (for example, the 1 in 16 dilution solution was prepared from the 1 in 8 dilution solution). All diluted sample solutions were vortex mixed (30 seconds) followed by centrifugation (4000xg, 20 minutes, 4°C). For each assay two extraction replicates were prepared at each dilution and four technical replicates were injected (two from each extraction replicate).

2.4.1.2 Porcine Plasma, Urine, And Tears

Sample extraction applied three different solutions; ice-cold 50/50 (v/v) methanol/water (for the aqueous C_{18} reversed-phase assay), ice cold IPA (for the lipidomics C_{18} reversed-phase assay) and 3:1 acetonitrile/methanol (for the HILIC assay).

For all plasma samples, protein precipitation was performed by mixing 50µL of plasma with 150µL of extraction solution followed by vortex mixing (30 seconds) and centrifugation (21,885 xg, 20 minutes, 4°C). For all neat urine samples, dilution was performed by mixing 50µL of urine with 150µL of extraction solution followed by vortex mixing (30 seconds) and centrifugation (21,885 xg, 20 minutes, 4°C).

For the porcine tear fluid samples, extraction was carried out using a biphasic liquid - liquid extraction method, where a chloroform/methanol/water 2:2:1.8 (v/v/v) solvent system[178] was used to separate polar metabolites from lipid metabolites whilst also precipitating out proteins[163]. Each tear fluid was weighed out, using the assumption that the density of tear fluid was equal to that of water; 1 g/mL. Once the samples had thawed on ice, 8 µl/mg tear fluid of cold methanol was added to the sample and vortexed. Samples were then transferred to 1.8mL vials where 8 µL/mg tear fluid of chloroform and 4 µl/mg tear fluid of water were added. The samples were then vortexed for 30s and then left at room temperature for 10 minutes. Samples were then centrifuged (2,500xg, 4°C) for 10 minutes. Following this, samples were left at room temperature for a further 5 minutes before removal of both the polar and lipid phases. The polar phase was dried down in a SPD11V SpeedVac sample concentrator (Thermo Scientific) at room temperature for 4 hours whilst the lipid phase was dried down by nitrogen blowdown at room temperature. Both phases were then stored in a freezer at -80°C, before being re-suspended in 40µL of solvent on the day of analysis.

2.4.1.3 Quality Control Samples

2.4.1.3.1 Plasma and Urine Samples

QC samples for each sample type were prepared by pooling sample aliquots and vortexing at top speed for 1 minute before aliquoting out into separate 50µL Eppendorfs. QC samples were then diluted 1 in 4 with either ice-cold 50/50 (v/v) methanol/water (for the aqueous C_{18} reversed-phase assay), ice cold IPA (for the lipidomics C_{18} reversed-phase assay) or 3:1 acetonitrile/methanol (for the HILIC assay). All diluted sample solutions were vortex mixed (30 seconds) followed by centrifugation (4000xg, 20 minutes, 4°C) before being stored in a -80°C freezer until ready for analysis.

2.4.1.3.2 Tear Samples

QC samples were only created for the MS/MS injections as there was not enough sample to be used to create QCs for assessment of data quality (plasma QCs were applied for this).

QC samples were created by taking a 5µL from each sample after reconstitution and pooling them together before mixing in a vortex (1 minute) before being separated out into five LC vials ready for analysis.

2.4.1.4 Blank Samples

2.4.1.4.1 Plasma and Urine Samples

Blank samples were created using the sample protocol as mentioned in section 2.1.1.2 but instead of sample, water was used.

2.4.1.4.2 Tear Samples

Blank samples for were created using the sample protocol as mentioned in section 2.1.1.2 for tear samples but instead of sample, water was used.

2.5 UHPLC Analysis Methods

For data acquisition completed at the University of Birmingham a Vanquish™ Horizon UHPLC System coupled to a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) was used. All data were collected separately in positive and negative ion modes, with data on the human samples collected first and the porcine data collected at a later date. Mass calibration was performed in both ion modes applying Pierce™ LTQ Velos ESI Positive Ion Calibration Solution and Pierce™ Negative Ion Calibration Solution to carry out the recommended manufacturer evaluation and calibration tests. 10 QC samples were injected at the start of the run for column equilibration. Before the data collection of each sample type, a further five QCs from that sample type for collection of MS/MS spectra. QC samples were also injected every sixth injection for data quality assessment during data analysis.

2.5.1 UHPLC-MS Methods Applied For Chapters 3 And 4 For Samples Analysed By Researchers At The University Of Birmingham

2.5.1.1 C18 Reversed-Phase Lipidomics (Lipids) Assay

Plasma samples were analysed in both positive and negative ion mode on both ID columns. Column details for each research chapter can be found below. For the 2.1mm ID column, solvent A was 10 mM ammonium formate in 60/40 acetonitrile/water (v/v) + 0.1% formic and solvent B was 10 mM ammonium formate in 90/10 IPA/acetonitrile (v/v) + 0.1% formic acid. The gradient elution used was 0 to 1 minute 99%A; 1 to 3 minutes 85%A with a curve = 5; 3 to 6 minutes 50%A with a curve = 5, 6 to 9 minutes; 5%A with a curve = 5, 9 to 10 minutes; 5%A, 10 to 10.5 minutes; 99%A with a curve = 5, 10.5 to 14 minutes; 99%A. The UHPLC parameters were flow rate=0.40 $mL.min^{-1}$, column temperature=55 °C and injection volume=2 µL. Mass spectrometer settings were scan range=150-2000 m/z, sheath gas flow rate=48; aux gas flow rate=0, spray voltage (kV) positive ion mode=3.2 and negative ion mode=2.7, capillary temp=350°C; S-lens RF level=60 and aux gas heater temperature=400°C. A high flow electrospray needle was used.

2.5.1.2 C_{18} Aqueous Reversed-Phase (RPaq) Assay

Urine samples were analysed in both positive and negative ion mode on both ID columns. The column used for this assay was a Thermo ScientificTM Hypersil GOLD™ aQ C_{18} (175 Å, 1.9 µm, 2.1 mm x 100 mm). Solvent A was water/formic acid (99.9/0.1), and solvent B was acetonitrile/formic acid (99.9/0.1). The gradient used was 0 to 0.5 minutes 99%A; 0.5 to 2 minutes 50%A with a curve = 5; 2 to 10.5 minutes 1%A with a curve = 5; 10.5 to 11 minutes 99%A with a curve = 8; 11-11.5 minutes 99%A; 11.5-15.0 minutes 99%A. The UHPLC parameters were: flow rate = 0.30$mL.min^{-1}$ for 0-11 minutes and 14.9-15.0 minutes and 0.4$ml.min^{-1}$ for 11-14.9 minutes. The column temperature was 40°C and the injection volume was 2µL. Mass spectrometer settings were scan range = 100-1500 m/z, sheath gas flow rate of 30, aux gas flow rate = 13, sweep gas flow rate = 0, spray voltage (kV) = 3.2 (positive ion mode) and 2.7 (negative ion mode), capillary temperature = 350°C, S-lens RF level = 40 and aux gas heater temperature = 400°C. A high flow electrospray needle was applied.

2.5.1.3 Hydrophilic Interaction Liquid Chromatography (HILIC) Assay

Plasma and urine samples were analysed in both positive and negative ion mode on both ID columns. The columns applied for this method can be found below in sections relating to each research chapter. The positive ion mode solvent A was 10 mM ammonium formate in 95/5 acetonitrile/water (v/v) + 0.1% formic acid; solvent B was 10 mM ammonium formate in 50/50 acetonitrile/water (v/v) + 0.1% formic acid. The Negative ion mode solvent A was 10 mM ammonium acetate in 95/5 acetonitrile/water (v/v) + 0.1% acetic acid; solvent B was 10 mM ammonium acetate in 50/50 acetonitrile/water (v/v) + 0.1% acetic acid. The gradient used was 0 to 0.5 minutes 99%A; 0.5 to 2 minutes 50%A with a curve

= 5; 2 to 10.5 minutes 1%A with a curve = 5; 10.5 to 11 minutes 99%A with a curve = 8; 11-11.5 minutes 99%A; 11.5-15.0 minutes 99%A. The UHPLC parameters were flow rate was 0.5mL.min^{-1} for 0-11 minutes and 14.9-15.0 minutes and 0.40ml.min^{-1} for 11-14.9 minutes. The column temperature was 40°C and the injection volume was 2µL. Mass spectrometer settings were scan range = 70-1050 m/z, sheath gas flow rate = 30, aux gas flow rate = 13, sweep gas flow rate = 0, spray voltage (kV) = 3.2 (positive ion mode) and 2.7 (negative ion mode), capillary temperature = 350°C, S-lens RF level = 40 and aux gas heater temperature = 400°C. A high flow electrospray needle was applied.

2.6 METHODS APPLIED TO RESEARCH FEATURED IN DEVELOPMENT OF MICROFLOW ULTRA HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY -MASS SPECTROMETRY METABOLOMIC ASSAYS FOR ANALYSIS OF MAMMALIAN BIOFLUIDS

2.6.1 C_{18} Reversed-Phase Lipidomics (Lipids) Assay Details

Human plasma samples were analysed in both positive and negative ion mode on both ID columns. Porcine plasma and tear samples were analysed on the 1.0mm ID column using both positive and negative ion mode. The columns used were a Waters® ACQUITY UPLC® BEH C_{18} (130Å, 1.7 µm, 2.1 mm x 100 mm) and a Waters® ACQUITY UPLC® BEH C_{18} (130Å, 1.7 µm, 1.0 mm x 100 mm).

For the 2.1mm ID column, method 2.5.1.1 was applied.

For the 1.0mmID column, solvent A was 10 mM ammonium formate in 60/40 acetonitrile/water (v/v) + 0.1% formic and solvent B was 10 mm ammonium formate in 90/10 IPA/acetonitrile (v/v) + 0.1% formic acid. The gradient elution used was 0 to 1 minute 80%A; 1 to 10 minutes 0%A with a curve = 5; 10 to 14 minutes 0%A; 14 to 14.5 minutes 80%A with a curve = 5; 14.5 to 25 minutes 80%A. The UHPLC parameters were flow rate = 0.12mL.min^{-1}, column temperature = 55°C and injection volume =2µL. Mass spectrometer settings were a scan range =150-2000 m/z, sheath gas flow rate =32 (positive ion mode) and 48 (negative ion mode), aux gas flow rate = 10 (positive ion mode) and 11 (negative ion mode), sweep gas flow rate = 1 (positive ion mode) and 2 (negative ion mode), spray voltage (kV) = 3.2 (positive ion mode) and 2.5 (negative ion mode), S-lens RF level of 50 and aux gas heater temp = 168°C (positive ion mode) and 413°C (negative ion mode). A high flow electrospray needle was used.

2.6.2 C_{18} Aqueous Reversed-Phase (RPaq) Assay

Human urine samples were analysed in both positive and negative ion mode on both ID columns. Porcine urine and tear samples were analysed on the 1.0mm ID column in both positive and negative ion mode. The columns used were a Thermo ScientificTM Hypersil GOLD™ aQ C_{18} (175 Å, 1.9 µm,

2.1 mm x 100 mm) and Thermo ScientificTM Hypersil GOLD™ aQ C_{18} (175 Å, 1.9 µm, 1.0 mm x 100 mm).

For the 2.1mm ID column, method 2.5.1.2 was applied.

For the 1.0mm ID column, solvent A was water/formic acid (99.9/0.1), and solvent B was acetonitrile/formic acid (99.9/0.1). The gradient used was 0 to 2 minutes 99%A; 2 to 7 minutes 50%A with a curve = 5; 7 to 10.5 minutes 1%A with a curve = 5; 10.5 to 11 minutes 1%A; 11-11.5 minutes 99%A with a curve = 5; 11.5-15.0 minutes 99%A. The flow rate was 0.13mL.min^{-1} = 0-11 minutes and 14.0-15 minutes and was 0.135mL.min^{-1} = 11-14.9 minutes. The column temperature was 40°C and the injection volume was 2µL. Mass spectrometer settings were scan range = 100-1500 m/z, sheath gas flow rate = 30, aux gas flow rate = 13, sweep gas flow rate = 0, spray voltage (kV) = 3.2 (positive ion mode) and 2.7 (negative ion mode), capillary temp = 350°C, S-lens RF level = 40 and aux gas heater temperature = 400°C. A high flow electrospray needle was used.

2.6.3 Hydrophilic Interaction Liquid Chromatography (HILIC) Assay

Huma plasma and urine samples were analysed in both positive and negative ion mode on both ID columns. Porcine plasma, urine and tear samples were analysed on the 1.0mm ID column in both positive and negative ion mode. The columns used were a Waters® ACQUITY UPLC® BEH HILIC (130Å, 1.7 µm, 2.1 mm x 100 mm) and a Waters® ACQUITY UPLC® BEH HILIC (130Å, 1.7 µm, 1.0 mm x 100 mm). For the 1.0mm ID column, Positive ion mode solvent A was 10 mM ammonium formate in 95/5 acetonitrile/water (v/v) + 0.1% formic acid; solvent B was 10 mM ammonium formate in 50/50 acetonitrile/water (v/v) + 0.1% formic acid. Negative ion mode solvent A was 10 mM ammonium acetate in 95/5 acetonitrile/water (v/v) + 0.1% acetic acid; solvent B was 10 mM ammonium acetate in 50/50 acetonitrile/water (v/v) + 0.1% acetic acid. The gradient used was 0 to 1 minute 99%A; 1 to 3 minutes 80%A with a curve = 3; 3 to 5 minutes 80%A; 5 to 5.5 minutes 75%A with a curve = 5; 5.5-9.5 minutes 5%A with a curve = 5; 9.5-10.0 minutes 5%A; 10-10.5 minutes 99%A with a curve of 5; 10.5-14 minutes 99%A with a curve of 5. The flow rate was 0.175mL.min-1 for 0-10.5 minutes and 13.9-14 minutes and was 0.225mL.min-1 for 10.5-13.9 minutes. The column temperature was 40°C and the injection volume was 2µL. Mass spectrometer settings were scan range = 70-1050 m/z, sheath gas flow rate = 10, aux gas flow rate = 5, sweep gas flow rate = 2, spray voltage (kV) = 3.2 (positive ion mode) and 2.7 (negative ion mode), capillary temp = 320°C, S-lens RF level = 30 and aux gas heater temperature = 440°C. A high flow electrospray needle was used.

2.6.4 Raw Data Processing, Metabolite Annotation

Raw data files (.RAW format) for the human plasma and urine samples were imported into Compound Discoverer ™ 3.0 (Thermo Fisher Scientific, San Jose, USA) where the data were processed by applying peak picking, peak integration and data integration across all samples at all dilutions for each sample type separately. One data matrix was constructed for each sample type/assay combination of compound (rows) and sample (columns) with the extracted ion chromatogram peak area for all metabolite features of the same compound reported where a compound was detected in a specific sample. After processing, the software was used to annotate the compounds detected using the Chemspider search modules in which MS1 data is matched to molecular formulae. All metabolites were annotated to level 3 as defined by the Metabolomics Standards Initiative [199]. MS2 data was not collected for this section of the work as the focus was the chromatography and number of compounds annotated rather than the annotations themselves. The Compound Discoverer ™ workflow and parameters used can be found in the appendix (Figures A1-13). Blank filtering was applied to remove compounds whose mean peak area in the blank samples was greater than 10% of the mean peak area in the QC samples.

The porcine sample data files were processed in a similar workflow, using Compound Discoverer ™ 3.1 (Thermo Scientific). After processing the software was used to annotate the compounds detected. MS1 data was matched to molecular formulae and metabolites using Chemspider searches. MS2 data were searched against mzCloud (Thermo Scientific), and an in-house MS/MS mass spectral library constructed with authentic chemical standards. All metabolites were annotated to level 2 or level 3 as defined by the Metabolomics Standards Initiative[199]. The workflow and parameters used can be found in the appendix (Figures B1-15). Blank filtering was applied to remove compounds whose mean peak area in the blank samples was greater than 10% of the mean peak area in the QC samples.

2.6.5 Data Analysis

2.6.5.1 Human Samples

Comparisons for the number of compounds reported, the mean peak area and mean peak width were carried out for each column at each sample dilution. This was completed for each dilution within a column so that the effect of sample dilution could be determined along with the effects caused by changing the column ID. For peak width and peak area comparisons, the compounds per file outputs from Compound Discoverer ™ were used to compare the full width half max (FWHM) and chromatographic peak area of peaks present across each dilution. Prior to comparisons, the datasets were filtered by removing peaks with a FWHM > 0.2 minutes on the basis that those peaks were most likely to be solvent peaks and/or were peaks of poor chromatographic resolution or shape. For comparing the number of compounds reported, filters removed peaks that were not present in 75% (or 3 of 4) of the technical replicates in the dilution being studied and which were reported with a relative standard deviation (RSD) for the peak area < 30%.

2.6.5.2 Porcine Samples

Outliers were discovered and removed post processing by summing the peak areas within each sample. Samples with total peak areas much lower than others of the same sample type were omitted from any further analysis. This was done to improve confidence in results. Filtering was then applied to each sample type to remove compounds that were not present in a minimum of two samples (regardless of how many samples were present). These samples were then filtered to remove any compounds that had an RSD > 30%.

To determine whether metabolites were detected in one, two or three of the sample types (plasma, urine, tear fluid), compounds which were within 0.001Da molecular weight of each other and the retention times were within 0.1 minute of each other, were classed as a matching compound. Datasets were checked for retention time shift before compound matching was conducted.

2.7 METHODS APPLIED TO IN-SILICO EXTENSION OF RT LIBRARIES FOR WATER-SOLUBLE METABOLITES AND LIPIDS WHEN APPLYING UHPLC-MS.

2.7.1 Retention Time Mapping Datasets – Reversed Phase

The datasets from Mayo clinic and University of Florida are biological datasets obtained from the metabolomics workbench[200]. The datasets from University of Birmingham[201] and University of Liverpool[202] are retention time libraries made from analytical standards.

2.7.2 Retention Time Mapping Datasets – HILIC

The dataset from Mayo clinic is a biological dataset obtained from metabolomics workbench[200]. The datasets from University of Birmingham[201], University of Liverpool[202], and Fiehn MassBank of North America (MoNA)[203] are retention time libraries made from analytical standards.

2.7.3 UHPLC Methods Applied By Different Research Groups

2.7.3.1 University of Birmingham UHPLC-MS Analysis

Data acquisition for C_{18} Aqueous Reversed-Phase (RPaq) and HILIC assays was performed during a two-week trip of Dr William Nash and Dr Elliott Palmer (University of Birmingham) to visit Dr Ioanna Ntai (Thermo Fisher Scientific, San Jose, USA) in February 2019. 875 chemical standards purchased from MetaSci[204] were prepared in 50 groups of varying numbers of metabolites. All solutions were analysed applying UHPLC-MS to acquire HCD MS2 data and UHPLC retention time data in both positive and negative ion mode using an aqueous C_{18} reversed-phase assay and a HILIC assay.[201]

Data acquisition for the lipids retention time library was completed by Dr Lukáš Najdekr at the University of Birmingham. A variety of sample types were analysed to allow for as many lipids to be detected as possible.

2.7.3.2 C_{18} Reversed-Phase Lipidomics (Lipids) Assay

Analytical standards were analysed in both positive and negative ion mode on a Thermo ScientificTM Hypersil GOLD™ C_{18} (175 Å, 1.9 µm, 2.1 mm x 100 mm). Solvent A was 10 mM ammonium formate in 60/40 acetonitrile/water (v/v) + 0.1% formic and solvent B was 10 mM ammonium formate in 90/10 IPA/acetonitrile (v/v) + 0.1% formic acid. The gradient elution used was 0 to 1 minute 99%A; 1 to 3 minutes 85%A with a curve = 5; 3 to 6 minutes 50%A with a curve = 5, 6 to 9 minutes; 5%A with a curve = 5, 9 to 10 minutes; 5%A, 10 to 10.5 minutes; 99%A with a curve = 5, 10.5 to 14 minutes; 99%A. The UHPLC parameters were flow rate=0.40 $mL.min^{-1}$, column temperature=55 °C and injection volume=2µL. Mass spectrometer settings were scan range=150-2000 m/z, sheath gas flow rate=48; aux gas flow rate=0, spray voltage (kV) positive ion mode=3.2 and negative ion mode=2.7, capillary temp=350°C; S-lens RF level=60 and aux gas heater temperature=400°C. A high flow electrospray needle was used.

2.7.3.3 C_{18} Aqueous Reversed-Phase (RPaq) Assay

Analytical standards were analysed in both positive and negative ion mode using method 2.5.1.2 with the change that flow rate was kept constant at 0.30 $mL.min^{-1}$.

2.7.3.4 Hydrophilic Interaction Liquid Chromatography (HILIC) Assay

Analytical standards were analysed in both positive and negative ion mode using a Thermo ScientificTM Accucore™ 150 Amide column (150Å, 2.6µm, 2.1 mm X 100 mm). Method 2.5.1.3 was applied.

2.7.4 University of Liverpool UHPLC-MS Analysis

2.7.4.1 LC Method 1 (Reversed-Phase) Accurate Mass Retention Time (AMRT) Database

Analytical standards were analysed on a Zorbax Eclipse Plus C_{18} column (2.1x100 mm, 1.8 µm, Agilent). Mobile phases were (A) water and (B) methanol both containing 31.6 mg/dL ammonium formate and 0.1% formic acid. The gradient used was 5% B at 0-1 min increasing linearly to 100% by 12 min, held at 100% B until 14 min, returning to 95% A for 5 min. The flow rate = 0.40 $mL.min^{-1}$ and column temperature = 60 °C.[202]

2.7.4.2 LC Method 2 (Reversed-Phase) Accurate Mass Retention Time (AMRT) Database

Analytical standards were analysed on an Atlantis dC_{18} column (3x100 mm, 3 µm, Waters, UK). Mobile phases were (A) water and (B) methanol both containing 31.6 mg/dL ammonium formate and 0.1% formic acid. The gradient used was 5% B at 0-1 min increasing linearly to 100% by 12 min, held at

100% B until 14 min, returning to 95% A for 5 min. The flow rate = 0.40 mL.min-[1] and the column temperature = 60 °C.[202]

2.7.4.3 LC Method 3 (Normal-Phase) Accurate Mass Retention Time (AMRT) Database

Analytical standards were analysed on a BEH amide column (3x150 mm, 1.7 µm, Waters). Mobile phases were (A) water and (B) acetonitrile both containing 0.1% formic acid. The elution gradient used was 99% B 0-1 min decreasing linearly to 30% from 1-12 min, held at 30% B until 12.6 min, returning to 99% B for 3.4 min. The flow rate = 0.60 mL.min-[1] and the column temperature = 40 °C.[202]

2.7.5 Fiehn HILIC MoNA UHPLC-MS Analysis Methods

Analytical standards were analysed on a Waters Acquity UPLC-BEH Amide column (150 mm × 2.1 mm; 1.7 µm) coupled to an Acquity UPLC BEH Amide VanGuard precolumn (5 × 2.1 mm2; 1.7 µm). Mobile phases were A) water with ammonium formate (10 mM) and formic acid (0.125%) and (B) acetonitrile: water (95:5, v/v) with ammonium formate (10 mM) and formic acid (0.125%). The elution gradient used was 0 min 100% B; 0−2 min 100% B; 2−7.7 min 70% B; 7.7−9.5 min 40% B; 9.5−10.25 min 30% B; 10.25−12.75 min 100% B; 12.75−17 min 100% B. The flow rate = 0.40 mL.min-[1] and the column temperature = 45 °C. [205]

2.7.6 University of Florida UHPLC-MS Analysis Methods

Metabolite separation for mammalian samples was achieved using a Thermo Scientific-Dionex Ultimate 3000 (ThermoFisher Scientific, San Jose, CA) coupled to a Thermo Scientific- Q Exactive. The column applied was an ACE Excel 2 C_{18}-PFP (100 x 2.1mm, 2.0 µm) with a Halo C_{18}-PFP guard attached. Mobile phases applied were: (A) H20 with 0.1% Formic Acid LC/MS grade and (B) Acetonitrile LC/MS grade. The gradient used had a starting flow rate of 0.35 ml.min-[1] and the initial starting conditions of 0% B. These initial conditions lasted until 3 minutes and then increased linearly to 80% by 13 minutes, were held at 80% B until 16 minutes, returning to 0% B for 0.5 minutes. Where the flow rate was increased to 0.6 ml.min-[1] at 0% B until 20 minutes. The flow rate was the reduced back to 0.35ml.min-1. The column temperature was 60°C. [200]

2.8 METHODS APPLIED TO AN INVESTIGATION INTO HOW THE TYPE AND INTENSITY OF EXERCISE PERTURBS METABOLISM

2.8.1 Experimental Visits

Participants from the study had a 50/50 female/male split. The total age range was 18.3-26.7 –-years (female age range 18.7-25.0, male age range 18.3-26.7). The BMI range across all participants was 20.8-25.8 (female BMI range 21.1-25.1, male BMI range 20.8-25.8). All participants took part in three experimental visits within two weeks, the order of each experimental visit was randomised across

participants to avoid experimental bias. Before starting exercise, participants rested for at least 10 minutes before resting blood samples were collected. All exercise protocols were carried out on an exercise bike at a self-selected cadence, including a 5-minute warm up, ending with 3 minutes of active recovery at 50 W and a seated 15-minute passive recovery period.

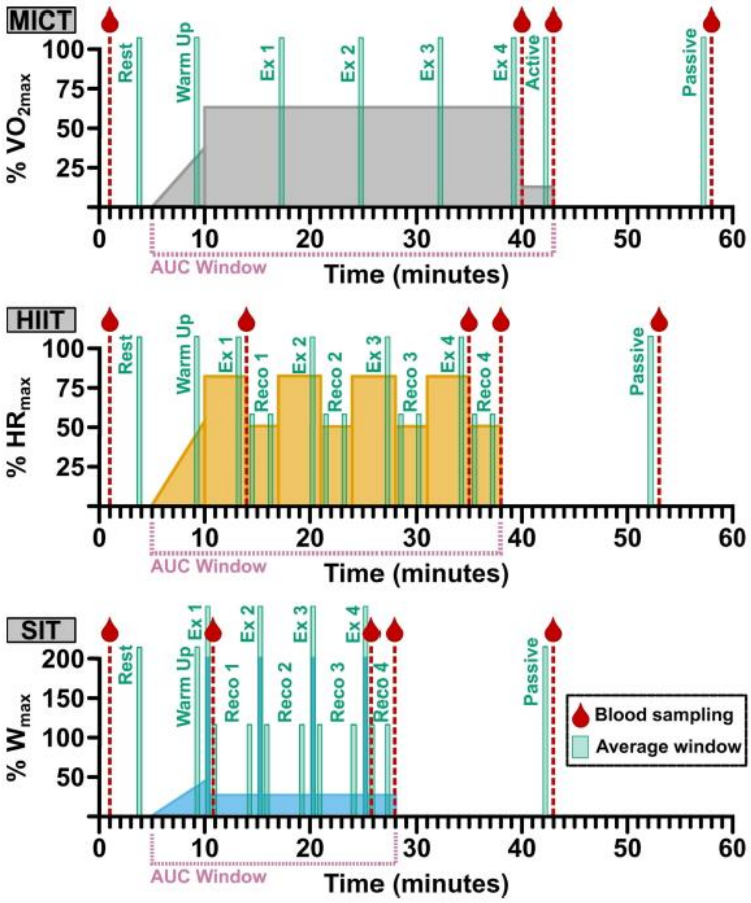


Figure 2-1: "Participants (50/50 male/female split) took part in three separate exercise visits in a randomized order, consisting of 30 min of moderate, steady-state exercise at 65% VO2peak (MICT); four 4 min high intensity intervals at 85% HRmax (HIIT); and four 30 s supramaximal sprint intervals at 200% Wmax (SIT). All exercise protocols were preceded by 5 min of seated rest and a 5 min warm-up, and then followed by 3 mins of active recovery at 50 W and 15 mins of seated passive recovery. Graph Key: The timing of blood sampling is marked by red blood droplets. VO2peak = maximum oxygen consumption; HRmax = maximum heart rate; Wmax = maximum aerobic power."[206] *Figure reconstructed from*[206].

2.8.1.1 MICT Protocol

During MICT, participants carried out 30 minutes of continuous exercise at 65% of VO_2peak. Resistance on the cycle ergometer was initially set between 55 and 60% of Wmax and adjusted once cycling had commenced in order to attain 65% of VO_2peak. Blood samples were collected using heparin anti-coagulant tubes at rest, at the end of each exercise protocol, end of active recovery and end of

passive recovery. A visualisation of this can be seen in Figure 2-1 where all sampling points are represented with a red droplet. Samples were immediately centrifuged at 5,000g for 10 min at 4°C before all acellular supernatant was aliquoted and frozen at −80°C for future analysis.

2.8.1.2 HIIT Protocol

HIIT consisted of four lots of 4 minute intervals at 85% HRmax separated by 3 minutes of active recovery at 50 W, based on guideline recommendations and definitions of HIIT in clinical settings[207]. Initial workload during interval bout 1 was set at ∼70% Wmax and monitored and adjusted to achieve 85% HRmax in a similar manner to the MICT workload. For remaining bouts, the initial intensity was set at the power output during the final 30s of the preceding 4 min bout and adjusted as necessary to maintain intensity at 85% HRmax. Blood samples were collected using heparin anti-coagulant tubes at rest, at end of the initial interval bout, end of each exercise protocol, end of active recovery, and end of passive recovery. A visualisation of this can be seen in Figure 2-1 where all sampling points are represented with a red droplet. Samples were immediately centrifuged at 5,000g for 10 min at 4°C before all acellular supernatant was aliquoted and frozen at −80°C for future analysis.

2.8.1.3 SIT Protocol

SIT intervals consisted of four lots of 30 second intervals at 200% of Wmax, separated by 4.5 minutes of active recovery at 50 W; except for the final recovery which was matched to MICT and HIIT at 3 min at 50 W. Following warm-up, the workload was reduced to 50W for 30s in order to ensure that the change in workload from pre-interval to sprint was matched across all four intervals. Workload was ramped up 5s prior to each sprint interval to account for a ∼5–7s delay in resistance on the cycle ergometer and intensity was dropped to 50W as soon as the 30s bout ended. Blood samples were collected using heparin anti-coagulant tubes at rest, at end of the initial interval bout, end of each exercise protocol, end of active recovery, and end of passive recovery. A visualisation of this can be seen in Figure 2-1 where all sampling points are represented with a red droplet. Samples were immediately centrifuged at 5,000g for 10 min at 4°C before all acellular supernatant was aliquoted and frozen at −80°C for future analysis.

2.8.2 Sample Extraction

Sample extraction applied the extraction techniques for plasma samples to be analysed using Lipidomics and HILIC assays as described in section 2.4.1.1. Sample randomisation was conducted as block randomisation where the participants were randomised and then all samples for that participant were randomised. QC samples were created using the method 2.4.1.3.1 and blank samples were created using the method described in 2.4.1.4.1.

2.8.3 UHPLC-MS Analysis

Analysis of the samples using UHPLC-MS was conducted using similar methods as applied in section 2.5.1. Along with MS1 data being collected MS2 data were also collected from QC samples to obtain further structural information about the metabolites present in the samples. Sample randomisation for analysis was conducted again after extraction. Block randomisation was completed where the participants were randomised and then all samples for that participant were randomised.

2.8.3.1 C_{18} Reversed-Phase Lipidomics (Lipids) Assay

Plasma samples were analysed in both positive and negative ion mode. The column used was a Thermo Scientific™ Hypersil GOLD™ C_{18} (175Å, 1.9 µm, 2.1 mm x 100 mm). The mobile phases used were the same as in method section 2.5.1.1. The gradient elution differed to that of 2.5.1.1 and was 0 to 9 minutes 80%A; 9 to 12.5 minutes 0%A with a curve = 5; 12.5 to 15 minutes 80%A with a curve = 5. The UHPLC parameters were flow rate = 0.40mL.min^{-1}, column temperature = 55°C and injection volume = 2µL. Mass spectrometer settings applied can be see in section 2.5.1.1. MS/MS data collection was completed using data dependent acquisition (DDA) in "discovery mode". Parameters for this were: mass resolution = 17,500 (FWHM at m/z 200); isolation width = 3.0 m/z; stepped normalised collision energies (stepped NCE) = 20, 40, 100 for positive ionisation mode and (stepped NCE) = 40, 60, 130 for negative ionisation mode. Spectra were acquired in five different mass ranges: 150–510 m/z; 500–710 m/z; 700–860 m/z; 850–1010 m/z; 1000–2000 m/z. The method was set using 'Discovery mode' with scan event cycle of 1 Full MS and 3 DDA MS/MS scans (TopN = 3).

2.8.3.2 Hydrophilic Interaction Liquid Chromatography (HILIC) Assay

Plasma samples were analysed in both positive and negative ion mode. The column used was a Thermo Scientific™ Accucore™ 150 Amide (150Å, 2.6µm, 2.1 mm x 100 mm). The method used for the HILIC assay was largely the same as the method written in section 2.5.1.3 but with a different elution gradient applied. Instead, the gradient consisted of 0 to 4 minutes 99%A; 4 to 7 minutes 85%A with a curve = 5; 7 to 10 minutes 50%A with a curve = 5; 10 to 11 minutes 5%A with a curve = 5; 11-11.5 minutes 5%A; 11.5-15.0 minutes 99%A. MS/MS data was collected on QC samples to allow for more structural data. This was completed using data dependent acquisition (DDA) in "discovery mode". Parameters for this were: mass resolution = 17,500 (FWHM at m/z 200); isolation width = 3.0 m/z; stepped normalised collision energies (stepped NCE) = 20, 50, 100 for positive ionisation mode and (stepped NCE) = 40, 60, 130 for negative ionisation mode. Spectra were acquired in five different mass ranges: 70–210 m/z; 200–310 m/z; 300–410 m/z; 400–510 m/z; 500–1050 m/z. The method was set using 'Discovery mode' with scan event cycle of 1 Full MS and 3 DDA MS/MS scans (TopN = 3).

2.8.4 Raw Data Processing, Metabolite Annotation

Raw data files (.RAW format) were imported into Compound Discoverer ™ 3.1 (Thermo Scientific, San Jose, USA) where the data were processed by applying peak picking, peak integration and data integration across all samples at all time points. Separate data matrices were constructed for each ion mode of each assay in the combination of compound (rows) and sample (columns) with the extracted ion chromatogram peak area for all metabolite features of the same compound reported where a compound was detected in a specific sample/assay. A compound is defined as the integration of multiple metabolite features and the peak area for the compound is the sum of the peak areas for all features. During processing, MS1 data was assigned molecular formulae and matched to a compound annotation by using the Chemspider node to search multiple databases such as HMDB, LipidMaps and Chemspider. MS/MS data were searched against mzCloud (Thermo Scientific) and in-house MS/MS mass spectral libraries constructed with authentic chemical standards[201]. All metabolites were annotated to level 2 or level 3 as defined by the Metabolomics Standards Initiative[208]. The Compound Discoverer ™ workflow and parameters used can be found in the appendix (Figures A1-15). To ensure the data used for statistical analysis and biological interpretation were reproducible, data quality checks were conducted. Any samples with a lower number of compounds detected (in comparison to other samples at the same time point, the difference was greater than 20%) were removed as outliers. Data for the first nine QC samples that are used to equilibrate the instrument were removed. The data was then filtered based on precision, this was done by calculating the relative standard deviation (RSD) for the reported responses for the remaining QC samples for each compound and removing data for any compounds with an RSD of ≥30%. Following this, the data was filtered on detection rate. Data for any compounds that were not present in ≥70% of the pooled QC samples were also removed. Finally, the data was then filtered based on percentage blank contribution. Data for compounds where the mean peak area in the blank samples was ≥5% of the mean peak area in the QC samples were removed.

2.8.4.1 Statistical Analysis

For statistical analysis, Metaboanalyst version 5.0 was used[209,210]. Metaboanalyst is a web-based platform that allows includes a number of simple yet widespread data analysis techniques that helps lead to informed biological interpretation to take place[211].

2.8.4.1.1 One-Way Repeated Measures ANOVA

The first statistical analysis that was completed was repeated one-way ANOVA (OWRMA). The data matrices were compounds (rows) and samples (columns). The first three rows of the matrix were always comprised of the same three rows: unique sample name, participant number, and time point. For each different research question asked, different data was used for the OWRMA. For question one, where each exercise protocol was separated, the time points where the points were samples were taken throughout the protocol. For question two and three, the time points were the matching time point

across each of the three exercise protocols, e.g., 1- HIIT time point 1, 2- MICT time point 1, 3- SIT time point 1. For question three, the samples included were only female or only male. Parameters were Data Filtering – none; Sample Normalisation – Normalisation by Sum; Data Transformation – None; Data Scaling – None. Due to repeated one-way ANOVA repeatedly testing the same data, false discovery rate (FDR) was used to adjust the p-values.

2.8.4.1.2 Fold Change Analysis

Fold change analysis was conducted on pairwise comparisons of each time point for each question as the one-way repeated measures ANOVA did not have any post-hoc analysis in the results. Parameters were Data Filtering – none; Sample Normalisation – Normalisation by Sum; Data Transformation – None; Data Scaling – None; Analysis type: Paired.

2.8.4.1.3 Paired Parametric T-Test

Paired parametric t-tests were conducted on pairwise comparisons of each time point for each question as the repeated one-way ANOVA did not have any post-hoc analysis in the results. Parameters were Data Filtering – none; Sample Normalisation – Normalisation by Sum; Data Transformation – None; Data Scaling – None; Analysis type: Paired.

2.8.4.1.4 Pathway Enrichment Analysis

Pathway enrichment analysis was conducted on the statistically significant metabolites for each question to see if there were any pathways that were significantly enriched. All possible annotations for compounds were entered into the software and matched where possible to KEGG, HMDB or PubChem IDs. Pathway analysis parameters were Visualisation method- scatter plot; Enrichment method – Hypergeometric test; Topology Analysis – Realtive-betweeness Centrality; and Reference metabolome – All compounds in the selected pathway library. The pathway library was the KEGG Homo sapiens library. To be considered a pathway of interest, the pathway required a match of three or more metabolites and a P-value <0.05.

2.9 METHODS APPLIED TO HEAT STRESS AND ACCLIMATION EFFECTS ON HUMAN METABOLISM

2.9.1 Sample Extraction

Sample extraction applied the extraction techniques for plasma and urine samples to be analysed using Lipidomic, aqueous reversed-phase and HILIC assays as described in section 2.4.1.1. Sample randomisation was conducted as block randomisation where the participants were randomised and then all samples for that participant were randomised. QC samples were created using the method 2.4.1.3.1 and blank samples were created using the method described in 2.4.1.4.1.

2.9.2 UHPLC-MS Analysis

Analysis of the samples using UHPLC-MS was conducted using similar methods as applied in section 2.5.1. Along with MS1 data being collected MS2 data were also collected from QC samples to obtain further structural information about the metabolites present in the samples. Sample randomisation for analysis was conducted again after extraction. Block randomisation was completed where the participants were randomised and then all samples for that participant were randomised.

2.9.2.1 C_{18} Reversed-Phase Lipidomics (Lipids) Assay

Plasma samples were analysed in both positive and negative ion mode using the method described in section 2.8.3.1.

2.9.2.2 C_{18} Aqueous Reversed-Phase (RPaq) Assay

Urine samples were analysed in both positive and negative ion mode on both ID columns. The column used was a Thermo Scientific™ Hypersil GOLD™ aQ C_{18} (175 Å, 1.9 µm, 2.1 mm X 100 mm) The method used for the C_{18} aqueous reversed-phase (RPaq) assay was largely the same as the method written in section 2.5.1.2 but with a different elution gradient applied. Instead, the elution gradient used was 0 to 2 minutes 99%A; 2 to 10.5 minutes 50%A with a curve = 5; 10.5 to 11 minutes 1%A with a curve = 8; 11-11.5 minutes 99%A with a curve =5; 11.5-15 minutes 99%A with a curve of 5. The UHPLC parameters were flow rate = 0.300mL.min-1 for 0-11 minutes, 0.4ml/min for 11-14.9 minutes, and 0.300mL.min-1 for 14.9-15.0 minutes. The column temperature was 50°C and the injection volume was 2µL. Mass spectrometer settings were scan range = 50-750 m/z, sheath gas flow rate of 30, aux gas flow rate = 13, sweep gas flow rate = 0, spray voltage (kV) = 3.2 (positive ion mode) and 2.7 (negative ion mode), capillary temperature = 350°C, S-lens RF level = 40 and aux gas heater temperature = 400°C. MS/MS data collection was completed using data dependent acquisition (DDA) in “discovery mode”. Parameters for this were: mass resolution = 17,500 (FWHM at m/z 200); isolation width = 3.0 m/z; stepped normalised collision energies (stepped NCE) = 20, 50, 100 for positive ionisation mode and (stepped NCE) = 40, 60, 130 for negative ionisation mode. Spectra were acquired in three different mass ranges: 50–250 m/z; 200–500 m/z; 500–750 m/z. The method was set using ‘Discovery mode’ with scan event cycle of 1 Full MS and 3 DDA MS/MS scans (TopN = 3).

2.9.2.3 Hydrophilic Interaction Liquid Chromatography (HILIC) Assay

Both Urine and plasma samples were analysed in both positive and negative ion mode using the method described in section 2.8.3.2.

2.9.3 Raw Data Processing, Metabolite Annotation

Details for data processing and metabolite annotation can be found in section 2.8.4.

2.9.4 Statistical Analysis

For statistical analysis, Metaboanalyst version 5.0 was used[209,210]. The statistical analysis applied was similar to that used in Section 2.4.1 with any changes in protocol made clear below.

2.9.4.1 One-Way Repeated Measures ANOVA

Parameters applied to the one-way repeated measures ANOVA (OWRMA) were the same as in section 2.4.1.1. Three different OWRMA were conducted, one for the heat stress test samples across time, one for the hypoxic tolerance tests across time, and one for the heat acclimation sessions across time. This was repeated for both sample types separately.

2.9.4.2 Fold Change Analysis

Fold change analysis was conducted in the same way as in section 2.4.1.2, on pairwise comparisons of each time point for of the three different research areas, separately for each samples type. This was conducted as the one-way repeated measures ANOVA did not have any post-hoc analysis in the results.

2.9.4.3 Paired Parametric T-Test

Paired parametric t-tests were conducted in the same way as in section 2.4.1.3, on pairwise comparisons of each time point for each of the heat stress tests, hypoxic tolerance tests and heat acclimation sessions, as the one-way repeated measures ANOVA did not have any post-hoc analysis in the results. This was conducted separately for each of the sample types.

2.9.5 Pathway Enrichment Analysis

Pathway enrichment analysis was conducted on the statistically significant metabolites for each question to see if there were any pathways that were significantly enriched. All possible annotations for compounds were entered into the software and matched where possible to KEGG, HMDB or PubChem IDs. Pathway analysis parameters were Visualisation method- scatter plot; Enrichment method – Hypergeometric test; Topology Analysis – Realtive-betweeness Centrality; and Reference metabolome – All compounds in the selected pathway library. The pathway library was the KEGG Homo sapiens library. To be considered a pathway of interest, the pathway required a match of three or more metabolites and a P-value <0.05.

Chapter Three

3 DEVELOPMENT OF MICROFLOW ULTRA HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY – MASS SPECTROMETRY METABOLOMIC ASSAYS FOR ANALYSIS OF MAMMALIAN BIOFLUIDS

Note: The majority of research reported in this chapter of being prepared as a manuscript for peer reviewed publication in Metabolomics journal.

Annie J. Harwood-Stamper, Caroline A. Rowland, and Warwick B. Dunn; Development of Microflow Ultra High-Performance Liquid Chromatography -Mass Spectrometry Metabolomic Assays for Analysis of Mammalian Biofluids

3.1 INTRODUCTION

Within metabolomic studies involving mammals, including humans, there is a vast range of biological sample types that are used. As discussed in Section 1.6 of Chapter One, common sample types include (but are not limited to) plasma[212], serum[213], whole blood[214], cerebrospinal fluid[215], saliva[167], kidney tissue[162], liver tissue[163], heart tissue[164], adipose tissue[165], urine[36], muscle tissue[166], hepaRG cell lines[171], and tear fluid[216]. Serum, plasma, and urine are the most commonly studied human-derived samples. These sample types are analysed using numerous different analytical platforms as no single platform would allow detection and quantification of an entire metabolome.

The volume or mass collected for these samples is typically not limited, as sufficient sample volume or mass can be collected from the study participants and therefore assay sensitivity is not normally a concern. However, some sample types are produced in smaller volumes or masses and therefore the assay sensitivity for untargeted assays have to be considered and include as examples tear fluid (5-10µL) and cell 3D spheroids (>10^4 cells)[217]. Additionally, the collection of biological samples of low volume or mass can provide several advantages including less invasive sampling, ability to gain samples outside of a clinical setting, reduce risk assessments, and provide opportunities for larger cohort studies

with less costs involved (e.g., dried blood spot collection (5-20µL collected) instead of venous blood collection (3-10mL collected)).

For untargeted metabolomics studies of low mass or volume samples, the development of more sensitive analytical assays to allow for the maximisation of the number of metabolites as detected, to maximise metabolome coverage is a priority. Although technically difficult for low volume or mass samples, this can be achieved through pre-concentration of metabolites during sample collection and preparation. Another strategy that can be applied is the use of liquid chromatography (LC) columns of smaller internal diameters compared to the analytical LC columns of 2.1mm internal diameter which are most routinely applied in metabolomics studies. Ultra-High Performance Liquid Chromatography-Mass Spectrometry (UHPLC-MS) is the most frequently applied analytical platform for untargeted metabolomics applications[84,218,219]. UHPLC has many advantages over the more traditionally applied HPLC as a result of using smaller diameter stationary phase particles (<2.5µm compared to 3 or 5 µm) and higher flow rates/velocities. As is discussed in detail in Chapter One, Section 1.2.2, using smaller particles and higher flow velocities helps to reduce eddy diffusion which subsequently reduces peak broadening, leading to a higher chromatographic resolution and a potentially higher sensitivity[220,221]. Different stationary phases can be applied in a complimentary manner to increase the number of metabolites detected; for example, HILIC and C_{18} reversed-phase assays prominently allow for the detection of water-soluble and lipid metabolites, respectively. Although it should be noted that lipids can be detected in HILIC assays and water-soluble metabolites detected in reversed-phase assays. (UHP)LC-MS uses four different classes of column based on their internal diameter: standard/analytical (≥2.1mm ID), microbore (1.0-2.1mm ID), capillary (0.1-1.0mm ID) and nano (0.075µm-0.1mm ID) [222-224].

The use of microbore columns instead of analytical columns can, theoretically, increase sensitivity. By decreasing the column ID and maintaining the same flow velocity, an increased sensitivity is typically observed in cases where the injection volume is not too high so as to cause column overloading[225]. However, a lower flow velocity for a smaller ID column, when compared to a larger ID column, can result in greater eddy diffusion, and reduce or negate any increase in sensitivity. The use of lower flow velocities can also impact on peak broadening in the LC or electrospray ionisation source if appropriate ID tubing and a minimisation of dead volumes is not considered.

In previous research, it has been shown that the use of 1.0mm ID UHPLC columns showed promise in providing assays of appropriate sensitivity and reproducibility. Studies describing the use of a 1.0mm ID column in metabolomics research include the investigation of *Escherichia coli* (*E. coli*)[226], dystrophic retinas in rat models[227], dried blood spots collected from mice[228], rat brain tissue [229] and urine[230]. Only a small number of studies have been reported which perform a direct comparison between

2.1mm and 1.0mm ID columns for untargeted metabolomics assays, though these assays applied a rapid three-minute chromatographic separation [231-234].

3.1.1 Research Objectives

The purpose of the study reported here was to:

- Develop three complementary UHPLC assays applying 1.0mm ID UHPLC columns and to subsequently compare their applicability to analytical UHPLC assays applying 2.1mm ID columns with the same stationary phase type and particle size.
- Apply the newly developed assays to commonly analysed mammalian biofluids (plasma and urine) as well as low volume samples (tear fluid samples <5µL) for assessment of the applicability of 1.0mm ID columns for untargeted metabolomics analysis of low volume samples.
- Apply the newly developed assays for the characterisation of tear fluid samples.

3.2 MATERIALS AND METHODS

Methods for this work can be found under section 2.1 in the methods chapter.

3.3 RESULTS AND DISCUSSION

3.3.1 Method Development Using Human Plasma And Urine Samples

Development of three complementary UHPLC assays (HILIC, RPaq and lipidomics reversed phase) applying 1.0mm ID UHPLC columns was conducted using human plasma and urine samples. These assays were then subsequently compared on their applicability to analytical UHPLC assays applying 2.1mm ID columns with the same stationary phase type and particle size.

3.3.1.1 Reproducibility

PCA scores plots were used to visualise the distribution of the data and identify any trends or outliers. As can be seen in Figures 3-1 to 3-3 below, as well as in appendices C1-C8, there is a clear clustering between different sample dilutions across all sample types, assays, and column IDs.

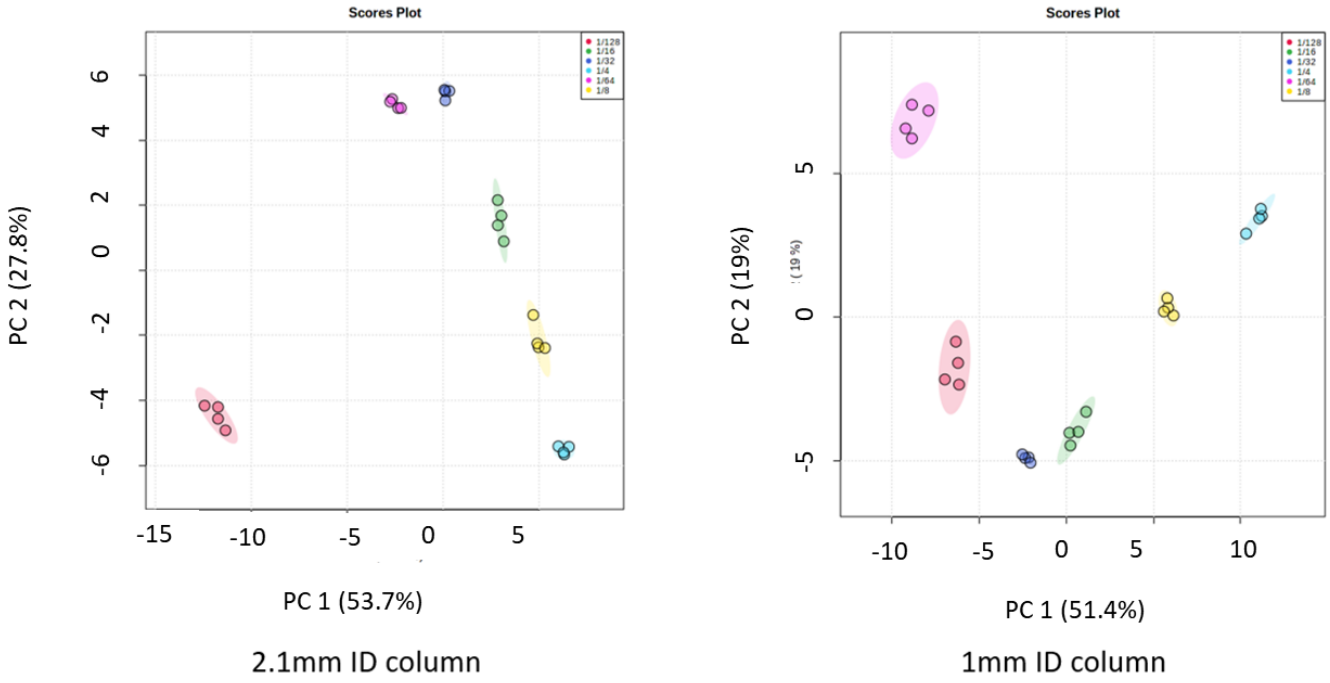


Figure 3-1: PCA Scores plots to show the variation between different dilutions when applying the 2.1mm ID column (left) or 1.0mm ID column (right) to lipidomic assays in negative ion mode using plasma samples

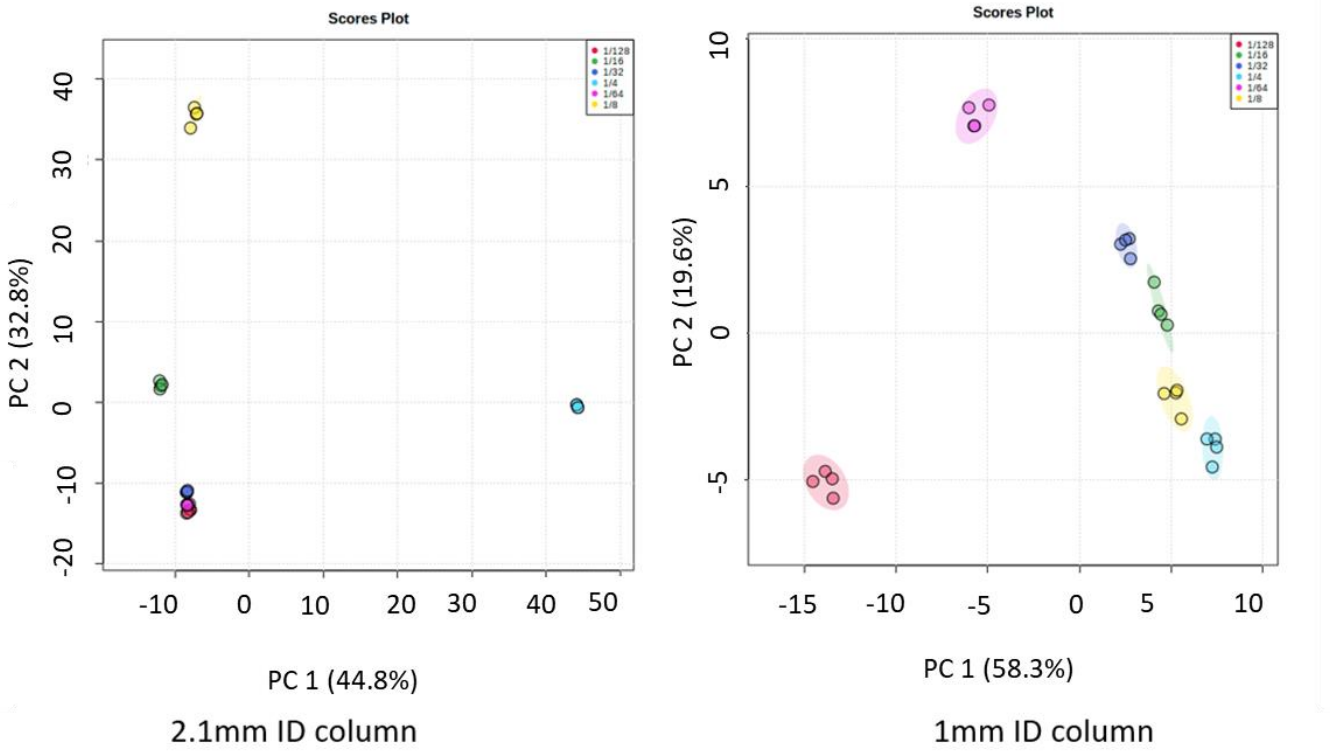


Figure 3-2: PCA Scores plots to show the variation between different dilutions when applying the 2.1mm ID column (left) or 1.0mm ID column (right) to lipidomic assays in negative ion mode using plasma samples

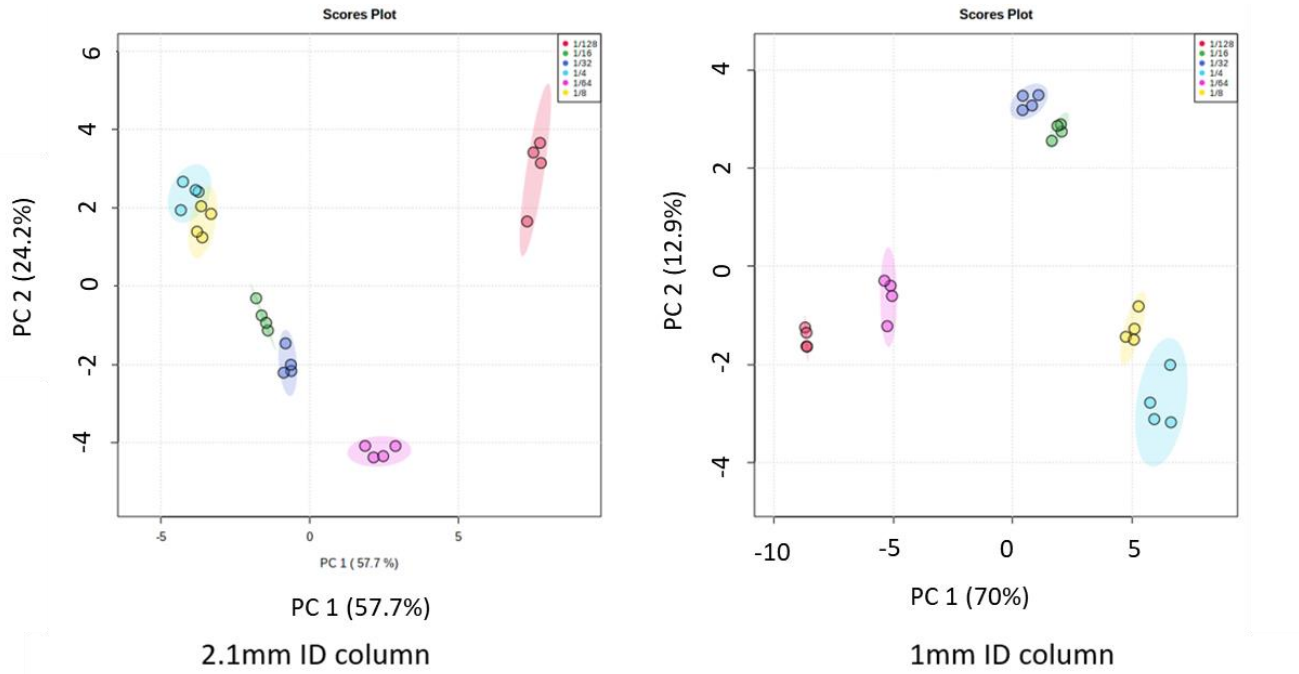


Figure 3-3: PCA Scores plots to show the variation between different dilutions when applying the 2.1mm ID column (left) or 1.0mm ID column (right) to reversed phase aqueous assays in negative ion mode using urine samples

The PCA plots for both the 2.1mm and the 1.0mm ID columns show similar levels of clustering and variance between sample dilutions for all three assays with clear differences between the sample dilutions. The only outlier to this was the lipidomics assay where the 2.1mm ID assay did not have a lot of variants between the 1/128, 1/64 and 1/32 dilutions. Relative standard deviations (RSD) for each compound were calculated at each dilution for each assay and sample type and can be seen below in Table 3-1. These demonstrated similar RSD ranges when comparing 2.1mm ID column data to 1.0mm ID column data. It can be concluded that a similar level of reproducibility for both column types is achievable.

Table 3-1: RSD range in each Dilution for each assay before filtering

Assay	**1/4**	**1/8**	**1/16**	**1/32**	**1/64**	**1/128**
2.1 Plasma HILIC NEG	0.27-83.4	0.20-98.2	0.14-111	0.52-84.5	1.0-79	1.93-55
1.0 Plasma HILIC NEG	0.23-136	0.28-138	0.19-123	0.006-126	0.65-68	0.075-136
2.1 Plasma HILIC POS	0.02-104	0.21-128	0.15-119	0.11-110	0.18-115	0.36-127
1.0 Plasma HILIC POS	0.08-107	0.003-122	0.13-119	0.24-139	0.31-87.9	0.31-109
2.1 Urine HILIC NEG	0.19-87.8	0.12-97.9	0.30-126	0.40-99.8	0.47-42.1	0.44-48.5
1.0 Urine HILIC NEG	0.1-122	0.02-143	0.09-104	0.1-99	0.55-105	0.7-112
2.1 Urine HILIC POS	0.21-140	0.17-137	0.23-133	0.21-169	0.07-99.8	0.05-115
1.0 Urine HILIC POS	0.003-131	0.19-114	0.23-124	0.21-144	0.2-134	0.2-132
2.1 Plasma LIPID NEG	0.21-140	0.17-137	0.23-133	0.21-170	0.07-99.8	0.05-115
1.0 Plasma LIPID NEG	0.02-192	0.12-194	0.31-193	0.07-168	0.09-191	0.04-166
2.1 Plasma LIPID POS	0.1-141	0.16-166	0.02-160	0.02-159	0.02-172	0.02-188
1.0 Plasma LIPID POS	0.13-185	0.08-196	0.18-154	0.003-158	0.52-150	0.43-190
2.1 Urine RPaq NEG	0.06-73.2	0.03-77.4	0.61-68	0.57-116	0.54-30.7	0.08-64.1
1.0 Urine RPaq NEG	0.45-86.5	0.85-84.0	0.91-68.9	0.46-49.3	1.34-85.1	0.44-37.3
2.1 Urine RPaq POS	0.08-120	0.02-95.4	0.39-136	0.29-149	0.07-117	1.00-127
1.0 Urine RPaq POS	0.11-139	0.20-120	0.07-112	0.39-114	0.04-129	0.65-129

3.3.1 Comparison Of The Number Of Detected Compounds

The number of compounds detected in each dilution for the same sample type and different column IDs were compared, results for which can be seen in Figure 3-4. The number of compounds reported in each dilution decreased as the dilution factor increased for both the 1.0mm and 2.1mm ID columns, defining that dilution effects were observed as expected, although this decrease was not always linear.

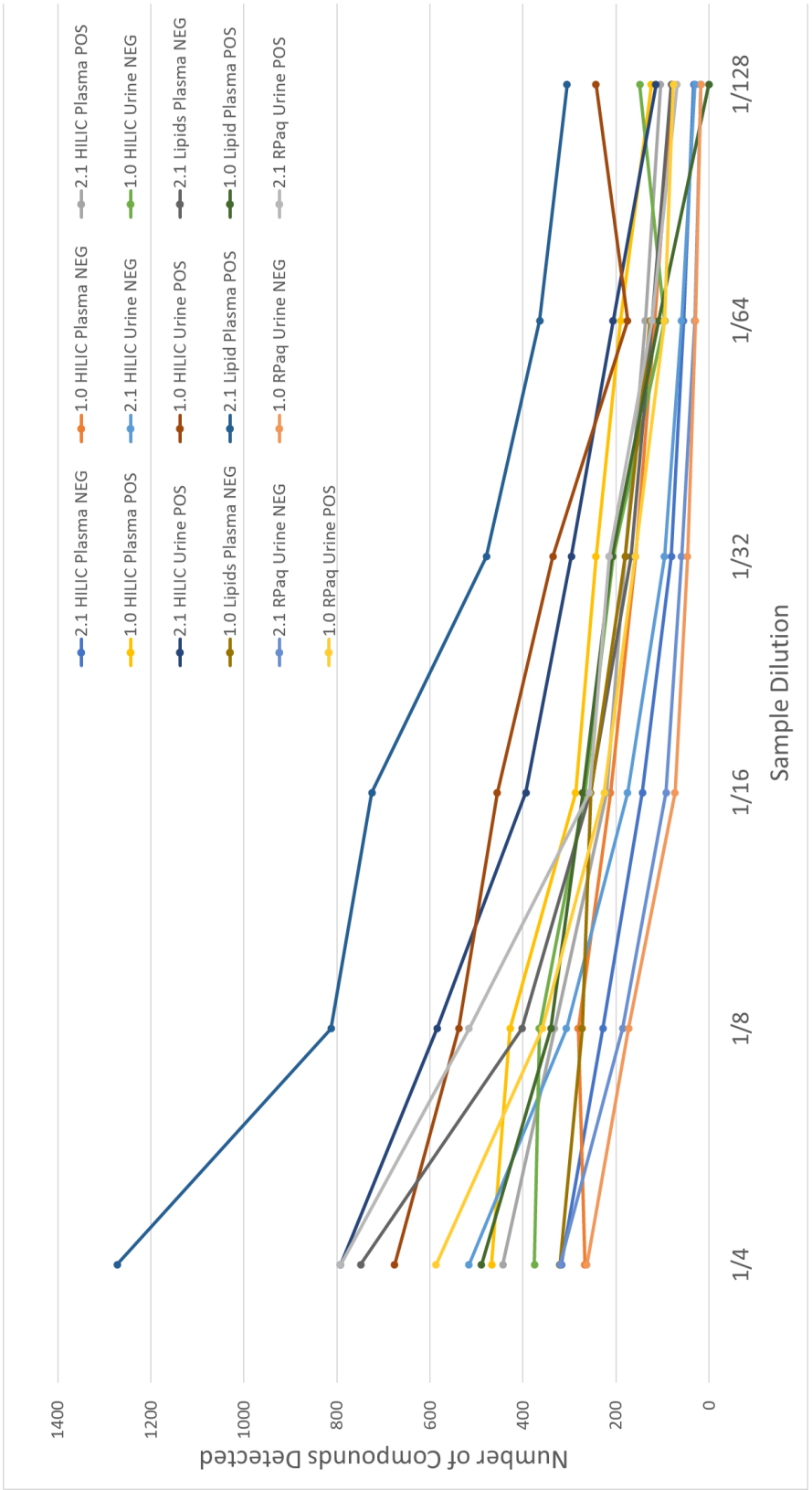


Figure 3-4:The number of reported compounds for each dilution in each assay. All compounds reported have a FWHM <0.2 minutes, were present in ¾ of injections and RSD <30% in each separate injection. The numbers in table format can be seen in appendix Figure D-1

When standard sample dilution methods (1 in 4 dilution) were investigated, the 1.0mm ID assays did not detect as many compounds when compared to data collected for the 2.1mm ID assays. There was one exception, the HILIC positive ion mode assay which for plasma detected a significantly higher number of compounds for the 1.0mm ID column (467 compared to 443 compounds). This can be seen in Figure 3-5 below.

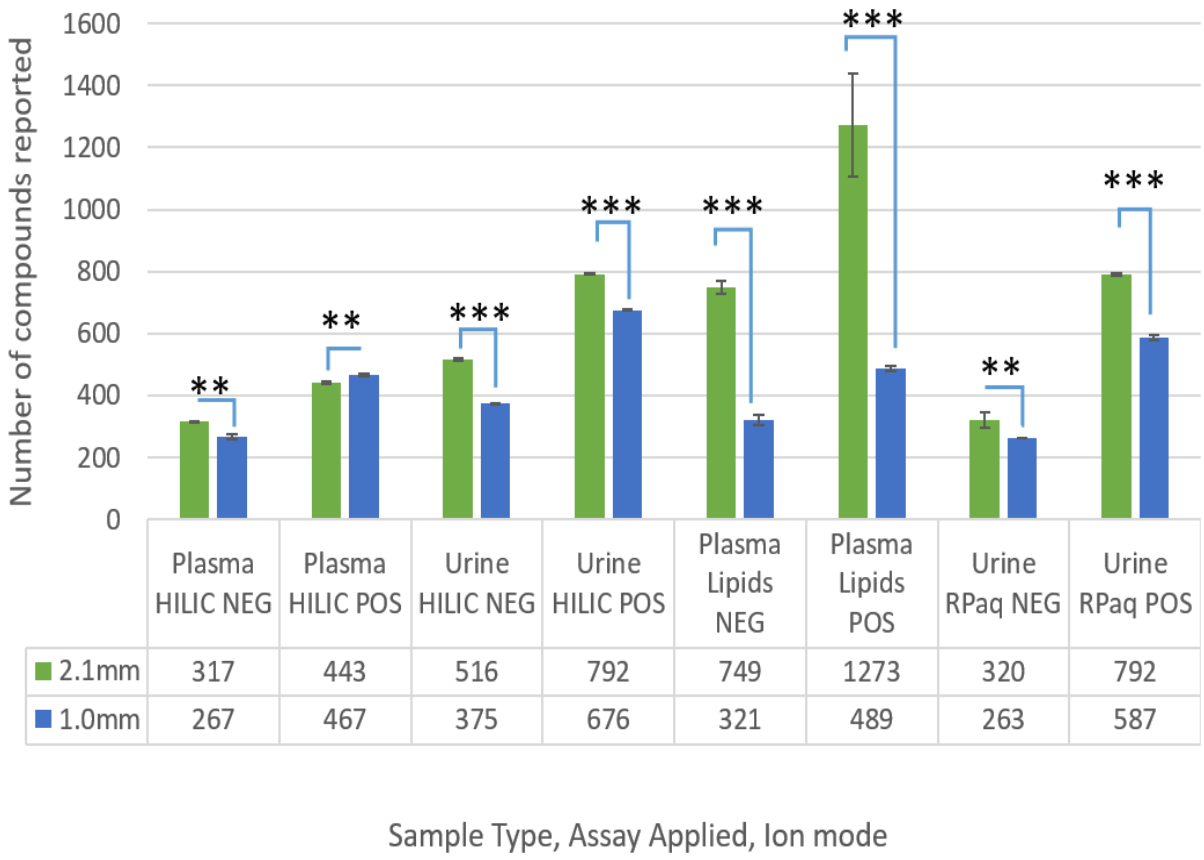


*Figure 3-5: The average number of reported compounds in human plasma and urine samples following a 1 in 4 dilution for each of three UHPLC-MS assays. Compounds which were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. 2.1mm ID column (green) and 1.0mm ID column (Blue) *** has a significance of <0.01, ** has a significance of <0.05.*

At higher dilutions no clear trend was observed across the three assays compared. To demonstrate this, data for the 1 in 16 dilution investigations were visualised in Figure 3-6 and the complete set of results are available in D2-16 in the appendix.

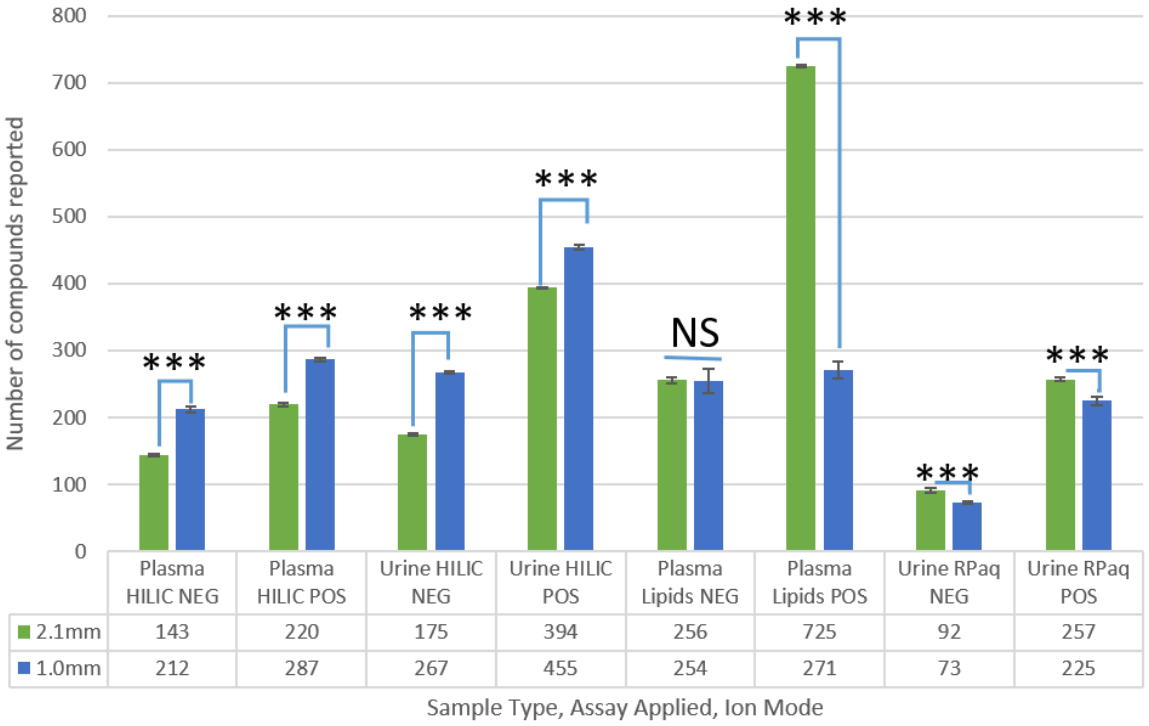


*Figure 3-6: The average number of reported compounds in human plasma and urine samples following a 1 in 16 dilution for each of three UHPLC-MS assays. Compounds which were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. 2.1mm ID column (green) and 1.0mm ID column (Blue) *** has a significance of <0.001, NS has no significance.*

The 1 in 16 dilution data (Figure 3-5) showed that all the HILIC assays for both plasma and urine detected significantly more compounds in the 1.0mm ID assay than the 2.1mm ID assay and this conclusion was observed for the 1 in 8 dilution data also for plasma (positive and negative ion mode) and urine (negative ion mode). This suggests that the 1.0mm ID HILIC assays could operate at higher sensitivity at larger sample dilutions which was what was theoretically expected. This trend in the HILIC assays continues with all further dilutions (1 in 32 to 1 in 128) containing more detected compounds in the 1.0mm ID assay compared to the 2.1mm ID assay. Both the lipidomics assays and reversed-phase aqueous assays, which both apply C_{18} columns but from different commercial sources and with different particle sizes, did not follow the same trend for detected compounds as for the HILIC assays. Instead, for these assays the 2.1mm ID assay detected either a similar number of compounds or more compounds in most of the sample dilutions studied. Overall, in relatively concentrated samples applied in typical metabolomic studies, using 1mm ID columns does not increase sensitivity whereas for more diluted samples the use of 1.0mm ID columns may appear more appropriate.

3.3.1.1 Chromatographic Peak Widths And Associated Chromatographic Resolution

As the expected increase in sensitivity was not observed for the 1.0mm ID assays compared to the 2.1mm ID assays, further investigations were performed to determine why. The full width half max (FWHM) chromatographic peak widths of all reported compounds were investigated to determine the effect of column internal diameter and the associated change in flow rate on chromatographic peak widths (Figure 3-7).

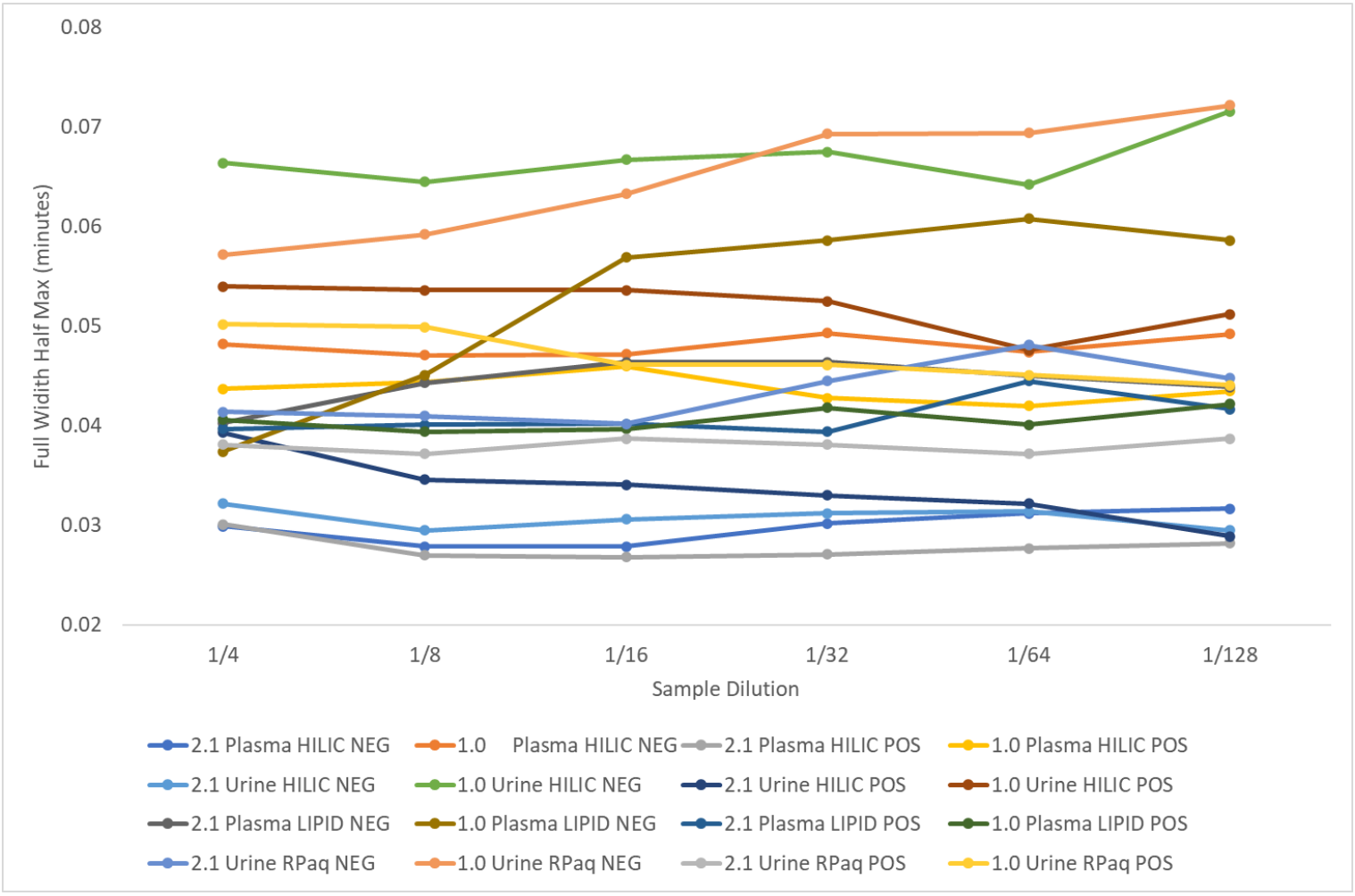


Figure 3-7: The average full width half max in minutes for each sample dilution in each assay. Compounds used were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. NEG = negative ion mode, POS = positive ion mode, SD = standard deviation.

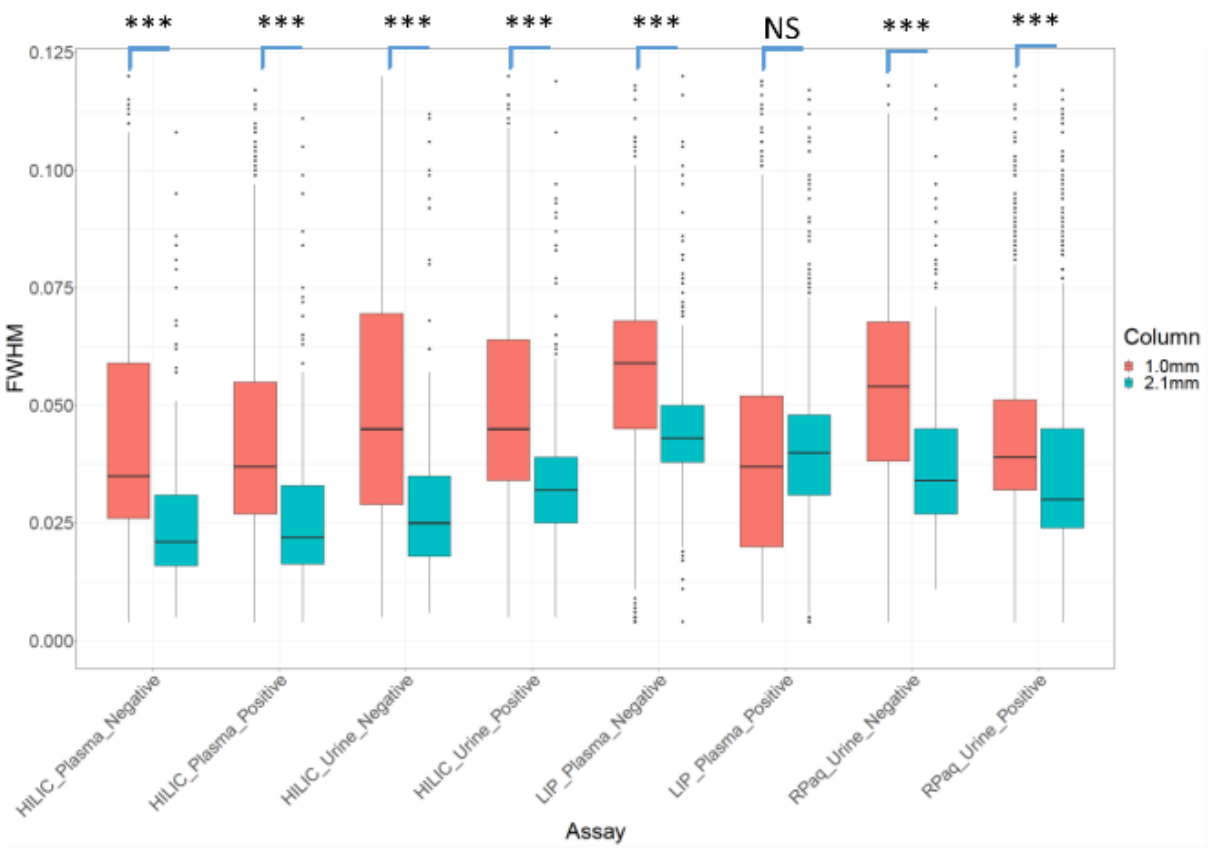


*Figure 3-8: The distribution of chromatographic peak widths in minutes for the 1 in 16 dilutions of each assay type for both the 2.1mm ID (blue) and 1mm ID (red) columns. Compounds which were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. *** has a significance of <0.001, NS has no significance.*

These trends were observed across all dilutions (as seen in appendix Figures E1-24) and so it can be concluded that the peak widths in the 1.0mm ID assays are larger than observed in the 2.1mm ID assays. There are many reasons why this could be occurring and the most likely is that these results are observed because of the flow rate and associated flow velocities applied for the 1.0mm ID columns. UHPLC chromatographic resolution is influenced by the flow velocity with lower flow velocities typically being sub-optimal for a given column ID and length. The smaller ID column was operated at lower flow rates to allow operation below the UHPLC pressure limit of 1200 BAR. The flow velocities for both ID column were calculated using:

$$\frac{\text{Volumetric flow rate } (\text{L.hr}^{-1})}{\left(\frac{\text{column crossectional area}(\text{cm}^2)}{1000}\right)} = \text{flow velocity } (\text{cm.hr}^{-1}) \qquad \text{Eq. 10}$$

The 1.0mm ID column has a higher flow velocity (1019 cm.hour^{-1} for a 0.13 ml.min^{-1} flow rate) compared to the 2.1mm ID column (693 cm.hour^{-1} for a 0.40 ml.min^{-1} flow rate). Therefore, a lower flow velocity is not a potential reason for wider chromatographic peaks as a result of eddy diffusion as the flow velocity is 1.5 times higher for the 1.0mm ID column compared to the 2.1mm ID column. Another possible reason for wider chromatographic peaks is band broadening occurring at the time of injection as the same injection volume (2µL) was applied for both column IDs. Injecting the same volume on both columns would mean that the smaller ID column would have an approximately 4.4 times wider band width just after injection than for the 2.1mm ID column. With the difference at detection being a lot smaller than the above, this suggests that the combination of column ID and the flow velocity is having an impact on the reducing the wide band that was injected.

For all assay-sample type combinations there were no significant changes in peak width as sample dilution increased, with one exception, the lipids negative assay applied to the analysis of plasma which showed the same peak width for the 2.1mm assay for all dilutions whereas the peak widths increased for the 1.0mm assay as sample dilution increased.

3.3.1.2 Chromatographic Peak Areas

As an alternative investigation into why an increase in sensitivity was not observed for the 1.0mm ID assays compared to the 2.1mm ID assays, peak areas were analysed. The distribution of peak areas was investigated to determine whether lower numbers of reported compounds was a result of lower reported peaks areas. Each assay displayed a small drop in average peak area as the dilution of the sample increased (Figure 3-9 below). Data for the 1 in 16 dilutions for all samples and assays is shown in Figure 3-10.

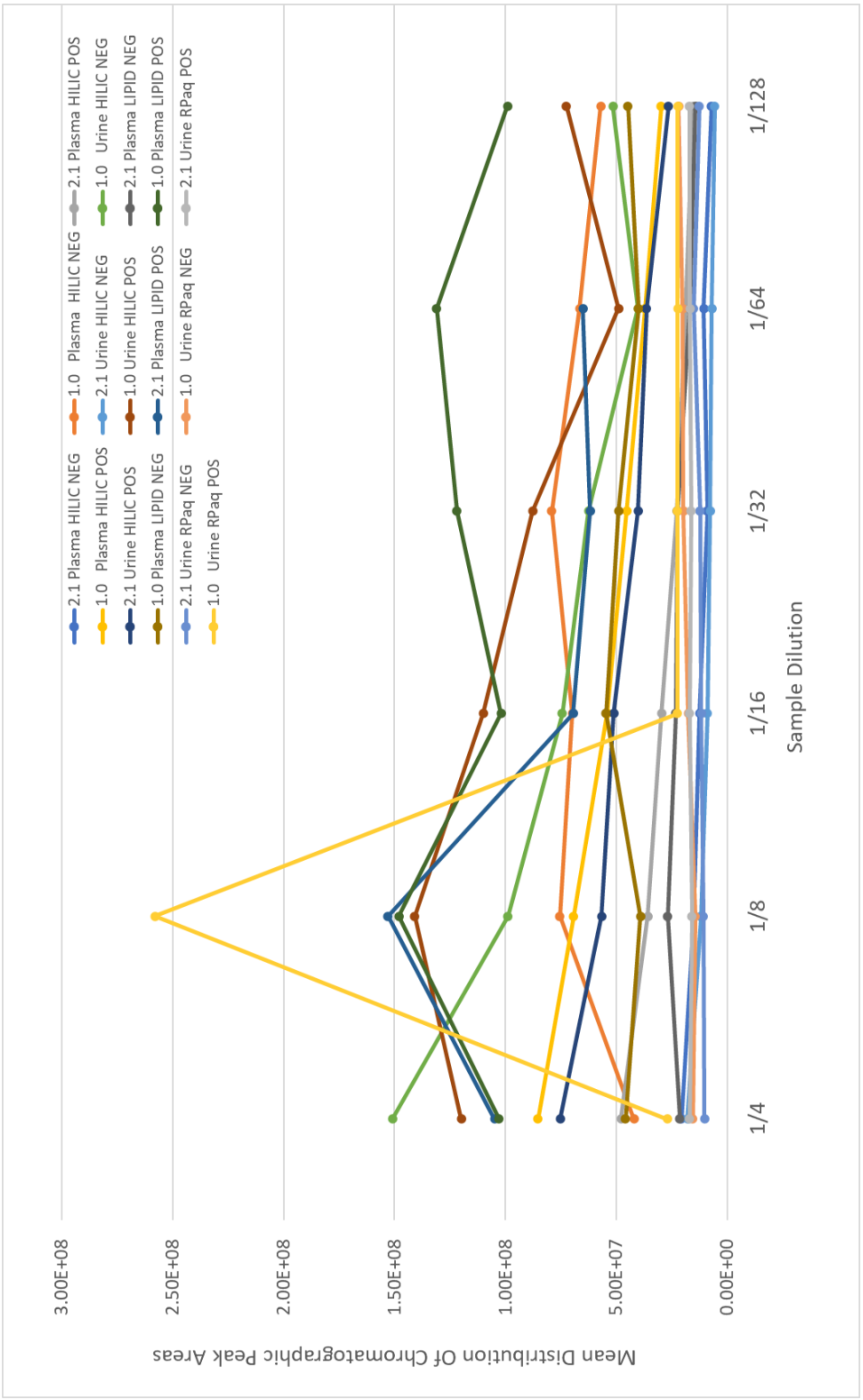


Figure 3-9: A summary of the mean distribution of chromatographic peak areas for all dilutions of 2.1mm ID vs. 1mm ID for all assays and sample types. Compounds used were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. NEG = negative ion mode, POS = positive ion mode, SD = standard deviation. A table displaying the information can be found in the appendix under E-1.

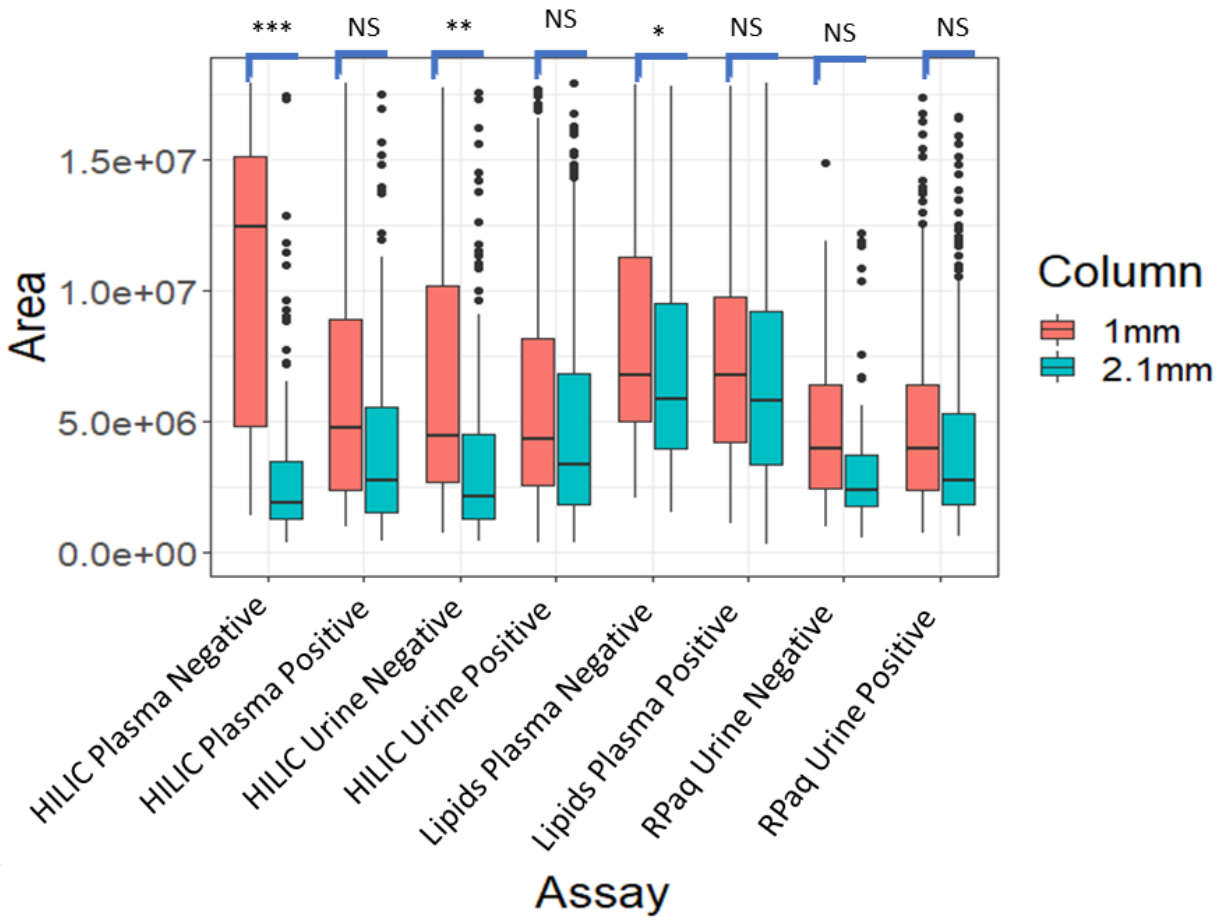


Figure 3-10: The distribution of chromatographic peak areas in the 1 in 16 dilutions of each assay type for both the 2.1mm ID (blue) and 1mm ID (red) columns. Compounds which were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. Most comparisons were not significant (NS) (P > 0.05), whilst one was significant (*) (P ≤ 0.05), one was moderately significant (**) (P≤ 0.01) and one was very significant (***) (P ≤ 0.001).

The results show that the median peak area is higher for the 1mm ID column of to the 2.1mm ID column for all assays and sample types. Although the distributions for the two column IDs overlap considerably, the peak areas for the 1.0mm ID columns are significantly higher when compared to the 2.1mm ID columns for HILIC assays (negative ion mode) and Lipidomics assay (negative ion mode). Most comparisons were not significant (NS) (P > 0.05), whilst one was significant (*) (P ≤ 0.05), one was moderately significant (**) (P≤ 0.01) and one was very significant (***) (P ≤ 0.001).

3.3.2 Application Of 1.0mm ID Columns To Study Different Biofluids

Newly developed assays were applied to commonly analysed mammalian biofluids (plasma and urine) as well as low volume samples (tears samples <5µL) for an assessment of applicability for 1.0mm ID columns to be applied for untargeted metabolomics analysis of low volume samples.

1.0mm ID columns were applied to analyse porcine plasma, urine and tear samples using three different assays. The lipidomics assay was applied to plasma and tear samples, the RPaq assay was applied to urine and tears, and the HILIC assay was applied to all three sample types. This research was conducted for (1) the characterisation of the metabolites present within the tear metabolome, (2) to assess the overlap between the tear, plasma, and urine metabolomes, and (3) to assess whether the 1.0mm ID assay could be applied to analyse small volume samples including tears. This would support research for whether tear fluid would be an appropriate sample type for metabolomics research. Results are available in Figures 3-11 and 3-12 along with Appendix Figures F-1 to F-4. All assays demonstrated that hundreds or low thousands of compounds can be detected in each sample type, similar to the numbers of compounds regularly detected in published untargeted metabolomic studies.

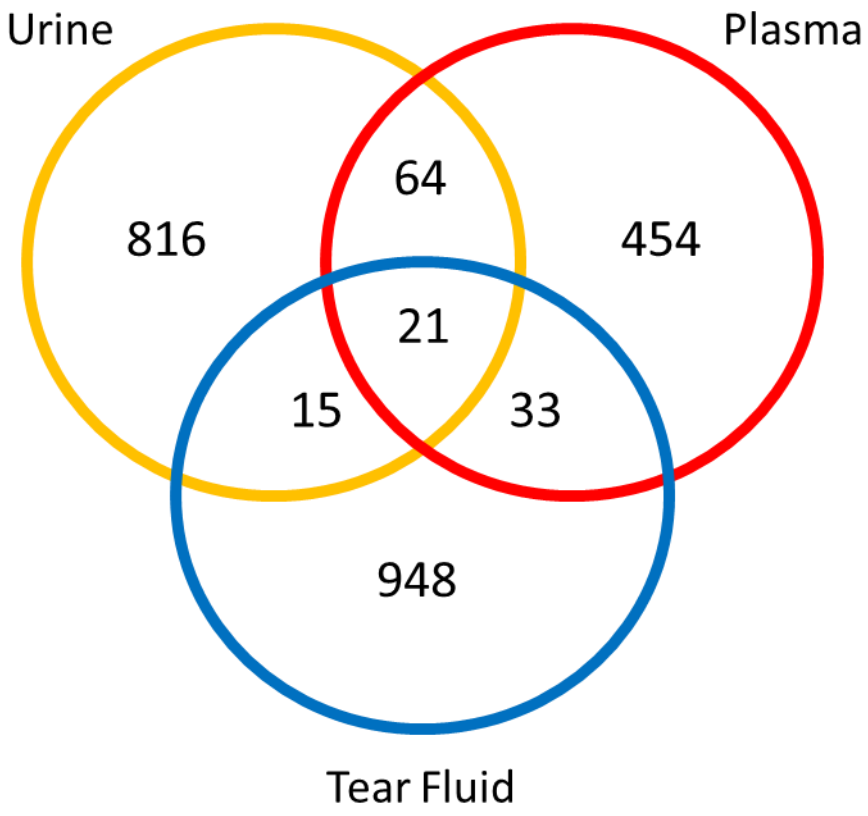


Figure 3-11: Overlap of porcine plasma (red), urine (blue), and tear fluid (yellow) compounds detected during analysis using HILIC assay in negative ion mode.

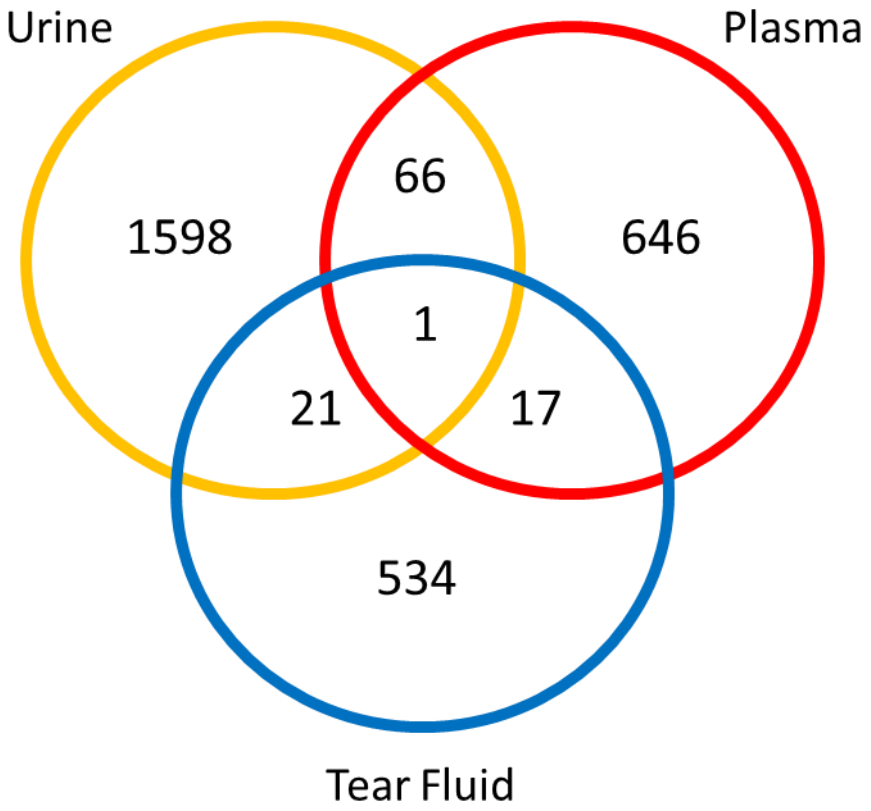


Figure 3-12: Overlap of porcine plasma (red), urine (blue), and tear fluid (yellow) compounds detected during analysis using HILIC assay in positive ion mode

Applying the HILIC assay to all three sample types showed both pairwise and 3-way overlaps of polar compounds. There was some concern that the urine -plasma overlap should have been larger than what was seen for both the positive and negative ion modes of this assay type. This is because on the human metabolome database (HMDB) there are 25372 plasma metabolites and 4237 urine metabolites of which 3668 metabolites are present in both metabolomes (though it should be noted that this includes lipid metabolites, which were not compared for the porcine plasma and urine during this research). For the non-polar assays, all showed a small overlap (<70) in compounds between the tear fluid and the plasma/urine, as can be seen in below in Appendix F-1 to F-4.

These results show hundreds of metabolites can be detected when applying the 1.0 mm ID assays and that the developed 1.0mm ID assays are therefore appropriate to be applied to untargeted metabolomic studies of low volume mammalian biofluids including low volume tear samples which were diluted during extraction by a factor 6.8 and then concentrated by a factor of 5 upon reconstitution for polar metabolites and a factor of 3 for non-polar metabolites.

The research also shows that for HILIC assays more metabolites are detected in tear fluid than the other biofluids for polar assays, which suggests that it is metabolite rich and so is an appropriate sample type to be studied in metabolomic studies. For both reversed-phases assays used to look at lipids and non-

polar compounds, there were less metabolites detected in tear fluid in comparison to the other sample types. This suggests that despite the two lipid layers in tears, that there were less lipids being detected.

As a there was only a small overlap between tear fluid and plasma, and tear fluid and urine, further data exploration was performed to investigate if there was a technical or biological hypothesis as to why these results were observed. Single ion chromatograms of common lipids were plotted to investigate the differences between tears and plasma, which showed that some of the commonly observed metabolites in plasma are not detected in tears. One possibility for the lack of detection could be due to significant differences in metabolite concentrations between the two sample types.

3.3.3 Putative Metabolite Annotations For Tear Fluid

Overall, there were 288 unique compounds annotated in tear fluid samples across all assays. To the best of my knowledge, this is still the largest coverage of the tear metabolome in an untargeted metabolomics study. In previous literature there has been limited characterisation of the metabolites present in tear fluid samples[152,216,235,236] as a lot of the data published has been targeted assays[237,238].

Despite there being multiple studies looking at the tear metabolome, currently the only on HMDB, there is only one metabolite known to be detected in tear fluid (H_2O). This shows that there is a lot of progress still to be made with tear fluid samples to make them easier to study and to help annotate metabolites detected with confidence.

3.4 CONCLUSIONS

The purpose of the study reported here was to:

- Develop three complementary UHPLC assays applying 1.0mm ID UHPLC columns and to subsequently compare their applicability to analytical UHPLC assays applying 2.1mm ID columns with the same stationary phase type and particle size.
- Apply the newly developed assays to commonly analysed mammalian biofluids (plasma and urine) as well as low volume samples (tears samples <5µL for assessment of the applicability of 1.0mm ID columns for untargeted metabolomics analysis of low volume samples.
- Apply the newly developed assays to commonly analysed mammalian biofluids (plasma and urine) as well as low volume samples (tears samples <5µL) for characterisation of tear fluid samples

Upon comparing the chromatographic peak widths and total peak areas, it was found that both were higher for the 1.0mm ID assays compared to the 2.1mm ID assays. UHPLC chromatographic resolution is influenced by the flow velocity with lower flow velocities typically being sub-optimal for a given column ID and length. The smaller ID column was operated at lower flow rates to allow operation

below the UHPLC column back pressure limit of 1200 BAR. We calculated the flow velocities for both ID columns and the 1.0mm ID column has a higher flow velocity (1019cm/hour for a 0.13ml/min flow rate) compared to the 2.1mm ID column (693 cm/hour for a 0.40ml/min flow rate). Therefore, non-optimal flow velocity is not a potential reason for wider chromatographic peaks as a result of eddy diffusion as the flow velocity is 1.5 times higher for the 1.0mm ID column compared to the 2.1mm ID column. Another possible reason for wider chromatographic peaks is band broadening occurring at the time of injection as the same injection volume (2µL) was applied for both column IDs. Injecting the same volume on both columns would mean that the smaller ID column would have an approximately 4.4 times wider band width just after injection than for the 2.1mm ID column.

The results from the 1.0mm and 2.1mm ID comparison showed that at commonly used dilutions, the 2.1mm ID assay was more sensitive than the 1.0mm ID assay. When applying higher dilutions during sample preparation as can be observed for small volume/mass samples, the HILIC assays resulted in an increase in the number of compounds detected (1 in 8 and larger dilutions), whilst the lipidomics and aqueous reversed-phase assays had a very similar number or a decrease in the number of compounds detected but an increase in peak areas when applying the 1.0mm ID columns. Therefore, in conclusion, although none of the 1mm ID assays show a true increase in sensitivity at a low dilution/high concentration, it can be said that when samples become more dilute, the 1.0mm ID HILIC assays would be suitable to apply and in most cases similar number of compounds were detected at higher dilutions when comparing 2.1mm ID and 1.0mm ID assays with the exception of the lipids positive assay where significantly greater numbers of compounds were detected for the 2.1mm ID assay when plasma was analysed. The reproducibility of 1.0mm ID assays were good and comparable with the 2.1mm ID assays for the same injection volume.

The 1.0mm ID assays were applied to analyse three pig biofluids and demonstrated their applicability for untargeted metabolomic studies. Importantly, the low volume tear samples (volume less than 20µL) which were diluted during extraction by a factor 6.8 and then concentrated by a factor of 5 upon reconstitution for polar metabolites and a factor of 3 for non-polar metabolites in combination with the 1.0mm ID column assays provided the detection of hundreds of metabolites. This led to the largest number of putatively annotated compounds in tear fluid to date, although further work applying *MS/MS* techniques to ensure the annotations are of high confidence should be conducted. In conclusion, the application of 1.0mm ID columns is recommended for the analysis of highly diluted samples where a similar number of or more compounds are typically detected. However, for samples where > 50µL volumes are available and where sample dilutions are ≥ 1 in 4 then 2.1mm ID column assays are recommended because they provide detection of a similar number or greater number of compounds.

3.5 FUTURE WORK

Further work surrounding the characterisation of metabolites present in tear fluid samples would be advantageous to gain higher confidence metabolite annotations to characterise the metabolites present further, and aid in building more knowledge around the sample type in online databases. Applying these assays to other low volume/mass sample types would also be beneficial to display how well the methods work with a more diverse set of sample types. Further work into low flow UHPLC techniques would be to analyse samples on a capillary UHPLC instrument instead as there are a small number of scientific publications which have demonstrated this for metabolomics applications. Although, it has not been investigated whether large scale studies could be applied on capillary UHPLC instrument robustly and reproducibly. Further work should also be investigated into the injection volumes applied. For this study as many parameters between the two ID methods were kept as similar as possible to make sure any differences that were seen were purely from the change in column ID. Investigating alternative injection volumes would allow for potential increases in peak area if the 1mm ID column is indeed being overloaded. A limitation of this study was the sample types applied, all sample types were biofluids, so it has not been investigated as to how well tissue sample types would be analysed when applying these assays. Future work should be conducted to discover if the trends seen in biofluids are the same as tissue or whether these methods are more applicable to alternative sample types.

Chapter 4

4 IN-SILICO EXTENSION OF RT LIBRARIES FOR WATER-SOLUBLE METABOLITES AND LIPIDS WHEN APPLYING UHPLC-MS.

4.1 INTRODUCTION

When it comes to annotating features and identifying metabolites in a dataset for untargeted metabolomics research there are multiple levels of confidence that can be applied to the annotations given[102,103]. These levels of confidence are:

1. Identified metabolites
2. Putatively annotated metabolites
3. Putatively characterized compound classes
4. Unknown metabolites

A compound that has no annotation would be kept and classed under level 4. A compound classed under level 3 would be based on *m/z* only and characterised from an open database on molecular weight only. A compound annotated to confidence level 2 would require a structural match from an online spectral library (such as mzCloud™) or to data collected for chemical standards outside your research laboratory [103]. To be annotated to level 1 confidence would require two independent data matches (such as RT and *m/z* or RT and NMR spectrum) with matching data from standards collected inside your research laboratory. Data from standards must also be collected within the same laboratory under matching experimental conditions.

Meeting confidence levels 2 and 1 are challenging since databases containing information on metabolites that are used for metabolite annotation and identification are commonly either very small and undistinguished or too large, but not accurate enough for specific metabolomes when compared to other -omics methods (such as proteomics). This is not only because the research area is less established, or the fact that many metabolites are yet to be discovered, but also because many metabolites are endogenous and are not available to purchase as chemical standards. Therefore, any level 1 identifications are very advantageous because they are the highest confidence identifications that can

be made, and this is often difficult to accomplish for most metabolites detected in an untargeted metabolomics dataset.

One of the challenges that hinders the ability to perform biological interpretation from acquired untargeted metabolomics datasets is the structural annotation or identification of the metabolites detected.[239] Metabolite annotation and identification is necessary for biological interpretation to take place and therefore for the research to fill a knowledge gap. Unfortunately, there are many pitfalls in the current annotation and identification process that means it is difficult to derive many metabolites with high confidence annotations or identifications. These pitfalls occur for many reasons including:

1) Identification of metabolites is currently complex because it is dependent on the analytical platform applied. This means that identifications cannot be transferred between different types of analytical method or between instruments/methods of the same analytical method type.

2) For metabolites not in a library (with both RT and MS/MS match), full structural identification of metabolites requires a lot of work (time and money) as there could be 100s of metabolites that are seen as biologically important, and therefore need structural confirmation of a match against an analytical standard.

3) Analytical standards to compare metabolites to for structural identification can be costly for small amounts, and in many cases are unable to be purchased from commercial companies or to be synthesised.

4) Retention time (RT) and MS/MS libraries that data are matched to do not have full coverage of the metabolome, and so not all metabolites can be matched to data in these libraries.

5) We do not know the full composition for metabolomes at present. There are a number of unknown unknowns for each sample type metabolome.

6) MS1 data on its own cannot be used to distinguish isomers and does not produce a singular annotation.

7) MS/MS data can sometimes (but not always) distinguish isomers.

One way to overcome the issues addressed above is to include chromatographic separation of isomers and utilise RT to provide extra confidence in identifications made. RT is the time taken for a solute to travel through a chromatography column. This data is useful as metabolites of different structure will take different lengths of time to travel through the chromatography column. Thus, meaning that for metabolites with similar molecular weights but very different structures, identification would be difficult if only utilising MS1 data, but with retention time data also included in the identification

process becomes easier. This is because if the two metabolites could have different retention time values due to their interactions with the column stationary phase.

Currently, RT libraries are limited in size and global abundance when compared to mass spectral libraries. This is because they are expensive to create and are non-transferrable between laboratories. Using RT data has some complications to overcome, mainly in that RT differs between laboratories as methods in metabolomics are still far from standardized and vary greatly between research groups (differences include solvents, columns, gradient elutions, instrument dead volumes etc.). Due to this, RT libraries are mainly constructed 'in-house' using chemical standards for each research group and are not shared outside of the research group unless identical assays are applied. Current research into ways to overcome this have been to predict retention time in-silico[104-106,240-246], this has been done with varying levels of success[107]. Several of these methods of retention time prediction used a quantitative structure retention relationship (QSRR) model[106,240,243,245],whilst others used a machine learning approach[104,105,241] One difference between retention time prediction and retention time mapping is that retention time mapping only predicts where data for chemical standards are available and retention time prediction can theoretically predict retention times for all metabolites. Research has also been conducted into retention order prediction as an alternative option[244]. Another alternative would be to try to map retention times[110-112]. So far this has been done using multivariate methods such as penalized constrained least-squares (PCLS) and generalized additive models[110], using regression modelling but only on one set of data[111] or by using mixture of machine learning models[112]. RT mapping is a lot less researched but provides some potential opportunities to better transfer retention time data between laboratories.

4.1.1 Research Objectives

The purpose of the study reported here was to:

- To assess whether RTs collected in two different laboratories applying aqueous reversed phase and HILIC assays can be mapped to allow transfer of RT data across libraries
- To develop an expanded RT library for use using standard methods applying multiple RT libraries.
- To develop a RT library for lipids for a specific assay/method using lipid annotations derived from MS1 and MS/MS and LipidSearch software (not from chemical standards)
- To assess whether RT elution trends can be modelled to provide RTs for lipids that are not present in the RT library
- To assess whether elution ranges for lipid classes can be applied to reduce the number of false annotations based on MS1 data

4.2 RESULTS AND DISCUSSION

4.2.1 Retention Time Mapping

4.2.1.1 Reversed Phase Assays

The first objective was to assess whether RTs collected in two different laboratories applying aqueous reversed phase and HILIC assays can be mapped to allow transfer of RT data across libraries. To do this, there had to be overlapping data for metabolites with RT data present in both datasets/libraries so that can be modelling could be achieved accurately. The number total number of metabolites in each dataset/library and the number of overlapping metabolites with other datasets/libraries can be seen in Table 4-1. After this, 10-fold cross-validation was conducted by separating 20 random overlapping metabolites and creating a model based on the remaining portion. The separated 20 were then used to test the model and see if the mapped metabolite retention times were accurate. This was repeated 10 times. If the model was accurate and reproducible, the model could be used to add in metabolites with RTs only available in one of the two libraries. As can be seen by the table below (Table 4-1), for some of the RT libraries there was potential to be expanded by almost 100%.

Table 4-1: The total number of unique metabolites within each dataset and/or library along with the number of overlapping metabolites between datasets for the reversed phase assays from different research groups. For example, the Mayo Clinic and University of Birmingham libraries/datasets have 676 and 536 metabolites, respectively but only eight metabolites which are present in both libraries/datasets.

	Mayo Clinic	University of Birmingham	University of Florida	Liverpool Method 1	Liverpool Method 2
Mayo Clinic	676				
University of Birmingham	8	536			
University of Florida	38	127	1727		
Liverpool Method 1	14	55	58	504	
Liverpool Method 2	14	55	58	504	504

4.2.1.1.1 University Of Liverpool (Method 1) Vs University Of Liverpool (Method 2)

The relationship between these two datasets hypothetically had the best chance for mapping as the only differences between the two methods was the column applied. This is reflected in the data below. Before filtering the data, there was a very strong relationship between the RTs of the two datasets, with the overlapping metabolites trend line having a high accuracy correlation coefficient value of 0.9843.

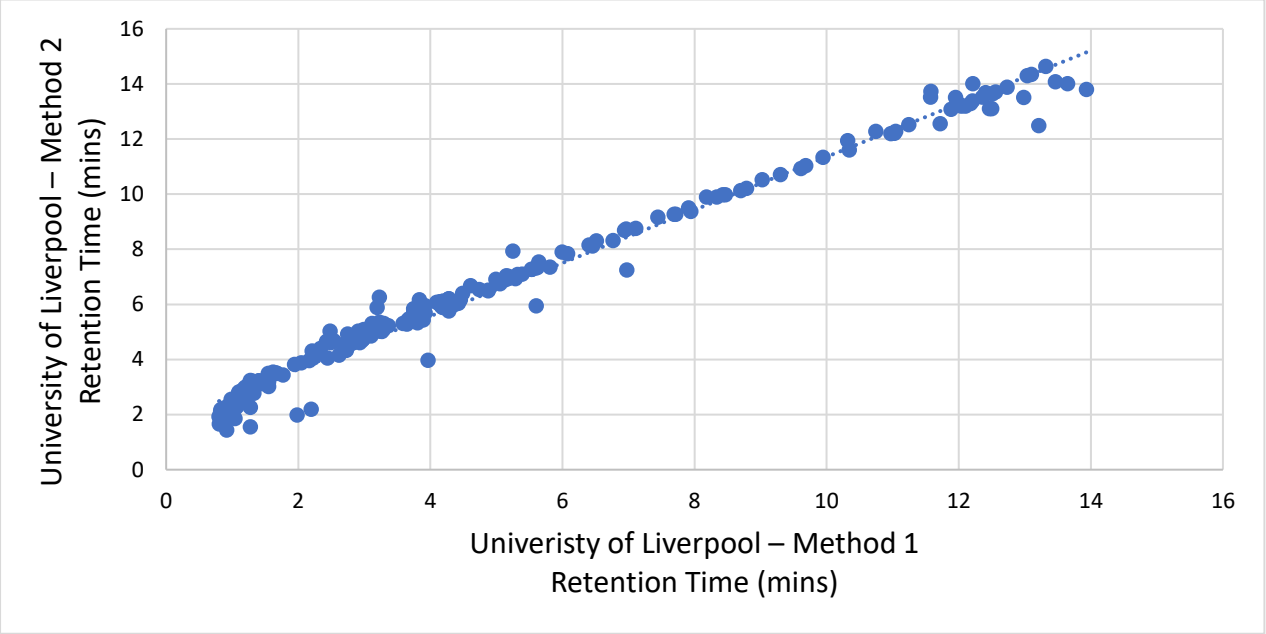


Figure 4-1: The relationship between retention times for the University of Liverpool method 1 dataset and the University of Liverpool method 2 dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time for both datasets was measured in minutes. Line of best fit: $y = 0.9639x + 1.7167$, $R^2 = 0.9843$

As the accuracy of this model was extremely high and there were 220 overlapping metabolites, 10-fold cross-validation models were completed to ensure that the data was reproducible (Appendix H-1).

Each validation model had a very high R^2 value (average of 0.9845) and low mean difference in retention time between the mapped retention time and the experimental retention time (average of 0.02401 minutes). This showed that no matter what metabolites were removed, the relationship between the retention times was always very robust.

This comparison of datasets shows that data collected using the same equipment in the same research group can be easily transferred across methods when method parameters such as columns are changed. This is useful as it allows for more variability in the methods used within research groups and means that internal libraries do not have to be reanalysed as different technology is developed.

4.2.1.1.2 University Of Birmingham Vs University Of Liverpool (Method 1)

The comparison of the dataset from the University of Birmingham and the University of Liverpool (method 1) allows for the data from two different analytical standards libraries to be compared. These libraries used methods that differed on column brand, column particle size, mobile phase, elution gradient, and column temperature. This would allow internal libraries from different groups to be transferred with little error and be used for higher confidence metabolite annotations to be made without the cost and time required to build large internal libraries.

The relationship between the two datasets was more complex, but a high accuracy trend line was still able to be deciphered ($R^2 = 0.9694$).

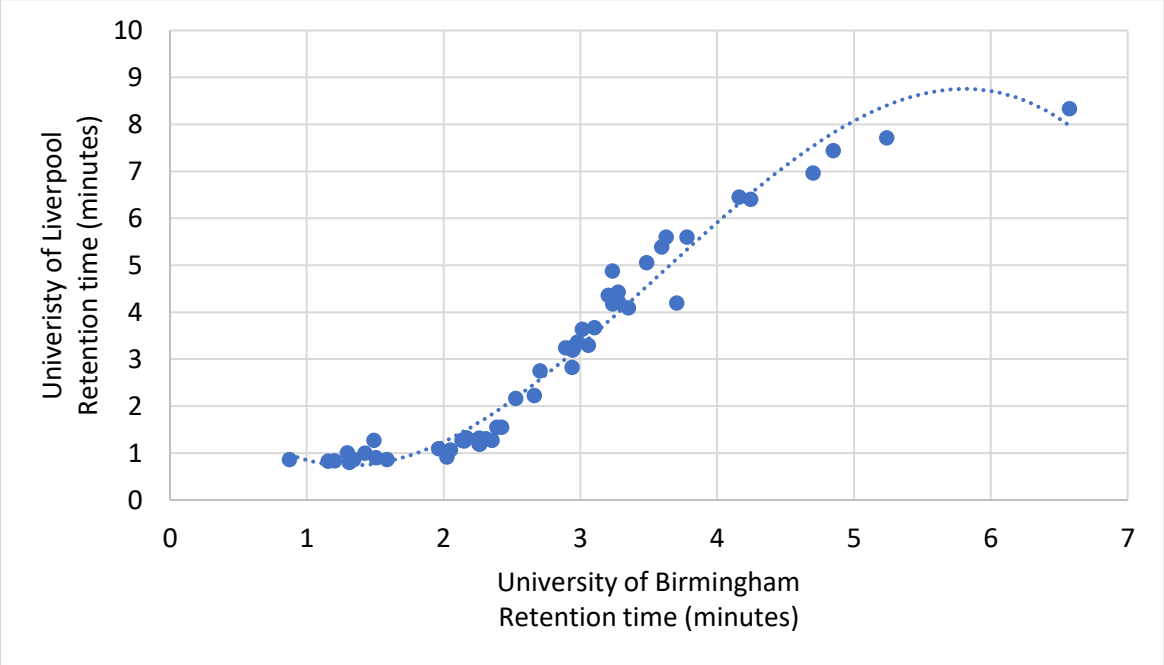


Figure 4-2: The relationship between retention times for the University Birmingham dataset and the University of Liverpool method 1 dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time for both datasets was measured in minutes. Line of best fit: y = -0.1765x3 + 1.8838x2 - 4.0261x + 3.1709 R² = 0.9694

To test the validity of the model, a subset of 20 random metabolites were removed and a new model created. For the 20 random metabolites, the retention time from the University of Birmingham method was mapped onto the University of Liverpool (method 1) using the equation from the trend line on the model (this can be seen in Figure 4-2). One of the limitations with RT mapping is that to create a good model, overlapping metabolites must elute across the whole elution gradient for both methods. Otherwise, there are gaps created in which data cannot be modelled for. An example of this is in this model where that there were very few metabolites retaining between seven and 11 minutes for the

university of Birmingham method. This meant that the model had to be adjusted to accommodate for that, and so all models for this comparison were only created for up to seven minutes of the university of Birmingham elution gradient.

The difference in retention time for the 20 random overlapping metabolites was calculated for each of the 10-fold cross-validations and the range of differences can be seen below in Appendix H-2. The mean difference in retention time was consistently small enough (mean difference 0.1658 minutes) with a high R^2 value (average 0.9724), which meant that the models worked well enough to accurately map the retention times of standards across from the University of Birmingham method to the University of Liverpool (method 1).

When using the regression model above to expand the two datasets, the University of Birmingham library was increased by 84% from 536 to 985 metabolites and the University of Liverpool (method 1) library would be increased by 95% from 504 to 985 metabolites.

4.2.1.1.3 University Of Birmingham Vs University Of Liverpool (Method 2)

The University of Birmingham vs University of Liverpool (method 2) dataset comparison compared two datasets where a wide variety of analytical standards had been analysed to form a retention time library. The methods applied by the different research groups varied multiple parameters including: UHPLC column applied, elution gradient, flow rate, column temperature, and mobile phases. Despite these differences, there was a clear relationship between the retention times of the two different methods and a trend line that fitted very well to form the model. This model suffered a similar issue to that of the university of Birmingham vs university of Liverpool (method 1) model, in that there was a large gap in metabolites eluted between seven – 11 minutes. This meant that any metabolites eluting in this time would potentially have a poorly fitted model, to avoid this, the model was only formed from zero to seven minutes (Figure 4-3).

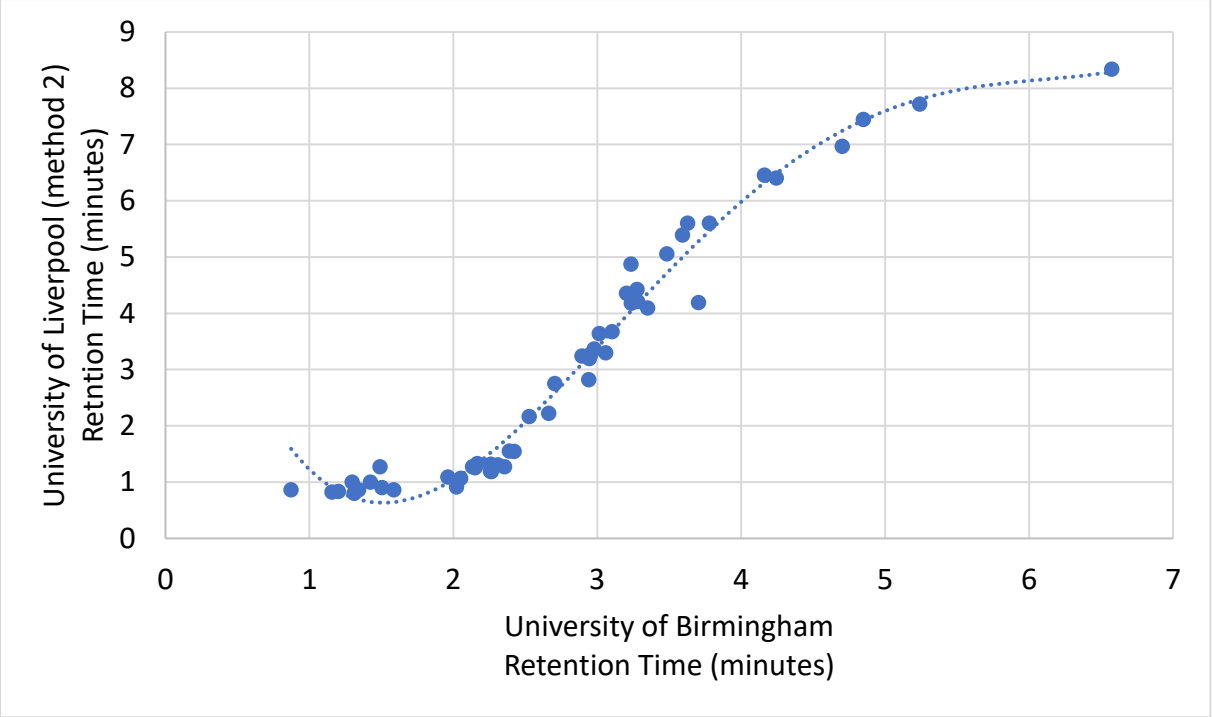


Figure 4-3: The relationship between retention times for the University Birmingham dataset and the University of Liverpool method 2 dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time for both datasets is recorded in minutes. Line of best fit: $y = 0.0455x4 - 0.8373x3 + 5.1526x2 - 10.462x + 7.3291$ $R^2 = 0.9773$

After the initial model has been created, the data was then subsetted into two groups to conduct 10-fold cross-validation. When completing this, 20 random overlapping metabolites were subsetted and the rest used to form a model. The model equation was then applied to the subsetted 20 overlapping metabolites to see if the mapped compound retention time was accurate to the experimental retention time using that method. This was repeated ten times. The results from the 10-fold cross-validation can be seen in appendix H-3.

When looking at the range of retention time differences (appendix H-3) this model is very accurate in retention time prediction when compound retention times are mapped across. This further shows that mapping retention times from dataset onto a different method is appropriate.

When expanding two datasets by mapping all the metabolites from one dataset onto the other using the model above, the University of Birmingham library was increased by 84% from 536 to 985 metabolites and the University of Liverpool (method 2) library would be increased by 95% from 504 to 985 metabolites.

4.2.1.1.4 University Of Florida Vs University Of Liverpool (Method 1)

For the comparison of datasets from the University of Florida and the University of Liverpool, the data was merged, and any non-overlapping metabolites removed. The retention time for the overlapping metabolites was then compared and a model formed. The data was then filtered to remove any obvious outliers and any metabolites eluting in the void volume of either method. This reduced the number of overlapping compounds from 58 to 43. The removal of these metabolites improved the accuracy of the model from a correlation coefficient of 0.6784 to 0.853.

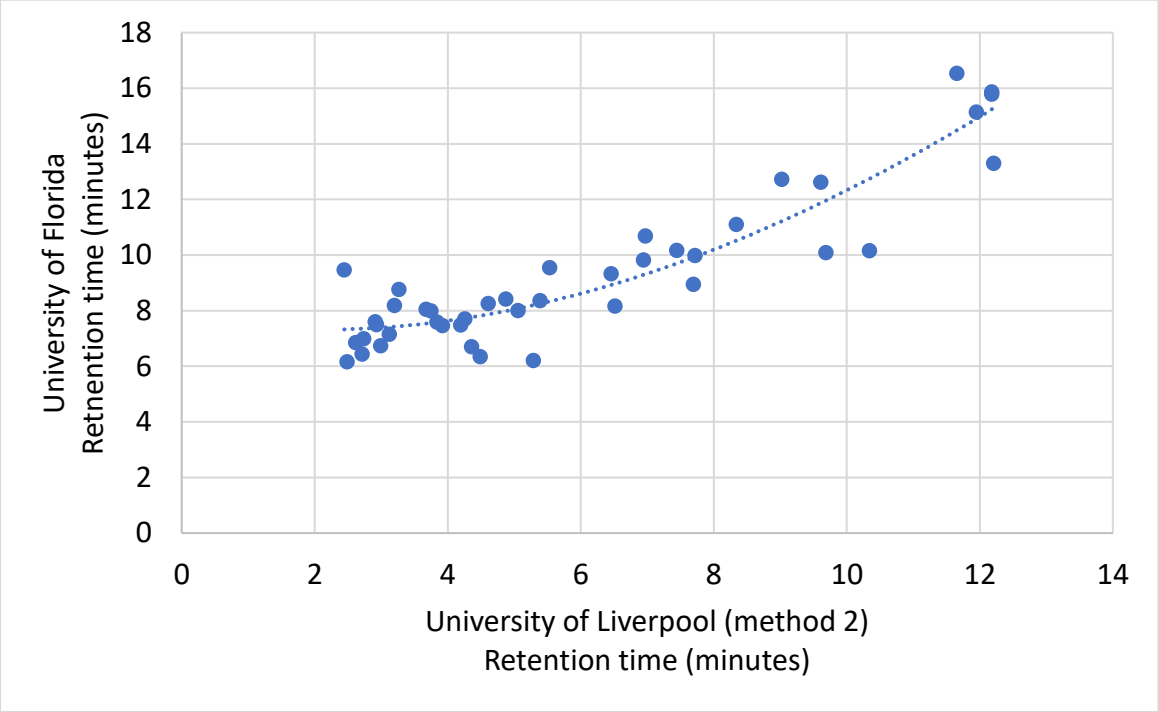


Figure 4-4: The relationship between retention times for the University of Liverpool (method 2) dataset and the University of Florida dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time for both datasets is recorded in minutes. $y = -0.001x3 + 0.0928x2 - 0.3663x + 7.6796$ $R^2 = 0.853$

As the retention times of the metabolites in the University of Liverpool-Method 2 had a clear relationship with the University of Florida dataset, the trend line from the model of the data could be applied to map the University of Liverpool-Method 1 retention times onto the University of Florida dataset (Figure 4-4). To validate this model, 10-fold cross-validation was completed by subsetting 20 random overlapping metabolites and a forming a model on the remaining metabolites. The model equation was then used to map the University of Liverpool-Method 2 metabolites onto the University of Florida method. This was completed 10 times. The range of differences can be seen in Appendix H-4. The 10-fold validation showed that the mean difference in retention time between the mapped retention time and the experimental retention time for that method was -0.81 to 0.49 minutes. This variance in mean across the 10 validations along with the wide range of R^2 value (0.72-0.93) suggests

that this model is not robust enough to be used confidently to map retention times across. This could be due to one dataset being complied of biological data and the other analytical standards or may be because only 23 metabolites were used for model formation.

4.2.1.1.5 University Of Florida Vs University Of Liverpool (Method 2)

For the comparison of datasets from the University of Florida and the University of Liverpool (method 2), data were merged, and any non-overlapping metabolites removed. The retention times for the overlapping metabolites were then compared and a model formed. The data was then filtered by the removal of any obvious outliers and any metabolites eluting in the void volume of either method. This meant that overlapping metabolites were reduced from 58 to 42. Before removal of metabolites eluting in the void volume and clear outliers from data overlapping both datasets, the correlation coefficient was 0.6501, this was then improved to 0.7611.

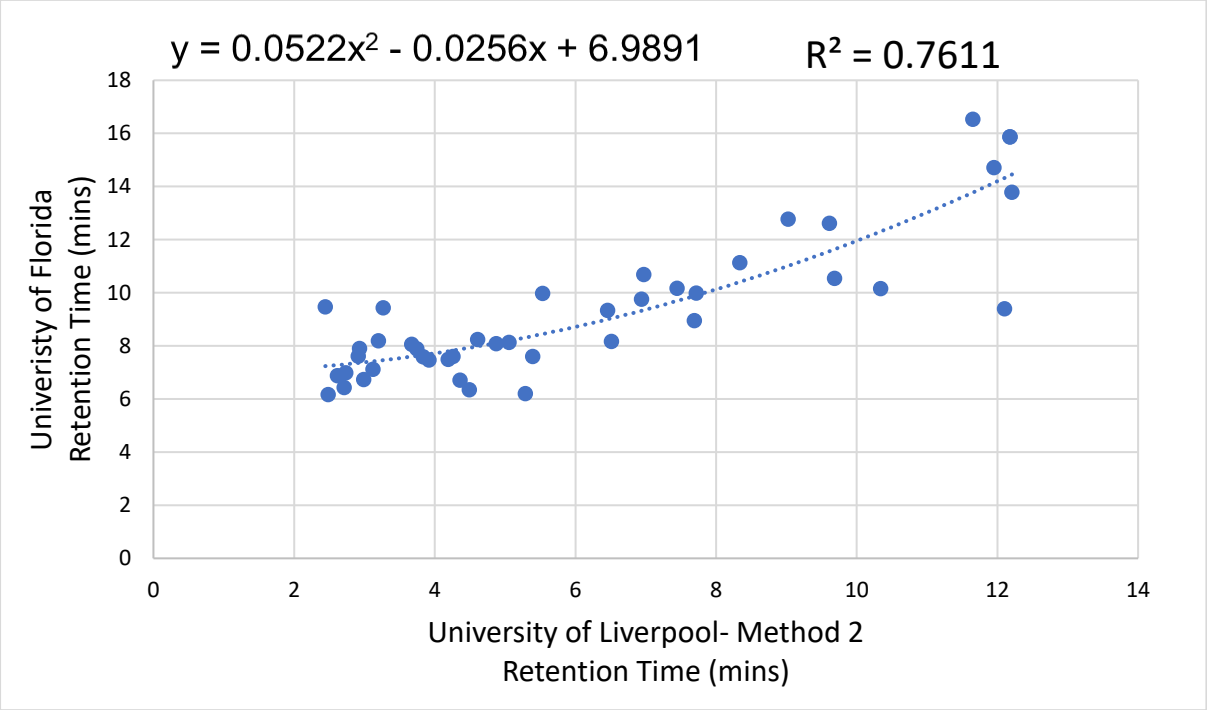


Figure 4-5: The relationship between retention times for the University of Birmingham dataset and the University of Florida dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time for both datasets is recorded in minutes.

As the retention times of the metabolites in the University of Liverpool-Method 2 had a clear relationship with the University of Florida dataset (Figure 4-5), the trend line from the model could be applied to map the University of Liverpool-Method 2 retention times onto the University of Florida dataset. To validate this model, 10-fold cross-validation was completed by subsetting 20 random overlapping metabolites forming a model based on the remaining metabolites. The validation results can be seen in Appendix H-5.

After the filtering and subsetting of 20 random metabolites there were only 23 overlapping metabolites left to use for constructing a model. This could be part of the reason that the model is less robust and varies more in accuracy. This is because with less data points across whole elution gradient time, the less data points there are to try and fit a model to. For this reason, it also explains the variation in R^2 value (0.75-0.93). When investigating the accuracy of the model, the mean differences in retention time were investigated. Within the 10-fold validation, the mean difference ranged from -0.90 to 0.58 minutes. This is a very wide range and further demonstrates the variability in the model validation. This current model would not be ideal for retention time mapping and to better improve this model more investigation would need to be completed to conclude whether the use of biological samples is the source of the variability, or the low number of overlapping metabolites used for model construction.

4.2.1.1.6 University Of Florida Vs University Of Birmingham

Datasets from both research groups were merged and filtered so that only the metabolites that were present in datasets/libraries were applied to construct a model. The retention times of the overlapping metabolites were then used to form a model by looking at the relationship of the retention times and forming a line of best fit through the data. This line of best fit would then be used to map the retention time from one research group onto the other research group. As can be seen from Figure 4-6, there is no clear relationship between the two sets of data, with a lot of the metabolites potentially being outliers and affecting the model accuracy. Due to this, no cross-validation of the model was undertaken as it was clear that any mapped metabolites would differ greatly to the actual recorded retention time.

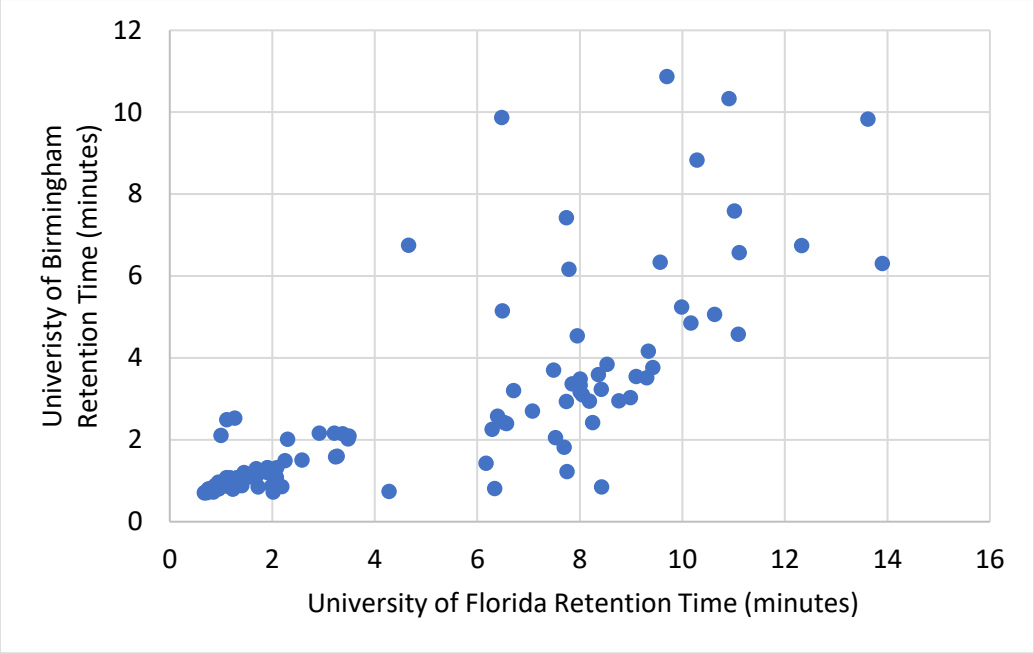


Figure 4-6: The relationship between retention times for the University of Birmingham dataset and the University of Florida dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time for both datasets is recorded in minutes.

4.2.1.2 HILIC Assays

The objective was to look at data from HILIC assays across different laboratories applying similar methods and investigate if an accurate model could be formed to allow for one set of data to be mapped onto the other, therefore providing an opportunity to expand the RT library of analytical standards analysed. Two different types of datasets were used, those that were composed of biological data collected (Mayo clinic) and those composed of data acquired from the analysis of analytical standards (Fiehn MoNA, University of Birmingham, and University of Liverpool).

Table 4-2: The total number of unique metabolites within each dataset and/or library along with the number of overlapping metabolites between datasets for the HILIC assays from different research groups. For example, the Mayo Clinic and University of Birmingham Negative ion mode libraries/datasets have 670 and 535 metabolites, respectively but only ten metabolites which are present in both libraries/datasets.

	Mayo Clinic	University Of Birmingham Negative Ion Mode	University Of Birmingham Positive Ion Mode	University Of Liverpool	Fiehn Mona
Mayo Clinic	670				
University Of Birmingham Negative Ion Mode	10	535			
University Of Birmingham Positive Ion Mode	11	317	496		
University Of Liverpool	3	79	79	371	
Fiehn Mona	10	162	203	56	1002

4.2.1.2.1 University Of Birmingham Positive Ion Mode Vs University Of Birmingham Negative Ion Mode

At the University of Birmingham, the Analytical and Clinical Metabolomics Group (ACMG) and Phenome Centre Birmingham (PCB) apply two different methods for positive ion mode and negative ion mode in order to try to separate and detect as many different metabolites as possible. For this reason, it was thought a good idea to compare the two different methods to investigate whether changes in mobile phase only while all other assay parameters are kept constant influence retention time mapping abilities.

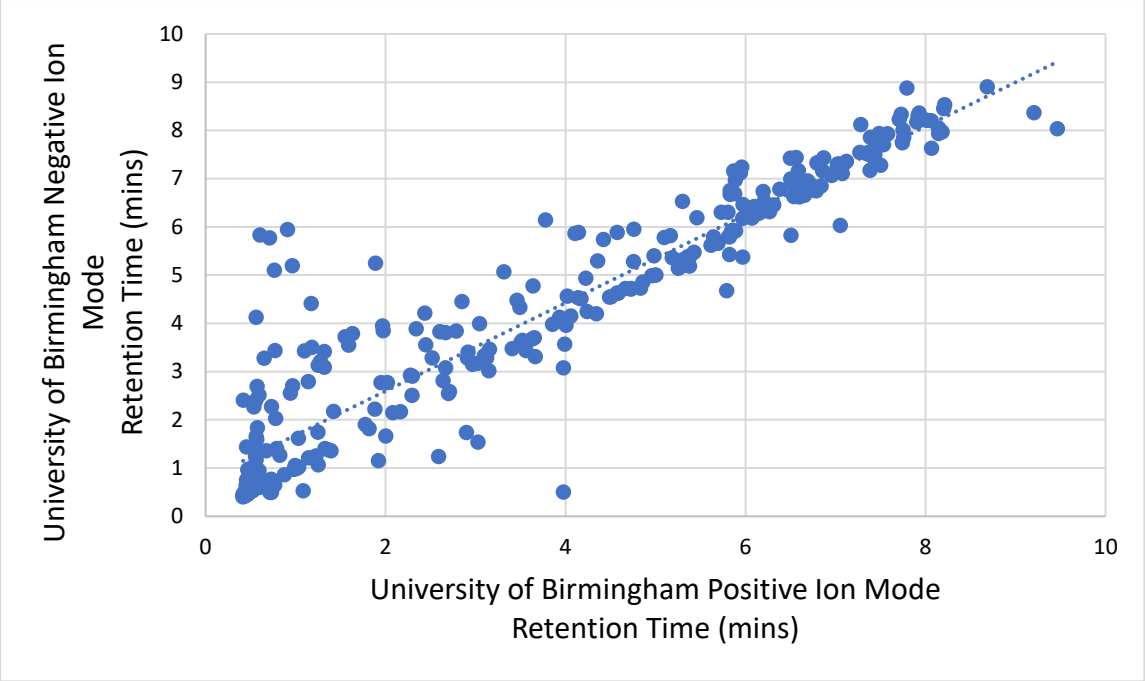


Figure 4-7: The relationship between retention times for the University of Birmingham positive ion mode dataset and the University of Birmingham negative ion mode dataset. The retention time for both datasets is recorded in minutes. Line of best fit equation: $y = 0.9144x + 0.7678$ R^2 value = 0.8801

In the above Figure, there is a relationship between the two different sets of retention times but the overlapping metabolites that elute in the void volume along with the outliers are skewing the data and lowering the accuracy of the model. To overcome this, the metabolites that eluted before two minutes within the positive ion mode assay were removed along with any clear outliers to produce the final model below.

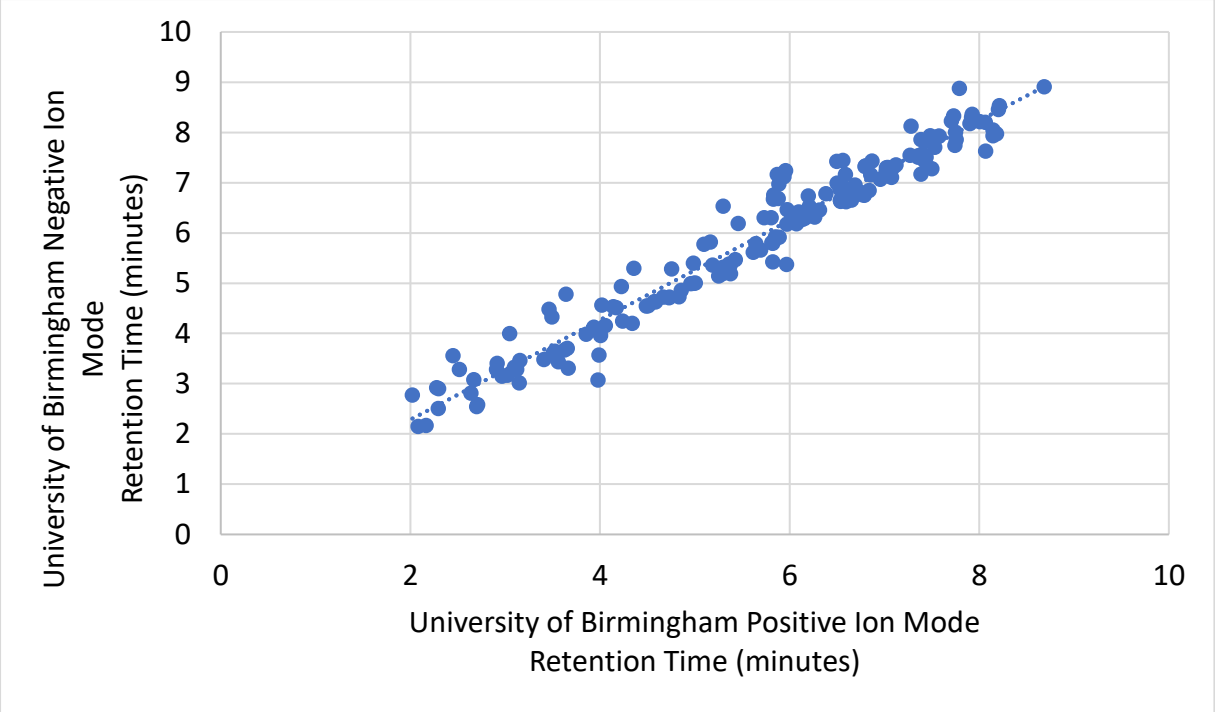


Figure 4-8: The relationship between retention times for University of Birmingham positive ion mode dataset and the University of Birmingham negative ion mode dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time both datasets is recorded in minutes. Line of best fit equation: y = 0.9914x + 0.3017 R^2 value= 0.9538

The removal of metabolites that eluted before two minutes within the positive ion mode assay along with any clear outliers improved the accuracy of the model from correlation coefficient = 0.8801 to correlation coefficient = 0.9538 and a clear relationship can be seen between the two different ion modes. The outliers removed can be seen below in Appendix H-6. Most of the outliers were of the same metabolite super-class (Organoheterocyclic compounds) and were structurally similar (cyclic). Suggesting that the difference in mobile phase between the two methods was affecting when the metabolites of that class eluted off the column.

To test the robustness of the model,10-fold cross-validation was completed by subsetting 20 of the overlapping metabolites and the constructing a model using the remaining overlapping metabolites. This was repeated 10 times. The results from which can be seen in Appendix H-7. The model applied produced a range of retention time differences, with a mean difference range of -0.12 to 0.16 minutes between the mapped retention time and the experimental retention time when applying the negative ion mode method. This high level of accuracy shows that it is possible to map between two different methods that differ on mobile phase, which was before thought not possible.

4.2.1.2.2 University Of Birmingham Positive Ion Mode Vs University Of Liverpool

The comparison of University of Birmingham Positive Ion mode and the University of Liverpool datasets showed the comparison of two different datasets comprised of analytical standards of metabolites. The relationship between the two datasets was clear with a correlation coefficient of 0.747. In attempt to improve this value, metabolites that eluted before two minutes within both assays were removed along with any clear outliers that were very different to the clear trend within the data. This increased the correlation coefficient to 0.834. Removing the void volume metabolites and outliers also influenced the retention time relationship between metabolites and the graph type became a more complex polynomial rather than linear.

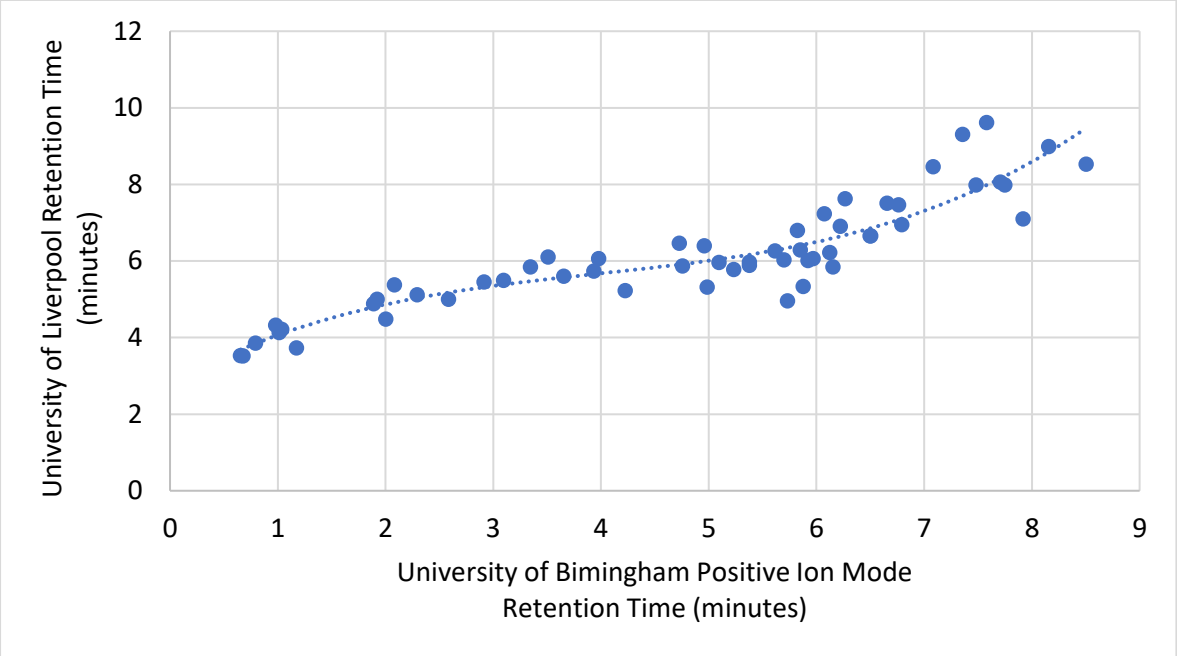


Figure 4-9: The relationship between retention times for University of Birmingham positive ion mode dataset and the University of Liverpool dataset after the metabolites that were eluting in the void volume were removed along with any obvious outliers. The retention time both datasets is recorded in minutes. Line of best fit equation: $y = 0.0264x3 - 0.3157x2 + 1.5621x + 2.79$; $R^2 = 0.834$

Validation of this model was conducted via 10-fold cross-validation to see if the retention times could be mapped across accurately without a close to perfect correlation coefficient. The cross-validation results can be seen below in Table 4-3.

Table 4-3: The mean, minimum, maximum and median differences in retention time in both seconds and minutes related to the validation test metabolites used to test the retention time mapping model based off of the relationship between University of Birmingham positive ion mode dataset and the University of Liverpool dataset. All differences were calculated in minutes.

	Mean Difference	Minimum Difference	Maximum Difference	Median Difference	Line Of Best Fit Equation	R^2 Value
Validation 1	-0.05078	0.034082	1.125953	0.004655	$Y = 0.0271x^3 - 0.3223x^2 + 1.555x + 2.8606$	0.807
Validation 2	-0.0209	0.01474	1.608522	-0.0378	$Y = 0.0245x^3 - 0.2829x^2 + 1.3981x + 2.9894$	0.8
Validation 3	0.05739	0.008987	1.746965	0.097085	$Y = 0.0209x^3 - 0.2506x^2 + 1.3317x + 3.0661$	0.8608
Validation 4	0.096307	0.012682	1.551295	0.013248	$Y = 0.0236x^3 - 0.2697x^2 + 1.3814x + 2.8997$	0.8256
Validation 5	0.269034	0.025747	1.651602	0.264891	$Y = 0.0254x^3 - 0.2954x^2 + 1.4959x + 2.8111$	0.8398
Validation 6	-0.25274	0.036684	1.92541	-0.22878	$Y = 0.0256x^3 - 0.3219x^2 + 1.6277x + 2.6937$	0.8747
Validation 7	-0.14012	0.014209	1.808738	-0.11935	$Y = 0.0297x^3 - 0.3756x^2 + 1.8409x + 2.4942$	0.8238
Validation 8	0.036212	0.109281	1.483383	-0.00616	$Y = 0.0182x^3 - 0.2059x^2 + 1.1746x + 3.0456$	0.8438
Validation 9	-0.31258	0.021863	1.716683	-0.32467	$Y = 0.0311x^3 - 0.3713x^2 + 1.7189x + 2.6501$	0.8914
Validation 10	0.248254	0.008785	1.493361	0.143445	$Y = 0.0252x^3 - 0.2904x^2 + 1.4649x + 2.8496$	0.8021

The resulting differences between the mapped university of Birmingham retention times and the experimental University of Liverpool assay retention times were not very consistent with validation models one, two, three, four, seven, and eight having small mean differences in retention times and validation models four, six, nine and ten having much larger mean differences. This larger mean difference in retention time was not reflected in how well the model fitted the data, as these validation models included two of the higher R^2 values. Even with the larger mean differences the model would still be appropriate for allowing the extension of both retention time libraries.

When expanding the two libraries using the above models, the number of metabolites in the university of Birmingham Positive Ion Mode library was increased by 59% from 496 unique metabolites to 788 unique metabolites. The University of Liverpool library was increased by 112% from 371 to 788 unique metabolites.

4.2.1.2.3 University Of Birmingham Negative Ion Mode Vs Fiehn MoNA

The comparison of University of Birmingham Negative Ion mode and the Fiehn MoNA datasets showed the comparison of two different datasets comprised of analytical standards of metabolites. The relationship between the two datasets showed a mixture of retention time relationships that had a clear trend and those that didn't conform to a clear trend. This can be seen clearly in the Figure below (Figure 4-10) and in the correlation coefficient of 0.6043.

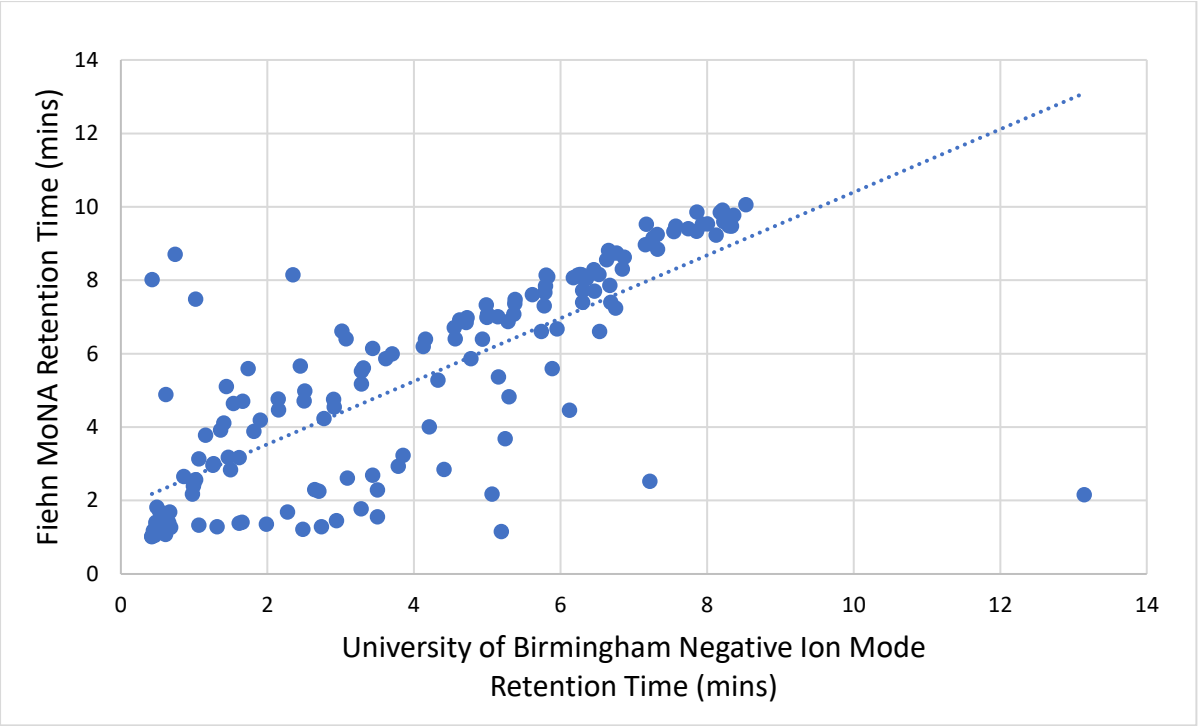


Figure 4-10: The relationship between retention times for the University of Birmingham negative ion mode dataset and the Fiehn MoNA dataset. The retention time for both datasets is recorded in minutes. Line of Best Fit equation: $y = 0.858x + 1.8163$; $R^2 = 0.6043$

To improve this value, metabolites that eluted before the first minute of the University of Birmingham assay were removed along with the metabolites plotted below the line of best fit. This increased the correlation coefficient to 0.9331. Although the above graph was showing two different retention time trends, with one was very high correlation coefficient and the other was lower but still noticeable as a pattern. To investigate this further, all the metabolites that were plotted below the trend line between 1-6 minutes on the x-axis were investigated into their structure to see if there were any similarities that would explain the trend seen above. The metabolites and their classes can be seen in appendix H-8.

This investigation resulted in the conclusion that there was no specific trend that separated the two different sets of retention time trends as there was overlapping classes and structural similarities in both sets of retention times. For HILIC separation techniques, there are multiple ways to separate metabolites including hydrogen bonding, dipole-dipole interaction, and ion exchange. All the metabolites above

either featured nitrogen groups, oxygen groups and/or aromatic rings. These all have a higher amount of hydrogen bonding capability, and thus their retention ability on the column will be easily affect by changes in method parameters, such as pH. Some of the metabolites above, also featured in the outliers for the University of Birmingham positive ion mode vs University of Birmingham negative ion mode comparison. Which furthers the possibility that the differences in mobile phase are influencing metabolite retention. Just as in section 4.1.2.1, the two different methods have different buffers present. Fiehn methods apply ammonium formate whilst University of Birmingham negative ion mode applies ammonium acetate.

Table 4-4: The mean, minimum, maximum and median differences in retention time in both seconds and minutes related to the validation test metabolites used to test the retention time mapping model based off of the relationship between University of Birmingham negative ion mode dataset and the Fiehn MoNA dataset. All differences are calculated in minutes.

	Mean difference	Minimum difference	Maximum difference	Median difference	Line of Best Fit Equation	R^2 Value
Validation 1	0.047619	0.000332	1.366248	-0.13521	y = 0.847x + 2.6992	0.9397
Validation 2	0.025664	0.060896	0.875341	-0.0948	y = 0.8625x + 2.6172	0.9323
Validation 3	0.198538	0.019379	1.138379	0.094828	y = 0.8509x + 2.7104	0.9258
Validation 4	-0.13293	0.015328	1.285676	-0.12832	y = 0.864x + 2.5772	0.9248
Validation 5	0.117088	0.031064	1.206589	-0.06911	y = 0.8462x + 2.7322	0.9371
Validation 6	0.035113	0.022292	1.257882	0.064796	y = 0.8658x + 2.6024	0.9329
Validation 7	0.080759	0.005122	1.080637	0.002924	y = 0.8615x + 2.6333	0.9325
Validation 8	-0.21108	0.036149	1.522608	-0.22582	y = 0.8615x + 2.574	0.934
Validation 9	0.15566	0.045431	1.225533	0.013135	y = 0.861x + 2.6512	0.9307
Validation 10	-0.03731	0.011201	0.781066	-0.13347	y = 0.8531x + 2.6516	0.9218

The above table (Table 4-4) displays the results to the 10-fold cross-validation for the two methods. The mean difference in RT between mapped Fiehn MoNA and experimental Fiehn Mona had a wide range (-0.21 - 0.20 minutes) despite having high correlation values (0.92-0.94). This shows there is a large multitude of variation within the model accuracy but that all models are still well fitted to the data. This difference in retention time is still small enough to be used for increasing annotation confidence and would be extremely useful for reducing the number of false positive annotations.

When expanding the two libraries using the above models, the number of metabolites in the university of Birmingham Negative Ion Mode library was increased by 157% from 535 unique metabolites to 1375 unique metabolites. The Fiehn MoNA library was increased by 37% from 1002 to 1375 unique metabolites.

4.2.1.2.4 University Of Birmingham Positive Ion Mode Vs Fiehn MoNA

The comparison of University of Birmingham Positive Ion mode and the Fiehn MoNA datasets showed the comparison of two different datasets comprised of analytical standards of metabolites. The relationship between the two datasets showed a clear trend. This can be seen clearly in the figure below (Figure 4-11) and in the correlation coefficient of 0.8496. This was very different to the comparison seen before in section 4.1.2.4 and suggests that the mobile phase is paramount in the ability to map metabolites well across methods. This is because in comparison to section 4.1.2.4, both methods here apply ammonium formate in their mobile phases.

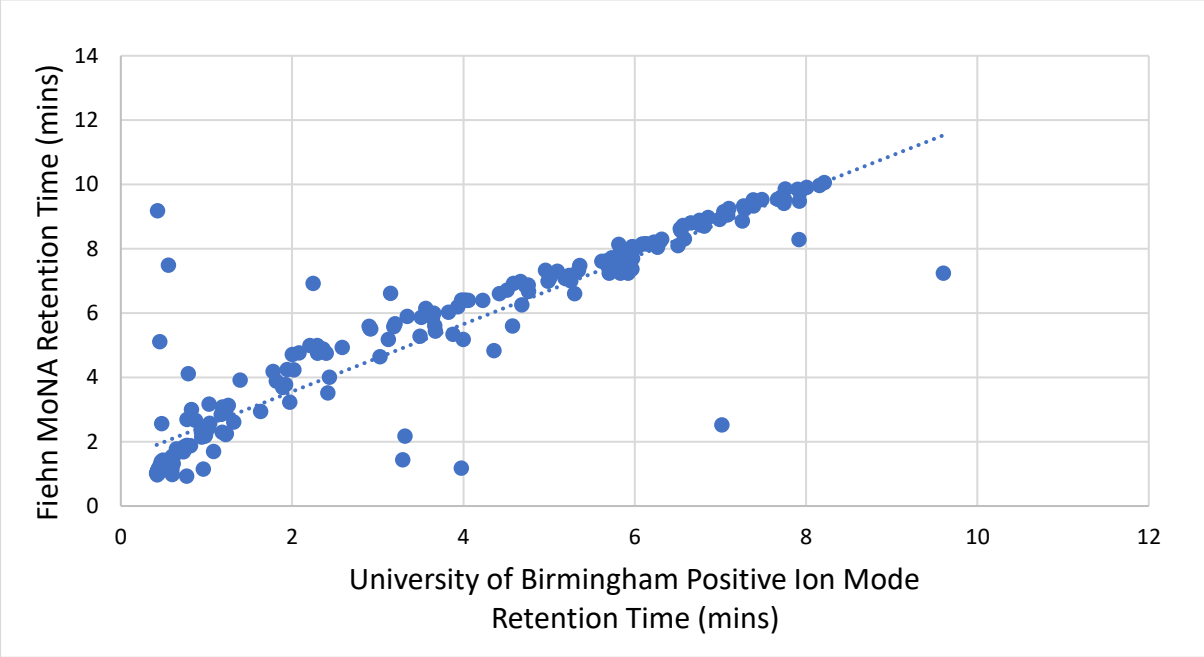


Figure 4-11: The relationship between retention times for the University of Birmingham positive ion mode dataset and the Fiehn MoNA dataset. The retention time for both datasets is recorded in minutes. Line of best fit equation: $y = 1.0483x + 1.4621$; $R^2 = 0.8496$

To remove any skew and bias from the data, metabolites that were eluting within the void volume of either method were removed along with any clear outliers. This improved the correlation coefficient to 0.9479. As this relationship produced a very clear model with a high correlation coefficient, 10-fold cross-validation was applied to demonstrate that the model would be applicable and robust. This was done by subsetting 20 of the overlapping metabolites from both datasets and using the remaining metabolites to form a model. The equation from the trend line of the relationship between the retention

times was then applied to the subsetted metabolites to map the Birmingham retention times onto the Fiehn MoNA retention method.

As can be seen in Appendix H-9, the range of mean differences between the mapped retention time and the experimental retention times cover a wide range (-0.11-0.12 minutes). Although the range is relatively large, the difference mean difference is actually very low. This in conjunction with the high R^2 values suggests that this model could be used to expand the Fiehn MoNA RT library by 29% as well as the University of Birmingham Positive Ion mode RT library by 161%. This would take the libraires from 1002 and 496 respectively to 1295 metabolites.

4.2.1.2.5 University Of Liverpool Vs Fiehn MoNA

The comparison of University of Liverpool and the Fiehn MoNA datasets showed the comparison of two different datasets comprised of analytical standards of metabolites. The relationship (Figure 4-12) between the two datasets showed a clear trend which when a trend line was applied had a correlation coefficient of 0.7481.

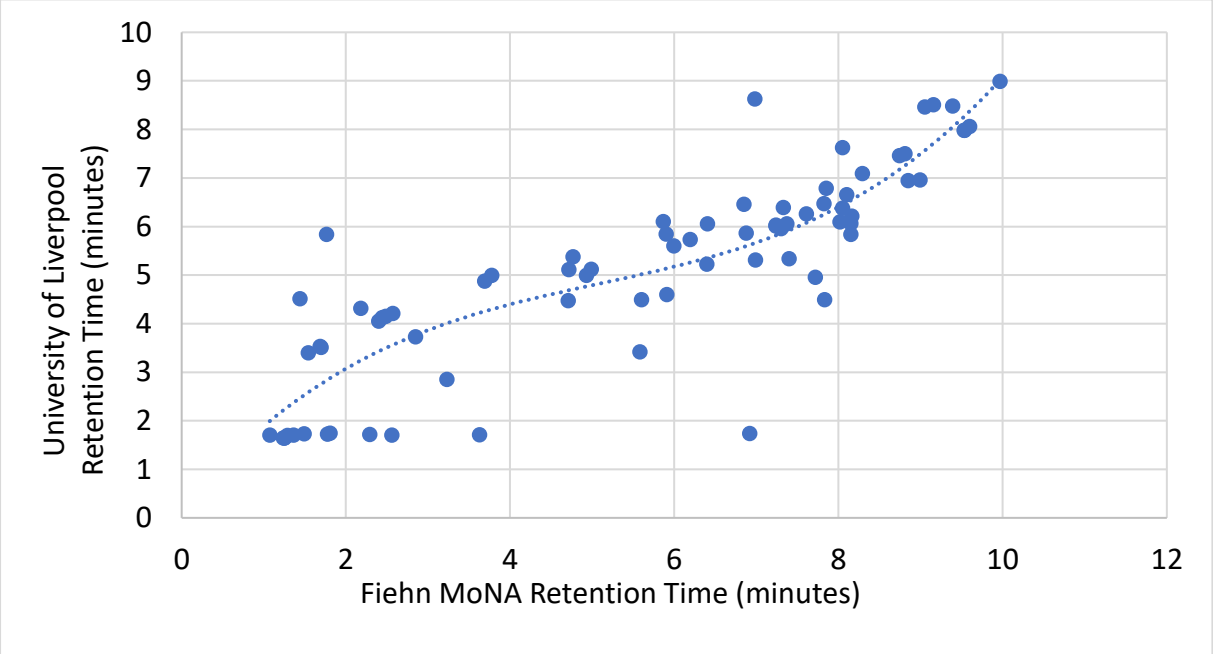


Figure 4-12: The relationship between retention times for University of Liverpool dataset and the Fiehn MoNA dataset. The retention time for both datasets is recorded in minutes. Line of best fit equation: y = 0.0208x3 - 0.3181x2 + 1.9889x + 0.1989; R² = 0.7481

This correlation coefficient was too low to form an accurate model for retention time mapping and so to remove any skew and bias from the data, metabolites that were eluting within the void volume of either method were removed along with any clear outliers. Metabolites that eluted before two minutes of the Liverpool assay and/or the first two minutes of the Fiehn assay were removed. This improved the correlation coefficient marginally to 0.8445.

Despite only gaining a small increase in the correlation coefficient, 10-fold cross-validation was conducted to see if retention time mapping could be successfully conducted on these datasets. 10-fold cross-validation was conducted by subsetting 20 random overlapping metabolites and using the remaining number to form a model. The equation of the trend line was then used to map the overlapping metabolites from the Fiehn assay onto the Liverpool assay. This was completed 10 times and results can be seen in Appendix H-10. Despite only a small increase in correlation coefficient before the 10-fold cross validation was conducted, the mean retention time difference between the mapped retention time and the experimental retention time was -0.14 - 0.17 minutes. This difference still accurate mapping to use to help annotate metabolites with a higher-level confidence of confidence. It would also be useful for the removal of possible false negative annotations. Expanding both these libraries would give using retention mapping models above would increase the Fiehn MoNA library by 31% from 1002 to 1317 and it would increase the Liverpool library by 255% from 371 to 1317 metabolites.

4.2.1.2.6 University Of Birmingham Negative Ion Mode Vs University Of Liverpool

Datasets from both research groups were merged and filtered so that only the metabolites that were present in datasets/libraries were applied to construct a model. The retention times of the overlapping metabolites were then used to form a model by looking at the relationship of the retention times and forming a line of best fit through the data. This line of best fit would then be used to map the retention time from one research group onto the other research group. As can be seen from Figure 4-13, although there was a relationship between the two sets of data, there was >30 outliers which were influencing the model accuracy. Because over 38% of the metabolites were outliers, no cross-validation of the model was undertaken as it was clear that any model would not be very robust.

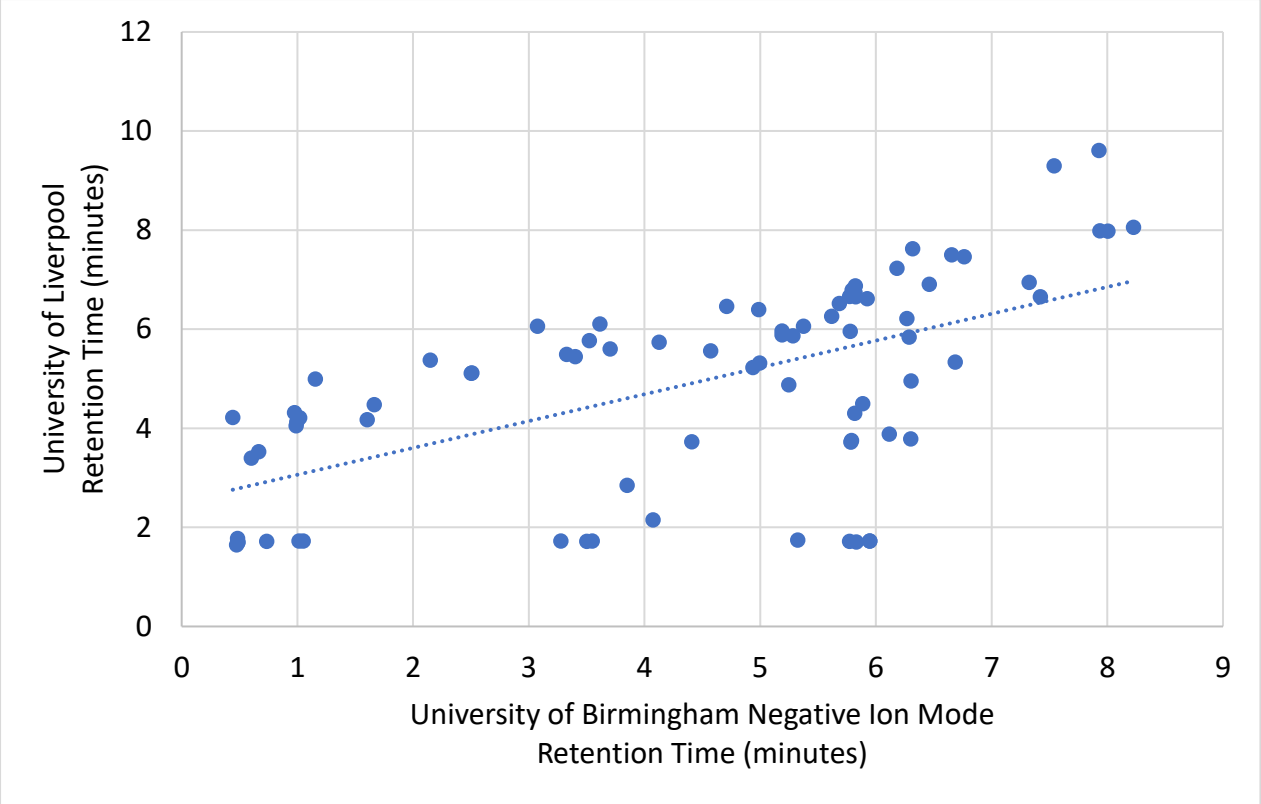


Figure 4-13: The relationship between retention times for University of Birmingham Negative Ion Mode and the University of Liverpool dataset. The retention time for both datasets is recorded in minutes. Line of best fit equation: $y = 0.5414x + 2.521$; $R^2 = 0.3547$

4.2.1.2.7 Mayo Clinic Comparisons

All comparisons where one of the datasets was from the Mayo Clinic produced unusable models. This is because for all comparisons there was a very low number of overlapping metabolites which meant that once any clear outliers were removed there was such a small amount left (<20 metabolites) that no validation models could take place. The data from the biological datasets was much more variable compared to analytical standards. This could have been due to many factors including complex matrices and dirty samples. One other factor that could have caused the low number of overlapping metabolites was the translation of metabolite name to InChIKey. For these studies, most of the translation was completed by hand because the software was not accepting the names provided. This remains a good example that retention time mapping works most accurately when a unique identifying naming system is used.

4.2.2 Retention Time Mapping Conclusions

Retention time mapping as a form of increasing metabolite annotation confidence is possible both for annotating metabolite matches, as well as reducing the number of false annotations made. This was a clear conclusion from both HILIC and reversed phase assays where many highly positive correlation coefficient models could be constructed. These models were then still valid after 10-fold cross-validation and were applicable to be made into expanded retention time libraries.

One observation made within this research chapter was that the studies that contained data from biological samples (as opposed to analytical standards) were harder to form accurate models for. This observation can be seen both in the table and the modelling graphs below. This observation was important because for this work to be applied in the future, it is paramount that the limits of this method be investigated. Retention time mapping should only be conducted using libraries where standards have been analysed and is not suitable for biological samples. This is because the variation in biological samples is too great and doesn't produce reliable, reproducible models to map retention times from.

Reversed phase assays were much more easily mapped across research groups in comparison with HILIC assays, with differences in column, mobile phase or elution gradient having little effect on the accuracy of the model. HILIC assays were difficult for retention time mapping and achieving accurate mapping models was temperamental. This could have been because the retention chemistry was more complicated or could be due to the mobile phases often being more complex with buffers being applied. Accurate mapping models can be created and the models with slightly lower accuracy are still applicable to be used for reducing the number of false annotations.

For models where only one parameter was changed between the two datasets/libraries a highly accurate model was able to be formed, although the HILIC assay where this took place contained more outliers, but this was concluded to be due to the different mobile phases as opposed to applying different instrumentation. This is because the same assay also produced a larger number of outliers with other assays.

RT mapping something which can be used in the future and would allow other laboratories/research groups to share retention time libraries and methods without the need for standardisation of methods globally. It would also allow for different UHPLC instrumentation to be utilised, which would allow for more research groups to undertake metabolomics studies and produce high level confidence data without the need for an expensive commitment to new instrumentation.

4.2.3 Lipidomics Retention Time Library

The aim of this research was to develop an RT library for lipids for a specific assay/method using lipid annotations derived from MS1 and MS/MS and LipidSearch^{TM} software and assess whether RT elution trends can be modelled to provide RTs for lipids not present in the RT library. Therefore, allowing for the investigation as to whether elution ranges for lipid classes can be applied to reduce the number of false annotations based on MS1 data.

Lipids from one dataset that contained multiple sample types (heart, kidney, liver, muscle, serum, milk, and chocolate) were separated out into different lipid classes and grouped together based on how many double bonds were present in the structure. Those groups were then ordered in ascending number of total carbons present using amalgamated nomenclature at a species level[4]. Mean retention times for each carbon number were then calculated, ensuring only metabolites with a minimum of three replicates and an RSD ≤ 5 were used. Regression models depicting the relationship between RT and the total number of carbons were produced and the best fitting trend line used for the retention time model. This equation was then used on the matching lipid class and carbon number of second dataset for validation. The predicted retention times were compared to experimental RTs to evaluate the accuracy of the in-silico library.

4.2.3.1 Acyl Carnitines (AcCA)

Acyl carnitines are a sub-class of fatty acyls that are commonly found in mammalian sample types. They are paramount for oxidative catabolism of fatty acids which allows for energy homeostasis within the human body. To study the elution times for acyl carnitines, lipids from this category were grouped by the number of double bonds present and then organised further into an ascending number of carbons. A different model was constructed for each set of acyl carnitines based on the number of double bonds present (Figure 4-14 and Table 4-5).

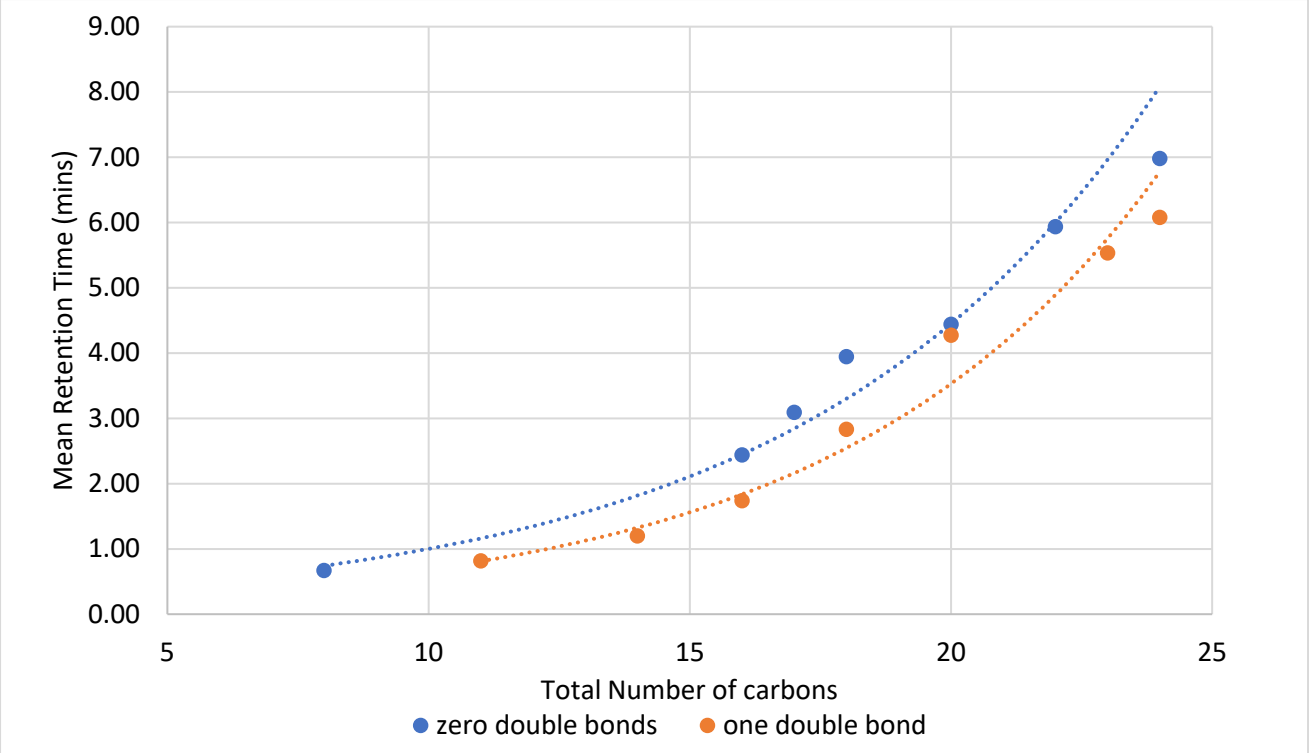


Figure 4-14: The trend in mean retention time as the number of carbons increased for all acyl carnitines with zero or one double bond present. Acyl carnitine structures with zero double bonds present are denoted by blue data points and orange data points denote acyl carnitine structures with one double bond present.

Table 4-5: Acyl Carnitine retention time model equations and correlation coefficients for each retention time trend based on the number of double bonds within the structure.

Number Of Double Bonds	Line Of Best Fit Equation	R^2 Value
Zero	$Y = 0.2249e^{0.1492x}$	0.9651
One	$Y = 0.1347e^{0.1633x}$	0.9637

For acyl carnitines with zero double bonds, applying the model to the second dataset for validation showed that the model worked reasonably well but was not accurate enough for the model to be used for metabolite annotation (Table 4-16). This is because there was a mean retention time difference of 0.59 minutes between the measured retention time and the predicted retention time.

For acyl carnitines with one double bond, when the model was applied to the validation dataset, the difference between the predicted retention time and the experimental retention time was very low at a mean of 0.13 minutes (Table 4-6). This small time difference would mean that this model would be applicable to be applied for metabolite annotation techniques. Not only is the mean retention time difference low, but the range of differences between the predicted retention time and the mean

experimental retention time for acyl carnitines with the same total number of carbons and double bonds was 0.02 – 0.21 minutes, which would have all been applicable error for metabolite annotation.

Table 4-6: The mean, minimum, maximum, and median difference in retention time in minutes between the predicted RT and the mean RT detected for Acyl carnitines with zero or one double bond present in the structure.

	Zero double bonds	One double bond
Minimum difference in retention time (minutes)	0.00	0.02
Maximum difference in retention time (minutes)	2.50	0.21
Mean difference in retention time (minutes)	0.59	0.13
Median difference in retention time (minutes)	0.37	0.13

Despite only a small number of data points, these models are still accurate and applicable for a retention time library. Although, the zero double bond model has the potential to be less accurate for structures containing eight to 16 carbons as there is a large gap in data here.

Only the acyl carnitine structures with zero and one double bond were modelled because all structures with a larger number of doubles bonds (such as two, three, four and five) all had a very low number of replicates. This meant that there was only one or two points to model with, which wouldn't have formed a very reliable model.

In summary, the acyl carnitine basic structure lends itself well to being mapped and an in-silico library developed and used for metabolite annotation. More work would be beneficial for the acyl carnitines with zero double bonds in order to try and make a more accurate model.

4.2.3.2 Lysophosphocholines (LPC)

Lysophosphocholines are a type of glycerophosphocholines, which are a subclass of glycerophospholipids. They are comprised of a glycerol backbone connected to a fatty acid and a phosphocholine molecule. To study the elution times for lysophosphocholines, lipids from this category were grouped by the number of double bonds present and then organised further into an ascending total number of carbons in each structure. A different model was constructed for each set of lysophosphocholines based on the number of double bonds present (Figure 4-15 and Table 4-7).

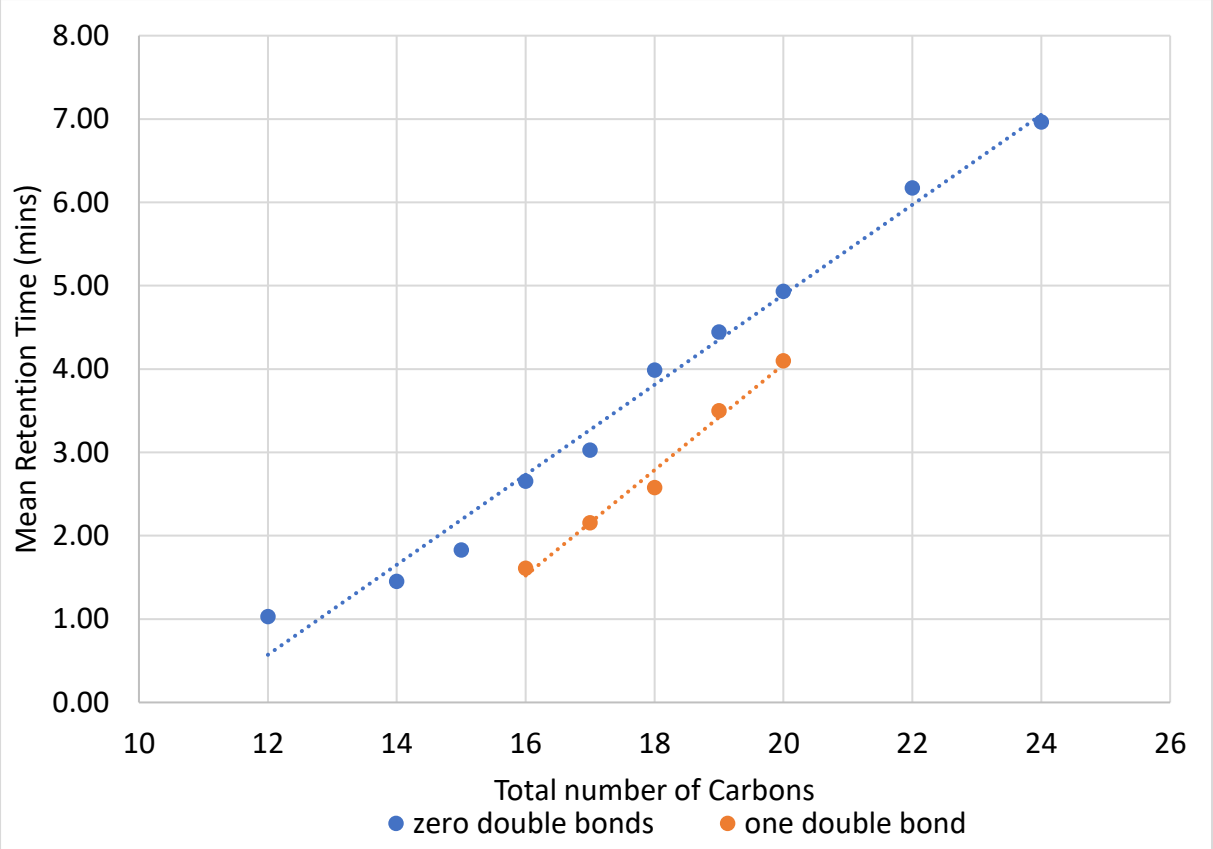


Figure 4-15: The trend in mean retention time as the total number of carbons increased in the lysophosphocholines containing zero or one double bond in the structure.

Table 4-7: Lysophosphocholine retention time model equations and correlation coefficients for each retention time trend based on the number of double bonds within the structure.

Number of Double Bonds	Line of Best Fit Equation	R^2 Value
Zero	y = 0.54x - 5.9078	0.9852
One	y = 0.6328x - 8.6026	0.9853

All lysophosphocholines that contained zero double bonds were ordered by the total number of carbons present. There was a wide range of total carbons that were used to form a model. This allowed for an accurate model to be formed without possibility of overfitting. When applied to the validation set, there was a wide range of differences in retention time between the predicted and experimental retention times of 0.07-10.30 minutes (Table 4-18). This range was extremely vast, but upon looking at the median and mean retention times they were relatively small in comparison to the range (0.74 and 1.45 respectively). This proposes that although the models looked to be accurate, the validation dataset highlighted that the model wasn't accurate enough to be used for predicating the retention times of lipids that weren't included in the model.

All lysophosphocholines that contained one double bond were ordered by the total number of carbons present. And a model formed on the relationship between the number of carbons present and the retention time detected. There was not an abundance of metabolites that could be used for creating a model for the metabolites that contained one double bond, and all these metabolites had a low number of total carbons. This meant that when applied to the validation model, any retention times for metabolites with larger total number of carbons, are more likely to be less accurate as the model was not designed with their inclusion. The mean retention time difference between the predicted and detected retention times in the validation dataset was 1.03 minutes (Table 4-8). This amount of error was too large to be used in annotation, as any uncertainty about annotation accuracy would be too high and the same annotation would apply to too many metabolites eluting off the column close to each other.

Table 4-8: The mean, minimum, maximum and median difference in retention time between the predicted RT and the mean RT detected for lysophosphocholines with zero or one double bond present in the structure.

	Minimum difference in retention time (minutes)	Maximum difference in retention time (minutes)	Mean difference in retention time (minutes)	Median difference in retention time (minutes)
Zero double bond	0.07	10.30	1.45	0.74
One double bond	0.00	6.45	1.03	0.52

Only lysophosphocholines with zero or one double bond in the structure were modelled because higher numbers of double bonds (two, three, four, five and six) all had too few data points to form a model.

In summary, despite only containing one fatty acid and therefore having a relatively simple structure, the models based on the relationship between the total number of carbons and the mean retention time detected was not accurate enough to be used for metabolite annotation when applied to other datasets. Therefore, the models should not be applied to construct a RT library for all carbon numbers in the range modelled above.

4.2.3.3 Ceramides (Cer)

Ceramides are a subclass of sphingolipids and are comprised of a sphingolipid base (most commonly sphingosine) linked to a second fatty acid via an amide linkage. The polar head of a ceramide can be simple (e.g., OH group) or more complex (e.g., phosphocholine). The possible variability in the lipid polar head group along with the differences in structure of each of the long chain fatty acids means that there are lots of possible structural causes to changes in retention time. To study the elution times for lysophosphocholines, lipids from this category were grouped by the number of double bonds present and then organised further into an ascending total number of carbons in each structure. A different

model was constructed for each set of lysophosphocholines based on the number of double bonds present (Figures 4-16 to 4-18 and Table 4-9).

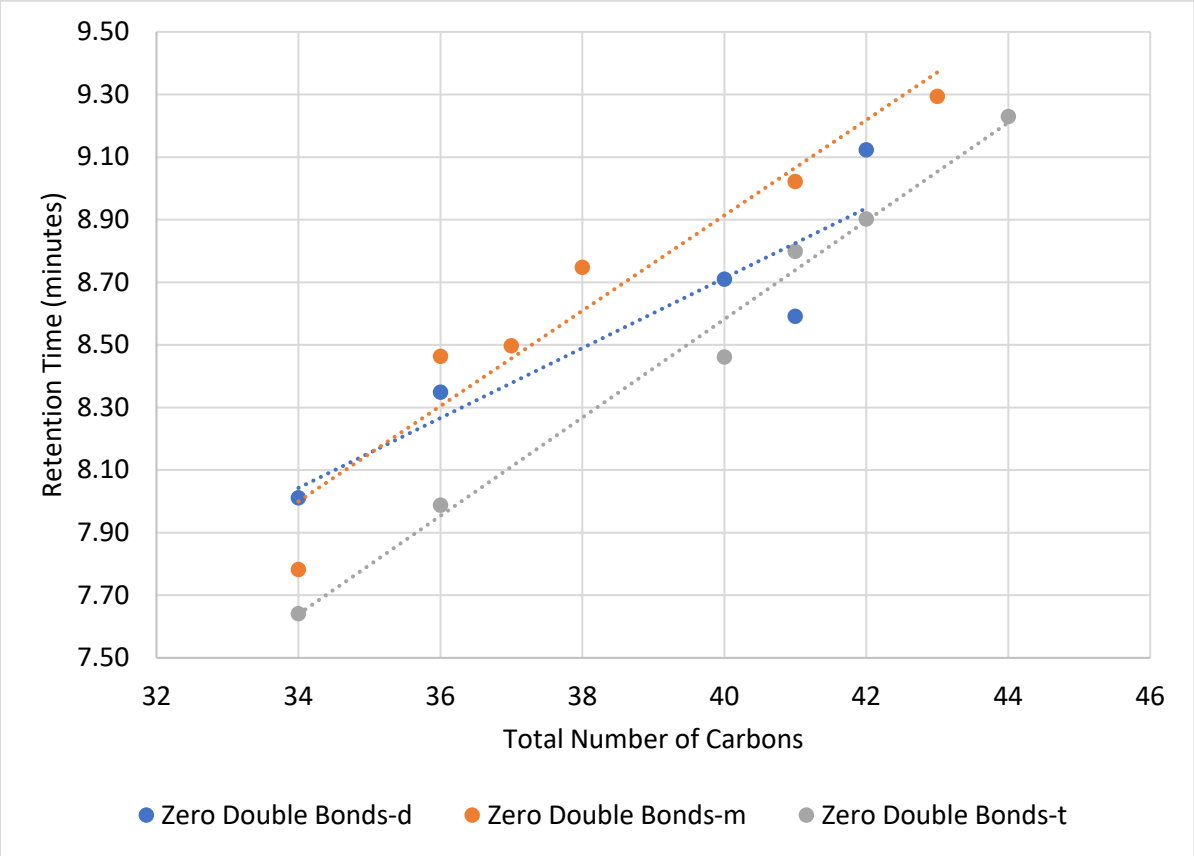


Figure 4-16: The trend in mean retention time as the total number of carbons increased in the ceramides containing zero. d, m, t denotes whether the ceramide base is m- mono hydroxy, d- dihydroxy, or t- trihydroxy.

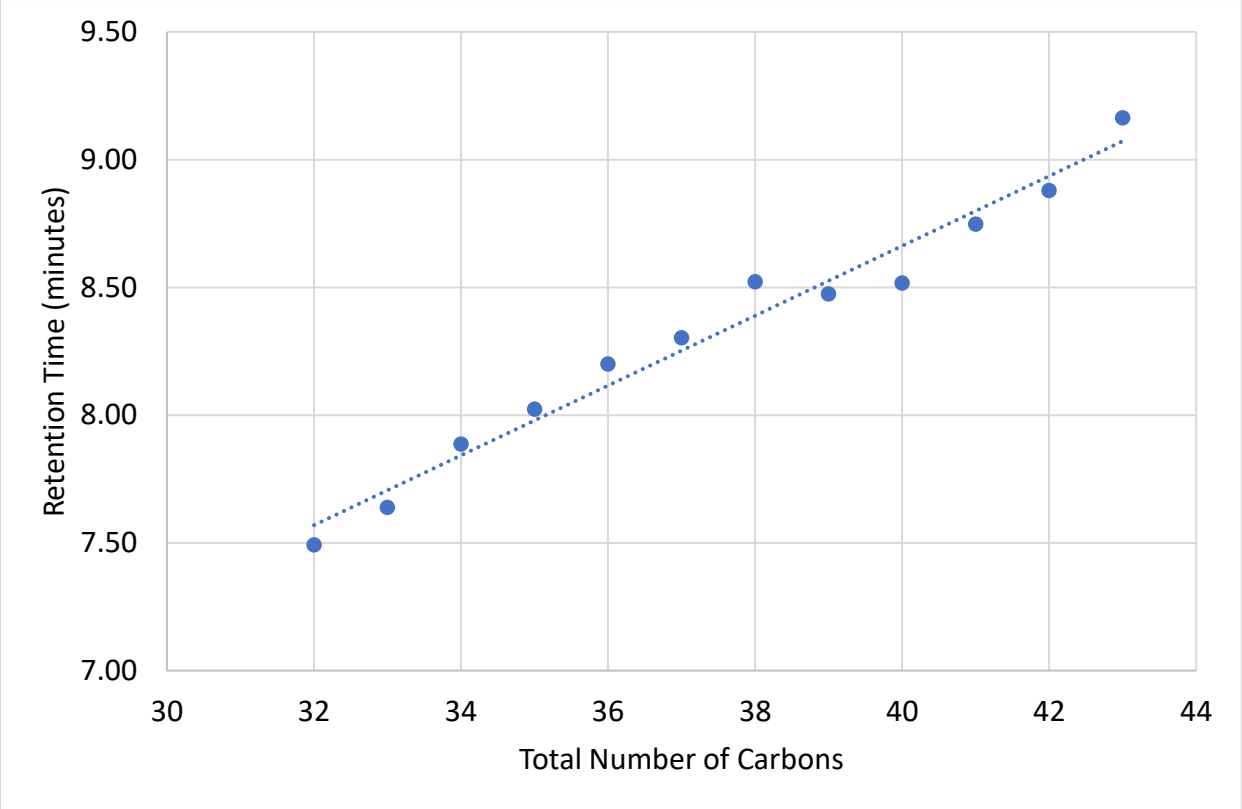


Figure 4-17: The trend in mean retention time as the total number of carbons increased in the ceramides containing one double bond that have a dihydroxyl base.

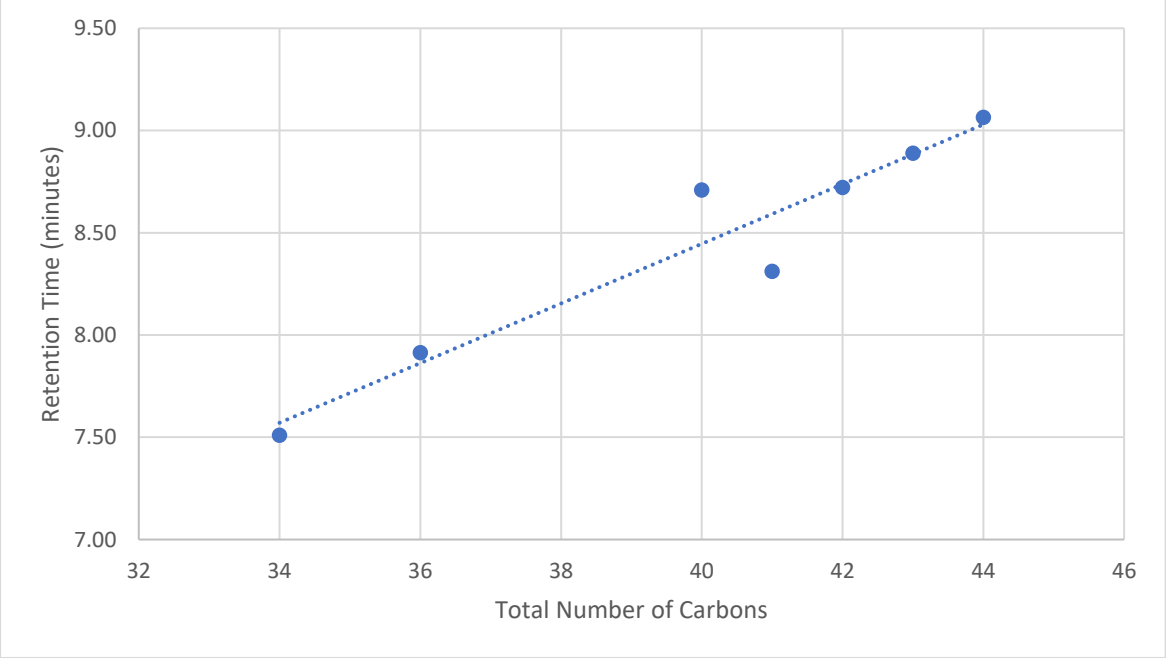


Figure 4-18: The trend in mean retention time as the total number of carbons increased in the ceramides containing two double bonds that have a dihydroxyl base.

Table 4-9: Ceramides retention time model equations and correlation coefficients for each retention time trend based on the number of double bonds within the structure. d, m, t denote whether the ceramide base is m- mono hydroxy, d- dihydroxy, or t-trihydroxy.

Number of Double Bonds	Line of Best Fit Equation	R^2 Value
Zero double bonds - d	y = 0.1117x + 4.2443	0.8587
Zero double bonds - m	y = 0.6328x - 8.6026	0.9853
Zero double bonds- t	y = 0.1523x + 2.8204	0.9263
One double bond- d	y = 0.1367x + 3.1957	0.9712
Two Double bonds- d	y = 0.1459x + 2.6111	0.9182

Table 4-10: The minimum, maximum, mean, and median difference in retention time between the predicted retention time and the detected retention time for ceramides with zero, one or two double bonds. d, m,t denote whether the ceramide base is m- mono hydroxy, d- dihydroxy, or t-trihydroxy.

	Minimum Difference In Retention Time	Maximum Difference In Retention Time	Mean Difference In Retention Time	Median Difference In Retention Time
Zero Double Bonds - d	0.93	1.05	0.96	0.95
Zero Double Bonds - m	0.75	0.99	0.86	0.87
Zero Double Bonds- t	0.90	1.15	1.05	1.09
One Double Bond- d	0.17	1.70	0.97	0.83
Two Double Bonds- d	0.03	1.87	1.00	0.88

Despite having a simple relationship with retention time and a model fitted very well to the data, all ceramides were unable to have a model that allowed for accurate annotations to be applied. All the mean retention times were too large as all were just under one minute in difference (see Table 4-10). This may have been because of the complex structure of ceramides with multiple possibilities for structure under the same number of carbons and double bonds, such as different polar head group possibilities along with the differences in structure of each of the long chain fatty acids. These differences within the ceramides would change retention on a column and therefore affect the retention time.

One benefit to the model is that it does allow for a reduction of false annotations by reducing the number of possible ceramides down to a smaller number of possibilities.

4.2.3.4 Hexosyl-2-ceramides (Hex2Cer)

Hexosyl-2-ceramides are a subclass of glycosphingolipids, which by extension are a subclass of glycerolipids. They are comprised of a ceramide base with two glucose molecules (lactose) attached to

the OH group at the polar head of the compound. All hexosyl-2-ceramides that contained one double bond were ordered by the total number of carbons present. A model was then formed on the relationship between the number of carbons present and the retention time detected (Figure 4-19).

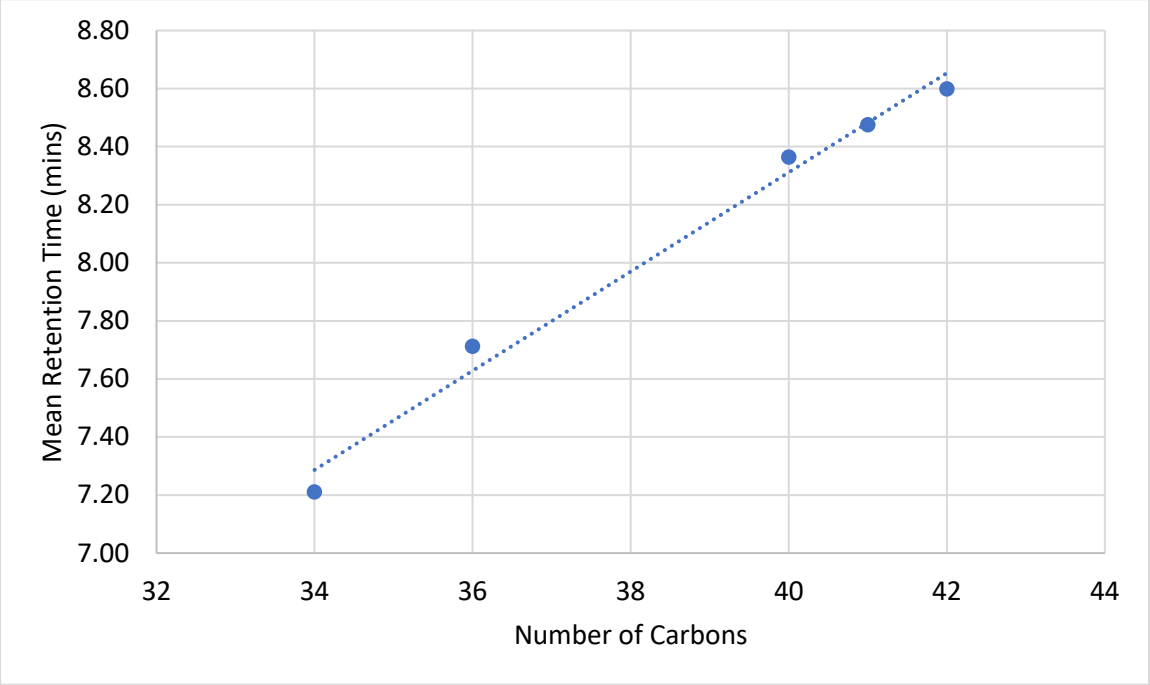


Figure 4-19: The trend in mean retention time as the number of carbons increased in the metabolites classed as Hexosyl-2-ceramides by LipidSearch ™. All metabolites contained one double bond in the structure and were dihydroxyl bases. Line of best fit equation: y = 0.1708x + 1.4776; R^2 = 0.9867

Since hexosyl-2-ceramides are a relatively non-abundant group of lipids there was again not many to make a model from. The model did not cover a wide range of compound sizes and so there was uncertainty about how accurate the model would be. Upon application to the validation set, it was clear that the difference in retention time was vast (Table 4-11). With a mean retention time of 0.81 minutes and a range of 0.68-1.09 minutes, any annotations made using this model would not be accurate.

Table 4-11: The mean, minimum, maximum, and median difference in retention time between the predicted RT and the mean RT detected for Hexosyl-2-ceramides with a dihydroxyl base and one double bond present in the structure.

	Difference In Minutes
Minimum Difference In Retention Time	0.68
Maximum Difference In Retention Time	1.09
Mean Difference In Retention Time	0.81
Median Difference In Retention Time	0.73

This model could have been inaccurate for multiple reasons, from the complex differences in structure between hexosyl-2-ceramides, or the fact that the number of ceramides the model was made from was limited. As only data were present for hexosyl-2-ceramides with one double bond, no other models for this lipids class were able to be constructed.

Using the model above to form a retention time library would not have been accurate for many reasons. One was that the number of carbons featured was very limited and still contained gaps. Another reason was that when validating the model, it was clear that the model was not as accurate as first thought with large differences in the retention time predicted using the model and experimental RTs detected.

4.2.3.5 Glycerophosphocholines (GPC)

Glycerophosphocholines are a subclass of glycerophospholipids. Their structure consists of a glycerol backbone attached to two fatty acids and a phosphocholine molecule. This means that there is a high level of variety in the different metabolites available as both fatty acids present can vary in length, hydroxylation, and number of double bonds. All of which affect the retention of a compound on a column. All glycerophosphocholines were grouped by the number of double bonds present in their structure and then were ordered by the total number of carbons present. To be included in the model, all metabolites must have had at least three replicates and an RSD < 5. Models were formed on the relationship between the number of carbons present and the mean retention time detected (Figure 4-20 and Table 4-12).

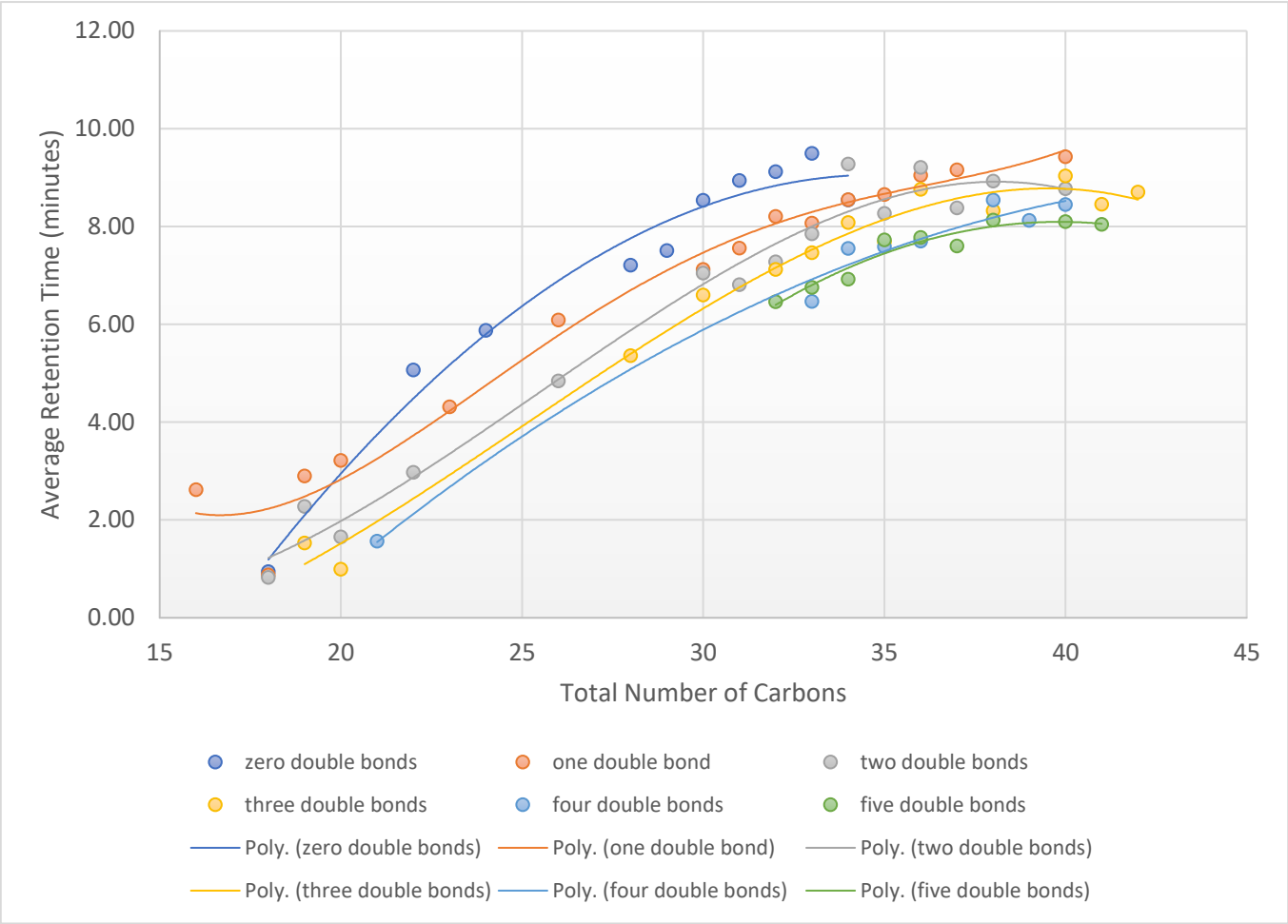


Figure 4-20: The trend in mean retention time as the number of carbons increased in the metabolites classed as Glycerophosphocholines with zero double bonds by LipidSearch ™.

Table 4-12: Trend line equations and correlation coefficient values for the relationships between the total number of carbons and mean retention time for each set of Glycerophosphocholines with a different number of double bonds.

	Trend Line Equation	Correlation Coefficient
Zero Double Bonds	$Y = -0.0277x^2 + 1.9318x - 24.602$	0.9709
One Double Bond	$Y = 1e\text{-}04x^4 - 0.0116x^3 + 0.5027x^2 - 8.8581x + 56.461$	0.9748
Two Double Bonds	$Y = -0.0011x^3 + 0.0818x^2 - 1.5717x + 9.2603$	0.9786
Three Double Bonds	$Y = -0.0008x^3 + 0.0609x^2 - 1.0313x + 4.2358$	0.9853
Four Double Bonds	$Y = -0.0115x^2 + 1.0683x - 15.817$	0.9858
Five Double Bonds	$Y = -0.0277x^2 + 2.2049x - 35.819$	0.9245

As can be seen from the figures above, multiple different models were able to be formed for glycerophosphocholines with a different number of double bonds. One limitation to these models is that there are gaps in the data, which means that any glycerophosphocholines with that number of carbons may have a lower accuracy in predicted retention time. Despite this, validation of the models was

conducted by applying the model equation to a second dataset and calculating the differences in predicted retention time to the experimental RT (Table 4-13).

Table 4-13: The minimum, maximum, mean and median differences in the retention time (RT) predicted by the models and the RT detected in minutes. All metabolites were Glycerophosphocholines with either zero, one, two, three, four, five, or six double bonds present in their structure.

	Minimum Difference In Retention Time (Mins)	Maximum Difference In Retention Time (Mins)	Mean Difference In Retention Time (Mins)	Median Difference In Retention Time (Mins)
Zero Double Bonds	0.22	2.47	1.61	1.78
One Double Bond	1.19	16.97	7.56	7.48
Two Double Bonds	0.01	5.25	1.48	0.51
Three Double Bonds	0.12	4.51	1.86	1.86
Four Double Bonds	0.01	4.10	1.70	1.26
Five Double Bonds	0.01	3.66	1.70	1.45

As can be seen from the data above, despite having a wide range of total number of carbons, and well-fitting trend lines, the models were not usable for predicting retention times. This is because the mean differences in the predicted compared to the experimental retention time were all greater than one minute.

This inability to predict an accurate retention time may have been due to the increase in variability of the structure of the compound. This work was done using the total number of carbons and double bonds and not on a more detailed structure comparison, so the variability in retention due to the placement of the double bond within the structure along with the different possible combinations of fatty acid chains would influence the retention of the compound on the column. Thus, causing differences in retention time for metabolites with the same total number of carbons. Therefore, using the models above to form a retention time library would not have been accurate. Further investigation into if the models could have been more accurate by focussing on different fatty acid combinations would be beneficial.

4.2.3.6 Glycerophosphoethanolamines (GPE)

Glycerophosphoethanolamines are a subclass of glycerophospholipids. They consist of a glycerol backbone linked to two fatty acids and a phosphoethanolamine molecule. Each fatty acid can differ in length, saturation, and hydroxylation. All glycerophosphoethanolamines that contained one double bond were ordered by the total number of carbons present, and a model formed on the relationship between the number of carbons present and the retention time detected (Figure 4-21).

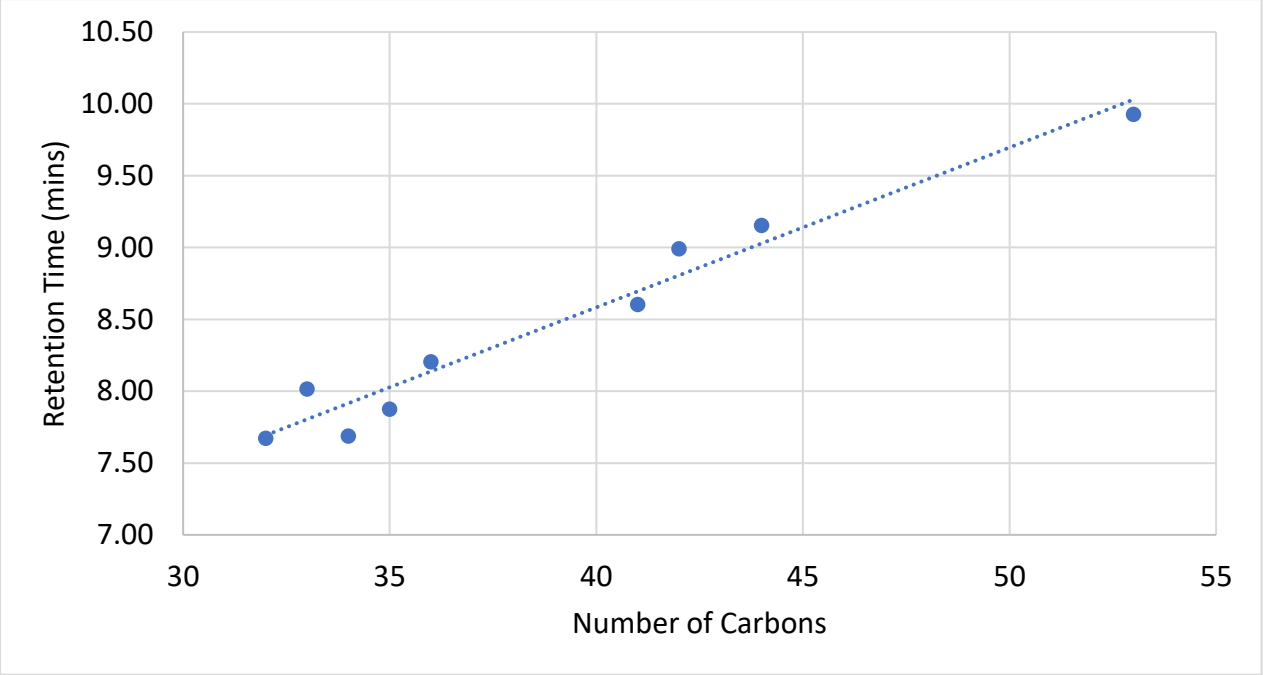


Figure 4-21: The trend in mean retention time as the number of carbons increased in the metabolites classed as Glycerophosphoethanolamines by LipidSearchTM. All metabolites contained one double bond in the structure. Line of best fit equation: $y = 0.1113x + 4.1328$; $R^2 = 0.9595$

When creating a model for the glycerophosphoethanolamines, there was some large gaps in the data. Although this didn't influence the ability to model the data, it did mean that any of the validation dataset that had glycerophosphoethanolamines that fitted into those large gaps would have potentially poor predictions. The validation set showed a great difference in the predicted RT compared to the experimental RT (Table 4-14).

Table 4-14: The mean retention time (RT) detected, the predicted RT and the difference between the two in minutes and in seconds for the Glycerophosphoethanolamines with one double bond present in the structure.

	RT Difference (Minutes)
Minimum Difference In Retention Time	0.81
Maximum Difference In Retention Time	1.49
Mean Difference In Retention Time	1.07
Median Difference In Retention Time	0.91

This meant that the model was not applicable for accurate metabolite annotations and should not be used to create a RT library but would be useful for reduction of false annotations as it helps to show the time range of where glycerophosphoethanolamines elute from the column.

4.2.3.7 Glycerophosphoinositols (GPI)

Glycerophosphoinositols are a subclass of glycerophospholipids. They consist of a glycerol backbone linked to two fatty acids and a phosphoinositol molecule. Each fatty acid can differ in length, saturation and hydroxylation. This can mean that despite two glycerophosphoinositols having the same molecular formulae, their structures can differ considerably. All glycerophosphoinositols that contained one, two or three double bonds were ordered by the total number of carbons present, and a model formed on the relationship between the number of carbons present and the retention time detected (Figure 4-22 and Table 4-15).

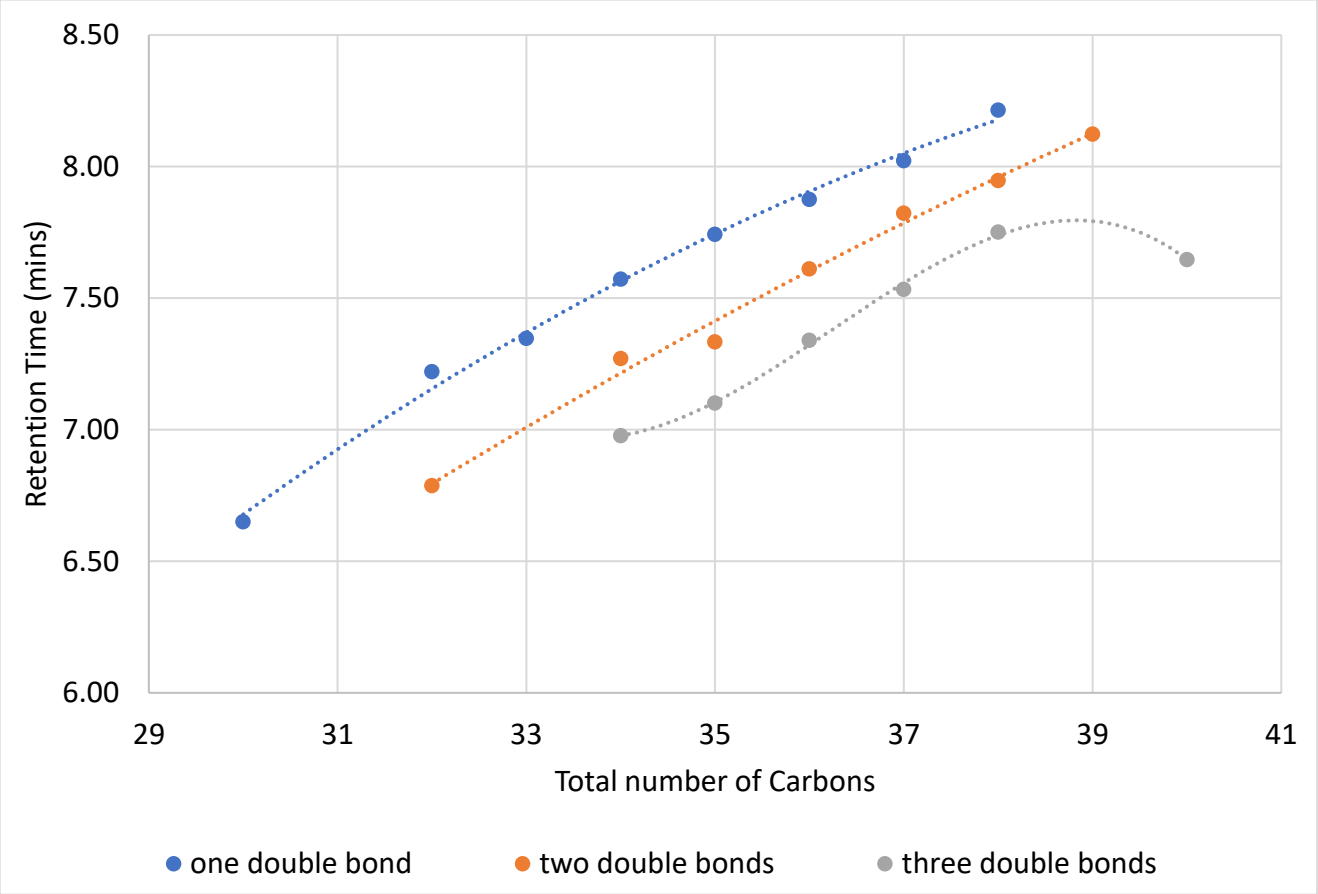


Figure 4-22: The trend in mean retention time as the number of carbons increased in the metabolites classed as Glycerophosphoinositols. The light blue data points represent the trend in the metabolites containing one double bond. The orange data points represent the trend in the metabolites containing two double bonds. The grey data points represent the trend in the metabolites containing three double bonds.

Table 4-15: Trend line equations and correlation coefficient values for the relationships between the total number of carbons and mean retention time for each set of Glycerophosphoinositols with a different number of double bonds.

Number Of Double Bonds Present In Lipid Structure	Trend Line Equation	Correlation Coefficient
One Double Bond	$Y = -0.0085x^2 + 0.7685x - 8.6894$	0.9952
Two Double Bonds	$Y = -0.004x^2 + 0.4756x - 4.3081$	0.9913
Three Double Bonds	$Y = -0.0121x^3 + 1.3132x^2 - 47.36x + 573.81$	0.9978

The retention time models formed from the mean retention time for metabolites with differing total numbers of carbons, but the same double bond can be seen above. These models looked to be very accurate with high correlation coefficients, despite the low number of data points.

When looking at the ranges of differences between the predicted retention times and the detected retention times in the validation data (Table 4-16), only the model for three double bonds was accurate enough to be used for a retention time library.

Table 4-16: The minimum, maximum, mean, and median differences in the retention time (RT) predicted by the models and the RT detected in minutes. All metabolites were Glycerophosphoinositols with either one, two, or three bonds present in their structure.

	Minimum Difference In Retention Time (Mins)	Maximum Difference In Retention Time (Mins)	Mean Difference In Retention Time (Mins)	Median Difference In Retention Time (Mins)
One Double Bond	0.99	1.02	1.00	1.01
Two Double Bonds	0.94	1.41	1.09	0.97
Three Double Bonds	0.09	0.33	0.23	0.24

From this is can be concluded that, similarly to other glycerophospholipid classes, there is a lot of variety in the retention time of each different structure of glycerophospholipid. Therefore, creating a retention time library is difficult as the structure changes so commonly with different combinations of fatty acids making up the same total number of carbons. One model that could be used to make a RT library was the three double bonds model for this class. Larger amounts of double bonds were present but not included in the models because there was not enough data available.

4.2.3.8 Triacylglycerides (TAG)

Triacylglycerides are a sub class of glycerolipids where the structure consists of a glycerol backbone linked to three fatty acids. These fatty acids can differ on length, saturation, and hydroxylation. Retention time models were formed from the mean retention time for metabolites with differing total numbers of carbons, but the same number of double bonds and can be seen below (Figures 4-23 and Table 4-17). These models looked to be accurate with high correlation coefficients, although the models with higher numbers of double bonds had less data points and so had potential to be over fitted.

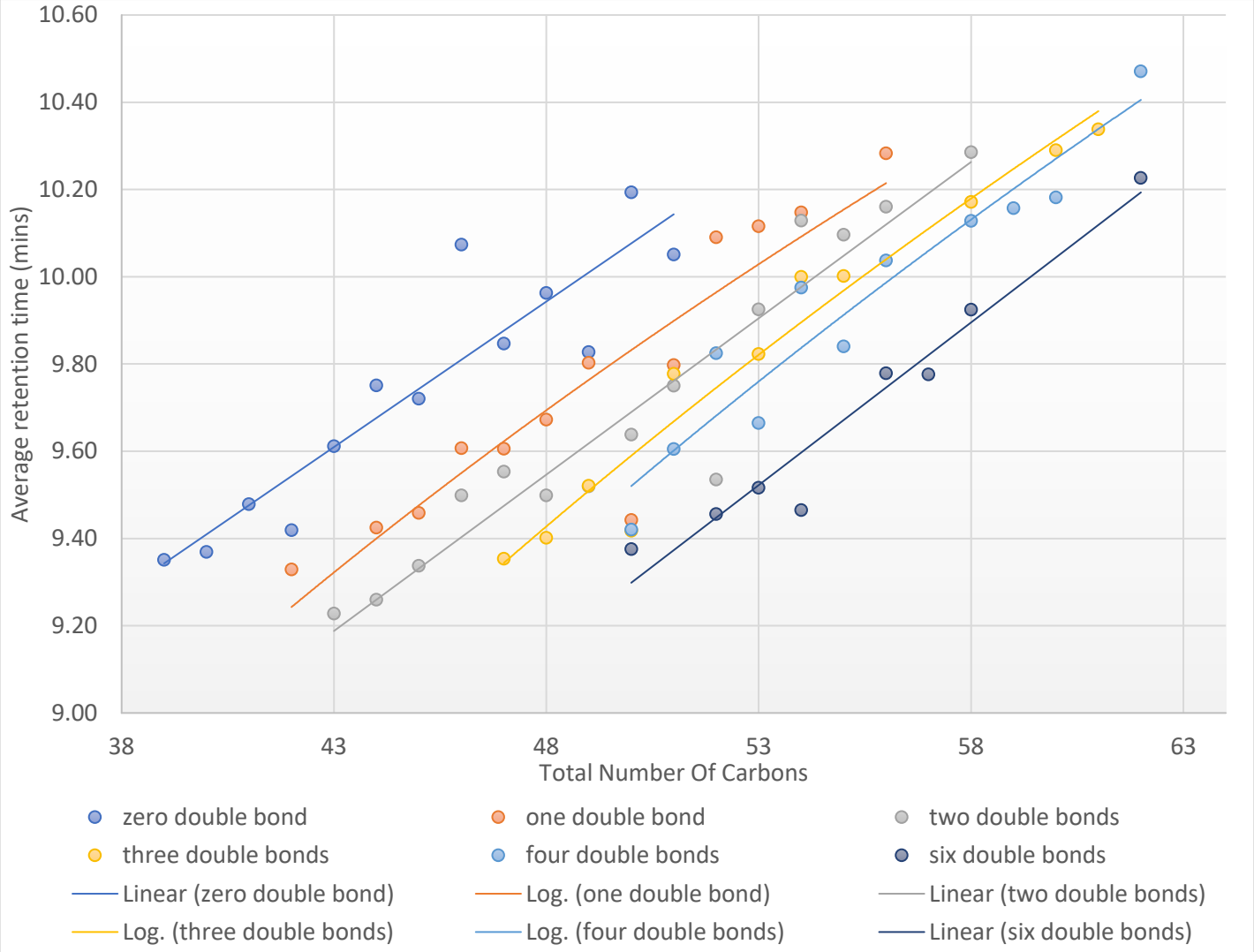


Figure 4-23: The trend in mean retention time as the number of carbons increased in the metabolites classed as Triacylglycerides with zero, one, two, three, four and six double bonds in the structure.

Table 4-17: Trend line equations and correlation coefficient values for the relationships between the total number of carbons and mean retention time for each set of Triacylglycerides with a different number of double bonds.

Number Of Double Bonds	Trend Line Equation	Correlation Coefficient
Zero Double Bonds	Y = 0.0667x + 6.7416	0.844
One Double Bond	Y = 3.3782ln(X) - 3.3839	0.8283
Two Double Bonds	Y = 0.0717x + 6.1064	0.9107
Three Double Bonds	Y = 3.9723ln(X) - 5.9503	0.9571
Four Double Bonds	Y = 4.1163ln(X) - 6.583	0.9121
Six Double Bonds	Y = 0.0746x + 5.5699	0.9522

Regression models differed between being linear and polynomial depending on the number of double bonds present in the model. This demonstrates how strong the influence of the structure of a lipid is when investigating the relationship between lipids and retention time. To validate the models above, the equations from the regression models were applied to a second dataset for cross-validation.

When looking at the ranges of differences between the predicted retention times and the detected retention times in the validation data (Table 4-18), none of the models were accurate enough to predict retention times that would allow for accurate annotations to be applied.

Table 4-18: The minimum, maximum, mean, and median differences in the retention time (RT) predicted by the models and the RT detected in minutes. All metabolites were Triacylglycerides with either zero, one, two, three, four, or six double bonds present in their structure.

	Minimum Difference In Retention Time (Minutes)	Maximum Difference In Retention Time (Minutes)	Mean Difference In Retention Time (Minutes)	Median Difference In Retention Time (Minutes)
Zero Double Bonds	0.44	0.89	0.77	0.78
One Double Bond	2.71	8.71	5.03	4.44
Two Double Bonds	2.42	11.27	5.32	4.97
Three Double Bonds	0.12	6.75	1.57	1.07
Four Double Bonds	5.26	10.38	7.82	7.62
Six Double Bonds	3.12	5.42	3.84	3.64

This inability to accurately predict retention times could be because there are three fatty acids that could be simultaneously changing in structure, which means that despite having the same molecular formulae, their retention on a column could differ significantly making predictions difficult. It is also another example that cross-validation is extremely important in retention time modelling for lipids because when one model looks accurate, there is a possibility that is not as robust as once hypothesised.

4.2.3.9 Classes Unable To Be Modelled

There were several lipid classes that were unable to be modelled for a variety of reasons. Diradylglycerolipids (DG), methylphosphocholine (MePC), Cardiolipins (CL), Hexosyl-1-ceramides (HEX-1-CER), Monoacylglycerophosphoglycerols (PG) and Glycerophosphates (GPA) were unable to be modelled due to the small number of data points available.

Sphingomyelins were unbale to be modelled as each subset of structures didn't seem to follow a clear trend in retention time. This may have been due to the structural differences between sphingomyelins with the same total number of metabolites. This is because the fatty acids can differ in length, the number of double bonds in either fatty acid or also the placement of the double bonds in fatty acid. Despite refining the metabolites to only use those where the mean retention time had an RSD <5, a usable model was not able to be formed

4.2.4 Lipidomics In-Silico Retention Time Library Conclusions

In summary, the complex structure of lipids makes their retention times very difficult to predict consistently. Out of all the retention time models presented above, only two were robust and accurate enough to be used for a retention time library. These were acylcarnitines (AcCA) with one double bond, and Glycerophosphoinositols (GPI) with three double bonds. The limitation of this research work was that all the data collected was experimental and so retention times of lipids may have shifted slightly due to matrix effects or dirty samples. Any shifts should have been minimal and were unlikely to have affected the work as only lipids with a minimum of three repeats and RSDs >5 were applied for the construction of models. The inaccurate prediction of lipids was most likely due to using the total number of carbons and number of double bonds to predict the retention time and would have possibly been more accurate if the structure had been further incorporated into the prediction model. There were several lipid classes that were unable to be modelled due to a lack of data or an unclear retention time trend. This too limited the research conducted, and for future work I would suggest that these classes be researched further to see if any models applicable for a RT library can be produced. One advantage to completing this work is that it is clear the retention time range for each class of lipids across multiple sample types. This means that when annotating compounds, the number of false annotations can be reduced by eliminating possible lipids that would not elute at the time of the compound in question. These elution ranges can be seen below (Figure 4-24).

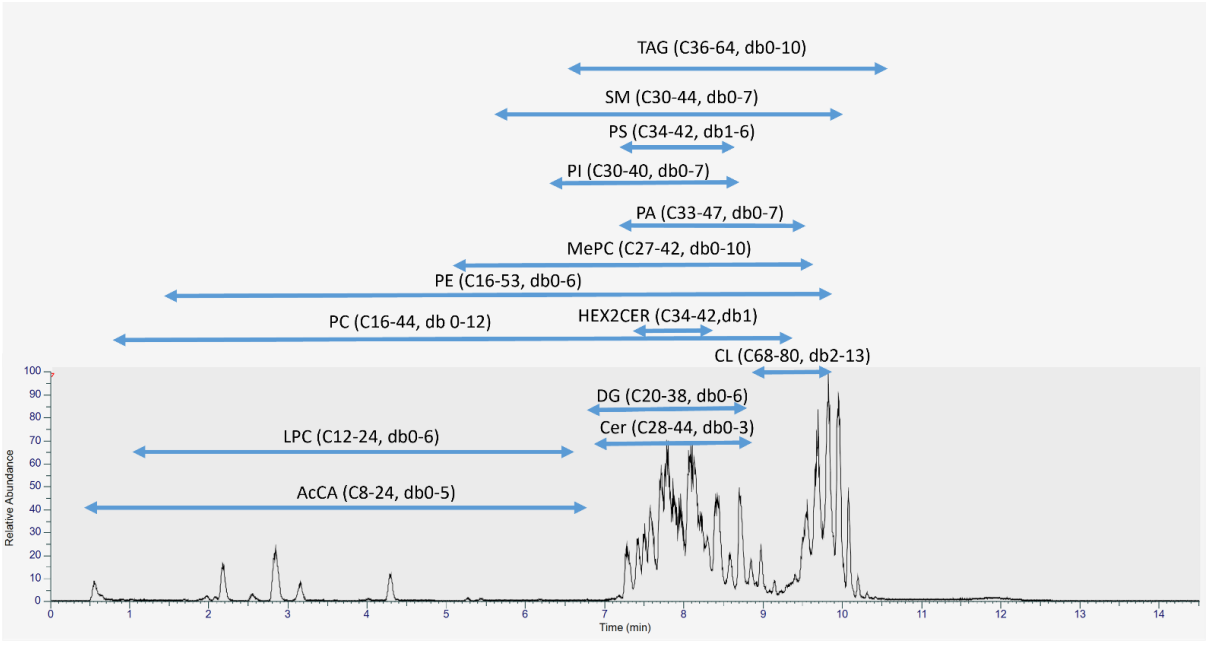


Figure 4-24: The different elution zones for different classes of lipids. AcCA = Acytlcarnitines, CL = Cardiolipins, DG = Diacylglycerols, HEX2CER =Hexosyl-2-ceramides, LPC = lyso- phosphocholines, MePC = Monoetherglycerophosphocholines, PA=Phosphatidates, PC = Phosphocholines, PE = Phosphatidylethanolamines, PI=Phosphatidylinositols, PS=Phosphatidylserines,

SM=Sphingomyelins, TAG=Triacylglycerides. The carbon range is denoted by CX-Y, and the double bond range is denoted by dbX-Y.

Through looking at the mean retention times of the different lipid classes, it can be determined when some of the classes will no longer be eluting from the column, and therefore any annotations of that class are no longer relevant or accurate.

4.3 CONCLUSIONS

4.3.1 Retention Time Mapping

Retention time mapping as a form of increasing metabolite annotation confidence is possible both for annotating metabolite matches, as well as reducing the number of false annotations made. Isomeric separation is possible depending on the assay gradient applied, with some libraries having better isomer separation than others. HILIC assays are problematic for retention time mapping due to the pH changes and the different buffers applied, which makes achieving accurate mapping models is temperamental. Accurate mapping models can be created and the models with slightly lower accuracy are still applicable to be used for reducing the number of false annotations. Reversed phase assays are easily mapped across groups, with differences in column, mobile phase or elution gradient having little effect on the accuracy of the model. To ensure an appropriate model is formed, analytical standards applied should separate well on both methods and should cover the whole elution gradient.

4.3.2 Lipid Library

In conclusion, complex lipid structure makes retention times very difficult to predict. This is most likely because this research was using the total number of carbons and number of double bonds to predict the retention time and would have possibly been more accurate if the structure had been further incorporated into the prediction model. What the research does allow for, is the removal false positive annotations.

4.4 FUTURE WORK

4.4.1 Retention Time Mapping

Future work on this area would be to apply the extended retention time libraries to biological data to see if any of the metabolites annotated under lower confidence levels could be improved and/or some of the false annotations removed. Work should also be completed to work out the limits of this method, so that it is clear what metabolites can and can't be mapped across well when applying this method. To translate this work from research into a useable product within the metabolomics community, there is opportunity here to create a large central RT library from which a small selection of analytical standards could be run by a research group to allow for the central RTs to be mapped onto their methods. This

would allow for many more groups to gain higher confidence annotations without needing an exact match of instrumentation and analytical methods.

4.4.2 Lipidomics In-Silico Retention Time Library

Future work here would be to investigate if working at a level which included more structural information would allow for the method to work better. Other future work should be to investigate how much this work has affected the ability to remove false negatives. Since this work has been completed (December 2020), similar work has been published[247]. Vaňková, et.al applied lipid standards from multiple lipid classes (NIST Standard reference material 1950 human plasma) from one sample type as opposed to biological annotations from multiple sample types. Their work concluded that multiple lipid classes could be modelled on their retention behaviour. Although their research also shows that for some classes, retention was affected by isomers of the same lipid, with their retention models separating out isomers of the same lipid (example PC 16:1/18:2 and PC 14:0/20:3 can both be denotated as PC 34:3).

Chapter 5

5 An Investigation Into How The Type And Intensity Of Exercise Perturbs Metabolism

---

Samples for this study were collected and provided by Samuel R. Weaver, and Samuel J. E. Lucas to study the influence of exercise on cerebral hemodynamic and neurotrophic factor responses[206]. Study design will therefore overlap substantially.

5.1 Introduction

Metabolomics has most commonly been applied within the medical research fields for:[248]

- Understanding of underlying molecular mechanisms of diseases[249,250]
- Sub-classification of diseases, and stratification of patients based on metabolic pathways impacted[251]
- Revealing biomarkers for drug response phenotypes and providing an effective means to predict variation in a subject's response to treatment (pharmacometabolomics)[252,253]
- Providing a means to monitor response and recurrence of diseases, such as cancers[179,254] and diabetes[181]
- Defining the molecular landscape in human performance applications and extreme environments. [190,193,255]

It is well known that physical exercise plays an important part in keeping people healthy[256-260]. Therefore, applying metabolism research into health and exercise science is paramount to fully understand how exercise affects the human metabolome at the cellular and biofluid levels. Exercise alters many metabolic pathways at once, for example lipolysis pathways[261], hormone regulation[262] and cell signalling[263]. Investigating the effects of exercise on the human body is not only important for health research but is also useful for sports performance and nutritional needs. It also aids in the understanding of how different exercise types affect the body at the pathophysiological level.

There are a wide variety of studies that have been conducted when applying metabolomics to study molecular aspects of exercise. These include but are not limited to: comparing metabolic profiles of athletes from a wide range of sporting disciplines[183], comparing different elite level athletes from different disciplines[183], investigating low intensity exercise[184], studying marathons/ultramarathons[185,186], investigating short duration high intensity exercise[187,188], and studying long duration high intensity exercise[189]. Much of the exercise-based metabolomics research is targeted metabolomics and only focusses on sets of metabolic pathways[264-270]. The issue with this is that metabolomic pathways work together as a metabolic network and not as small pathways. So looking at what changes are occurring globally within the metabolome can provide a much clearer picture of the metabolic changes taking place and the affects that exercise have on the human body[190]. The most common sample types applied in exercise studies are plasma[137,191-193], urine[194-196] and sweat[168,197,198].

Exercise is a very broad term as there are many ways to exercise from resistance training, to cycling, to playing individual or team sports. Three very common forms of exercise are moderate-intensity continuous training (MICT), high-intensity interval training (HIIT) and sprint interval training (SIT). MICT exercise can be summarised as 30-60 minutes of exercise where the peak heart rate sits at about 64-76% in comparison to HIIT and SIT exercise protocols where peak heart rate reaches 90% and 170% respectively[271] . Interval training types are used to push the participant to exercise at a higher peak heart rate for short periods of time, separated by short recovery periods. SIT is the highest intensity of the three (supramaximal exertion) followed by HIIT (near maximal exertion) and then MICT.

Variations of HIIT exercise have been claimed to be around since the early 20th Century, although it was not popularised in scientific literature until the 2010s[272-274]. There are many ways to undertake HIIT exercise but there are two main categories: bodyweight[275] (such as Tabata and boot camp) and aerobic (running and cycling)[276,277]. SIT exercise protocols are similar to that of, and sometimes included in HIIT. The main difference between HIIT and SIT is the level of exertion reached in the exercise period. Due to this, SIT exercise protocols are much shorter than HIIT.

Previously there has been a lot of research where the study focused on one[187,278-280] type of exercise and others that have focused on two types of exercise[281-286], but there are not many studies that look in at all three exertion areas[287-289]. These studies are frequently looking at targeted pathways or are interested in a disease and the different affects that different types of exercise can have. There is a clear research gap where there is limited knowledge relating to the metabolic differences between different exercise types undertaken by healthy human participants.

5.1.1 Research Objectives

This research study will apply untargeted metabolomics to determine metabolic changes over time related to each exercise type to investigate how the metabolism changes across different time points and throughout recovery afterwards. Another objective of this research was to see if there any

differences between exercise protocols at the same time points; this will show if there are any major differences between protocols and which exercise protocol puts the most metabolic stress on the body. Finally, the last objective of this research was to determine if there are any metabolic differences between males and females in their response to each different exercise protocol. This is important because it could help to define more about the health requirements of both sexes.

5.2 MATERIALS AND METHODS

5.2.1 Study Design

This samples used in this study were originally collected as part of a previous study published by Samuel R. Weaver[206]. In which, the full experimental protocol can be found. Description of each exercise visit can be found in section 2.8.1 of the Materials And Methods chapter.

5.2.2 Methods

Methods for this research chapter can be found in the Materials And Methods chapter. The ethics statement is in section 2.2.2, sample collection details are under section 2.3.2, sample extraction details are under section 2.4, UHPLC-MS methods are in section 2.8.3, and data processing and metabolite annotation can be found in section 2.8.3.

5.3 RESULTS AND DISCUSSION

5.3.1 What Are The Metabolic Changes Over Time Related To Exercise Type?

To answer this question, each exercise protocol (HIIT, MICT, and SIT) was explored independently looking at exercise sampling time points *vs.* participant. This was completed by applying a one-way repeated measures (within subjects) ANOVA (OWRMA). After this, exercise time points were individually compared against each other using paired t-tests to see at which time point the compound's response were changing significantly.

5.3.1.1 MICT

Compounds that were changing statistically over time were filtered down to those annotated and with a clear time point where the statistically significant change could be reported. After filtering there were four water soluble compounds (detected in the HILIC assay) that statistically changed over time, none of which were linked together in a single metabolic pathway. In the lipid classes reported, five fatty acids and three acyl carnitines significantly changed in relative concentration over time, with seven increasing in concentration over time. The fold changes for the compounds are reported in Table 5-1.

The majority of the changes in relative concentration occurred either from at rest to active recovery or at rest to passive recovery.

Table 5-1: The fold changes for compounds that were found to be statistically significant in the one-way repeated measures ANOVA as well as in the separate t-tests from the MICT protocol. Fold changes had a statistical significance of $P<0.05$ and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. All changes that were not statistically significant are labelled as NSS. Table Key: OWRMA = one-way repeated measures ANOVA fold change, T1 = at rest, T2 = end of exercise, T3 = active recovery, T4 = passive recovery.

		OWRMA		T1vsT2	T1vsT3	T1vsT4	T2vsT3	T2vsT4	T3vsT4
Compound Class	Compound Name	F-value	Adjusted P-val	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change
Fatty Acyls	Hydroxydodecanoic acid	7.53	1.31E-02	NSS	NSS	2.61	NSS	4.18	3.08
Fatty Acyls	Oleic acid; Elaidic Acid; trans-Vaccenic acid; cis-Vaccenic acid; Octadecenoic acid; Petroselinic acid; Isooleic acid; Ethyl (9E)-9-hexadecenoate; 11-Cycloheptylundecanoic acid; cis-10-oleic acid; 2-Methyl-16-heptadecenoic acid	13.27	5.27E-04	NSS	2.45	2.55	NSS	NSS	NSS
Fatty Acyls	Palmitic Acid; Lauryl isobutyrate; Hexyl decanoate; 4,8,12-Trimethyltridecanoic acid; Dodecyl butyrate; Methylpentadecanoic acid; 2-Heptylnonanoic acid; 2-Butyl-2-ethyldecanoic acid; 2,4-Dimethyltetradecanoic acid; 4-Hexyldecanoic acid; 2-Propyltridecanoic acid; Dimethyltetradecanoic acid; 6-Ethyltetradecanoic acid; 3-Ethyl-3-methyltridecanoic acid	8.45	6.57E-03	NSS	NSS	2.60	NSS	NSS	NSS
Fatty Acyls	Ricinoleic Acid; ketostearic acid; METHYL PALMOXIRATE; Oxooctadecanoic acid; Hydroxyoctadecenoic acid; Lactaric Acid; oxostearic acid; cis-6,7-epoxystearic acid; Isoricinoleic Acid; hydroxyoleic acid; 8-(5-Hexyltetrahydro-2-furanyl)octanoic acid; Methyloxoheptadecanoic acid	13.61	4.82E-04	NSS	NSS	NSS	NSS	3.27	2.69
Fatty Acyls	Decanoylcarnitine; 4145305	11.21	1.03E-03	2.12	NSS	NSS	NSS	NSS	NSS
Fatty Acyls	floionic acid; Sorbitan, monododecanoate; Hydroperoxy-dihydroxy-octadecenoic acid;	6.31	4.54E-02	0.50	NSS	0.40	NSS	NSS	NSS
Fatty Acyls	Tetradecadiencarnitine	9.16	3.34E-02	NSS	2.23	NSS	NSS	NSS	NSS
Fatty Acyls	Hexadecadienoylcarnitine	8.96	3.39E-02	NSS	2.12	NSS	NSS	NSS	NSS
Water Soluble	Hypoxanthine	12.55	6.06E-04	NSS	NSS	2.75	NSS	NSS	NSS
Water Soluble	Inosine	13.81	3.58E-04	NSS	NSS	2.64	NSS	NSS	NSS
Water Soluble	Dimenhydrinate	7.99	9.25E-03	3.12	5.34	6.08	NSS	NSS	NSS
Water Soluble	Retinyl acetate	10.73	1.97E-03	NSS	NSS	NSS	NSS	2.06	NSS

Following investigation of the trends seen in the OWRMA, t-tests of singular time point comparisons were inspected to see if more changes in relative concentration can be seen and if more compounds are statistically significant. As most changes occurred when comparing the "at rest" time point and the recovery time points, those t-tests were isolated and investigated further.

5.3.1.1.1 At Rest Vs Active Recovery

Focussing on the t-test between the at rest and active recovery time points revealed a lot more significant changes, in comparison to the OWRMA results. Compounds were filtered down to only those with annotations, this meant that 185 compounds were narrowed down to 54. 34 were water soluble compounds and 20 were lipid compounds. Pathway enrichment analysis was conducted on the water-soluble metabolites and the purine metabolism pathway had 3/65 metabolites present with a p-value of 6.63E-04 (Figure5-2). The purine metabolism pathway plays an important role during exercise as it helps to balance the adenine nucleotide pool (adenosine triphosphate (ATP), adenosine diphosphate (ADP) and adenosine monophosphate (AMP)) which are degraded during exercise for energy in the mitochondria of the cell. So the increase in relative concentration of adenosine as ATP, ADP and AMP were degraded for energy was not unexpected.

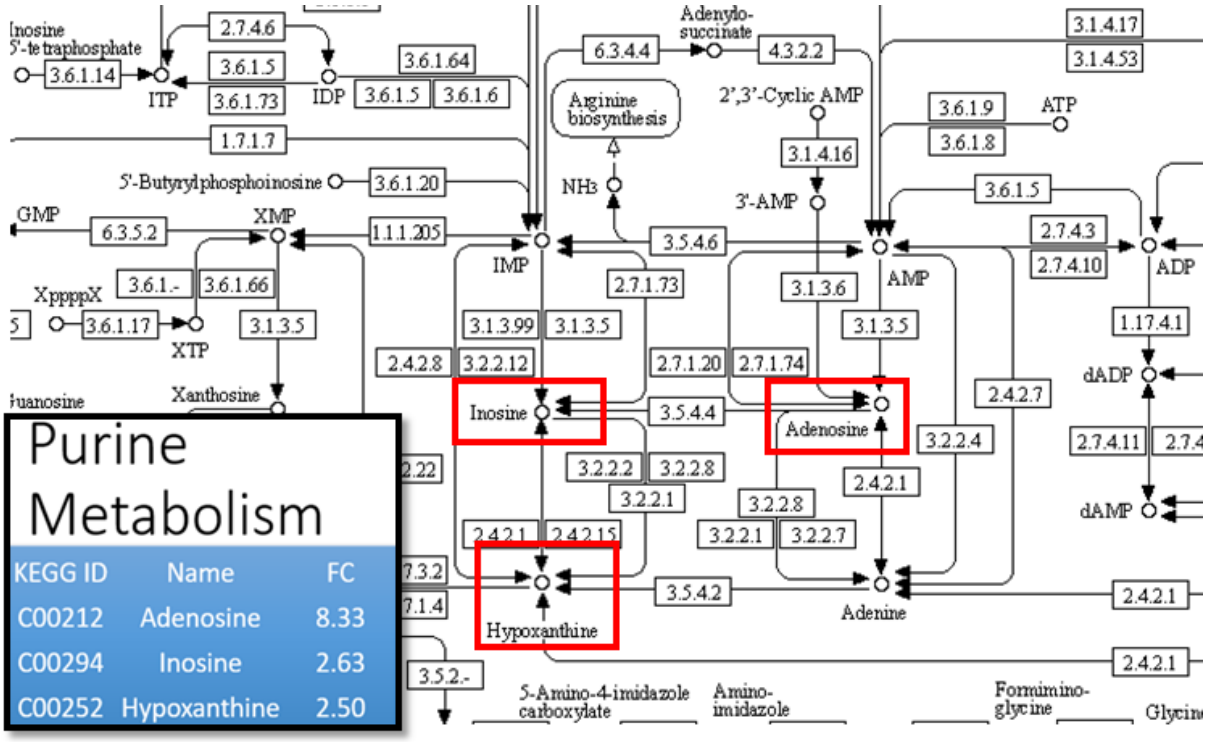


Figure 5-1: The purine metabolism pathway as recreated from KEGG Pathway Database[290,291] and adapted to highlight the changing metabolites, along with the fold change increases seen in the paired t-test between starting rest vs active recovery. Fold change calculations were active recovery / starting rest.

Out of all the lipids that were changing the majority were fatty acyls, which increased in relative concentration (see Table in Appendix H-1). This increase is regularly recorded across literature[292-294] as the cells require more energy with increased metabolic output. This increase in relative concentration of fatty acyls comes from lipolysis when triglycerides are broken down to release free fatty acids which are then into acyl carnitines where they are transported to the mitochondria of the cell before undergoing

ß-oxidation. During ß-oxidation the acyl carnitines are transformed into fatty-acyl-CoA and then acetyl-CoA which are used in the TCA cycle to produce ATP for energy.

5.3.1.1.2 At rest vs Passive recovery

Focusing on the t-test for the at rest vs passive recovery (table in appendix H-2) showed some similarities to the results of the comparison above. There was once again an overall increase in fatty acyls although there was a much larger mix of increases and decreases of relative concentration across the larger lipid classes (glycerophospholipids and triglycerides).

5.3.1.2 HIIT

For the HIIT protocol, compounds that were found to have significantly changed in relative concentration over time in the OWRMA results were filtered down to those that had annotations and one or more time points where a statistically significant change(s) were reported in the post-hoc analysis. Once filtered (Table 5-2), there were 20 water soluble compounds that statistically changed over time. However, after pathway enrichment analysis had been conducted there was not any pathways which were significantly enriched. This is most likely due to the low number of significant water-soluble compounds detected. It could also be due to the compounds being present all being from different pathways, or only one to two compounds in a pathway being detected. There were several lipid classes that contained compounds that were significantly changing in concentration, most commonly fatty acyls and glycerophospholipids (Figure 5-2). Fatty acyls decreased over time, whilst concentrations of glycerophospholipids initially increased before some started to decrease in concentration. This result is different from the MICT protocol where fatty acyls increased. The difference here could be due to the higher amount of energy required in a smaller amount of time for the HIIT protocol, with the rapid need for energy the cells switch from using lipolysis for energy to using glucose. Although this is the probable cause, upon investigation into if any changes in glucose metabolism were occurring, none were detected.

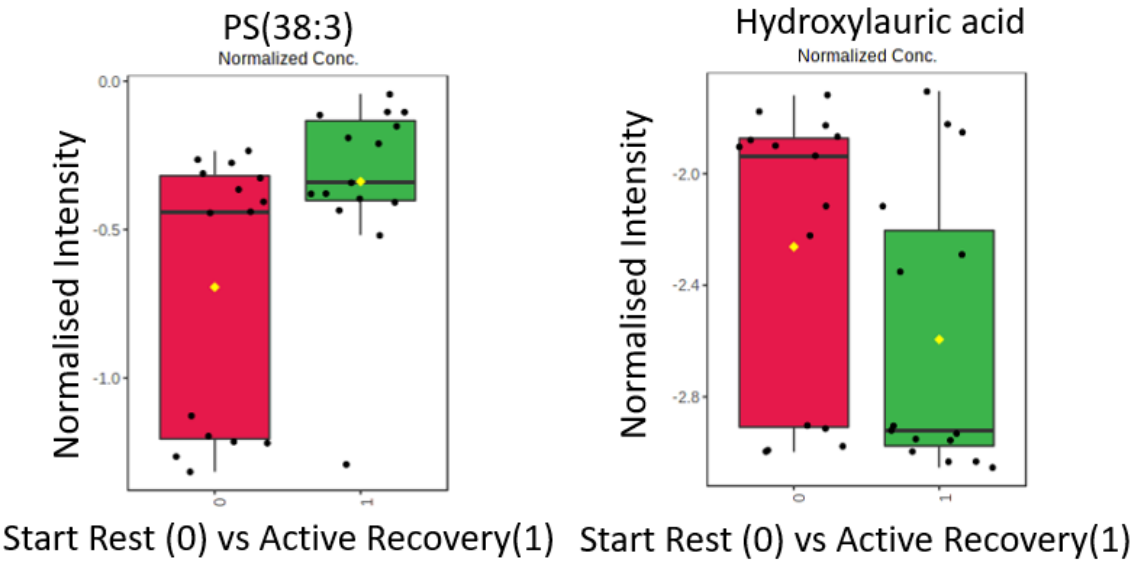


Figure 5-2: Normalised box plots of two different lipids that increased in relative concentration when compared at rest (0) and active recovery (1)

Table 5-2: The fold changes for lipid compounds that were found to be statistically significant in the one-way repeated measures ANOVA as well as in the t-tests for Starting Rest (T1) vs Active Recovery (T4) and Starting Rest (T1) and Passive Recovery (T5) time points from the HIIT protocol. Fold changes had a statistical significance of $P<0.05$ and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. All changes that were not statistically significant are labelled as NSS.

		overall	T1vsT2	T1vsT3	T1vsT4	T1vsT5	T2vsT3	T2vsT4	T2vsT5	T3vsT4	T3vsT5	T4vsT5
Metabolite Class	Compound Name	F-value	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change
Fatty Acyl	Oxotetradecanoic acid;Ketomyristic acid;(octyloxiranyl)butanoic acid;hydroxyoxotetradecanal;hydroxytetradecenoic acid;L-Menthyl (R,S)-3-hydroxybutyrate	6.3563	0.34	NSS	NSS	NSS	NSS	NSS	2.73	NSS	NSS	NSS
Fatty Acyl	Oleic acid;Isooleic acid;Petroselinic acid;Octadecenoic acid;Cycloheptylundecanoic acid;Methylheptadecenoic acid;Vaccenic acid;Elaidic Acid;Ethyl (9E)-9-hexadecenoate	9.608	NSS	NSS	NSS	NSS	NSS	2.45	2.47	NSS	NSS	NSS
Fatty Acyl	Acetyl-L-carnitine	9.9096	NSS	0.42	2.36	NSS	2.47	2.47	2.02	NSS	NSS	NSS
Fatty Acyl	(5-hydroxy-7-methoxy-2,2-dimethyl-3,4-dihydro-2H-1-benzopyran-4-yl)oxidanesulfonic acid;5-(3-methoxyphenyl)-4-(sulfooxy)pentanoic acid;5-[3-methoxy-4-(sulfooxy)phenyl]pentanoic acid;((4-[(4-methoxyphenyl)methoxy]-4-oxobutan-2-yl)oxy)sulfonic acid;(4-[(4-methoxyphenyl)methoxy]-4-oxobutoxy)sulfonic acid;((1-[(4-methoxyphenyl)methoxy]-1-oxobutan-2-yl)oxy)sulfonic acid	10.583	NSS	NSS	NSS	0.47	NSS	NSS	0.46	NSS	NSS	NSS
Glycerophospholipids	PS(38:3)	6.9276	2.52	NSS	2.27	2.17	NSS	NSS	NSS	NSS	NSS	NSS
Glycerophospholipids	LysoPC(17:0)	5.4249	NSS	NSS	NSS	NSS	NSS	NSS	0.47	NSS	NSS	NSS
Glycerophospholipids	LysoPE(18:1)	4.7856	NSS	NSS	NSS	0.46	NSS	NSS	0.45	NSS	NSS	NSS
Water Soluble	Hydroxylauric acid;Hydroxydodecanoic acid	8.4478	0.40	2.91	0.47	NSS	NSS	NSS	2.10	NSS	2.46	NSS
Water Soluble	Dimenhydrinate	7.6673	NSS	NSS	NSS	NSS	NSS	3.27	4.18	3.30	4.31	NSS
Water Soluble	Retinyl acetate;Docosahexaenoic acid;3546;6?-methylprogesterone;Zuretinol acetate;O94MU16592;Grifolin;2?-methylprogesterone;5-Methyl-4-[(2E,6E)-3,7,11-trimethyl-2,6,10-dodecatrien-1-yl]-1,3-benzenediol;7alpha-Methylprogesterone;10,13,16-Docosatrienoic acid;2-Methyl-5-[(8Z,11E)-8,11,14-pentadecatrien-1-yl]-1,3-benzenediol;8,11,14-Docosatrienoic acid;3,9,15-Docosatrienoic acid;(4E,7E,10E,13E,16Z,19Z)-4,7,10,13,16,19-Docosahexaenoic acid	10.168	NSS	2.07	NSS	NSS	NSS	NSS	2.24	NSS	2.73	NSS
Water Soluble	Ribose-1-arsenate	7.8866	NSS	0.49	NSS	NSS	NSS	NSS	NSS	NSS	NSS	NSS
Water Soluble	Tricin 7-[p-coumaroyl-(->2)-glucuronyl-(1->2)-glucuronide];Chrysoeriol 7-[feruloyl-(->2)-glucuronyl-(1->2)-glucuronide]	6.6888	NSS	NSS	NSS	NSS	NSS	NSS	3.37	NSS	2.64	2.04
Water Soluble	17-beta-estradiol 3-sulfate-17-(beta-D-glucuronide)	7.0149	NSS	NSS	NSS	2.12	NSS	2.62	2.77	NSS	NSS	NSS
Water Soluble	22-Acetylpriverogenin B	5.6331	0.35	3.15	NSS	NSS	NSS	3.52	4.23	3.91	4.71	NSS
Water Soluble	3,4,5-trihydroxy-6-{4-[(6E)-3-oxo-7-phenylhepta-4,6-dien-1-yl]phenoxy}oxane-2-carboxylic acid;3,4,5-trihydroxy-6-{3-[(6E)-3-oxo-7-phenylhepta-4,6-dien-1-yl]phenoxy}oxane-2-carboxylic acid;3,4,5-trihydroxy-6-{4-[(1E)-5-oxo-7-phenylhepta-1,3-dien-1-yl]phenoxy}oxane-2-carboxylic acid;3,4,5-trihydroxy-6-{3-[(1E)-5-oxo-7-phenylhepta-1,3-dien-1-yl]phenoxy}oxane-2-carboxylic acid;3,4,5-trihydroxy-6-{[(6E)-3-oxo-1,7-diphenylhepta-4,6-dien-1-yl]oxy}oxane-2-carboxylic acid;3,4,5-trihydroxy-6-[(3-oxo-1,7-diphenylhepta-4,6-dien-2-yl)oxy]oxane-2-carboxylic acid	5.7798	NSS	NSS	NSS	NSS	NSS	3.37	3.29	2.72	2.60	NSS
Water Soluble	N-(4-{[1-hydroxy-3-(4-hydroxy-3-methoxyphenyl)propylidene]amino}butyl)-3-(4-hydroxy-3-methoxyphenyl)prop-2-enimidic acid	10.092	2.80	NSS	NSS	NSS	NSS	3.35	5.54	2.92	4.68	NSS
Water Soluble	3-Methoxyestra-tetraene; [(Methylphenyl)hexanyl]phenol	4.7346	2.80	0.48	2.09	2.14	NSS	NSS	NSS	NSS	NSS	NSS
Water Soluble	Methylguanine	6.1229	0.47	NSS	NSS	NSS	2.16	2.19	2.13	NSS	NSS	NSS
Water Soluble	Pantothenic acid	85.219	NSS	0.49	2.13	NSS	NSS	NSS	NSS	NSS	NSS	NSS
Water Soluble	Hypoxanthine;2H-pyrazolo[4,3-d]pyrimidin-7-ol;Allopurinol	7.454	NSS	0.49	2.05	NSS	NSS	NSS	NSS	NSS	NSS	NSS

5.3.1.2.1 At Rest Vs Active Recovery

Paired t-test results from the comparison of relative concentration of compounds at rest and after active recovery showed a concentration decrease in the fatty acyls and a mixture of concentration increases and decreases in the glycerophospholipids (Table 5-3). This change is again different to the pattern of results seen in the MICT protocol. One reason for this could be the body switching from lipolysis as a source of energy to glucose, this would occur because the body requires more energy at a higher pace for the amount of effort being exerted. Here the decrease in fatty acyls (mainly fatty acids) along with the increase in larger molecular weight lipids suggests that the fatty acyls are being used to produce glycerophospholipids.

Table 5-3: The fold changes for compounds that were found to be statistically significant in t-tests for At Rest (T1) vs Active Recovery (T4) time points from the HIIT protocol. All fold changes had a statistical significance of P<0.05 and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. The glycerophospholipids are denoted by Class (number of carbons: number of double bonds). GPC= Glycerophosphocholines, GPA=Glycerophosphates, GPE/ GPE-NMe/ GPE-NMe2 = Glycerophosphoethanolamines, LysoGPE= Glycerophosphoethanolamines, GPS= Glycerophosphoserines.

Metabolite Class	Compound Name	Fold Change	Raw p-value
Fatty Acyl	Oleic acid; Elaidic Acid; Vaccenic acid; Octadecenoic acid; Petroselinic acid; Isooleic acid; Ethyl-hexadecenoate; Cycloheptylundecanoic acid; Methyl-heptadecenoic acid	2.49	2.15E-02
Fatty Acyl	Montecristin	2.09	8.51E-03
Fatty Acyl	floionic acid; Sorbitan, monododecanoate	2.48	4.75E-03
Fatty Acyl	Acetylcarnitine	0.42	1.54E-04
Fatty Acyl	Pantothenic acid	0.47	1.80E-10
Glycerolipids	MG(24:1)	0.33	4.13E-02
Glycerophospholipids	LysoGPC(18:2/0:0); ({[hydroxy[octadecadienoyloxy]propylphosphonato]oxy}ethyl) trimethylazanium	0.48	4.88E-02
Glycerophospholipids	GPC(40:7)	0.38	4.37E-02
Glycerophospholipids	LysoGPE(22:6)	0.47	4.24E-02
Glycerophospholipids	GPC(40:8); GPE-NMe(42:8)	2.63	2.29E-02
Glycerophospholipids	GPS(38:3)	0.44	1.73E-02
Glycerophospholipids	GPE(38:5)	0.27	5.81E-03
Glycerophospholipids	GPE(40:6)	3.42	3.06E-03
Prenol Lipids	Ubiquinol-10	0.40	1.89E-02
Sterol Lipids	Cholestadienone;Dehydrodesmosterol; Secocholestatetraenol; zymosterol intermediate 2	3.32	2.22E-02
Water Soluble	Urobilinogen;Urobilin;	0.43	4.17E-02
Water Soluble	Cytosine; Imexon	2.46	4.02E-02
Water Soluble	Oxagrelate; [Hydroxypropanoyl]tryptophan	0.37	1.06E-02

Water Soluble	Aspartylphenylalanine	0.18	5.69E-03
Water Soluble	Isothiocyanato(methylsulfinyl)hexane; Lipoamide	0.43	4.15E-03
Water Soluble	Hypoxanthine; Allopurinol; 2H-pyrazolo[4,3-d]pyrimidin-7-ol	0.49	2.27E-04

5.3.1.2.2 At Rest Vs Passive Recovery

Paired t-test results from the comparison of relative concentration of compounds at rest and after passive recovery (table in appendix H-3) showed a concentration decrease in the fatty acyls and a mixture of concentration increases and decreases in the glycerophospholipids. This change is different to the pattern of results seen in the MICT protocol. One reason for this could be the body switching from lipolysis as a source of energy to glucose, this would occur because the body requires more energy at a higher pace for the amount of effort being exerted.

5.3.1.3 SIT

For the SIT protocol, compounds that were found to have changed significantly in relative concentration over time in the OWRMA results were filtered down to those that had annotations and one or more time points where a significant change in relative concentration could be seen. Once filtered (Table 5-4), there were 40 water soluble compounds that significantly changed over time. After pathway enrichment analysis there was no pathway that was statistically enriched. This may have been because not enough compounds were detected, or because there were not enough compounds present in the pathways for it to be classed as significant. There were several lipid classes that contained compounds that were significantly changing in relative concentration, most commonly fatty acyls and glycerophospholipids. Compounds in both lipid classes decreased in relative concentration over time, which was different to both HIIT and MICT protocols and suggests that the main metabolite pathway that was being used for energy was the carbohydrate pathway where glycolysis occurs.[295] Although there was no evidence of this in the plasma samples analysed. This assumption comes from the fact that signs of lipolysis were not occurring in a high enough quantity to produce an increase in relative concentration in fatty acids or acyl carnitines, both of which occurs when lipolysis ensues.

The decrease in fatty acyls was much more prominent in the at rest vs passive recovery t-test (figure 5-3). This suggests that metabolism is changing in the time that the body has stopped moving.

Table 5-4: The fold changes for lipid compounds that were found to be statistically significant in the one-way repeated measures ANOVA as well as in the t-tests for Starting Rest (T1) vs Active Recovery (T3) and Starting Rest (T1) and Passive Recovery (T4) time points from the SIT protocol. All fold changes had a statistical significance of P<0.05 and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. All changes that were not statistically significant are labelled as NSS. The sphingolipids and glycerophospholipids are denoted by Class (number of carbons: number of double bonds). LysoGPA= LysoGlycerophosphates, LysoGPC= LysoGlycerophosphocholines, GPS= Glycerophosphoserines, SM= Sphingomyelins.

Lipid Class	Compound Name	T1vsT4	T1vsT5
Fatty Acyls	Decanoylcarnitine	0.48	0.27
Fatty Acyls	Hydroxyoctanoic Acid	NSS	2.23
Fatty Acyls	Tetradecadiencarnitine	NSS	0.33
Fatty Acyls	Hydroxydodecanoylcarnitine	0.48	0.29
Fatty Acyls	Tetradecenoylcarnitine	NSS	0.42
Fatty Acyls	Decanoylcarnitine	0.46	0.25
Fatty Acyls	Hexanoylcarnitine	NSS	0.27
Fatty Acyls	Hydroxylauric Acid	0.38	0.27
Fatty Acyls	Hydroxytetradecenoic Acid	0.45	0.20
Fatty Acyls	Dihydroxystearic acid	NSS	NSS
Fatty Acyls	Hydroxyoctadecenoic acid	NSS	0.38
Glycerophospholipids	LysoGPA(20:4)	0.35	NSS
Glycerophospholipids	LysoGPC(20:4)	0.48	NSS
Glycerophospholipids	GPS(38:6)	0.45	NSS
Sphingolipids	SM(42:2)	0.47	NSS
Steroids and steroid derivatives	Glycochenodeoxycholic acid	NSS	0.31
Steroids and steroid derivatives	Tauroursocholic acid	NSS	2.18

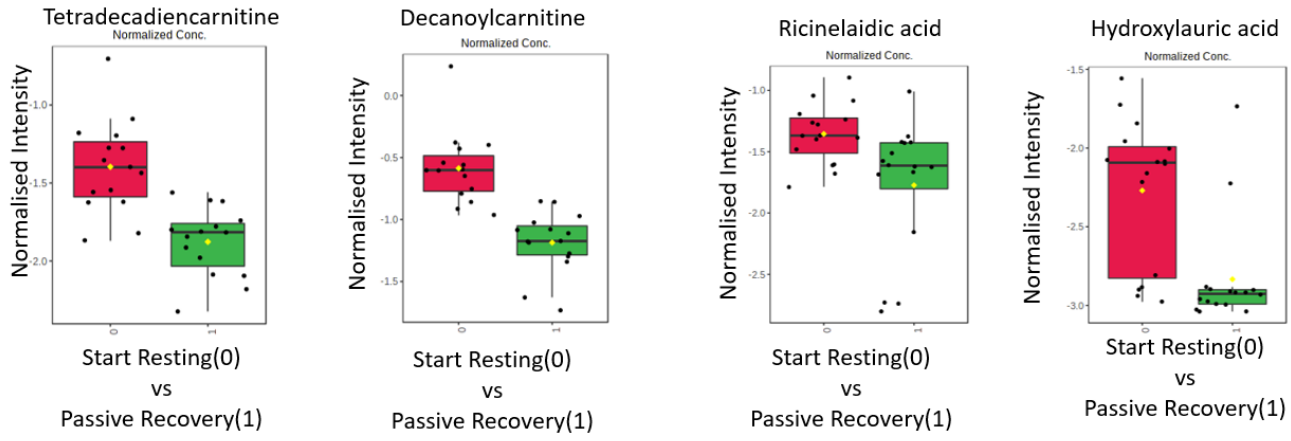


Figure 5-3: Normalised box plots of different fatty acyls that decreased in relative concentration when compared at rest (0) and passive recovery (1).

5.3.1.3.1 At Rest Vs Active Recovery

When disregarding the OWRMA results and focusing on the results of the pairwise t-test there was a larger number of changes in relative concentration over time, these changes contradicted what was seen in the OWRMA. As can be seen in Table 5-5, the main lipid class that was changing was glycerophospholipids, followed by glycerolipids and fatty acyls. The concentration changes of compounds were different from the OWRMA results with a large mix of fold change increases and decreases in the glycerophospholipids.

Table 5-5: The fold changes for lipid compounds that were found to be statistically significant in t-tests for Starting Rest (T1) vs Active Recovery (T4) time points from the SIT protocol. All fold changes had a statistical significance of P<0.05 and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. Fold change decreases are calculated as T4/T1. The sphingolipids, glyceroplipids and glycerophospholipids are denoted by Class (number of carbons: number of double bonds). LysoPA= LysoGlycerophosphates, LysoPC= LysoGlycerophosphocholines, LysoPE= Lyso Glycerophosphoethanolamines, PA=Glycerophosphates, PE/PE-NMe/PE-NMe2= Glycerophosphoethanolamines, PI= Glycerophosphoinositols, CL= Glycerophosphoglycerophosphoglycerols (Cardiolipins), CDP-DG= cytidine diphosphate diacylglycerol, PS= Glycerophosphoserines, SM= Sphingomyelins, MG= Monoradylglycerolipids, TG= Triradylglycerolipids.

Compound Class	Name	Fold Change	Raw p-value
Fatty Acyl	Acetylcarnitine	0.30	3.96 E-02
Fatty Acyl	Decanoylcarnitine; Heptadecenedioic acid	2.16	2.51 E-03
Fatty Acyl	Hydroxydodecanoylcarnitine; [(Hydroxydodecanoyl)oxy](trimethylammonio)butanoate	2.10	1.27 E-02
Fatty Acyl	Hydroxylauric acid; Hydroxydodecanoic acid	2.66	1.19 E-02
Fatty Acyl	Hydroxyoxotetradecanal; Hydroxytetradecenoic Acid; Oxotetradecanoic Acid; (Octyloxiranyl)Butanoic Acid; Hydroxytetradecenoic Acid; Ketomyristic Acid; Hdroxytetradecenoic Acid; Menthyl Hydroxybutyrate	2.20	4.46 E-02
Fatty Acyl	Icosatetraenoic Acid	2.23	2.92 E-02
Fatty Acyl	Lauroylcarnitine	2.04	3.00 E-02
Glycerophospholipids and Glycerolipids	PA(33:0);DG(40:9)	2.82	8.30 E-03
Glycerolipids	MG(24:1)	3.82	2.82 E-02
Glycerolipids	TG(52:1)	0.12	4.70 E-02
Glycerophospholipids	CDP-DG(36:4)	2.70	4.16 E-02
Glycerophospholipids	CL(63:2)	2.68	4.60 E-03
Glycerophospholipids	CL(64:2)	2.17	1.55 E-02
Glycerophospholipids	Glycerophosphorylethanolamine	0.37	4.42 E-04

Glycerophospholipids	LysoPA(20:4); Hydroxy(phosphonatooxy)propylicosatrienoate	2.09	7.27 E-03
Glycerophospholipids	LysoPC(22:5);Enniatin B1	2.33	4.18 E-02
Glycerophospholipids	LysoPE(16:1);Cytochalasin Opho	2.82	3.21 E-02
Glycerophospholipids	LysoPE(22:5)	4.85	5.00 E-03
Glycerophospholipids	PC(32:1)	0.28	1.24 E-02
Glycerophospholipids	PC(44:10)	0.32	3.57 E-02
Glycerophospholipids	PE(34:2);PE-NMe(33:2);PE-NMe2(32:2)	0.12	1.73 E-02
Glycerophospholipids	PI(38:4)	0.36	2.79 E-02
Prenol Lipids	Congmunoside XIII	2.52	1.50 E-03
Prenol Lipids	Phytoene	2.79	1.49 E-02
Sphingolipids	Ganglioside GA2 (44:2)	2.24	4.46 E-03
Sphingolipids	SM(42:2)	2.11	4.01 E-02
Steroids and steroid derivatives	CE(DiMe(9,3)); Dimethyl(Methylheptanyl)Tetracycloheptadecenyl Butylmethylfuranyl)Nonanoate; Dimethyl(Methylheptanyl)Tetracycloheptadecenyl (Heptylfuranyl)Heptanoate; Dimethyl(Methylheptanyl)Tetracycloheptadecenyl (Hexylfuranyl)Octanoate; Dimethyl(Methylheptanyl)Tetracycloheptadecenyl Hydroxyoctadecatrienoate	0.39	9.91 E-03
Water Soluble	1-Methylguanine	6.58	2.11 E-04
Water Soluble	Acetyl Proline	0.11	2.74 E-05
Water Soluble	acetylaminoadipic acid	2.59	2.68 E-02
Water Soluble	Acetylcytidine	3.13	4.96 E-03
Water Soluble	Adenosine; Deoxyguanosine; Vidarabine; Zidovudine	0.25	2.03 E-02
Water Soluble	Aspartyl-phenylalanine	0.41	2.07 E-02
Water Soluble	Benzoic acid; Hydroxybenzaldehyde; Salicylaldehyde; Tropolone; (Furyl)acrylaldehyde	2.50	1.75 E-02
Water Soluble	Bilirubin;Bilirubin Ixa; Dihydrobiliverdin	3.23	1.73 E-04
Water Soluble	Biliverdin	4.65	2.31 E-04
Water Soluble	Chymopapain	3.88	5.96 E-06
Water Soluble	Convallamarogenin; Dideoxydehydroecdysone; Dihydroxyoxocalciol; Dihydroxyoxovitamin D3; Dihydroxypropanyl Tetracosahexaenoate; Dihydroxypropyl Tetracosahexaenoate; Dioxocholestanoic Acid; Epoxysecocholestatrienetriol; Hecogenin; Hydroxyoxocholestenoic Acid; Hydroxytetramethyl(Oxopentanyl)Oxapentacyclononadecanone; Kryptogenin; Nuatigenin;Ruscogenin; Secocholestatetraenetetrol; Trihydroxysecocholestatrienone	0.45	1.80 E-02
Water Soluble	Creatine	0.28	2.33 E-02
Water Soluble	Creatine Riboside	2.26	1.00 E-02
Water Soluble	Dihydrothymine	0.21	2.33 E-03
Water Soluble	Glycidyl Oleate	0.42	3.53 E-02
Water Soluble	HC Blue 2	2.01	2.69 E-02

5.3.1.3.2 At Rest Vs Passive Recovery

This time point comparison was conducted by using a pairwise t-test to look at the differences in relative concentration of compounds present at rest and after passive recovery. As can be seen in Table 5-6, the main lipid classes that was changing were glycerophospholipids, fatty acyls and glycerolipids. The concentration changes of compounds were a large mix of fold change increases and decreases in the glycerophospholipids. The decrease in fatty acyls suggests that the glycerophospholipids are no longer being used for energy, as there would be an increase in fatty acyls as the free fatty acids are reacted with carnitine before they cross the cell membrane, and here all the acylcarntines except acetyl carnitine (compounds ending in carnitine in Table 5-6) demonstrate fold change decreases in relative concentration from the starting rest (T1) to the passive recovery time point (T5).

Table 5-6: The fold changes for lipid compounds that were found to be statistically significant in t-tests for Starting Rest (T1) vs Passive Recovery (T5) time points from the SIT protocol. All fold changes had a statistical significance of P<0.05 and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. Fold change decreases are calculated as T5/T1. The glyceroplipids and glycerophospholipids are denoted by Class (number of carbons: number of double bonds). LysoPE= Lyso Glycerophosphoethanolamines, GPE/GPE-NMe/GPE-NMe2= Glycerophosphoethanolamines, GPC= Glycerophosphocholines, CL= Glycerophosphoglycerophosphoglycerols (Cardiolipins), TG= Triradylglycerolipids.

Compound Class	Compound Name	Fold Change	Raw p-value
Fatty Acyls	Acetylcarnitine	0.37	9.31E-03
Fatty Acyls	Decanoylcarnitine	3.64	3.16E-07
Fatty Acyls	Dodecenoylcarnitine	3.67	2.89E-04
Fatty Acyls	Hexanoylcarnitine	3.75	4.55E-04
Fatty Acyls	Hydroxydodecanoylcarnitine; [(Hydroxydodecanoyl)Oxy](Trimethylammonio) Butanoate	3.43	2.09E-04
Fatty Acyls	Hydroxyoxotetradecanal; Hydroxytetradecenoic Acid; (Octyloxiranyl)Butanoic Acid; Ketomyristic Acid; Menthylhydroxybutyrate	2.76	2.44E-02
Fatty Acyls	Hydroxytetradecenoylcarnitine	2.01	2.21E-02
Fatty Acyls	Isobutyrylcarnitine;C4-Carnitine;butyrylcarnitine; N-(tert-Butoxycarbonyl)leucine	2.36	3.63E-04
Fatty Acyls	Lauroyl Carnitine	2.76	4.51E-04
Fatty Acyls	LysoPE(22:5)	3.58	9.91E-03
Fatty Acyls	Tetradecadiencarnitine	3.03	2.60E-05
Glycerolipids	TG(62:2)	2.74	2.18E-02
Glycerolipids	TG(62:7)	3.05	3.93E-02
Glycerophospholipids	CL(63:2)	2.30	3.68E-03
Glycerophospholipids	GPC(40:4);GPE-NMe(42:4)	2.20	3.56E-02
Glycerophospholipids	GPE(34:2);GPE-NMe(33:2);GPE-NMe2(32:2)	0.15	2.79E-02
Prenol Lipids	Congmunoside XIII	2.32	7.24E-03

Prenol Lipids	Ganoderiol C	0.20	3.04E-02
Steroids And Steroid Derivatives	Glycoursodeoxycholic acid; N-Dihydroxyoxocholanyl]glycine; Glycochenodeoxycholic acid; Glycodeoxycholic acid	3.25	2.90E-02
Water Soluble	(Indolyl)acrylic acid	0.37	1.60E-02
Water Soluble	1-Methylguanine	6.60	1.29E-04
Water Soluble	Acetoxyhydroxylimonin	0.31	3.39E-02
Water Soluble	Acetyl Proline	0.12	7.13E-05
Water Soluble	Aspartylphenylalanine	0.39	1.21E-02
Water Soluble	Bilirubin	2.77	1.77E-02
Water Soluble	Biliverdin	2.77	2.92E-02
Water Soluble	Chymopapain	3.47	1.05E-04
Water Soluble	Creatine	2.63	3.58E-02
Water Soluble	Histidine	3.29	4.79E-02
Water Soluble	Hypoxanthine; Pentyl Hydrodisulfide; Pyrazolopyrimidinol; Allopurinol; Methyldithiahexane; Ethyl Propyl Disulfide	0.39	1.76E-09
Water Soluble	Inosine	0.49	4.01E-03
Water Soluble	Inosine	0.47	9.07E-05
Water Soluble	Lipoamide; 6-HITC	0.28	3.52E-04
Water Soluble	N-Arachidonoyl-L-Serine	2.83	3.97e-02
Water Soluble	Oxagrelate	0.12	2.84e-05
Water Soluble	Pantothenic Acid	0.38	7.58e-11
Water Soluble	Pentobarbital	0.30	1.17e-02
Water Soluble	Ribose-1-Arsenate	0.25	6.31e-07
Water Soluble	Threonylglutamine	4.42	1.26e-03
Water Soluble	Uric Acid	0.48	2.81e-02

5.3.2 What Are The Metabolic Changes When Comparing All Exercise Types At Each Time Point?

To answer this question the same time points were compared across exercise protocols. This was completed by using a one-way repeated measures (within subjects) ANOVA (OWRMA). After this, exercise protocols were individually compared against each other using post hoc paired t-tests to see which ones contained compounds that were significantly different in relative concentration. Final outputs from these statistical tests were filtered down to show annotated compounds only and with a clear time point where the significant change could be seen.

5.3.2.1 <u>At Rest</u>

There was only one metabolite that was significantly lower in the HIIT protocol (Table 5-7). The body was expected to be in a similar metabolic state at the start of each exercise session, so the absence of significantly altering compounds supports that theory.

Table 5-7: The compound that was found to be statistically significant in relative concentration when comparing all three exercise protocols across the at rest time point. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. Fold change increases are highlighted in green. Fold change decreases were calculated as SIT/MICT, MICT/HIIT and SIT/HIIT.

	OWRMA		MICT vs SIT		HIIT vs MICT		HIIT vs SIT	
Compound Name	F-Value	Adjusted P-Value	Fold Change	Raw P-Value	Fold Change	Raw P-Value	Fold Change	Raw P-Value
Methylhistidine	15.33	1.46E-02	NA	NA	3.21	1.02E-03	3.32	8.90E-04

5.3.2.2 End Of Exercise

At the end of exercise there were five different compounds that were statistically different in concentration between the three different protocols. Most of the compounds were present at the highest concentration in the SIT protocol (see Table 5-8). The highest number of differences were observed when comparing the MICT vs SIT protocols. Seeing as these protocols are the most different in effort required and energy uptake, it correlates to see the highest number of significant differences between these two exercise types.

Table 5-8: Fold change values for the relative concentration of the compounds that were found to be statistically significant when comparing all three exercise protocols across the end of exercise time point. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. Fold change decreases are highlighted in orange, and any fold change increases are highlighted in green. Fold change decreases are calculated as SIT/MICT, MICT/HIIT and SIT/HIIT. The glycerophospholipids are denoted by Class (number of carbons: number of double bonds). LysoGPE= Lyso Glycerophosphoethanolamines.

		OWRMA		MICT Vs SIT		HIIT Vs MICT		HIIT Vs SIT	
Metabolite Class	Compound Name	F-Value	Adjusted P-Value	Fold Change	Raw P-Value	Fold Change	Raw P-Value	Fold Change	Raw P-Value
Fatty Acyl	Decanoylcarnitine	16.42	1.80e-03	2.50	1.15e-04	NSS	NSS	NSS	NSS
Glycerophospholipids	LysoGPE(18:1)	9.19	1.83E-02	0.40	8.59E-03	NSS	NSS	0.40	7.20E-03
Glycerophospholipids	Lysope(18:2)	11.26	8.04E-03	5.68	3.25E-04	0.32	1.70E-03	NSS	NSS
Prenol Lipids	Ganoderiol C	8.44	2.65E-02	2.65	1.72E-03	NSS	NSS	NSS	NSS
Prenol Lipids	Soyasapogenol A; Ganoderiol A; Camelliagenin A	11.83	8.04E-03	5.27	1.35E-04	NSS	NSS	2.84	3.29E-03
Water Soluble	Pantothenic Acid	70.85	1.22e-09	0.47	3.31e-08	NSS	NSS	NSS	NSS
Water Soluble	Chymopapain	18.64	2.93e-03	2.61	2.83e-03	NSS	NSS	2.21	4.20e-02
Water Soluble	Hypoxanthine; Allopurinol; Pyrazolopyrimidinol	66.98	1.84e-09	2.84	2.77e-07	NSS	NSS	2.39	5.70e-07

5.3.2.3 Active Recovery

There were seven water soluble compounds which were statistically different across the three exercise protocols. There were also nine fatty acyls (five acyl carnitines and four fatty acids) with a 2-3 fold increase in concentration in the MICT protocol when in comparison to the SIT protocols (Table 5-9). This showed that with the use of lipolysis as a form of energy production rather than glycolysis, the levels of fatty acyls are much higher. Although there were no significant pathways in the SIT protocol over time, here it can be seen that hypoxanthine concentrations are 2.70-fold higher in SIT than in MICT. Hypoxanthine is a purine derivative which suggests that there may be changes to the purine metabolism pathway in SIT protocol as well as in MICT. There were not enough compounds detected to make the pathway significant under pathway enrichment analysis.

Table 5-9: Fold change values for the relative concentration of the compounds that were found to be statistically significant when comparing all three exercise protocols across the end of active recovery time point. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. Fold change decreases are highlighted in orange, and any fold change increases are highlighted in green. Fold change decreases are calculated as SIT/MICT, MICT/HIIT and SIT/HIIT.

		OWRMA		MictvsSIT		HIITvsMICT		HIITvsSIT	
metabolite class	Metabolite name	F-value	Adjusted P-value	Fold Change	raw p-value	Fold Change	raw p-value	Fold Change	raw p-value
Fatty Acyl	Linoleic acid;Malvalic acid; Methylheptadecadienoic acid; Octadecadienoic acid;Octadecynoic Acid; Rumenic acid; Stearolic acid; Tariric acid; Chaulmoogric Acid; dihomolinoleic acid; Ethyl hexadecadienoate; Isolinoleic acid; laballenic acid;Linalyl caprylate	23.27	1.92E-04	0.33	4.56E-06	NSS	NSS	NSS	NSS
Fatty Acyl	Icosatetraenoic acid;Icosadiynoic acid; Abietadienediol;acutilol A; Arachidonic acid; Dihydroabietic acid; Dimethylmethylene(methyloxopentyl)decahydronaphthalenecarbaldehyde; drostanolone;Hydroxykauranal; Hydroxymethylandrostanone; Kaurenediol;Ketoepimanool; Mesterolone; Methandriol; methoxy-androstenol; Methyl(tetramethyloctahydronaphthalenyl) pentenoic acid; Methyl(trimethylmethylenedecahydronaphthalenyl)pentenoic acid; Methylnonadecatetraenoic acid;oryzalexin D;oryzalexin E; Pimaradienediol;taxadiendiol	10.68	3.82E-02	0.36	3.66E-03	NSS	NSS	0.40	7.59E-03
Fatty Acyl	Ricinoleic Acid;oxostearic acid;Oxooctadecanoic acid; Methyloxoheptadecanoic acid; Methyl palmoxirate; Lactaric Acid; ketostearic acid; Isoricinoleic Acid; hydroxyoleic acid; Hydroxyoctadecenoic acid;epoxystearic acid; (Hexyltetrahydrofuranyl)octanoic acid	15.31	4.62E-03	0.26	1.79E-03	NSS	NSS	0.43	4.15E-02
Fatty Acyl	Oleic acid;Vaccenic acid; Petroselinic acid; Octadecenoic acid; Cycloheptylundecanoic acid; Elaidic Acid; Ethyl hexadecenoate; Isooleic acid; Methylheptadecenoic acid	37.54	2.64E-06	0.37	1.60E-05	NSS	NSS	NSS	NSS
Fatty Acyl	Tetradecadiencarnitine	22.29	7.94E-04	0.30	1.67E-04	NSS	NSS	NSS	NSS
Fatty Acyl	Hexadecadienoylcarnitine	11.35	3.33E-02	0.36	1.81E-03	NSS	NSS	NSS	NSS
Fatty Acyl	Tetradecenoylcarnitine; O-(tetradecenoyl)carnitine; [Tetradecenoyloxy]-(trimethylammonio)butanoate	10.56	3.33E-02	0.43	5.30E-03	NSS	NSS	NSS	NSS
Fatty Acyl	Hydroxytetradecenoylcarnitine	9.57	2.30E-02	0.24	7.18E-04	2.43	4.18E-03	NSS	NSS
Fatty Acyl	Decanoylcarnitine	24.03	6.65E-05	0.31	1.55E-05	NSS	NA	NSS	NSS
Water Soluble	O-(Hydroxycinnamoyl) astragalin; O-Coumaroylastragalin; O-Coumaroyltrifolin; O-Caffeoylcosmosiin; Gallocatechin-epicatechin; Piperitoside; Epigallocatechin-catechin; Buddlenoid A; Epicatechin-gallocatechin	21.09	1.07E-03	0.19	2.24E-04	NSS	NSS	0.33	1.84E-02
Water Soluble	Biliverdin	12.06	2.59E-02	0.16	1.18E-03	NSS	NSS	0.20	1.20E-04
Water Soluble	Eriodictyol (coumaroylglucoside)	28.41	2.58E-04	0.46	5.15E-06	NSS	NSS	NSS	NSS
Water Soluble	glycerophosphorylethanolamine; Glycerylphosphorylethanolamine	8.51	3.52E-02	2.35	2.68E-03	NSS	NSS	2.51	1.10E-03
Water Soluble	Epiguanine;Methylguanine	17.48	6.54E-04	0.20	1.52E-04	NSS	NSS	0.20	2.56E-04
Water Soluble	Pantothenic acid	89.72	6.19E-11	2.39	7.34E-10	NSS	NSS	NSS	NSS
Water Soluble	Hypoxanthine;Allopurinol; pyrazolopyrimidinol	138.57	3.96E-13	2.70	6.42E-10	NSS	NSS	2.05	5.41E-10

5.3.2.4 Passive Recovery

When comparing all three protocols at the passive recovery time point (Table 5-10), there were nine water soluble compounds that were significantly different across the three exercise protocols. There were also 11 different fatty acyls (six acyl carnitines, five fatty acids) that were changing across the exercise protocols. All the fatty acids demonstrated a 3-6-fold lower concentration in the SIT protocol compared to the MICT protocol and a 2-3-fold lower concentration in the SIT protocol compared to in the HIIT protocol. The acyl carnitines were at higher concentrations in the MICT protocol than in the HIIT protocol.

Table 5-10: Fold change values for the relative concentration of the compounds that were found to be statistically significant when comparing all three exercise protocols across the end of passive recovery time point. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. Fold change decreases are highlighted in orange, and any fold change increases are highlighted in green. Fold change decreases are calculated as SIT/MICT, MICT/HIIT and SIT/HIIT.

		overall		MictvsSIT		HIITvsMICT		HIITvsSIT	
Metabolite Class	name	F-value	Adjusted P-value	Fold Change	raw p-value	Fold Change	raw p-value	Fold Change	raw p-value
Fatty Acyl	Hydroxylauric acid;Hydroxydodecanoic acid	19.43	8.26E-04	0.16	8.01E-05	NSS	NSS	0.30	5.07E-03
Fatty Acyl	Oleic acid; Vaccenic acid; Petroselinic acid; Octadecenoic acid; Cycloheptylundecanoic acid; Elaidic Acid; Ethyl hexadecenoate; Isooleic acid; Methylheptadecenoic acid	14.37	3.78E-03	0.32	9.83E-04	NSS	NSS	0.43	1.08E-02
Fatty Acyl	Palmitic Acid;Butylethyldecanoic acid; Dimethyltetradecanoic acid;Propyltridecanoic acid; Trimethyltridecanoic acid;Dodecyl butyrate; Ethylmethyltridecanoic acid;Ethyltetradecanoic acid; Heptylnonanoic acid;Hexyl decanoate; Hexyldecanoic acid;Lauryl isobutyrate; Methylpentadecanoic acid	18.14	1.14E-03	0.22	3.29E-04	NSS	NSS	0.29	5.33E-03
Fatty Acyl	Ricinoleic Acid;oxostearic acid;Oxooctadecanoic acid; Methyloxoheptadecanoic acid; METHYL PALMOXIRATE;Lactaric Acid; ketostearic acid;Isoricinoleic Acid; hydroxyoleic acid;Hydroxyoctadecenoic acid; epoxystearic acid; (Hexyltetrahydrofuranyl)octanoic acid	18.22	1.14E-03	0.30	5.61E-03	NSS	NSS	0.36	1.02E-02
Fatty Acyl	Tetradecadiencarnitine	19.09	1.63E-03	NSS	NSS	0.36	6.73E-04	NSS	NSS
Fatty Acyl	hydroxydodecanoylcarnitine; [(Hydroxydodecanoyl)oxy](trimethylammonio)butanoate	11.09	2.58E-02	NSS	NSS	0.38	8.97E-03	NSS	NSS
Fatty Acyl	Hexadecadienoylcarnitine	10.31	3.20E-02	NSS	NSS	0.49	5.64E-03	NSS	NSS
Fatty Acyl	Tetradecenoylcarnitine;Tetradecenoylcarnitine; O-(tetradecenoyl)carnitine; [Tetradecenoyloxy](trimethylammonio)butanoate	10.90	2.60E-02	NSS	NSS	0.48	8.16E-03	NSS	NSS
Fatty Acyl	Decanoylcarnitine	45.42	1.18E-07	0.24	9.94E-07	NSS	NSS	0.45	6.35E-05
Fatty Acyl	Hexanoylcarnitine	13.41	1.18E-02	NSS	NSS	0.26	3.26E-03	0.32	5.67E-03
Fatty Acyl	Hydroxytetradecenoylcarnitine	10.99	7.81E-03	0.22	3.58E-04	NSS	NSS	NSS	NSS
Water Soluble	Menthyl acetoacetate	22.13	4.17E-04	0.18	1.63E-05	NSS	NSS	0.33	1.95E-03
Water Soluble	Histidine;Pyrazole-alanine	11.05	1.60E-02	0.18	3.35E-03	3.54	2.57E-02	NSS	NSS
Water Soluble	Biliverdin	18.06	2.21E-03	NSS	NSS	0.31	2.22E-02	0.16	6.44E-05
Water Soluble	Epiguanine;Methylguanine	8.06	3.43E-02	0.25	3.40E-03	NSS	NSS	0.34	1.46E-02
Water Soluble	Asparagine;Glycylglycine;Ureidopropionic acid; Methylnitrosourethane; METHYLHYDANTOIC ACID; Methylazoxymethanol Acetate	8.66	2.63E-02	NSS	NSS	NSS	NSS	NSS	NSS
Water Soluble	ACETYL PROLINE;Tiglylglycine;Methylcrotonyl Glycine;Paramethadione; Ethadione; Hydroxyethylclavam	12.92	3.37E-03	4.48	2.17E-04	0.47	3.75E-02	2.16	7.54E-03
Water Soluble	biotin	12.24	4.25E-03	NSS	NSS	4.44	6.41E-05	3.40	6.00E-03
Water Soluble	Pantothenic acid	107.23	6.08E-12	2.09	1.02E-09	NSS	NSS	NSS	NSS
Water Soluble	Hypoxanthine;Allopurinol; pyrazolopyrimidinol	189.26	5.49E-15	3.04	1.14E-10	NSS	NSS	2.36	1.40E-10

These results align with what was seen in the earlier results section where it was clear that the SIT protocol was causing different metabolic changes to occur in comparison to MICT and HIIT seemed to be a mixture of both MICT and SIT in terms of metabolic response.

5.3.3 Are There Any Differences In Metabolic Response To Exercise When It Comes To Exercise Type Vs Sex?

To answer this question, the same time points were compared across exercise protocols for just male or female participants to see if any sex differences were present. This was completed by using a one-way repeated measures (within subjects) ANOVA (OWRMA). After this, exercise protocols were individually compared against each other using paired t-tests to see which ones contained the compounds where significantly different in concentration.

5.3.3.1 At Rest

When at rest, neither males nor females had any compounds that were significantly different across the three exercise protocols. This shows that both sexes had metabolic states that were similar at the start of all three exercise protocols. It is important that this is the case for each sex because this time point was taken as the baseline time point those comparisons were made against.

5.3.3.2 End Of Exercise

At the end of exercise (time point 2 in MICT and time point 3 in HIIT and SIT), there were three compounds that significantly differed in relative concentration between exercises in female participants (Table 5-11) whilst only two compounds were significantly different in relative concentration in the male participants (Table 5-12). Two of these compounds overlapped between the males and females and both compounds were present at the highest relative concentration of compounds in the SIT protocol followed by HIIT and then MICT.

Table 5-11: The compounds that were statistically significantly changing in relative concentration between exercises at the end of exercise time point (time point 2 in MICT and time point 3 in HIIT and SIT), within the female participants in the study. Fold changes were calculated as MICT/SIT, MICT/HIIT, and SIT/HIIT. Fold change increases are highlighted in green and fold change decreases are highlighted in orange. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. The p-value for the repeated one-way ANOVA was adjusted by applying FDR correction.

		Repeated One-way ANOVA		MICT vs SIT		HIIT vs MICT		HIIT vs SIT	
Compound Class	Compound Name	F-value	Adjusted P-Value	Fold Change	Raw P-Value	Fold Change	Raw P-Value	Fold Change	Raw P-Value
Water Soluble	Ribose arsenate	31.38	1.30E-02	NSS	NSS	NSS	NSS	NSS	NSS
Water Soluble	Hypoxanthine; Allopurinol; Pyrazolopyrimidinol	53.45	7.87E-05	0.40	2.30E-05	NSS	NSS	0.50	2.33E-06
Water Soluble	Pantothenic acid	56.41	7.87E-05	0.46	3.13E-05	NSS	NSS	NSS	NSS

Table 5-12: The compounds that were statistically significantly changing in relative concentration between exercises at the end of exercise time point (time point 2 in MICT and time point 3 in HIIT and SIT), within the male participants in the study. Fold changes were calculated as MICT/SIT, MICT/HIIT, and SIT/HIIT. Fold change increases are highlighted in green and fold change decreases are highlighted in orange. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. The p-value for the repeated one-way ANOVA was adjusted by applying FDR correction.

		Repeated One-way ANOVA		MICT vs SIT		HIIT vs MICT		HIIT vs SIT	
Compound Class	Compound Name	F-Value	Adjusted P-Value	Fold Change	Raw P-Value	Fold Change	Raw P-Value	Fold Change	Raw P-Value
Water Soluble	Pantothenic acid	23.80	8.78E-03	0.47	6.59E-04	NSS	NSS	NSS	NSS
Water Soluble	Hypoxanthine; Allopurinol; Pyrazolopyrimidinol	70.19	2.83E-05	0.44	3.33E-05	NSS	NSS	NSS	NSS

5.3.3.3 Active Recovery

At the end of active recovery time point females had four different compounds (Table 5-13) which were significantly different in intensity between exercise protocols, whilst males had three (Table 5-14). Two of these compounds overlapped between the males and females and showed the highest relative concentration in the SIT protocol followed by HIIT and then MICT.

Table 5-13: The compounds that were statistically significantly changing in relative concentration between exercises at the active recovery time point within the female participants in the study. Fold changes were calculated as MICT/SIT, MICT/HIIT, and SIT/HIIT. Fold change increases are highlighted in green and fold change decreases are highlighted in orange. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. The p-value for the repeated one-way ANOVA was adjusted by applying FDR correction.

		Repeated One-way ANOVA		MICT vs SIT		HIIT vs MICT		HIIT vs SIT	
Compound Class	Compound Name	F-value	Adjusted P-value	Fold Change	raw p-value	Fold Change	raw p-value	Fold Change	raw p-value
water soluble	Pantothenic acid	23.80	8.78E-03	0.47	6.59E-04	NSS	NSS	NSS	NSS
water soluble	Hypoxanthine;Allopurinol; pyrazolopyrimidin-7-ol	70.19	2.83E-05	0.44	3.33E-05	NSS	NSS	NSS	NSS
Fatty Acyl	Linoleic acid; Malvalic acid; Rumenic acid;Stearolic acid; Tariric acid; Chaulmoogric Acid; dihomolinoleic acid; Ethyl hexadecadienoate; laballenic acid; Linalyl caprylate; Methylheptadecadienoic acid; Octadecadienoic acid; Octadecynoic acid	22.48	7.50E-03	3.06	3.76E-03	NSS	NSS	2.96	6.09E-03
Fatty Acyl	Ricinoleic Acid; ricinelaidic acid; oxostearic acid; Oxooctadecanoic acid; Methyloxoheptadecanoic acid; Methyloxoheptadecanoic acid; Methyl Palmoxirate; Lactaric Acid; ketostearic acid; Isoricinoleic Acid; hydroxyoleic acid; hydroxyoctadecenoic acid; epoxystearic acid; (Hexyltetrahydrofuranyl)octanoic acid	16.37	3.33E-02	3.06	2.46E-02	NSS	NSS	2.81	4.80E-02
Fatty Acyl	Oleic acid; Vaccenic acid; Octadecenoic acid; Petroselinic acid; Elaidic Acid; Ethyl hexadecenoate; Isooleic acid; Methylheptadecenoic acid; Cycloheptylundecanoic acid	34.43	1.90E-03	2.85	5.64E-03	NSS	NSS	2.61	9.23E-03

Table 5-14: The compounds that were statistically significantly changing in relative concentration between exercises at the active recovery time point within the male participants in the study. Fold changes were calculated as MICT/SIT, MICT/HIIT, and SIT/HIIT. Fold change increases are highlighted in green and fold change decreases are highlighted in orange. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. The p-value for the repeated one-way ANOVA was adjusted by applying FDR correction.

		Repeated One-way ANOVA		MICT vs SIT		HIIT vs MICT		HIIT vs SIT	
Compound Class	Compound Name	F-Value	Adjusted P-Value	Fold Change	Raw P-Value	Fold Change	Raw P-Value	Fold Change	Raw P-Value
Water Soluble	Epiguanine; Methylguanine	15.43	2.70 E-02	5.75	6.10 E-03	NSS	NSS	6.05	5.88 E-04
Water Soluble	Pantothenic acid	32.24	1.62 E-03	0.42	1.13 E-04	NSS	NSS	NSS	NSS
Water Soluble	Hypoxanthine; Allopurinol; pyrazolopyrimidinol	51.63	1.94 E-04	0.37	6.49 E-05	NSS	NSS	0.48	9.98 E-05

5.3.3.4 Passive Recovery

At the end of passive recovery, females had six fatty acyls (Table 5-15) that differed in relative concentration between exercise protocols whilst men had five that changed significantly (Table 5-16). These were four water soluble and one fatty acyl. There was still the one overlapping water-soluble compound between the two sexes (hypoxanthine) that was similar in intensities between the exercise protocols. This suggests that there may be some differences in the way that males and females undergo passive recovery after exercise.

Table 5-15: The compounds that were statistically significantly changing in relative concentration between exercises at the passive recovery time point within the female participants in the study. Fold changes were calculated as MICT/SIT, MICT/HIIT, and SIT/HIIT. Fold change increases are highlighted in green and fold change decreases are highlighted in orange. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. The p-value for the repeated one-way ANOVA was adjusted by applying FDR correction.

		Repeated One-way ANOVA		MICT vs SIT		HIIT vs MICT		HIIT vs SIT	
Compound Class	name	F-value	Adjusted P-value	Fold Change	raw p-value	Fold Change	raw p-value	Fold Change	raw p-value
Fatty Acyl	Decanoylcarnitine	31.82	1.17E-03	4.43	2.76E-04	NSS	NSS	2.37	9.24E-04
Fatty Acyl	Linoleic acid;Malvalic acid; Rumenic acid;Stearolic acid; Tariric acid;Chaulmoogric Acid;dihomolinoleic acid; Ethyl hexadecadienoate; laballenic acid; Linalyl caprylate; Methylheptadecadienoic acid; Octadecadienoic acid; Octadecynoic acid	33.40	3.03E-03	4.08	2.30E-03	NSS	NSS	4.80	2.44E-03
Fatty Acyl	Oleic acid;Vaccenic acid; Octadecenoic acid; Petroselinic acid; Elaidic Acid; Ethyl hexadecenoate; Isooleic acid; Methylheptadecenoic acid; Cycloheptylundecanoic acid	16.11	3.33E-02	3.81	3.11E-03	NSS	NSS	3.05	4.43E-02
Fatty Acyl	Palmitic Acid; Butylethyldecanoic acid; Dimethyltetradecanoic acid; Dodecyl butyrate; Ethylmethyltridecanoic acid; Ethyltetradecanoic acid; Heptylnonanoic acid; Hexyl decanoate; Hexyldecanoic acid;Lauryl isobutyrate; Methylpentadecanoic acid; Propyltridecanoic acid; Trimethyltridecanoic acid	18.15	2.55E-02	3.52	1.99E-02	NSS	NSS	3.84	1.59E-02
Fatty Acyl	Ricinoleic Acid;ricinelaidic acid; oxostearic acid; Oxooctadecanoic acid; Methyloxoheptadecanoic acid; Methyloxoheptadecanoic acid; METHYL PALMOXIRATE; Lactaric Acid; ketostearic acid; Isoricinoleic Acid; hydroxyoleic acid; hydroxyoctadecenoic acid; epoxystearic acid; (Hexyltetrahydrofuranyl)octanoic acid	17.76	2.55E-02	NSS	NSS	NSS	NSS	2.58	4.75E-02
Fatty Acyl	Octaneperoxoic acid; Hydroxycaprylic acid; Hydroxyoctanoic acid; Hydroxyvalproic acid; (Isobutyldioxolanyl)methanol; Ethyl hydroxyhexanoate; Isoamyl lactate; Rengyol	12.18	4.70E-02	0.44	8.51E-03	NSS	NSS	NSS	NSS
Water Soluble	Hypoxanthine;Allopurinol; pyrazolopyrimidinol	315.10	9.11E-09	5.36	9.65E-07	NSS	NSS	4.49	1.30E-06

Table 5-16: The compounds that were statistically significantly changing in relative concentration between exercises at the passive recovery time point within the male participants in the study. Fold changes were calculated as MICT/SIT, MICT/HIIT, and SIT/HIIT. Fold change increases are highlighted in green and fold change decreases are highlighted in orange. All comparisons were which were significant had a p-value < 0.05 and any that were not significant were labelled as NSS. The p-value for the repeated one-way ANOVA was adjusted by applying FDR correction.

		OWRMA		MICT vs SIT		HIIT vs MICT		HIIT vs SIT	
Compound Class	Compound Name	F-value	Adjusted P-value	Fold Change	raw p-value	Fold Change	raw p-value	Fold Change	raw p-value
Water Soluble and Fatty Acyls	Dihydrothymine; pyroglutamine; Oxopyrrolidinecarboxamide; Methylpiperazinedione; Nitrilonorvaline; Aminocyanobutanoic acid	20.56	0.038621	9.68	8.49E-05	NSS	NSS	6.76	0.006783
Water Soluble	biotin	13.34	0.034054	NSS	NSS	0.21	0.002619	0.18	0.000506
Fatty Acyls	Decanoylcarnitine	15.44	0.020165	3.94	0.002558	NSS	NSS	2.09	0.023605
Water Soluble	Pantothenic acid	73.46	7.05E-06	0.45	1.14E-05	NSS	NSS	NSS	NSS
Water Soluble	Hypoxanthine;Allopurinol; pyrazolopyrimidinol	73.69	7.05E-06	0.31	3.58E-05	NSS	NSS	0.43	0.000057

From these results it appears that there is very little metabolic difference between the two sexes. This is contrasting to the results found by S.J. Weaver, et.al.[206] who found a significant difference in response between males and females during the SIT and MICT protocols. This is not the only study that this is contrasting to, as there are multiple studies focussed around the differences in metabolism between sexes both during exercise[296-298], and in general[299-301].

5.4 CONCLUSIONS

5.4.1 What Are The Metabolic Changes Over Time Related To Exercise Type?

MICT protocol metabolic changes saw changes in relative concentration in fatty acyls and glycerophospholipids that were synonymous with changes seen in lipolysis when glycerophospholipids are broken down into fatty acids which are than reacted with carnitine when transferred across the cell wall to the mitochondria to be used as energy. This protocol also detected changes in purine metabolism, which showed that ATP concentrations were changing rapidly.

HIIT protocol metabolic changes saw a decrease in fatty acyls and glycerophospholipids over the exercise time points, although by passive recovery some gylcerophosolipids had increased in relative concentration. This suggested that glycerophospholipids were originally being broken down and used for energy before the body switched to using carbohydrates as a quicker source of energy.

SIT protocol metabolic changes saw a decrease in fatty acyls, glycerophospholipids and glycolipids across all time points. This suggests that lipolysis was originally occurring as an energy source, but that

the body used a lot of carbohydrates as energy source for the fact that the exercise protocol required more effort.

Overall, the lower intensity exercises saw changes in the metabolism where lipolysis was happening in order to release more ATP for the body to use as energy. Whilst the higher intensity exercise protocols didn't show these changes, it suggests that alternative metabolic pathways (such as glycolysis) were being activated to reach a higher level of ATP quicker.

5.4.2 What Are The Metabolic Changes When Comparing All Exercises At Each Time Point?

When at rest there was minimal differences between the three exercise protocols, which was to be expected. At end of exercise, compounds with statistically significant differences had highest concentration in the SIT exercise protocol whilst the HIIT and MICT were similar in relative concentration of overlapping compounds. At active recovery, the overlapping compounds with significant differences between the three different protocols had highest concentration in the MICT protocol, followed by the HIIT protocol and then SIT was most commonly lowest in concentration. At passive recovery, the results were similar to that of active recovery with the MICT protocol, followed by the HIIT protocol and then SIT was most commonly lowest in concentration of overlapping compounds with statistically significant differences.

Overall, it appears that MICT and SIT protocols are the most different in terms of energy requirements and that HIIT is somewhere in between the two. The lipolysis changes in MICT where very high because the increases were still present at the passive recovery and were much higher in concentration than in the HIIT and SIT protocols were there had been a decrease in lipids.

5.4.3 Are There Any Differences When It Comes To Exercise Type Vs Sex?

When at rest, neither males nor females had any compounds that were statistically significantly different across the three exercise protocols. At end of exercising, there were two overlapping compounds between the males and females. Both compounds behaved in a similar manner in both the sexes, being in the highest concentration in the SIT protocol. At the end of active recovery both the males and females had three different compounds were significantly different in relative concentration between exercise protocols. The two overlapping compounds behaved in a similar manner in both the sexes, being in the highest concentration in the SIT protocol. At the end of passive recovery, females had more fatty acyls that differed in intensity between exercise protocols than males. There was still a number of overlapping compounds between the two sexes that were similar in intensities between the exercise protocols.

Overall, there were not any many metabolic differences between the male and female participants. It should be taken into consideration that there were only eight participants of each sex, and thus with more participants there would be more differences detected as significant.

5.5 FUTURE WORK

For future work leading on from this research I would suggest that the study be repeated with a larger number of participants of each sex. This is because the results from the collaborating research found great differences between sexes[206] and I think this should be further investigated from a metabolomics perspective.

I would also suggest that HIIT and SIT exercise protocols are investigated further using HILIC or other polar LC methods to gain a better understanding of what changes are occurring over time in those exercise protocols where the change in energy requirement means there are metabolic changes occurring in pathways which are predominantly water-soluble compounds.

6 HEAT STRESS AND ACCLIMATION EFFECTS ON HUMAN METABOLISM

This work was completed in collaboration with Dr Ella Walker, Dr Caroline Rowland, Dr Nicola Armstrong, and Dr Sarah Goodchild at the Defence science and technology laboratory (Dstl) alongside Charlotte Stevens, Prof. Mike Tipton and Dr Jo Corbett from the School of Sport, Health and Exercise Science at the University of Portsmouth.

It should be taken into consideration that this study has been altered due to the confinement of COVID-19 regulations in order for the study to continue taking place. Instead of both the traditional and condensed heat acclimation trials occurring, only the condensed took place with a smaller number of participants than originally designed with.

6.1 INTRODUCTION

When working in a hot climate, the human body is at risk of becoming subject to a reduced work capacity[302] along with many different heat related illnesses[303,304] including heat syncope, heat edema, heat tetany, heat cramps, and heat exhaustion[305]. One way to help overcome the risk of succumbing to heat related illness is to partake in heat acclimation. Heat acclimation is the collection of physiological changes that occur when the body is repeatedly subjected to a rise in core/skin temperature from exposure to exercise, hot climates or both[306-308]. The beneficial physiological effects that occur include "improved sweating, improved skin blood flow, lowered body temperatures, reduced cardiovascular strain, improved fluid balance, altered metabolism, and enhanced cellular protection"[306]. Heat acclimation has obvious benefits for elite level athletes[306,309-311] that compete in a vast array of climates during international events but also has great benefits for personnel conducting work in adverse climate conditions. Military personnel are a good example of people who are often required to operate in adverse conditions and may be exposed to prolonged periods of high ambient heat and humidity[312].

Traditionally people would undertake heat acclimation protocols to adapt their bodies to the new adverse conditions around them, with the typical daily heat exposure of 1 to 2 hours over ~7 to 10 consecutive days[313-316]. This method of heat acclimation can be time costly, logistically difficult to implement and as well as impractical[310] in situations where individuals are only present in the climate for a short period of time (athletes) or where they are rapidly deployed to the adverse climate (military). Current government protocol recommends military personnel to undergo an 8-day fixed-exercise intensity protocol (JSP 375 Chapter 41 Annex E). Therefore, heat acclimation approaches with a

reduction in the number of days required to achieve acclimation may have utility for both elite athletes and military personnel.

Previous studies have indicated that there is a possible dose-response relationship between heat exposure and the magnitude of the adaptive response to heat[317,318], and even within the traditional 7-10 day heat acclimation programme ~75% of the adaptive response to heat is obtained between 4-6 daily exposures.[315,316] Another relationship that has been hypothesised is the relationship between heat acclimation and the cross tolerance to hypoxia[319]. Previous research in this area suggests that heat acclimated species are more tolerant of hypoxic conditions[320-323].

Currently, studies on metabolomics in conjunction with both heat and/or cold acclimation[324] have been mainly conducted on plants, with only one study conducted on *Drosophila melanogaster*.[325] There have not been any published metabolomics studies for heat acclimation in humans or mammals, which is of great surprise due to the amount of changes that the physiological phenotype undergoes throughout the acclimation process. There has been a wide range of research into heat stress metabolism with most of it exercise focused[326-329]. Often, the literature does not centre around metabolomics and so applies a more targeted approach featuring certain metabolites or pathways (e.g. glucose[330], ATP[331], urate metabolism[332] and cytokine production[333]). Understanding how the human metabolome changes with heat stress and/or heat acclimation provides an opportunity to better comprehend how the body is dealing with the levels of stress applied, and which pathways are most affected by the acclimation process. This may also allow metabolic biomarkers indicative of heat acclimatisation or heat stress to be identified and detected before changes to the physiological phenotype are detected.

6.1.1 Research Objectives

This study will look at the dynamic relationship between heat acclimation and human metabolism, investigating if there are any differences in how the plasma, and urine metabolome reacts to heat stress and hypoxic tolerance before and after undergoing a condensed set of heat acclimation sessions. Research questions include:

1. **Are There Any Metabolic Differences In Samples Collected Before And After Heat Acclimation As Demonstrated By Two Independent Heat Stress Tests (HSTs)?**

   This question was answered by comparing the changes occurring within each of the HSTs which were conducted before and after the heat acclimation sessions using a repeated one-way ANOVA and filtering the results to only include annotated metabolites. This was repeated separately for the plasma and urine samples.

2. **Are There Any Metabolic Differences In Samples Collected Before And After Heat Acclimation As Demonstrated By Two Independent Hypoxic Tolerance Tests (HTTs)?**

   This question was answered by comparing the changes occurring in the plasma samples collected at time points across the two HTTs before and after the heat acclimation sessions. This was completed by applying a repeated one-way ANOVA and filtering the results to only include annotated metabolites.

3. **Are There Any Metabolic Changes During One Heat Acclimation Session As Well Across An Entire Day Of The Heat Acclimation Sessions?**

   This question was answered in two different ways:

   By comparing the changes occurring in the samples across all the Post Condensed Heat Acclimation (Cond HA) sessions to investigate what metabolic changes are occurring across the whole Cond HA window.

   By comparing the first Cond HA sessions from each day to see what metabolic changes occur within one Cond HA session, and if there are any differences between those two sessions due to the following Cond HA sessions on day 0.

6.2 STUDY DESIGN

This study was designed and conducted by Charlotte Stevens and Dr Jo Corbett from the School of Sport, Health and Exercise at the University of Portsmouth in conjunction with Dstl. The original study design was planned to investigate the traditionally applied heat acclimation timeline (8 days) and compare it to a newly developed condensed heat acclimation timeline (2 days) that would allow for the same heat acclimation to be achieved in a much shorter time period. Unfortunately, due to COVID-19, adaptions to the study had to be made. This meant that only the data for the condensed heat acclimation study were collected, and the number of participants included was reduced from 16 to 6, with the addition of the two pilot study participants to increase statistical power. The timeline for the condensed heat acclimation can be seen below (Table 6-1), whilst the timeline for the traditionally applied heat acclimation timeline can be found in the appendix A-1.

Table 6-1: Timeline for the condensed heat acclimation trial. Key: GXT = Graded Exercise Test; HypTT = Hypoxic Tolerance Test; HST = Heat Stress Test; Cond HA = Heat Acclimation.

Day	Test Undertaken
-6	Graded Exercise Test pre and Anthropometry
-5	Hypoxic Tolerance Test Pre
-4	OFF
-3 or -2	Heat Stress Test Pre
-1	OFF
0	Condensed Heat Acclimation
1	Condensed Heat Acclimation
2	OFF
3 or 4	Heat Stress Test Post
5	OFF
6	Graded Exercise Test Post
7	Hypoxic Tolerance Test Post
8 - 21	OFF
22 - 25	Heat Stress Test Decay

6.2.1.1 Hypoxic Tolerance Test

The hypoxic tolerance test (HTT) (summarised in figure 6-1) applied in this study is one taken and from Lunt et al.[334]. Participants entered the temperate (~ 21°C, 50 % R.H.) environmental chamber where they were immediately connected to an external air supply, enabling them to breathe normoxic[335] air (F_iO_2 =0.21) from outside of the laboratory. Once connected to the instrumentation, participants rested for 10 minutes, breathing normoxic air. After this period, a baseline venous blood sample was drawn and collected in a heparin coated tube for plasma analysis. Thereafter, participants were exposed to the hypoxic[336] air mixture (F_iO_2 =0.15) within the chamber. The hypoxic air mixture produced a partial pressure of oxygen that is similar to that which occurs at a natural (hypobaric hypoxia) altitude of ~2,500m.

During hypoxic exposure, participants rested on the cycle ergometer for 10 minutes breathing the hypoxic air mixture before undertaking 10 minutes exercise at 100 W (fixed cadence of 60 revs.min^{-1}). Upon cessation of exercise a second blood sample was obtained. Blood samples were spun in a centrifuge and acellular supernatant was collected and stored at -80°C. Any cellular material was then disposed of.

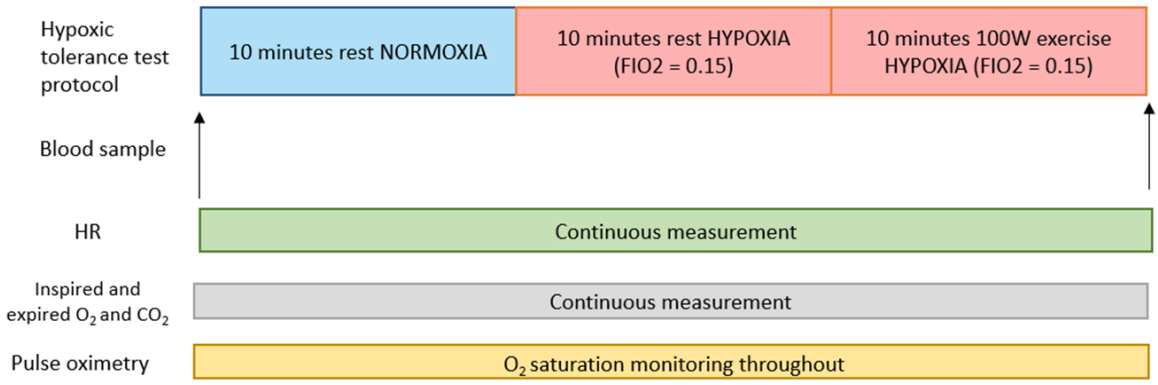


Figure 6-1: Summary of the hypoxic tolerance test protocol performed in temperate conditions (21°C, 50%). Participants were exposed to a hypoxic mixture of 15% O2 (FIO2 = 0.15). Venous blood samples were taken before and after hypoxic exposure. Heart rate, inspired and expired oxygen and carbon dioxide and oxygen saturation were monitored and recorded throughout the protocol.

6.2.1.2 Heat Stress Test

The heat stress test (HST) (summarised in Figure 6-2) assessed the thermo-physiological responses to a standard exercise-thermal challenge (40°C, 50 % R.H.). Urine samples were collected before the participants entered the laboratory. Upon entry, participants rested for 10 minutes in the temperate laboratory environment. At the end of the rest period a baseline venous blood sample was obtained in a heparin lined tube for plasma analysis. Participants then entered the environmental chamber where they rested for fifteen minutes. After this, the HST commenced. Participants completed 60 minutes of cycle ergometer exercise (fixed cadence of 70 revs.min^{-1}) at a work rate equivalent to 35% of the peak power output achieved during an initial graded exercise test, as has been used in previous heat acclimation work[316]. Upon cessation of exercise, participants left the environmental chamber where the sweat patch was then removed and nude body mass recorded. Thereafter, a second venous blood sample was obtained once the participant had sat for a 10-minute period. The first urine produced following test termination was collected for subsequent analysis. Samples were spun in a centrifuge and the acellular supernatant collected before being stored at -80°C until further analysis.

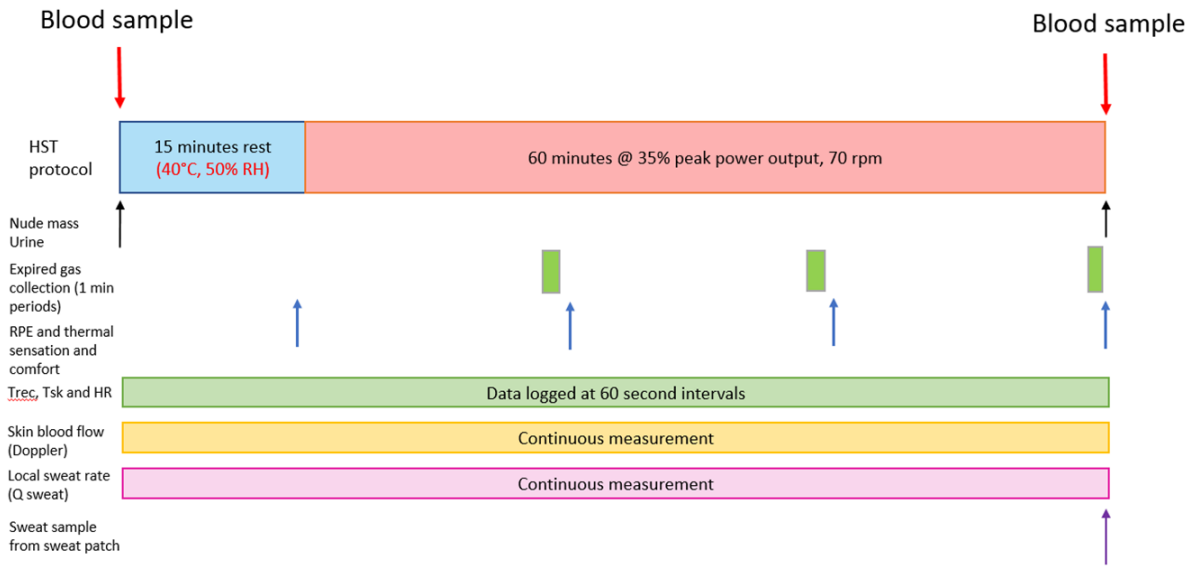


Figure 6-2: Summary of the heat stress test (HST) protocol undertaken in hot conditions (40 degrees, 50% RH). Nude mass and a urine sample were recorded before and after entering the chamber. Data from the rectal thermistor, skin thermistors and heart rate were logged at 60 second intervals throughout the HST. Skin blood flow using a Doppler probe and local sweat rate from the right scapula using a Q sweat were measured continuously. Oxygen uptake and carbon dioxide production were measured at 20, 40and 60 minutes of each exercise work rate. At the same time points, perceptual ratings of RPE, thermal sensation and thermal comfort were taken. A custom patch containing a cotton wool swab to absorb sweat were attached during the rest period and removed following the exercise test for sampling. Blood samples were taken before the heat stress teat and after exercise is complete as indicated by the red arrows.

6.2.1.3 Condensed Heat Acclimation Sessions

The heat acclimation sessions employed an isothermal strain approach, adapted from a method initially developed by Fox et al.[317]. The isothermal strain model targets rapid achievement and subsequent maintenance of a specified deep body temperature over the course of each heat acclimation session, in this instance a rectal temperature of ~38.5±0.2 °C. Each individual heat acclimation session lasted 75 minutes, with the condensed heat acclimation (Cond HA) programme applying four sessions per day over two consecutive days.

Participants dressed in their clothing along with water vapour impermeable clothing before entering the hot (45 °C, 20-30 % R.H.) environmental chamber. Initial elevation in deep body temperature was achieved by a single ten-minute bout of low intensity stepping exercise (18 steps per minute on to a 25 cm bench). Water vapour impermeable clothing inhibited the evaporation of sweat, thereby restricting the loss of the metabolic heat to the environment and facilitating a rapid rise in deep body temperature. Thereafter, maintaining a target deep body temperature of 38.5 °C was achieved with the removal of

clothing if required. Upon completion of each heat acclimation session the participants removed their vapour impermeable clothing, left the environmental chamber, towel dried, and rested in the laboratory (~ 20 °C, 50 % R.H.) for a 60-minute period. At the end of the 60-minute rest period they re-entered the environmental chamber and the next session commenced.

Venous blood samples were obtained before and after the first and fifth heat acclimation sessions along with after each of the second, fourth, sixth and eighth sessions. Sweat samples were obtained from sweat patches worn during the first, fourth, fifth and eighth heat acclimation sessions. Urine samples were collected before the first and fifth heat acclimation sessions and after the fourth and eighth heat acclimation session. The condensed heat acclimation protocol is summarised in figure 6-3.

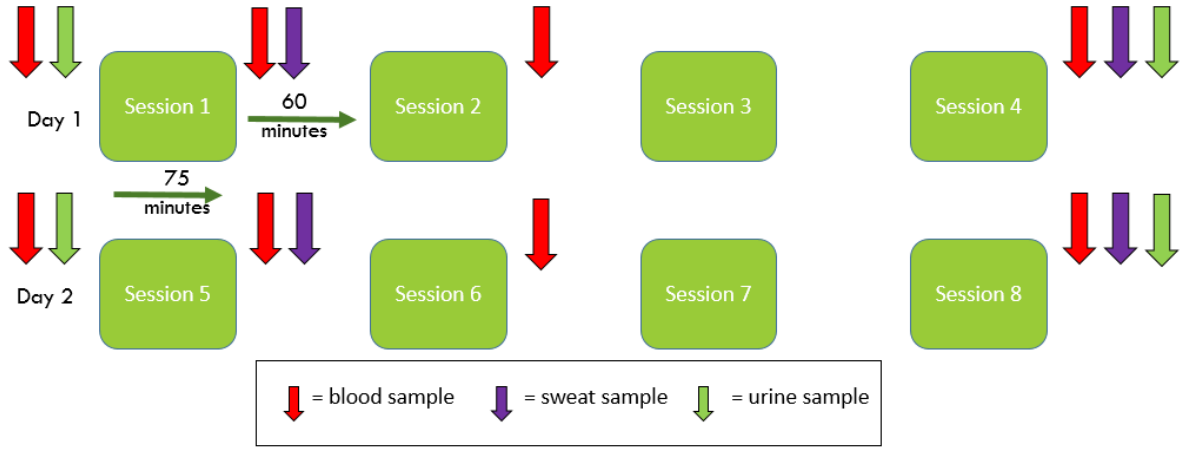


Figure 6-3: Summary of the condensed heat acclimation (Cond HA) protocol undertaken by participants. Two days of heat acclimation (HA) sessions were conducted with each HA session was 75 minutes followed by a 60 minute rest period. Blood samples were collected before and after the first HA session of each day, along with after the second, fourth, sixth and eighth HA sessions (red arrow). Urine samples were collected at the start and end of each day (green arrow), whilst sweat samples (not analysed in this thesis) were collected during the first and last HA session of each day (purple arrow).

6.3 MATERIALS AND METHODS

Methods for this research chapter can be found in the Materials And Methods chapter. The ethics statement is in section 2.2.3, sample collection details are under section 2.3.3, sample extraction details are under section 2.4, UHPLC-MS methods are in section 2.9.2, and data processing and metabolite annotation can be found in section 2.9.3.

6.4 RESULTS AND DISCUSSION

6.4.1 Are There Any Metabolic Differences In Samples Collected Before And After Heat Acclimation As Demonstrated By Two Independent Heat Stress Tests?

This question was answered by comparing the changes occurring within each of the HSTs conducted before and after the heat acclimation sessions using a repeated one-way ANOVA and filtering the results to only include annotated metabolites. This was repeated separately for the plasma and urine samples. The number of compounds which were statistically significant in plasma decreased from 43 to 14, whilst the number of statistically significant compounds in urine decreased from 261 to 115.

6.4.1.1 Plasma Samples

After the repeated one-way ANOVA was conducted, and initial filtering completed, the annotated compounds left were filtered to focus on the subset that had changes pre/post HST1, HST1 Post/ HST2 post and/or pre/post HST2. After filtering there was only 4 compounds that were statistically significant. None of which were water soluble. Changes in lipid classes were observed, with the three fatty acids being significant and one glycerophospholipid. As can be seen in Table 6-2, there was a decrease in fatty acyls in both HSTs. When focussing on the two post HST samples, the relative concentration of fatty acyls was a lot less in the second HST, and the level of glycerophosolipid was higher. This would suggest that there was a change in the amount of energy required by the body. This could be due to the heat acclimatisation that the participants completed between HSTs. Although there were changes observed, there were not enough compounds to define that any clear metabolic changes were occurring.

Table 6-2: One Way Repeated Measures ANOVA (OWRMA) results featuring annotated compounds with statistically significant fold changes across time points pre/post heat stress test 1 (HST1PostvsHST1Pre), heat stress test 1 Post/ heat stress test 2 post (HST1PostvsHST2Post) and pre/post heat stress test 2 HST2PostvsHST2Pre. Positive fold changes are highlighted green, negative fold changes are highlighted orange and any non-statistically significant changes are labelled as NSS. LysoPE = Lysophosphatidylethanolamine. FC= Fold Change. OWRMA= one way repeated measures ANOVA.

		OWRMA		HST1 Post vs HST1 Pre		HST2 Post Vs HST2 Pre		HST1Post Vs HST2Post	
Compound Class	Compound Name	F value	Raw P-value	FC	Raw P-value	FC	Raw P-value	FC	Raw P-value
Fatty Acyls	Dihydroxyoctadecanoic acid	5.59	8.89E-03	NSS	NSS	NSS	NSS	0.24	2.71E-02
Fatty Acyls	Octadecenoic acid	12.89	1.99E-04	0.34	1.53E-02	0.40	2.58E-02	0.48	3.19E-02
Fatty Acyls	Hexadecanoic acid	7.90	2.15E-03	0.34	6.16E-03	NSS	NSS	NSS	NSS
Gylcerophospholipids	LysoPE(16:0)	9.36	1.82E-03	NSS	NSS	NSS	NSS	3.76	2.37E-02

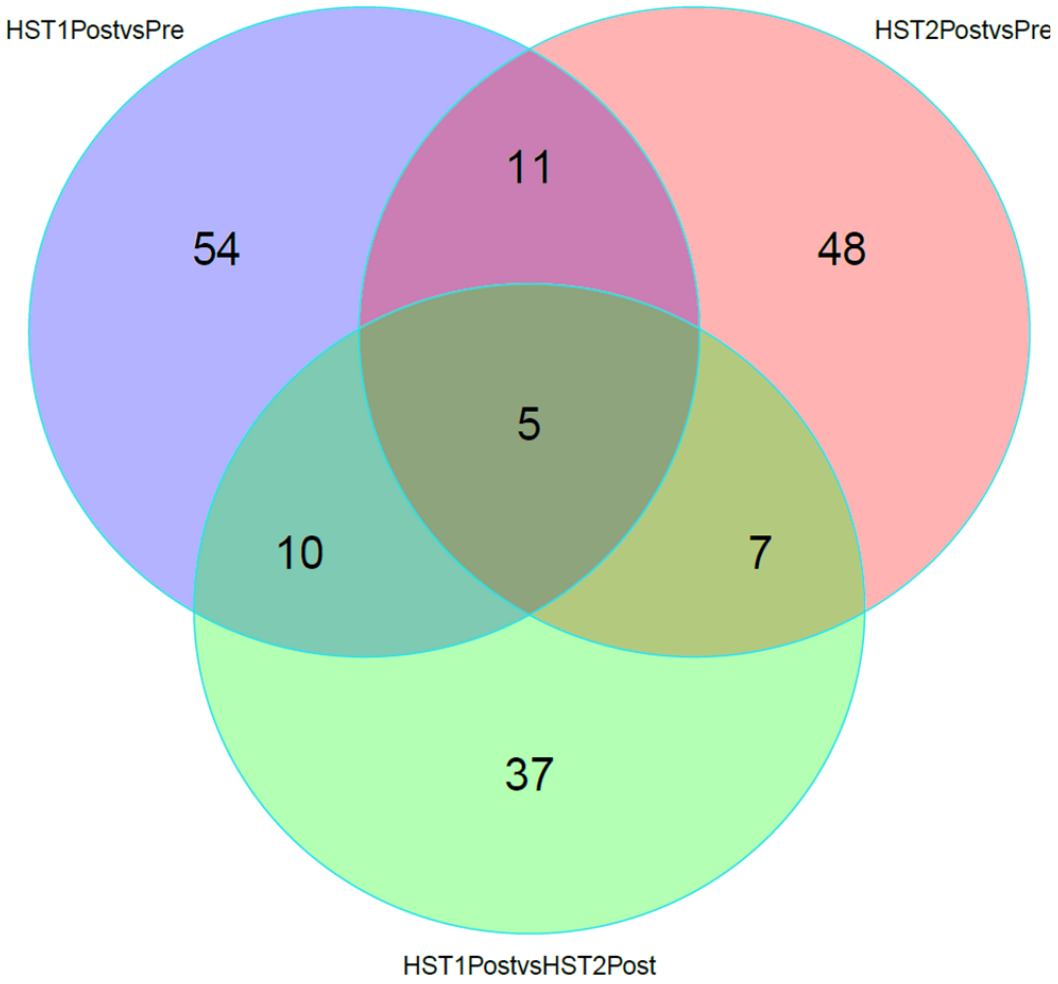


Figure 6-4: Number of overlapping compounds detected in plasma samples from each separate t-test for heat stress tests conducted before and after the condensed heat acclimation sessions. HST 1= heat stress test one, conducted before heat acclimation sessions. HST2 = heat stress test two, conducted after the heat acclimation sessions.

The total number of compounds in each separate t-test was also compared to see the number of overlapping compounds (Figure 6-4). It was expected that the number of overlapping compounds would be high for the two pre/post t-tests and low for the HST1 Post/ HST2 post t-test. This is because it is thought that the same pathways and metabolites would be affected across both heat stress tests, but the metabolites that are significantly different in relative concentration after the HSTs might differ from those changing across the heat stress tests. Instead, the number of overlapping compounds was low for all three. This may have been because the number of compounds that were statistically significantly in all 3 of the separate t-tests was low, with the maximum being 80 compounds.

6.4.1.2 Urine Samples

In comparison to the plasma samples, there were more statistically significant compounds present in urine (66 in urine vs 4 in plasma). With this increase of detected and annotated compounds came more opportunity for biological interpretation. Despite there being 46 unique water-soluble compounds with multiple possible annotations, there were not any pathways that were significantly enriched. This was because there were not enough matches to a single pathway (minimum of three) to ensure a pathway was truly enriched rather than a singular metabolite fluctuation in relative concentration.

Table 6-3: One Way Repeated Measures ANOVA (OWRMA) results for annotated lipid compounds featuring statistically significant fold changes across time points pre/post heat stress test 1 (HST1PostvsHST1Pre), heat stress test 1 Post/ heat stress test 2 post (HST1PostvsHST2Post), pre/post heat stress test 2 HST2PostvsHST2Pre and heat stress test 1 pre/ heat stress test 2 pre (HST1PrevsHST2Pre). All interactions had a statistical significance of P<0.05 and any non-significant interactions are labelled as NSS. Positive fold changes are highlighted green, negative fold changes are highlighted orange and any non-statistically significant changes are labelled as NSS. OWRMA = one-way repeated measures ANOVA. FC= fold change.

	ANOVA	HST1Pre Vs HST2Pre	HST1Post Vs HST1Pre	HST1Post Vs HST2Post	HST2Post Vs HST2Pre
name	F-value	FC	FC	FC	FC
Acetylcarnitine	5.03	NSS	NSS	NSS	0.16
Methylbutyrylcarnitine	3.72	NSS	NSS	NSS	2.01
Hydroxyoctanoylcarnitine	4.00	NSS	NSS	NSS	2.98
3,4-Methylenesebacic acid	5.42	NSS	0.14	NSS	NSS
Dihydroxyoctanoic acid	6.60	NSS	3.13	NSS	NSS
3,4-Methylene suberic acid	9.12	NSS	3.76	NSS	NSS
3-Methylsuberic acid; Azelaic acid; 2,4-Dimethylpimelic acid	5.34	2.75	NSS	NSS	NSS

3-Hydroxytetradecanedioic acid	4.08	4.51	NSS	NSS	NSS
Butyrylcarnitine; Isobutyrylcarnitine	5.45	NSS	NSS	0.48	NSS
2-Hydroxyvaleric acid	3.61	NSS	NSS	2.27	NSS
Nonanoylcarnitine	3.72	NSS	NSS	3.36	NSS
3-Oxo-4-pentenoic acid; cis-2-hydroxypenta-2,4-dienoic acid; cis-acetylacrylic acid	5.69	0.16	NSS	NSS	NSS
Decatrienoic acid	3.18	0.25	NSS	NSS	NSS
Citraconic acid; Glutaconic acid; Dioxopentanoic acid	5.8102	0.36358	NSS	NSS	NSS

There were also lipid metabolism changes present with 14 different fatty acyls (Table 6-3), and four unique steroids and steroid derivatives that significantly changed over time in the repeated one-way ANOVA analysis. Within these 14 fatty acyls, there were clear separated changes between the different comparisons. When comparing the two samples taken before the heat stress tests (HST1 pre/HST2pre) there were five fatty acids that were significant, with a mixture of fold change increases and decreases in relative concentration. Across the first heat stress test there were three different fatty acids that significantly changed in concentration. Again, these were a mixture of fold change increases and decreases in relative concentration. For the second heat stress test, taken after heat acclimation sessions had occurred, there were three acyl carnitines that significantly differed in relative concentration. When comparing the two post heat stress test time points (HST1post/HST2 post) there was a further three unique fatty acyls that were statistically significant.

Changes in fatty acyls are not unexpected as a major part of the heat stress test is exercise. A metabolic response to exercise is changes in fatty acyl concentrations (as can be seen in chapter 5). One complication with the HSTs is that it is difficult to know which of the metabolic responses are due to the heat stress or the exercise, or a combined effect of exercising in the heat as both occurring at the same time. It could be suggested that the two are intangible, as heat is a by-product of exercise and even

when the body is exposed to heat, there will also be affects from the stress the body is being subjected to.

When looking at the number of overlapping compounds (figure 6-5), there are very few overlapping compounds between the different comparisons. It leads to the suggestion that the Cond HA sessions are causing a lasting effect on the human metabolism as the metabolites affected the first time are not significant in the second test HST.

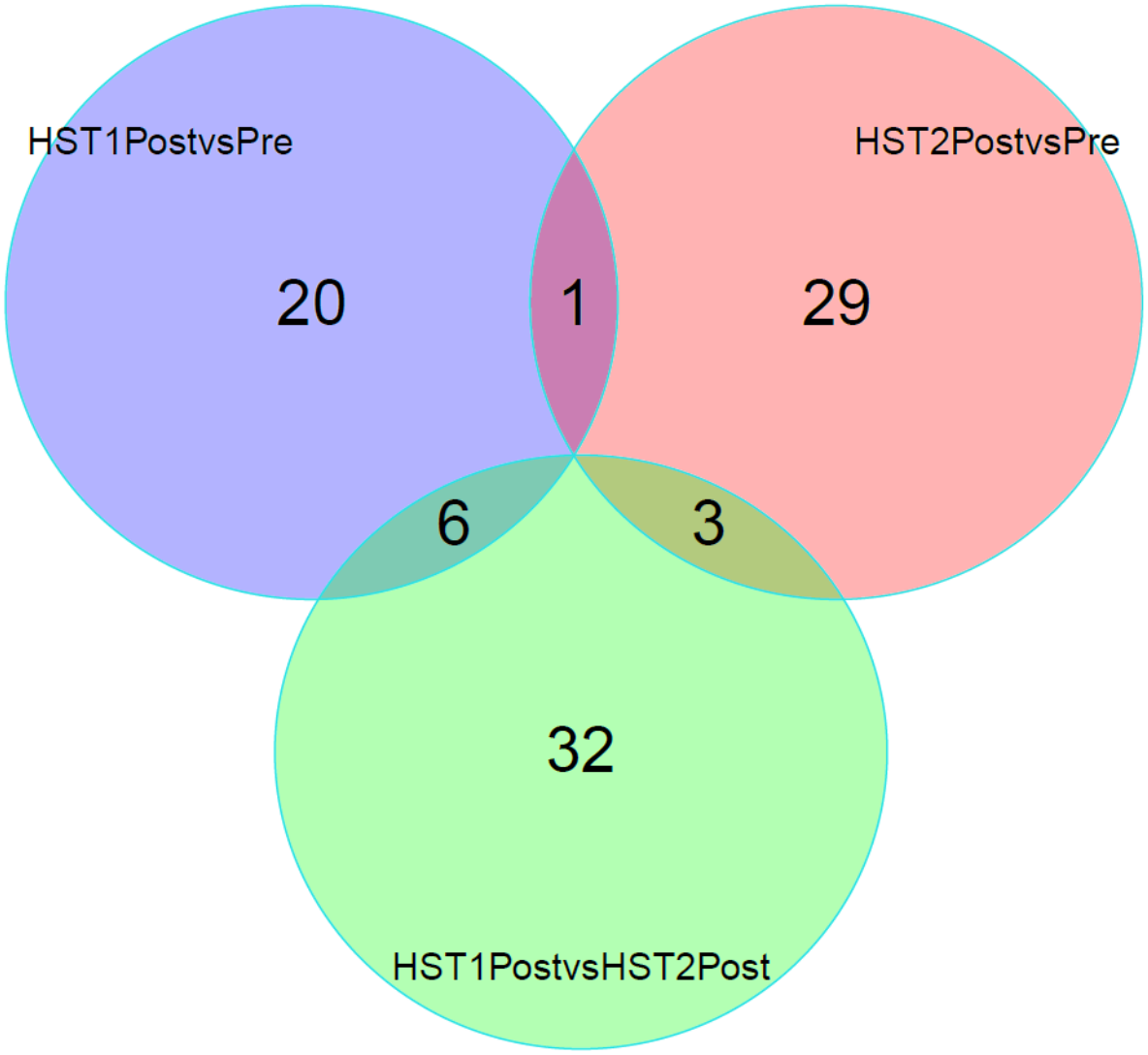


Figure 6-5: Number of overlapping annotated compounds that were detected in urine samples in each separate t-test conducted before and after the condensed heat acclimation sessions. HST 1= heat stress test one, conducted before heat acclimation sessions. HST2 = heat stress test two, conducted after the heat acclimation sessions.

Overall, it can be said that HSTs elicit a bodily response similar to that in exercise studies, which is not unexpected as exercise is a big feature of the tests. To answer the research question: Are there any metabolic differences in samples collected before and after heat acclimation as demonstrated by two independent heat stress tests? Yes, there are clear differences the two different heat stress tests that are caused by how the condensed heat acclimation sessions have influenced the body's metabolism pathways. Further research is required to improve the validity of the statistical analysis and to investigate further into which metabolite pathways are being most affected.

6.4.2 Are There Any Metabolic Differences In Samples Collected Before And After Heat Acclimation As Demonstrated By Two Independent Hypoxic Tolerance Tests?

This question was answered by comparing the changes occurring in the plasma samples collected at time points across the HTTs before and after the heat acclimation sessions. This was completed by applying a repeated one-way ANOVA and filtering the results. Any metabolites that were not annotated and/or statistically significant (P>0.05) were removed, reducing the number of compounds that were changing statistically significantly from 89 to 2. The two remaining compounds were fatty acyls, with one being an acylcarnitine and the other a fatty acid (Table 6-4). These compounds were also only statistically significant in the first HTT test.

Table 6-4: Repeated one-way ANOVA results, comparing fold changes across the compounds that were statistically significant at one or more time point comparison. All interactions had a statistical significance of P<0.05 and any non-significant interactions are labelled as NSS. Fold chance increases in relative concentration are filled in green and any fold change decreases in relative concentration are filled in orange. OWRMA = one way repeated measures ANOVA.

		OWRMA	HTT1Post Vs HTT1Pre	HTT2Post Vs HTT2Pre	HTT1Post Vs HTT2Post
Compound Class	name	F value	Fold Change	Fold Change	Fold Change
Fatty Acyl	Tetradecadiencarnitine	3.49	0.26	NSS	NSS
Fatty Acyl	Glycodeoxycholic acid; Glycochenodeoxycholic acid	1.79	NSS	5.80	NSS

When looking at the separate ad-hoc t-tests in more detail, there were more compounds that were significant. Two in HTT1 Post vs HTT1 Pre, five in HTT2 Post vs HTT2 Pre and three in HTT1 Post vs HTT2 Post. This low number of statistically significant compounds meant that biological interpretation was extremely limited. It can be noted that there were no overlapping compounds between the three time point comparisons and that six of the compounds were lipids. This means that there is a possibility that the Cond HA sessions influenced how the body deals with hypoxic conditions, but that a lot more further research would be required to support and confirm this hypothesis.

6.4.3 Research Question 2: Are There Metabolic Changes Over Time During The Heat Acclimation Sessions?

This question was answered in two different ways:

1. By comparing the changes occurring in the samples across all the Post Condensed Heat Acclimation (Cond HA) sessions to investigate what metabolic changes are occurring across the whole Cond HA window.

2. By comparing the first Cond HA sessions from each day to see what metabolic changes occur within one Cond HA session, and if there are any differences between those two sessions due to the following Cond HA sessions on day zero.

This was completed by applying a repeated one-way ANOVA and filtering the results by removing ay metabolites that were not annotated. This vastly reduced the number of compounds that were changing statistically significantly. This reduction of number of compounds left after the statistical analysis meant that biological interpretation limited.

6.4.3.1 Plasma Samples

When comparing the plasma samples taken after each heat acclimatisation session, the number of compounds that were statistically significant in the repeated one-way ANOVA was very low at 16, which was then reduced further to two when filtering to compounds that were annotated (Table 6-5). For both compounds, all the statistical changes occurred over the first day of heat acclimation sessions.

Table 6-5: OWRMA results for compounds that were statistically significant during the HTTs conducted before and after heat acclimation sessions. All interactions had a statistical significance of P<0.05 and any non-significant interactions are labelled as NSS. Fold chance increases in relative concentration are filled in green and any fold change decreases in relative concentration are filled in orange. OWRMA= one way repeated measures ANOVA. CHA = condensed heat acclimnation.

		OWRMA	CHA1Post Vs CHA1Pre	CHA4Post Vs CHA1Pre	CHA1 post Vs CHA2 post	CHA1 Vs CHA4
Compound Class	Compound Name	F value	Fold Change	Fold Change	Fold Change	Fold Change
Water Soluble	Methylxanthine; Oxoguanine	8.20	NSS	0.46	NSS	2.25
Water Soluble	Allantoin	5.91	0.34	NSS	0.15	NSS

Unsurprisingly there were not any metabolic pathways that were that were reported as being significantly enriched, as there were too few compounds left to identify any metabolic pathways.

Examining the pre and post session samples from the first heat acclimation session of each day allowed the changes that occur over one session to be investigated, and whether the first day of heat acclimation had had any lasting effect on the body. On the first day, there were 50 compounds that changed statistically significantly of which six were annotated. One of the annotated compounds was water soluble, with the rest being a mixture of lipid classes (three fatty acids and two glycerophospholipids). On the second day, there were 16 different compounds that were statistically significant, of which four were annotated. One was a glycerophospholipid and three were fatty acyls. The fatty acids detected all dropped in relative concentration, which is similar to day one where 2/3 of the fatty acids decreased in relative concentration. Due to the lack of significant annotated compounds, any biological interpretation was limited.

From the resulting metabolic changes in plasma samples, the majority of the statistical differences in relative concentration occurred on the first day of the Cond HA sessions, and even with similar changes occurring in sessions on day two, these changes are not as large as on day one. Suggesting that any heat acclimatisation is occurring on the first day of the Cond HA sessions.

6.4.3.2 Urine Samples

There were a lot more compounds that were statistically significant in urine, in comparison to the plasma. After filtering there were 87 annotated compounds detected that were significant in the one-way repeated measures ANOVA. 77 were water soluble and 10 were fatty acyls. Significant changes were detected in multiple pathways (Table 6-6).

Table 6-6: Metabolite pathways that were significantly enriched with multiple metabolites detected in urine samples taken across time points during the condensed heat acclimation sessions changing in relative concentration.

Pathway Name	Match Status	p-value
Alanine, aspartate and glutamate metabolism	5/28	2.47E-03
Arginine biosynthesis	3/14	1.17E-02
Butanoate metabolism	3/15	1.42E-02
Pyrimidine metabolism	4/39	4.64E-02

The Alanine, aspartate and glutamate metabolism pathway, Arginine biosynthesis pathway and the Butanoate metabolism pathway are all closely connected to the citrate cycle, which is central to the metabolic pathways network where the catabolic reactions end and many anabolic reactions begin. Changes to this pathway suggest that there are changes to the energy requirements of the body. Table 6-7 highlights the metabolites of interest in the alanine, aspartate and glutamate metabolism pathway and the fold changes across each of the t-tests. Succinate, which is incredibly important in both the citric acid cycle and cell respiratory chain, shows a large fold decrease in the day one time points when compared to the same time point on day zero. There is also a large fold decrease in relative concentration when comparing the end of the last heat acclimation session with the start of the first heat acclimation session. Other studies on succinate in heat stress and heat acclimation show a range of responses in succinate to changes in heat[337-340]. Other metabolite changes occurring within the metabolism pathway are similar on both days, with the changes being less substantial on the second day. This would suggest that the CHA sessions on day one were influencing the metabolism. Although it should be taken into consideration that all annotations in the pathway were once again level 3 annotations and therefore should not be used with absolute confidence as they are only matched on MS1 data and/or molecular formulae.

Table 6-7: Repeated one-way ANOVA results for the compounds detected as significant in the alanine, aspartate, and glutamate metabolism pathway. All compounds in red are the compound annotations featured in the pathway, other annotations are possible annotations for that compound based on the MS1 database searches. All interactions had a statistical significance of P<0.05 and any non-significant interactions are labelled as NSS. Fold chance increases in relative concentration are filled in green and any fold change decreases in relative concentration are filled in orange.

		2-amino-3-oxoadipic acid; N-Formyl-L-glutamic acid; N-Acetylaspartic acid	Iminoaspartic acid; 2-oxosuccinamic acid	Ureidoisobutyric acid; Glutamine	Succinate	5-Hydroxy-2,4-dioxopentanoic acid; 2-Methoxy-2-butenedioic acid; alpha-Ketoglutaric acid; 4-Methyl 2-oxobutanedioate; 2-methyl-3-oxosuccinic acid
OWRMA	F value	4.26	10.56	5.27	11.62	10.7
	Raw P-value	2.88E-02	1.10E-03	1.50E-02	7.31E-04	1.04E-03
CHA1 Vs CHA4	FC	0.37	4.01	NSS	NSS	8.34
	Raw P-value	3.73E-02	5.57E-03	NSS	NSS	1.18E-03
CHA1 Vs CHA5	FC	NSS	NSS	NSS	0.22	NSS
	Raw P-value	NSS	NSS	NSS	4.50E-02	NSS
CHA1 Vs CHA8	FC	0.17	NSS	3.58	0.11	4.96
	Raw P-value	2.98E-02	NSS	3.04E-02	2.09E-03	2.70E-02
CHA4 Vs CHA5	FC	NSS	0.25	0.12	NSS	0.13
	Raw P-value	NSS	3.48E-03	3.34E-02	NSS	2.05E-04
CHA4 Vs CHA8	FC	NSS	NSS	NSS	0.14	NSS
	Raw P-value	NSS	NSS	NSS	5.71E-03	NSS
CHA5 Vs CHA8	FC	NSS	NSS	5.02	NSS	2.51
	Raw P-value	NSS	NSS	1.15E-03	NSS	9.11E-03

There were ten fatty acyl lipids (eight fatty acids and two acyl carnitines) that were significantly changing in relative concentration across the condHA sessions (Table 6-8). These large fold change increases in relative concentration occurred over each day, but when comparing the relative concentrations at the end of each day, only two fatty acyls were significantly different in relative concentration. Fatty acyls that were significant over both days had larger fold increases over day one (CHA5vsCHA8) compared to day zero (CHA1vsCHA4). This increase in fatty acids suggests that the metabolism pathways that utilise ß-oxidation of fatty acids may be increased.

Table 6-8: OWRMA results for lipids that were statistically significant during the condensed heat acclimation sessions conducted. All interactions had a statistical significance of P<0.05 and any non-significant interactions are labelled as NSS. All fold change increases are highlighted green, whilst all fold change decreases are highlighted orange. OWRMA = one-way repeated measures ANOVA, CHA = Condensed heat acclimation. FC= fold change.

		OWRMA	CHA1 Vs CHA4	CHA1 Vs CHA5	CHA1 Vs CHA8	CHA4 Vs CHA5	CHA4 Vs CHA8	CHA5 Vs CHA8
Compound Class	Compounds	F value	FC	FC	FC	FC	FC	FC
Fatty Acyls	Dimethyl-octadienedioic acid; 2,3-Methylene suberic acid	5.74	NSS	NSS	NSS	NSS	NSS	9.97
Fatty Acyls	3-Methylpimelic acid; Dimethyl adipate; Suberic acid; 2,4-Dimethyladipic acid; 2,4-Dimethyladipic acid	6.80	3.90	NSS	7.44	NSS	NSS	5.22
Fatty Acyls	Nonadienoic acid	16.37	5.62	NSS	5.17	0.15	NSS	7.07
Fatty Acyls	2-Hydroxyvaleric acid	9.34	0.16	NSS	0.24	6.66	NSS	0.22
Fatty Acyls	Decenedioic acid; Dihydroxy-decynoic acid; Dioxodecanoic acid	11.82	3.08	NSS	4.51	0.20	NSS	7.54
Fatty Acyls	Decenoylcarnitine	16.54	NSS	0.39	NSS	0.29	NSS	NSS
Fatty Acyls	Hydroxyisovalerylcarnitine	5.08	4.11	NSS	NSS	0.11	NSS	NSS

Fatty Acyls; water soluble	Isobutyric acid; Butyric acid	4.64	6.84	NSS	NSS	NSS	NSS	NSS
Fatty Acyls; water soluble	N-valeryl-Glycine; 3-dehydrocarnitine; 5-Acetamidopentanoic acid; Isovalerylglycine; N-Acetylvaline; 2-Methylbutyrylglycine	6.13	0.12	NSS	NSS	NSS	4.02	NSS
Fatty Acyls; water soluble	3,4-Methylenesebacic acid	3.89	3.86	NSS	NSS	NSS	0.23	NSS

To answer the research question: are there metabolic changes over time during the heat acclimation sessions? Yes, there are strong results in both plasma and urine sample types that show that condensed heat acclimation sessions cause metabolic changes to occur within the human body. It is clear in both sample types that a lot of the changes are occurring on the first day of heat acclimation sessions (day zero) and although similar changes can be seen on the second day (day one) in urine. Urine also shows that the condHA sessions are affecting the TCA cycle and that the body is requiring less energy throughout the sessions and throughout the second day.

6.5 CONCLUSIONS

The purpose of the study reported here was to investigate the metabolic changes occurring during heat stress tests, hypoxic tolerance tests and heat acclimation sessions, in conjunction with investigating any possible metabolic markers that might be able to be used to predict that heat acclimation has occurred.

Results indicated that metabolic changes in heat stress happen in a similar manner to exercise studies, with signs of lipolysis occurring as relative concentrations of fatty acyls increased as some glycerophospholipids decreased. As exercise was undertaken throughout the heat stress test, this conclusion was expected as it is difficult to determine what a marker of heat stress was and what a marker of exercise stress was. From the urine samples there were metabolic changes occurring in the heat stress test conducted before the heat acclimation sessions that do not occur in the heat stress test conducted after heat acclimation sessions, thus showing that the condensed heat acclimation sessions did have a lasting effect on the metabolism. It can be concluded, that condensed heat acclimation influenced the metabolic response to the heat stress tests, with relative concentrations on the second test being significantly different to the relative concentrations on the first test. Further work should be conducted into pathways that are affected by the heat stress tests to improve statistical reliability and reinforce the changes that are displayed here.

Research question two (are there metabolic changes over time during the heat acclimation sessions?) displayed that there could be potential for heat acclimation sessions to have affected how the human metabolism reacts to hypoxic conditions. Although due to there only being a maximum of five annotated compounds that were statistically significant any biological interpretation was extremely limited. From the trends seen here, it should be suggested that the study should be repeated with a larger number of participants so that there are more trends that might come up as significant, and there would be potential for biological interpretation of any changes seen.

There were many metabolic changes occurring across the condensed heat acclimation sessions with many more changes detected in urine samples than in plasma samples (87 statistically significant compounds in urine vs 2 statistically significant compounds in plasma). Urine samples displayed metabolic changes to the alanine, aspartate, and glutamate metabolism pathway, Arginine biosynthesis pathway, Pyrimidine metabolism pathway and the Butanoate metabolism pathway across both days of CHA sessions. These changes show that heat acclimation sessions clearly influenced the body more within the first four sessions with day zero having more significant changes than day one. There did not appear to be any single clear markers of heat acclimatisation that could be applied with confidence. For future work it was clear that urine samples provide a lot more biological information than the plasma samples, so it would be advised to choose that sample type for future research.

One limitation throughout this study was the lack of participants. This meant that sometimes the variation in data meant that only trends could be determined. If the originally planned study had taken place, then the comparisons would have allowed for more precise comparisons and a higher level of confidence in occurring changes, along with more reproducible results and statistical analysis. This study was a useful pilot study to explore what changes are occurring but should be repeated with a larger number of participants to ensure that statistical analysis can be more appropriately applied.

In conclusion, heat acclimation does affect the human metabolism, and that there may be some metabolic changes that could be markers of heat acclimation. Further work is needed to be sure that any changes seen from heat acclimation are occurring because the participants are heat acclimated. It would also be recommended that this study be repeated with a larger number of participants so that more accurate statistical tests can be applied with higher levels of confidence to ensure that the conclusions drawn are accurate.

7 FINAL CONCLUSIONS

The aims of the research presented in this thesis were to develop and optimize UHPLC-MS methods to allow for a wider range of sample types to be analysed by UHPLC-MS along with expanding the bottlenecked area of metabolite annotation, to allow for better characterized metabolomes of a wider range of sample types, along with higher confidence annotations whilst addressing knowledge gaps in how exercise and heat acclimation perturbs the human metabolome.

Chapter Three

Chapter three addressed the issue of optimising UHPLC-MS methods to allow for a wider range of sample types to be studied applying an untargeted metabolomics strategy. Specifically, I:

(1) *Developed three complementary UHPLC assays applying 1.0mm ID UHPLC columns and subsequently*

(2) Compared the applicability of the 1.0mm ID UHPLC-MS assays to UHPLC assays applying 2.1mm ID columns with the same stationary phase type and particle size.

(3) Applied the newly developed assays to commonly studied mammalian biofluids (plasma and urine) as well as low volume samples (tears samples <5µL) for assessment of the applicability of 1.0mm ID columns for untargeted metabolomics analysis of low volume samples.

When developing the three complimentary assays (aqueous reversed phase, lipidomic C_{18} reversed phase and HILIC assays), chromatographic peak widths, total peak areas, and the number of compounds detected were compared to investigate whether applying a smaller ID column had allowed for an increase in sensitivity. The number of compounds detected were compared across both columns for the aqueous reversed phase and lipidomics reversed phase assays with the 2.1mm ID column detecting larger numbers of compounds across all dilutions. For the HILIC assays only, the number of compounds detected where higher for the 2.1mm ID column for the 1 in 4 dilution for plasma and urine (the typically used dilution in biological studies) but were higher for the 1.0mm ID column for the 1 in 8 and higher dilutions. It was found that both chromatographic peak widths and total peak areas were higher for the 1.0mm ID assays compared to the 2.1mm ID assays. To investigate why this was happening, flow velocities for the two column IDs were compared. The results of which concluded that the 1.0mm ID column applied a higher flow velocity of 1.5 times higher than the 2.1mm ID column (1019cm/hour for a 0.13 $mL.min^{-1}$ flow rate using the 1.0mm ID columns compared to 693 cm/hour for a 0.40 $mL.min^{-1}$ flow rate on a 2.1mm ID column). Therefore, non-optimal flow velocity was disregarded as potential reason for wider chromatographic peaks because of eddy diffusion.

In conclusion, the 1.0mm and 2.1mm ID comparison showed that at commonly used sample dilutions, the 2.1mm ID assay was more sensitive than the 1.0mm ID assay. However, when applying higher dilutions during sample preparation as can be observed for small volume/mass samples (such as tear fluid), the 1.0mm ID HILIC assays resulted in an increase in the number of compounds detected for 1 in 8 and higher dilutions. It can be said that when samples become more dilute, the 1.0mm ID HILIC assays would be suitable to apply and, in most cases, similar or greater numbers of compounds were detected at higher dilutions when comparing 2.1mm ID and 1.0mm ID assays.

Following this, the 1.0mm ID assays were applied to analyse three pig-derived biofluids (plasma, urine, and tear fluid) and to demonstrate the assays applicability for untargeted metabolomic studies. The tear samples (volume less than 20µL), which were diluted during extraction by a factor of 6.8 and then concentrated by a factor of 5.0 upon reconstitution for polar metabolites and a factor of 3.0 for non-polar metabolites, in combination with the 1.0mm ID column assays provided the detection of 288 unique compounds in tear fluid. On multiple assays, the tear fluid had more metabolites detected than the plasma or urine samples and therefore defining that it is a metabolite rich sample type that would be applicable for future metabolomic studies. The number of putatively annotated metabolites was higher in tear fluid when comparing to previously published literature. Results from this study concluded that the application of 1.0mm ID columns is recommended for the analysis of highly diluted samples where a similar number of or more compounds are typically detected. However, for samples where > 50µL volumes are available and where sample dilutions are ≤ 1 in 4 then 2.1mm ID column assays are recommended because they provide detection of a similar number or greater number of compounds.

Limitations in this study were the injection volume, and sample types, flow rate, system pressure limits. Throughout the comparison of the 2.1mm and 1.0mm ID columns as many parameters were kept the same across both. This included, column length, column stationary phase, column brand, mobile phase, and injection volume. Applying the same injection volume (2µL) onto both column IDs meant that the 1.0 ID column had a band width that was ~4.4 times wider just after injection than for the 2.1mm ID column and therefore band width on injection appears to be the reason why chromatographic peaks widths are broader for the 1.0mm ID columns. Although this did not have a significant impact by the time that the analytes reached the detector it is something that should be investigated in future research to ensure that the 1.0mm ID columns are not being overloaded. When using smaller columns, smaller ID tubing between the column and mass spectrometer must be applied to stop any band broadening occurring after the analytes are eluted from the column. This smaller ID tubing meant that there was more backpressure applied to both the column and UHPLC system, therefore meaning that flow rate had to be lowered. Future work to overcome this limitation would be to analyse samples on a capillary UHPLC instrument instead and there are a small number of scientific publications which have demonstrated this for metabolomics applications. Although, it has not been investigated whether large

scale studies could be applied on capillary UHPLC instrument robustly and reproducibly. Another limitation of this study was the sample types applied, all sample types were biofluids, so it has not been investigated as to how well tissue sample types would be analysed when applying these assays. Future work should be conducted to discover if the trends seen in biofluids are the same as tissue or whether these methods are more applicable to alternative sample types.

Chapter Four

Chapter Four addressed the issues of metabolite annotation and identification in two different classes of metabolites. The first approach was to develop retention time mapping techniques using regression modelling that would allow for retention times available in RT libraries to be mapped from one UHPLC-MS method to another UHPLC-MS method using a set of metabolites present in both RT libraries. This approach was applied primarily for water-soluble and not lipophilic metabolites. The second approach was to develop a retention time library specifically for lipids and a specific assay/method using lipid annotations derived from MS1 and MS/MS data and LipidSearch software. It is important to note that the library was developed using data from biological samples and knowledge of MS1 and MS/MS data for lipids and was not constructed applying chemical standards. From this it was then assessed whether retention time elution trends could be applied to construct models to provide retention times for the lipids not present in the constructed RT library and so to increase the number of lipids which could be annotated.

Retention time mapping methods were developed by obtaining different RT libraries/datasets and translating metabolite names into InCHIKey using the Chemical Translation Service (CTS) or by hand. After this, overlapping metabolites from both libraries/datasets were matched across based on their InChIKey. Regression models were created based on the retention times of the overlapping metabolites from each RT library/dataset. Outliers for each regression model were defined by elution order and were removed from the model to ensure it was accurate and reproducible. Outliers occur due to analyte retention behavior differing between methods, as changing UHPLC parameters such as mobile phase and column stationary phase will affect retention behavior of metabolites. Therefore, both RT libraries/datasets required similar elution orders for the overlapping metabolites for the RT mapping method to be successful and accurate. These models were then validated using 10-fold cross-validation. Once models had passed validation, the original model including all overlapping metabolites from both RT libraries/datasets can then be applied to the whole library, therefore expanding the RT library/dataset. Results from the development of retention time mapping techniques concluded that retention time mapping as a form of increasing metabolite annotation confidence is possible for annotating metabolite matches. An example of this is in the University of Birmingham and Fiehn RT libraries, which when mapped allowed for the University of Birmingham RT libraries to be expanded 840 metabolites for HILIC Negative ion mode and 799 metabolites for HILIC Positive ion mode. It is

also expected that this technique will also help to reduce the number of false annotations made. Retention time mapping should only be conducted using libraries where standards have been analysed and is not suitable for biological samples. HILIC assays are challenging for retention time mapping due to the pH changes and the different buffers applied, which makes achieving accurate mapping models difficult and they are often less robust and reproducible. This being said, accurate mapping models can be created and the models with slightly lower accuracy are still applicable to be used for reducing the number of false annotations. Reversed phase assays are easily mapped across groups, with differences in column, mobile phase or elution gradient having little effect on the accuracy of the model. To ensure an appropriate model is formed, analytical standards applied should separate well on both methods and should cover the whole elution gradient.

One limitation to this work was that there are only a small number of publicly available RT libraries/datasets. This is because there is a lack of standardisation of what data should be made publicly available upon publication of work. To further this work, more investigation should be conducted with larger RT/libraries and datasets. Future work on this area would be to apply the extended retention time libraries to biological data to see if any of the metabolites annotated under lower confidence levels could be improved and/or some of the false annotations removed. Other future work should be to define whether isomeric separation reproducible as during a small investigation on the datasets applied this looked to be possible depending on the assay gradient applied, with some libraries having better isomer separation than others. To translate this work from research into a useable product within the metabolomics community, there is opportunity here to create a large central RT library from which a small selection of analytical standards could be run by a research group to allow for the central RTs to be mapped onto their methods. This would allow for many more groups to gain higher confidence annotations without needing an exact match of instrumentation and analytical methods.

Standards for lipids are less available and more expensive than for water metabolites but there is previous research to show that the RT of lipids can be predicted in silico. Research here was conducted to create a RT library for use at the University of Birmingham that would allow for lipids not present in the constructed RT library to have retention times predicted and so to increase the number of lipids which could be annotated. For this work, two datasets were applied, one for RT model construction and one for model validation. Both datasets were collected on the same instrument, in a similar time frame, applying the same methods. The first dataset that that was applied for RT model construction contained multiple sample types (heart, kidney, liver, muscle, serum, milk, and chocolate) were separated out into different lipid classes and grouped together based on how many double bonds were present in the structure. Those groups were then ordered in ascending number of total carbons present using amalgamated nomenclature at a species level[4]. Mean retention times for each carbon number were then calculated, ensuring only metabolites with a minimum of three replicates and an RSD ≤ 5 were used. Regression models depicting the relationship between RT and the total number of carbons were

produced and the best fitting trend line used for the retention time model. This equation was then used on the matching lipid class and carbon number of second dataset for validation. The predicted retention times were compared to experimental RTs to evaluate the accuracy of the in-silico library. Results declared that only two of the retention time models (acylcarnitines (AcCA) with one double bond, and Glycerophosphoinositols (PI) with three double bonds) were accurate and robust enough to be used for a retention library. The inaccurate prediction of lipids was most likely due to using the total number of carbons and number of double bonds to predict the retention time and would have possibly been more accurate if the structure had been further incorporated into the prediction model. There were also several lipid classes that were unable to be modelled (Diradylglycerolipids (DG), methylphosphocholine (MePC), Cardiolipins (CL), Hexosyl-1-ceramides (HEX-1-CER), Monoacylglycerophosphoglycerols (PG) and Glycerophosphates (PA)) due to a lack of data or an unclear retention time trend. It was concluded that complex lipid structure makes retention times very difficult to predict. This is most likely because this research was using the total number of carbons and number of double bonds to predict the retention time and would have possibly been more accurate if the structure had been further incorporated into the prediction model. What the research does allow for, is the removal false positive annotations.

There were multiple limitations within this research, the first of which was the use of only biological data. To overcome this, research should be completed using chemical standards. Since this work has been completed (December 2020), similar work has been published[247]. Vaňková, et.al applied lipid standards from multiple lipid classes (NIST Standard reference material 1950 human plasma) from one sample type as opposed to biological annotations from multiple sample types. Their work concluded that multiple lipid classes could be modelled on their retention behaviour. Although their research also shows that for some classes, retention was affected by isomers of the same lipid, with their retention models separating out isomers of the same lipid (example PC 16:1/18:2 and PC 14:0/20:3 can both be denotated as PC 34:3). Another limitation was the number of classes able to be modelled due to a lack in data, this should be overcome in future work by investigating the limited classes further to see if retention time models can be constructed and applied within a retention time library. Future work should also be completed on the lipid retention time library would be to investigate if using a more detailed nomenclature that differentiates between fatty acid combinations would allow for the method to work better. This work allowed for the retention time range for each class of lipids across multiple sample types, future work should be conducted to investigate how much this work has affected the ability to remove false negatives of lipids in chromatographic data.

Chapter Five

Chapter five investigated the knowledge gap of how the type and intensity of exercise perturbs metabolism. This was investigated by applying untargeted metabolomics to determine metabolic changes over time during and after exercise related to each exercise type to investigate how metabolism changes across different time points and during recovery after exercise. Another objective of this research was to explore any differences between exercise intensity/duration at the same time points to show any major differences between protocols and which exercise protocol puts the most metabolic stress on the body. Finally, the last objective of this research was to determine if there were any differences in the metabolic responses observed between males and females.

When analysing the differences from each exercise protocol over time, the lower intensity exercises saw changes in lipolysis pathway with fatty acid and acyl carnitine increases over time. This was happening as triacylglycerides were being reduced into glycerol and fatty acids, which were then transported across the cell membrane by carnitine, which forms acylcarnitines in the process. This lipolysis pathway allows the release more of ATP for the body to use as energy. As the higher intensity exercise protocols didn't show these changes, it suggests that alternative metabolic pathways (such as glycolysis) were being activated to maintain required ATP concentrations.

When comparing each of the exercises at the same time point, it can be concluded that MICT and SIT protocols are the most different in terms of energy requirements and that HIIT is somewhere in between the two. The lipolysis changes in MICT where much larger in concentration than in the other two protocols, with the large concentration differences between MICT and HIIT/SIT increases still being present at the passive recovery time point where the HIIT and SIT protocols detected a decrease in lipids.

When investigating the difference in response between the male and female participants there were not many metabolic differences. This was contrasting to the differences seen by S.J. Weaver, et.al.[206] who found a significant difference in response between males and females during the SIT and MICT protocols It should be taken into consideration that there were only eight participants of each sex, and thus with more participants there would be more differences detected as significant.

Throughout the study there was an expectation that with such high intensity exercise being conducted, that more areas of the metabolism would have been perturbed. This lack of changes may have resulted from the statistical analysis being very robust in only using FDR corrected values.

Limitations of this study were the BMI, age range, and number of the participants included in the study. To overcome these limitations, future would be completed by expanding the BMI, age range and number of participants in the study. This would ensure that the biological interpretation was more accurate for a larger portion of the global population. A larger number of participants for each sex would

also allow for more trends to become significant and therefore better biological interpretation can then take place. I would also suggest that HIIT and SIT exercise protocols are investigated further using HILIC or other polar LC methods to gain a better understanding of what changes are occurring over time in those exercise protocols where the change in energy requirement means there are metabolic changes occurring in pathways which are predominantly water-soluble compounds.

Chapter Six

Heat acclimation is a common requirement across multiple careers including athletes and military personnel, with traditional heat acclimation protocols requiring ten days of heat acclimation sessions. Condensing this protocol would save time and money required for this process. The sixth chapter investigated heat stress and heat acclimation effects on human metabolism. The original aim of the study was to explore the metabolic changes occurring during heat stress tests, hypoxic tolerance tests and heat acclimation sessions for 16 participants across traditional ten day heat acclimation protocols used by the MoD followed by a two day condensed heat acclimation protocol for a different set of 16 participants developed by the University of Portsmouth. Due to covid-19 this research was condensed to eight participants, applying only the condensed heat acclimation protocol only.

Results for the heat stress tests indicated that metabolic changes occur in a similar manner to exercise studies, with signs of lipolysis occurring as relative concentrations of fatty acyls increased. As exercise was undertaken throughout the heat stress test, this conclusion was expected as it is difficult to determine what a marker of heat stress was and what a marker of exercise stress was. From the urine samples there were metabolic changes occurring in the heat stress test conducted before the heat acclimation sessions that do not occur in the heat stress test conducted after heat acclimation sessions, thus showing that the condensed heat acclimation sessions did have a lasting effect on the metabolism. It can be concluded, that condensed heat acclimation influenced the metabolic response to the heat stress tests, with relative concentrations on the second test being significantly different to the relative concentrations on the first test.

Results for the hypoxic tolerance tests indicated that metabolic changes might occurred differently before and after heat acclimation as there were differences in fatty acids that were changing significantly in the HTT before but not after heat acclimation sessions had taken place. Although due to there only being a maximum of five annotated compounds that were statistically significant any biological interpretation was extremely limited. From the trends reported in this thesis, it should be suggested that the study should be repeated with a larger number of participants so that there are more trends that might result in a significant difference, and there would be potential for biological interpretation of any changes seen.

There were many metabolic changes occurring across the condensed heat acclimation sessions with many more changes detected in urine samples than in plasma samples (87 statistically significant compounds in urine vs 2 statistically significant compounds in plasma). Urine samples displayed metabolic changes to the alanine, aspartate, and glutamate metabolism pathway, Arginine biosynthesis pathway, Pyrimidine metabolism pathway and the Butanoate metabolism pathway across both days of CHA sessions. These changes show that heat acclimation sessions clearly influenced the body more within the first four sessions with day zero having more significant changes than day one. There did not appear to be any single clear markers of heat acclimatisation that could be applied with confidence. For future work it was clear that urine samples provide a lot more biological information than the plasma samples, so it would be advised to choose that sample type for future research.

The main limitation of this study was the number of participants. Using only six participants meant that sometimes the variation in data meant that only trends could be determined, and that non-FDR corrected data had to be used in order to explore the data for changes to metabolism. If the originally planned study had taken place, then the comparisons would have allowed for more precise comparisons and a higher level of confidence in occurring changes, along with more reproducible results and statistical analysis. This study was useful as a pilot study to explore what changes are occurring but should be repeated with a larger number of participants to ensure that statistical analysis can be more appropriately applied. To overcome this limitation, I would recommend this work be expanded with a larger number of participants. Another limitation to this study was the inability to be able to compare the condensed protocol metabolic changes to the traditional protocol, from this comparison more biological interpretation on how the body reacts to heat acclimation would have been better investigated. A third limitation was that only male participants were used. Further work including female participants would allow for more in-depth analysis to be conducted and exploration on how the different sexes are affected by heat stress and heat acclimation. Another limitation to this study was the time points at which the samples were taken. For the plasma samples, the collection time was up to 20 minutes post acclimation session. As metabolism is extremely dynamic there may have been a number of changes that were missed as the body reverted from a stressed state to a healthy one. In future work I would suggest a much shorter length of time between end of session and sample collection.

In conclusion, heat acclimation does affect the human metabolism, and that there may be some metabolic changes that could be markers of heat acclimation. Further work is needed to be sure that any changes seen from heat acclimation are occurring because the participants are heat acclimated. It would also be recommended that this study be repeated with a larger number of participants so that more accurate statistical tests can be applied with higher levels of confidence to ensure that the conclusions drawn are accurate. Research into how the traditional heat acclimation protocol affects metabolism is still ongoing with Covid-19 restrictions.

How Does The Research Performed Impact On Scientific Research In The Future And What Are The Proposed Next Research Steps?

To conclude this thesis, chapter three has defined that only when samples being injected are more dilute or if smaller sample volumes/masses are diluted during sample preparation, then using a 1mm ID column has advantages in the number of metabolites detected. Otherwise, a 2.1mm ID column should be applied for untargeted metabolomics analysis. Newly developed 1mm ID assays applied to porcine plasma, urine, and tear fluid samples demonstrated that tear fluid is a metabolite rich sample type that is applicable for future metabolomics research, allowing the possibility of a less intrusive sample type which has multiple benefits such as the possibility to conduct studies away from a clinical environment.

Chapter four explored retention time mapping methods and forming a RT library to allow for RT data of lipids with missing RT data. Retention time mapping methods were developed and cross validated before concluding that RT mapping methods proved that it is possible to transfer retention times between research groups, even when alternative UHPLC-MS instrumentation, col[206]umn stationary phases and mobile phases were applied. This allowed for the University of Birmingham reversed phase RT libraries to be expanded by up to 95%, the University of Birmingham HILIC positive ion mode RT library to be expanded by 161% and the University of Birmingham HILIC negative ion mode RT library to be expanded by 157%. Further work is required to complete these mapped retention time libraries with accurate error ranges. In-house lipids retention time models were constructed from regression models for each lipid class with the aim to predict the retention time for lipids of the same class where chemical standards are not available. Conclusions for this were that only two classes were able to be accurately modelled, whilst a lot of the classes had structures that were too complex to be modelled in such a simplistic way, with more structural information required to create accurate models.

Finally, two different biological studies were conducted to investigate how the human metabolome responds to different intensities of exercise and heat acclimation, respectively. For the first study, results showed that as the intensity of exercise increases, the bodily response changes as there is a larger requirement for energy. The second study followed participants through a condensed heat acclimation protocol, where plasma and urine samples were collected at regular time points throughout. Results concluded that heat acclimation influences biofluid metabolomes, as urine samples displayed metabolic changes to the alanine, aspartate, and glutamate metabolism pathway, Arginine biosynthesis pathway, Pyrimidine metabolism pathway and the Butanoate metabolism pathway across both days of condensed heat acclimation sessions.

Proposed next steps have been discussed above and should be conducted in future work to overcome the limitations here as well as expand forward from the conclusions above.

8 REFERENCES

1 Wishart, D. S. *et al.* HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Research* **37**, D603-610, doi:10.1093/nar/gkn810 (2009).

2 Fahy, E. *et al.* A comprehensive classification system for lipids. *Journal of Lipid Research* **46**, 839-861, doi:10.1194/jlr.E400004-JLR200 (2005).

3 Fahy, E., Cotter, D., Sud, M. & Subramaniam, S. Lipid classification, structures and tools. *Biochimica et Biophysica Acta (BBA)* **1811**, 637-647, doi:10.1016/j.bbalip.2011.06.009 (2011).

4 Liebisch, G. *et al.* Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures. *Journal of Lipid Research* **61**, 1539-1555, doi:10.1194/jlr.S120001025 (2020).

5 Voelker, D. R. in *Encyclopedia of Biological Chemistry* (eds William J. Lennarz & M. Daniel Lane) 412-418 (Academic Press, 2013).

6 Blanco, A. & Blanco, G. in *Medical Biochemistry* (eds Antonio Blanco & Gustavo Blanco) 99-119 (Academic Press, 2017).

7 Dufourc, E. J. Sterols and membrane dynamics. *Journal of Chemical Biology* **1**, 63-77, doi:10.1007/s12154-008-0010-6 (2008).

8 Donato, P., Dugo, P. & Mondello, L. in *Liquid Chromatography* (eds Salvatore Fanali, Paul R. Haddad, Colin F. Poole, & Marja-Liisa Riekkola) 201-243 (Elsevier, 2017).

9 Agin, A. *et al.* Metabolomics – an overview. From basic principles to potential biomarkers (part 1). *Médecine Nucléaire* **40**, 4-10, doi:10.1016/j.mednuc.2015.12.006 (2016).

10 Dalgliesh, C. E., Horning, E. C., Horning, M. G., Knox, K. L. & Yarger, K. A gas-liquid-chromatographic procedure for separating a wide range of metabolites occuring in urine or tissue extracts. *Biochemical Journal* **101**, 792-810, doi:10.1042/bj1010792 (1966).

11 Horning, E. C. & Horning, M. G. Metabolic profiles: gas-phase methods for analysis of metabolites. *Clinical Chemistry* **17**, 802-809 (1971).

12 Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. Systematic functional analysis of the yeast genome. *Trends Biotechnol* **16**, 373-378, doi:10.1016/s0167-7799(98)01214-1 (1998).

13 Yang, Q. *et al.* Metabolomics biotechnology, applications, and future trends: a systematic review. *RSC Advances* **9**, 37245-37257, doi:10.1039/c9ra06697g (2019).

14 Bundy, J. G., Davey, M. P. & Viant, M. R. Environmental metabolomics: a critical review and future perspectives. *Metabolomics* **5**, 3, doi:10.1007/s11306-008-0152-0 (2008).

15 Kumar, R., Bohra, A., Pandey, A. K., Pandey, M. K. & Kumar, A. Metabolomics for Plant Improvement: Status and Prospects. *Front Plant Sci* **8**, 1302, doi:10.3389/fpls.2017.01302 (2017).

16 Ramirez, T. *et al.* Metabolomics in toxicology and preclinical research. *ALTEX* **30**, 209-225, doi:10.14573/altex.2013.2.209 (2013).

17 Gowda, G. A. *et al.* Metabolomics-based methods for early disease diagnostics. *Expert Review of Molecular Diagnostics* **8**, 617-633, doi:10.1586/14737159.8.5.617 (2008).

18 Klupczynska, A., Derezinski, P. & Kokot, Z. J. Metabolomics in Medical Sciences--Trends, Challenges and Perspectives. *Acta Poloniae Pharmaceutica* **72**, 629-641 (2015).

19 Puchades-Carrasco, L. & Pineda-Lucena, A. Metabolomics Applications in Precision Medicine: An Oncological Perspective. *Current Topics in Medicinal Chemistry* **17**, 2740-2751, doi:10.2174/1568026617666170707120034 (2017).

20 Azad, R. K. & Shulaev, V. Metabolomics technology and bioinformatics for precision medicine. *Briefings in Bioinformatics* **20**, 1957-1971, doi:10.1093/bib/bbx170 (2019).

21 Jayaraman, S. P. *et al.* Metabolomics and Precision Medicine in Trauma: The State of the Field. *Shock* **50**, 5-13, doi:10.1097/SHK.0000000000001093 (2018).

22 Christopher, K. B. Nutritional metabolomics in critical illness. *Current Opinion in Clinical Nutrition & Metabolic Care* **21**, 121-125, doi:10.1097/MCO.0000000000000451 (2018).

23 Veenstra, T. D. Metabolomics: the final frontier? *Genome Medicine* **4**, 40, doi:10.1186/gm339 (2012).

24 Dunn, W. B. *et al.* Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* **9**, 44-66, doi:10.1007/s11306-012-0434-4 (2012).

25 Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology* **13**, 263-269, doi:10.1038/nrm3314 (2012).

26 Alonzo, A., Marsal, S. & Julia, A. Analytical methods in untargeted metabolomics: state of the art in 2015. *frontiers in bioengineering and biotechnology* **3**, 23, doi:10.3389/fbioe.2015.00023 (2015).

27 Floegel, A. *et al.* Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* **62**, 639-648, doi:10.2337/db12-0495 (2013).

28 Bain, J. R. Targeted metabolomics finds its mark in diabetes research. *Diabetes* **62**, 349-351, doi:10.2337/db12-1189 (2013).

29 Leiherer, A. *et al.* P4768Targeted metabolomics identifies elevated serotonin levels in carriers of a TCF7L2 diabetes-risk allele. *European Heart Journal* **39**, doi:10.1093/eurheartj/ehy563.P4768 (2018).

30 Pitt, J. Metabolomic screening in neonates. *Pathology* **42**, S32, doi:10.1097/01268031-201042001-00066 (2010).

31 Scolamiero, E. *et al.* Targeted metabolomics in the expanded newborn screening for inborn errors of metabolism. *Molecular BioSystems* **11**, 1525-1535, doi:10.1039/c4mb00729h (2015).

32 Shyur, L. F. & Yang, N. S. Metabolomics for phytomedicine research and drug development. *Current Opinion in Chemical Biology* **12**, 66-71, doi:10.1016/j.cbpa.2008.01.032 (2008).

33 Griffiths, W. J. *et al.* Targeted Metabolomics for Biomarker Discovery. *Angewandte Chemie International Edition* **49**, 5426-5445, doi:10.1002/anie.200905579 (2010).

34 Castillo-Peinado, L. S. & Luque de Castro, M. D. Present and foreseeable future of metabolomics in forensic analysis. *Analytica Chimica Acta* **925**, 1-15, doi:10.1016/j.aca.2016.04.040 (2016).

35 St John-Williams, L. *et al.* Targeted metabolomics and medication classification data from participants in the ADNI1 cohort. *Scientific Data* **4**, 170140, doi:10.1038/sdata.2017.140 (2017).

36 Bouatra, S. *et al.* The Human Urine Metabolome. *Plos One* **8**, 28, doi:10.1371/journal.pone.0073076 (2013).

37 Wang, H. *et al.* UHPLC-MS-Based Serum and Urine Metabolomics Reveals the Anti-Diabetic Mechanism of Ginsenoside Re in Type 2 Diabetic Rats. *Molecules* **26**, doi:10.3390/molecules26216657 (2021).

38 Wang, L. L. *et al.* Mechanism of the Effect of High-Intensity Training on Urinary Metabolism in Female Water Polo Players Based on UHPLC-MS Non-Targeted Metabolomics Technique. *Healthcare (Basel)* **9**, doi:10.3390/healthcare9040381 (2021).

39 Palmer, E. A., Cooper, H. J. & Dunn, W. B. Investigation of the 12-Month Stability of Dried Blood and Urine Spots Applying Untargeted UHPLC-MS Metabolomic Assays. *Analytical Chemistry* **91**, 14306-14313, doi:10.1021/acs.analchem.9b02577 (2019).

40 Kanani, H., Chrysanthopoulos, P. K. & Klapa, M. I. Standardizing GC-MS metabolomics. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **871**, 191-201, doi:10.1016/j.jchromb.2008.04.049 (2008).

41 Garcia, A. & Barbas, C. in *Metabolic Profiling: Methods and Protocols* (ed Thomas O. Metz) 191-204 (Humana Press, 2011).

42 Beale, D. J. *et al.* Review of recent developments in GC-MS approaches to metabolomics-based research. *Metabolomics* **14**, 152, doi:10.1007/s11306-018-1449-2 (2018).

43 Stettin, D., Poulin, R. X. & Pohnert, G. Metabolomics Benefits from Orbitrap GC-MS-Comparison of Low- and High-Resolution GC-MS. *Metabolites* **10**, 143, doi:10.3390/metabo10040143 (2020).

44 Kuhara, T. *et al.* Identification of new biomarkers of pyridoxine-dependent epilepsy by GC/MS-based urine metabolomics. *Analytical Biochemistry* **604**, 113739, doi:10.1016/j.ab.2020.113739 (2020).

45 Misra, B. B. & Olivier, M. High Resolution GC-Orbitrap-MS Metabolomics Using Both Electron Ionization and Chemical Ionization for Analysis of Human Plasma. *Journal of Proteome Research* **19**, 2717-2731, doi:10.1021/acs.jproteome.9b00774 (2020).

46 Paglia, G., Smith, A. J. & Astarita, G. Ion mobility mass spectrometry in the omics era: Challenges and opportunities for metabolomics and lipidomics. *Mass Spectrometry Reviews* **n/a**, doi:10.1002/mas.21686 (2021).

47 Paglia, G. & Astarita, G. Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nature Protocols* **12**, 797-813, doi:10.1038/nprot.2017.013 (2017).

48 Chouinard, C. D., Nagy, G., Smith, R. D. & Baker, E. S. in *Advances in Ion Mobility-Mass Spectrometry: Fundamentals, Instrumentation and Applications* Vol. 83 *Comprehensive Analytical Chemistry* (eds W. Alexander Donald & James S. Prell) 123-159 (Elsevier, 2019).

49 Mairinger, T., Causon, T. J. & Hann, S. The potential of ion mobility-mass spectrometry for non-targeted metabolomics. *Current Opinion in Chemical Biology* **42**, 9-15, doi:10.1016/j.cbpa.2017.10.015 (2018).

50 Ramautar, R. in *Advances in Clinical Chemistry* Vol. 74 (ed Gregory S. Makowski) 1-34 (Elsevier, 2016).

51 Monton, M. R. & Soga, T. Metabolome analysis by capillary electrophoresis-mass spectrometry. *Journal of Chromatography A* **1168**, 237-246; discussion 236, doi:10.1016/j.chroma.2007.02.065 (2007).

52 Sugimoto, M., Wong, D. T., Hirayama, A., Soga, T. & Tomita, M. Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics* **6**, 78-95, doi:10.1007/s11306-009-0178-y (2010).

53 Soga, T. *et al.* Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *Journal of Proteome Research* **2**, 488-494, doi:10.1021/pr034020m (2003).

54 Soga, T. *et al.* Metabolomic profiling of anionic metabolites by capillary electrophoresis mass spectrometry. *Analytical Chemistry* **81**, 6165-6174, doi:10.1021/ac900675k (2009).

55 Udo, R. *et al.* Urinary charged metabolite profiling of colorectal cancer using capillary electrophoresis-mass spectrometry. *Scientific Reports* **10**, 21057, doi:10.1038/s41598-020-78038-2 (2020).

56 Markley, J. L. *et al.* The future of NMR-based metabolomics. *Current Opinion in Biotechnology (COBIOT)* **43**, 34-40, doi:10.1016/j.copbio.2016.08.001 (2017).

57 Wishart, D. S. NMR metabolomics: A look ahead. *Journal of Magnetic Resonance* **306**, 155-161, doi:10.1016/j.jmr.2019.07.013 (2019).

58 Abreu, A. C., Navas, M. M., Fernandez, C. P., Sanchez-Santed, F. & Fernandez, I. NMR-Based Metabolomics Approach to Explore Brain Metabolic Changes Induced by Prenatal Exposure to Autism-Inducing Chemicals. *ACS Chemical Biology* **16**, 753-765, doi:10.1021/acschembio.1c00053 (2021).

59 Yang, B. *et al.* Novel Metabolic Signatures of Prostate Cancer Revealed by (1)H-NMR Metabolomics of Urine. *Diagnostics (Basel)* **11**, 149, doi:10.3390/diagnostics11020149 (2021).

60 Mishra, R. K., Cherusseri, J., Bishnoi, A. & Thomas, S. in *Spectroscopic Methods for Nanomaterials Characterization* (eds Sabu Thomas, Raju Thomas, Ajesh K. Zachariah, & Raghvendra Kumar Mishra) 369-415 (Elsevier, 2017).

61 Kyle, J. E. *et al.* Uncovering biologically significant lipid isomers with liquid chromatography, ion mobility spectrometry and mass spectrometry. *Analyst* **141**, 1649-1659, doi:10.1039/c5an02062j (2016).

62 Cuyckens, F. *et al.* Product ion mobility as a promising tool for assignment of positional isomers of drug metabolites. *Rapid Commun Mass Spectrom* **25**, 3497-3503, doi:10.1002/rcm.5258 (2011).

63 Li, H., Bendiak, B., Siems, W. F., Gang, D. R. & Hill, H. H., Jr. Carbohydrate structure characterization by tandem ion mobility mass spectrometry (IMMS)2. *Analytical Chemistry* **85**, 2760-2769, doi:10.1021/ac303273z (2013).

64 Simmler, C., Napolitano, J. G., McAlpine, J. B., Chen, S. N. & Pauli, G. F. Universal quantitative NMR analysis of complex natural samples. *Current Opinion in Biotechnology (COBIOT)* **25**, 51-59, doi:10.1016/j.copbio.2013.08.004 (2014).

65 David, A. & Rostkowski, P. in *Environmental Metabolomics* (eds Diana Álvarez-Muñoz & Marinella Farré) 35-64 (Elsevier, 2020).

66 Ettre, L. S. & Sakodynskii, K. I. M. S. Tswett and the discovery of chromatography I: Early work (1899–1903). *Chromatographia* **35**, 223-231, doi:10.1007/bf02269707 (1993).

67 Martin, A. J. & Synge, R. L. A new form of chromatogram employing two liquid phases: A theory of chromatography. 2. Application to the micro-determination of the higher monoamino-acids in proteins. *Biochemical Journal* **35**, 1358-1368, doi:10.1042/bj0351358 (1941).

68 Guillaume, Y. & Guinchard, C. Study and Optimization of Column Efficiency in HPLC: Comparison of Two Methods for Separating Ten Benzodiazepines. *Journal of Liquid Chromatography* **17**, 1443-1459, doi:10.1080/10826079408013172 (1994).

69 Ikegami, T. & Tanaka, N. Monolithic columns for high-efficiency HPLC separations. *Current Opinion in Chemical Biology* **8**, 527-533, doi:10.1016/j.cbpa.2004.08.009 (2004).

70 Tanaka, N. *et al.* Monolithic Silica Columns for HPLC, Micro-HPLC, and CEC. *Journal of High Resolution Chromatography* **23**, 111-116, doi:10.1002/(sici)1521-4168(20000101)23:1<111::Aid-jhrc111>3.0.Co;2-h (2000).

71 Issaq, H. J., Abbott, E. & Veenstra, T. D. Utility of separation science in metabolomic studies. *Journal of Separation Science* **31**, 1936-1947, doi:10.1002/jssc.200700601 (2008).

72 Issaq, H. J., Van, Q. N., Waybright, T. J., Muschik, G. M. & Veenstra, T. D. Analytical and statistical approaches to metabolomics research. *Journal of Separation Science* **32**, 2183-2199, doi:10.1002/jssc.200900152 (2009).

73 Sorensen, M. J. & Kennedy, R. T. Capillary ultrahigh-pressure liquid chromatography-mass spectrometry for fast and high resolution metabolomics separations. *Journal of Chromatography A* **1635**, 461706, doi:10.1016/j.chroma.2020.461706 (2021).

74 Lough, W. J. & Wainer, I. W. *High Performance Liquid Chromatography: Fundamental Principles and Practice*. (Taylor & Francis, 1995).

75 van Deemter, J. J., Zuiderweg, F. J. & Klinkenberg, A. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chemical Engineering Science* **5**, 271-289, doi:10.1016/0009-2509(56)80003-1 (1956).

76 Murphy, A. T., Berna, M. J., Holsapple, J. L. & Ackermann, B. L. Effects of flow rate on high-throughput quantitative analysis of protein-precipitated plasma using liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom* **16**, 537-543, doi:10.1002/rcm.606 (2002).

77 Uliyanchenko, E., van der Wal, S. & Schoenmakers, P. J. Challenges in polymer analysis by liquid chromatography. *Polymer Chemistry* **3**, 2313-2335, doi:10.1039/c2py20274c (2012).

78 Jandera, P. & Henze, G. in *Ullmann's Encyclopedia of Industrial Chemistry* (2011).

79 Jankevics, A., Jenkins, A., Dunn, W. B. & Najdekr, L. An improved strategy for analysis of lipid molecules utilising a reversed phase C30 UHPLC column and scheduled MS/MS acquisition. *Talanta* **229**, 122262, doi:10.1016/j.talanta.2021.122262 (2021).

80 Cajka, T. & Fiehn, O. Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. *Trends in Analytical Chemistry* **61**, 192-206, doi:10.1016/j.trac.2014.04.017 (2014).

81 Rustam, Y. H. & Reid, G. E. Analytical Challenges and Recent Advances in Mass Spectrometry Based Lipidomics. *Analytical Chemistry* **90**, 374-397, doi:10.1021/acs.analchem.7b04836 (2018).

82 Buszewski, B. & Noga, S. Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique. *Analytical and Bioanalytical Chemistry* **402**, 231-247, doi:10.1007/s00216-011-5308-5 (2012).

83 Alpert, A. J. Effect of salts on retention in hydrophilic interaction chromatography. *Journal of Chromatography A* **1538**, 45-53, doi:10.1016/j.chroma.2018.01.038 (2018).

84 Perez de Souza, L., Alseekh, S., Scossa, F. & Fernie, A. R. Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research. *Nature Methods* **18**, 733-746, doi:10.1038/s41592-021-01116-4 (2021).

85 Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71, doi:10.1126/science.2675315 (1989).

86 Rajawat, J. & Jhingan, G. in *Data Processing Handbook for Complex Biological Data Sources* (ed Gauri Misra) 1-20 (Academic Press, 2019).

87 El-Aneed, A., Cohen, A. & Banoub, J. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Applied Spectroscopy Reviews* **44**, 210-230, doi:10.1080/05704920902717872 (2009).

88 de Koster, C. G. & Schoenmakers, P. J. in *Hyphenations of Capillary Chromatography with Mass Spectrometry* (eds Peter Q. Tranchida & Luigi Mondello) 279-295 (Elsevier, 2020).

89 Raffaelli, A. & Saba, A. Atmospheric pressure photoionization mass spectrometry. *Mass Spectrometry Reviews* **22**, 318-331, doi:10.1002/mas.10060 (2003).

90 Scalbert, A. *et al.* Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* **5**, 435-458, doi:10.1007/s11306-009-0168-0 (2009).

91 Ren, J.-L., Zhang, A.-H., Kong, L. & Wang, X.-J. Advances in mass spectrometry-based metabolomics for investigation of metabolites. *RSC Advances* **8**, 22335-22350, doi:10.1039/c8ra01574k (2018).

92 Dunn, W. B., Bailey, N. J. & Johnson, H. E. Measuring the metabolome: current analytical technologies. *Analyst* **130**, 606-625, doi:10.1039/b418288j (2005).

93 Thomson, J. J. XL. Cathode Rays. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **44**, 293-316, doi:10.1080/14786449708621070 (2009).

94 Lössl, P., van de Waterbeemd, M. & Heck, A. J. The diverse and expanding role of mass spectrometry in structural and molecular biology. *EMBO J* **35**, 2634-2657, doi:10.15252/embj.201694818 (2016).

95 Savaryn, J. P., Toby, T. K. & Kelleher, N. L. A researcher's guide to mass spectrometry-based proteomics. *Proteomics* **16**, 2435-2443, doi:10.1002/pmic.201600113 (2016).

96 Goeddel, L. C. & Patti, G. J. Maximizing the value of metabolomic data. *Bioanalysis* **4**, 2199-2201, doi:10.4155/bio.12.210 (2012).

97 Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395, doi:10.1186/1471-2105-11-395 (2010).

98 Waters. *MassLynx MS Software*, <https://www.waters.com/waters/en_US/MassLynx-MS-Software/nav.htm?locale=en_US&cid=513662> (

99 Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry* **78**, 779-787, doi:10.1021/ac051437y (2006).

100 Scientific, T. F. *Compound Discoverer Software*, <https://www.thermofisher.com/uk/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/compound-discoverer-software.html> (

101 Holman, J. D., Tabb, D. L. & Mallick, P. Employing ProteoWizard to Convert Raw Mass Spectrometry Data. *Current Protocols in Bioinformatics* **46**, 13 24 11-19, doi:10.1002/0471250953.bi1324s46 (2014).

102 Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211-221, doi:10.1007/s11306-007-0082-2 (2007).

103 Creek, D. J. *et al.* Metabolite identification: are you sure? And how do your peers gauge your confidence? *Metabolomics* **10**, 350-353, doi:10.1007/s11306-014-0656-8 (2014).

104 Domingo-Almenara, X. *et al.* The METLIN small molecule dataset for machine learning-based retention time prediction. *Nature Communications* **10**, 5811, doi:10.1038/s41467-019-13680-7 (2019).

105 Bonini, P., Kind, T., Tsugawa, H., Barupal, D. K. & Fiehn, O. Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics. *Analytical Chemistry* **92**, 7515-7522, doi:10.1021/acs.analchem.9b05765 (2020).

106 Naylor, B. C., Catrow, J. L., Maschek, J. A. & Cox, J. E. QSRR Automator: A Tool for Automating Retention Time Prediction in Lipidomics and Metabolomics. *Metabolites* **10**, 237, doi:10.3390/metabo10060237 (2020).

107 Bouwmeester, R., Martens, L. & Degroeve, S. Comprehensive and Empirical Evaluation of Machine Learning Algorithms for Small Molecule LC Retention Time Prediction. *Analytical Chemistry* **91**, 3694-3703, doi:10.1021/acs.analchem.8b05820 (2019).

108 Aalizadeh, R., Nika, M. C. & Thomaidis, N. S. Development and application of retention time prediction models in the suspect and non-target screening of emerging contaminants. *Journal of Hazardous Materials* **363**, 277-285, doi:10.1016/j.jhazmat.2018.09.047 (2019).

109 Yang, J. J. *et al.* Streamlined MRM method transfer between instruments assisted with HRMS matching and retention-time prediction. *Analytica Chimica Acta* **1100**, 88-96, doi:10.1016/j.aca.2019.12.002 (2020).

110 Stanstrup, J., Neumann, S. & Vrhovsek, U. PredRet: prediction of retention time by direct mapping between multiple chromatographic systems. *Analytical Chemistry* **87**, 9421-9428, doi:10.1021/acs.analchem.5b02287 (2015).

111 Vaughan, A. A. *et al.* Liquid chromatography-mass spectrometry calibration transfer and metabolomics data fusion. *Analytical Chemistry* **84**, 9848-9857, doi:10.1021/ac302227c (2012).

112 Bouwmeester, R., Martens, L. & Degroeve, S. Generalized Calibration Across Liquid Chromatography Setups for Generic Prediction of Small-Molecule Retention Times. *Analytical Chemistry* **92**, 6571-6578, doi:10.1021/acs.analchem.0c00233 (2020).

113 Witting, M., Ruttkies, C., Neumann, S. & Schmitt-Kopplin, P. LipidFrag: Improving reliability of in silico fragmentation of lipids and application to the Caenorhabditis elegans lipidome. *PLoS One* **12**, e0172311, doi:10.1371/journal.pone.0172311 (2017).

114 Kind, T. *et al.* LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nature Methods* **10**, 755-758, doi:10.1038/nmeth.2551 (2013).

115 Ma, Y. *et al.* An in silico MS/MS library for automatic annotation of novel FAHFA lipids. *Journal of Cheminformatics* **7**, 53, doi:10.1186/s13321-015-0104-4 (2015).

116 Tsugawa, H. *et al.* Comprehensive identification of sphingolipid species by in silico retention time and tandem mass spectral library. *Journal of Cheminformatics* **9**, 19, doi:10.1186/s13321-017-0205-3 (2017).

117 Aicheler, F. *et al.* Retention Time Prediction Improves Identification in Nontargeted Lipidomics Approaches. *Analytical Chemistry* **87**, 7698-7704, doi:10.1021/acs.analchem.5b01139 (2015).

118 Hong, S. & Lynn, H. S. Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction. *BMC Medical Research Methodology* **20**, 199, doi:10.1186/s12874-020-01080-1 (2020).

119 Misra, B. B. Data normalization strategies in metabolomics: Current challenges, approaches, and tools. *Eur J Mass Spectrom (Chichester)* **26**, 165-174, doi:10.1177/1469066720918446 (2020).

120 Di Guida, R. *et al.* Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. *Metabolomics* **12**, 93, doi:10.1007/s11306-016-1030-9 (2016).

121 Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Analytical Chemistry* **78**, 4281-4290, doi:10.1021/ac051632c (2006).

122 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289-300, doi:10.1111/j.2517-6161.1995.tb02031.x (1995).

123 Bonferroni, C. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze* **8**, 3-62 (1936).

124 Smeraglia, J., Baldrey, S. F. & Watson, D. Matrix effects and selectivity issues in LC-MS-MS. *Chromatographia* **55**, S95-S99, doi:10.1007/BF02493363 (2002).

125 Zhou, W., Yang, S. & Wang, P. G. Matrix effects and application of matrix effect factor. *Bioanalysis* **9**, 1839-1844, doi:10.4155/bio-2017-0214 (2017).

126 Yu, Z. *et al.* Differences between human plasma and serum metabolite profiles. *PLoS One* **6**, e21230, doi:10.1371/journal.pone.0021230 (2011).

127 Sotelo-Orozco, J., Chen, S. Y., Hertz-Picciotto, I. & Slupsky, C. M. A Comparison of Serum and Plasma Blood Collection Tubes for the Integration of Epidemiological and Metabolomics Data. *Frontiers in Molecular Biosciences* **8**, 682134, doi:10.3389/fmolb.2021.682134 (2021).

128 Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research* **46**, D608-D617, doi:10.1093/nar/gkx1089 (2018).

129 Foschi, C. *et al.* Urine metabolome in women with Chlamydia trachomatis infection. *PLoS One* **13**, e0194827, doi:10.1371/journal.pone.0194827 (2018).

130 Armbruster, M. *et al.* Metabolite profiles evaluated, according to sex, do not predict resting energy expenditure and lean body mass in healthy non-obese subjects. *Eur J Nutr* **58**, 2207-2217, doi:10.1007/s00394-018-1767-1 (2019).

131 Lutz, U., Bittner, N., Lutz, R. W. & Lutz, W. K. Metabolite profiling in human urine by LC-MS/MS: method optimization and application for glucuronides from dextromethorphan metabolism. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **871**, 349-356, doi:10.1016/j.jchromb.2008.04.018 (2008).

132 Chen, Y. *et al.* Combination of injection volume calibration by creatinine and MS signals' normalization to overcome urine variability in LC-MS-based metabolomics studies. *Analytical Chemistry* **85**, 7659-7665, doi:10.1021/ac401400b (2013).

133 Chetwynd, A. J., Abdul-Sada, A., Holt, S. G. & Hill, E. M. Use of a pre-analysis osmolality normalisation method to correct for variable urine concentrations and for improved metabolomic analyses. *Journal of Chromatography A* **1431**, 103-110, doi:10.1016/j.chroma.2015.12.056 (2016).

134 Vogl, F. C. *et al.* Evaluation of dilution and normalization strategies to correct for urinary output in HPLC-HRTOFMS metabolomics. *Analytical and Bioanalytical Chemistry* **408**, 8483-8493, doi:10.1007/s00216-016-9974-1 (2016).

135 Warrack, B. M. *et al.* Normalization strategies for metabonomic analysis of urine samples. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **877**, 547-552, doi:10.1016/j.jchromb.2009.01.007 (2009).

136 Choy, C. K. M., Cho, P., Chung, W. Y. & Benzie, I. F. F. Water-soluble antioxidants, in human tears: Effect of the collection method. *Investigative Ophthalmology & Visual Science* **42**, 3130-3134 (2001).

137 Peake, J. M. *et al.* Metabolic and hormonal responses to isoenergetic high-intensity interval exercise and continuous moderate-intensity exercise. *The American Journal of Physiology-Endocrinology and Metabolism* **307**, E539-552, doi:10.1152/ajpendo.00276.2014 (2014).

138 Ghosh, A. & Nishtala, K. Biofluid lipidome: a source for potential diagnostic biomarkers. *Clinical and Translational Medicine* **6**, 10, doi:10.1186/s40169-017-0152-7 (2017).

139 Vanhaeringen, N. J. CLINICAL BIOCHEMISTRY OF TEARS. *Survey of Ophthalmology* **26**, 84-96, doi:10.1016/0039-6257(81)90145-4 (1981).

140 Filik, J. & Stone, N. Analysis of human tear fluid by Raman spectroscopy. *Analytica Chimica Acta* **616**, 177-184, doi:10.1016/j.aca.2008.04.036 (2008).

141 Srinivasan, S., Thangavelu, M., Zhang, L., Green, K. B. & Nichols, K. K. iTRAQ quantitative proteomics in the analysis of tears in dry eye patients. *Investigative Ophthalmology & Visual Science* **53**, 5052-5059, doi:10.1167/iovs.11-9022 (2012).

142 Versura, P. & Campos, E. Disease Update on Human Tear Proteome. *European Ophthalmic Review* **7**, 36-41, doi:10.17925/EOR.2013.07.01.36 (2013).

143 Versura, P. *et al.* Tear proteomics in evaporative dry eye disease. *Eye* **24**, 1396-1402, doi:10.1038/eye.2010.7 (2010).

144 Ananthi, S. *et al.* Comparative proteomics of human male and female tears by two-dimensional electrophoresis. *Experimental Eye Research* **92**, 454-463, doi:10.1016/j.exer.2011.03.002 (2011).

145 Manicam, C. *et al.* Proteomics Unravels the Regulatory Mechanisms in Human Tears Following Acute Renouncement of Contact Lens Use: A Comparison between Hard and Soft Lenses. *Scientific Reports* **8**, 11526, doi:10.1038/s41598-018-30032-5 (2018).

146 Perumal, N., Funke, S., Pfeiffer, N. & Grus, F. H. Proteomics analysis of human tears from aqueous-deficient and evaporative dry eye patients. *Scientific Reports* **6**, 29629, doi:10.1038/srep29629 (2016).

147 Rossi, C. *et al.* Multi-Omics Approach for Studying Tears in Treatment-Naive Glaucoma Patients. *International Journal of Molecular Sciences* **20**, doi:10.3390/ijms20164029 (2019).
148 English, J. T., Norris, P. C., Hodges, R. R., Dartt, D. A. & Serhan, C. N. Identification and Profiling of Specialized Pro-Resolving Mediators in Human Tears by Lipid Mediator Metabolomics. *Prostaglandins Leukot Essent Fatty Acids* **117**, 17-27, doi:10.1016/j.plefa.2017.01.004 (2017).
149 Chen, J., Nichols, K. K., Wilson, L., Barnes, S. & Nichols, J. J. Untargeted lipidomic analysis of human tears: A new approach for quantification of O-acyl-omega hydroxy fatty acids. *The Ocular Surface* **17**, 347-355, doi:10.1016/j.jtos.2019.02.004 (2019).
150 Yazdani, M. *et al.* Tear Metabolomics in Dry Eye Disease: A Review. *Int. J. Mol. Sci.* **20**, doi:10.3390/ijms20153755 (2019).
151 Brown, M. *et al.* Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets. *Bioinformatics* **27**, 1108-1112, doi:10.1093/bioinformatics/btr079 (2011).
152 Benitez Del Castillo, J. M. *et al.* Tear 1H Nuclear Magnetic Resonance-Based Metabolomics Application to the Molecular Diagnosis of Aqueous Tear Deficiency and Meibomian Gland Dysfunction. *Ophthalmic Research* **64**, 297-309, doi:10.1159/000510211 (2021).
153 Shrestha, G. S., Vijay, A. K., Stapleton, F., Pickford, R. & Carnt, N. Human Tear Metabolomics Using Liquid Chromatography-Q Exactive-HF Mass Spectrometry. *bioRxiv*, 2021.2001.2018.424378, doi:10.1101/2021.01.18.424378 (2021).
154 Zhou, L. & Beuerman, R. W. Tear analysis in ocular surface diseases. *Progress in Retinal and Eye Research* **31**, 527-550, doi:10.1016/j.preteyeres.2012.06.002 (2012).
155 Esmaeelpour, M., Cai, J., Watts, P., Boulton, M. & Murphy, P. J. Tear sample collection using cellulose acetate absorbent filters. *Ophthalmic and Physiological Optics* **28**, 577-583, doi:10.1111/j.1475-1313.2008.00603.x (2008).
156 Sebbag, L., Harrington, D. M. & Mochel, J. P. Tear fluid collection in dogs and cats using ophthalmic sponges. *Veterinary Ophthalmology*, doi:10.1111/vop.12502 (2017).
157 Sia, R. K. *et al.* Non-stimulated Tear Sample Collection Using Polyvinyl Alcohol (PVA) Foam and Polyester Wick. *International Journal of Ophthalmology and Clinical Research* **3**, 1-5 (2016).
158 Stuchell, R. N., Feldman, J. J., Farris, R. L. & Mandel, I. D. THE EFFECT OF COLLECTION TECHNIQUE ON TEAR COMPOSITION. *Investigative Ophthalmology & Visual Science* **25**, 374-377 (1984).
159 Vanhaeringen, N. J. & Glasius, E. COLLECTION METHOD DEPENDANT CONCENTRATIONS OF SOME METABOLITES IN HUMAN TEAR FLUID, WITH SPECIAL REFERENCE TO GLUCOSE IN HYPERGLYCEMIC CONDITIONS. *Albrecht Von Graefes Archiv Fur Klinische Und Experimentelle Ophthalmologie* **202**, 1-7 (1977).
160 Nattinen, J., Aapola, U., Jylha, A., Vaajanen, A. & Uusitalo, H. Comparison of Capillary and Schirmer Strip Tear Fluid Sampling Methods Using SWATH-MS Proteomics Approach. *Translational Vision Science & Technology* **9**, 16, doi:10.1167/tvst.9.3.16 (2020).
161 Bachhuber, F., Huss, A., Senel, M. & Tumani, H. Diagnostic biomarkers in tear fluid: from sampling to preanalytical processing. *Scientific Reports* **11**, 10064, doi:10.1038/s41598-021-89514-8 (2021).
162 Leuthold, P. *et al.* Comprehensive Metabolomic and Lipidomic Profiling of Human Kidney Tissue: A Platform Comparison. *Journal of Proteome Research* **16**, 933-944, doi:10.1021/acs.jproteome.6b00875 (2017).
163 Wu, H., Southam, A. D., Hines, A. & Viant, M. R. High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Analytical Biochemistry* **372**, 204-212, doi:10.1016/j.ab.2007.10.002 (2008).
164 Maekawa, K. *et al.* Global metabolomic analysis of heart tissue in a hamster model for dilated cardiomyopathy. *Journal of Molecular and Cellular Cardiology* **59**, 76-85, doi:10.1016/j.yjmcc.2013.02.008 (2013).
165 Ji, B. *et al.* Transcriptomic and metabolomic profiling of chicken adipose tissue in response to insulin neutralization and fasting. *BMC Genomics* **13**, 441, doi:10.1186/1471-2164-13-441 (2012).

166 Ramsay, T. G., Stoll, M. J., Shannon, A. E. & Blomberg, L. A. Metabolomic analysis of longissimus from underperforming piglets relative to piglets with normal preweaning growth. *Journal of Animal Science and Biotechnology* **9**, 36, doi:10.1186/s40104-018-0251-3 (2018).
167 Dame, Z. T. *et al.* The human saliva metabolome. *Metabolomics* **11**, 1864-1883, doi:10.1007/s11306-015-0840-5 (2015).
168 Kujala, U. M. *et al.* Long-term leisure-time physical activity and serum metabolome. *Circulation* **127**, 340-348, doi:10.1161/CIRCULATIONAHA.112.105551 (2013).
169 Kutyshenko, V. P., Molchanov, M., Beskaravayny, P., Uversky, V. N. & Timchenko, M. A. Analyzing and mapping sweat metabolomics by high-resolution NMR spectroscopy. *PLoS One* **6**, e28824, doi:10.1371/journal.pone.0028824 (2011).
170 Calderon-Santiago, M. *et al.* Human sweat metabolomics for lung cancer screening. *Analytical and Bioanalytical Chemistry* **407**, 5381-5392, doi:10.1007/s00216-015-8700-8 (2015).
171 Van den Eede, N. *et al.* Metabolomics analysis of the toxicity pathways of triphenyl phosphate in HepaRG cells and comparison to oxidative stress mechanisms caused by acetaminophen. *Toxicology in Vitro* **29**, 2045-2054, doi:10.1016/j.tiv.2015.08.012 (2015).
172 Ramirez, T. *et al.* Prediction of liver toxicity and mode of action using metabolomics in vitro in HepG2 cells. *Archives of Toxicology* **92**, 893-906, doi:10.1007/s00204-017-2079-6 (2018).
173 Starnes, J. W. *et al.* Exercise-Induced Alterations in Skeletal Muscle, Heart, Liver, and Serum Metabolome Identified by Non-Targeted Metabolomics Analysis. *Metabolites* **7**, doi:10.3390/metabo7030040 (2017).
174 Ferrarini, A. *et al.* Metabolomic Analysis of Liver Tissues for Characterization of Hepatocellular Carcinoma. *Journal of Proteome Research* **18**, 3067-3076, doi:10.1021/acs.jproteome.9b00185 (2019).
175 Rui, L. Energy metabolism in the liver. *Comprehensive Physiology* **4**, 177-197, doi:10.1002/cphy.c130024 (2014).
176 Cajka, T. & Fiehn, O. Increasing lipidomic coverage by selecting optimal mobile-phase modifiers in LC–MS of blood plasma. *Metabolomics* **12**, 34, doi:10.1007/s11306-015-0929-x (2016).
177 Southam, A. D. *et al.* Assessment of human plasma and urine sample preparation for reproducible and high-throughput UHPLC-MS clinical metabolic phenotyping. *Analyst* **145**, 6511-6523, doi:10.1039/d0an01319f (2020).
178 Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology* **37**, 911-917, doi:10.1139/o59-099 (1959).
179 Hall, Z. *et al.* Lipid Remodeling in Hepatocyte Proliferation and Hepatocellular Carcinoma. *Hepatology* **73**, 1028-1044, doi:10.1002/hep.31391 (2021).
180 Savic, D. *et al.* L-Carnitine Stimulates In Vivo Carbohydrate Metabolism in the Type 1 Diabetic Heart as Demonstrated by Hyperpolarized MRI. *Metabolites* **11**, 191 (2021).
181 Sowton, A. P., Griffin, J. L. & Murray, A. J. Metabolic Profiling of the Diabetic Heart: Toward a Richer Picture. *Front Physiol* **10**, 639, doi:10.3389/fphys.2019.00639 (2019).
182 Hsu, Y. H. *et al.* Integrating untargeted metabolomics, genetically informed causal inference, and pathway enrichment to define the obesity metabolome. *International Journal of Obesity* **44**, 1596-1606, doi:10.1038/s41366-020-0603-x (2020).
183 Al-Khelaifi, F. *et al.* A pilot study comparing the metabolic profiles of elite-level athletes from different sporting disciplines. *Sports Medicine - Open* **4**, 2, doi:10.1186/s40798-017-0114-z (2018).
184 Muhsen Ali, A. *et al.* Metabolomic Profiling of Submaximal Exercise at a Standardised Relative Intensity in Healthy Adults. *Metabolites* **6**, 9, doi:10.3390/metabo6010009 (2016).
185 Howe, C. C. F. *et al.* Untargeted Metabolomics Profiling of an 80.5 km Simulated Treadmill Ultramarathon. *Metabolites* **8**, 14, doi:10.3390/metabo8010014 (2018).
186 Stander, Z. *et al.* The altered human serum metabolome induced by a marathon. *Metabolomics* **14**, 150, doi:10.1007/s11306-018-1447-4 (2018).
187 Zhao, J. *et al.* Integration of metabolomics and proteomics to reveal the metabolic characteristics of high-intensity interval training. *Analyst* **145**, 6500-6510, doi:10.1039/D0AN01287D (2020).

188 Danaher, J., Gerber, T., Wellard, R. M., Stathis, C. G. & Cooke, M. B. The use of metabolomics to monitor simultaneous changes in metabolic variables following supramaximal low volume high intensity exercise. *Metabolomics* **12**, 7, doi:10.1007/s11306-015-0883-7 (2015).
189 Pechlivanis, A. *et al.* 1H NMR study on the short- and long-term impact of two training programs of sprint running on the metabolic fingerprint of human serum. *Journal of Proteome Research* **12**, 470-480, doi:10.1021/pr300846x (2013).
190 Heaney, L. M., Deighton, K. & Suzuki, T. Non-targeted metabolomics in sport and exercise science. *Journal of Sports Sciences* **37**, 959-967, doi:10.1080/02640414.2017.1305122 (2019).
191 Morville, T., Sahl, R. E., Moritz, T., Helge, J. W. & Clemmensen, C. Plasma Metabolome Profiling of Resistance Exercise and Endurance Exercise in Humans. *Cell Rep* **33**, 108554, doi:10.1016/j.celrep.2020.108554 (2020).
192 Grapov, D. *et al.* Exercise plasma metabolomics and xenometabolomics in obese, sedentary, insulin-resistant women: impact of a fitness and weight loss intervention. *The American Journal of Physiology-Endocrinology and Metabolism* **317**, E999-E1014, doi:10.1152/ajpendo.00091.2019 (2019).
193 Chorell, E., Svensson, M. B., Moritz, T. & Antti, H. Physical fitness level is reflected by alterations in the human plasma metabolome. *Molecular BioSystems* **8**, 1187-1196, doi:10.1039/c2mb05428k (2012).
194 Mukherjee, K. *et al.* Whole blood transcriptomics and urinary metabolomics to define adaptive biochemical pathways of high-intensity exercise in 50-60 year old masters athletes. *PLoS One* **9**, e92031, doi:10.1371/journal.pone.0092031 (2014).
195 Daskalaki, E. *et al.* A study of the effects of exercise on the urinary metabolome using normalisation to individual metabolic output. *Metabolites* **5**, 119-139, doi:10.3390/metabo5010119 (2015).
196 Pechlivanis, A. *et al.* (1)H NMR-based metabonomic investigation of the effect of two different exercise sessions on the metabolic fingerprint of human urine. *Journal of Proteome Research* **9**, 6405-6416, doi:10.1021/pr100684t (2010).
197 Hussain, J. N., Mantri, N. & Cohen, M. M. Working Up a Good Sweat - The Challenges of Standardising Sweat Collection for Metabolomics Analysis. *The Clinical Biochemist Reviews* **38**, 13-34 (2017).
198 Harshman, S. W. *et al.* Metabolomic stability of exercise-induced sweat. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **1126-1127**, 121763, doi:10.1016/j.jchromb.2019.121763 (2019).
199 Fiehn, O. *et al.* The metabolomics standards initiative (MSI). *Metabolomics* **3**, 175-178, doi:10.1007/s11306-007-0070-6 (2007).
200 *The metabolomics workbench*, <https://www.metabolomicsworkbench.org/> (
201 Nash, D. W., Palmer, D. E. & Dunn, P. W. *UHPLC retention time and MS/MS mass spectral fragmentation libraries*, <https://more.bham.ac.uk/bamcg/resources/uhplc-retention-time-and-ms-ms-mass-spectral-fragmentation-libraries/> (
202 Norman, B. *et al. Three accurate mass retention time (AMRT) databases generated from IROA Technologies Metabolite Library of Standards by LC-QTOF-MS analysis (accessible via: https://doi.org/10.6084/m9.figshare.c.4378235.v2).* (2019).
203 Horai, H. *et al.* MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry* **45**, 703-714, doi:10.1002/jms.1777 (2010).
204 *MetaSci*, <https://www.metasci.ca/> (
205 Fiehn, O.
206 Weaver, S. R. *et al.* Cerebral Hemodynamic and Neurotrophic Factor Responses Are Dependent on the Type of Exercise. *Front Physiol* **11**, 609935, doi:10.3389/fphys.2020.609935 (2020).
207 Weston, K. S., Wisloff, U. & Coombes, J. S. High-intensity interval training in patients with lifestyle-induced cardiometabolic disease: a systematic review and meta-analysis. *British Journal of Sports Medicine* **48**, 1227-1234, doi:10.1136/bjsports-2013-092576 (2014).
208 Gray, N., Lewis, M. R., Plumb, R. S., Wilson, I. D. & Nicholson, J. K. High-Throughput Microbore UPLC-MS Metabolic Phenotyping of Urine for Large-Scale Epidemiology Studies. *Journal of Proteome Research* **14**, 2714-2721, doi:10.1021/acs.jproteome.5b00203 (2015).

209 Pang, Z. *et al.* MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research* **49**, W388-W396, doi:10.1093/nar/gkab382 (2021).
210 *MetaboAnalyst*, <https://www.metaboanalyst.ca/> (
211 Xia, J., Psychogios, N., Young, N. & Wishart, D. S. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research* **37**, W652-660, doi:10.1093/nar/gkp356 (2009).
212 Abela, L. *et al.* Plasma metabolomics reveals a diagnostic metabolic fingerprint for mitochondrial aconitase (ACO2) deficiency. *PLoS One* **12**, e0176363, doi:10.1371/journal.pone.0176363 (2017).
213 Hao, J. *et al.* Serum Metabolomics Analysis Reveals a Distinct Metabolic Profile of Patients with Primary Biliary Cholangitis. *Scientific Reports* **7**, 784, doi:10.1038/s41598-017-00944-9 (2017).
214 Zukunft, S., Sorgenfrei, M., Prehn, C., Möller, G. & Adamski, J. Targeted Metabolomics of Dried Blood Spot Extracts. *Chromatographia* **76**, 1295-1305, doi:10.1007/s10337-013-2429-3 (2013).
215 Wishart, D. S. *et al.* The human cerebrospinal fluid metabolome. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **871**, 164-173, doi:10.1016/j.jchromb.2008.05.001 (2008).
216 Chen, L. Y., Zhou, L., Chan, E. C. Y., Neo, J. & Beuerman, R. W. Characterization of The Human Tear Metabolome by LC-MS/MS. *Journal of Proteome Research* **10**, 4876-4882, doi:10.1021/pr2004874 (2011).
217 Rusz, M., Rampler, E., Keppler, B. K., Jakupec, M. A. & Koellensperger, G. Single Spheroid Metabolomics: Optimizing Sample Preparation of Three-Dimensional Multicellular Tumor Spheroids. *Metabolites* **9**, 304, doi:10.3390/metabo9120304 (2019).
218 Zhou, B., Xiao, J. F., Tuli, L. & Ressom, H. W. LC-MS-based metabolomics. *Molecular Biosystems* **8**, 470-481, doi:10.1039/c1mb05350g (2012).
219 Wang, X., Sun, H., Zhang, A., Wang, P. & Han, Y. Ultra-performance liquid chromatography coupled to mass spectrometry as a sensitive and powerful technology for metabolomic studies. *Journal of Separation Science* **34**, 3451-3459, doi:10.1002/jssc.201100333 (2011).
220 Wu, N., Liu, Y. & Lee, M. L. Sub-2 microm porous and nonporous particles for fast separation in reversed-phase high performance liquid chromatography. *Journal of Chromatography A* **1131**, 142-150, doi:10.1016/j.chroma.2006.07.042 (2006).
221 Plumb, R. *et al.* Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **18**, 2331-2337, doi:10.1002/rcm.1627 (2004).
222 Chen, J.-H. & Singer, S. in *The Handbook of Metabonomics and Metabolomics* (eds John C. Lindon, Jeremy K. Nicholson, & Elaine Holmes) 149-169 (Elsevier Science B.V., 2007).
223 Gerard Rozing, Maria Serwe, Hans-Georg Weissgerber & Bernd Glatz. in *American Laboratory* (2001).
224 Rozing, G. in *RECENT DEVELOPMENTS IN LC COLUMN TECHNOLOGY* (2003).
225 Hopfgartner, G., Bean, K., Henion, J. & Henry, R. Ion spray mass spectrometric detection for liquid chromatography: A concentration- or a mass-flow-sensitive device? *Journal of Chromatography A* **647**, 51-61, doi:10.1016/0021-9673(93)83323-k (1993).
226 Yanes, O., Tautenhahn, R., Patti, G. J. & Siuzdak, G. Expanding coverage of the metabolome for global metabolite profiling. *Analytical Chemistry* **83**, 2152-2161, doi:10.1021/ac102981k (2011).
227 Wang, J., Westenskow, P. D., Fang, M., Friedlander, M. & Siuzdak, G. Quantitative metabolomics of photoreceptor degeneration and the effects of stem cell-derived retinal pigment epithelium transplantation. *Philosophical Transactions of the Royal Society A* **374**, 20150376, doi:10.1098/rsta.2015.0376 (2016).
228 Rahavendran, S. V. *et al.* Discovery pharmacokinetic studies in mice using serial microsampling, dried blood spots and microbore LC-MS/MS. *Bioanalysis* **4**, 1077-1095, doi:10.4155/bio.12.85 (2012).
229 Lanckmans, K. *et al.* Use of microbore LC-MS/MS for the quantification of oxcarbazepine and its active metabolite in rat brain microdialysis samples. *Journal of Chromatography B:*

Analytical Technologies in the Biomedical and Life Sciences **831**, 205-212, doi:10.1016/j.jchromb.2005.12.003 (2006).

230 Esmans, E. L., Geboes, P., Luyten, Y. & Alderweireldt, F. C. Direct liquid introduction LC/MS microbore experiments for the analysis of nucleoside material present in human urine. *Biological Mass Spectrometry* **12**, 241-245, doi:10.1002/bms.1200120511 (1985).

231 Gray, N. *et al.* Development of a Rapid Microbore Metabolic Profiling Ultraperformance Liquid Chromatography-Mass Spectrometry Approach for High-Throughput Phenotyping Studies. *Analytical Chemistry* **88**, 5742-5751, doi:10.1021/acs.analchem.6b00038 (2016).

232 King, A. M. *et al.* Development of a rapid profiling method for the analysis of polar analytes in urine using HILIC-MS and ion mobility enabled HILIC-MS. *Metabolomics : Official journal of the Metabolomic Society* **15**, 17-17, doi:10.1007/s11306-019-1474-9 (2019).

233 Letertre, M. *et al.* Metabolic Phenotyping Using UPLC–MS and Rapid Microbore UPLC–IM–MS: Determination of the Effect of Different Dietary Regimes on the Urinary Metabolome of the Rat. *Chromatographia* **83**, 853-861, doi:10.1007/s10337-020-03900-4 (2020).

234 t'Kindt, R., Alaerts, G., Vander Heyden, Y., Deforce, D. & Van Bocxlaer, J. Broad-spectrum separations in metabolomics using enhanced polar LC stationary phases: a dedicated evaluation using plant metabolites. *Journal of Separation Science* **30**, 2002-2011, doi:10.1002/jssc.200700077 (2007).

235 Tsuji, F. & Kawazu, K. Biomarker Identification of Tear Fluid. *Metabolomics* **2** (2012).

236 Ahamad, S. R., Raish, M., Yaqoob, S. H., Khan, A. & Shakeel, F. Metabolomics and Trace Element Analysis of Camel Tear by GC-MS and ICP-MS. *Biological Trace Element Research* **177**, 251-257, doi:10.1007/s12011-016-0889-7 (2017).

237 Cicalini, I. *et al.* Integrated Lipidomics and Metabolomics Analysis of Tears in Multiple Sclerosis: An Insight into Diagnostic Potential of Lacrimal Fluid. *International Journal of Molecular Sciences* **20**, doi:10.3390/ijms20061265 (2019).

238 Urbanski, G. *et al.* Tear metabolomics highlights new potential biomarkers for differentiating between Sjogren's syndrome and other causes of dry eye. *The Ocular Surface* **22**, 110-116, doi:10.1016/j.jtos.2021.07.006 (2021).

239 Frainay, C. *et al.* Mind the Gap: Mapping Mass Spectral Databases in Genome-Scale Metabolic Networks Reveals Poorly Covered Areas. *Metabolites* **8**, 51, doi:10.3390/metabo8030051 (2018).

240 Creek, D. J. *et al.* Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction. *Analytical Chemistry* **83**, 8703-8710, doi:10.1021/ac2021823 (2011).

241 Wolfer, A. M. *et al.* UPLC–MS retention time prediction: a machine learning approach to metabolite identification in untargeted profiling. *Metabolomics* **12**, 8, doi:10.1007/s11306-015-0888-2 (2015).

242 Hagiwara, T. *et al.* HPLC Retention time prediction for metabolome analysi. *Bioinformation* **5**, 255-258, doi:10.6026/97320630005255 (2010).

243 Falchi, F. *et al.* Kernel-Based, Partial Least Squares Quantitative Structure-Retention Relationship Model for UPLC Retention Time Prediction: A Useful Tool for Metabolite Identification. *Analytical Chemistry* **88**, 9510-9517, doi:10.1021/acs.analchem.6b02075 (2016).

244 Bach, E., Szedmak, S., Brouard, C., Bocker, S. & Rousu, J. Liquid-chromatography retention order prediction for metabolite identification. *Bioinformatics* **34**, i875-i883, doi:10.1093/bioinformatics/bty590 (2018).

245 Gorynski, K. *et al.* Quantitative structure-retention relationships models for prediction of high performance liquid chromatography retention time of small molecules: endogenous metabolites and banned compounds. *Analytica Chimica Acta* **797**, 13-19, doi:10.1016/j.aca.2013.08.025 (2013).

246 Giese, S. H., Sinn, L. R., Wegner, F. & Rappsilber, J. Retention time prediction using neural networks increases identifications in crosslinking mass spectrometry. *Nature Communications* **12**, 3237, doi:10.1038/s41467-021-23441-0 (2021).

247 Vankova, Z. *et al.* Retention dependences support highly confident identification of lipid species in human plasma by reversed-phase UHPLC/MS. *Analytical and Bioanalytical Chemistry* **414**, 319-331, doi:10.1007/s00216-021-03492-4 (2022).
248 Beger, R. D. *et al.* Metabolomics enables precision medicine: "A White Paper, Community Perspective". *Metabolomics* **12**, 149, doi:10.1007/s11306-016-1094-6 (2016).
249 Wang, X., West, J. A., Murray, A. J. & Griffin, J. L. Comprehensive Metabolic Profiling of Age-Related Mitochondrial Dysfunction in the High-Fat-Fed ob/ob Mouse Heart. *Journal of Proteome Research* **14**, 2849-2862, doi:10.1021/acs.jproteome.5b00128 (2015).
250 Ramautar, R., Berger, R., van der Greef, J. & Hankemeier, T. Human metabolomics: strategies to understand biology. *Current Opinion in Chemical Biology* **17**, 841-846, doi:10.1016/j.cbpa.2013.06.015 (2013).
251 Hassan-Smith, G., Wallace, G. R., Douglas, M. R. & Sinclair, A. J. The role of metabolomics in neurological disease. *Journal of Neuroimmunology* **248**, 48-52, doi:10.1016/j.jneuroim.2012.01.009 (2012).
252 Castelli, F. A. *et al.* Metabolomics for personalized medicine: the input of analytical chemistry from biomarker discovery to point-of-care tests. *Analytical and Bioanalytical Chemistry* **414**, 759-789, doi:10.1007/s00216-021-03586-z (2022).
253 Wishart, D. S. Emerging applications of metabolomics in drug discovery and precision medicine. *Nature Reviews Drug Discovery* **15**, 473-484, doi:10.1038/nrd.2016.32 (2016).
254 McCartney, A. *et al.* Metabolomics in breast cancer: A decade in review. *Cancer Treatment Reviews* **67**, 88-96, doi:10.1016/j.ctrv.2018.04.012 (2018).
255 Chorell, E., Moritz, T., Branth, S., Antti, H. & Svensson, M. B. Predictive metabolomics evaluation of nutrition-modulated metabolic stress responses in human blood serum during the early recovery phase of strenuous physical exercise. *Journal of Proteome Research* **8**, 2966-2977, doi:10.1021/pr900081q (2009).
256 Gabriel, B. M. & Zierath, J. R. The Limits of Exercise Physiology: From Performance to Health. *Cell Metabolism* **25**, 1000-1011, doi:10.1016/j.cmet.2017.04.018 (2017).
257 Fletcher, G. F. *et al.* Promoting Physical Activity and Exercise: JACC Health Promotion Series. *Journal of the American College of Cardiology* **72**, 1622-1639, doi:10.1016/j.jacc.2018.08.2141 (2018).
258 Myers, J. Cardiology patient pages. Exercise and cardiovascular health. *Circulation* **107**, e2-5, doi:10.1161/01.cir.0000048890.59383.8d (2003).
259 Anderson, E. & Durstine, J. L. Physical activity, exercise, and chronic diseases: A brief review. *Sports Medicine and Health Science* **1**, 3-10, doi:10.1016/j.smhs.2019.08.006 (2019).
260 Fentem, P. H. ABC of sports medicine. Benefits of exercise in health and disease. *BMJ* **308**, 1291-1295, doi:10.1136/bmj.308.6939.1291 (1994).
261 Wolfe, R. R. in *Skeletal Muscle Metabolism in Exercise and Diabetes* (eds Erik A. Richter, Bente Kiens, Henrik Galbo, & Bengt Saltin) 147-156 (Springer US, 1998).
262 Mangine, G. T. *et al.* Exercise-Induced Hormone Elevations Are Related to Muscle Growth. *The Journal of Strength and Conditioning Research* **31**, 45-53, doi:10.1519/JSC.0000000000001491 (2017).
263 Pasiakos, S. M. Exercise and amino acid anabolic cell signaling and the regulation of skeletal muscle mass. *Nutrients* **4**, 740-758, doi:10.3390/nu4070740 (2012).
264 Zielinski, J., Slominska, E. M., Krol-Zielinska, M., Krasinski, Z. & Kusy, K. Purine metabolism in sprint- vs endurance-trained athletes aged 2090 years. *Scientific Reports* **9**, 12075, doi:10.1038/s41598-019-48633-z (2019).
265 Melancon, M. O., Lorrain, D. & Dionne, I. J. Exercise increases tryptophan availability to the brain in older men age 57-70 years. *Medicine & Science in Sports & Exercise* **44**, 881-887, doi:10.1249/MSS.0b013e31823ede8e (2012).
266 Nitzsche, N., Lenz, J. C., Voronoi, P. & Schulz, H. Adaption of Maximal Glycolysis Rate after Resistance Exercise with Different Volume Load. *Sports Medicine International* **4**, E39-E44, doi:10.1055/a-1146-4236 (2020).
267 Wolfe, R. R., Klein, S., Carraro, F. & Weber, J. M. Role of triglyceride-fatty acid cycle in controlling fat metabolism in humans during and after exercise. *American Journal of Physiology* **258**, E382-389, doi:10.1152/ajpendo.1990.258.2.E382 (1990).

268 Zheng, F. & Cai, Y. Concurrent exercise improves insulin resistance and nonalcoholic fatty liver disease by upregulating PPAR-gamma and genes involved in the beta-oxidation of fatty acids in ApoE-KO mice fed a high-fat diet. *Lipids in Health and Disease* **18**, 6, doi:10.1186/s12944-018-0933-z (2019).

269 Mika, A., Macaluso, F., Barone, R., Di Felice, V. & Sledzinski, T. Effect of Exercise on Fatty Acid Metabolism and Adipokine Secretion in Adipose Tissue. *Front Physiol* **10**, 26, doi:10.3389/fphys.2019.00026 (2019).

270 Mourtzakis, M. & Graham, T. E. Glutamate ingestion and its effects at rest and during exercise in humans. *Journal of Applied Physiology (1985)* **93**, 1251-1259, doi:10.1152/japplphysiol.00111.2002 (2002).

271 Williams, C. J. *et al.* A Multi-Center Comparison of O2peak Trainability Between Interval Training and Moderate Intensity Continuous Training. *Frontiers in Physiology* **10**, 19 (2019).

272 Little, J. P., Safdar, A., Bishop, D., Tarnopolsky, M. A. & Gibala, M. J. An acute bout of high-intensity interval training increases the nuclear abundance of PGC-1alpha and activates mitochondrial biogenesis in human skeletal muscle. *The American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **300**, R1303-1310, doi:10.1152/ajpregu.00538.2010 (2011).

273 Little, J. P., Safdar, A., Wilkin, G. P., Tarnopolsky, M. A. & Gibala, M. J. A practical model of low-volume high-intensity interval training induces mitochondrial biogenesis in human skeletal muscle: potential mechanisms. *The Journal of Physiology* **588**, 1011-1022, doi:10.1113/jphysiol.2009.181743 (2010).

274 Boutcher, S. H. High-intensity intermittent exercise and fat loss. *Journal of Obesity* **2011**, 868305, doi:10.1155/2011/868305 (2011).

275 Doleman, B. Short-Term, Equipment-Free High Intensity Interval Training Elicits Significant Improvements in Cardiorespiratory Fitness Irrespective of Supervision in Early Adulthood. (2021).

276 Perry, C. G., Heigenhauser, G. J., Bonen, A. & Spriet, L. L. High-intensity aerobic interval training increases fat and carbohydrate metabolic capacities in human skeletal muscle. *Applied Physiology, Nutrition, and Metabolism* **33**, 1112-1123, doi:10.1139/H08-097 (2008).

277 Mallol, M. *et al.* Physiological Response Differences between Run and Cycle High Intensity Interval Training Program in Recreational Middle Age Female Runners. *Journal of Sports Science and Medicine* **19**, 508-516 (2020).

278 Sun, L. *et al.* Effects of high-intensity interval training on adipose tissue lipolysis, inflammation, and metabolomics in aged rats. *Pflugers Arch* **472**, 245-258, doi:10.1007/s00424-020-02351-y (2020).

279 Liu, X. & Wang, G. The Effect of High-Intensity Interval Training on Physical Parameters, Metabolomic Indexes and Serum Ficolin-3 Levels in Patients with Prediabetes and Type 2 Diabetes. *Experimental and Clinical Endocrinology & Diabetes* **129**, 740-749, doi:10.1055/a-1027-6511 (2021).

280 Kuehnbaum, N. L., Gillen, J. B., Gibala, M. J. & Britz-McKibbin, P. Personalized metabolomics for predicting glucose tolerance changes in sedentary women after high-intensity interval training. *Scientific Reports* **4**, 6166, doi:10.1038/srep06166 (2014).

281 Jimenez-Maldonado, A., Garcia-Suarez, P. C., Renteria, I., Moncada-Jimenez, J. & Plaisance, E. P. Impact of high-intensity interval training and sprint interval training on peripheral markers of glycemic control in metabolic syndrome and type 2 diabetes. *Biochim Biophys Acta Mol Basis Dis* **1866**, 165820, doi:10.1016/j.bbadis.2020.165820 (2020).

282 Ryan, B. J. *et al.* Moderate-Intensity Exercise and High-Intensity Interval Training Affect Insulin Sensitivity Similarly in Obese Adults. *The Journal of Clinical Endocrinology and Metabolism* **105**, e2941-e2959, doi:10.1210/clinem/dgaa345 (2020).

283 Wang, N., Liu, Y., Ma, Y. & Wen, D. High-intensity interval versus moderate-intensity continuous training: Superior metabolic benefits in diet-induced obesity mice. *Life Sciences* **191**, 122-131, doi:10.1016/j.lfs.2017.08.023 (2017).

284 Motiani, K. K. *et al.* Effects of short-term sprint interval and moderate-intensity continuous training on liver fat content, lipoprotein profile, and substrate uptake: a randomized trial.

Journal of Applied Physiology (1985) **126**, 1756-1768, doi:10.1152/japplphysiol.00900.2018 (2019).

285 Cocks, M. *et al.* Sprint interval and moderate-intensity continuous training have equal benefits on aerobic capacity, insulin sensitivity, muscle capillarisation and endothelial eNOS/NAD(P)Hoxidase protein ratio in obese men. *The Journal of Physiology* **594**, 2307-2321, doi:10.1113/jphysiol.2014.285254 (2016).

286 Wewege, M., van den Berg, R., Ward, R. E. & Keech, A. The effects of high-intensity interval training vs. moderate-intensity continuous training on body composition in overweight and obese adults: a systematic review and meta-analysis. *Obesity Reviews* **18**, 635-646, doi:10.1111/obr.12532 (2017).

287 Langan, S. P. & Grosicki, G. J. Exercise Is Medicine…and the Dose Matters. *Frontiers in Physiology* **12**, 664 (2021).

288 Goto, C. *et al.* Effect of different intensities of exercise on endothelium-dependent vasodilation in humans: role of endothelium-dependent nitric oxide and oxidative stress. *Circulation* **108**, 530-535, doi:10.1161/01.CIR.0000080893.55729.28 (2003).

289 Keating, S. E., Johnson, N. A., Mielke, G. I. & Coombes, J. S. A systematic review and meta-analysis of interval training versus moderate-intensity continuous training on body adiposity. *Obesity Reviews* **18**, 943-964, doi:10.1111/obr.12536 (2017).

290 Qiu, Y.-Q. in *Encyclopedia of Systems Biology* (eds Werner Dubitzky, Olaf Wolkenhauer, Kwang-Hyun Cho, & Hiroki Yokota) Ch. Chapter 472, 1068-1069 (Springer New York, 2013).

291 Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Science* **28**, 1947-1951, doi:10.1002/pro.3715 (2019).

292 Lundsgaard, A. M., Fritzen, A. M. & Kiens, B. The Importance of Fatty Acids as Nutrients during Post-Exercise Recovery. *Nutrients* **12**, doi:10.3390/nu12020280 (2020).

293 Nikolaidis, M. G. & Mougios, V. Effects of exercise on the fatty-acid composition of blood and tissue lipids. *Sports Medicine* **34**, 1051-1076, doi:10.2165/00007256-200434150-00004 (2004).

294 Purdom, T., Kravitz, L., Dokladny, K. & Mermier, C. Understanding the factors that effect maximal fat oxidation. *Journal of the International Society of Sports Nutrition* **15**, 3, doi:10.1186/s12970-018-0207-1 (2018).

295 Hargreaves, M. & Spriet, L. L. Skeletal muscle energy metabolism during exercise. *Nature Metabolism* **2**, 817-828, doi:10.1038/s42255-020-0251-4 (2020).

296 Vislocky, L. M., Gaine, P. C., Pikosky, M. A., Martin, W. F. & Rodriguez, N. R. Gender impacts the post-exercise substrate and endocrine response in trained runners. *Journal of the International Society of Sports Nutrition* **5**, 7, doi:10.1186/1550-2783-5-7 (2008).

297 Wallis, G. A., Dawson, R., Achten, J., Webber, J. & Jeukendrup, A. E. Metabolic response to carbohydrate ingestion during exercise in males and females. *The American Journal of Physiology-Endocrinology and Metabolism* **290**, E708-715, doi:10.1152/ajpendo.00357.2005 (2006).

298 Wu, B. N. & O'Sullivan, A. J. Sex differences in energy metabolism need to be considered with lifestyle modifications in humans. *Journal of Nutrition and Metabolism* **2011**, 391809, doi:10.1155/2011/391809 (2011).

299 Rosa-Caldwell, M. E. & Greene, N. P. Muscle metabolism and atrophy: let's talk about sex. *Biology of Sex Differences* **10**, 43, doi:10.1186/s13293-019-0257-3 (2019).

300 Dearden, L., Bouret, S. G. & Ozanne, S. E. Sex and gender differences in developmental programming of metabolism. *Molecular Metabolism* **15**, 8-19, doi:10.1016/j.molmet.2018.04.007 (2018).

301 Chowen, J. A., Freire-Regatillo, A. & Argente, J. Neurobiological characteristics underlying metabolic differences between males and females. *Progress in Neurobiology* **176**, 18-32, doi:10.1016/j.pneurobio.2018.09.001 (2019).

302 Junge, N., Jorgensen, R., Flouris, A. D. & Nybo, L. Prolonged self-paced exercise in the heat - environmental factors affecting performance. *Temperature (Austin)* **3**, 539-548, doi:10.1080/23328940.2016.1216257 (2016).

303 American College of Sports, M. *et al.* American College of Sports Medicine position stand. Exertional heat illness during training and competition. *Medicine & Science in Sports & Exercise* **39**, 556-572, doi:10.1249/MSS.0b013e31802fa199 (2007).

304 Leon, L. R. & Helwig, B. G. Heat stroke: role of the systemic inflammatory response. *Journal of Applied Physiology (1985)* **109**, 1980-1988, doi:10.1152/japplphysiol.00301.2010 (2010).

305 Bezruchka, S. A. in *The Travel and Tropical Medicine Manual* (eds Elaine C. Jong & Christopher Sanford) 132-151 (W.B. Saunders, 2008).

306 Periard, J. D., Racinais, S. & Sawka, M. N. Adaptations and mechanisms of human heat acclimation: Applications for competitive athletes and sports. *Scandinavian Journal of Medicine & Science in Sports* **25 Suppl 1**, 20-38, doi:10.1111/sms.12408 (2015).

307 Taylor, N. A. Human heat adaptation. *Comprehensive Physiology* **4**, 325-365, doi:10.1002/cphy.c130022 (2014).

308 Kenny, G. P., Wilson, T. E., Flouris, A. D. & Fujii, N. in *Handbook of Clinical Neurology* Vol. 157 (ed Andrej A. Romanovsky) 505-529 (Elsevier, 2018).

309 Garrett, A. T., Creasy, R., Rehrer, N. J., Patterson, M. J. & Cotter, J. D. Effectiveness of short-term heat acclimation for highly trained athletes. *European Journal of Applied Physiology* **112**, 1827-1837, doi:10.1007/s00421-011-2153-3 (2012).

310 Casadio, J. R., Kilding, A. E., Cotter, J. D. & Laursen, P. B. From Lab to Real World: Heat Acclimation Considerations for Elite Athletes. *Sports Medicine* **47**, 1467-1476, doi:10.1007/s40279-016-0668-9 (2017).

311 Lorenzo, S., Halliwill, J. R., Sawka, M. N. & Minson, C. T. Heat acclimation improves exercise performance. *Journal of Applied Physiology (1985)* **109**, 1140-1147, doi:10.1152/japplphysiol.00495.2010 (2010).

312 Parsons, I. T., Stacey, M. J. & Woods, D. R. Heat Adaptation in Military Personnel: Mitigating Risk, Maximizing Performance. *Frontiers in Physiology* **10**, doi:10.3389/fphys.2019.01485 (2019).

313 Racinais, S. *et al.* Consensus recommendations on training and competing in the heat. *British Journal of Sports Medicine* **49**, 1164-1173, doi:10.1136/bjsports-2015-094915 (2015).

314 Tyler, C. J., Reeve, T., Hodges, G. J. & Cheung, S. S. The Effects of Heat Adaptation on Physiology, Perception and Exercise Performance in the Heat: A Meta-Analysis. *Sports Medicine* **46**, 1699-1724, doi:10.1007/s40279-016-0538-5 (2016).

315 Pandolf, K. B. Time course of heat acclimation and its decay. *International Journal of Sports Medicine* **19 Suppl 2**, S157-160, doi:10.1055/s-2007-971985 (1998).

316 Neal, R. A., Massey, H. C., Tipton, M. J., Young, J. S. & Corbett, J. Effect of Permissive Dehydration on Induction and Decay of Heat Acclimation, and Temperate Exercise Performance. *Front Physiol* **7**, 564, doi:10.3389/fphys.2016.00564 (2016).

317 Fox, R. H., Goldsmith, R., Kidd, D. J. & Lewis, H. E. Acclimatization to heat in man by controlled elevation of body temperature. *The Journal of Physiology* **166**, 530-547, doi:10.1113/jphysiol.1963.sp007121 (1963).

318 Lind, A. R. & Bass, D. E. Optimal exposure time for development of acclimatization to heat. *Federation Proceedings* **22**, 704-708 (1963).

319 Ely, B. R., Lovering, A. T., Horowitz, M. & Minson, C. T. Heat acclimation and cross tolerance to hypoxia: Bridging the gap between cellular and systemic responses. *Temperature (Austin)* **1**, 107-114, doi:10.4161/temp.29800 (2014).

320 Yacobi, A., Stern Bach, Y. & Horowitz, M. The protective effect of heat acclimation from hypoxic damage in the brain involves changes in the expression of glutamate receptors. *Temperature (Austin)* **1**, 57-65, doi:10.4161/temp.29719 (2014).

321 Heled, Y. *et al.* Heat acclimation and performance in hypoxic conditions. *Aerospace Medicine and Human Performance* **83**, 649-653, doi:10.3357/asem.3241.2012 (2012).

322 Rendell, R. A. *et al.* Effects of 10 days of separate heat and hypoxic exposure on heat acclimation and temperate exercise performance. *The American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **313**, R191-R201, doi:10.1152/ajpregu.00103.2017 (2017).

323 Lee, B. J., Miller, A., James, R. S. & Thake, C. D. Cross Acclimation between Heat and Hypoxia: Heat Acclimation Improves Cellular Tolerance and Exercise Performance in Acute Normobaric Hypoxia. *Frontiers in Physiology* **7**, 78 (2016).

324 Gordon, K. *et al.* Seven days of cold acclimation substantially reduces shivering intensity and increases nonshivering thermogenesis in adult humans. *Journal of Applied Physiology (1985)* **126**, 1598-1606, doi:10.1152/japplphysiol.01133.2018 (2019).

325 MacMillan, H. A. *et al.* Cold acclimation wholly reorganizes the Drosophila melanogaster transcriptome and metabolome. *Scientific Reports* **6**, 28999, doi:10.1038/srep28999 (2016).

326 Febbraio, M. A. Alterations in energy metabolism during exercise and heat stress. *Sports Medicine* **31**, 47-59, doi:10.2165/00007256-200131010-00004 (2001).

327 Havenith, G., Coenen, J. M., Kistemaker, L. & Kenney, W. L. Relevance of individual characteristics for human heat stress response is dependent on exercise intensity and climate type. *European Journal of Applied occupational Physiology* **77**, 231-241, doi:10.1007/s004210050327 (1998).

328 Febbraio, M. A., Snow, R. J., Stathis, C. G., Hargreaves, M. & Carey, M. F. Effect of heat stress on muscle energy metabolism during exercise. *Journal of Applied Physiology (1985)* **77**, 2827-2831, doi:10.1152/jappl.1994.77.6.2827 (1994).

329 Nielsen, B., Savard, G., Richter, E. A., Hargreaves, M. & Saltin, B. Muscle blood flow and muscle metabolism during exercise and heat stress. *Journal of Applied Physiology (1985)* **69**, 1040-1046, doi:10.1152/jappl.1990.69.3.1040 (1990).

330 Hargreaves, M., Angus, D., Howlett, K., Conus, N. M. & Febbraio, M. Effect of heat stress on glucose kinetics during exercise. *Journal of Applied Physiology (1985)* **81**, 1594-1597, doi:10.1152/jappl.1996.81.4.1594 (1996).

331 Febbraio, M. A., Carey, M. F., Snow, R. J., Stathis, C. G. & Hargreaves, M. Influence of elevated muscle temperature on metabolism during intense, dynamic exercise. *American Journal of Physiology* **271**, R1251-1255, doi:10.1152/ajpregu.1996.271.5.R1251 (1996).

332 Knochel, J. P., Dotin, L. N. & Hamburger, R. J. Heat stress, exercise, and muscle injury: effects on urate metabolism and renal function. *Annals of Internal Medicine* **81**, 321-328, doi:10.7326/0003-4819-81-3-321 (1974).

333 Starkie, R. L., Hargreaves, M., Rolland, J. & Febbraio, M. A. Heat stress, cytokines, and the immune response to exercise. *Brain, Behavior, and Immunity* **19**, 404-412, doi:10.1016/j.bbi.2005.03.005 (2005).

334 Lunt, H. C., Barwood, M. J., Corbett, J. & Tipton, M. J. 'Cross-adaptation': habituation to short repeated cold-water immersions affects the response to acute hypoxia in humans. *The Journal of Physiology* **588**, 3605-3613, doi:10.1113/jphysiol.2010.193458 (2010).

335 Molkov, Y. I. *et al.* in *Progress in Brain Research* Vol. 212 (eds Gert Holstege, Caroline M. Beers, & Hari H. Subramanian) 1-23 (Elsevier, 2014).

336 Manninen, P. H. & Unger, Z. M. in *Complications in Neuroanesthesia* (ed Hemanshu Prabhakar) 169-180 (Academic Press, 2016).

337 Briere, J. J. *et al.* Succinate dehydrogenase deficiency in human. *Cellular and Molecular Life Sciences CMLS* **62**, 2317-2324, doi:10.1007/s00018-005-5237-6 (2005).

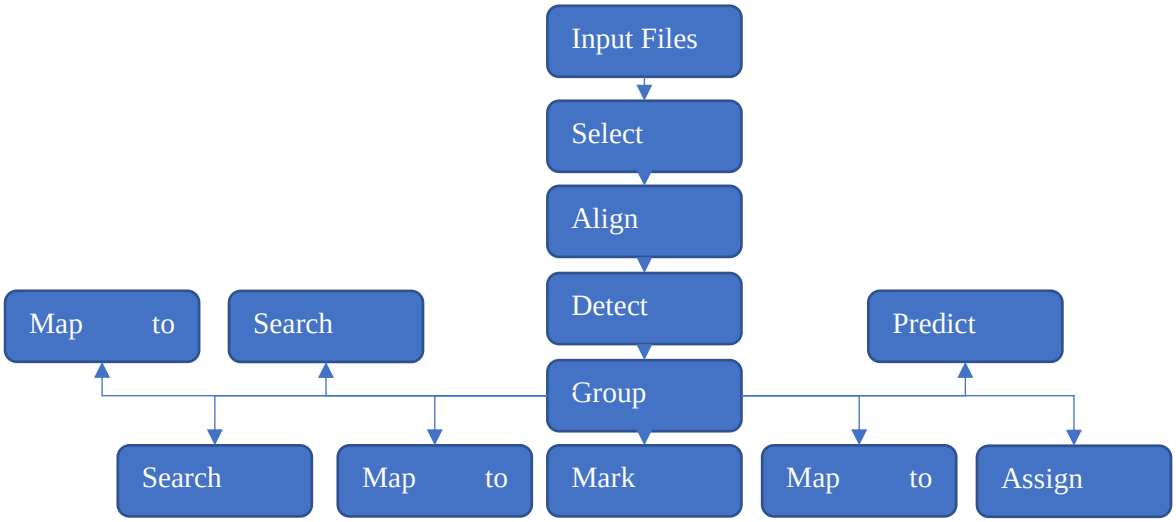
338 Ye, Y. *et al.* Global metabolomic responses of Escherichia coli to heat stress. *Journal of Proteome Research* **11**, 2559-2566, doi:10.1021/pr3000128 (2012).

339 Rabi, T. & Cassuto, Y. Metabolic adaptations in brown adipose tissue of the hamster in extreme ambient temperatures. *American Journal of Physiology* **231**, 153-160, doi:10.1152/ajplegacy.1976.231.1.153 (1976).

340 Hazel, J. R. The effect of temperature acclimation upon succinic dehydrogenase activity from the epaxial muscle of the common goldfish (Carassius auratus L)—I. Properties of the enzyme and the effect of lipid extraction. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **43**, 837-861, doi:10.1016/0305-0491(72)90230-1 (1972).

9 APPENDIX

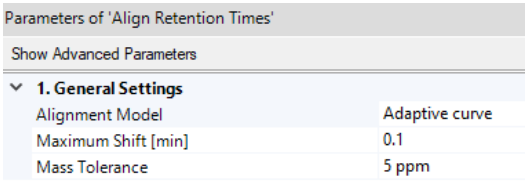
A-1: Compound 3.0 processing workflow



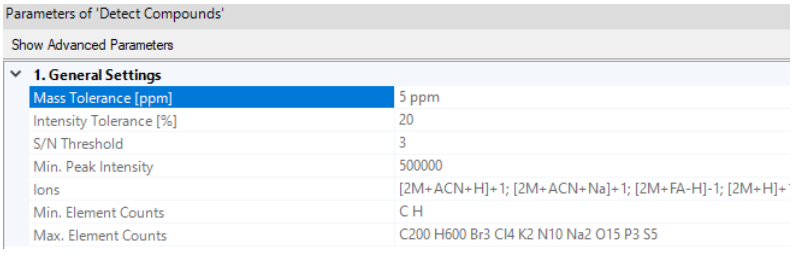
A-2: Compound 3.0 processing workflow -Select Spectra node parameters

Parameters of 'Select Spectra'	
Show Advanced Parameters	
1. General Settings	
Precursor Selection	Use MS(n - 1) Precursor
Provide Profile Spectra	Automatic
2. Spectrum Properties Filter	
Lower RT Limit	0
Upper RT Limit	0
First Scan	0
Last Scan	0
Ignore Specified Scans	
Lowest Charge State	0
Highest Charge State	0
Min. Precursor Mass	100 Da
Max. Precursor Mass	5000 Da
Total Intensity Threshold	0
Minimum Peak Count	1
3. Scan Event Filters	
Mass Analyzer	(Not specified)
MS Order	Any
Activation Type	(Not specified)
Min. Collision Energy	0
Max. Collision Energy	1000
Scan Type	Any
Polarity Mode	(Not specified)
4. Peak Filters	
S/N Threshold (FT-only)	1.5
5. Replacements for Unrecognized Properties	
Unrecognized Charge Replacements	1
Unrecognized Mass Analyzer Replacements	ITMS
Unrecognized MS Order Replacements	MS2
Unrecognized Activation Type Replacements	CID
Unrecognized Polarity Replacements	+
Unrecognized MS Resolution@200 Replacemer	60000
Unrecognized MSn Resolution@200 Replacem	30000

A-3: Compound 3.0 processing workflow -Align Retention Times node parameters



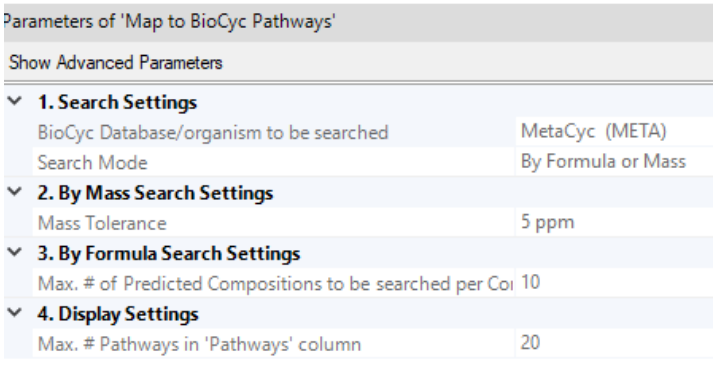
A-4: Compound 3.0 processing workflow -Detect Compounds node parameters



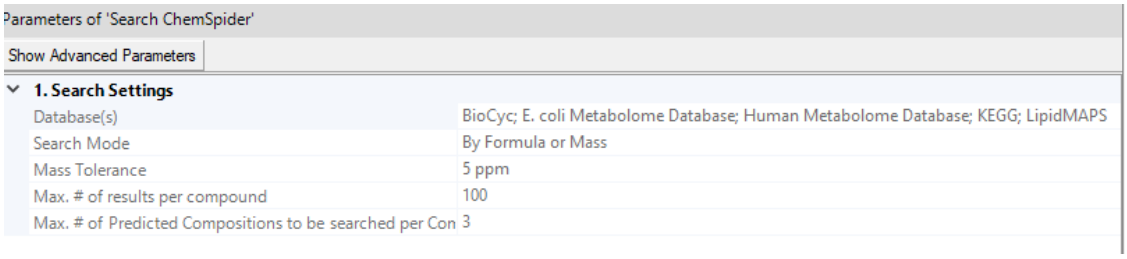
A-5: Compound 3.0 processing workflow- Group Compounds node Parameters



Input Files



A-7: Compound 3.0 processing workflow Search ChemSpider node Parameters



Parameters of 'Mark Background Compounds'

Show Advanced Parameters

1. General Settings	
Max. Sample/Blank	20
Max. Blank/Sample	0
Hide Background	True

Parameters of 'Map to Metabolika Pathways'

Show Advanced Parameters

1. Search Settings	
Metabolika Pathways	\(3R)-linalool biosynthesis.metabolika|\2-nitrobenzoate degradation I.metabolika|\2-oxob
Search Mode	By Formula or Mass
2. By Mass Search Settings	
Mass Tolerance	5 ppm
3. By Formula Search Settings	
Max. # of Predicted Compositions to be searched per Con	3
4. Display Settings	
Max. # Pathways in 'Pathways' column	20

Parameters of 'Map to KEGG Pathways'

Show Advanced Parameters

1. Search Settings	
Search Mode	By Formula or Mass
2. By Mass Search Settings	
Mass Tolerance	5 ppm
3. By Formula Search Settings	
Max. # of Predicted Compositions to be searched per Con	3
4. Display Settings	
Max. # Pathways in 'Pathways' column	20

A-11: Compound 3.0 processing workflow - Assign Compound Annotations node Parameters

Parameters of 'Assign Compound Annotations'

Show Advanced Parameters

1. General Settings	
Mass Tolerance	5 ppm
2. Data Sources	
Data Source #1	Predicted Compositions
Data Source #2	ChemSpider Search
Data Source #3	BioCyc Search
Data Source #4	
Data Source #5	
Data Source #6	
Data Source #7	
3. Scoring Rules	
Use mzLogic	True
Use Spectral Distance	True
SFit Threshold	20
SFit Range	20

A-12: Compound 3.0 processing workflow -Predict Compositions node Parameters

Parameters of 'Predict Compositions'

Show Advanced Parameters

1. Prediction Settings	
Mass Tolerance	5 ppm
Min. Element Counts	C H
Max. Element Counts	C200 H600 Br3 Cl4 N10 O18 P3 S5
Min. RDBE	0
Max. RDBE	40
Min. H/C	0.1
Max. H/C	4
Max. # Candidates	10
2. Pattern Matching	
Intensity Tolerance [%]	20
Intensity Threshold [%]	0.1
S/N Threshold	3
Use Dynamic Recalibration	True
3. Fragments Matching	
Use Fragments Matching	True
Mass Tolerance	5 ppm
S/N Threshold	3

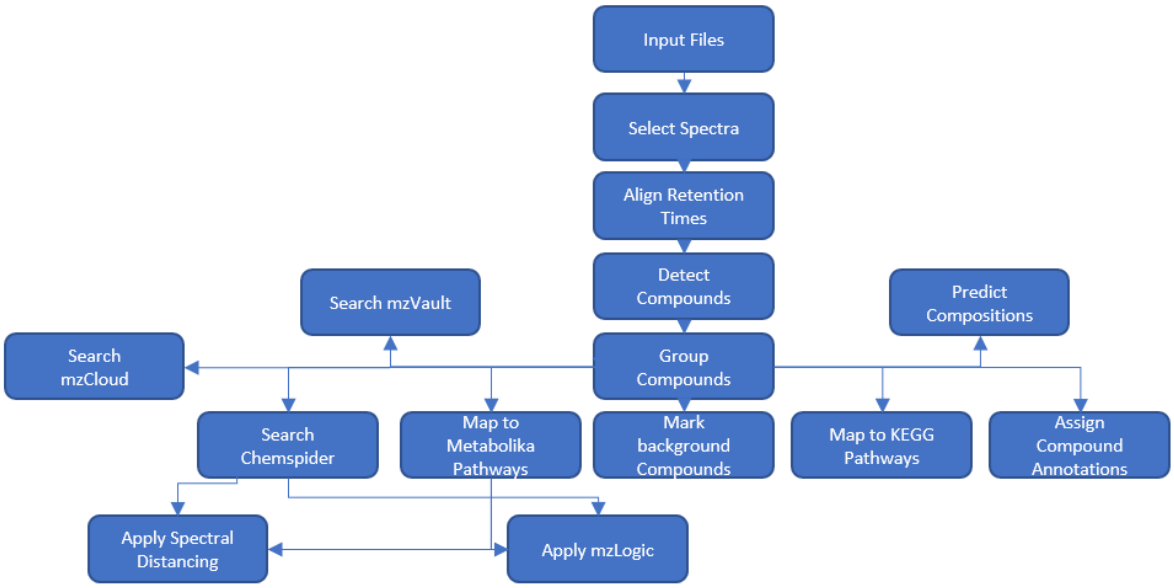
A-13: Compound 3.0 processing workflow -Search mzVault node Parameters

Parameters of 'Search mzVault'

Show Advanced Parameters

1. Search Settings	
mzVault Library	\20190513_MetaSci_POS_Will.db
Compound Classes	All
Match Ion Activation Type	True
Match Ion Activation Energy	Match with Tolerance
Ion Activation Energy Tolerance	20
Match Ionization Method	True
Apply Intensity Threshold	True
Precursor Mass Tolerance	10 ppm
Match Analyzer Type	True
Search Algorithm	HighChem HighRes
Match Factor Threshold	50
RT Tolerance [min]	2
Use Retention Time	False

B-1: Compound 3.1 processing workflow



B-2: Compound 3.1 processing workflow - Select Spectra node parameters

Parameters of 'Select Spectra'

Show Advanced Parameters

1. General Settings	
Precursor Selection	Use MS(n - 1) Precursor
Provide Profile Spectra	Automatic
2. Spectrum Properties Filter	
Lower RT Limit	0
Upper RT Limit	0
First Scan	0
Last Scan	0
Ignore Specified Scans	
Lowest Charge State	0
Highest Charge State	0
Min. Precursor Mass	0 Da
Max. Precursor Mass	5000 Da
Total Intensity Threshold	0
Minimum Peak Count	1
3. Scan Event Filters	
Mass Analyzer	(Not specified)
MS Order	Any
Activation Type	(Not specified)
Min. Collision Energy	0
Max. Collision Energy	1000
Scan Type	Any
Polarity Mode	(Not specified)
4. Peak Filters	
S/N Threshold (FT-only)	1.5
5. Replacements for Unrecognized Properties	
Unrecognized Charge Replacements	1
Unrecognized Mass Analyzer Replacements	ITMS
Unrecognized MS Order Replacements	MS2
Unrecognized Activation Type Replacements	CID
Unrecognized Polarity Replacements	+
Unrecognized MS Resolution@200 Replacements	60000
Unrecognized MSn Resolution@200 Replacements	30000

B-3: Compound 3.1 processing workflow-Align Retention Times node parameters

Parameters of 'Align Retention Times'

Show Advanced Parameters

1. General Settings	
Alignment Model	Adaptive curve
Maximum Shift [mir	0.2
Mass Tolerance	5 ppm

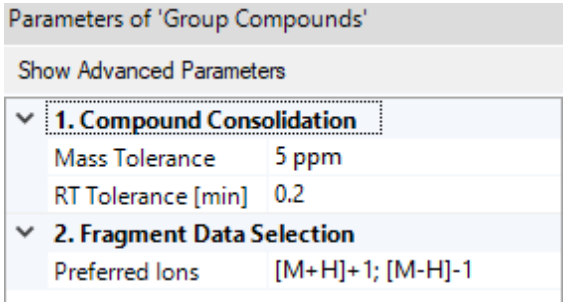
B-4: Compound 3.1 processing workflow -Detect Compounds node Parameters

Parameters of 'Detect Compounds'

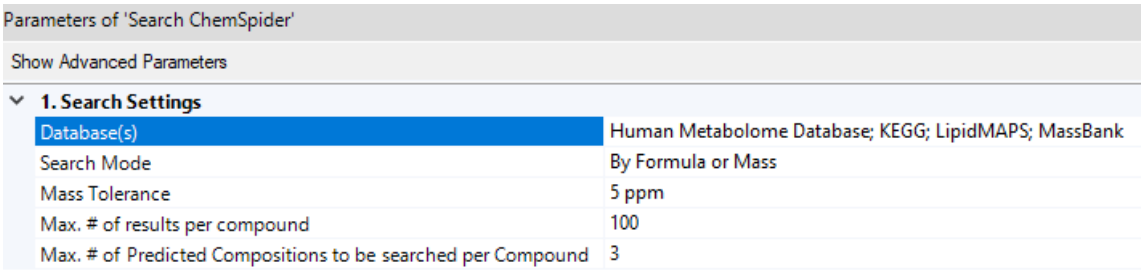
Show Advanced Parameters

1. General Settings	
Mass Tolerance [ppm]	5 ppm
Intensity Tolerance [%]	30
S/N Threshold	3
Min. Peak Intensity	5000
Ions	[2M+ACN+H]+1; [2M+ACN+Na]+1; [2M+FA-H]-1; [2M+H]+1; [2M+K]+1;
Min. Element Counts	C H
Max. Element Counts	C90 H190 Br3 Cl4 K2 N10 Na2 O18 P3 S5

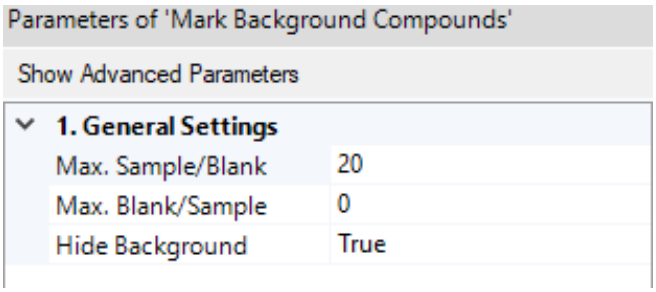
B-5: Compound 3.1 processing workflow -Group Compounds node Parameters



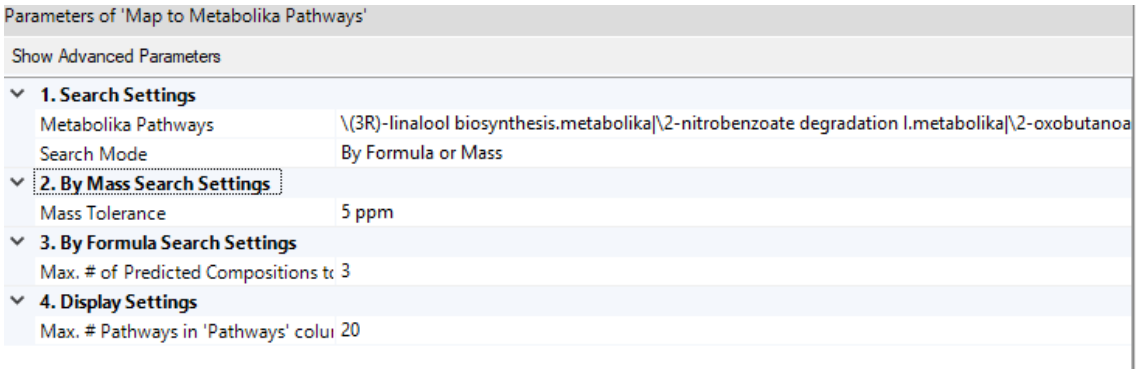
B-6: Compound 3.1 processing workflow- search ChemSpider node Parameters



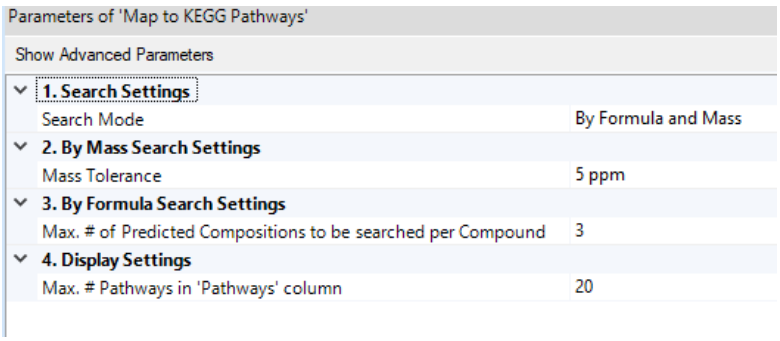
B-7: Compound 3.1 processing workflow- Mark Background Compounds Node Parameters



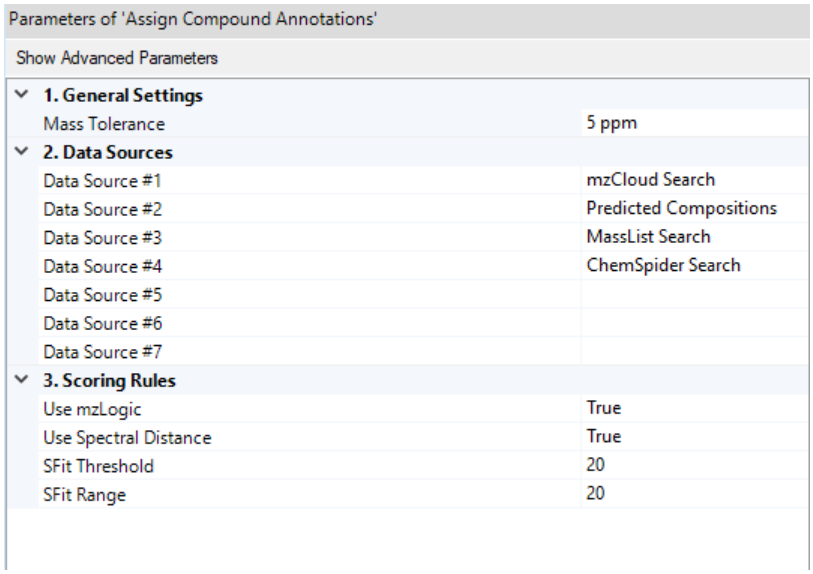
B-8: Compound 3.1 processing workflow-Map to Metabolika Pathways Parameters



B-9: Compound 3.1 processing workflow-Map to KEGG pathways node Parameters



B-10: Compound 3.1 processing workflow- Assign Compound Annotations node Parameters

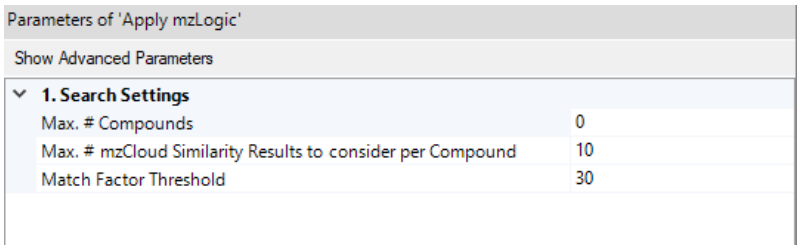


Parameters of 'Assign Compound Annotations'

Show Advanced Parameters

1. General Settings	
Mass Tolerance	5 ppm
2. Data Sources	
Data Source #1	mzCloud Search
Data Source #2	Predicted Compositions
Data Source #3	MassList Search
Data Source #4	ChemSpider Search
Data Source #5	
Data Source #6	
Data Source #7	
3. Scoring Rules	
Use mzLogic	True
Use Spectral Distance	True
SFit Threshold	20
SFit Range	20

B-11: Compound 3.1 processing workflow-Apply mzLogic node Parameters

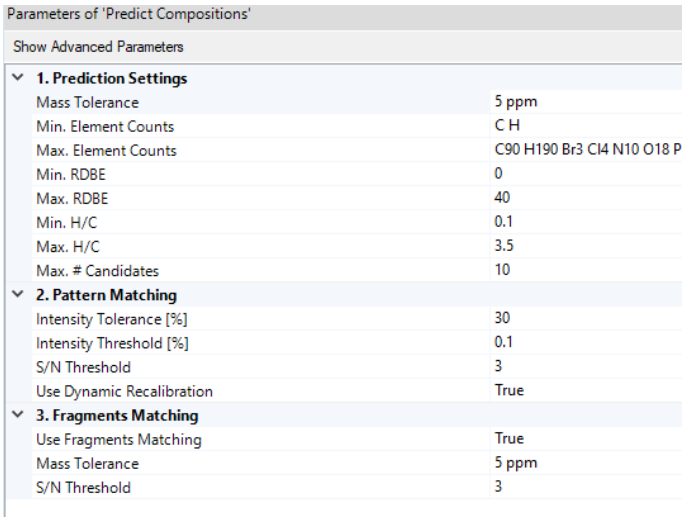


Parameters of 'Apply mzLogic'

Show Advanced Parameters

1. Search Settings	
Max. # Compounds	0
Max. # mzCloud Similarity Results to consider per Compound	10
Match Factor Threshold	30

B-12: Compound 3.1 processing workflow - Predict Compositions node Parameters

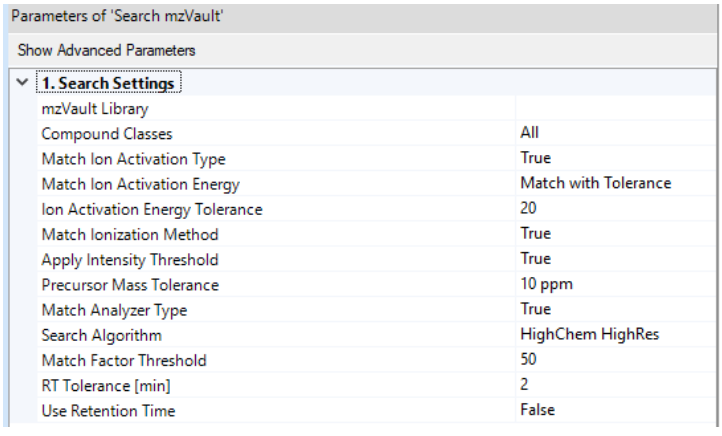


Parameters of 'Predict Compositions'

Show Advanced Parameters

1. Prediction Settings	
Mass Tolerance	5 ppm
Min. Element Counts	C H
Max. Element Counts	C90 H190 Br3 Cl4 N10 O18 P
Min. RDBE	0
Max. RDBE	40
Min. H/C	0.1
Max. H/C	3.5
Max. # Candidates	10
2. Pattern Matching	
Intensity Tolerance [%]	30
Intensity Threshold [%]	0.1
S/N Threshold	3
Use Dynamic Recalibration	True
3. Fragments Matching	
Use Fragments Matching	True
Mass Tolerance	5 ppm
S/N Threshold	3

B-13: Compound 3.1 processing workflow-Search mzVault node Parameters



Parameters of 'Search mzVault'

Show Advanced Parameters

1. Search Settings	
mzVault Library	
Compound Classes	All
Match Ion Activation Type	True
Match Ion Activation Energy	Match with Tolerance
Ion Activation Energy Tolerance	20
Match Ionization Method	True
Apply Intensity Threshold	True
Precursor Mass Tolerance	10 ppm
Match Analyzer Type	True
Search Algorithm	HighChem HighRes
Match Factor Threshold	50
RT Tolerance [min]	2
Use Retention Time	False

B-14: Compound 3.1 processing workflow- Search mzCloud node Parameters

Parameters of 'Search mzCloud'

Show Advanced Parameters

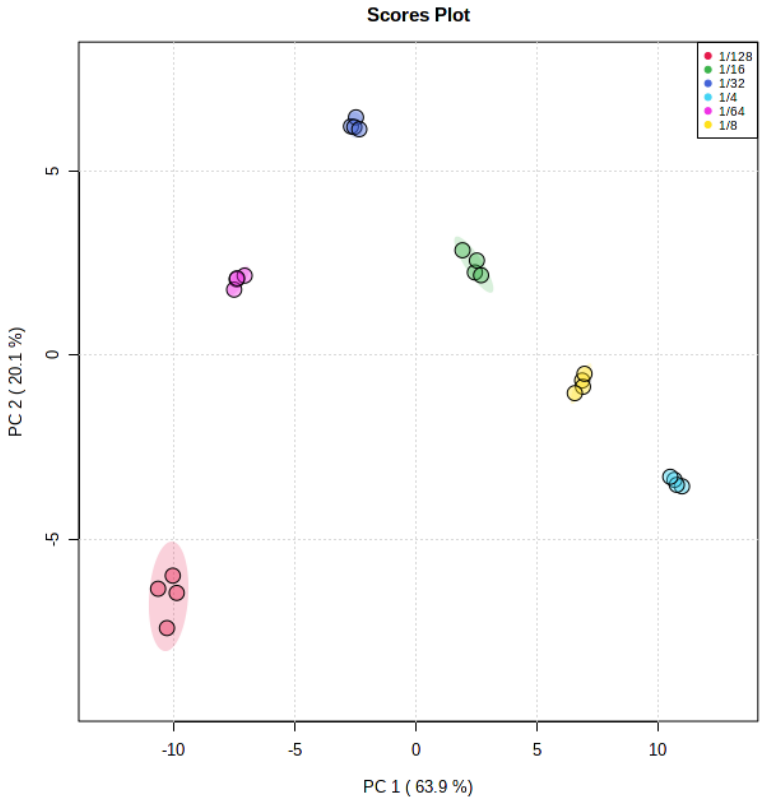
1. General Settings	
Compound Classes	All
Library	Autoprocessed; Reference
2. DDA Search	
Identity Search	HighChem HighRes
Match Activation Type	True
Match Activation Energy	Match with Tolerance
Activation Energy Tolerance	20
Apply Intensity Threshold	True
Similarity Search	None
Match Factor Threshold	60
3. DIA Search	
Use DIA Scans for Search	False
Max. Isolation Width [Da]	500
Match Activation Type	False
Match Activation Energy	Any
Activation Energy Tolerance	100
Apply Intensity Threshold	False
Match Factor Threshold	20

B-15: Compound 3.1 processing workflow-Apply Spectral Distance node Parameters

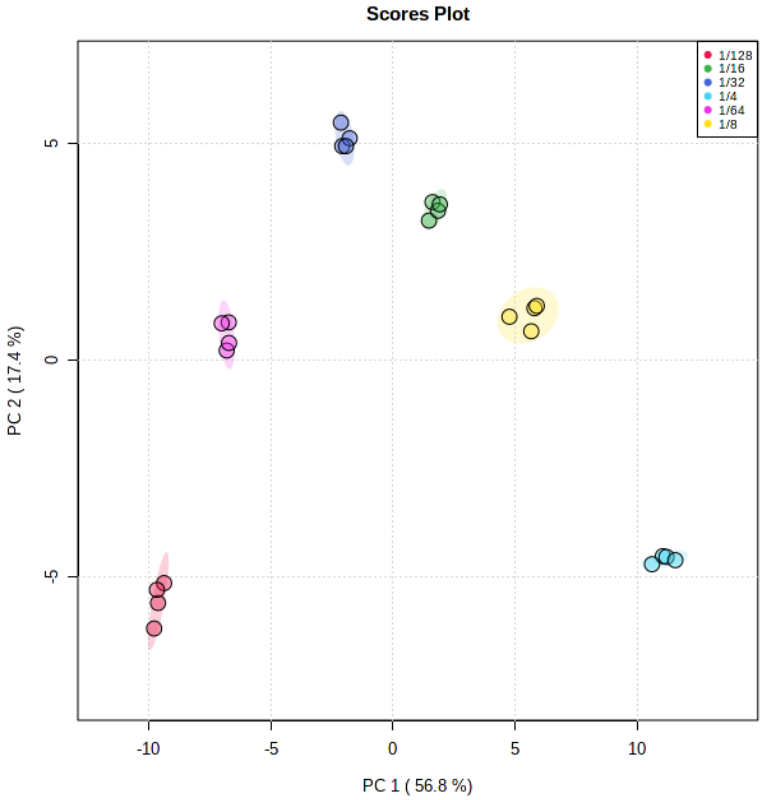
Show Advanced Parameters

1. General Settings	
Compound Classes	All
Library	Autoprocessed; Reference
2. DDA Search	
Identity Search	HighChem HighRes
Match Activation Type	True
Match Activation Energy	Match with Tolerance
Activation Energy Tolerance	20
Apply Intensity Threshold	True
Similarity Search	None
Match Factor Threshold	60
3. DIA Search	
Use DIA Scans for Search	False
Max. Isolation Width [Da]	500
Match Activation Type	False
Match Activation Energy	Any
Activation Energy Tolerance	100
Apply Intensity Threshold	False
Match Factor Threshold	20

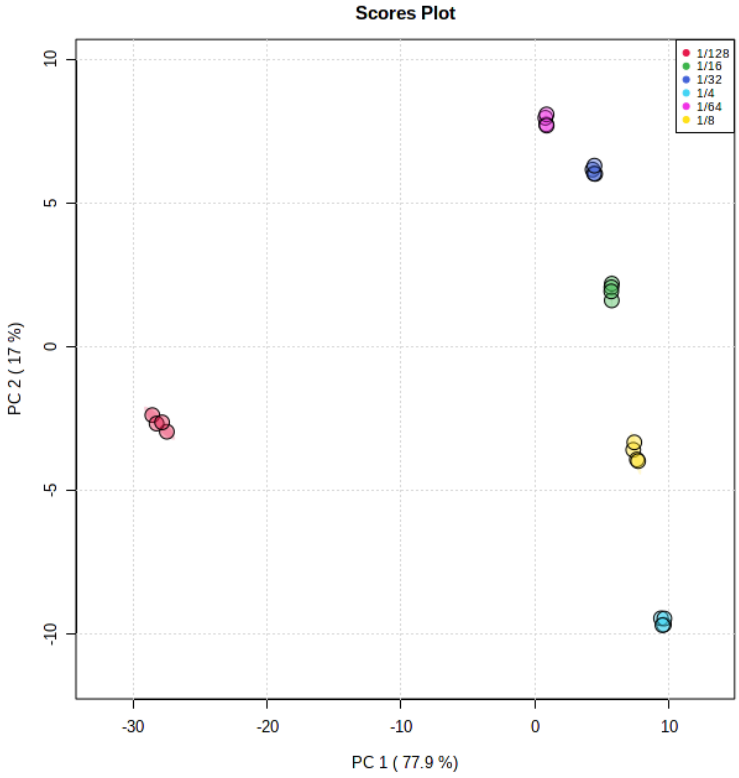
C-1: Variation between different dilutions when applying the 2.1mm ID column to HILIC assays in positive ion mode using Plasma samples



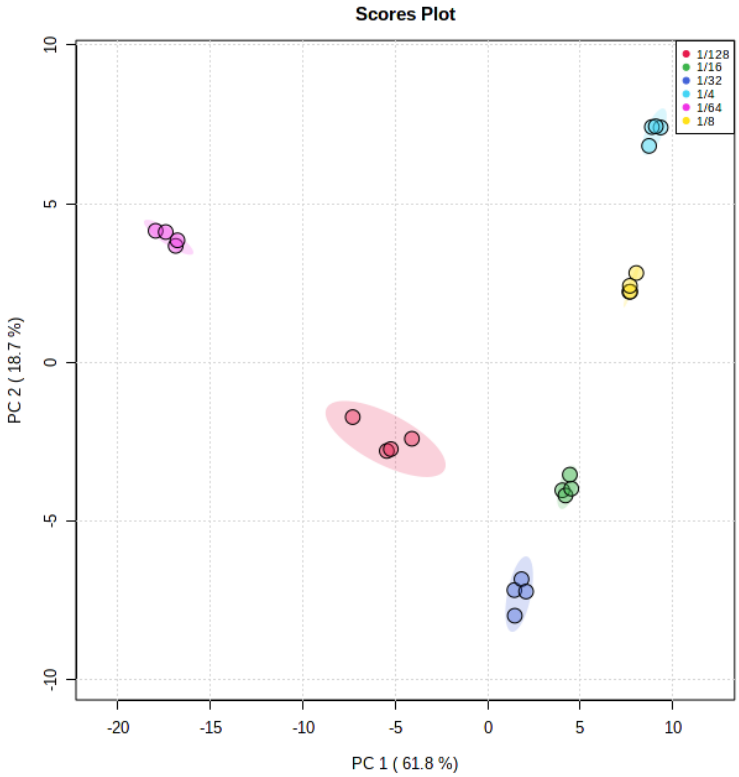
C-2: Variation between different dilutions when applying the 1mm ID column to HILIC assays in positive ion mode using Plasma samples



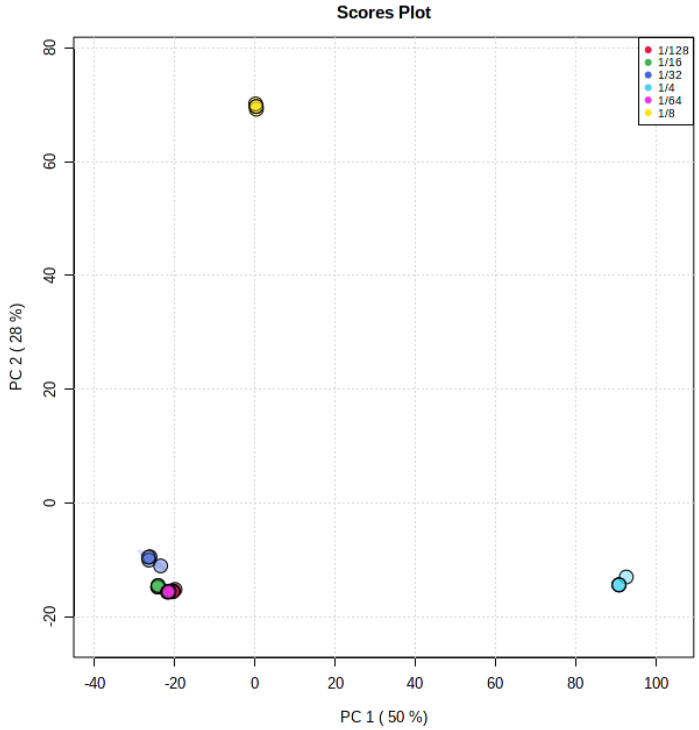
C-3 : Variation between different dilutions when applying the 2.1mm ID column to HILIC assays in positive ion mode using urine samples



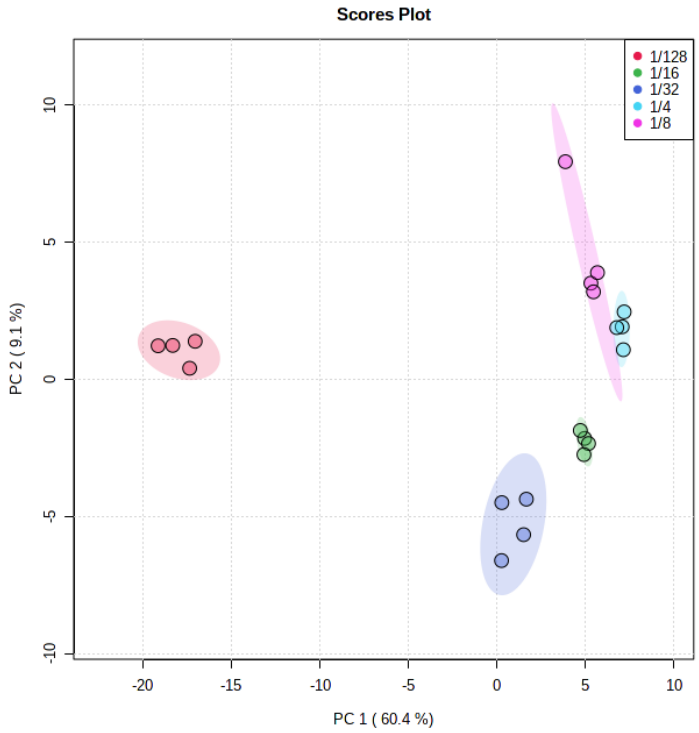
C-4: Variation between different dilutions when applying the 1mm ID column to HILIC assays in positive ion mode using urine samples



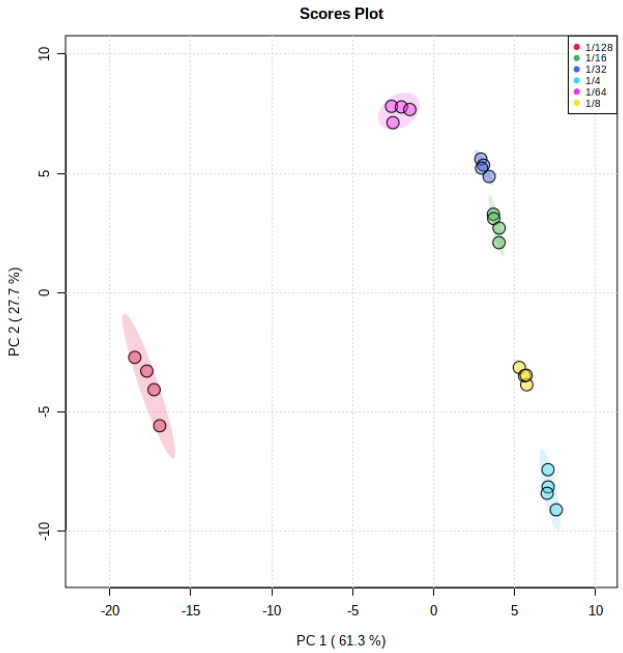
C-5: Variation between different dilutions when applying the 2.1mm ID column to Lipidomic assays in positive ion mode using Plasma samples



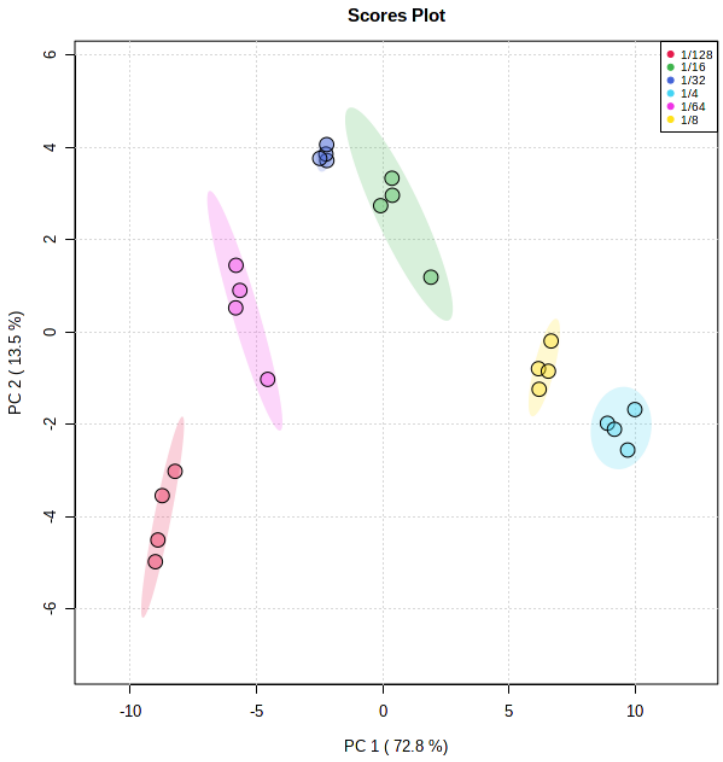
C-6: Variation between different dilutions when applying the 1mm ID column to Lipidomic assays in positive ion mode using Plasma samples



C-7: Variation between different dilutions when applying the 2.1mm ID column to Reversed Phase aqueous assays in positive ion mode using urine samples



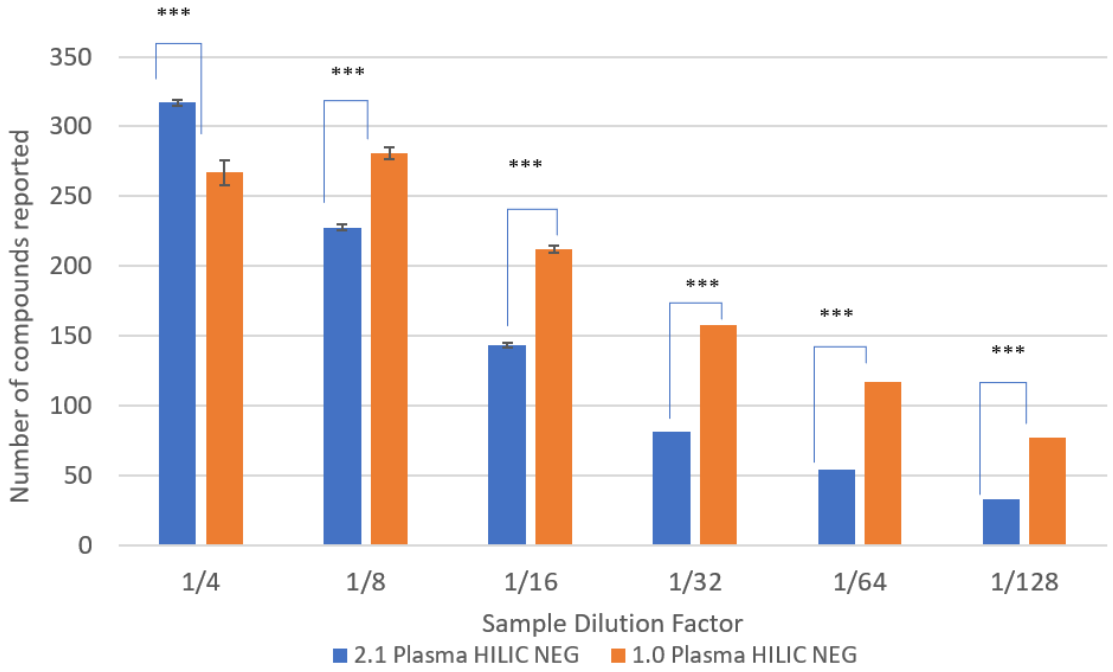
C-8: Variation between different dilutions when applying the 1mm ID column to Reversed Phase aqueous assays in positive ion mode using urine samples



D-1: Summary Table for the number of reported compounds for each dilution in each assay. All compounds reported have a FWHM <0.2 minutes, were present in ¾ of injections and RSD <30% in each separate injection

	1/4	1/4 SD	1/8	1/8 SD	1/16	1/16 SD	1/32	1/32 SD	1/64	1/64 SD	1/128	1/128 SD
2.1 Plasma HILIC NEG	317	2.12	228	2.49	143	1.87	81	2.17	55	1.66	33	1.64
1.0 Plasma HILIC NEG	267	8.79	281	1.92	212	4.21	158	3.81	117	2.38	78	1.66
2.1 Plasma HILIC POS	443	4.71	332	2.77	220	3.20	180	3.00	137	0.71	104	3.50
1.0 Plasma HILIC POS	467	5.10	428	2.12	287	2.50	243	1.41	190	3.20	124	0.87
2.1 Urine HILIC NEG	516	3.49	307	1.87	175	2.24	96	0.83	59	1.30	30	0.43
1.0 Urine HILIC NEG	375	2.49	365	3.03	267	1.30	206	1.58	95	2.86	149	2.96
2.1 Urine HILIC POS	792	2.06	585	1.48	394	1.12	296	2.17	206	1.92	114	1.30
1.0 Urine HILIC POS	676	1.92	538	4.49	455	3.49	336	2.45	175	2.06	243	2.96
2.1 Plasma LIPID NEG	749	20.00	402	8.70	256	5.12	167	3.11	123	1.50	81	0.71
1.0 Plasma LIPID NEG	321	17.17	273	5.79	254	18.23	180	1.66	126	4.03	80	2.74
2.1 Plasma LIPID POS	1273	2.05	813	7.12	725	165.55	479	31.58	364	6.04	305	2.92
1.0 Plasma LIPID POS	489	7.89	340	21.98	271	12.75	208	2.87	108	4.15	NA	NA
2.1 Urine RPaq NEG	320	25.62	186	13.75	92	4.09	59	2.77	31	1.50	18	0.71
1.0 Urine RPaq NEG	263	1.41	173	2.05	73	1.79	46	1.22	29	1.48	17	0.43
2.1 Urine RPaq POS	792	4.15	516	3.08	257	3.11	215	4.72	123	1.41	69	3.40
1.0 Urine RPaq POS	587	7.89	358	2.06	225	6.22	159	1.22	96	9.65	77	3

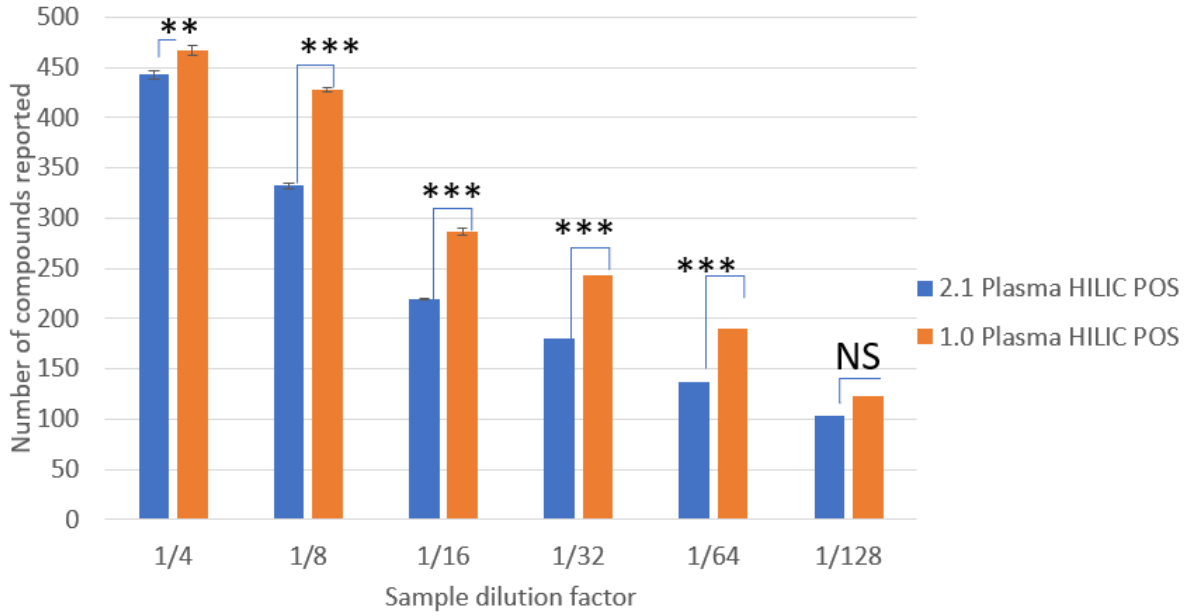
D-2: Graph of number of compounds reported for both 2.1mm and 1mm ID columns when applied to HILIC assays in negative ion mode using Plasma samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained. The difference



D-3: Table of number of compounds reported for both 2.1mm and 1mm ID columns when applied to HILIC assays in negative ion mode using Plasma samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained.

	2.1 Plasma HILIC NEG	1.0 Plasma HILIC NEG
1/4	317	267
1/4 SD	2.12	8.79
1/8	228	281
1/8 SD	2.49	1.92
1/16	143	212
1/16 SD	1.87	4.21
1/32	81	158
1/32 SD	2.17	3.81
1/64	55	117
1/64 SD	1.66	2.38
1/128	33	78
1/128 SD	1.64	1.66

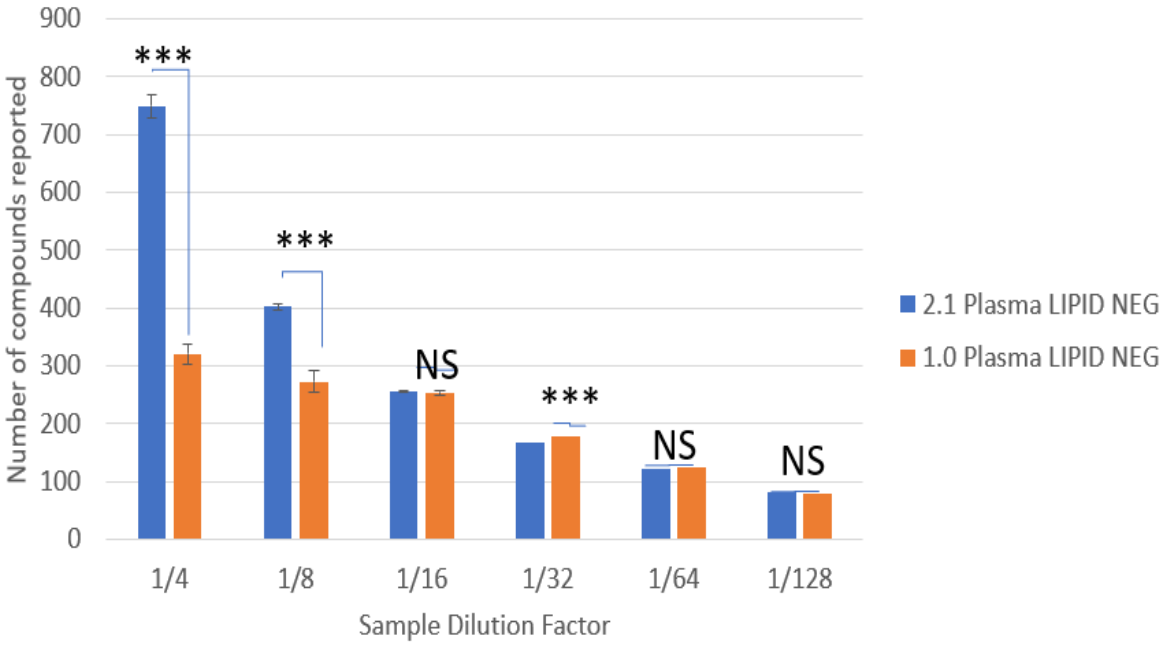
D-4: Graph of number of compounds reported for both 2.1mm and 1mm ID columns when applied to HILIC assays in positive ion mode using Plasma samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained. The difference in the number of compounds detected by each column ID was moderately significant (P<0.01), and then each dilution was of a very significant (P<0.001), until the final one which was not significant (P>0.05)



D-5: Table of number of compounds reported for both 2.1mm and 1mm ID columns when applied to HILIC assays in positive ion mode using Plasma samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained.

	2.1 Plasma HILIC POS	1.0 Plasma HILIC POS
1/4	443	467
1/4 SD	4.71	5.10
1/8	332	428
1/8 SD	2.77	2.12
1/16	220	287
1/16 SD	3.20	2.50
1/32	180	243
1/32 SD	3.00	1.41
1/64	137	190
1/64 SD	0.71	3.20
1/128	104	124
1/128 SD	3.50	0.87

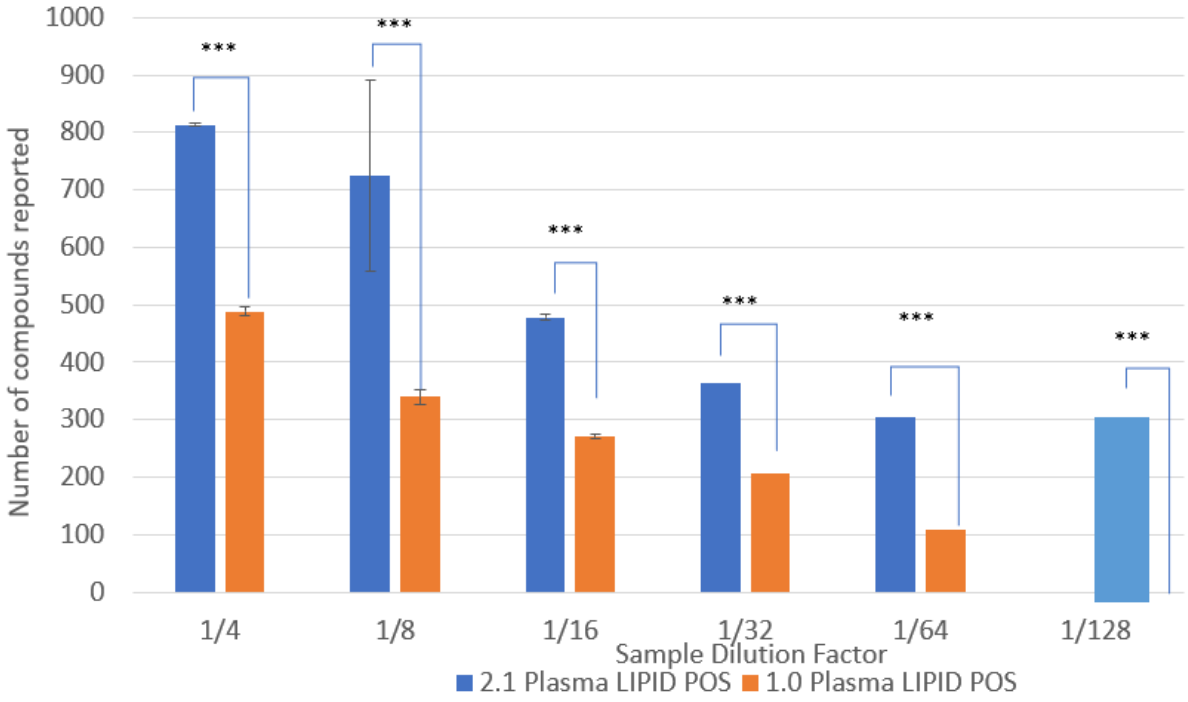
D-6: Graph of number of compounds reported for both 2.1mm and 1mm ID columns when applied to Lipidomic assays in negative ion mode using Plasma samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained. The difference in the number of compounds detected from each column ID was very significant (P<0.001), or not significant (P>0.05).



D-7: Table of number of compounds reported for both 2.1mm and 1mm ID columns when applied to Lipidomic assays in negative ion mode using Plasma samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained.

	2.1 Plasma LIPID NEG	1.0 Plasma LIPID NEG
1/4	749	321
1/4 SD	20.00	17.17
1/8	402	273
1/8 SD	8.70	5.79
1/16	256	254
1/16 SD	5.12	18.23
1/32	167	180
1/32 SD	3.11	1.66
1/64	123	126
1/64 SD	1.50	4.03
1/128	81	80
1/128 SD	0.71	2.74

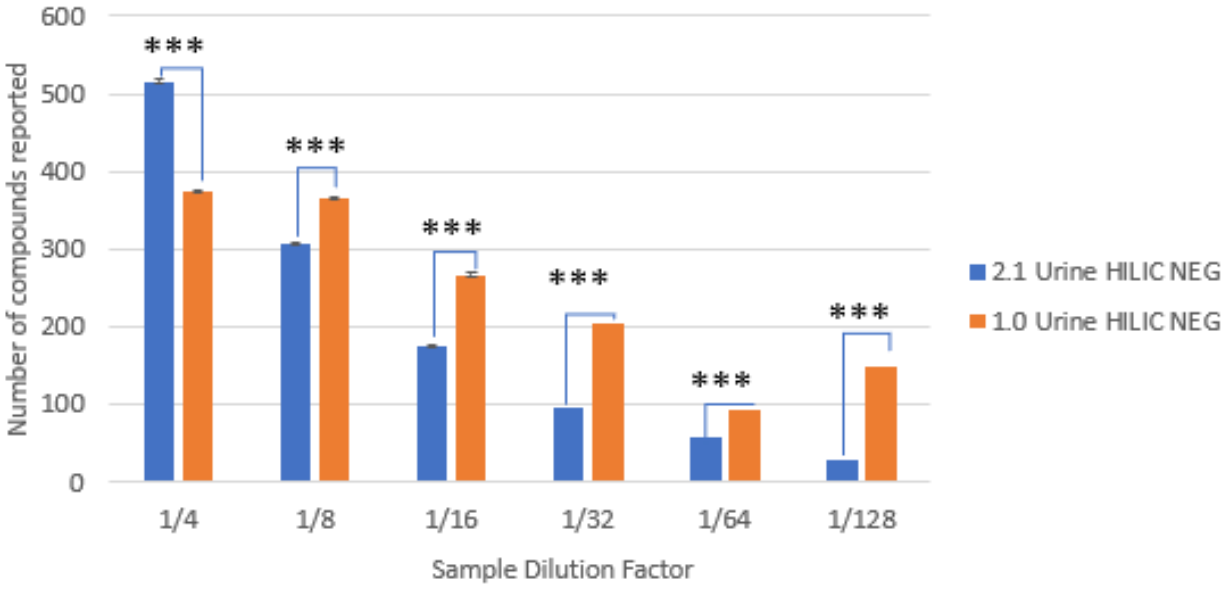
D-8: Graph of number of compounds reported for both 2.1mm and 1mm ID columns when applied to Lipidomic assays in positive ion mode using Plasma samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained. The difference in the number of compounds detected from each column ID was very significant (P<0.001).



D-9: Table of number of compounds reported for both 2.1mm and 1mm ID columns when applied to Lipidomic assays in positive ion mode using Plasma samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained.

	2.1 Plasma LIPID POS	1.0 Plasma LIPID POS
1/4	1273	271
1/4 SD	2.05	7.89
1/8	813	340
1/8 SD	7.12	21.98
1/16	725	489
1/16 SD	165.55	12.75
1/32	479	208
1/32 SD	31.58	2.87
1/64	364	108
1/64 SD	6.04	4.15
1/128	305	NA
1/128 SD	2.92	NA

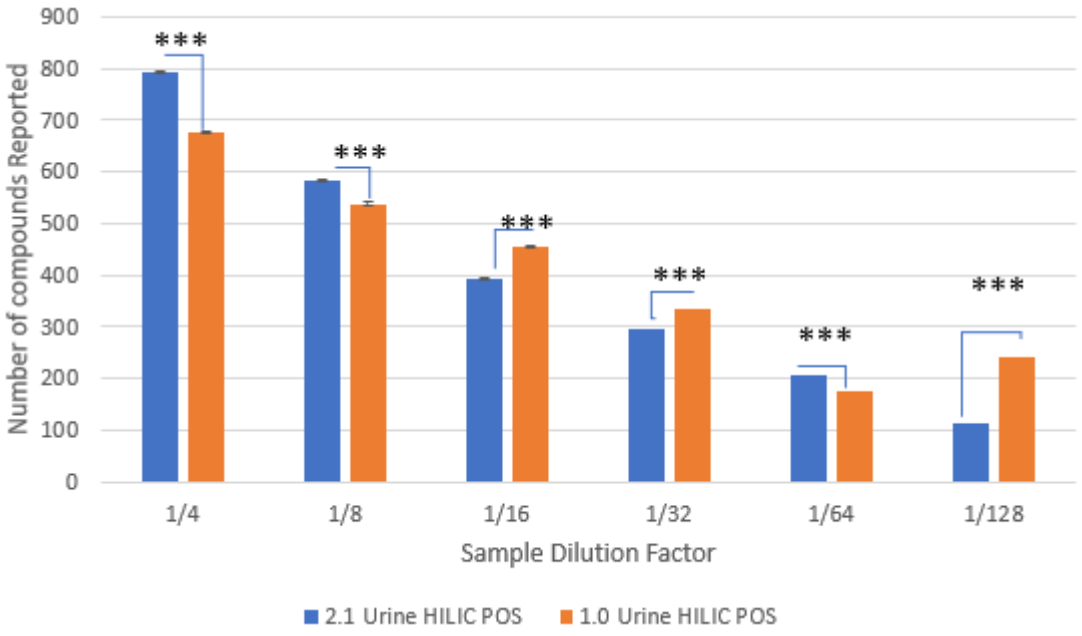
D-10: Graph of number of compounds reported for both 2.1mm and 1mm ID columns when applied to HILIC assays in negative ion mode using Urine samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained. The difference in the number of compounds at each dilution was of a very significant (P<0.001)



D-11: Table of number of compounds reported for both 2.1mm and 1mm ID columns when applied to HILIC assays in negative ion mode using Urine samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained.

	2.1 Urine HILIC NEG	1.0 Urine HILIC NEG
1/4	516	375
1/4 SD	3.49	2.49
1/8	307	365
1/8 SD	1.87	3.03
1/16	175	267
1/16 SD	2.24	1.30
1/32	96	206
1/32 SD	0.83	1.58
1/64	59	95
1/64 SD	1.30	2.86
1/128	30	149
1/128 SD	0.43	2.96

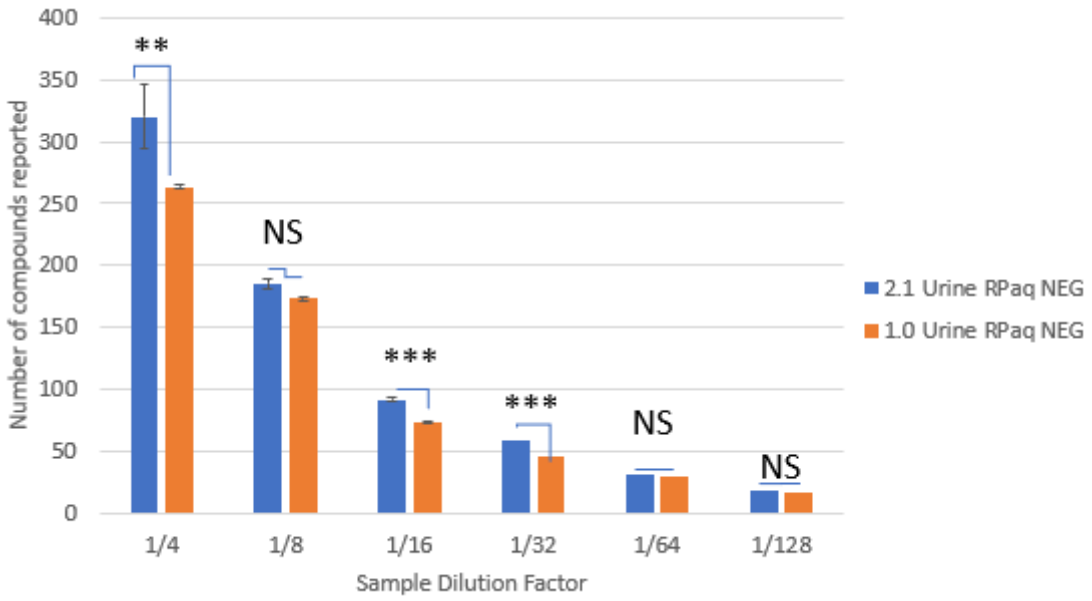
D-12: Graph of number of compounds reported for both 2.1mm and 1mm ID columns when applied to HILIC assays in positive ion mode using Urine samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained. The difference in the number of compounds at each dilution was of a very significant (P<0.001)



D-13: Table of number of compounds reported for both 2.1mm and 1mm ID columns when applied to HILIC assays in positive ion mode using Urine samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained.

	2.1 Urine HILIC POS	1.0 Urine HILIC POS
1/4	792	676
1/4 SD	2.06	1.92
1/8	585	538
1/8 SD	1.48	4.49
1/16	394	455
1/16 SD	1.12	3.49
1/32	296	336
1/32 SD	2.17	2.45
1/64	206	175
1/64 SD	1.92	2.06
1/128	114	243
1/128 SD	1.30	2.96

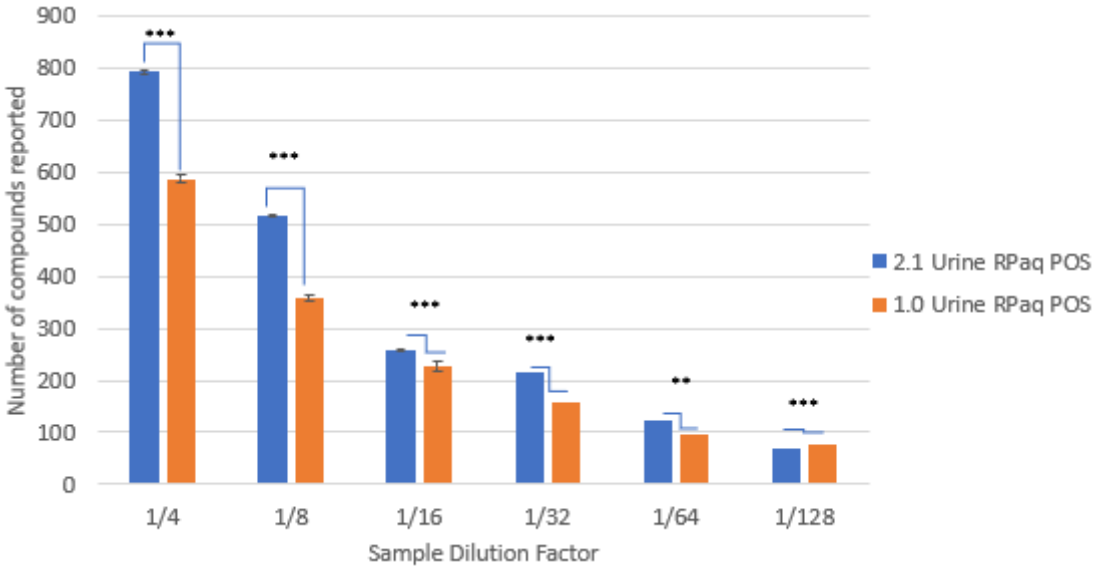
D-14: Graph of number of compounds reported for both 2.1mm and 1mm ID columns when applied to Reversed Phase aqueous assays in negative ion mode using Urine samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained. The difference in the number of compounds at each dilution was of a very significant (P<0.001)



D-15: Table of number of compounds reported for both 2.1mm and 1mm ID columns when applied to Reversed Phase aqueous assays in negative ion mode using Urine samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained.

	2.1 Urine RPaq NEG	1.0 Urine RPaq NEG
1/4	320	263
1/4 SD	25.62	1.41
1/8	186	173
1/8 SD	13.75	2.05
1/16	92	73
1/16 SD	4.09	1.79
1/32	59	46
1/32 SD	2.77	1.22
1/64	31	29
1/64 SD	1.50	1.48
1/128	18	17
1/128 SD	0.71	0.43

D-16: Graph of number of compounds reported for both 2.1mm and 1mm ID columns when applied to Reversed Phase aqueous assays in positive ion mode using Urine samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained. The difference in the number of compounds at each dilution was of a very significant (P<0.001)



D-17: Table of number of compounds reported for both 2.1mm and 1mm ID columns when applied to Reversed Phase aqueous assays in positive ion mode using Urine samples. Compounds were filtered so that only peaks with FWHM <0.2 minutes, were present in ¾ injections in each dilution and had an RSD of <30% remained.

	2.1 Urine RPaq POS	1.0 Urine RPaq POS
1/4	792	587
1/4 SD	4.15	7.89
1/8	516	358
1/8 SD	3.08	2.06
1/16	257	225
1/16 SD	3.11	6.22
1/32	215	159
1/32 SD	4.72	1.22
1/64	123	96
1/64 SD	1.41	9.65
1/128	69	77
1/128 SD	3.40	3

E-1: A summary of the mean distribution of chromatographic peak areas for all dilutions of 2.1mm ID vs. 1mm ID for all assays and sample types. Compounds used were present in 75% of replicates and had a relative standard deviation (RSD) <30% are reported. NEG = negative ion mode, POS = positive ion mode, SD = standard deviation

	1/4	1/4 SD	1/8	1/8 SD	1/16	1/16 SD	1/32	1/32 SD	1/64	1/64 SD	1/128	1/128 SD
2.1 Plasma HILIC NEG	2.07 E+07	1.29 E+08	1.50 E+07	8.04 E+07	1.24 E+07	6.45 E+07	8.98 E+06	3.79 E+07	1.06 E+07	3.34 E+07	7.26 E+06	2.16 E+07
1.0 Plasma HILIC NEG	4.21 E+07	1.87 E+08	7.55 E+07	4.32 E+08	6.99 E+07	3.48 E+08	7.92 E+07	3.95 E+08	6.65 E+07	2.84 E+08	5.70 E+07	3.01 E+08
2.1 Plasma HILIC POS	4.79 E+07	2.88 E+08	3.58 E+07	2.07 E+08	2.97 E+07	1.58 E+08	2.27 E+07	1.14 E+08	1.87 E+07	8.47 E+07	1.54 E+07	6.57 E+07
1.0 Plasma HILIC POS	8.53 E+07	4.76 E+08	6.93 E+07	3.76 E+08	5.41 E+07	2.53 E+08	4.53 E+07	1.92 E+08	3.72 E+07	1.59 E+08	2.99 E+07	1.01 E+08
2.1 Urine HILIC NEG	1.79 E+07	7.86 E+07	1.19 E+07	4.32 E+07	9.09 E+06	2.70 E+07	7.77 E+06	2.02 E+07	6.99 E+06	1.69 E+07	5.75 E+06	1.19 E+07
1.0 Urine HILIC NEG	1.51 E+08	7.03 E+08	9.89 E+07	4.86 E+08	7.44 E+07	2.94 E+08	6.25 E+07	2.32 E+08	4.03 E+07	1.31 E+08	5.16 E+07	1.81 E+08
2.1 Urine HILIC POS	7.53 E+07	6.76 E+08	5.67 E+07	5.01 E+08	5.12 E+07	4.34 E+08	4.03 E+07	3.27 E+08	3.64 E+07	2.85 E+08	2.66 E+07	1.95 E+08
1.0 Urine HILIC POS	1.20 E+08	7.45 E+08	1.41 E+08	1.72 E+09	1.10 E+08	1.34 E+09	8.77 E+07	9.67 E+08	4.91 E+07	5.11 E+08	7.26 E+07	7.84 E+08
2.1 Plasma LIPID NEG	2.14 E+07	6.58 E+07	2.70 E+07	1.11 E+08	2.32 E+07	6.98 E+07	2.26 E+07	4.78 E+07	1.69 E+07	3.60 E+07	1.43 E+07	2.49 E+07
1.0 Plasma LIPID NEG	4.59 E+07	2.01 E+08	3.92 E+07	1.42 E+08	5.46 E+07	1.68 E+08	4.91 E+07	1.41 E+08	4.01 E+07	9.86 E+07	4.49 E+07	1.14 E+08

2.1 Plasma LIPID POS	1.53 E+08	9.23 E+08	1.05 E+08	6.28 E+08	6.28 E+08	5.39 E+08	6.94 E+07	3.44 E+08	6.19 E+07	2.22 E+08	6.51 E+07	2.60 E+08
1.0 Plasma LIPID POS	1.03 E+08	5.53 E+08	1.48 E+08	9.67 E+08	1.02 E+08	5.11 E+08	1.22 E+08	5.67 E+08	1.31 E+08	5.58 E+08	9.91 E+07	3.31 E+08
2.1 Urine RPaq NEG	1.02 E+07	2.54 E+07	1.08 E+07	2.48 E+07	1.18 E+07	2.65 E+07	1.23 E+07	2.92 E+07	1.53 E+07	3.31 E+07	1.27 E+07	2.63 E+07
1.0 Urine RPaq NEG	1.56 E+07	3.20 E+07	1.44 E+07	2.89 E+07	1.76 E+07	3.79 E+07	1.98 E+07	3.58 E+07	2.02 E+07	4.03 E+07	2.19 E+07	4.25 E+07
2.1 Urine RPaq POS	1.71 E+07	8.62 E+07	1.60 E+07	6.52 E+07	1.71 E+07	5.25 E+07	1.64 E+07	4.69 E+07	1.69 E+07	4.58 E+07	1.71 E+07	4.43 E+07
1.0 Urine RPaq POS	2.71 E+07	2.58 E+08	2.58 E+07	1.82 E+08	2.26 E+07	1.03 E+08	2.26 E+07	1.03 E+08	2.24 E+07	6.09 E+07	2.23 E+07	5.45 E+07

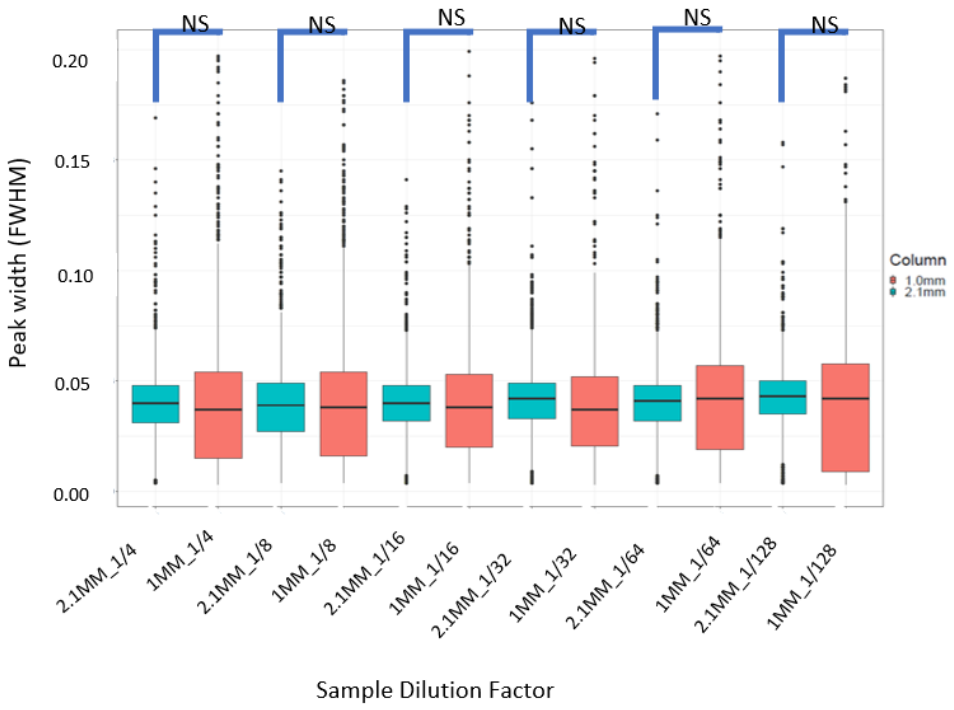
E-2: A summary of the full width half max for each dilution applying 2.1mm ID column to a Lipidomics assay in positive ion mode using Plasma samples.

	No. compounds	FWHM (mean)	FWHM (SD)	FWHM (QUART1)	FWHM (QUART/median)	FWHM (QUART3)	Total peak area (mean)	Total peak area (SD)
1/4	2181	0.0397	0.0296	0.0150	0.0370	0.0540	1.53E+08	9.23E+08
1/8	1521	0.0401	0.0306	0.0160	0.0380	0.0540	1.05E+08	6.28E+08
1/16	1154	0.0402	0.0293	0.0200	0.0380	0.0530	6.28E+08	5.39E+08
1/32	815	0.0394	0.0299	0.0205	0.0370	0.0520	6.94E+07	3.44E+08
1/64	734	0.0445	0.0331	0.0190	0.0420	0.0570	6.19E+07	2.22E+08
1/128	754	0.0417	0.0322	0.0090	0.0420	0.0578	6.51E+07	2.60E+08

E-3: A summary of the full width half max for each dilution applying 1mm ID column to a Lipidomics assay in positive ion mode using Plasma samples.

	No. compounds	FWHM (mean)	FWHM (SD)	FWHM (QUART1)	FWHM (QUART/median)	FWHM (QUART3)	Total peak area (mean)	Total peak area (SD)
1/4	1189	0.0406	0.0184	0.0310	0.0400	0.0480	1.03E+08	5.53E+08
1/8	2192	0.0394	0.0199	0.0270	0.0390	0.0490	1.48E+08	9.67E+08
1/16	1312	0.0397	0.0182	0.0318	0.0400	0.0480	1.02E+08	5.11E+08
1/32	1231	0.0418	0.0189	0.0330	0.0420	0.0490	1.22E+08	5.67E+08
1/64	1410	0.0401	0.0180	0.0320	0.0410	0.0480	1.31E+08	5.58E+08
1/128	937	0.0422	0.0185	0.0350	0.0430	0.0500	9.91E+07	3.31E+08

E-4: Distribution of Full Width Half Max between all dilutions applying both the 2.1mm ID and 1mm ID columns to Lipidomics assay in positive ion mode on plasma samples. Comparisons between columns were not statistically



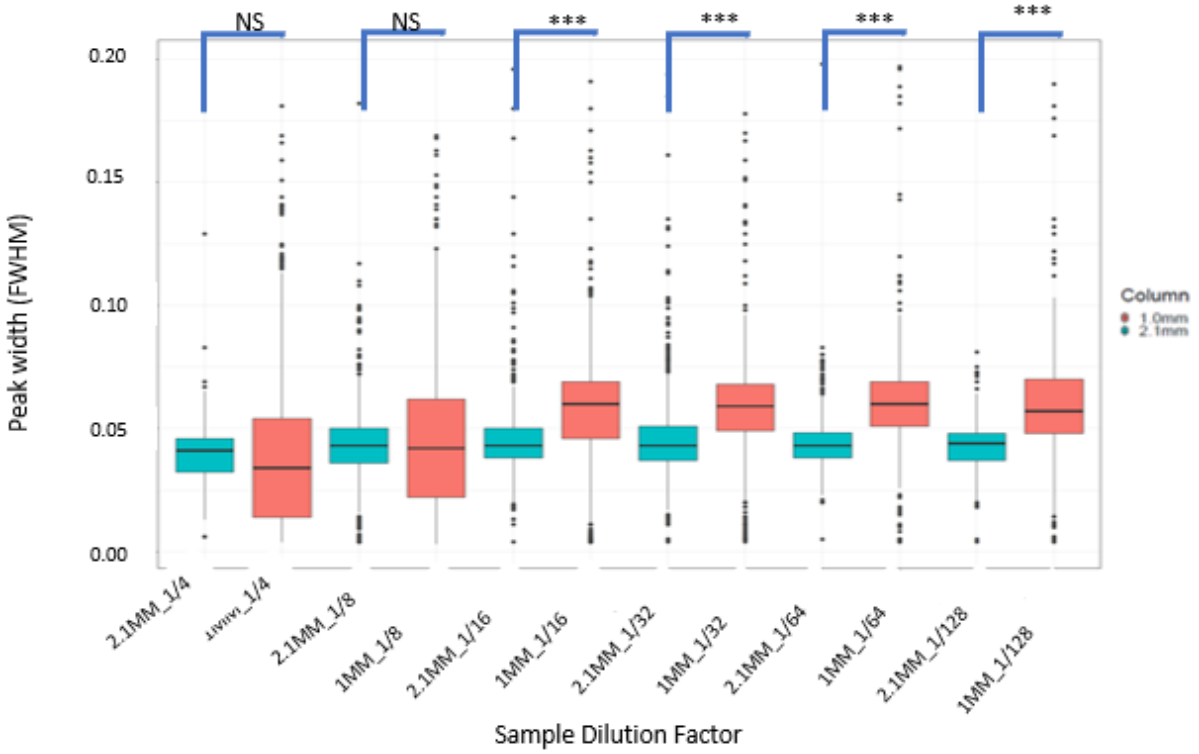
E-5: A summary of the full width half max for each dilution applying 2.1mm ID column to a Lipidomics assay in negative ion mode using Plasma samples.

	No. compounds	FWHM (mean)	FWHM (SD)	FWHM (QUART1)	FWHM (QUART/median)	FWHM (QUART3)	Total peak area (mean)	Total peak area (SD)
1/4	319	0.0403	0.0120	0.03225	0.041	0.046	2.14E+07	6.58E+07
1/8	675	0.0443	0.0161	0.036	0.043	0.05	2.70E+07	1.11E+08
1/16	529	0.0464	0.0179	0.038	0.043	0.05	2.32E+07	6.98E+07
1/32	526	0.0464	0.0201	0.037	0.043	0.051	2.26E+07	4.78E+07
1/64	284	0.0450	0.0145	0.038	0.043	0.04825	1.69E+07	3.60E+07
1/128	129	0.0439	0.0130	0.037	0.044	0.048	1.43E+07	2.49E+07

E-6: A summary of the full width half max for each dilution applying 1mm ID column to a Lipidomics assay in negative ion mode using Plasma samples.

	No. compounds	FWHM (mean)	FWHM (SD)	FWHM (QUART1)	FWHM (QUART/median)	FWHM (QUART3)	Total peak area (mean)	Total peak area (SD)
1/4	1251	0.0374	0.0277	0.014	0.034	0.054	4.59E+07	2.01E+08
1/8	960	0.0451	0.0298	0.022	0.042	0.062	3.92E+07	1.42E+08
1/16	535	0.0569	0.0279	0.046	0.06	0.069	5.46E+07	1.68E+08
1/32	463	0.0586	0.0242	0.049	0.059	0.068	4.91E+07	1.41E+08
1/64	364	0.0608	0.0261	0.051	0.06	0.069	4.01E+07	9.86E+07
1/128	342	0.0586	0.0258	0.048	0.057	0.07	4.49E+07	1.14E+08

E-7: Distribution of Full Width Half Max between all dilutions applying both the 2.1mm ID and 1mm ID columns to Lipidomics assay in negative ion mode on plasma samples. Comparisons between columns were very significant (P≤0.001) from 1 in 16 onwards. The two most concentrated dilutions were not statistically significant (P>0.05)



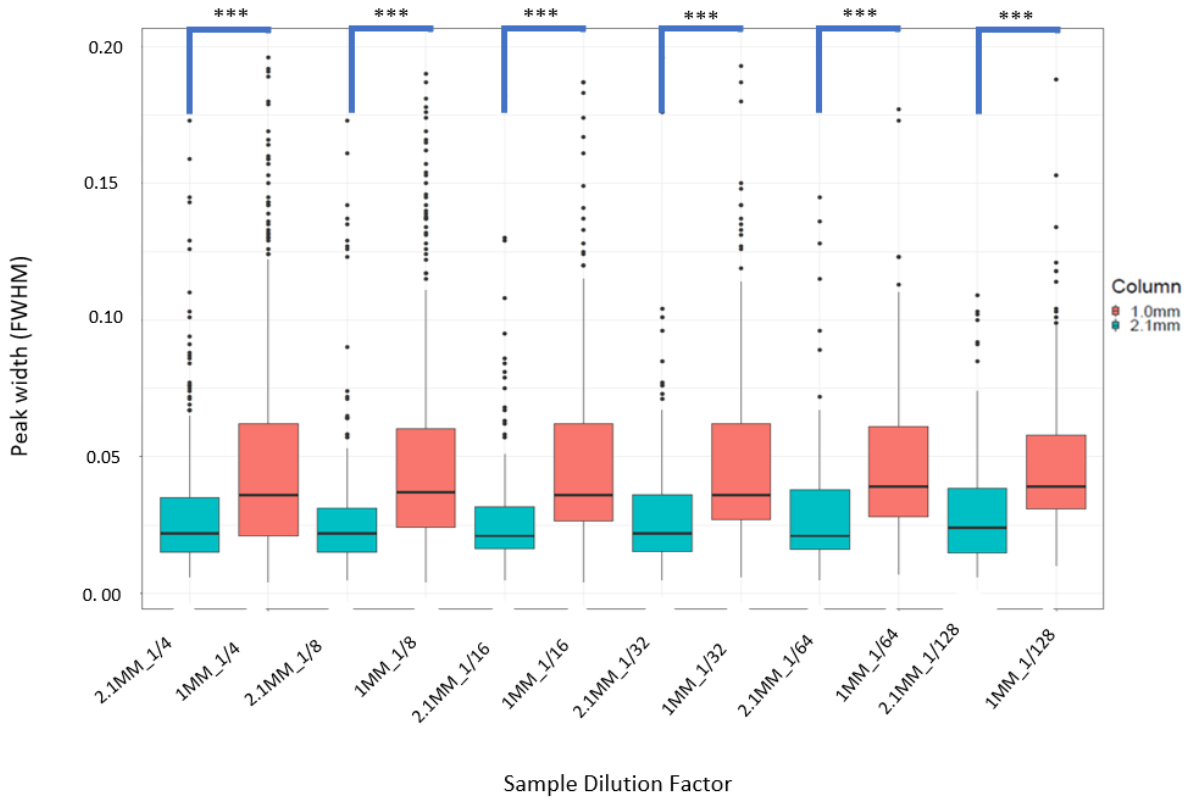
E-8: A summary of the full width half max for each dilution applying 2.1mm ID column to a HILIC assay in negative ion mode using Plasma samples.

	No. compounds	FWHM (mean)	FWHM (SD)	FWHM (QUART1)	FWHM (QUART/ median)	FWHM (QUART3)	Total peak area (mean)	Total peak area (SD)
1/4	398	0.0299	0.0269	0.015	0.022	0.035	2.07E+07	1.29E+08
1/8	296	0.0279	0.0243	0.015	0.022	0.03125	1.50E+07	8.04E+07
1/16	199	0.0279	0.0231	0.01625	0.021	0.03175	1.24E+07	6.45E+07
1/32	142	0.0302	0.0276	0.01525	0.022	0.036	8.98E+06	3.79E+07
1/64	99	0.0312	0.0271	0.016	0.021	0.038	1.06E+07	3.34E+07
1/128	84	0.0317	0.0246	0.01475	0.024	0.0385	7.26E+06	2.16E+07

E-9: A summary of the full width half max for each dilution applying 1mm ID column to a HILIC assay in negative ion mode using Plasma samples.

	No. compounds	FWHM (mean)	FWHM (SD)	FWHM (QUART1)	FWHM (QUART/ median)	FWHM (QUART3)	Total peak area (mean)	Total peak area (SD)
1/4	484	0.0482	0.0383	0.021	0.036	0.062	4.21E+07	1.87E+08
1/8	460	0.0471	0.0369	0.024	0.037	0.06025	7.55E+07	4.32E+08
1/16	331	0.0472	0.0335	0.0265	0.036	0.062	6.99E+07	3.48E+08
1/32	255	0.0493	0.0351	0.027	0.036	0.062	7.92E+07	3.95E+08
1/64	175	0.0474	0.0285	0.028	0.039	0.061	6.65E+07	2.84E+08
1/128	122	0.0492	0.0300	0.031	0.039	0.058	5.70E+07	3.01E+08

E-10: Distribution of Full Width Half Max between all dilutions applying both the 2.1mm ID and 1mm ID columns to HILIC assay in negative ion mode on plasma samples. Comparisons between columns were very significant (P≤0.001)



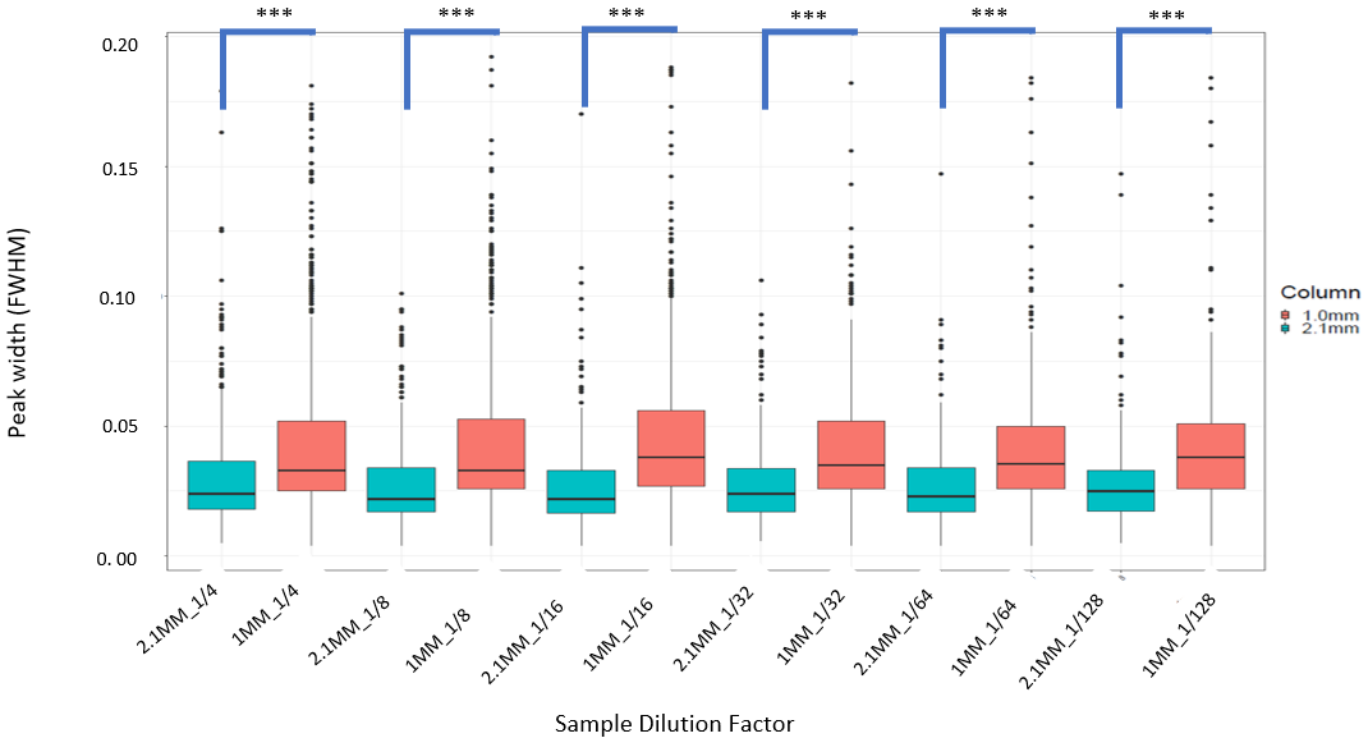
E-11: A summary of the full width half max for each dilution applying 2.1mm ID column to a HILIC assay in positive ion mode using Plasma samples.

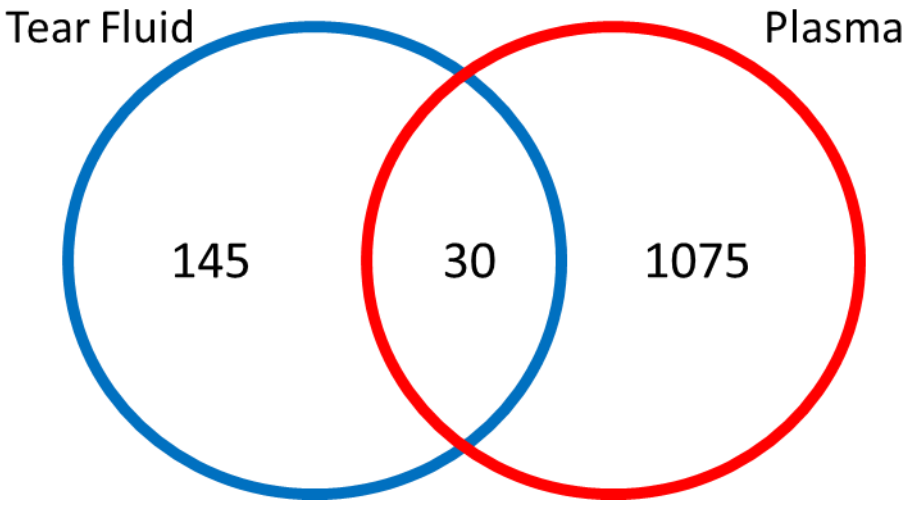
	No. compounds	FWHM (mean)	FWHM (SD)	FWHM (QUART1)	FWHM (QUART/ median)	FWHM (QUART3)	Total peak area (mean)	Total peak area (SD)
1/4	579	0.0301	0.0205	0.018	0.024	0.0365	4.79E+07	2.88E+08
1/8	495	0.0270	0.0162	0.017	0.022	0.034	3.58E+07	2.07E+08
1/16	371	0.0268	0.0171	0.0165	0.022	0.033	2.97E+07	1.58E+08
1/32	350	0.0271	0.0157	0.017	0.024	0.03375	2.27E+07	1.14E+08
1/64	243	0.0277	0.0171	0.017	0.023	0.034	1.87E+07	8.47E+07
1/128	246	0.0282	0.0183	0.01725	0.025	0.033	1.54E+07	6.57E+07

E-12: A summary of the full width half max for each dilution applying 1mm ID column to a HILIC assay in positive ion mode using Plasma samples.

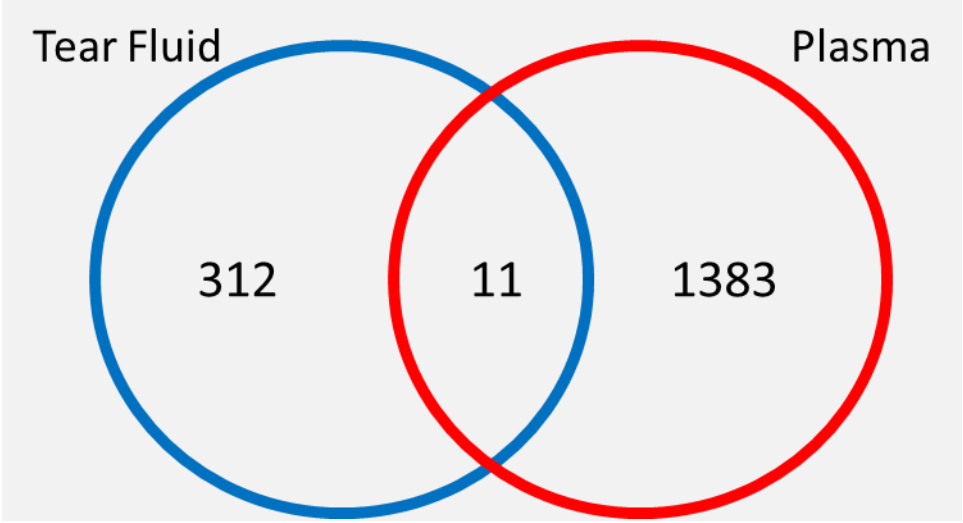
	No. compounds	FWHM (mean)	FWHM (SD)	FWHM (QUART1)	FWHM (QUART/ median)	FWHM (QUART3)	Total peak area (mean)	Total peak area (SD)
1/4	741	0.0437	0.0321	0.025	0.033	0.052	8.53E+07	4.76E+08
1/8	606	0.0444	0.0299	0.026	0.033	0.05275	6.93E+07	3.76E+08
1/16	567	0.0460	0.0302	0.027	0.038	0.056	5.41E+07	2.53E+08
1/32	374	0.0428	0.0253	0.026	0.035	0.052	4.53E+07	1.92E+08
1/64	322	0.0420	0.0264	0.026	0.0355	0.05	3.72E+07	1.59E+08
1/128	229	0.0435	0.0273	0.026	0.038	0.051	2.99E+07	1.01E+08

E-13: Distribution of Full Width Half Max between all dilutions applying both the 2.1mm ID and 1mm ID columns to HILIC assay in positive ion mode on plasma samples. Comparisons between columns were very significant (P≤0.001)

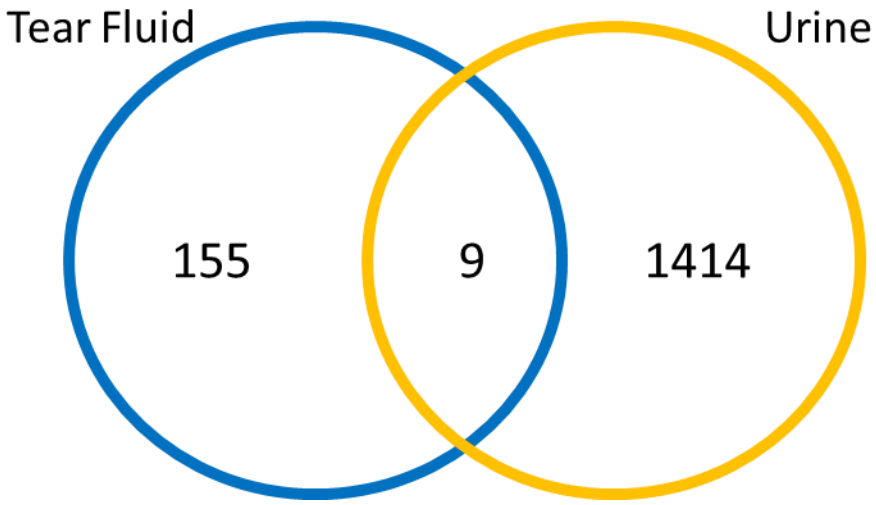




F-1: Overlap of porcine plasma (red) and tear fluid (blue) compounds detected during analysis using lipidomics assay in negative ion mode.



F-2:Overlap of porcine plasma (red) and tear fluid (blue) compounds detected during analysis using lipidomics assay in positive ion mode.

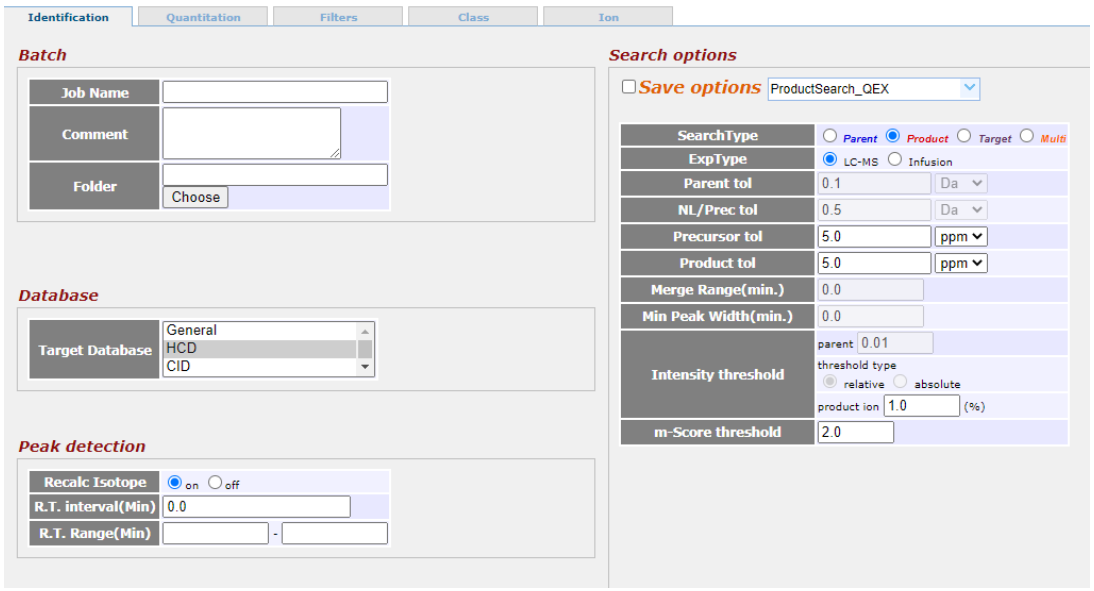


F-3: Overlap of porcine urine (red) and tear fluid (blue) compounds detected during analysis using reversed phase aqueous assay in negative ion mode

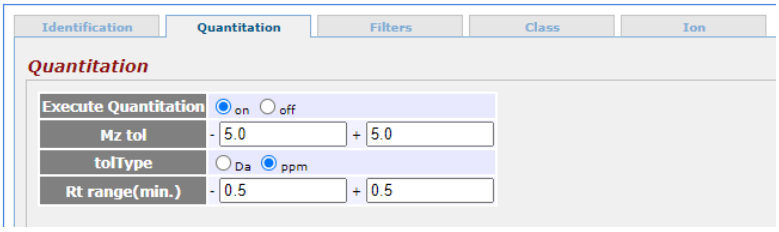


F-4: Overlap of porcine urine (red) and tear fluid (blue) compounds detected during analysis using reversed phase aqueous assay in positive ion mode

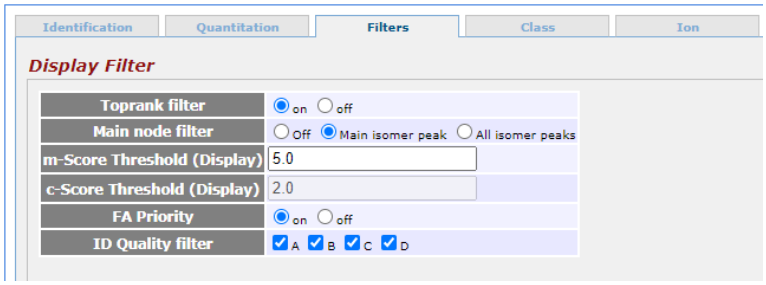
G-1: LipidSearch “Batch Search” settings applied for peak picking and annotation



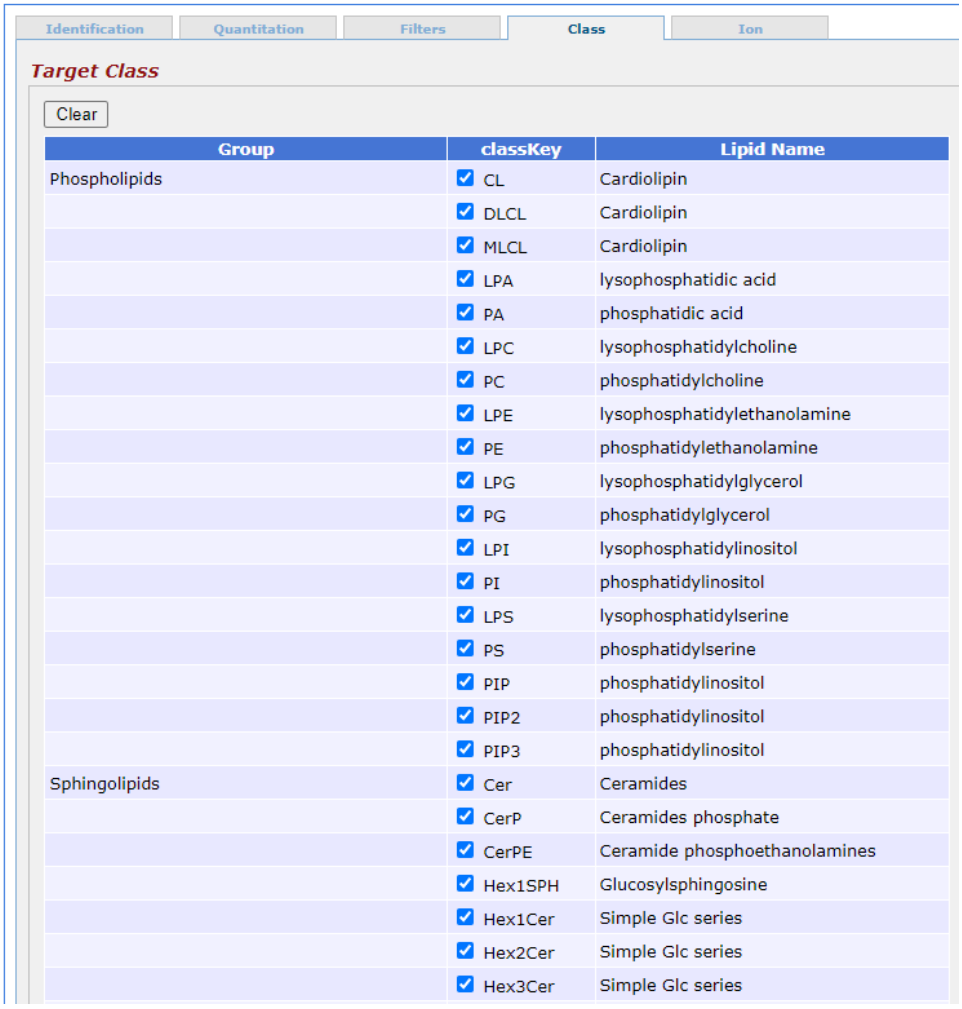
G-2: LipidSearch “Batch Search” settings applied for peak picking and annotation continued, with settings for Quantitation



G-3: LipidSearch “Batch Search” settings applied for peak picking and annotation continued, with settings for Display Filters



G-4: LipidSearch “Batch Search” settings applied for peak picking and annotation continued, with settings for the Target Classes (page 1 of 3)



Identification | Quantitation | Filters | Class | Ion

Target Class

Clear

Group	classKey	Lipid Name
Phospholipids	CL	Cardiolipin
	DLCL	Cardiolipin
	MLCL	Cardiolipin
	LPA	lysophosphatidic acid
	PA	phosphatidic acid
	LPC	lysophosphatidylcholine
	PC	phosphatidylcholine
	LPE	lysophosphatidylethanolamine
	PE	phosphatidylethanolamine
	LPG	lysophosphatidylglycerol
	PG	phosphatidylglycerol
	LPI	lysophosphatidylinositol
	PI	phosphatidylinositol
	LPS	lysophosphatidylserine
	PS	phosphatidylserine
	PIP	phosphatidylinositol
	PIP2	phosphatidylinositol
	PIP3	phosphatidylinositol
Sphingolipids	Cer	Ceramides
	CerP	Ceramides phosphate
	CerPE	Ceramide phosphoethanolamines
	Hex1SPH	Glucosylsphingosine
	Hex1Cer	Simple Glc series
	Hex2Cer	Simple Glc series
	Hex3Cer	Simple Glc series

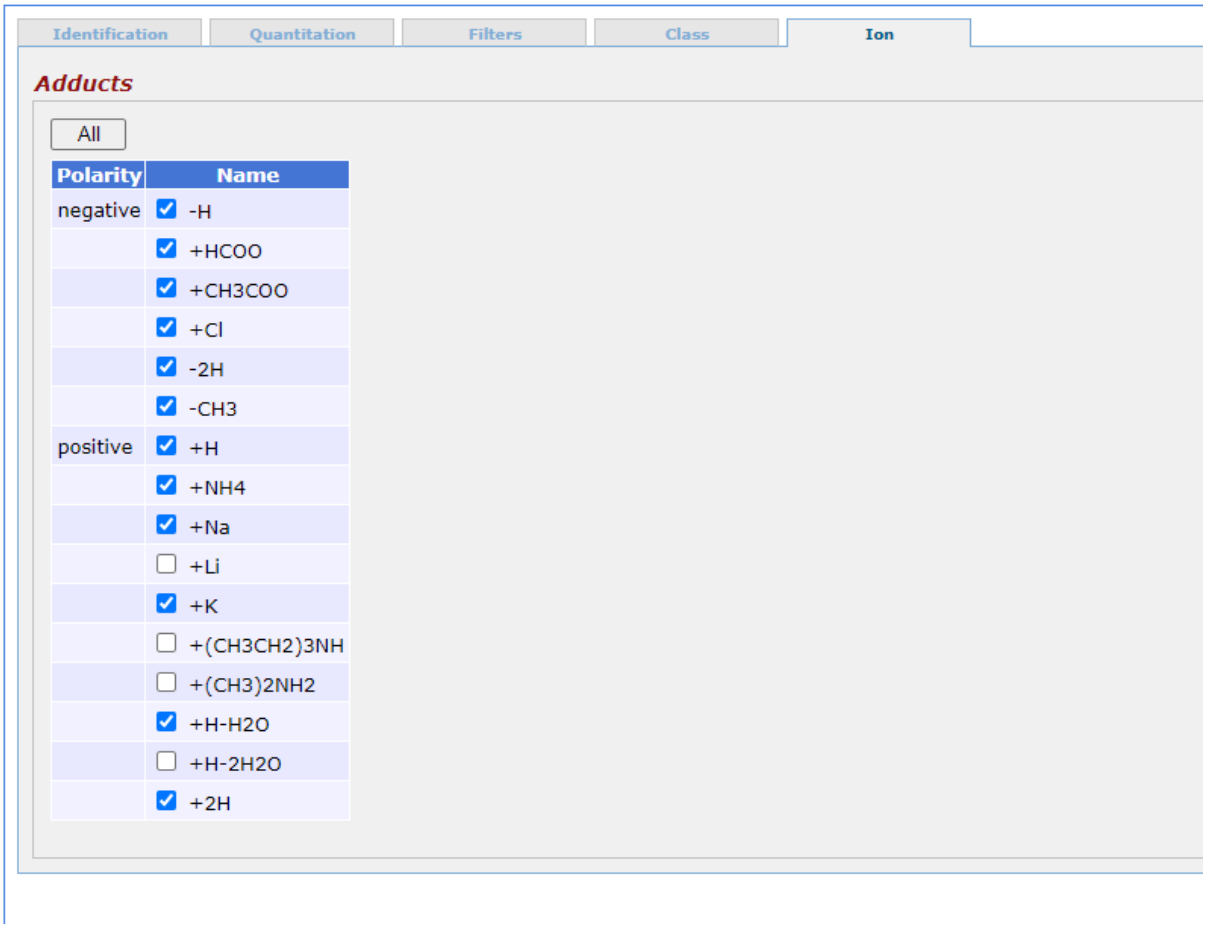
G-5: LipidSearch “Batch Search” settings applied for peak picking and annotation continued, with settings for the Target Classes (page 2 of 3)

	☑ CerG2GNAc1	Simple Glc series
	☑ CerG3GNAc1	Simple Glc series
	☑ CerG3GNAc2	Simple Glc series
	☑ GM3	Gangliosides
	☑ GM2	Gangliosides
	☑ GM1	Gangliosides
	☑ GD1a	Gangliosides
	☑ GD1b	Gangliosides
	☑ GD2	Gangliosides
	☑ GD3	Gangliosides
	☑ GT1a	Gangliosides
	☑ GT1b	Gangliosides
	☑ GT1c	Gangliosides
	☑ GT2	Gangliosides
	☑ GT3	Gangliosides
	☑ GQ1c	Gangliosides
	☑ GQ1b	Gangliosides
	☑ LSM	lysosphingomyelin
	☑ phSM	sphingomyelin(phytosphingosine)
	☑ SM	sphingomyelin
	☑ SPH	Sphingosine
	☑ SPHP	Sphingosine phosphate
	☑ ST	Sulfatide
Neutral lipids	☑ AcHexSiE	AcylGlcSitosterol ester
	☑ AcHexStE	AcylGlcStigmasterol ester
	☑ AcHexZyE	AcylGlcZymosterol ester
	☑ AcHexCmE	AcylGlcCampesterol ester
	☑ AcHexChE	AcylGlcCholesterol ester
	☑ ChE	Cholesterol Ester
	☑ D7ChE	Deuterated Cholesterol Ester
	☑ CmE	Campesterol Ester
	☑ DG	diglyceride
	☑ D5DG	Deuterated diglyceride
	☑ MG	monoglyceride
	☑ SiE	Sitosterol Ester

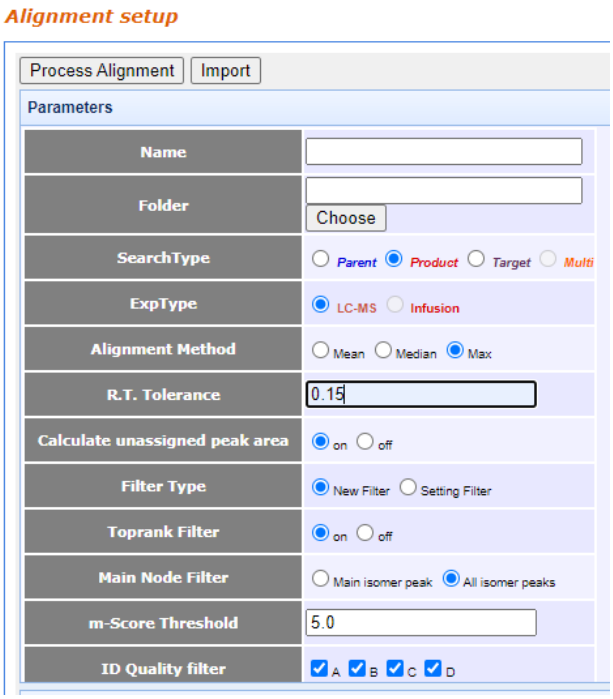
G-6: LipidSearch “Batch Search” settings applied for peak picking and annotation continued, with settings for the Target Classes (page 3 of 3)

	☑ StE	Stigmasterol Ester
	☑ TG	triglyceride
	☑ D5TG	Deuterated triglyceride
	☑ ZyE	Zymosterol Ester
Fatty acyl and other lipids	☑ AcCa	Acyl Carnitine
	☑ AEA	N-Acylethanolamine
	☑ Co	Coenzyme
	☑ cPA	cyclic phosphatidic acid
	☑ FA	fatty acid
	☑ LPEt	lysophosphatidylethanol
	☑ LPMe	lysophosphatidylmethanol
	☑ OAHFA	OAcyl-(gamma-hydroxy)FA
	☑ PAF	platelet-activating factor
	☑ PEt	phosphatidylethanol
	☑ PMe	phosphatidylmethanol
	☑ SL	Seminolipid
	☑ WE	wax esters
Glycoglycerolipids (plants)	☑ MGMG	Monogalactosylmonoacylglycerol
	☑ MGDG	Monogalactosyldiacylglycerol
	☑ DGMG	Digalactosylmonoacylglycerol
	☑ DGDG	Digalactosyldiacylglycerol
	☑ SQMG	Sulfoquinovosylmonoacylglycerol
	☑ SQDG	Sulfoquinovosyldiacylglycerol
Derivatized lipids (biotinylation, diazomethane)	☑ BiotinylPE	glycerophosphoethanolamine-N-(biotinyl)
	☑ BisMeLPA	Bis-methyl lysophosphatidic acid
	☑ BisMePA	Bis-methyl phosphatidic acid
	☑ BisMePE	Bis-methyl phosphatidylethanolamine
	☑ BisMePG	Bis-methyl phosphatidylglycerol
	☑ BisMePS	Bis-methyl phosphatidy lserine
	☑ MePC	Methyl phosphatidylcholine
	☑ dMePE	dimethylphosphatidylethanolamine
	☑ LdMePE	lysodimethylphosphatidylethanolamine

G-7: LipidSearch “Batch Search” settings applied for peak picking and annotation continued, with settings for the different possible adducts. When applying for a lipidomics assay, the acetate ion is not selected.



G-7: LipidSearch alignment settings applied for aligning together multiple annotated spectra into one large data matrix.



H-1: 10-fold cross- validation results that show the mean, minimum, maximum and median differences in retention time in both seconds and minutes related to the validation test metabolites used to test the retention time mapping model based off of the relationship between the Liverpool method 1 dataset and the University of Liverpool method 2 dataset comparison. Each difference is calculated in minutes.

	Mean Difference	Minimum Difference	Maximum Difference	Median Difference	Line Of Best Fit Equation	R^2 Value
Validation 1	0.004077	0.00586	1.374245	-0.05871	y = 0.9675x + 1.7001	0.9828
Validation 2	0.150288	0.008016	1.212515	0.059724	y = 0.968x + 1.7112	0.9836
Validation 3	-0.1087	0.043105	0.562319	-0.22452	y = 0.9614x + 1.7182	0.9837
Validation 4	0.049681	0.015261	0.869201	-0.0455	y = 0.9643x + 1.7194	0.9835
Validation 5	0.177096	0.021684	2.021006	-0.04353	y = 0.9686x + 1.7108	0.9851
Validation 6	-0.13892	0.09835	0.936836	-0.17905	y = 0.9639x + 1.7038	0.9847
Validation 7	0.17499	0.044833	1.411066	0.059917	y = 0.9603x + 1.7491	0.9855
Validation 8	-0.06796	0.026857	1.434402	-0.06757	y = 0.9635x + 1.7123	0.9843
Validation 9	0.156666	0.019754	1.667781	-0.04665	y = 0.9558x + 1.7648	0.985
Validation 10	-0.15709	0.031217	1.566506	-0.17712	y = 0.9557x + 1.7482	0.9867

H-2: 10-fold cross validation results that show the mean, minimum, maximum and median differences in retention time in both seconds and minutes related to the validation test metabolites used to test the retention time mapping model based off of the relationship between University of Birmingham dataset and the University of Liverpool (method 1) dataset. Each difference is calculated in minutes.

	Average Difference	Minimum Difference	Maximum Difference	Median Difference	Line Of Best Fit Equation	R^2 Value
Validation 1	0.139369	0.014949	0.657173	0.043746	$Y = -0.1601x^3 + 1.7054x^2 - 3.4786x + 2.7567$	0.969
Validation 2	-0.36634	0.018044	5.923608	-0.00291	$Y = -0.3213x^3 + 3.1824x^2 - 7.592x + 6.0814$	0.9747
Validation 3	0.120179	0.009154	0.859534	0.013024	$Y = -0.1833x^3 + 1.9786x^2 - 4.3798x + 3.5713$	0.9654
Validation 4	-0.30911	0.001654	5.103496	-0.08482	$Y = -0.2939x^3 + 2.908x^2 - 6.7169x + 5.222$	0.9853
Validation 5	0.064021	0.071581	1.048037	0.009375	$Y = -0.2038x^3 + 2.2083x^2 - 5.0986x + 4.1688$	0.9783
Validation 6	-0.0288	0.00312	1.014392	-0.13908	$Y = -0.1574x^3 + 1.6773x^2 - 3.3843x + 2.5868$	0.9683
Validation 7	0.034648	0.001088	0.963767	-0.05521	$Y = -0.1655x^3 + 1.7486x^2 - 3.5309x + 2.658$	0.9731
Validation 8	-0.2533	0.053208	4.879933	-0.06856	$Y = -0.2942x^3 + 2.9204x^2 - 6.7559x + 5.2512$	0.9609
Validation 9	0.086635	0.026395	0.954581	-0.05045	$Y = -0.1903x^3 + 2.0452x^2 - 4.5214x + 3.5868$	0.9664

Validation 10	-0.25543	0.021755	7.097603	0.0433	$Y = -0.3465x^3 + 3.3445x^2 - 7.7073x + 5.8207$	0.9826

H-3: 10-fold cross validation results that show the mean, minimum, maximum, and median differences in retention time in minutes related to the 10-fold cross-validation test metabolites used to test the retention time mapping model based off of the relationship between University of Birmingham dataset and the University of Liverpool Method 2 dataset. Each difference is calculated in minutes.

	Average Difference	Minimum Difference	Maximum Difference	Median Difference	Line Of Best Fit Equation	R^2 Value
Validation 1	-0.25181	0.001871	6.291139	-0.00026	$Y = -0.3207x^3 + 3.1143x^2 - 7.1236x + 5.4076$	0.98
Validation 2	0.12007	0.035398	0.677145	0.073484	$Y = -0.1614x^3 + 1.7243x^2 - 3.5602x + 2.8505$	0.9691
Validation 3	-0.41426	3.6E-05	5.160881	-0.18668	$Y = -0.2934x^3 + 2.9087x^2 - 6.7638x + 5.3006$	0.9789
Validation 4	0.144534	0.026811	1.045232	0.050619	$Y = -0.1882x^3 + 2.0231x^2 - 4.4588x + 3.5641$	0.9772
Validation 5	-0.00773	0.040559	1.075878	-0.15891	$Y = -0.2003x^3 + 2.1479x^2 - 4.7794x + 3.6789$	0.9756
Validation 6	-0.20439	0.004012	4.250293	0.028118	$Y = -0.2567x^3 + 2.5456x^2 - 5.6962x + 4.4571$	0.9758
Validation 7	-0.01406	0.009966	1.071427	0.028545	$Y = -0.1775x^3 + 1.9254x^2 - 4.2954x + 3.5613$	0.9752
Validation 8	0.144465	0.001903	0.987897	-0.00784	$Y = -0.1628x^3 + 1.7185x^2 - 3.4436x + 2.6345$	0.974
Validation 9	0.047513	0.029413	0.612385	0.044458	$Y = -0.1642x^3 + 1.7456x^2 - 3.5704x + 2.7587$	0.9642
Validation 10	-0.47192	0.012436	8.384623	-0.0528	$Y = -0.3683x^3 + 3.5564x^2 - 8.495x + 6.7487$	0.9683

H-4: 10-fold cross validation results that show the mean, minimum, maximum, and median differences in retention time between the mapped University of Liverpool method 1 metabolites and the measured retention time using the university of Florida method. All differences were measured in minutes.

	Mean	Minimum	Maximum	Median	Line Of Best Fit Equation	R^2 Value
Validation 1	-0.64312	6.84E-05	2.725097	-0.65656	$Y = 0.0112x^3 - 0.137x^2 + 0.875x + 5.3926$	0.9117
Validation 2	0.250806	0.047047	3.067657	0.248037	$Y = -0.0197x^3 + 0.446x^2 - 2.32x + 11.117$	0.8334
Validation 3	-0.81308	0.01428	4.00195	-0.36104	$Y = -0.0034x^3 + 0.058x^2 + 0.4196x + 5.153$	0.8342
Validation 4	0.494865	0.042001	3.373887	0.007501	$Y = -0.0129x^3 + 0.3871x^2 - 2.4867x + 12.119$	0.929
Validation 5	-0.28849	0.052918	2.39425	-0.42905	$Y = 0.0006x^3 + 0.0781x^2 - 0.4905x + 8.3083$	0.7543
Validation 6	0.337823	0.049035	3.519478	0.059202	$Y = -0.0127x^3 + 0.3209x^2 - 1.5336x + 9.266$	0.9311

Validation 7	-0.69342	0.02907	3.834445	-0.40691	$Y = -0.0039x^3 + 0.118x^2 - 0.4156x + 7.6971$	0.7239
Validation 8	0.067466	0.02782	2.727221	0.051997	$Y = 0.009x^3 - 0.1219x^2 + 1.0644x + 4.8344$	0.889
Validation 9	0.154679	0.194425	2.672967	-0.09101	$Y = -0.0127x^3 + 0.3537x^2 - 2.1712x + 11.506$	0.8218
Validation 10	-0.44991	0.061016	3.023666	-0.34486	$Y = 0.0037x^3 - 0.041x^2 + 0.7887x + 4.7098$	0.8843

H-5: 10-fold cross validation results that show the mean, minimum, maximum and median differences in retention time in minutes related to the validation metabolites used to test the retention time mapping model based off of the relationship between the university of Liverpool method 2 dataset and the university of Florida dataset. All differences were measured in minutes.

	Mean	Minimum	Maximum	Median	Line of Best Fit Equation	R^2 Value
Validation 1	0.543418	0.065853	4.003293	0.156233	$y = -0.0179x^3 + 0.4604x^2 - 2.5802x + 11.406$	0.9171
Validation 2	-0.462	0.00902	3.038444	-0.0325	$y = 0.0117x^3 - 0.2054x^2 + 1.6401x + 3.7689$	0.8462
Validation 3	-0.08655	0.054108	2.809872	0.166597	$y = 0.0091x^3 - 0.0657x^2 + 0.125x + 7.828$	0.8718
Validation 4	-0.14806	0.070582	3.238864	-0.55388	$y = -0.0127x^3 + 0.2989x^2 - 1.2357x + 8.258$	0.9178
Validation 5	-0.90162	0.019998	4.137017	-0.39392	$y = -0.0056x^3 + 0.1339x^2 - 0.3589x + 7.2645$	0.7509
Validation 6	0.579441	0.004348	3.741233	0.109248	$y = -0.0107x^3 + 0.3025x^2 - 1.5693x + 9.6135$	0.9326
Validation 7	0.457278	0.076219	2.623048	0.11969	$y = -0.0019x^3 + 0.1341x^2 - 0.7753x + 8.8675$	0.8449
Validation 8	-0.67426	0.090907	2.68628	-0.6287	$y = -0.0023x^3 + 0.1069x^2 - 0.3746x + 7.3434$	0.8706
Validation 9	-0.71524	0.169189	3.471418	-0.6775	$y = -0.0211x^3 + 0.5157x^2 - 3.1318x + 12.924$	0.7739
Validation 10	-0.0133	0.06291	3.2348	-0.20084	$y = 0.0145x^3 - 0.2647x^2 + 2.1937x + 2.2435$	0.9188

H-6: The metabolites from the University of Birmingham positive ion mode dataset and the University of Birmingham negative ion mode dataset comparison that were removed as outliers from the initial model.

Metabolite Name	**Cyclic Or Not**	**Metabolite Super-Class**
Epinephrine	Cyclic	Benzenoids
Glycoursodeoxycholic Acid	Cyclic	Lipids And Lipid-Like Molecules
Glycodeoxycholic Acid	Cyclic	Lipids And Lipid-Like Molecules
Glycochenodeoxycholic Acid	Cyclic	Lipids And Lipid-Like Molecules
Adenosine 5'-Diphosphate	Cyclic	Nucleosides, Nucleotides, And Analogues
Guanosine 5'-Diphosphate	Cyclic	Nucleosides, Nucleotides, And Analogues

N-Acetyl-L-Tyrosine	Cyclic	Organic Acids And Derivatives
N-Formyl-L-Methionine	Non-Cyclic	Organic Acids And Derivatives
3-Ureidopropionic Acid	Non-Cyclic	Organic Acids And Derivatives
2-Aminooctanoic Acid	Non-Cyclic	Organic Acids And Derivatives
L-Histidinol	Cyclic	Organic Nitrogen Compounds
4-Hydroxybenzaldehyde	Cyclic	Organic Oxygen Compounds
Chlorogenic Acid	Cyclic	Organic Oxygen Compounds
Nizatidine	Cyclic	Organoheterocyclic Compounds
Cimetidine	Cyclic	Organoheterocyclic Compounds
Famotidine	Cyclic	Organoheterocyclic Compounds
1-Methyluric Acid	Cyclic	Organoheterocyclic Compounds
Nicotinic Acid	Cyclic	Organoheterocyclic Compounds
Pyridoxamine	Cyclic	Organoheterocyclic Compounds
Amoxicillin	Cyclic	Organoheterocyclic Compounds
Doxycycline Monohydrate	Cyclic	Phenylpropanoids And Polyketides
Rutin	Cyclic	Phenylpropanoids And Polyketides

H-7: 10-fold cross validation results that show the mean, minimum, maximum and median differences in retention time in both seconds and minutes related to the validation test metabolites used to test the retention time mapping model based off of the relationship between University of Birmingham positive ion mode dataset and the University of Birmingham negative ion mode dataset. All differences are calculated in minutes.

	Mean Difference	Minimum Difference	Maximum Difference	Median Difference	Line Of Best Fit Equation	R^2 Value
Validation 1	0.163367	0.049891	0.869014	0.148829	Y = 0.9183x + 0.7626	0.904
Validation 2	0.108508	0.004907	0.685097	0.066124	Y = 0.9925x + 0.3096	0.9513
Validation 3	-0.04331	0.017376	0.982364	0.095996	Y = 0.9941x + 0.2819	0.9566
Validation 4	-0.01271	0.016342	1.164223	0.071685	Y = 0.9944x + 0.2835	0.9553
Validation 5	0.046062	0.012038	1.024039	0.119517	Y = 0.9916x + 0.306	0.9548
Validation 6	-0.1118	0.012018	1.054174	-0.02295	Y = 0.9898x + 0.2977	0.9536
Validation 7	-0.05024	0.049423	0.816958	-0.01009	Y = 0.9867x + 0.3226	0.9519
Validation 8	0.078156	0.025042	0.686804	0.201493	Y = 0.9339x + 0.6918	0.9167
Validation 9	0.001976	0.01972	1.038022	0.124506	Y = 0.9994x + 0.2575	0.9553
Validation 10	0.145074	0.010552	0.661932	0.128355	Y = 0.9811x + 0.3762	0.9484

H-8: The classes and structural similarities between the overlapping metabolites in the University of Birmingham negative ion mode dataset and the Fiehn MoNA dataset that are plotted above the line of best fit in Figure 4-10 and below the line of best fit in Figure 4-10.

Above Line Of Best Fit			Under Line Of Best Fit		
Metabolite Name	Metabolite Super-Class	Cyclic Or Non-Cyclic	Metabolite Name	Metabolite Super-Class	Cyclic Or

					Non-Cyclic
Olmesartan Medoxomil	Benzenoids	Cyclic	Gentisic Acid	Benzenoids	Cyclic
2-[(4-Hydroxyphenyl)Formamido]Acetic Acid	Benzenoids	Cyclic	5-Aminosalicylic Acid	Benzenoids	Cyclic
Dopamine	Benzenoids	Cyclic	Syringic Acid	Benzenoids	Cyclic
3-Methyl-2-Oxobutanoic Acid	Lipids And Lipid-Like Molecules	Non-Cyclic	4-Hydroxyphenylpyruvic Acid	Benzenoids	Cyclic
3-Hydroxyisovaleric Acid	Lipids And Lipid-Like Molecules	Non-Cyclic	3-Hydroxyphenylacetic Acid	Benzenoids	Cyclic
Thymidine	Nucleosides, Nucleotides, And Analogues	Cyclic	Phthalic Acid	Benzenoids	Cyclic
2'-Deoxyadenosine	Nucleosides, Nucleotides, And Analogues	Cyclic	DL -3,4-Dihydroxymandelic Acid	Lipids And Lipid-Like Molecules	Cyclic
5-Methyluridine	Nucleosides, Nucleotides, And Analogues	Cyclic	2-Hydroxy-3-Methylbutyric Acid	Lipids And Lipid-Like Molecules	Non-Cyclic
Adenosine	Nucleosides, Nucleotides, And Analogues	Cyclic	2'-Deoxyuridine	Nucleosides, Nucleotides, And Analogues	Cyclic
Uridine	Nucleosides, Nucleotides, And Analogues	Cyclic	Valsartan	Organic Acids And Derivatives	Cyclic
2'-Deoxycytidine	Nucleosides, Nucleotides, And Analogues	Cyclic	N-Acetyl-L-Phenylalanine	Organic Acids And Derivatives	Cyclic
Inosine	Nucleosides, Nucleotides, And Analogues	Cyclic	N-Acetyl-L-Leucine	Organic Acids And Derivatives	Non-Cyclic
2'-Deoxyguanosine	Nucleosides, Nucleotides, And Analogues	Cyclic	N-Iso Valerylglycine	Organic Acids And Derivatives	Non-Cyclic
L-Tryptophan	Nucleosides, Nucleotides, And Analogues	Cyclic	N-Acetyl-L-Methionine	Organic Acids And Derivatives	Non-Cyclic
Cytidine	Nucleosides, Nucleotides, And Analogues	Cyclic	N-Acetyl-L-Alanine	Organic Acids And Derivatives	Non-Cyclic
Guanosine	Nucleosides, Nucleotides, And Analogues	Cyclic	Succinic Acid	Organic Acids And Derivatives	Non-Cyclic
N-Acetyl-D-Mannosamine	Nucleosides, Nucleotides, And Analogues	Cyclic	3-Ureidopropionic Acid	Organic Acids And Derivatives	Non-Cyclic
N-Acetyl-L-Cysteine	Organic Acids And Derivatives	Non-Cyclic	DL-Threitol	Organic Oxygen Compounds	Non-Cyclic
L-Norleucine	Organic Acids And Derivatives	Non-Cyclic	Ketorolac	Organic Oxygen Compounds	Cyclic

L-Leucine	Organic Acids And Derivatives	Non-Cyclic	Indole-3-Acetic Acid	Organoheterocyclic Compounds	Cyclic
N-Propionylglycine	Organic Acids And Derivatives	Non-Cyclic	Indole-3-Pyruvic Acid	Organoheterocyclic Compounds	Cyclic
3-Nitro-L- Tyrosine	Organic Acids And Derivatives	Cyclic	Pyrrole-2-Carboxylic Acid	Organoheterocyclic Compounds	Cyclic
Nicotinuric Acid	Organic Acids And Derivatives	Cyclic	DL-Indole-3-Lactic Acid	Organoheterocyclic Compounds	Cyclic
Ranitidine	Organic Nitrogen Compounds	Cyclic	5-Hydroxyindole-3-Acetic Acid	Organoheterocyclic Compounds	Cyclic
L-Kynurenine	Organic Oxygen Compounds	Cyclic	Nicotinic Acid	Organoheterocyclic Compounds	Cyclic
D- (-)-Salicin	Organic Oxygen Compounds	Cyclic	Biotin	Organoheterocyclic Compounds	Cyclic
Aripiprazole	Organoheterocyclic Compounds	Cyclic	1-Methyluric Acid	Organoheterocyclic Compounds	Cyclic
Pyridoxine	Organoheterocyclic Compounds	Cyclic	Urocanic Acid	Organoheterocyclic Compounds	Cyclic
Nizatidine	Organoheterocyclic Compounds	Cyclic	(+)-Catechin	Phenylpropanoids And Polyketides	Cyclic
Carvedilol	Organoheterocyclic Compounds	Cyclic	Trans-4-Hydroxycinnamic Acid	Phenylpropanoids And Polyketides	Cyclic
Adenine	Organoheterocyclic Compounds	Cyclic	Sinapic Acid	Phenylpropanoids And Polyketides	Cyclic
Famotidine	Organoheterocyclic Compounds	Cyclic	Myricetin	Phenylpropanoids And Polyketides	Cyclic
Hypoxanthine	Organoheterocyclic Compounds	Cyclic			
Amlodipine	Organoheterocyclic Compounds	Cyclic			
5-Methylcytosine	Organoheterocyclic Compounds	Cyclic			
Guanine	Organoheterocyclic Compounds	Cyclic			
Amiloride	Organoheterocyclic Compounds	Cyclic			
Riboflavin	Organoheterocyclic Compounds	Cyclic			
Penciclovir	Organoheterocyclic Compounds	Cyclic			
Isoxanthopterin	Organoheterocyclic Compounds	Cyclic			
Naringin	Phenylpropanoids And Polyketides	Cyclic			
Kynurenic Acid	Phenylpropanoids And Polyketides	Cyclic			
(-)-Epicatechin	Phenylpropanoids And Polyketides	Cyclic			

H-9: 10-fold cross validation results that show the mean, minimum, maximum and median differences in retention time in both seconds and minutes related to the validation test metabolites used to test the

retention time mapping model based off of the relationship between University of Birmingham positive ion mode dataset and the Fiehn MoNA dataset. All differences are calculated in minutes.

	Mean difference	Minimum difference	Maximum difference	Median difference	Line of Best Fit Equation	R^2 Value
Validation 1	0.000325	0.019378	0.702485	-0.08124	y = 0.9103x + 2.4913	0.942
Validation 2	0.019745	0.021306	0.681405	-0.0412	y = 0.9127x + 2.4817	0.9444
Validation 3	0.12104	0.023129	1.636446	-0.04934	y = 0.9226x + 2.4476	0.9668
Validation 4	0.063912	0.001769	1.197453	-0.04933	y = 0.9035x + 2.5382	0.944
Validation 5	-0.11171	0.071124	0.691301	-0.19258	y = 0.9233x + 2.4025	0.9436
Validation 6	0.066312	0.000424	0.947938	-0.01145	y = 0.9208x + 2.4475	0.9479
Validation 7	-0.08721	0.027569	1.021405	-0.15376	y = 0.9352x + 2.3438	0.9462
Validation 8	0.053901	0.000838	1.640522	-0.07684	y = 0.9054x + 2.5266	0.9516
Validation 9	-0.0483	0.009214	1.263823	-0.08122	y = 0.9066x + 2.5022	0.9511
Validation 10	-0.06134	0.062337	0.688622	-0.15842	y = 0.9113x + 2.475	0.9416

H-10: 10-fold cross validation results that show the mean, minimum, maximum, and median differences in retention time in both seconds and minutes related to the validation test metabolites used to test the retention time mapping model based off of the relationship between University of Liverpool dataset and the Fiehn MoNA dataset. All differences are calculated in minutes.

	Mean difference	Minimum difference	Maximum difference	Median difference	Line of Best Fit Equation	R^2 Value
Validation 1	-0.14463	0.02848	1.186603	-0.05101	$y = 0.0531x^3 - 1.0191x^2 + 6.839x - 10.275$	0.8702
Validation 2	-0.02772	0.006622	0.903768	-0.02029	$y = 0.0328x^3 - 0.5922x^2 + 3.9438x - 3.9445$	0.8596
Validation 3	0.03247	0.037081	2.170251	0.046134	$y = 0.0099x^3 - 0.0794x^2 + 0.207x + 4.8464$	0.8496
Validation 4	0.111856	0.005722	0.853506	-0.01523	$y = 0.038x^3 - 0.6793x^2 + 4.4212x - 4.7961$	0.8541
Validation 5	0.052371	0.029553	1.681822	-0.08669	$y = 0.0273x^3 - 0.445x^2 + 2.66x - 0.3373$	0.8531
Validation 6	-0.05503	0.039647	0.835431	-0.12641	$y = 0.0487x^3 - 0.8923x^2 + 5.7636x - 7.5157$	0.8085
Validation 7	0.064873	0.001017	1.120371	0.012087	$y = 0.0462x^3 - 0.8521x^2 + 5.5329x - 6.9172$	0.8633
Validation 8	0.170146	0.066702	2.089876	0.134023	$y = 0.0051x^3 + 0.0138x^2 - 0.3133x + 5.636$	0.8416
Validation 9	0.019496	0.029362	0.885236	-0.05182	$y = 0.036x^3 - 0.6696x^2 + 4.5611x - 5.5293$	0.8654
Validation 10	0.060913	0.023329	1.090231	0.037023	$y = 0.0227x^3 - 0.3345x^2 + 1.8414x + 1.5468$	0.8463

I-1: Table of Fatty acyls and glycerophospholipids that had a statistical significance of $P<0.05$ when comparing their relative intensity at rest vs active recovery in a paired t-test. Fold change increases in relative concentration are highlighted green, whilst any fold change decreases in relative concentration

are highlighted orange. The glycerophospholipids are denoted by Class (number of carbons: number of double bonds). GPS = Glycerophosphoserines, GPC = Glycerophosphocholines, LysoGPC=lysoGlycerophosphocholines, GPE/ GPE-NMe/ GPE-NMe2 = Glycerophosphoethanolamines, CDP-DG = cytidine diphosphate diacylglycerol

Lipid Class	Compound Name	Fold Change	Raw P-Value
Fatty Acyls	Linolenic acid	2.32	0.086806
Fatty Acyls	Tetradecadiencarnitine	2.693458	0.009535
Fatty Acyls	Oxohexacosyl-CoA; Oxohexacosanoyl-CoA	0.434631	0.053457
Fatty Acyls	Hexadecadienoylcarnitine	2.118554	0.022491
Fatty Acyls	Decenoylcarnitine	2.875133	0.031574
Fatty Acyls	Acetylcarnitine	2.241147	0.002357
Fatty Acyls	Butyrylcarnitine; Carnitine; Isobutyrylcarnitine; (tert-Butoxycarbonyl)leucine	2.531325	0.038254
Fatty Acyls	Arachidonoylserine	2.424478	0.062491
Fatty Acyls	Oleic acid; Octadecenoic acid; Cycloheptylundecanoic acid; Methylheptadecenoic acid; Vaccenic acid; Elaidic Acid; Ethylhexadecenoate; Isooleic acid; Petroselinic acid	2.45032	0.002383
Fatty Acyls	Propionylcarnitine; Carnitine	3.417285	0.009186
Glycerophospholipids	GPS(36:3)	0.464253	0.046944
Glycerophospholipids	CDP-DG(36:4)	0.45159	0.086054
Glycerophospholipids	LysoGPC(18:2/0:0);	2.458452	0.010566
Glycerophospholipids	GPC(33:1);PE(36:1);GPE-NMe(35:1);GPE-NMe2(34:1)	0.331422	0.019463
Glycerophospholipids	GPC(O-38:0)	2.248606	0.041185
Glycerophospholipids	GPE(34:2);GPE-NMe(33:2);GPE-NMe2(32:2)	2.05225	0.020393
Glycerophospholipids	GPE(35:3);GPE-NMe(34:3);GPE-NMe2(33:3); GPC(32:3)	2.125941	0.087495

I-2: Lipids that had a statistical significance of P<0.05 when comparing their relative intensity at rest vs passive recovery in a paired t-test. Fold change increases in relative concentration are highlighted green, whilst any fold change decreases in relative concentration are highlighted orange. The glycerophospholipids, Glyceroplipids, and Phosphosphingolipids are denoted by Class (number of carbons: number of double bonds). GPC= Glycerophosphocholines, GPA=Glycerophosphates, GPE/ GPE-NMe/ GPE-NMe2 = Glycerophosphoethanolamines, TG=Triradylglycerolipids, GPS= Glycerophosphoserines, SM= Sphingomyelins.

Lipid Class	Compound name	Fold Change	Raw P Value
Fatty Acyls	Hydroxydodecanoic acid; Hydroxylauric acid	2.61	0.041272
Fatty Acyls	Tetradecadiencarnitine	3.37	0.014655
Fatty Acyls	Acetyl-carnitine	2.18	0.010858
Fatty Acyls	Tetradecenoylcarnitine	2.06	0.04024
Fatty Acyls	Nervonic acid	2.04	0.048509
Fatty Acyls	Oleic acid	2.55	0.0080356
Fatty Acyls	Palmitic Acid	2.60	0.0012451
Glycerophospholipids	GPA(20:4)	0.40	0.02021
Glycerophospholipids	GPC(32:2);GPE(35:2);GPE-NMe(34:2);GPE-NMe2(33:2)	0.39	0.038123
Glycerophospholipids	GPC(34:1);GPE(37:1);GPE-NMe(36:1);GPE-NMe2(35:1)	0.11	0.0083928
Glycerophospholipids	GPE(38:5)	3.62	0.031369
Glycerophospholipids	GPS(36:3)	0.42	0.016397
Glycerolipids	TG(60:8)	4.17	0.013404
Phosphosphingolipids	SM(d18:1/26:0)	0.47477	0.017226

I-3: The fold changes for compounds that were found to be statistically significant in t-tests for Starting Rest (T1) vs Passive Recovery (T5) time points from the HIIT protocol. All fold changes had a statistical significance of P<0.05 and increases in relative concentration are highlighted in green, whilst any fold change decreases are highlighted in orange. The glycerolipids and glycerophospholipids are denoted by Class (number of carbons: number of double bonds). PA=Glycerophosphates, PE/ PE-NMe/ PE-NMe2 = Glycerophosphoethanolamines, LysoPE= Glycerophosphoethanolamines, PS= Glycerophosphoserines, TG= Triradylglycerolipids, DG= Diradylglycerolipids, MG= Monoradylglycerolipids.

Compound Class	Compound Name	Fold Change	raw p-value
Fatty acyl	Palmitic Acid;Lauryl isobutyrate;Hexyl decanoate; Trimethyltridecanoic acid;Dodecyl butyrate; Methylpentadecanoic acid;Heptylnonanoic acid; Butylethyldecanoic acid;Dimethyltetradecanoic acid; Propyltridecanoic acid;Ethylmethyltridecanoic acid; Ethyltetradecanoic acid	0.45	4.98E-02
Fatty acyl	9-Decenoylcarnitine	2.32	4.87E-02
Fatty acyl	3-hydroxyoctanoylcarnitine	2.48	9.91E-03
Glycerolipids	MG(24:1)	0.34	4.75E-02
Glycerolipids	TG(52:2)	0.20	3.75E-02
Glycerolipids	DG(38:4)	3.35	1.67E-02
Glycerophospholipids	PE(34:2);PE-NMe(33:2);PE-NMe2(32:2)	2.78	2.29E-02
Glycerophospholipids	PS(38:3)	0.46	1.72E-02
Glycerophospholipids	LysoPE(18:1)	2.16	9.67E-03
Glycerophospholipids and Glycerolipids	PA(33:0);DG(40:9)	3.37	4.24E-03
Prenol lipids	2-hydroxy-6-[(8Z,11Z)-pentadeca-8,11,14-trien-1-yl]benzoic acid	0.39	3.88E-02
Water Soluble	Ubiquinol-10	0.46	3.99E-02
Water Soluble	Methoxyestratetraene; [(Methylphenyl)hexanyl]phenol	0.47	3.55E-02
Water Soluble	Phyllanthusol B	2.40	3.03E-02
Water Soluble	Armillarin;Armillaripin;Clausarinol;Eplerenone;O-Methylmelleolide; hydroxy(hydroxymethylbutanoyl)dimethylpentylpyranochromenone; hydroxy(hydroxypentyl)dimethyl(methylbutanoyl)pyranochromenone; hydroxy(ydroxypentyl)dimethyl(methylbutanoyl)pyranochromenone; hydroxy[(hydroxymethyl)butanoyl]dimethylpentylpyranochromenone; hydroxy(hydroxymethyl)methyl(methylbutanoyl)pentylpyranochromenone; ({[(dihydroxydimethylnonadienyl]methyloxiranyl}methoxy)chromenone; {[(dimethyloxiranyl)dihydroxydimethylnonadienyl]oxy}chromenone; {[dihydroxy(hydroxymethyl)dimethyldodecatrienyl]oxy}chromenone; {[dihydroxytrimethyldodecatrienyl]oxy}hydroxychromenone; {[hydroxy[(hydroxymethylpentenyl)methyloxiranyl]methylpentenyl]oxy}chromenone; {[trihydroxytrimethyldodecatrienyl]oxy}chromenone;dihydroxy-(hydroxymethyl)dimethyldodecatrienyl]oxy}chromenone	0.36	2.67E-02
Water Soluble	(Trihydroxytetramethyldecahydronaphthalenyl)butenone	2.46	2.57E-02
Water Soluble	({[hydroxydimethylnonatetraenyl]dimethylcyclohexenyl}methoxy)sulfonic acid; ({methyl[methyl(trimethylcyclohexenyl)hexatrienyl]oxiranyl}methoxy)sulfonic acid; {[(hydroxytrimethylcyclohexenyl)dimethylnonatetraenyl]oxy}sulfonic acid; {[[(hydroxymethyl)dimethylcyclohexenyl]dimethylnonatetraenyl]oxy}sulfonic acid; {[{[methyl(trimethylcyclohexenyl)butadienyl]oxiranyl}butenyl]oxy}sulfonic acid;nonatetraenyl]oxy}sulfonic acid; {[dimethyloxo(trimethylcyclohexenyl)nonatrienyl]oxy}sulfonic acid; {[hydroxy[methyl(trimethylcyclohexenyl)hexatrienyl]butenyl]oxy}sulfonic acid; {[hydroxydimethyl(trimethylcyclohexenyl)nonatetraenyl]oxy}sulfonic acid; {[hydroxydimethylnonatetraenyl]trimethylcyclohexenyl}oxidanesulfonic acid; {[hydroxymethyl[(trimethylcyclohexenyl)ethenyl]octatrienyl]oxy}sulfonic acid; {[methyl[(trimethylcyclohexenyl)oxiranyl]octatrienyl]oxy}sulfonic acid; {[methyl{methyl[(trimethylcyclohexenyl)ethenyl]oxiranyl}pentadienyl]oxy}sulfonic acid	0.43	2.54E-02
Water Soluble	Phe-Phe;Chloropanaxydiol	2.71	2.45E-02
Water Soluble	(22Z)-19-Oxo-22-octacosenoic acid	2.57	2.35E-02
Water Soluble	LysoPE(18:1)	2.94	2.13E-02
Water Soluble	Congmunoside XIII	2.45	1.39E-02
Water Soluble	Oxagrelate	0.37	1.26E-02
Water Soluble	Histidine; Pyrazole-1-alanine	3.99	1.21E-02
Water Soluble	biotin	3.28	6.18E-03
Water Soluble	L-alpha-Aspartyl-L-phenylalanine	0.20	5.34E-03
Water Soluble	(hydroxymethoxydimethyldihydrobenzopyranyl)oxidanesulfonic acid; (methoxyphenyl)(sulfooxy)pentanoic acid;[methoxy(sulfooxy)phenyl]pentanoic acid; {[(methoxyphenyl)methoxy]oxobutoxy}sulfonic acid; ({[(methoxyphenyl)methoxy]oxobutanyl}oxy)sulfonic acid	2.13	4.85E-04