# Metabolic profiling and pathway mapping of cardiovascular disease

Thesis submitted by

*Panagiotis Vorkas*

for the degree of Doctor of Philosophy (PhD)
of Imperial College London
in Biochemistry

Department of Surgery and Cancer
Faculty of Medicine
Imperial College London

2014

# *Abstract*

In this thesis, metabolic profiling (MP) platforms were utilised to interrogate the manifestation of cardiovascular disease and provide candidate biomarkers. A number of LC-MS and NMR methodologies were employed. Data processing was followed by assessment using multivariate (MVDA) and univariate (UV) statistics. MP is applied under three cardiovascular disease themes: 1) plaque rupture, 2) plaque formation, and 3) arterial ectopic calcification. Statistically significant features were structurally assigned. Identified metabolites were mapped to their corresponding biochemical pathways.

For MP of ruptured plaque, tissue from symptomatic and asymptomatic patients for stroke was used. After detection of statistically significant features and structural assignment, two biochemical pathways showed dysregulations: the arachidonic acid pathway, indicating increased levels of inflammation, and the β-oxidation pathway with increased levels of three acyl-carnitines.

Tissue extracts were used to investigate plaque formation. Arterial intima tissue, incorporating plaque lesions (carotid and femoral), was compared to intimal thickening tissue. Intima thickening demonstrated distinct MP compared to plaques. Plaques from different anatomical locations also demonstrated altered MP. After metabolite assignment, pathway mapping revealed dysregulations common to both anatomical locations. These were cholesterol, ceramide, purine, pyrimidine and β-oxidation pathways. These pathways are related to inflammation and apoptosis. A metabolite previously unassociated to atherogenesis was detected with strong statistical significance (t-test; $p \geq 9.8 \times 10^{-12}$), namely phosphatidylethanolamine-ceramide. It also demonstrated high correlations to cholesterol, a well-established risk-factor of atherosclerosis.

The third theme of the project explores ectopic cardiovascular calcification. Experiments were conducted on blood serum. Patients with coronary artery and aortic valve calcification were compared with non-calcified controls. Phosphatidylcholine moieties and sphingomyelins were the major discriminating metabolites between cases and controls. These are involved in inflammation and apoptosis. The two diseases manifested different profiles with only three commonly dysregulated metabolites.

A number of experiments using additional samples and bottom-up approaches will follow to provide validation of results.

# *Table of Contents*

To everyone who helped me accomplish this,

mentors, family and friends.

# *Acknowledgments*

I would like to thank my supervisors Prof. Elaine Holmes, Dr Elizabeth J Want and Prof. Jeremy K Nicholson, for their supervision and guidance. I am grateful for the opportunity they offered me, and the chance to work in their lab and interact with them, but also with scientists possessing invaluable expertise, ready to share and help each and every one of us in order advance.

I would like to acknowledge the Royal Society of Chemistry (RSC) for supporting my PhD studentship. My PhD studentship was supported through an RSC award to my supervisor Prof Elaine Holmes. For this I am thankful as it gave me the chance to focus on my PhD and accomplish maximum potential.

Special thanks to all the people who took the time to teach and assist during the first steps I took learning to use mass spectrometry (MS). Despite my previous experience in chromatography, MS was a totally new world, that only with proper guidance one can fully explore. Many many thanks to Matt Lewis, the person who introduced me to the world of MS. I am thankful to Dina Spagou and Alex Pechlivanis for their help with the UPLC-MS instrumentation. The guys were always there and ready to assist with troubleshooting. They were always keen to share their experiences, and this helped all of us to progress.

A big thank you to our XCMS man, Paul Benton. Paul was always there for all of us and always ready to help with informatics and coding. He was never bored to instantly put together a custom made code to fit our needs. Above all I am thankful for introducing me to the world of surf, and being my surferdude!

Many thanks to Renaud Mestdagh and Olaf Beckonert for teaching me the proper NMR spectroscopy experimental setup, as well as data processing. I am also grateful to Claire Boulange for her help and informative explanations on multivariate statistics.

I would like to thank Waters Corporation for offering their instrumentation for analysing my samples. Many thanks to Giorgis Isaac for providing his lipid profiling method and assistance while running my samples in Milford. Many thanks to the rest of the team there: Henry Shion, Steve McDonalds and Alan Miller. Last but definitely not least; I am very thankful to John Shockcor. Thank you John for being such a good friend and so Greek. I will never forget your hospitality and incredible time we had at Milford. Many thanks for all the discussions we had, science oriented or not. I've learned so much from you! I am very grateful to you for your explanations on triglyceride identification. It was the first lesson for the endless world of lipids.

I would also like to thank my collaborators from the clinic, providing samples for my studies. Prof. Alun Davies, Muzaffar Anwar and Joseph Shalhoub for the plaque projects. Many thanks to Prof. Michael Henein for the samples for studying calcific disease. A big, huge thanks to all the patients who even on a time of distress showed greatness and offered the specimens for the greater good of the society.

Many thanks to several friends in my lab. They all assisted one way or another especially by exchanging ideas and knowledge. Although I will definitely forget a lot of you guys I will attempt to put some names down: Jasmina Saric, Jon Bowen, Jia Li, Anisha Wijeyesekera, Gabriel Valbuena, Yotsawat Pomyen, Michael Kyriakides, Florian Geier, Lea Maitre....

I had some amazing times with friends and family from the lab. Kathy, Charis, little Vassilis, Matt, Dina, Alex, Renaud, Claire, Paul, Gabriel Valbuena, Michael, Gianna, Florian, Lea, Jessie Miller. We had some amazing times dudes; in London, but also Denver, Las Vegas, San Diego, Vancouver, Milford, Boston, Greece....

I can't thank enough my parents for their support since my first steps in science. They were always there and always patient. Thank you for helping me up when I was down, and holding me down when I was dangerously flying in the clouds. A big kiss to all my siblings; in order of appearance Maria, Constantinos, Kyriakos, Melpo and Varnavas. Lots of kisses to my niece Valeria and nephew Christos; my brothers-in-law Giannis and Achilleas and my sister-in-law Chrysi. The times we have back home are incredible and whenever you came to visit was just another taste of home for an eternally homesick man.

To my friends, older friends (some since kindergarten), or more recent; you guys were my inspiration to progress in my studies and achieve levels I never thought I could reach.

A huge kiss to my girlfriend Anita for her support, for putting me in order and helping me reach a bearable level of maturity, as I would always leave the child inside me take over. Above all I want to thank her for her love.

My PhD journey ends with numerous experiences, collaborators and friends, but above all a tremendous amount of acquired knowledge.

*Panagiotis A. Vorkas*

*London, February 2014*

# *Declaration of Originality*

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

# *Copyright Declaration*

# *List of figures*

# *List of tables*

## *List of Abbreviations*

| | |
|---|---|
| 2-D | two-dimensional |
| AA | arachidonic acid |
| AcC | acyl-carnitine |
| ACN | acetonitrile |
| AmAc | ammonium acetate |
| APCI | atmospheric pressure chemical ionisation |
| Asympt | asymptomatic carotid plaque tissue |
| CAR | carotid plaque tissue |
| CAVD | calcific aortic valve disease |
| CCAD | calcific coronary artery disease |
| CE | cholesterol ester |
| Cer | ceramide |
| Cho | cholesterol |
| CM | cell membrane |
| COSY | correlation spectroscopy |
| CS | calcium Score |
| CV-ANOVA | cross validation - analysis of variance |
| CVC | cardiovascular calcification |
| CVD | cardiovascular disease |
| DA | discriminant analysis |
| DCM | dichloromethane |
| DG | diglyceride |
| EC | epithelial cell |
| ESI | electrospray ionisation |
| FA | formic acid |
| FAC | fatty acyl chain |
| FEM | femoral plaque tissue |
| FFA | free fatty acid |
| FID | free induction decay |
| GC | gas chromatography |

| | |
|---|---|
| HexCer | hexosyl-ceramide |
| HILIC | hydrophilic interaction (liquid) chromatography |
| HPLC | high performance liquid chromatography |
| INT | intimal thickening tissue |
| LC-MS | liquid chromatography coupled to mass spectrometry |
| LoA | level of assignment |
| m/z | mass-to-charge ratio |
| MALDI | matrix-assisted laser desorption/ionisation |
| MeOH | methanol |
| MP | metabolic profiling |
| MS | mass spectrometry *or* mass spectrometer |
| MSI | mass spectrometry imaging |
| MVDA | multivariate data analysis |
| NMR | nuclear magnetic resonance spectroscopy |
| OAc | acetic acid |
| o-PLS | orthogonal - projection to latent structures |
| oTOF | orthogonal-time of flight |
| oxCE | oxidised cholesterol ester |
| PC | phosphatidylcholine |
| PCA | principal component analysis |
| PComp | principal component |
| PE | phosphatidylethanolamine |
| PE-Cer | phosphatidylethanolamine-ceramide |
| PG | phosphatidylglycerol |
| PI | phospatidylinositol |
| PLS | partial least squares; projection to latent structures* |
| ppm | parts per million |
| PS | phosphatidylserine |
| qTOF | quadrupole coupled to time-of-flight |
| R | resolution |
| RF | radiofrequency |
| RP | reversed phase |

| | |
|---|---|
| SM | sphingomyelin |
| SPE | solid phase extraction |
| Sympt | symptomatic carotid plaque tissue |
| TG | triglyceride |
| TOF | time-of-flight |
| UPLC | ultra performance liquid chromatography |

* Both refer to the same method

# Chapter 1 Introduction

## 1.1 Metabolic Profiling

Metabolic profiling (MP), also referred to as Metabonomics, is defined as "the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification"[1]. MP aims, by employing analytical chemical techniques, towards covering, ideally, the full 'spectrum' of metabolites (metabolome) in a living system. Metabolites "serve as direct signatures of biochemical activity"[2] and this activity provides current and historical information of the biological processes taking place in a cell or organism. Other –omic approaches are interrogating the parts of genetic information flow which can be subjected to further processes and modifications, such as genes, transcripts and proteins (genomics, transcriptomics and proteomics). However, the metabolome represents the real end-points of (patho-)biological reactions (Figure 1. 1) and can report on exogenous factors introduced by diet or gut microbiota in addition to the mammalian metabolism[3]. For these reasons metabolites can correlate closely to phenotype and deliver disease related information.

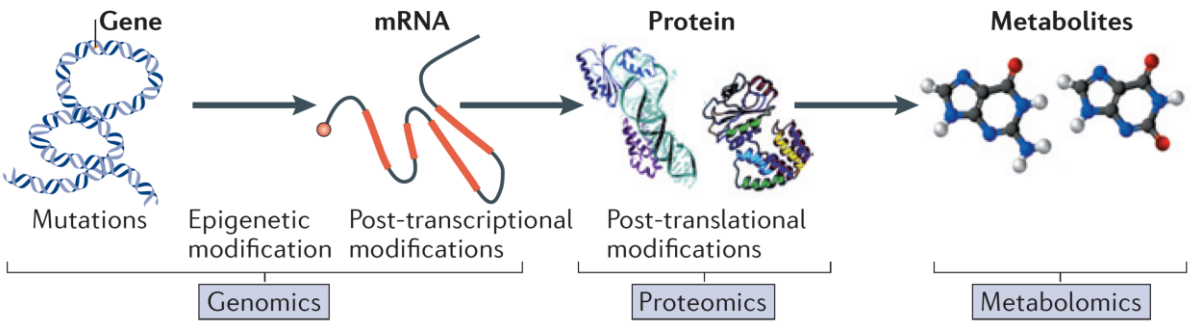


*Figure 1. 1: The central biological dogma indicating metabolites as the final products of genetic information flow. Metabolites excel compared to the information provided by proteins as being closer to the phenotype, since they are not subjected to further modifications. They practically represent an updated snapshot of the current status of a biological system. Adapted from Patti el al[2].*

MP can define this "multiparametric (metabolic) response" using multivariate data analysis (MVDA) methods. However, for this, metabolite measurements should be able to provide at least relative quantification. Therefore, techniques used should be responsive to differential metabolite concentrations, and have a good dynamic range in order to capture, as far as it can be possible, the dynamic range of living systems. MP, with the help of technological advances, has evolved into a powerful tool for discovering novel biomarkers for disease diagnosis[4] and prognosis[5]. It can also be

implemented for pharmaceutical target discovery, and elucidation of disease pathways[6, 7], paving the way for more individualised diagnostics and treatment schedules[5, 8].

Metabolome coverage has become extremely broad, due to advances in the analytical platforms employed e.g. the introduction of cryoprobes in NMR, high resolution mass spectrometers providing superior sensitivity, and Ultra Performance Liquid Chromatography (UPLC)[9]. Further, recent statistical developments, allowing data integration of different analytical matrices, have provided new insights in gene[10] and enzyme function[11, 12] studies and could potentially lead to cost reduction in the drug development pipeline [8], for example, by improving mechanistic understanding of drug toxicity[5, 8].

It has been shown that metabolic profiling of tissue, and/or of the biofluids that interact closely with the tissue of interest, delivers clinically and biologically significant results[13]. Consequently, metabolic profiling techniques are now being applied to homogenised or even intact tissues, cells, as well as biofluids[14-17]. Screening of biological samples, cells or tissues can nowadays be performed in order to obtain biomarker/metabolic pathway information relating to disease. Various spectroscopic techniques are employed[8, 18, 19], such as Nuclear Magnetic Resonance (NMR) spectroscopy[14, 15], Mass Spectrometry coupled to Liquid Chromatography (LC-MS)[16, 20], Gas Chromatography (GC)-MS[21], and less frequently Capillary Electrophoresis[22]. It is of supreme importance, when applying MP, to employ more than a single method or platform in order to maximise the range of the metabolites detected. Findings can then be mapped to reference pathways in online available databases[23, 24] and literature. In cases where traditional biochemistry cannot deliver, data processing methods such as correlation networks can be employed to aid biochemical interpretation of results. This is generally the issue when it comes to metabolite classes such as lipids[25].

When approaches implementing homogenised tissue are employed, the significant parameter of spatial distribution of the biomolecules through the tissue is disregarded. Therefore, new screening techniques for untargeted determination of biomolecules, in a manner that incorporates the spatial distribution, have been considered[26, 27]. Applications of imaging techniques that can deliver chemical profiles related to organ or tissue topography have recently been demonstrated for MP[28-32] (Figure 1. 2). Mass Spectrometry Imaging (MSI) techniques appear to hold what is needed for metabolite *in situ* screening. Magnetic resonance imaging (MRI), able to analyse intact tissue

samples along with providing spatial information of the collected spectra, is the alternative in NMR spectroscopy.

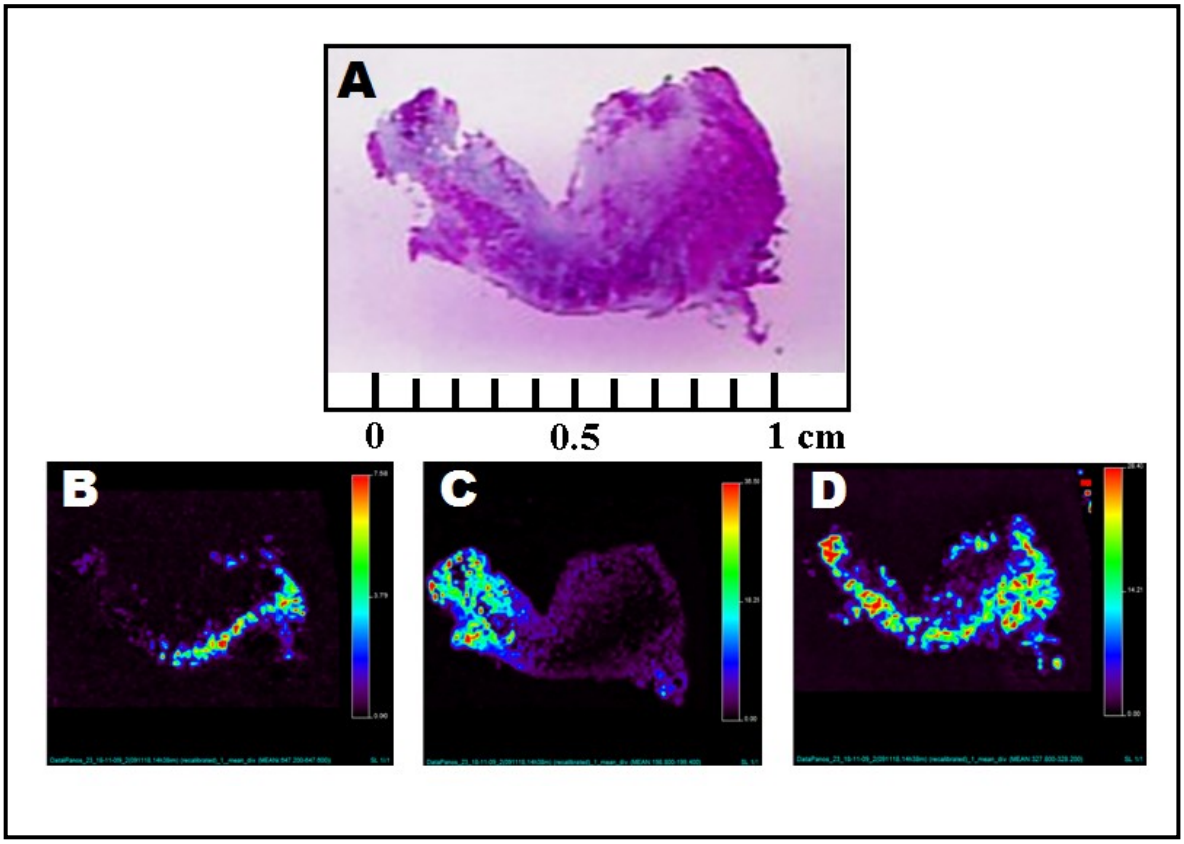


*Figure 1. 2: A calcified human valve section after A) H&E staining and B), C) and D) after representation of three selected ions acquired after MALDI-MSI of the tissue section.*

Nonetheless, MP is now well-recognised[6, 8, 10, 33], and in combination with other –omic technologies[2, 19, 34], usually transcriptomics, or metagenomics[34], can provide good coverage of metabolic and signalling pathways. It should be generally followed by appropriate validation, according to the function of the biomarker discovered. Biomarkers for diagnosis are generally validated to an independent cohort of samples, while novel pathways are validated by bottom-up approaches[35, 36] (Note: A bottom-up approach is defined as the process when prior knowledge of a biological system exists, and the analyst explores the behaviour of this known network and tests hypotheses[37]. Bottom-up approach is also used to describe cell-based approaches, as opposed to top-down approaches used to describe whole organism, multicellular system applications[33]. Top-down can also be used to describe the process where no prior knowledge of a system is used, and by the use of statistics and bioinformatics, modelling of a biological system is obtained). Still, in every case the biological basis of molecule serving as biomarkers, should be determined.

### *1.1.1 Lipid Profiling*

Lipid profiling approaches specifically target the lipidome, which can be considered a sub-category of the metabolome. The lipid moieties comprising the lipidome are calculated to several hundreds of thousands[38-40]. Enzymes known for their participation in lipid synthesis do not reflect the "enormous diversity of the lipidome"[39, 41]. Factors affecting lipidome variation could be environmental, such as the diet, and symbiotic microbiota[39], which can obviously be to an extend independent to the genetic information of an individual. Nonetheless, the lipidome appears to play a very important role in biological systems and disease[42, 43]. There are currently two widespread approaches for perform lipid profiling: LC-MS, and direct MS infusion (shotgun lipidomics)[44]. However, the disadvantage of obtaining lipids as biomarkers lays to the difficulty of comprehensively place them in biological pathways. In order to assist mapping of the detected lipid moieties, correlation analysis can be used[25], as well as combining analyses from polar metabolites and other levels of integration, such as the proteome and/or genome. At the same time this "disadvantage" of the lipidome depicts the importance for elucidating its biological potential in elucidating human disease.

## 1.2 Cardiovascular Disease

The term cardiovascular disease (CVD) refers to all diseases affecting the heart or blood vessels (arteries, veins, or capillaries). CVD is the leading cause of death in the UK and the USA according to the World Health Organisation (WHO) report [45]. The two most frequent causes of death of CVD are mainly related to atherosclerosis and hypertension. The theme of the current thesis is to elucidate the metabolic perturbations causing atherosclerosis-related diseases; such as plaque formation, rupture and arterial ectopic calcification.

### 1.2.1 Atherosclerosis

Atherosclerosis comprises of the formation of a lipid-rich lesion in the inner wall of the arteries. These lesions are generally referred to as (atherosclerotic) plaques. Firstly, this can cause limited flow to important organs such as the heart (angina or infarction), brain (transient ischemic attack, stroke, or microvascular disease) and kidneys (renal failure). In some cases the plaque can rupture and develop a blood clot on its surface. This clot can dislodge itself from the plaque and travel in the circulatory system blocking the blood supply to the brain or heart (thromboembolism), causing health and life-threatening events (e.g. stroke and heart attack). The important fact about atherosclerosis is that despite modern achievements, biologically as well as clinically, it still remains the leading cause of mortality and morbidity, in the western world[46].

#### 1.2.1.1 Low and High- Density Lipoproteins (LDL and HDL)

LDL and HDL are the lipoproteins responsible for transferring lipophilic molecules throughout the body, using the bloodstream. Most abundant lipophilic molecules in lipoproteins are cholesterol, cholesterol esters, triglycerides and phospholipids. Due to their lipophilic nature, these molecules cannot be dissolved in the blood. Therefore, proteins with appropriate structure (apolipoproteins) function as carriers of these molecules, by incorporating them in their core. LDL is considered to be the “bad” lipoprotein for human health. LDL provides the cells with cholesterol by attaching to appropriate LDL-receptors. On the other hand, HDL, the smallest and densest of the lipoprotein particles, is considered the “good” lipoprotein. For reasons that will be explained in the following paragraph, these two lipoproteins are routinely assessed as risk factors for atherosclerosis. As it would be expected by their pseudonyms, high circulating LDL and low HDL are considered to be highly correlated to the risk of atherosclerosis.

#### *1.2.1.2 The process of atherosclerosis*

The whole process of plaque formation begins with the retention of LDL in the endothelial cells of the arterial wall[47] and it occurs prior to monocyte recruitment in the area of plaque formation[48,49]. The trapped LDL soon starts to become oxidised. At this point, inflammatory[48] and adhesion molecules such as chemokines[50] are produced by the endothelial cells and stimulate the recruitment of leukocytes, primarily monocyte, into the endothelium. After recruitment to the lesion, monocytes are converted into macrophages. There macrophages will engulf the LDL particles forming foam cells (named after the foamy appearance of their cytoplasm) and initiating the atheroma.

Macrophages further contribute to LDL oxidation. The protein part of LDL is also oxidatively modified, an event that diminishes the affinity to the LDL-receptors. This eventually leads to bypassing of the feedback loop, causing the cell to create more LDL-receptors in order to capture and retain more LDL in the cell. This can cause cholesterol overload in the endothelium and macrophages, known to be a characteristic of plaques. Other evidence also suggests that oxidised-LDL can be recognised with higher affinity by appropriate acetyl-LDL receptors from the macrophages, leading them to form more foam cell populations[47, 51].

The HDL lipoprotein can help reduce the concentration of cholesterol by a mechanism called "reverse cholesterol transport"[46]. With this mechanism, excess cholesterol is transferred to the liver where it will be secreted in the gastrointestinal track. HDL may also mitigate plaque formation due to anti-oxidative properties of two HDL associated enzymes[47]. The whole procedure described herein is schematically represented in Figure 1. 3.

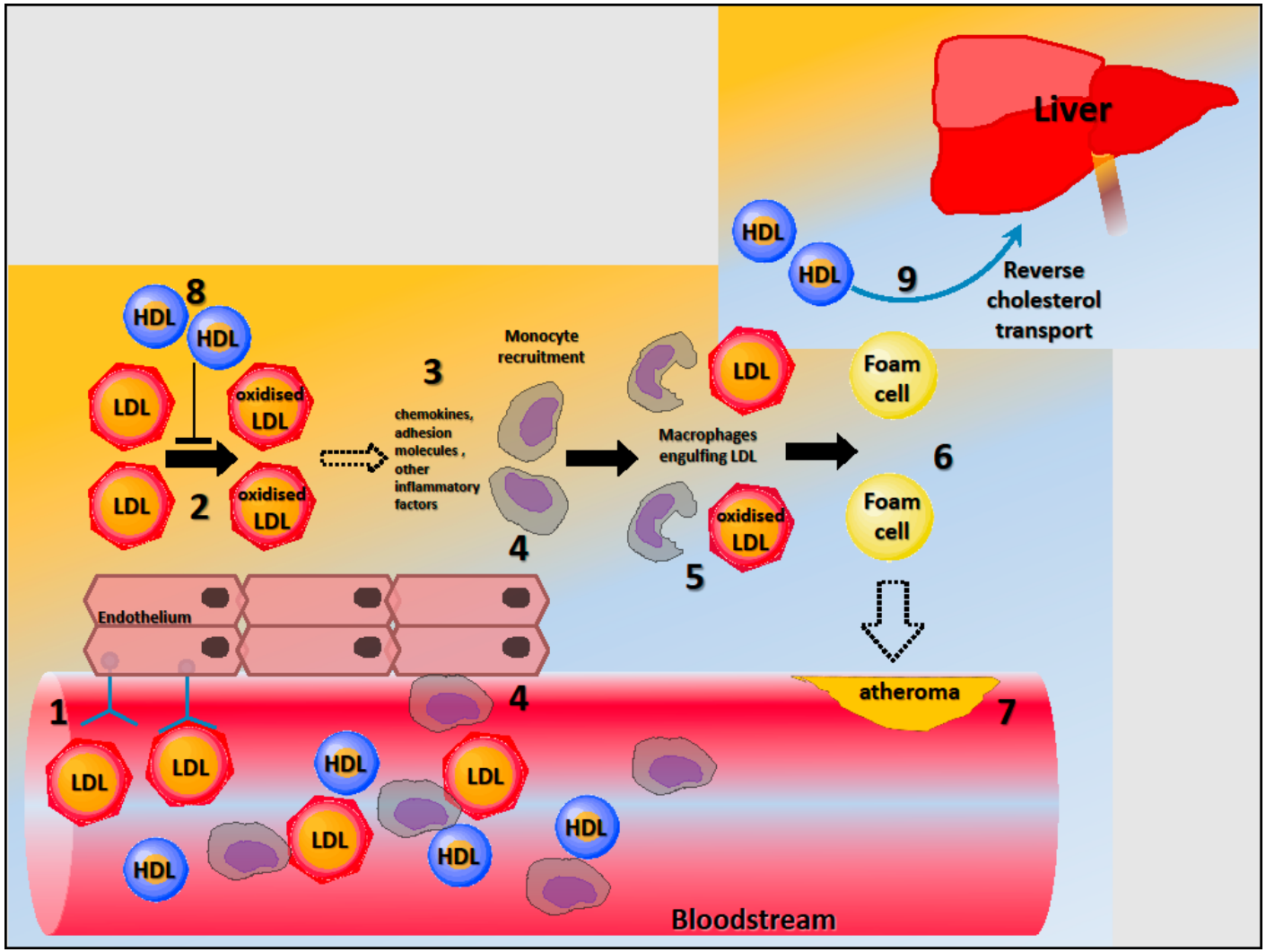


*Figure 1. 3: A schematic representation of atherogenesis. 1) LDL-receptors are produced by endothelial cells and capture LDL, 2) LDL is oxidised, 3) LDL and ox-LDL induce production of chemokines, adhesion and inflammatory molecules, 4) leukocytes and primarily monocytes are recruited, 5) monocytes are transformed to macrophages which engulf LDL and ox-LDL becoming 6) foam cells and forming the 7) atheroma (plaque). 8) HDL has anti-oxidative abilities and can reduce the oxidation of LDL. 9) HDL has the ability of incorporating plaque cholesterol and transports it to the liver.*

Atherosclerotic lesions presenting evidence of lipid accumulation are preceded by a stage known as intimal thickening. During this stage, production of proteoglycans is observed[52]. Intimal thickening is generally found at points of bifurcation, providing evidence indicating that it is a consequence of hemodynamic stress. Points of bifurcation are known areas of application of low wall shear stress[53]. Shear stress is also known to play a role in monocyte adhesion[53]. However, atherosclerotic lesions can also be observed in areas without bifurcations or areas exposed to low hemodynamic stress[48].

Inflammatory mediators in the presence of loss of smooth muscle cells support can lead to weakening of the fibrous cap of the atheroma[54]. This can lead to the rupture of the atheroma and initiation of the coagulation cascade with platelet adhesion. Eventually a thrombus is formed which (Figure 1. 4) can migrate and cause thromboembolism to parts of the subsequent arterial tree. According to the location of the plaque, this could cause a stroke (usually carotid plaque), or heart attack (coronary plaque). Thrombi can also occur from an external erosion of the atheroma[54] (Figure 1. 4).

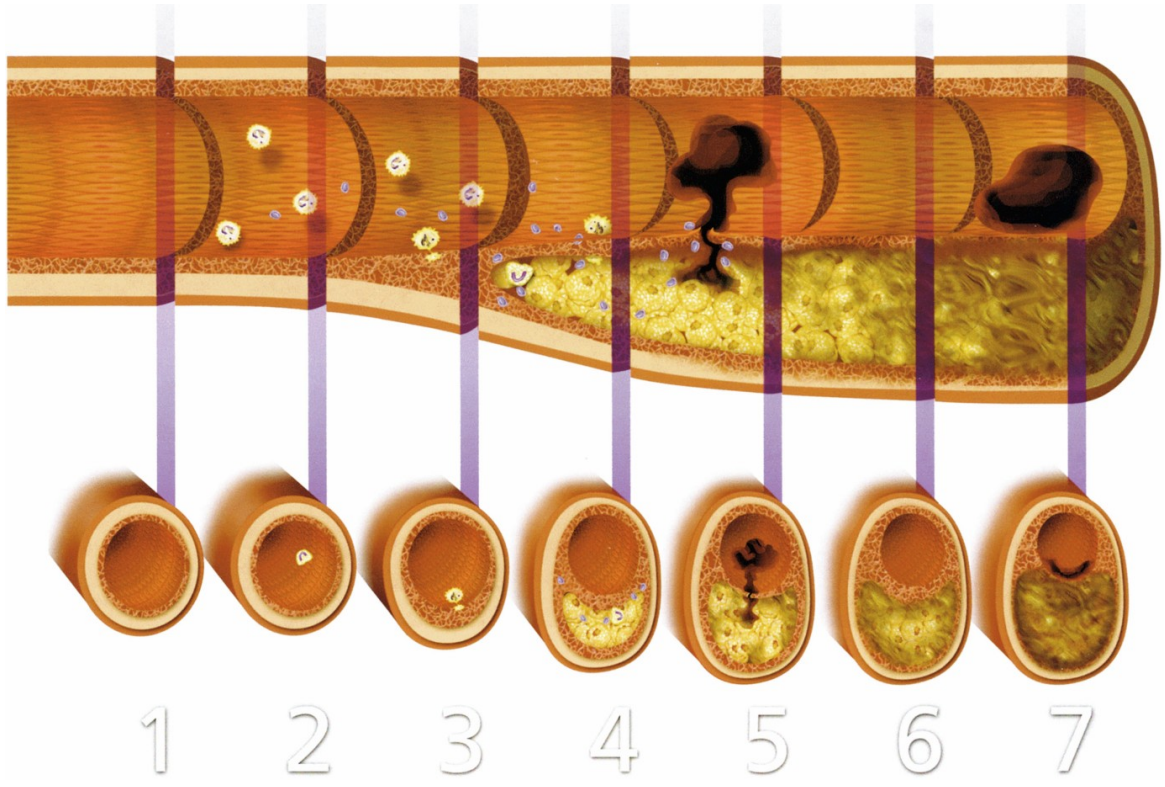


*Figure 1. 4: The basic stages of plaque formation, from normal artery (1), to plaque formation (4) and rupture (5). Adapted from Libby et al [54].*

The basic stages of plaque formation could be synopsised to:

1. Normal artery[54].
2. Lesion initiation[54]: First evidence of tissue injury and intimal thickening.
3. Fibro-fatty stage[54]: Monocytes conversion into macrophages which engulf modified-LDL becoming foam cells. Intima layer becomes thicker.
4. Progression to the atheroma stage[54].
5. Rupture of fibrous cap[54].
6. Thrombus resorption[54]. Evolution towards advance fibrous and calcified plaque.
7. Thrombus from superficial erosion[54].

### *1.2.2 Cardiovascular Calcification*

Cardiovascular calcification is the process where vessels become ossified. It is believed that the calcification process is an active biological process rather than passive deposition with similarities to bone formation[55]. It does not occur in normal vessels. On the contrary, there are evidence that calcification occurs after vessel wall injury[48, 56]. Calcification is generally considered a progression from atherosclerosis and if present, it will classify a lesion as "advanced"[48]. However, some evidence of independent mechanisms also exist[57].

Calcification is an advanced degenerative process. It is generally present, and increases in magnitude with age. However, it should not be considered an age-related degenerative process, as it can occur at any age and can progress with time[58]. Figure 1. 5 demonstrates an X-ray of the post-mortem examination of a young adult, where multicentric and advance calcification is obvious in the coronary artery[48].

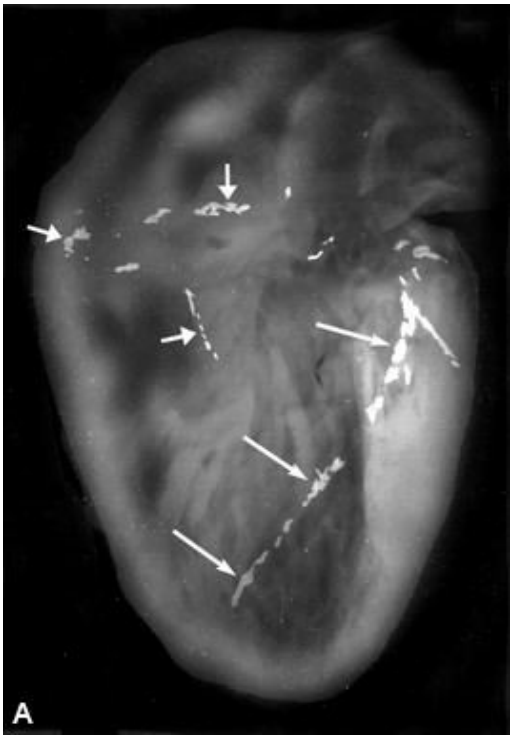


*Figure 1. 5: A post-mortem X-ray film of the heart of a young adult with extensive ectopic calcification on coronary arteries. Adapted from Frink et al [48].*

Within Chapter 5, results from experiments on vascular but also valvural calcification will be presented and discussed. The elucidation of the pathophysiology of vascular calcification is important due to contradictory reports of prognostic significance of calcification in cardiovascular events[57, 59, 60]. On the other hand, aortic valve calcification (Figure 1. 6) is the second most common indication for cardiac surgery[61], and the leading cause for valve replacement in the United States[62]. Calcific aortic valve disease is also associated with atherosclerosis[62] and hemodynamic stress[63]. It is also important to state that there is evidence of the reversibility of the disease, e.g. in malignancies[56] and after statin treatment[64], making the quest of an appropriate pharmaceutical target more essential.

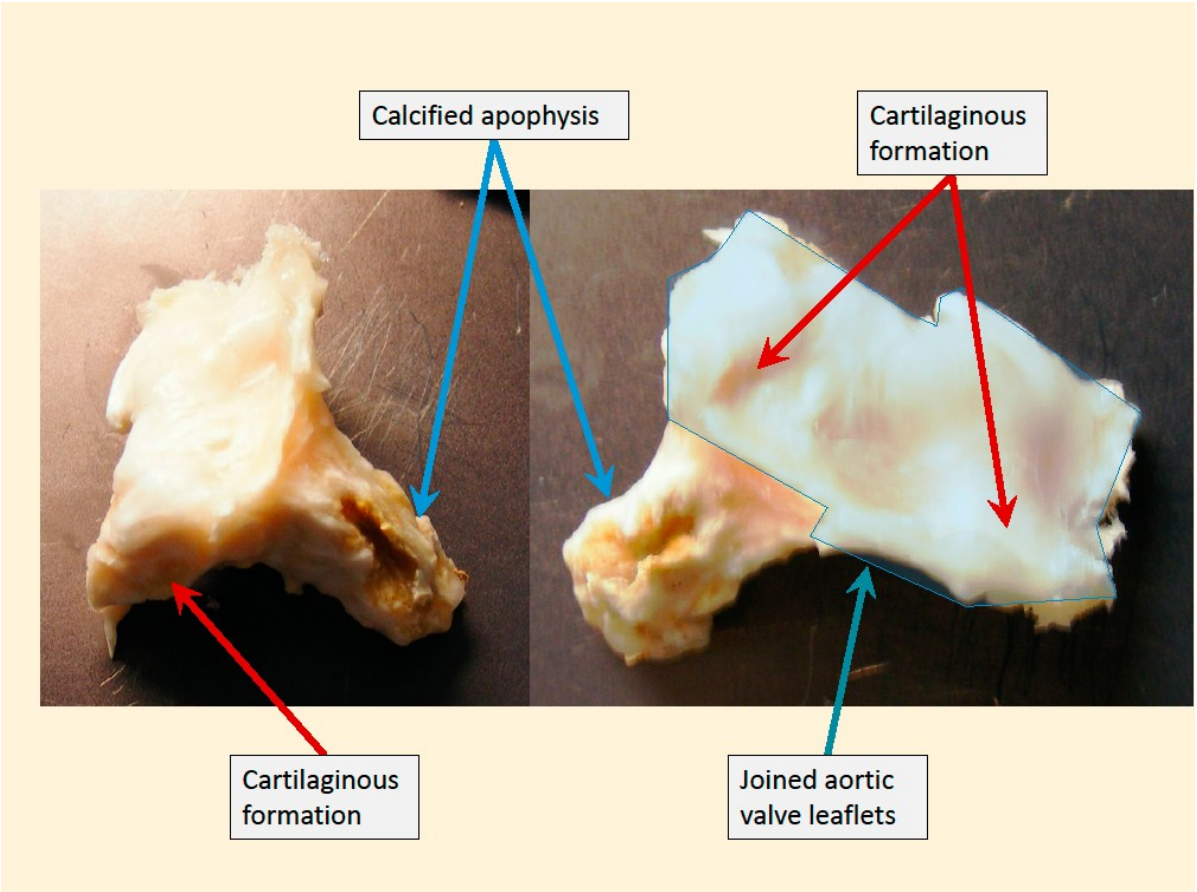


*Figure 1. 6: An image of two conjugated aortic heart valves. Valves show evidence of fibrous and calcium formation.*

## 1.3 Metabolic profiling studies on cardiovascular disease

In this section, a review of the current literature relating to metabolic profiling applications on cardiovascular disease is performed. The PubMed search engine (http://www.ncbi.nlm.nih.gov/pubmed) was used for identifying relevant papers. The search quote used was: (metabolomics OR metabonomics OR "metabolic profiling") AND (atherosclerosis OR "cardiovascular disease"). Searching with this quote retrieved 141 results. However, only original research articles are reviewed and discussed. Review, editorials, commentaries or perspective articles were excluded.

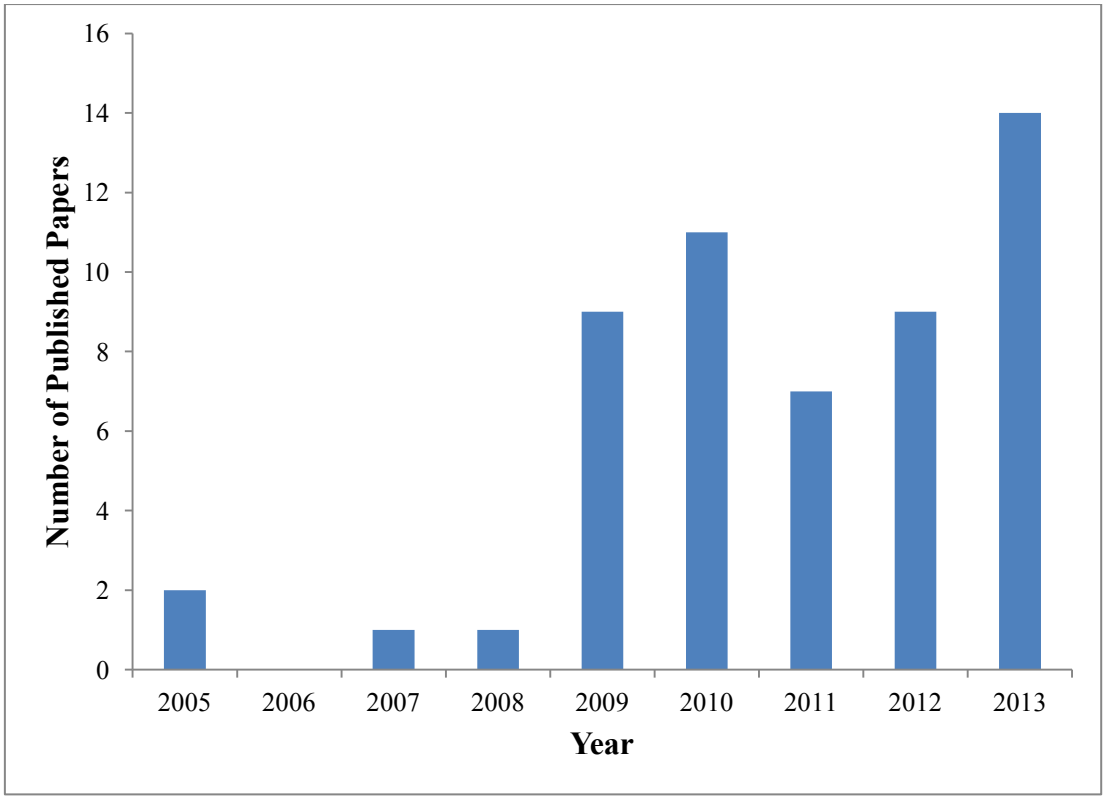


*Figure 1. 7: Number of published research articles relevant to metabolic profiling applications on cardiovascular disease, per calendar year.*

Metabolic profiling studies have been applied for cardiovascular disease and atherosclerosis-related research since 2005. Although until 2009 not many original research articles were published, there were a large number of reviews and perspective articles. This demonstrated very well the unmet need of the scientific community for new technologies able to provide novel candidate biomarkers, as well as aid the effort for elucidation of the dysregulated biological pathways.

Application of metabolic profiling was initiated on pilot/feasibility studies[65], and using recognised animal models[66-72]. Intervention studies conducted using metabolic profiling tools and relating to atherosclerosis, are also present in literature[69, 73-77]. Inevitably, - and due to the widespread interest for cardiovascular disease and importance for acquisition of relevant risk-factors – multi-centre, large population studies were conducted as soon as the technology could sustain the analysis and processing of the required sample numbers. The INTERMAP (INTERnational collaborative study of MAcronutrients, micronutrients and blood Pressure)[78], ARIC (Atherosclerosis RIsk in Communities)[79, 80] and LIPGENE[81] studies are some of the noticeable examples in literature.

Lastly, only five articles in published literature apply metabolic profiling using tissue as the biological matrix[65, 70, 82-84]. Moreover, only one is utilising the technology on human tissue[84] rather than tissue from animal models. This depicts the need for further applications that can provide

important biological information that can only be obtained on tissue level rather than biofluids. Furthermore, the need for developing appropriate methodologies, in order to cover for the need of tissue applications, is evident.

Overall, the great need for studies applying metabolic profiling on cardiovascular disease is obvious. These studies should work on elaborating on a global, systemic level, maximising the information obtained. Finally, more studies should be applied on tissue samples, a matrix that can provide invaluable information about the pathophysiology of the disease and dysregulation of biological pathways.

# Chapter 2 Analytical Strategies

## 2.1 Introduction

As discussed in the preceding chapter, metabolic profiling uses a pallet of techniques and methods in order to maximise metabolome coverage. The most frequently used techniques are NMR spectroscopy and LC-MS. In both platforms several methodologies are used to cover a different range of compounds. For NMR it could be a range of sequences, such as CPMG and diffusion, and resonance of different nucleus such as proton, $^{31}P$ and $^{13}C$. On the other hand for LC-MS a range in chromatographic columns such as reversed-phase and HILIC columns to cover for the different lipophilicity of metabolites, or even different sources for the ionisation of the compounds prior to detection, such electrospray ionisation (ESI) or atmospheric pressure chemical ionisation (APCI). In the following sections the two platforms, NMR and UPLC-MS, are discussed since they are the two techniques applied for obtaining results for studies relevant to this thesis. Statistical tools applied to deal with these holistic analytical approaches are also discussed.

## 2.2 Ultra Performance Liquid Chromatography coupled to Mass Spectrometry (UPLC-MS)

To address the complexity of the metabolome, platforms are needed that can provide enhanced resolution in order to detect and (semi-)quantify the wide range of compounds in a biological sample (such as urine, blood, tissues). In such cases coupled methodologies can provide the wide range of coverage required by holistic analytical approaches. This section describes the platform utilising the advantages of liquid chromatography coupled to mass spectrometry, which is becoming more ubiquitous as an analytical tool in metabolic profiling studies.

Ultra Performance Liquid Chromatography – Mass Spectrometry (UPLC-MS) describes the coupling of these two very powerful and widely used techniques: liquid chromatography and mass spectrometry[85, 86] (Figure 2. 1). Using liquid chromatography compounds are subjected to separation to be further ionised on elution and detected by the mass spectrometer. On one hand, liquid chromatography has become an invaluable separation tool in analysis, with numerous applications in bio-, as well as food[87, 88] and drug analysis[89-92]. On the other hand, mass spectrometers, as compared to other detectors coupled with liquid chromatography, offer the additional specificity that liquid chromatographic separation need, especially when it comes to

complex matrices. Currently, mass spectrometry constitutes the heavyweight tool in the 'arsenal' of untargeted metabolic profiling technologies, offering additional sensitivity in a wide range of analytes.

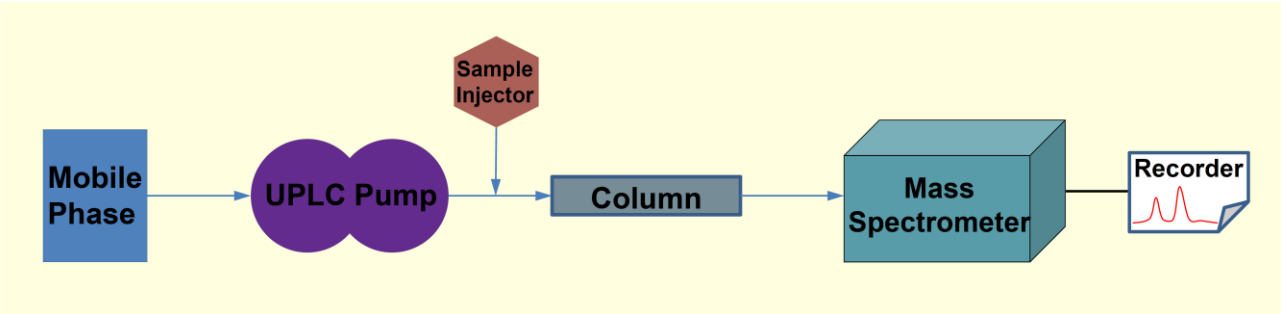


*Figure 2. 1: A diagram of an LC-MS format. The heart of the LC is the pump. The pump is supplying the mobile phase to the system at a specific flow. The mobile phase will pass through the column and reach the detector, an MS in this case, which is followed by a recorder. Between the pump and the column, there is the sample injector which is where the sample is loaded and injected into the flow to be separated.*

## 2.2.1 Ultra Performance Liquid Chromatography (UPLC)

### 2.2.1.1 Liquid Chromatography

Liquid chromatography is based on the chemical interactions of the analyte to: 1) the solid material that a chromatographic column is packed with, called the *stationary phase,* where the analyte is essentially partitioned, and 2) the *mobile phase*, which is constituted by the solvents that flow through the column. A correctly set up chromatographic method should firstly allow for the analyte to be retained in the column due to the interactions and physicochemical forces to the stationary phase, that are greater than to the mobile phase. This will allow for 'non-specific' molecules and salts to flow through the column. By adjustments to the composition of the mobile phase, the physicochemical forces of the stationary phase to the analyte will be surpassed by the ones between the analyte and mobile phase. Therefore the analyte will elute from the column. When the composition of the mobile phase is adjusted through the analysis then this is called a *gradient elution*. However, there is ability for the system, after proper adjustments in solvent composition, to retain the analyte and elute without changing the composition of the mobile phase through the run. This is called *isocratic elution*. There are both advantages and disadvantages in both types of elution. The gradient elution is generally used when the analytes in a sample have a wide range of physicochemical properties. This is the case in metabolic profiling, when untargeted methodologies are used. The immediate issue from this is the long re-equilibration of the system to the initial conditions, i.e. initial mobile phase compositions.

The efficiency of the chromatographic separation of a method/technique is measured by the resolution (R) of the system. This would be the measurement of separation of two peaks, proportional to their difference in retention time, and inversely proportional to the average width of the two peaks (Equation 2. 1)[93].Figure 2. 2, and Equation 2. 1 describe the concept of chromatographic resolution. Another measure of chromatographic efficiency is the number of chromatographic plates which is inversely proportional to peak width.

$$\boldsymbol{R} = \frac{t_2 - t_1}{{}^{1}/_{2}\,(t_{w1} + t_{w2})}$$

*Equation 2. 1: The equation describing the calculation of chromatographic resolution: $t_1$ and $t_2$ represent the time of each peak at the apex, $t_{w1}$ and $t_{w2}$ represent the width of each peak.*

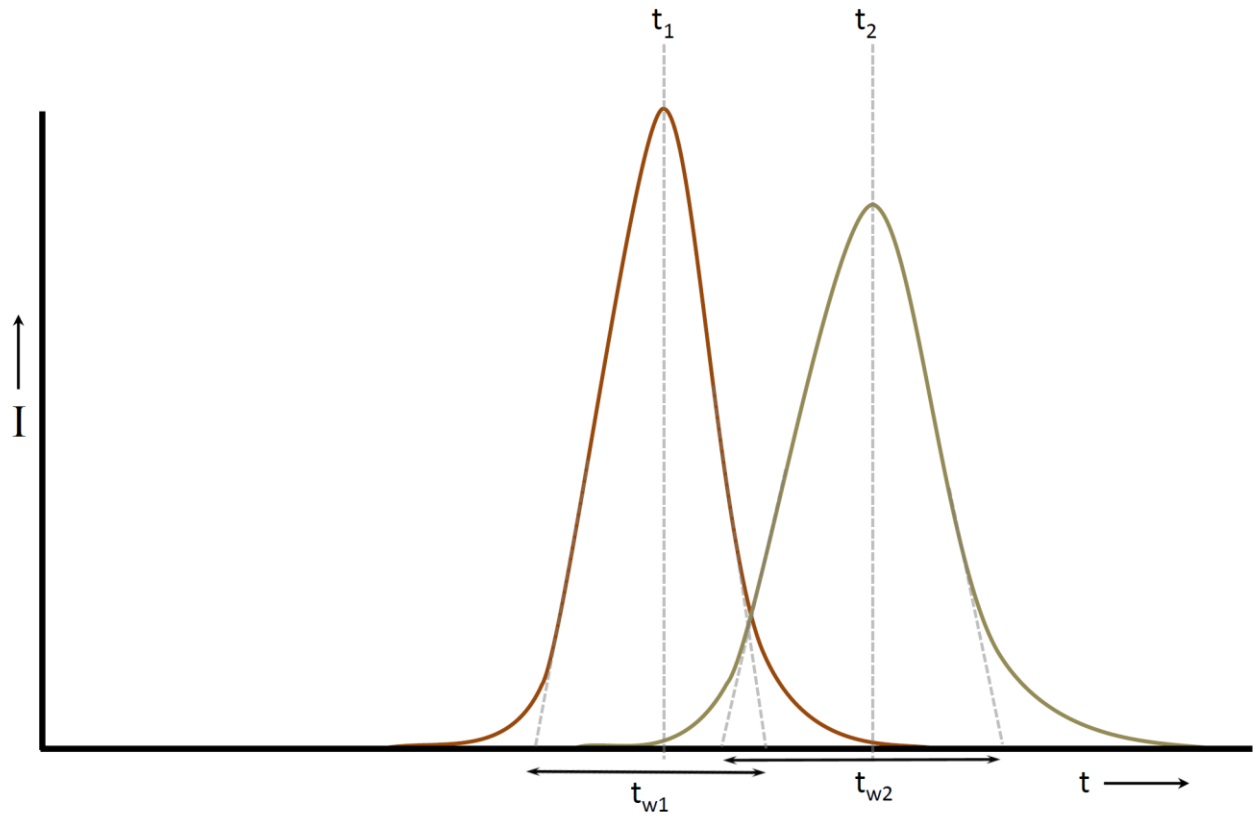


*Figure 2. 2: A schematic representation of two chromatographic peaks, along with their characteristics important for calculation chromatographic resolution.*

The invention and coining of the term chromatography, in attributed to Mikhail Tsvet (1903). Since then, multiple contributions to chromatography are reported and recognised. However, the

introduction of the first commercial HPLC (High Performance Liquid Chromatography) is credited to Waters in 1969, with the ALC100 HPLC. The HPLC came with advance resolution capabilities. This was due to the ability to support a stationary phase consisting of particles typically in a diameter of 5μm or larger[93]. This increased the surface of interaction, and provided tighter packing of the column. However, tighter packing came with increased backpressure. Thus, HPLC columns came with the advanced ability to tolerate such pressures, and at the same time providing constant and reproducible flowrate.

The heart of the liquid chromatography instrumentation is the pumps. They are constructed with proviso to resist corrosion from salts and strong organic solvents which usually comprise the mobile phase. They are also designed to provide a pulse-free output. Reciprocating pumps are the ones generally used. According to the pump format single-, dual-, or triple- head pumps maybe used to reduce flow pulsation. Also pulse damping is sometimes used to compensate for the small fluctuations in flow and pressure[93].

#### *2.2.1.2 Ultra Performance Liquid Chromatography*

HPLC was the separation technique of choice for over 30 years. Ten years ago the new concept of Ultra Performance Liquid Chromatography, UPLC, was introduced[94, 95]. It was introduced by Water Corporation and now the copyrights for the name UPLC belong to this company. Other companies use different names to describe UPLC, such as RRLC, for rapid resolution liquid chromatography (Agilent). Therefore, the scientific community is currently using the term ultra high performance liquid chromatography (UHPLC). For the purposes of this thesis the term UPLC is used.

The advancement that made UPLC a higher efficiency chromatographic technique was the reduction of particle size of the stationary phase and tighter packing. The theoretical background is that particles of less than 2.5μm in diameter can increase theoretical chromatographic plates and keep this high efficiency even at high flowrates. UPLC is using sub-2μm particles in diameter, generally 1.7 - 1.8μm.

However, with smaller particle size and tighter packing, along came the increase in backpressure. This was another challenge that was resolved by the UPLC systems. A UPLC system and columns, can typically tolerate up to 15,000psi of system pressure. Additional provisions are in place to avoid any cause of high pressure. Such protective module is the pulse-free sample injection.

In general, the UPLC can provide improved separation compared to its predecessor, HPLC. Additionally, the minimal dispersion of analytes provides higher sensitivity. Lastly, as discussed,

the analyst has the ability to go for faster analyses without significantly compromising chromatographic resolution. These are the advantages that make UPLC the separation platform of choice in today's bio- (or other) analysis applications.

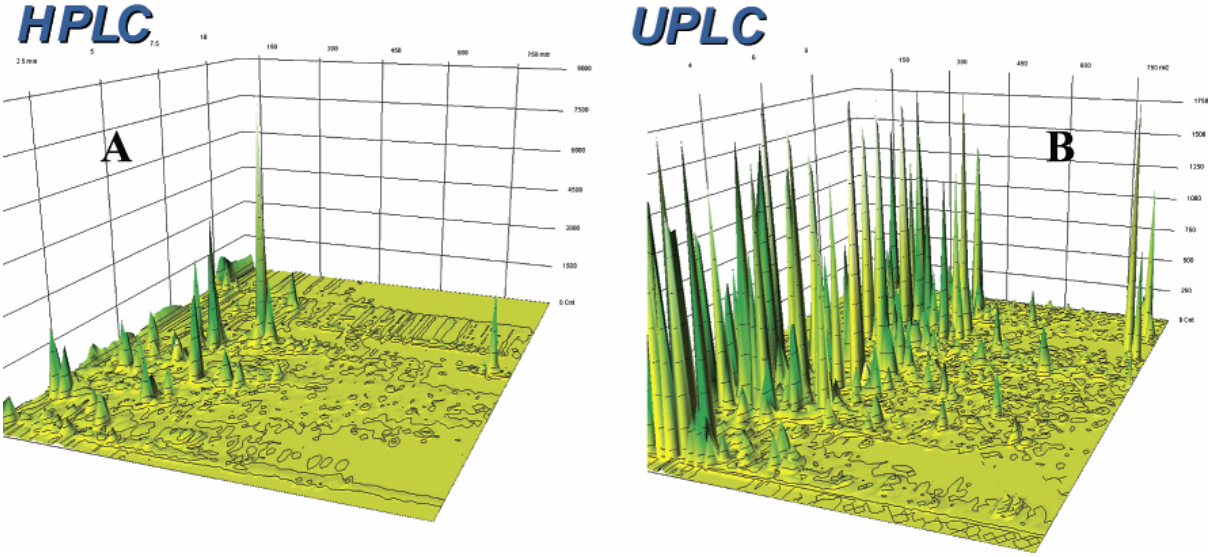


*Figure 2. 3. Three-Dimensional Maps of (A) HPLC-MS and (B) UPLC-MS, showing retention time, m/z and intensity. The superiority of UPLC, especially in terms of sensitivity is obvious. The figure is adapted from Wilson ID et al[96] and is obtained from the analysis of mouse urine.[96]*

### *2.2.1.3 UPLC Column chemistries*

In order to optimise the analyte-stationary phase chemical interactions, different column chemistries are used. This means that the material of the stationary phase will vary according to the physicochemical properties of the analyte, such as their structural groups and lipophilicity. An optimal interaction would mean that the analyte would elute without major interferences or background, in a rather pure form, and with a minimal diffusion within the system, translated to a sharp peak when detected.

Two are the major concepts applied in column chemistries today in UPLC-MS untargeted metabolic profiling, reversed-phase and HILIC:

#### *2.2.1.3.1 Reversed-phase Chromatography*

Reversed-phase (RP) chromatography got its name due to the inverse principles applied compared to the classical mode that was initially used for chromatographic applications. This initial form was coined normal phase and it used to describe chromatographic systems where the mobile phase is less polar than the stationary phase. Typically for normal phase, the stationary phase is silica.

Consequently, RP describes the exact opposite. The mobile phase is more polar than the stationary phase. This is generally achieved by irreversible binding of aliphatic chains to the silica. The most common chains used are C8 (saturated aliphatic chain of 8 carbons) and C18 (saturated aliphatic chain of 18 carbons).

For the current study two RP columns were used. The first was a $C_{18}$ HSS (High Strength Silica) T3 column (Figure 2. 4A). The high strength silica material is made in order to provide high mechanical stability to the silica. This allows for its application in high pressure methods. In high pressures the silica material can be crushed, thus restricting the flow and altering the profile of the analysis. This kind of stability is needed in metabolic profiling especially due to the long runs with large sample sets, in order to provide the required reproducibility[97].

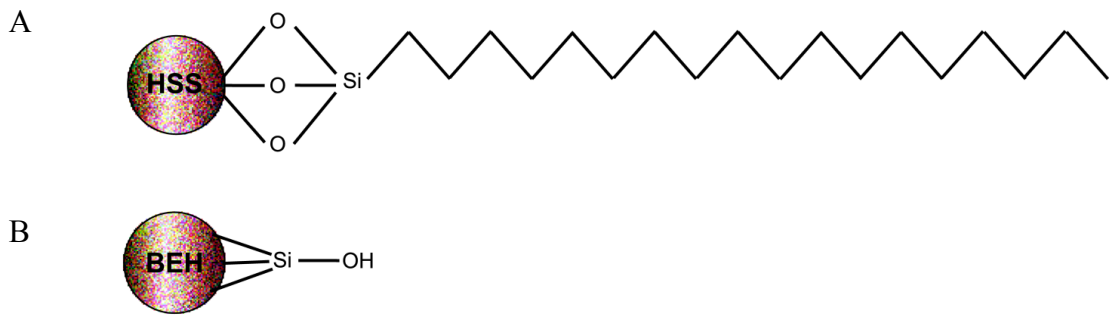


*Figure 2. 4: The chemical structures of an HSS T3 column and a BEH HILIC.*

The second column used is a C18 CSH (Charge Surface Hybrid) column. This column uses chemically stable silica material in an ethylene bridged hybrid (BEH) structure (Figure 2. 5). It provides the high chemical stability that silica was missing and in a wide range of pH, as well as increased mechanical stability. On top of the BEH silica particles a low-level charge is applied on the particle surface. Then the C18 chain is bonded to the silica. Apart from improved chromatographic efficiency, the CSH format provides increased loading capacity, tolerance to high pH and fast equilibration time[98].

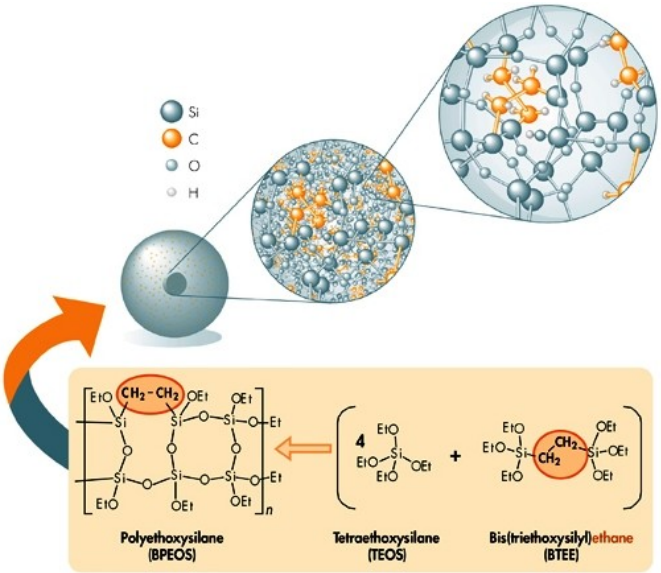


*Figure 2. 5: The chemical structure of BEH column indicating the ethylene bridge in the silica structure for added stability. In the bottom is the reaction to produce this bridged polymer. Adapted from Waters Corp[99].*

### *2.2.1.3.2 Hydrophilic Interaction (Liquid) Chromatography (HILIC)*

Hydrophilic Interaction (Liquid) Chromatography (HILIC) can be considered a variant of normal phase chromatography. HILIC uses silica as a stationary phase. For our experiments BEH HILIC columns were used. The ethylene bridge links of BEH, as discussed in Section 2.2.1.3.1 , provide extra chemical stability to the stationary phase. As expected, HILIC can manage very well with polar compounds (a property lacking in RP), and provide the retention needed. However, for the format of this approach, in contrast to normal phase, some water is required in the mobile phase. HILIC functions by "partitioning of the analyte between a water-enriched layer of stagnant eluent on a hydrophilic stationary phase and a relatively hydrophobic bulk eluent"[100]. This is demonstrated in Figure 2. 6. On top of the advantage of superior retention and separation of polar compounds[101, 102], HILIC also provides better ionisation when coupled to MS, than normal phase. This is due to the eluents used, which are not totally organic and non-polar[100]. Additionally, preparation on mobile phase is less laborious as water does not need to be totally omitted[100]. Lastly, the solvents used, water and acetonitrile, are more compatible to methods already applied (e.g. RP methods) which minimises the need of extensive priming of the UPLC system.

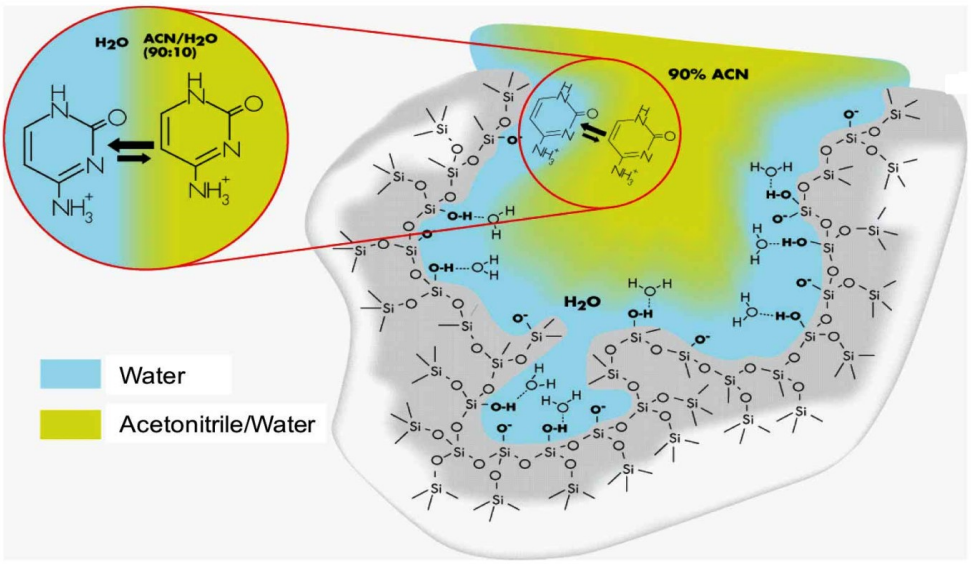


*Figure 2. 6: The partitioning of an analyte (cytosine) between the water-enriched layer, formed by the interaction to silica, and the mobile phase. Adapted from Neue et al [103].*

### 2.2.2 Mass Spectrometry

UPLC provides the improved separation needed for bioanalysis applications and especially for holistic analytical methods applied in metabolic profiling[96]. However, the short width of the UPLC peak profile (usually in the range of 3-10 s) demands detectors with high-speed scanning abilities, while at the same time, holistic, untargeted approaches, demand a detector that can provide specificity, sensitivity in a wide range of molecules and adequate structural information on the separated compounds.

Mass Spectrometry (MS) is a physical analytical technique. It can function as a detector for UPLC, as it can accurately measure analyte mass, and with multiple scans per second. MS can provide identification of compounds, according to their mass-to-charge ratio (m/z), and fragmentation pattern (MS/MS). This makes MS an attractive tool for applications in the field of metabolic profiling, and has actually become the method of choice compared to other platforms[104], especially when coupled to UPLC. Nowadays, MS instruments come with enhanced sensitivity and specificity. Their universal format is summarised in Figure 2. 7, and it includes the source, analyser, detector and recorder. The detectors are usually using multichannel plates (MCP), which use the principle of secondary emission, to amplify the signal from each ion colliding on their surface.

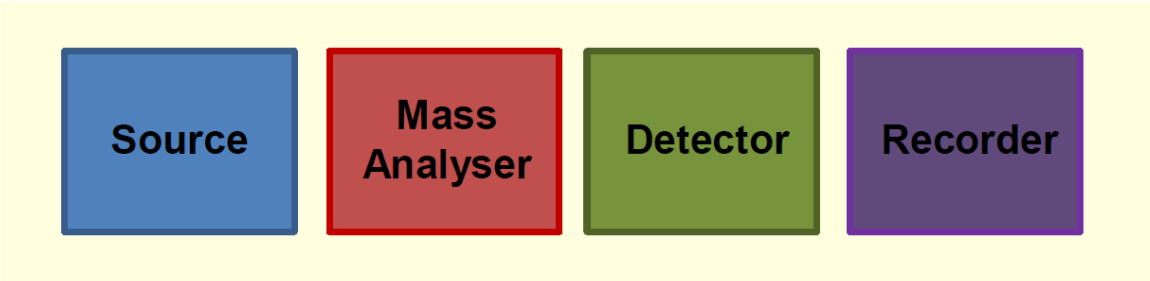


*Figure 2. 7: A diagram indicating the basic parts constituting an MS instrument.*

The high resolution abilities of modern MS instruments are a major contributor to the increased specificity. MS resolution (*R*) is a measure of separation between two m/z peaks. In most MS instruments this is considered a separation at 50% of two m/z peaks. It is therefore calculated as the m/z of the peak (*m*) divided by the difference between two resolved peaks (*Δm*) (Equation 2. 2). Resolution can be calculated for an isolated m/z peak by using as Δm the full width at half maximum (FWHM) of the peak.

$$R = \frac{m}{\Delta m}$$

*Equation 2. 2: The equation for calculation of MS resolution: m is the m/z of the ion used for resolution calculation and Δm is the m/z difference between two separated ions.*

Mass accuracy is the measure of accuracy of the detected m/z of a specific ion. It is describing the error compared to the theoretical m/z, and it is important to be as low as possible for untargeted studies as it can make the structural assignment easier and the methodology more specific. It is expressed as parts per million (ppm) and calculated as the difference of the experimental m/z from the theoretical, divided by the theoretical and multiplied by $10^6$ (Equation 2. 3). The mass accuracy of an instrument is closely related to the resolution. Only high resolution instruments can provide a high mass accuracy[105] such as Fourier transform ion cyclotron resonance (FTICR) mass spectrometers.

$$Mass\ Accuracy = \frac{m/z_{theor} - m/z_{exper}}{m/z_{theor}} \times 10^6$$

*Equation 2. 3: Calculation of mass accuracy of a detected ion.*

#### *2.2.2.1 Types of mass spectrometers for determination of m/z*

Several types of mass spectrometers exist, each relying on different principles and instrumental setup, in order to determine the m/z of an analyte. Some of the most widely used mass spectrometers for targeted and untargeted metabolic profiling (utilized for applications for this thesis) are described in the following paragraphs.

##### *2.2.2.1.1 Time of flight (TOF) mass spectrometers*

The most widely used in metabolic profiling are the time-of-flight (TOF) mass spectrometers. For the TOF-MS instrument to calculate the m/z of a charged molecule, an electric pulse (voltage) is applied to the molecule while guided in a field-free region. All molecules have the same initial energy according to the voltage applied ($E_v$=zV). The molecules travel in the TOF tube while under a vacuum, but have different velocities due to their differences in mass. This is due to the kinetic energy which is proportional to mass and velocity ($E_k$=½$mv^2$). Provided the initial energy $E_v$ is equal to the kinetic $E_k$, the time for the molecule to flight in the tube is analogous to the root of m/z.

Thus, smaller ions will reach the detector faster than larger ions when introduced in the TOF, provided that they have the same charge. The advantage of the TOF-MS is that charged molecules reach the detector within nanoseconds, and therefore a full, wide-range, m/z scan can be obtained in high-speed. Typically, 5 scans are required per second when MS is coupled with UPLC, and TOF analysers can deliver this scan frequency, and in theory even as fast as 30 scans per second.

Modern TOFs can also provide high resolution, especially since the introduction of reflectrons and orthogonal TOFs, and improved mass accuracy, enhanced by the use of lock-mass correction (Figure 2. 9). Reflectrons, also referred to as ion mirrors, are a set of parallel opposite charged metal rods, where ions can be reflected. This way the TOF separation is increased by elongating the ion path without physically making the TOF tube longer. This means that less space is needed for the analyser but also less power to pump down. Orthogonal TOFs (oTOF) were introduced to solve the issue with TOF-separation affecting ions prior to entering the TOF tube. This separation can cause widening of the ion peaks and reduce resolution. The oTOFs can increase resolution by orthogonally introducing ions into the TOF tube since they apply a voltage perpendicular to their current velocity. These advantages, and the relative low cost of instrumentation and maintenance, are what make TOFs preferred for untargeted metabolic profiling.

*2.2.2.1.2 Quadrupole mass spectrometers*

On the other hand, quadrupoles use a different principle for calculating m/z. Quadrupole analysers use oscillating electric fields to isolate charged molecules according to m/z. This is done by usually two pairs of parallel cylindrical rods with opposite charge (Figure 2. 8). Trajectories followed by the charged molecules will differ according to their m/z and oscillation frequencies. Ions with m/z not focused to pass through the quadrupole, will eventually collide with the charged rods and neutralise. This apart from the obvious specificity, it can provide higher sensitivity when targeted approaches are applied. Additionally, the cost of instrument can be low. However, when a wide m/z range is required, a full scan can take up to a second[106]. Quadrupoles can also suffer in mass accuracy. These disadvantages restrain the application of quadrupoles for untargeted approaches, though they are preferred for targeted.

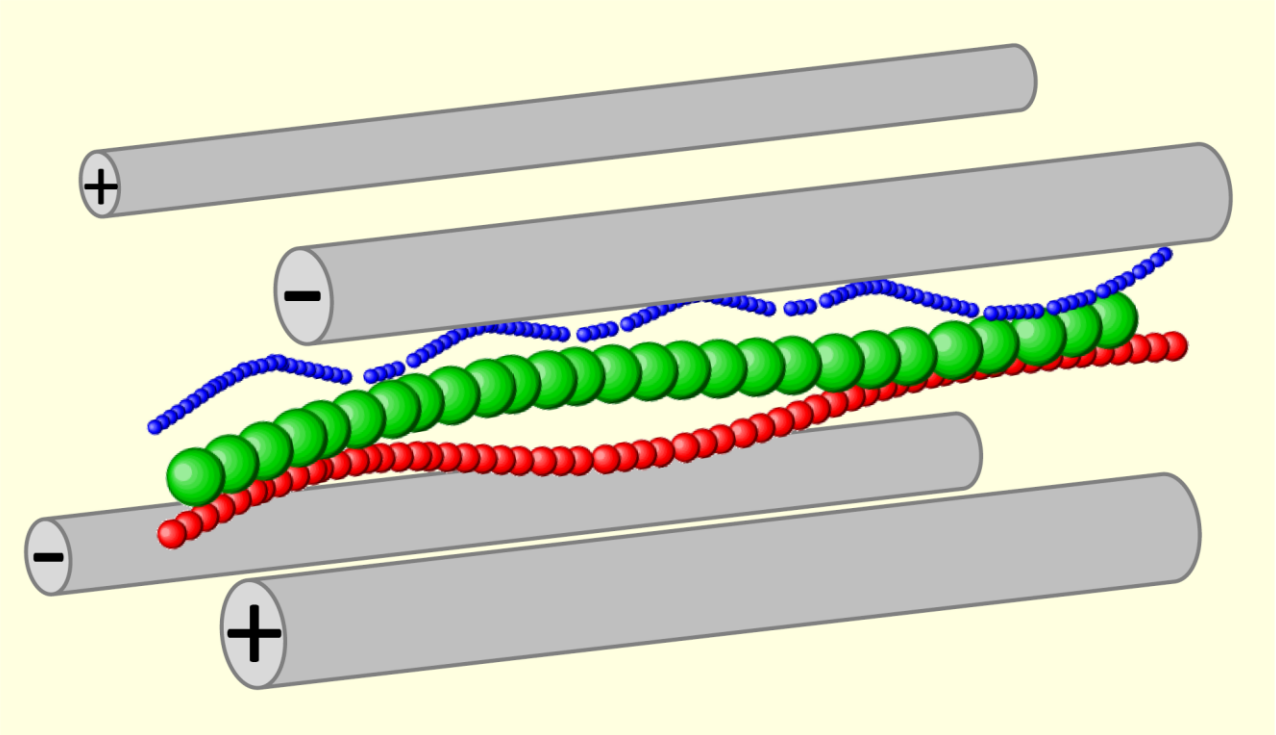


*Figure 2. 8: A schematic representation of a quadrupole. The two pairs of parallel cylindrical rods bear opposite charge and are under constant electric field oscillation, according to the desired ion. According to the oscillation, ions with different m/z will differ in their trajectories. Non-specific ions will eventually collide to the charged rods and neutralise.*

#### *2.2.2.1.3 Quadrupole-TOF*

A hybrid of the two MS instrumentations that combines the best features from both is the instrument of choice for several laboratories working on metabolic profiling. This instrument is the quadrupole-TOF (qTOF) and has the quadrupole followed by a collision cell (where fragmentation of molecules occurs), and ending with a TOF for m/z determination and recording (Figure 2. 9). This provides the flexibility for TOF analyzer to be used with untargeted applications. The quadrupole can be used in order to select a specific m/z to be subjected to fragmentation, and the full m/z spectrum can be recorded again after TOF separation.

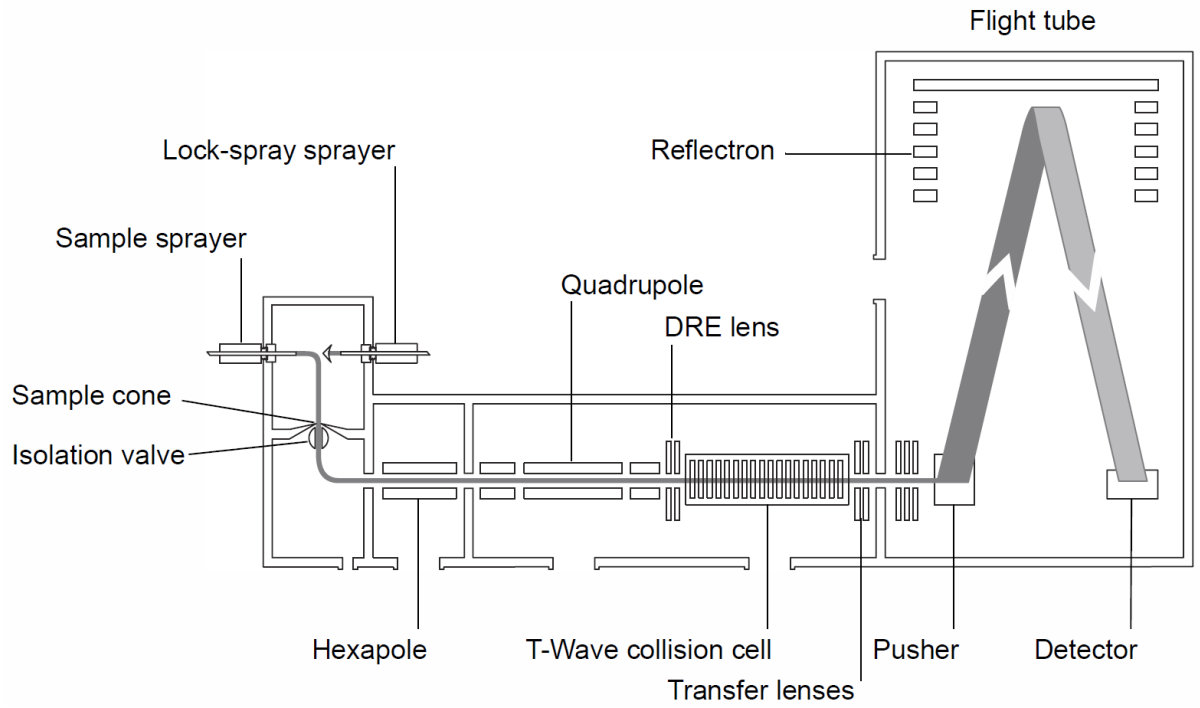


*Figure 2. 9: A diagram of a Xevo G2. The sample is sprayed into the source by the sample sprayer. The ions are the guided through a z-shape course and by passing through ion optics, a quadrupole and a collision cell are entering the TOF tube to be subjected to separation and detection according to their m/z.* [107]

### *2.2.2.2 Ionisation Sources – Electrospray ionisation (ESI)*

The ionisation source is the part of an MS instrument that converts molecules to a charged state. It also transfers charged molecules into the gas-phase, appropriate for detection by MS analysers. The mostly used ionisation technique coupled to UPLC is ESI. ESI, introduced by *Fenn et al*[108], is the preferred ionisation technique for UPLC-MS metabolic profiling studies, due to the 'soft' ionisation it can provide[109]. It can interface very well UPLC and MS, as it can cope with transferring separated molecules, coming from the UPLC, into the gas-phase.

ESI occurs in essentially ambient atmospheric pressure, and after application of strong electric field to a liquid. The liquid gets to the source through a capillary. The high voltage is applied between the capillary and the counter-electrode, in close proximity. Generally voltages range between 1-3kV and vary in polarity according to the charge of the molecules detected. The droplets created at the end of the capillary are subjected to high charge accumulation due to the electric field. A gas flow usually nitrogen is applied coaxially to the sample flow, with the intension of restraining the droplets axially and assisting to the production of fine aerosols. On top of that, the desolvation gas is applied ($N_2$), usually at high temperature and flow, which causes solvents to evaporate (Figure 2. 10). The droplet will experience greater opposite forces on the surface as the solvent evaporates and it gets smaller. This will cause a further explosion of the droplets into smaller ones, and will keep

occurring until no solvent is left and the analyte is in the gas-phase (Figure 2. 11). This way the accumulated charge on the surface of the droplet is passed to the molecules. The exact mechanism of transfer of the charge from the droplet to the molecules is not clear, but its principle is that the charge of the droplet remains unchanged as solvent evaporates[110]. The ions enter the MS analyser guided by the forces from the application of opposite charge voltage. This comes on top of the suction forces applied as the analyser constantly functions under a vacuum. Usually, the flow of the source and the entrance of the analyser (sample cone) are orthogonally located (Figure 2. 9 and Figure 2. 10). This avoids the direct infusion of neutral molecules into the analyser. In Figure 2. 9 it can be seen that the ions are guided through another orthogonal route. This works as additional improvement in ion-charge specificity, and eventually increases the sensitivity of the instrument by reducing the non-specific ions reaching the detector. At the same time it protects the instrument from unnecessary accumulation of debris. Perhaps the greater disadvantage of ESI and MS in general, is that molecules are competing for the same amount of charge. This causes a phenomenon called ion suppression and is the major restriction of sensitivity in MS technologies.

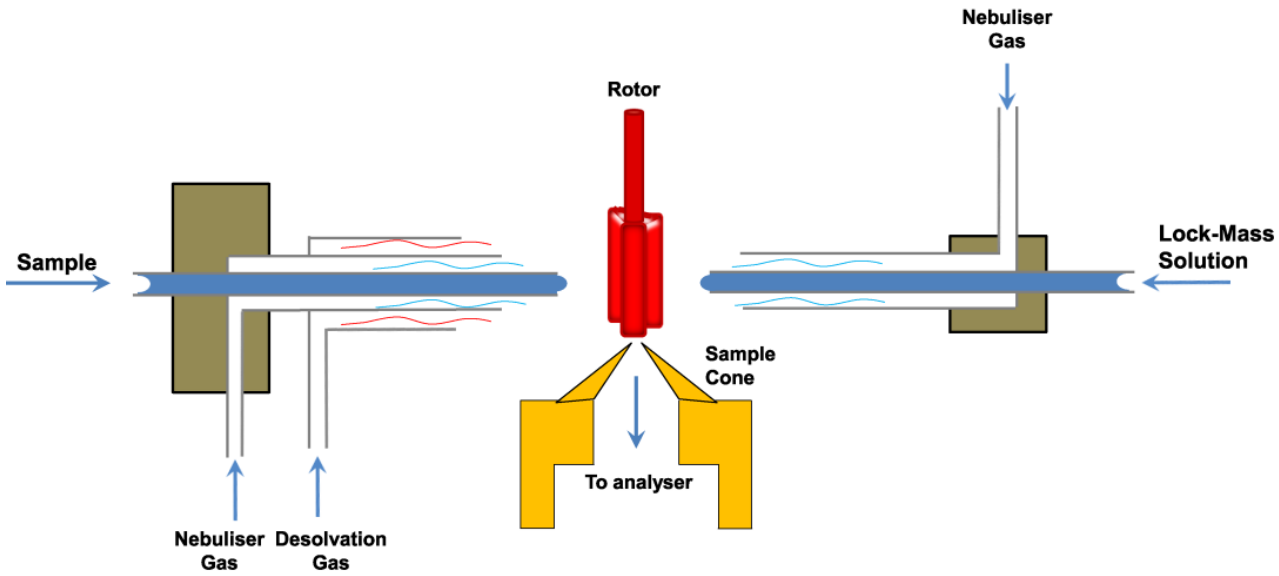


*Figure 2. 10: The design of an ESI source. On the left is the probe where the sample is injected. On the right is the lock-spray, where the lock-mass is injected for constant adjustments to the recorded mass. The rotor is constantly switching between blocking either the sample entering the analyser or the lock-mass.*

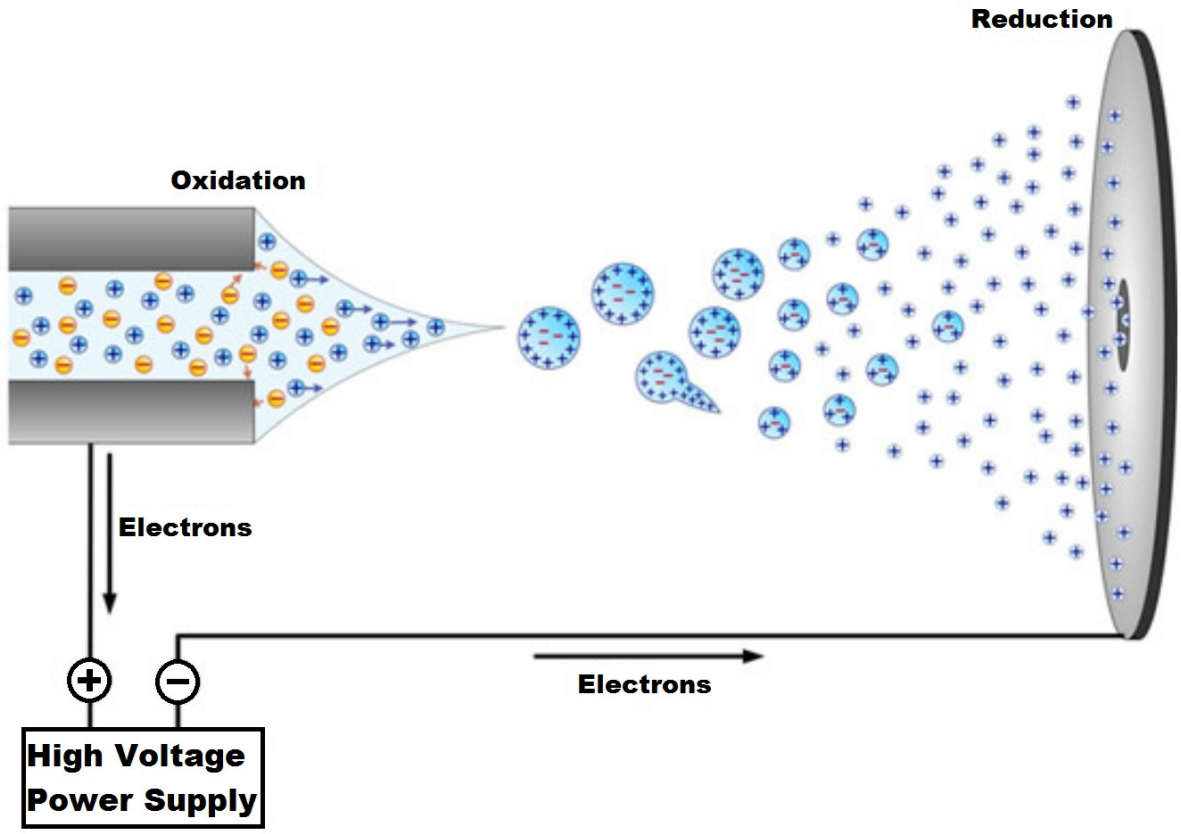


*Figure 2. 11: A representation of the mechanism of electrospray ionisation in the source of an MS analyser. Adapted from Waters Corp[110].*

### *2.2.2.3 Tandem MS*

Tandem MS or MS/MS can provide structural information of a molecule according to the fragments produced, or by matching of the fragmentation pattern to a standard compound. As previously mentioned one of the most common instruments used for MS/MS acquisition is the QTOF, and it is the instrument used in this study as well. In order to obtain an MS/MS spectrum of an ion using a QTOF, this ion, called the parent ion, needs to be isolated first. This is done by a quadrupole, but in this case the analyser is not followed by a detector. Instead, the ion will pass through a collision cell where it will be fragmented by a neutral gas (such as nitrogen or helium) and application of voltage. The fragments will then pass into the TOF analyser and detected, providing only ions originated only from the precursor ion. An example of how MS/MS can provide structural information and be matched to a compound can be found in Figure 2. 12.

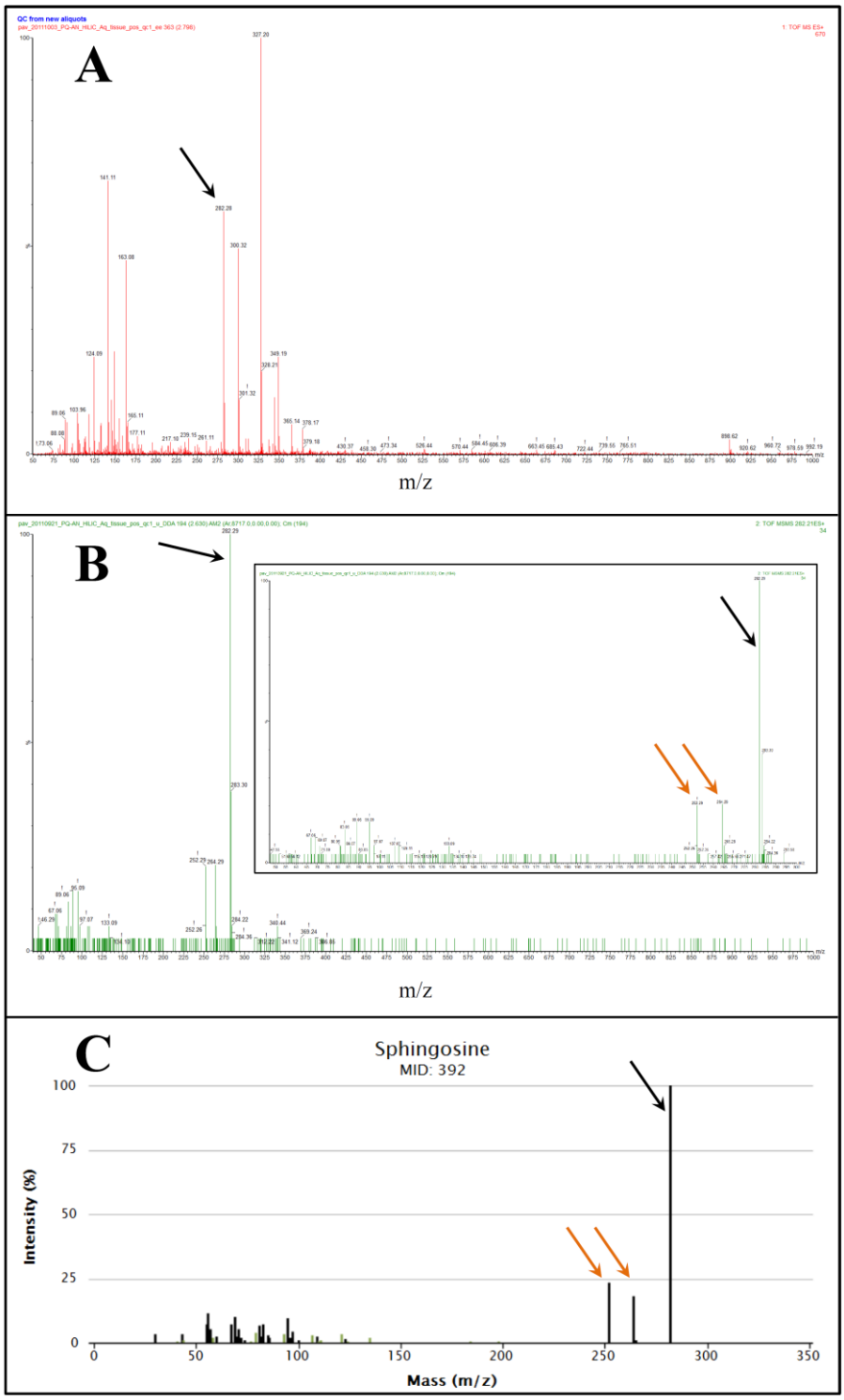


*Figure 2. 12: Example of how tandem mass spectrometry (MS/MS) by using a QTOF mass analyser, can assist in structural assignment of a compound. Data from studies described in Chapter 4 are used, illustrating the identification of sphingosine from ms/ms data. (A) The MS function of a fraction eluted from a column at a given retention time shows multiple ions. (B) A specific m/z (282; black arrow) is chosen for MS/MS, and ion-specific fragments (orange arrows) after collision-induced dissociation are detected (Insert: Expansion of 50-300m/z). (C) Ion-specific fragments can be matched to online MS databases (in this case Metlin[111]).*

The introduction of tandem MS acquisition modes such as data depended acquisition (DDA) has assisted to a great extent the assignment of statistically significant molecules in metabolic profiling. DDA can be set to automatically acquire MS/MS spectra from ions, usually according to an intensity threshold.

Another mode which can be characterised as a variation of tandem MS is the $MS^E$. It is different to $MS^n$ as there is no initial parent ion isolation. All molecules will be fragmented as they pass from the collision cell, usually after ramping of the collision energy. It is generally used after UPLC separation and data are generally collected along with an $MS^1$ acquisition scan[85]. The $MS^E$ function is specific to Waters instruments.

These approaches allow the scientist to collect, during the initial untargeted run, enough fragmentation information to assign a great number of molecules prior to knowing the outcome of the analysis. This saves time and the frustration of matching different retention times to the same analyte, as minor shifting may occur. Such fluctuations in retention time, observed between runs, are usually small, and are mostly due to small differences in the preparation of mobile phase or column degradation. Sometimes they can be quite large differences due to variations in tubing length.

### *2.2.3 UPLC-MS data processing in metabolic profiling studies*

In untargeted UPLC-MS metabolic profiling methods, a vast amount of information is collected. In the case of UPLC-MS this information comes in the form of detected chromatographic peaks, attributed to a specific m/z. In order to handle this wealth of information, appropriate automation is required to transform the data into a concise and representative format, enabling the application of subsequent statistical analysis tools. Generally, the scientist requires a matrix consisting of all the samples as columns, and as rows all the detected peaks. The peaks are represented as features describing the intensity of a specific retention time and m/z. Generally peaks are characterised by the area under the peak, or height, and there are some basic steps to be applied in order to get to that point.

Two are the major algorithms that need to be applied in order to obtain the information needed for statistical analysis. The first is peak-picking, which is comprised of an algorithm that identifies chromatographic peaks, and results in a number of identified peaks, with their own individual characteristics (retention time, m/z, area) per sample. This is followed by grouping, which involves the identification of peaks that describe the same feature. An example of the matrix generated by such algorithms in UPLC-MS can be seen in Figure 2. 13.

| Primary ID | 848.6401_12.9581 | 848.6553_9.3234 | 848.6585_11.057 | 848.69_14.6416 | 848.697_13.7563 | 848.7717_15.4354 |
|---|---|---|---|---|---|---|
| 73 | 10839 | 8564 | 0 | 11889 | 0 | 1.43679e+006 |
| 211 | 11373 | 0 | 0 | 5737 | 0 | 1.07911e+006 |
| 233 | 13314 | 0 | 0 | 10244 | 0 | 970055 |
| 277 | 9539 | 11080 | 0 | 11943 | 0 | 1.04614e+006 |
| 305 | 12190 | 0 | 0 | 11965 | 0 | 1.52072e+006 |
| 308 | 14955 | 7367 | 0 | 9636 | 0 | 529209 |
| 358 | 12597 | 0 | 0 | 19090 | 0 | 1.46445e+006 |
| 373 | 11112 | 0 | 0 | 10043 | 0 | 829686 |
| 395 | 8217 | 10038 | 0 | 6172 | 0 | 690561 |
| 407 | 12039 | 1355 | 0 | 19067 | 0 | 1.44246e+006 |
| 408 | 5483 | 0 | 0 | 9320 | 0 | 1.34823e+006 |
| 414 | 4433 | 0 | 0 | 0 | 0 | 162770 |
| 425 | 17282 | 0 | 0 | 29109 | 0 | 1.84708e+006 |
| 452 | 11305 | 11236 | 6899 | 6545 | 0 | 601862 |
| 481 | 12832 | 0 | 4065 | 0 | 0 | 178634 |
| 494 | 7635 | 0 | 0 | 27811 | 0 | 1.6966e+006 |
| 507 | 10787 | 10105 | 942 | 8029 | 0 | 640901 |
| 556 | 10089 | 10867 | 0 | 7695 | 0 | 381318 |
| 583 | 8008 | 0 | 0 | 11658 | 0 | 1.57512e+006 |
| 592 | 9363 | 11231 | 0 | 7916 | 0 | 658820 |
| 634 | 11489 | 8504 | 10414 | 4166 | 0 | 232035 |
| 725 | 7831 | 3727 | 0 | 9029 | 0 | 642226 |
| 768 | 10406 | 0 | 0 | 13008 | 0 | 1.0675e+006 |
| 794 | 10351 | 12114 | 0 | 1722 | 0 | 176779 |
| 841 | 12878 | 10981 | 0 | 12371 | 4062 | 827401 |

*Figure 2. 13: A matrix (partial) generated from UPLC-MS data after data processing with peak-picking and grouping algorithms. Columns represent features describing m/z and retention time, and rows represent samples.*

There are a number of packages applying algorithms that generate these matrices of processed data. Some of them offer more flexibility and have more abilities than others. Two packages that are used in this study for data processing are the XCMS package[112] and MarkerLynx.

XCMS is an open-source package[112], written in the R programming language. Apart from the peak-picking and grouping, it can provide abilities such as retention time correction and filling of missing peak data. It also provides the flexibility of choosing from different algorithms for different processes. Disadvantages could be considered the low user friendliness, and necessity of transformation from the initial format (related to the software of the instrument company) to a compatible format to the package (e.g. NetCDF)

MarkerLynx is a package from Waters Corporation that is built-in their central software MassLynx, used primarily to interface the instrument and for data visualisation. It makes processing easier as the scientist can move directly towards applying data processing algorithms. The disadvantage is the limited flexibility, and reduced abilities which don't include processes such as retention time correction and missing peak filling.

The last part of data processing is normalisation. Normalisation is a per sample adjustment of features in order to remove systematic variation occurring from biologically irrelevant factors. Such variations may include reduction in instrumental sensitivity through a run, inconsistencies in pipetting, and different sample dilutions (e.g. in urine). This way samples become comparable to each other, and variation is more likely to be biologically related. There are different approaches on how to properly normalise a dataset but which one is more appropriate is debated[113]. For this study, total area normalisation is applied for UPLC-MS studies. Total area normalisation accounts very well for reduction in analyser sensitivity (due to accumulation of debris and other sources of contamination, generally in the instrument source) which is a common issue in long runs[114].

### *2.2.4 UPLC-MS metabolite identification*

Structural assignment of metabolites can be a very tedious task, when untargeted approaches are applied. It is sometimes referred to as "the bottleneck of metabolic profiling". The success of metabolite identification is, to a great extend, laying on the skills of the analyst and requires a wide range on knowledge. For this thesis, descriptions of modules used to identify compounds are described in each chapter. In this section, guidelines for: accomplishing structural assignment and appropriate reporting for the level of assignment are described.

Metabolite assignment is essential no matter what the ultimate target of the analysis is; elucidation of a biological pathway, or diagnostic biomarker discovery. For revealing pathway dysregulations, pathway mapping can aid biological interpretation and placing differential metabolites into context with the studied disease. Altered biological processes contributing towards the pathological phenotype can be identified and targeted for further clarification of their dysfunctionality. On the other hand, although there is an impression that diagnostic biomarkers or profiling may not necessitate relating to pathways of disease, it is actually of supreme importance and is a part of diagnostic biomarker validation.

#### *2.2.4.1 Minimum reporting for metabolite assignments*

Concerns arising from the need to appropriately report evidence for structural assignments of metabolites, led several groups working on metabolic profiling to reach and publish a consensus on minimum reporting. The Metabolomics Standard Initiative has published a paper entitled "Proposed minimum reporting standards for chemical analysis"[115] reporting suggested levels of metabolite identification:

- Level 1: Identified compounds (two independent and orthogonal data relative to an *authentic compound* analyzed under *identical experimental conditions*)

- Level 2: Putatively annotated compounds (e.g. without chemical reference standards, based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries)
- Level 3: Putatively characterized compound classes (e.g. based upon characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class)
- Level 4: Unknown compounds — although unidentified or unclassified these metabolites can still be differentiated and quantified based upon spectral data

There of course intermediate levels that could be achieved or one could even surpass the top levels if it is considered necessary. For example, while two independent and orthogonal measurements, matched to authentic standards, are considered sufficient for level 1 assignment, this would mean that m/z and retention time could be enough to accomplish top level of assignment. However, ms/ms spectra could provide more confidence to the assignment combined with m/z and retention time measurements. In rare cases it might be considered of importance to derivatise the compound and if the expected derivative is produced, provide additional support to the assignment, though this could usually be considered excessive.

#### *2.2.4.2 Workflow for accomplishing structural assignment in untargeted UPLC-MS metabolic profiling experiments*

There are several steps towards accomplishing the structural assignment. These steps are to guide the analyst towards the corresponding structure and at the same time provide support for the assignment:

1. Accurate mass database search
2. MS polarity mode of detection consistent with structure
3. Biological matrix and species
4. Sample preparation
5. Retention time
6. Charge number (z)
7. Mass deficiency
8. Elemental composition
9. MS/MS and $MS^E$
10. Authentic standards – matching to retention time and MS/MS of sample feature
11. Considerations for the possibility of structural isomers

Although matching to accurate mass, retention time, ms/ms, and authentic standards are the most important steps of the workflow, there is still valuable information to be gained from every suggested step.

#### *2.2.4.2.1 Accurate mass database search*

Searching on relevant public databases for a specific accurate mass can provide vital direction towards the appropriate structure of an unassigned feature. Databases used should be relevant to the samples analysed, meaning that the analyst should use databases dedicated for biological matrices. The Human Metabolome Project (HMDB)[116], Metlin[111] and LipidMaps[117] are some of the databases available online, relevant to biological matrices and offering searching based on mass accuracy measurements (Figure 2. 14).

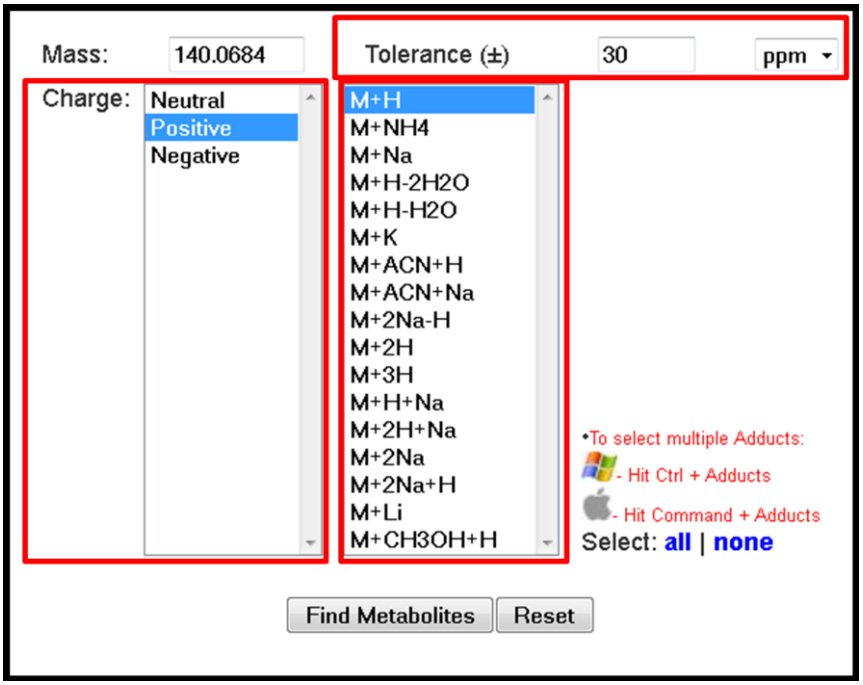


*Figure 2. 14:The options offered by the Metlin database[111], for searching using an accurate mass measurement.*

Online databases would generally offer options for polarity mode, relevant adducts and a mass window in Dalton (Da) and/or ppm (Figure 2. 14). Some of these databases offer calculations for dimer ions, which can also be the case for small molecules. They would not, however, provide matching of any isotopes, and they are focusing mostly on primary and parent ions. All of these manifestations of a molecule in the ionic state (adducts, dimers, isotopes) could make approaching a conclusive result more difficult, relaying on the experience of the analyst to rationally narrow down probable matching metabolites.

In Figure 2. 15, an example of database matching (Metlin database[111]) using accurate mass is illustrated for a feature detected with m/z of 140.0684. When covering a greater range of potential adducts (in this case adducts of ammonium, sodium, water and the protonated ion form) a total of 36 metabolites match the searching criteria. However, when the potential adduct forms are restricted to only sodium, only 17 hits match the searching criteria. The analyst can acquire this information and restrict the search criteria by inspecting the m/z spectrum at the specific retention time. As can be seen in Figure 2. 15C another ion at 118.0852, is matching the possibility of the protonated ion. This can be further confirmed by chromatography (Figure 2. 15D) where it becomes obvious that the two ions at m/z 140.0684 and 118.0852 are coeluting, demonstrating with high confidence that these two ions are manifestations of the same molecule.

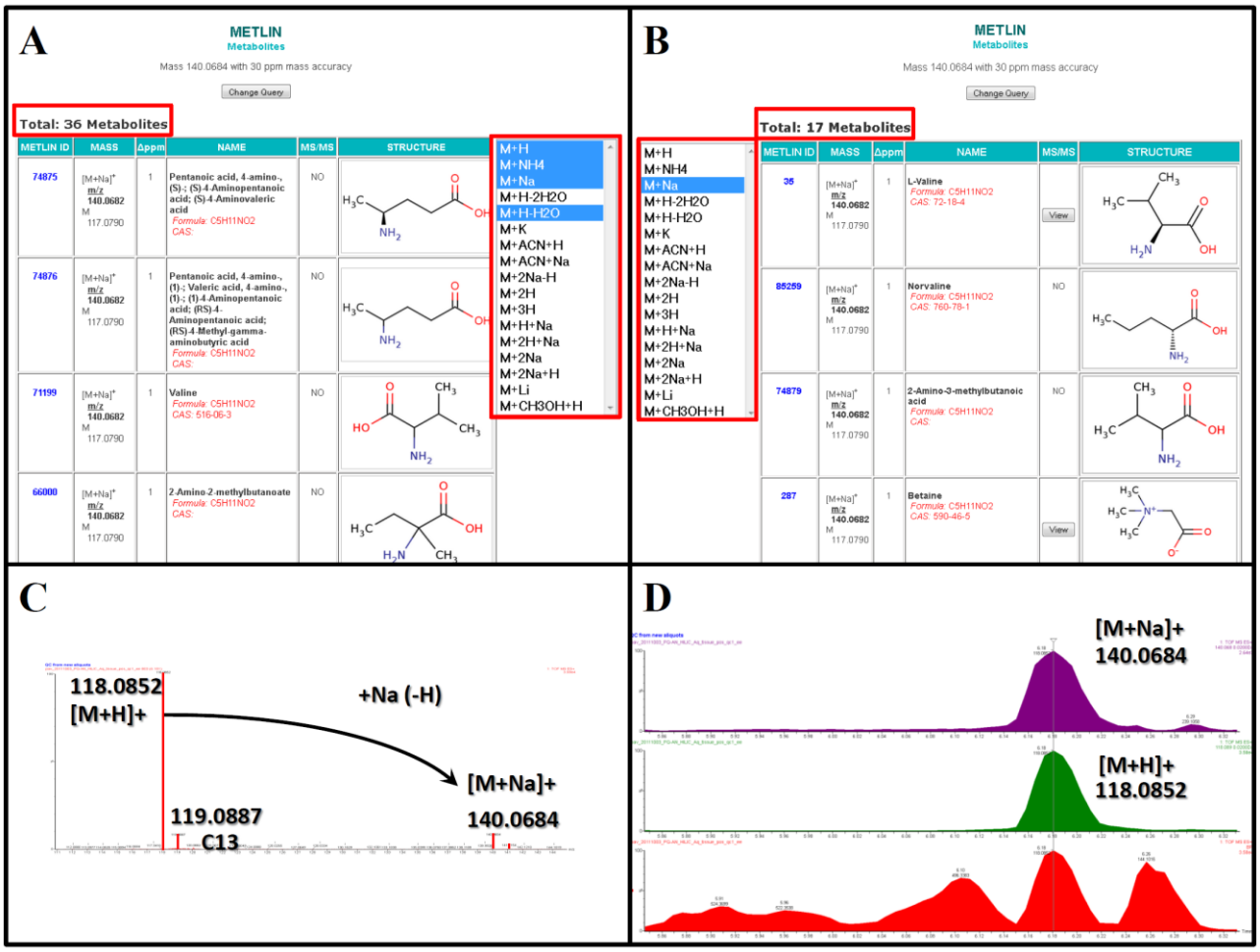


*Figure 2. 15: Reducing possible hits for metabolites matching the criteria for m/z 140.0684[111]. A) Using four possible adducts a total of 36 hits are obtained. B) When using only the possibility of sodium adduct hits are reduced to 17. C) Information about the adduct form can be obtained by the spectrum and D) chromatography can help verify that the two ions originated from the same molecule.*

### *2.2.4.2.2 ESI polarity mode of MS detection*

When obtaining the search results from databases, the analyst should first enquire if the possible matching metabolites can actually ionise in the detected polarity mode. This piece of information is not available by databases, which might be just presenting possible calculations for aiding the user. Figure 2. 16 shows all the possible ion adducts as calculated by the Metlin database[111] for stearic acid. Although the protonated ion is also demonstrated, it is known that carboxylic acids do not sufficiently ionise on positive mode. Therefore, the analyst should critically review results from databases and exclude matching to ions that cannot be practically formed.

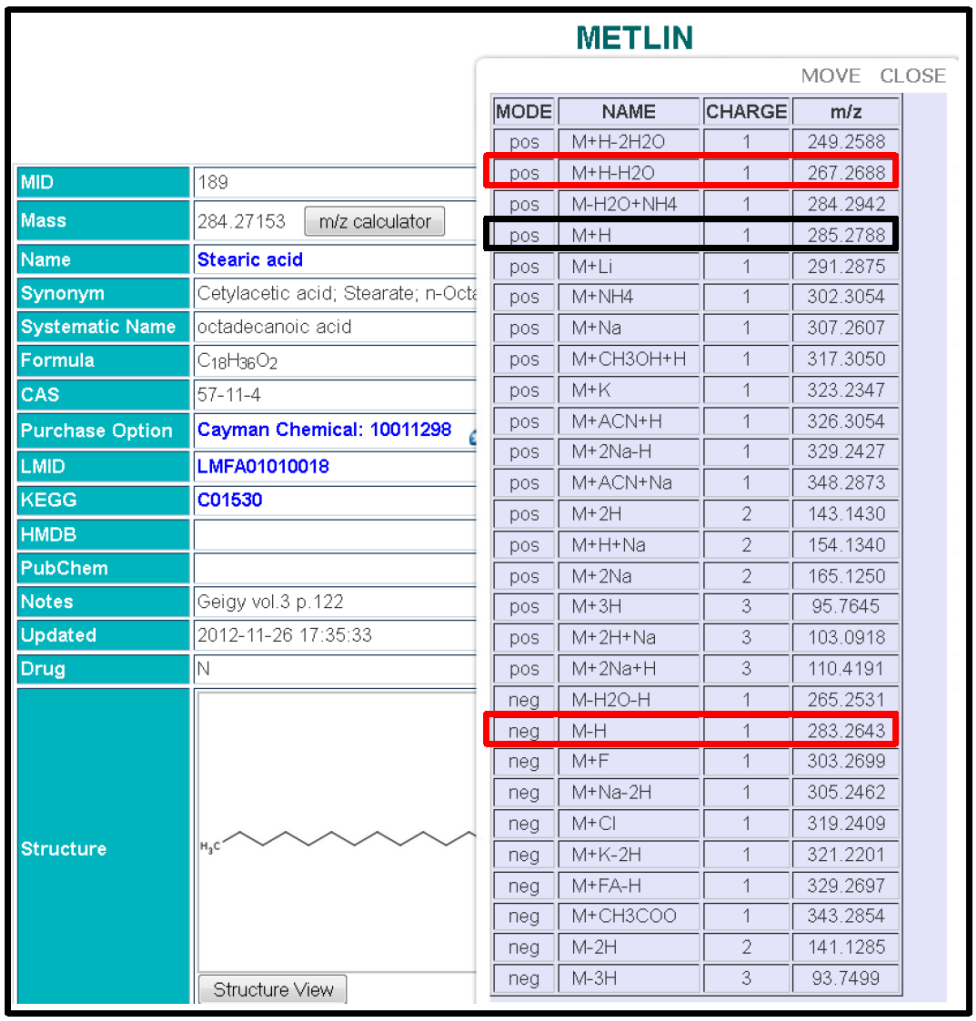


*Figure 2. 16: An example of a database (Metlin[111]) calculating adducts that might not be necessarily formed for the polarity MS mode used for analysis.*

### *2.2.4.2.3 Biological matrix and species*

It is also essential to make sure that the matched metabolite is compatible with the analysed biological matrix and species. When it comes to the biological matrix, knowledge of basic biochemistry can be very helpful. For example, the detected m/z of 314.1241 was matched to two potential metabolites. One of them and with less mass error is a glucuronide. Keeping in mind that

the biological matrix in this case was a urine sample, where numerous non-polar compounds are excreted as glucuronides, the analyst could focus more on proving this as a possibility rather than a less polar match (norcisapride).

On the other hand, differences on expressed metabolites between species, could also work on the analyst's advantage. For example, the metabolome of humans (*Homo sapiens*), *Drosophila melanogaster*, bacteria species, should be expected to have substantial diversion. The lipid class of sphingolipids, for instance, would manifest different lengths of backbone, with the 18 carbon chain largely detected in humans, whereas the 14 carbon chain would be detected in *Drosophila melanogaster*. For both species, the 16 carbon chain can be detected as the sphingolipid backbone. Lastly, before one excludes any metabolite as abnormal for a species/matrix, should first ask if they could be present due to environmental interactions, symbiosis or any sort of intervention.

##### *2.2.4.2.4 Sample preparation*

Sample preparation can also assist towards narrowing down possible metabolites, matched by databases. The most frequently applied extraction methodologies nowadays are liquid-liquid extraction and solid phase extraction (SPE). For liquid-liquid extraction, the solvents used can provide an idea of the physicochemical properties of the molecules extracted with high yield. On the other hand, if SPE is applied, the stationary phase and eluents can provide this information.

Sample preparation can induce features due to materials and handling. In this case, the analyst can easily determine and exclude such ions by running preparation blank samples. This means that the same solvents and procedure would be applied without the additions of sample. In this manner, impurities from solvents and compounds generated from the procedure, such as plasticisers, can be identified.

##### *2.2.4.2.5 Chromatographic retention time*

The retention time of an ion could provide valuable information and guidance towards structural assignment, reflecting metabolite physicochemical properties and structure. This could be a tedious task as it requires experience, critical review of the chromatographic system used, especially stationary and mobile phase, and experimental characterisation of the method applied. However, some simple rules for specific chromatographic systems could further assist the analyst. For example, when a reversed phase system is used, one can expect that less polar compounds to be retained longer in the column (Figure 2. 17). On the other hand, when a HILIC system is used, molecules with higher polarity would generally be less retained in the column (Figure 2. 18); thought partition mechanisms with HILIC systems can be more complicated. A practical option

would be for the analyst to run mixtures of standards in order to understand the interactions occurring in chromatographic systems.

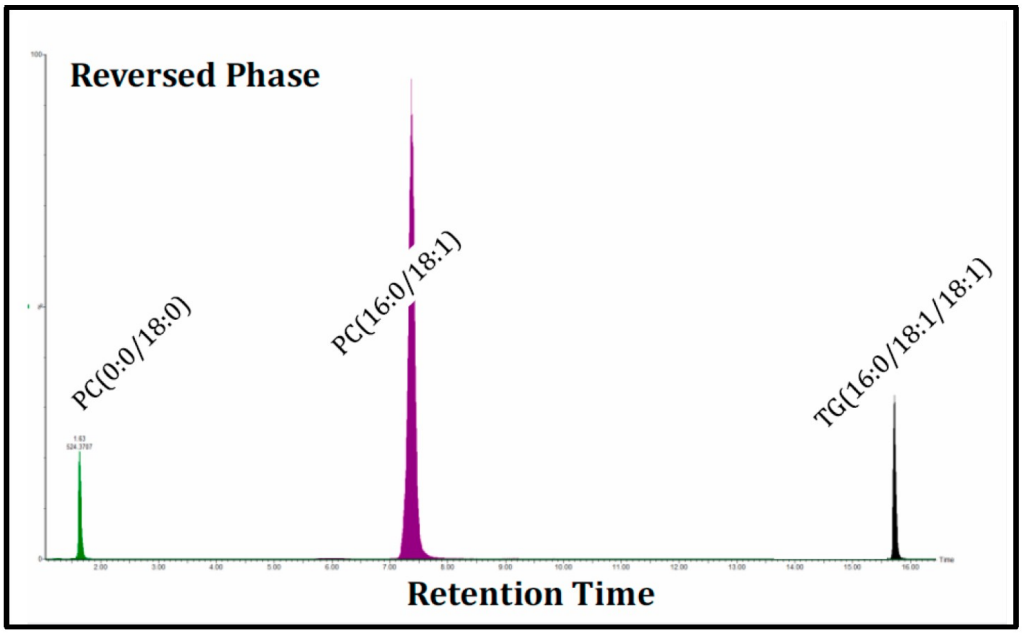


*Figure 2. 17: An example of reversed phase UPLC-MS analysis, where compounds with higher lipophilicity are retained longer in the column.*

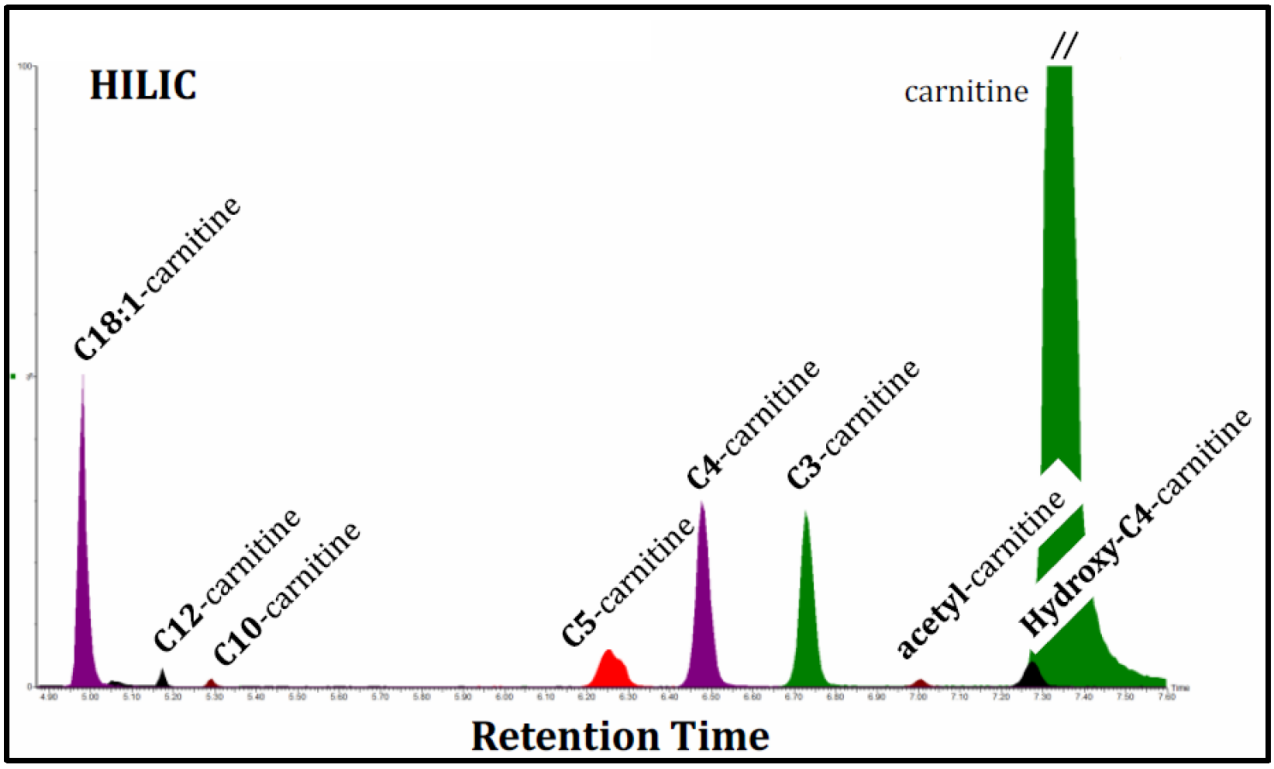


*Figure 2. 18: An example of HILIC UPLC-MS analysis, where compounds with less lipophilicity are retained longer in the column.*

Finally, measurements such as the partition coefficient (logP), which is a measure of lipophilicity of a molecule, can provide a rough idea of the extent of the interactions that can occur in a chromatographic system. It can be used as a guide especially when comparing between compounds. Figure 2. 19 demonstrates how the logP value is consistent with the retention time in a reversed phase chromatographic system.

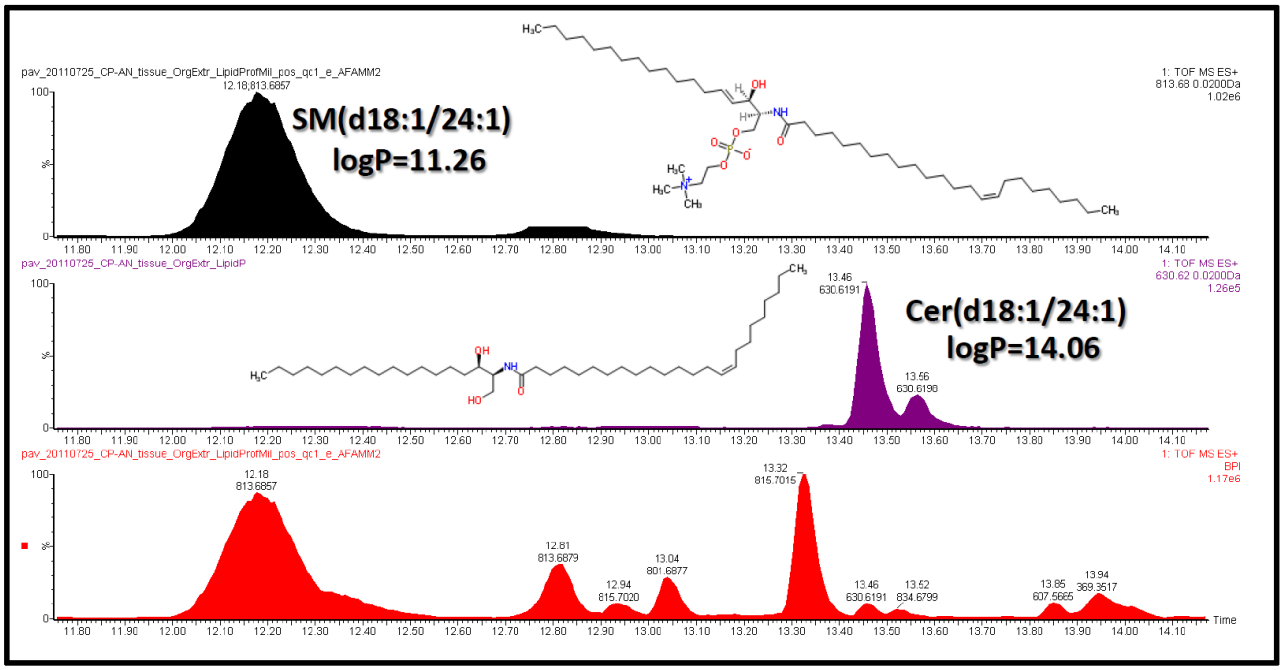


*Figure 2. 19: A demonstration of two compounds with different logP, and how this provides estimation for their distribution between the two phases formed in the column.*

### *2.2.4.2.6 Charge number*

Although in metabolomics the scientists generally deal with singly charged molecules, it is sometimes possible for a molecule to have a double (or higher) charge. For this to occur, more than one sites with structural groups that could ionise, are required. This is generally the case with macromolecules (Figure 2. 20A). Nonetheless, this is essential information for identifying the structure of a molecule and can be easily obtained by observing the isotopic profile of an ion. Keeping in mind that the MS measures the mass-to-charge ratio (m/z) a doubly charged molecule, unlike singly charged, will have the $^{13}C$ isotope at an additional m/z of approximately 0.5 rather than 1 (Figure 2. 20B). This can provide additional assistance for structural assignment of detected ions.

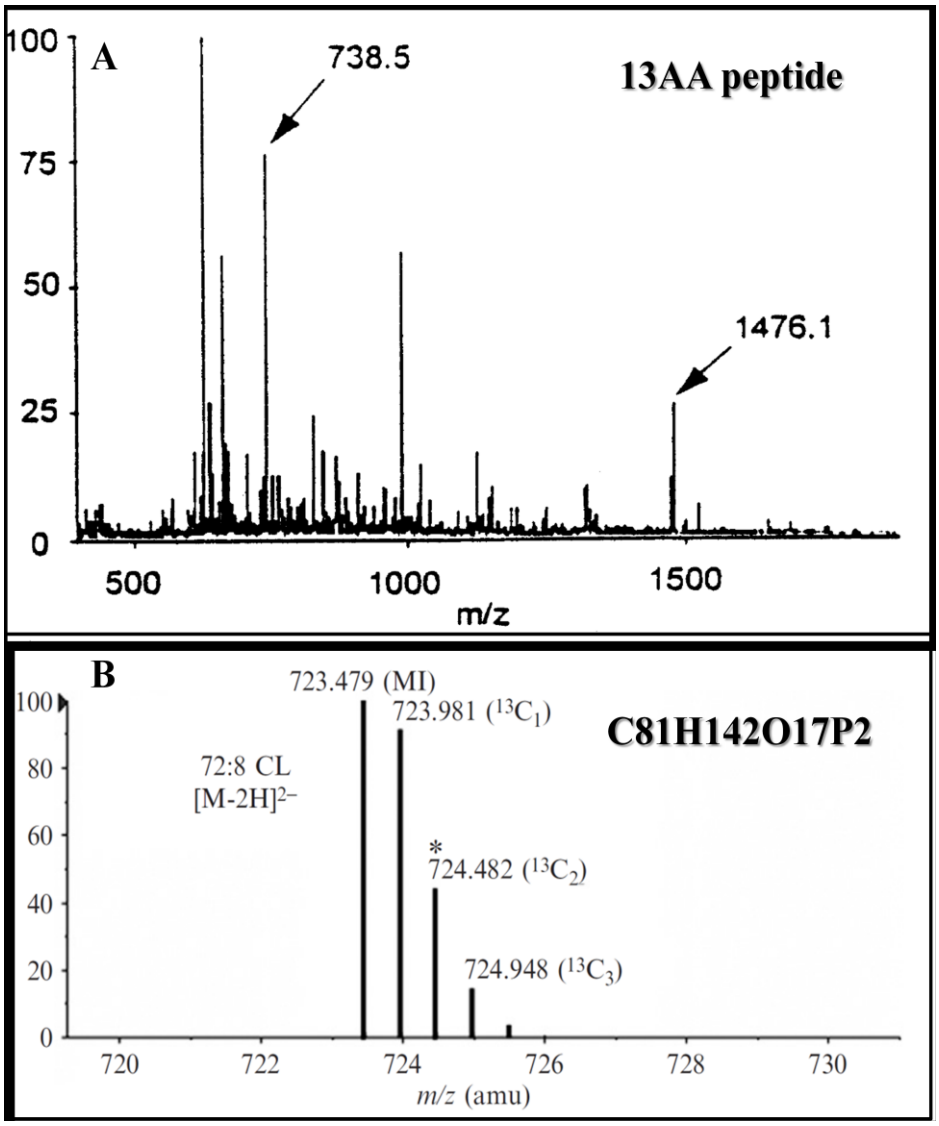


*Figure 2. 20: A) A polypeptide can form doubly charged ions due to several number of structural groups that can ionise. B) An example of a metabolite (cardiolipin), demonstrating that when there is a double charge, the second isotopic signal will be located at M+0.5 m/z.*

### *2.2.4.2.7 Mass deficiency*

Mass deficiency can be described as the difference between a compound's exact mass and its nominal mass. It is yet another benefit of high resolution/high mass accuracy MS. Mass defect can provide tremendous assistance in metabolite structural identification, from aiding assigning the corresponding elemental composition to detecting adducts. An example frequently encountered in metabolomics is assigning alkyl ether or alkyl ester to phosphatidylcholines (PC). The PC(O-18:0/18:0) with molecular formula of C44H91NO7P+ (protonated ion), and PC(17:0/18:0) with molecular formula C43H87NO8P+ both have a nominal mass of 776. Additionally, they would experience similar range in retention times (provided that an untargeted method is used), and present similarities in characteristic fragments. However, their exact mass would be significantly different (ΔDa=0.0364), and easily detected by modern TOF instruments. The analyst could

therefore gain guidance on the structure and specifically the elemental composition of the PC detected.

*2.2.4.2.8 Elemental composition and isotopic pattern*

Calculating the elemental composition or even just being able to confirm the existence of a specific atom in a molecule could prove significantly useful. The existence of $^{13}C$ will display an additional ion after the main, parent ion. Knowing that $^{13}C$ is present at a percentage of 1.1% the number of carbons in a molecule can be calculated simple by dividing the proportion of 2$^{nd}$ to 1$^{st}$ parent ion isotopes by 1.1. Additionally, characteristic isotopic patterns can be obtained if chloride (Figure 2. 21), bromide and sulphur are present. This would be the case even if they are present as adducts.

Recently, software that can suggest an elemental composition for a specific isotopic pattern has been made available. However, the user has to be careful and check the quality of the data provided. High intensity saturated ion signals, as well as low intensity signals that their isotopic signals cannot rise above noise, could lead to a wrong result. Additionally, such software also uses mass accuracy for their calculations (generally provided as a 'score'), so the user should make sure that mass accuracy is satisfactory.

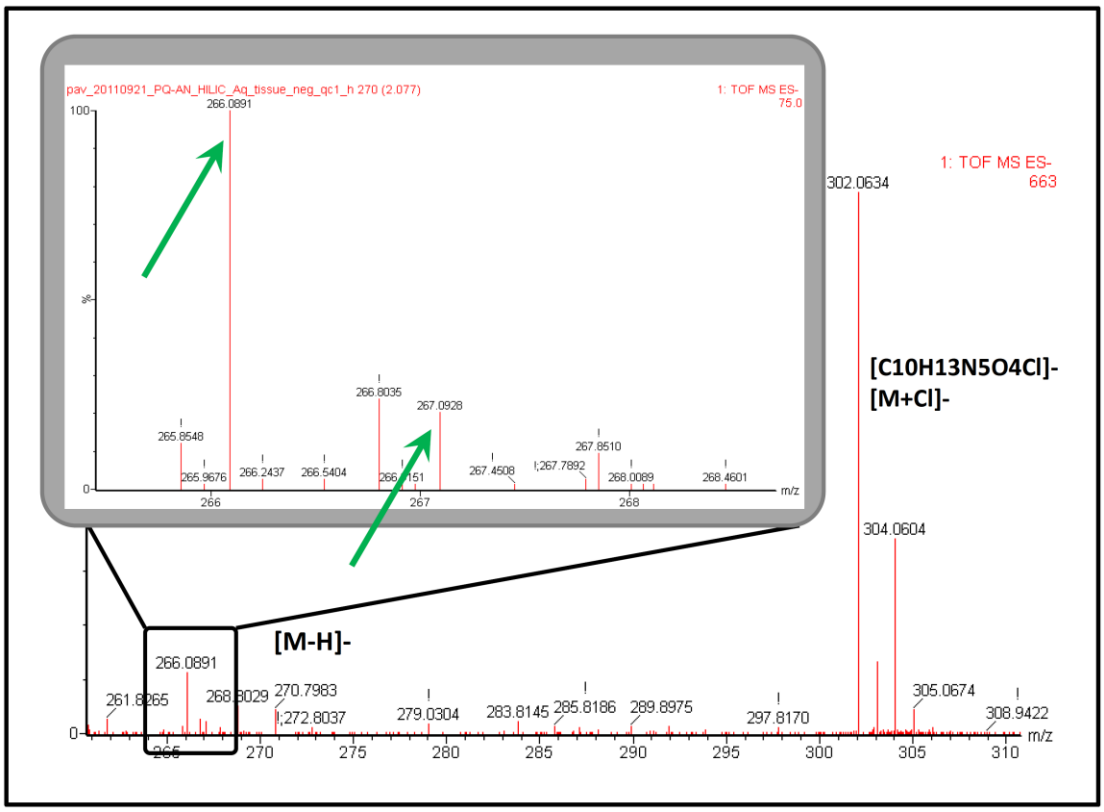


*Figure 2. 21: An example for the spectrum of adenosine in negative MS mode, where a chloride adduct is formed demonstrating a characteristic isotopic pattern.*

*2.2.4.2.9 Tandem MS*

Metabolite fragments and fragmentation patterns can be characteristic and very informative for the structure of a molecule. The information they can provide can be specific and by further use of tandem MS databases and/or authentic standards, could lead to structural assignment. The MS/MS and $MS^E$ are the most frequently used approaches applied, as described in Section 2.2.2.3 . In Figure 2. 22 the example for ion 140.0684 from Section 2.2.4.2.1 (identified as sodium adduct) is used to illustrate how using fragments from the MS/MS spectrum of the ion, and matching them to fragments from an online database (Metlin[111]), led to the structural assignment for betaine. However, if such matching is not possible, then it lays on the skills of the analyst to decipher the structure of the analyte. Software and algorithms using basic fragmentation rules can assist when assembling a molecule from fragments is needed.

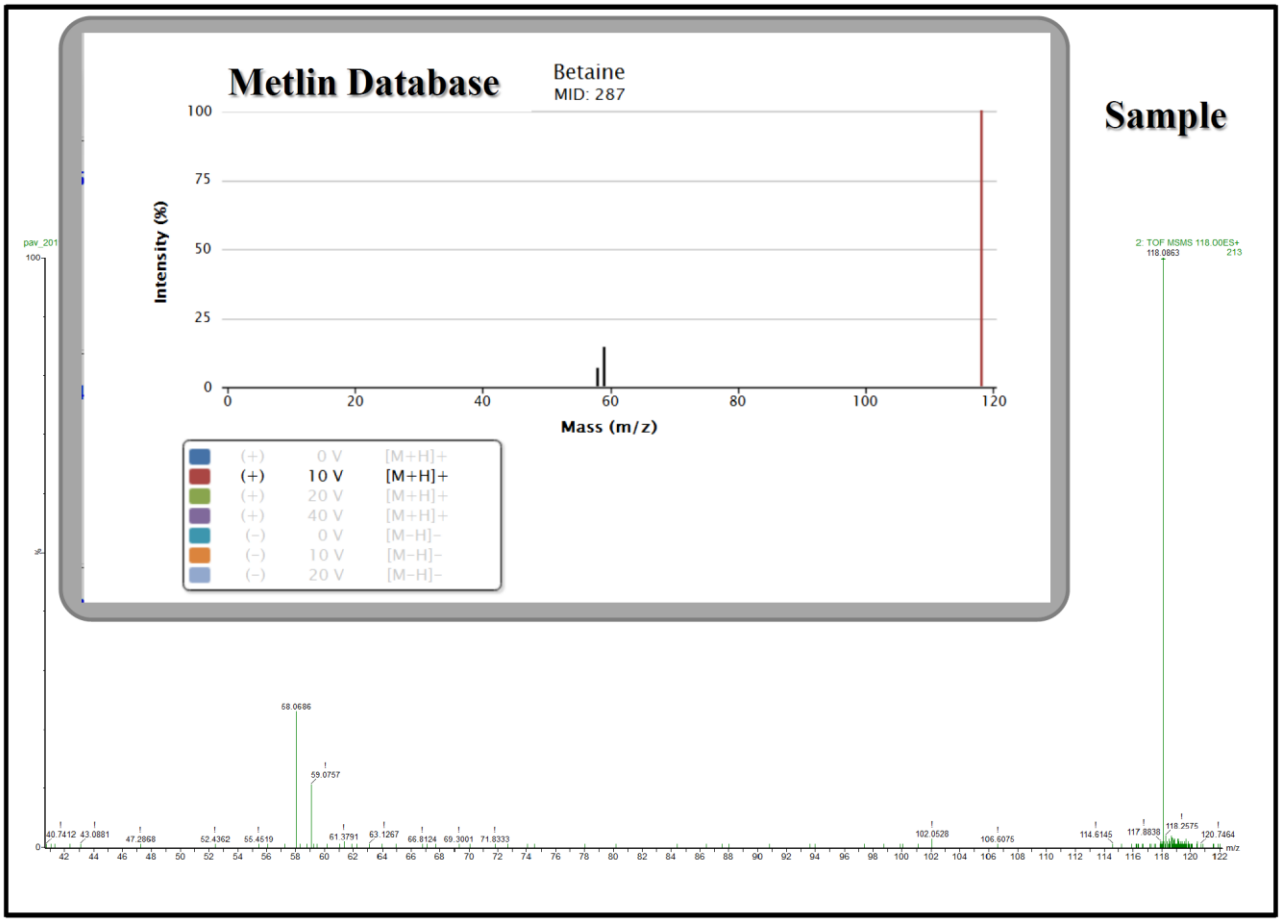


*Figure 2. 22: Matching the MS/MS spectrum of betaine to the MS/MS spectrum from Metlin database[111].*

*2.2.4.2.10 Matching to authentic standards*

In order to report a metabolite structure with confidence, it is of importance to match the detected metabolite to the authentic standard in terms of tandem MS spectrum, and retention time, when chromatography is used. Firstly, it is advised to opt for matching retention time, and then for

matching to tandem MS spectrum. This is due to occasional difficulties of obtaining MS/MS spectra from samples as the concentration of the analyte can be low. In Figure 2. 23 the example for ion 140.0684 from Section 2.2.4.2.1 (identified as a sodium adduct), demonstrates matching the ion of betaine to the corresponding authentic standard by retention time and MS/MS spectrum. This approach can provide the confidence needed for an assignment.

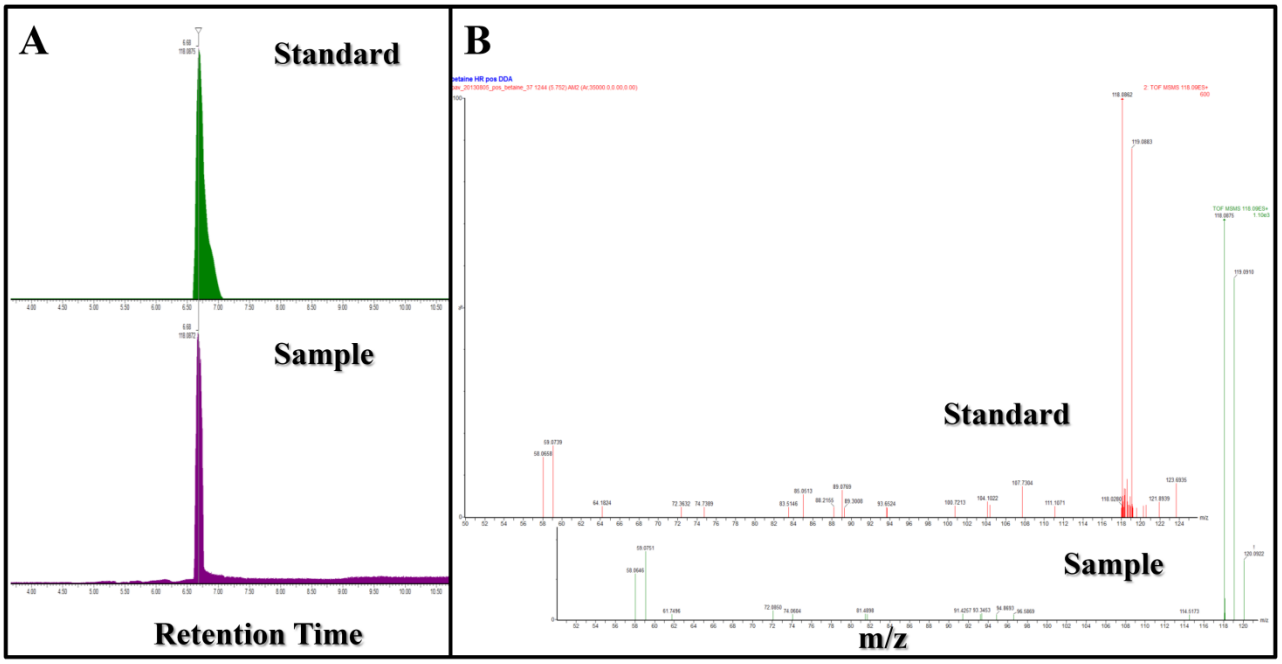


*Figure 2. 23: Matching betaine to the authentic standard. A) Matching retention time of betaine from standard and sample, and B) matching MS/MS spectra.*

### *2.2.4.2.11 Structural isomers*

An important parameter that needs to be addressed by the analyst is whether existing isomers (including stereoisomers) can be identified. This can be very important as they can show numerous biological interpretations. There are several occasions where isomers can be coeluting when chromatography is applied. Therefore, the analyst should be able to discriminate between different structural forms of isomers. This can occasionally be possible from MS/MS experiments. Figure 2. 24A demonstrates how the two isomers of citrate and isocitrate present differences in their fragmentation pattern. However, this is not the case with butyrylcarnitine and isobutyrylcarnitine which have practically identical MS/MS spectra (Figure 2. 24B). In this case it should be determined if chromatographic separation can be achieved (Figure 2. 24C). If not, the possibilities should be investigated and reported.

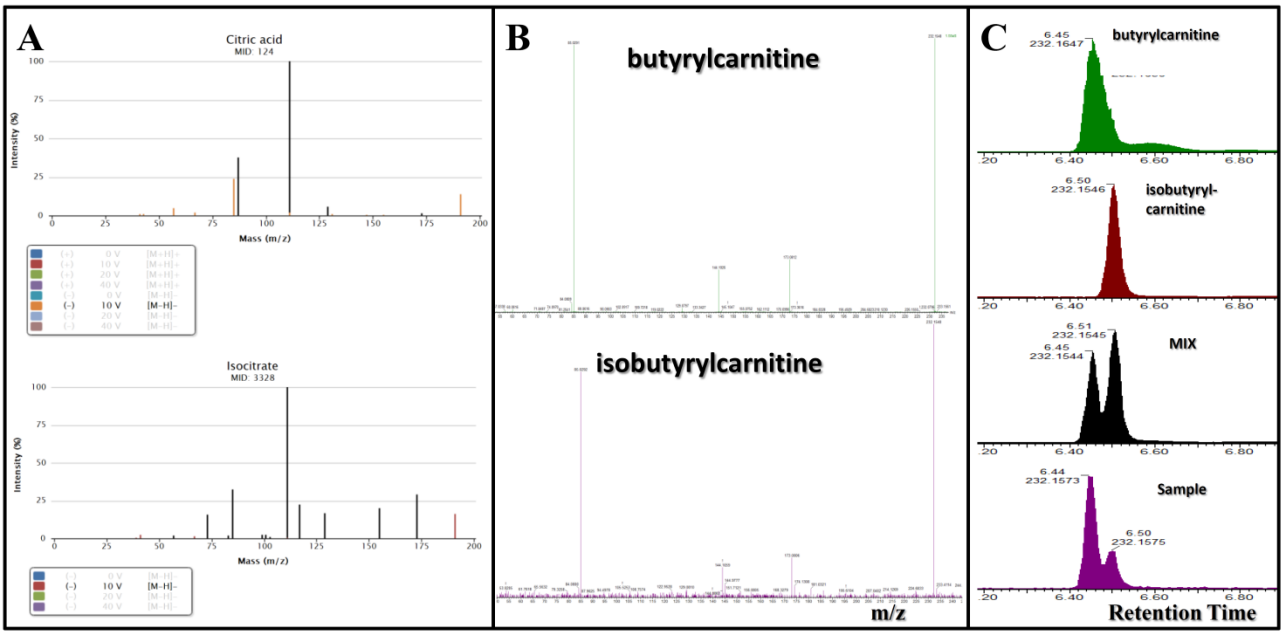


*Figure 2. 24: A) Spectra of citric acid and isocitrate[111] present differences in formed fragments. B) Spectra of authentic standards for butyrylcarnitine and isobutyrylcarnitine present the same fragmentation pattern. C) Isomers for butyrylcarnitine and isobutyrylcarnitine can be resolved with chromatography.*

### *2.2.4.2.12 Examples of structural assignments*

#### *2.2.4.2.12.1 Phosphatidylcholines*

Lipophilic compounds such as phosphatidylcholines (PC) can ionise very well with ESI-MS. They are therefore frequently detected by untargeted metabolic/lipid profiling analyses. Characteristic fragments for PCs can provide structural information for assigning. As an example the ion 784.59 will be used. This ion is detected with two chromatographic peaks (Figure 2. 25). A database search (Metlin[111]) for this m/z provided 65 metabolites that could be matching in terms of mass (Da). These included phosphatidylcholines, phosphatidylethanolamines, phosphatidylserines and phosphatidic acids. PCs can be easily detected by their characteristic fragment in tandem MS, of the loss of phosphocholine head-group (184.07 Da). This narrows results down to 43 PCs. What remain to be determined are the fatty acyl chains of the two detected PCs. Although in positive mode this can be possible by detecting the positively charged fragments after neutral loss of one of the chains, in negative mode the loss of the fatty chain can be sensitively detected. This presumes that both polarity modes are conducted with the same method, in order to easily identify by retention time, identical metabolites. For this example, the first ion was identified as PC(18:2/18:1) by positive mode MS/MS fragments (Figure 2. 26), to be further confirmed by negative mode (Figure 2. 27). The second ion was identified as PC(16:0/20:3) in the same way. In positive mode MS/MS spectra (Figure 2. 28) the difficulty for obtaining the length of the chains can be observed, due to the

reduced ability of the formed ions to ionise. On the contrary, in negative mode fatty acyl chain fragments can ionise very well and are well above the noise level (Figure 2. 29).

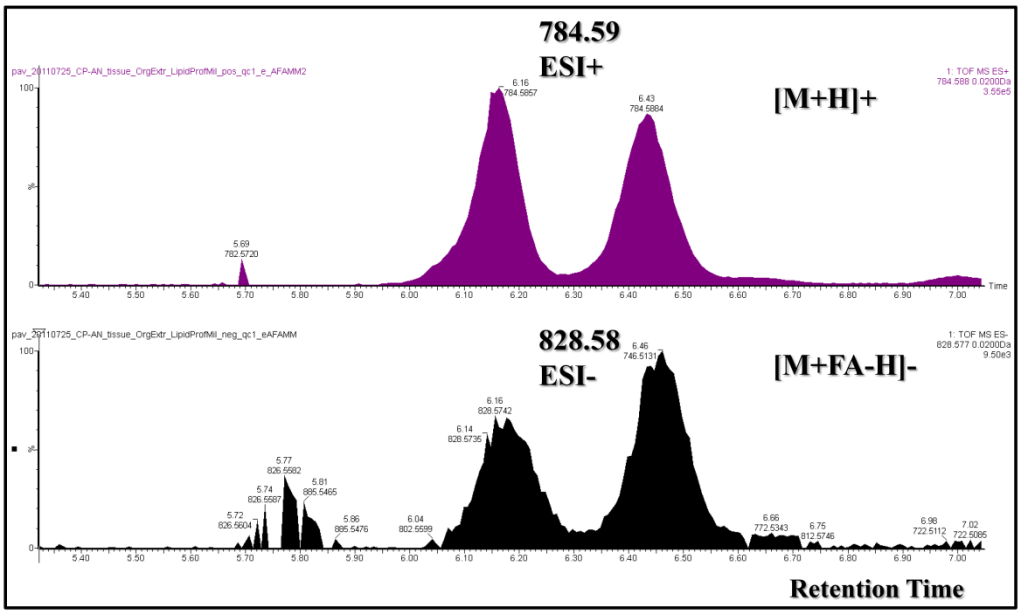


*Figure 2. 25: Extracted ion chromatograms for to isotopes detected with m/z of 784.59 in positive mode and 828.58 in negative mode.*

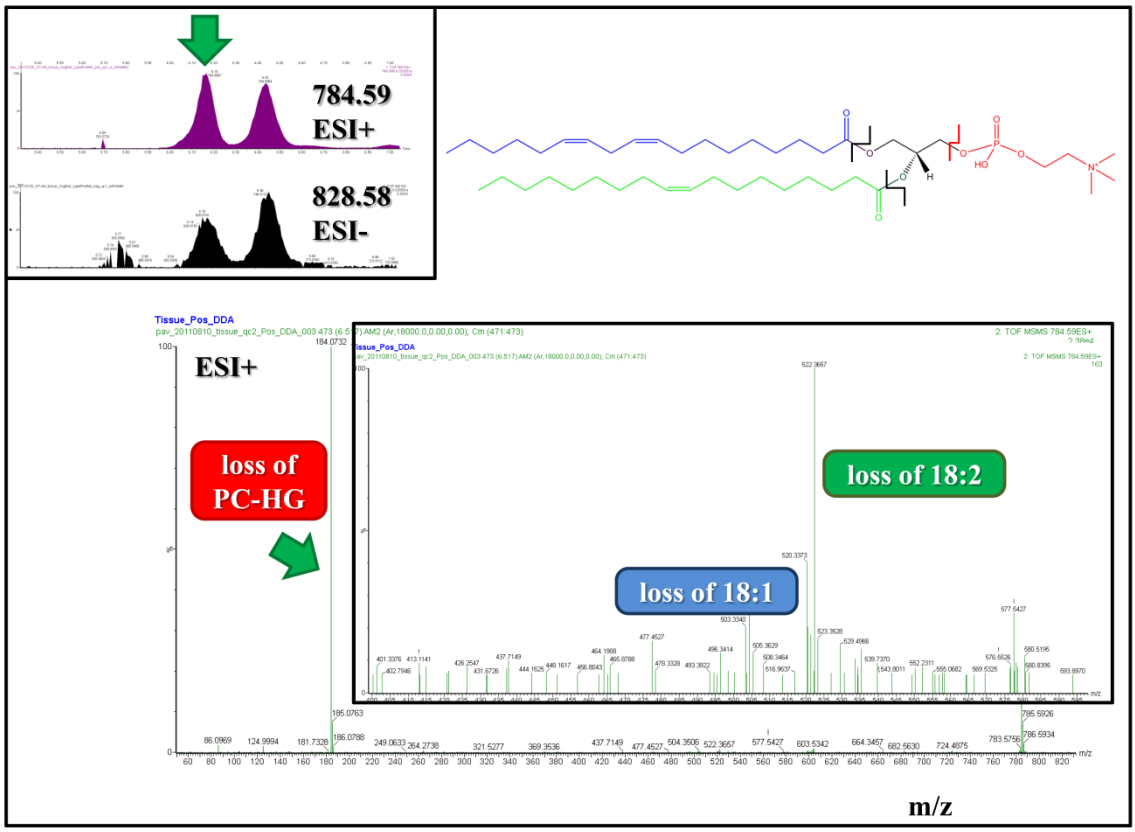


*Figure 2. 26: The MS/MS spectrum of the first ion on positive mode. This ion can be easily identified as a phosphatidylcholine (PC). This information can be obtained from the ion formed by the loss of the phosphocholine head group (PC-HG). For identifying the fatty acyl chains the spectrum is noisy, but indicates the molecule is the PC(18:2/18:1).*

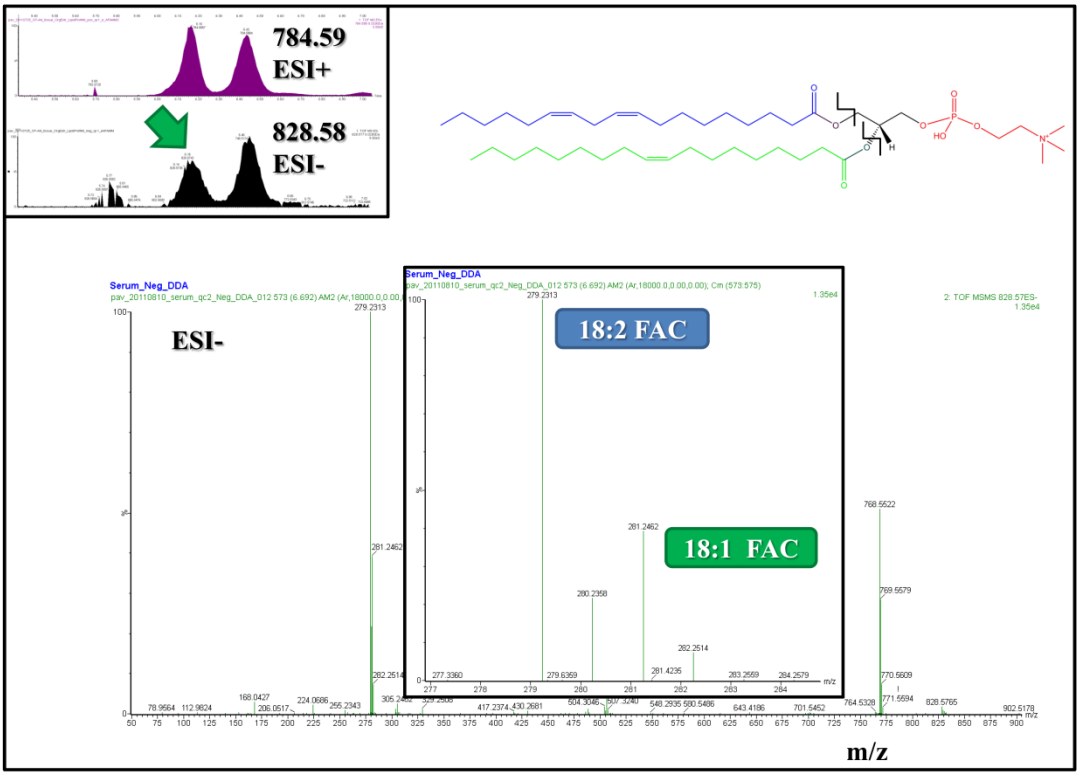


*Figure 2. 27: The MS/MS spectrum of the first ion on negative mode. This spectrum cannot assist for identifying the class of the lipid. However, it can provide clear information about the fatty acyl chains.*

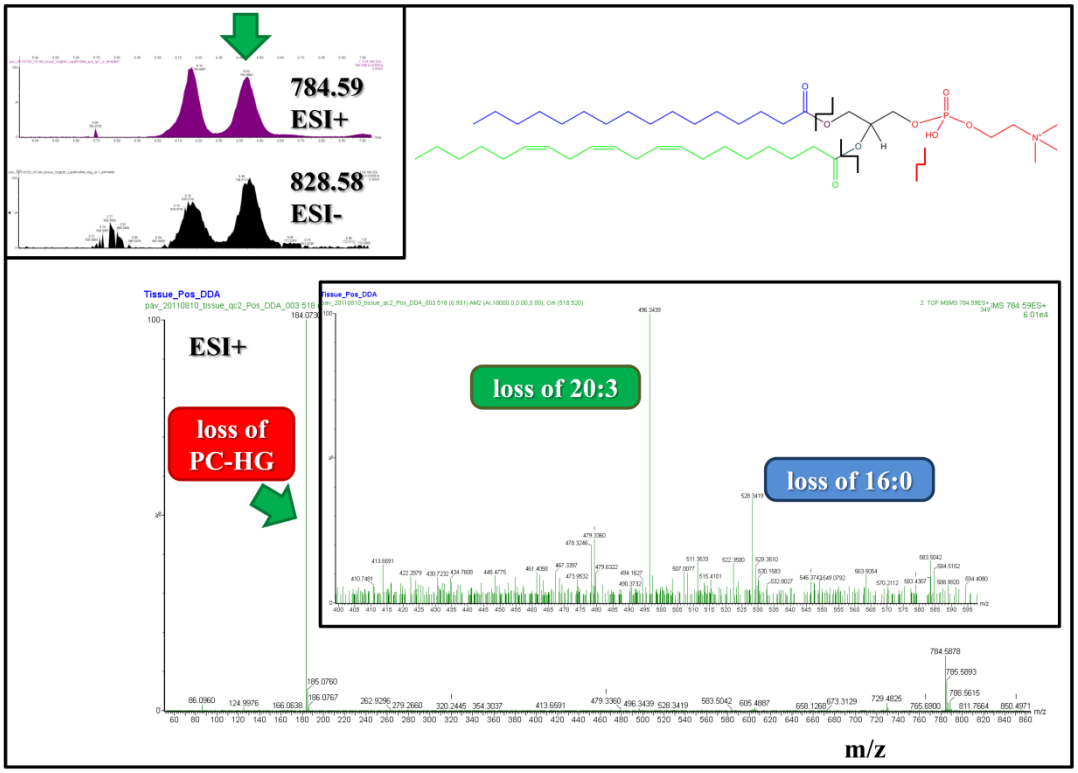


*Figure 2. 28: The MS/MS spectrum of the second ion on positive mode. This ion can be easily identified as a phosphatidylcholine (PC). This information can be obtained from the ion formed by the loss of the phosphocholine head group (PC-HG). For identifying the fatty acyl chains the spectrum is noisy, but indicates the molecule is the PC(16:0/20:3).*

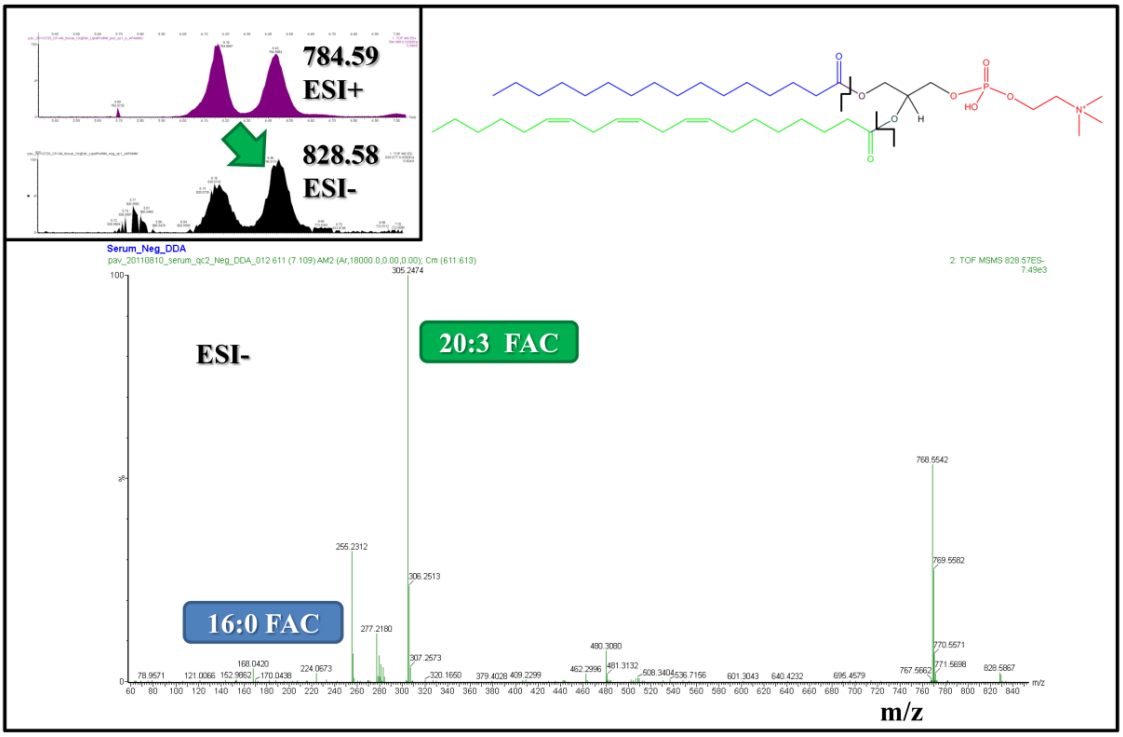


*Figure 2. 29: The MS/MS spectrum of the second ion on negative mode. This spectrum cannot assist for identifying the class of the lipid. However, it can provide clear information about the fatty acyl chains.*

### *2.2.4.2.12.2 Sphingomyelins*

The lipid subclass of sphingomyelins (SM) is another group of metabolites detected with high sensitivity in ESI-MS. Like PCs, SMs have the characteristic fragment of phosphocholine head-group loss, in MS/MS experiment. It is very easy to be discriminated from PCs since they will be detected as odd m/z values. This would be expected as they have two nitrogens atoms in the structure (nitrogen rule). Further on, sphingomyelins will present a loss of their fatty amide (Figure 2. 30). Additionally, as it is the case with all sphingolipids, a characteristic fragment derived from the backbone of the molecule (Figure 2. 30) will complete the information needed to be able to assign a SM.

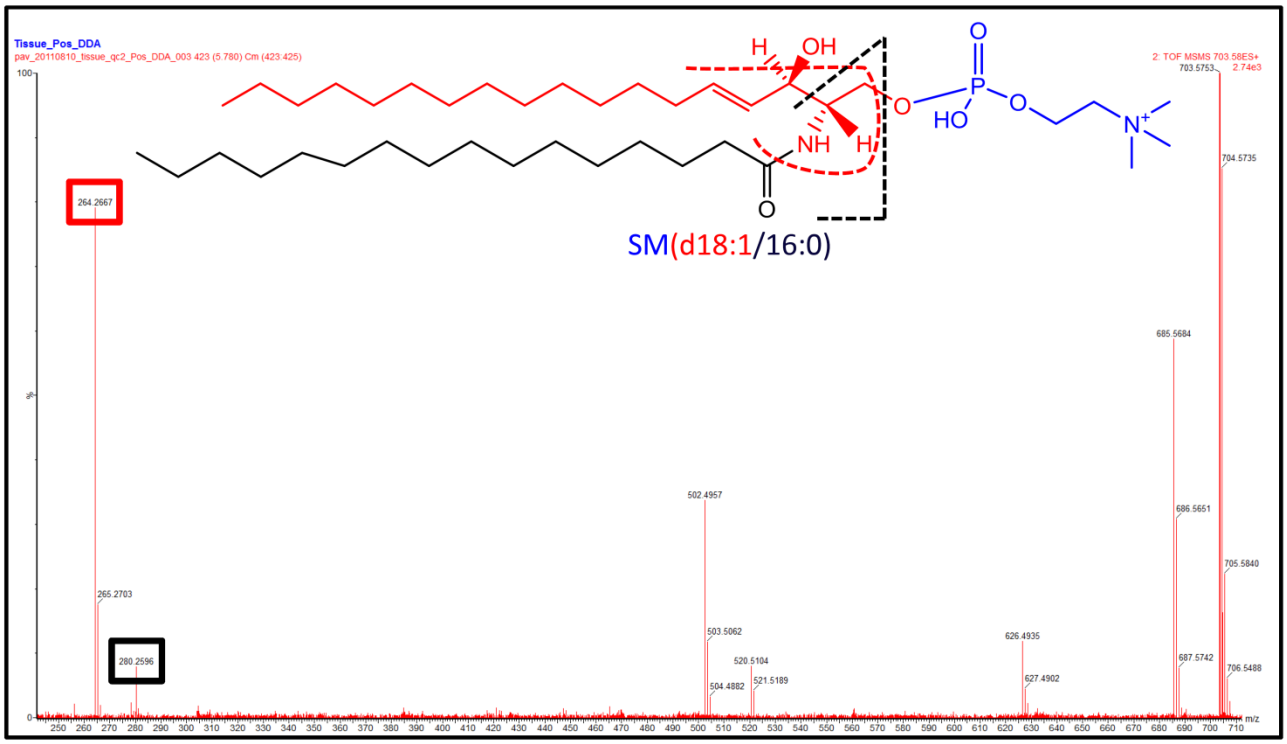


*Figure 2. 30: The MS/MS spectrum of a sphingomyelin (SM) on positive mode. Characteristic fragments that can assist on structural assignment are the loss of the phosphocholine head group (not shown in figure), the loss of the d18:1 ceramide backbone (m/z 264) and the loss of the fatty amide (m/z 280). This fragmentation pattern indicates that the molecule is the SM(d18:1/16:0).*

### *2.2.4.2.12.3 Structural assignment of lipids*

Several lipids have characteristic fragments that can provide information of their structure, lipid group, and chains incorporated. Table 2. 1 demonstrates characteristic fragments of lipid classes that can aid structural assignment. Additionally, retention time can assist as demonstrated in Section 2.2.4.2.5 Lipid classes like triglycerides may not have a characteristic fragment to define their class, but their high lipophilicity, translated to an increase in retention time. They also have a fragmentation pattern that can provide the needed information for the fatty acyl chains incorporated, by providing positively charged ions after loss of one of their chains. By combining these fragments the needed information can be provided for assignment.

*Table 2. 1: Table of tandem MS fragments and neutral losses, characteristic to specific lipid classes and subclasses.*

| **Lipid Class** | **Characteristic Fragments (m/z)** |
|---|---|
| **ESI+** | |
| Phosphatidylcholines<br>Sphingomyelins | 184 |
| Cholesterol esters | 369 |
| Phosphatidylethanolamines | 141 (neutral loss) |
| acyl-carnitines | 85 |
| | |
| **ESI-** | |
| Phosphatidylglycerols | 171/153 |
| Phosphatidylinositols | 241 |
| Phosphatidylserines | 87 (neutral loss) |

### *2.2.5 UPLC-MS applications in metabolic profiling*

UPLC-MS was well-received from the metabolic profiling community from its early days[96]. UPLC-MS provided the superior separation, robustness, reproducibility and sensitivity demanded for the metabolic profiling concept to move forward[114]. The fast turnaround time per sample turned UPLC-MS into a tool appropriate for epidemiological studies greatly needed in the field of metabolic phenotype to surpass issues of biological variation, such as differences due to different gender, age, or ethnic background[118]. In combination with automation in data processing it is now considered a recognised platform for metabolic profiling studies with a wide range of applications [119-121] and in a wide range of biological matrices[16, 122].

## 2.3 Nuclear Magnetic Resonance spectroscopy (NMR)

Nuclear magnetic resonance (NMR) is based on the physical phenomenon occurring with specific nuclei when they experience a magnetic field (**Bo**). This phenomenon occurs only to nuclei that have a spin quantum number (**I**) different to zero. This number is determined by the number of protons and neutrons of an atom (Table 2. 2). A nucleus with spin quantum number of **I** will have 2**I**+1 energy levels in a magnetic field. These energy levels are described by the magnetic quantum number $\mathbf{m_I}$=I, I-1, I-2, ..., -I.

*Table 2. 2: Characteristics of different nuclei and effect on quantum number*[123]

| Number of protons | Number of neutrons | Spin quantum number mode | Example |
|---|---|---|---|
| Even | Even | No spin | $^{12}C(0)$, $^{16}O(0)$ |
| Odd | Even | Half-integer spin | $^{1}H(½)$, $^{31}P(½)$, $^{23}Na(^{3}/_{2})$ |
| Even | Odd | Half-integer spin | $^{13}C(½)$, $^{17}O(^{5}/_{2})$ |
| Odd | Odd | Integer spin | $^{2}H(1)$, $^{14}N(1)$ |

The most frequently used atom for NMR, in metabolic profiling, is hydrogen ($^{1}H$). Hydrogen has I=1/2, which means that in a magnetic field its nucleus has either of two energy levels with $m_I$ of 1/2 and -1/2. The higher energy state (**α**) with $\mathbf{m_I}$=-1/2 is the one opposed to Bo, while the lower state (**β**) with $\mathbf{m_I}$=+1/2 is aligned with Bo (Figure 2. 31).

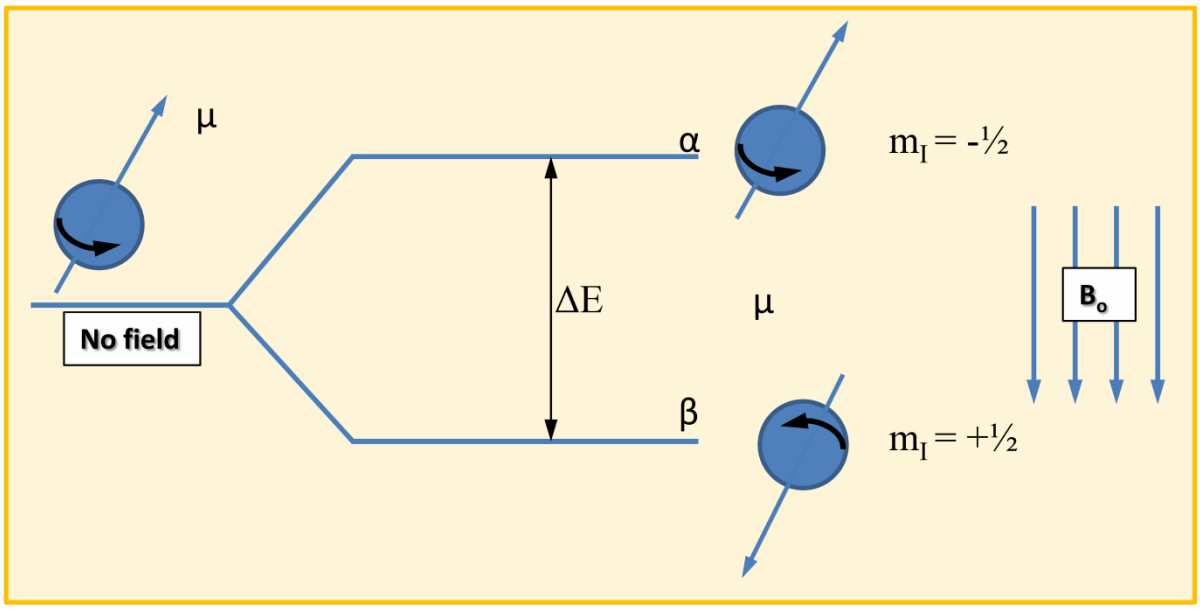


*Figure 2. 31: Scheme of energy states of a nucleus with I=1/2 (such as proton). Two energy states are present when the nucleus is inside a magnetic field. The highest energy α ($m_I$=-1/2), and lowest energy β ($m_I$=+1/2). The difference between the two energy states is described as ΔE.*

When a population of protons are within a magnetic field, some of them will occupy the **α** state and some the **β** state. The population distribution between these two states is given by the Boltzmann distribution (Equation 2. 4).

$$\frac{N_\alpha}{N_\beta} = e^{-\frac{\Delta E}{kT}}$$

*Equation 2. 4: Boltzmann distribution of nuclei populations between two energy states in a magnetic field.*

Where **ΔE** the difference in energy between the two energy levels, **k** the Boltzmann constant and **T** the temperature in degrees Kelvin.

**ΔE** can be described as **hν**, where h is the Plank's constant and ν is the frequency. This means that, keeping in mind that an exact energy 'package' is needed for a nucleus to get to a higher state (resonate), an exact frequency should be applied. Also, provided that the energy difference is related to the applied magnetic field (**$\Delta E = \gamma \hbar B_o$**; **γ** is a distinctive constant for each nucleus called magnetogyric ratio) the resonance frequency will also be proportional to the magnetic field (**$\nu_{res} = \gamma B_o / 2\pi$**).

Energy is provided to the nuclei in a magnetic field using a short radiofrequency (**RF**) pulse. The nuclei will resonate and after the pulse stops the system will return to equilibrium. Equilibrium will be reached after a specific amount of time, called the relaxation time. During this time the energy of the nucleus will oscillate emitting an RF which referred to as the free induction decay (FID). This FID is converted into frequency (Hz) by Fourier transformation (Figure 2. 31). Note here that, as $B_o$ can affect the resonance frequency, these are always normalised to account for different magnets. This is done by dividing the frequency of a nucleus to the frequency the magnet used. The resulting number is called chemical shift and is usually given in parts per million, **ppm** (or **δ**).

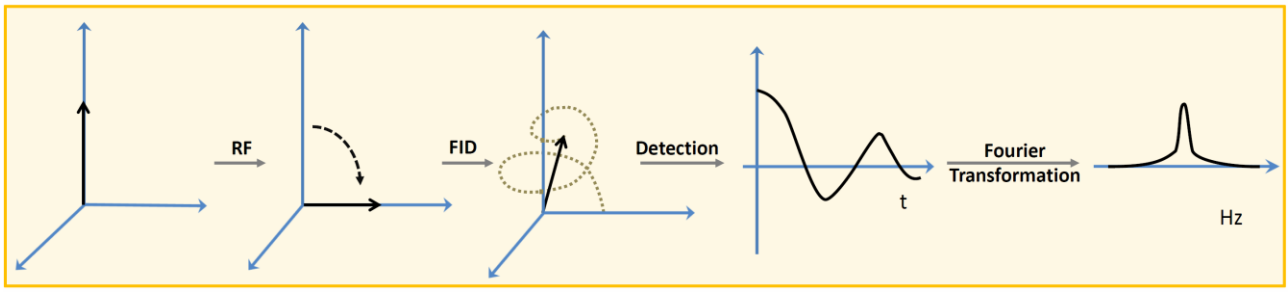


*Figure 2. 32: Stages of NMR acquisition procedure. From left to right: orientation of a nucleus in a magnetic field, application of RF and transition to a higher energy state, free induction decay to settle back to equilibrium, acquisition and Fourier transformation.*

When a magnetic field $B_o$ is applied to a nucleus, the magnetic field experienced by that nucleus is not of the same magnitude in all chemical environments. This is because of the magnetic fields created by electrons circulating the nucleus. This local magnetic field (**B´**) is opposed to $B_o$ and varies according to electrons surrounding the interrogated nucleus. This difference in electron environment is what provides the different chemical shifts that allow definition of the chemical environment for each nucleus. For example, for benzene, electrons are located, due to p orbitals, parallel to the benzene ring. However, protons will be located in the same plane as benzene. Therefore, protons will not be 'shielded' by this local magnetic field from electrons and experience a higher overall field. This is why benzene protons will resonate in higher frequencies. This phenomenon aids the structural assignment of a molecule.

Another feature of NMR that can provide additional structural information is spin-spin coupling. This phenomenon occurs intensively when two nuclei are in close proximity, and function as small independent magnetic fields. These two protons will interact with each other in addition to the magnetic field and this causes splitting of the signal. In proton NMR, the most frequently observed coupling is between proton nuclei that are three bonds away. A proton, $\mathbf{H_A}$, will experience a magnetic field from a proton, $\mathbf{H_B}$. However, the spin from a single proton (I=1/2) can have two energy states. Therefore, two signals will be detected for $\mathbf{H_A}$ (Figure 2. 33). Respectively, if two (equivalent) $\mathbf{H_B}$ protons existed then the possible energy state combinations would be four, with two being equal, and three peaks would be detected (Figure 2. 33).

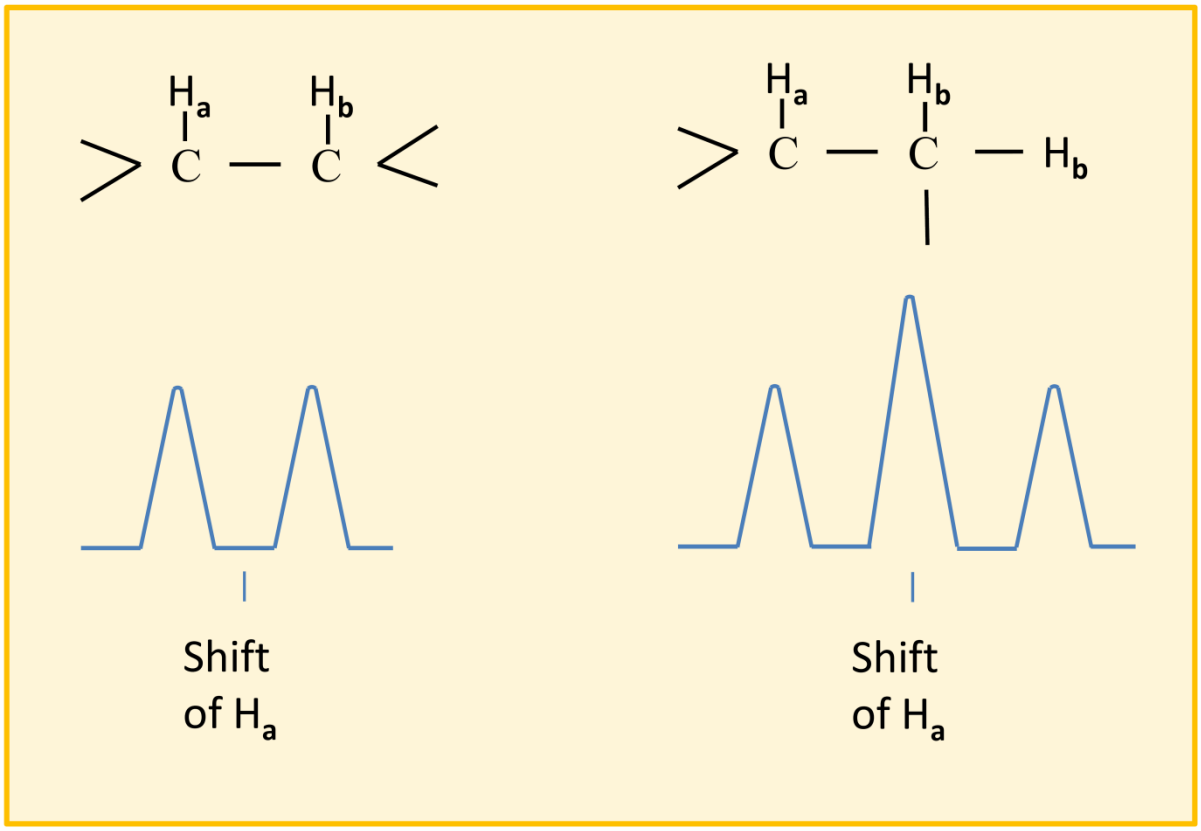


*Figure 2. 33: Representation of spin-spin coupling effect as detected from a proton (Ha), due a proton (Hb), three bonds away. On the left, only one Hb exists and therefore only two spin states. This will lead to a doublet chemical shift. On the right, two equivalent protons exist. Thus, in this case, four different spin orientations can occur. This will give a triplet structure. As two energy states are equal the middle peak which represents those, is larger.*

When additional structural information is needed 2-dimensional (2-D) NMR sequences are employed. 2-D NMR sequences, such as COSY (**CO**rellation **S**pectroscop**Y**), can provide relationships between signals of the same molecule. This makes them really useful for complex solutions, such as matrices commonly analysed by metabolic profiling methods.

Overall, NMR is a useful tool in metabolic profiling approaches. Firstly, signal intensity is proportional to analyte concentration. Additionally, instrument stability and robustness, the non-destructive sample format, the simultaneous structural information obtained, and the absence of signal suppression have it the method of choice for several years. However, some disadvantages are the low sensitivity in a number of compounds, the large sample quantity generally needed, and unavoidable slow scanning times needed due to the slow relaxation of small molecules. Still, NMR technology keeps advancing, with cryoprobe and other developments allowing enhanced sensitivity and it remains an irreplaceable tool for metabolic profiling methodologies.

## 2.4 Sample preparation for metabolic profiling studies

Sample preparation is an important step preceding analysis with the described platforms (UPLC-MS and NMR). It involves applications performed in order to transform the samples in a state that allows achieving the full potential of the applied technique. This can be the case with debris and nonprecipitated proteins. In UPLC-MS this can cause blocking of the column, leading to poor chromatographic resolution, and increase in backpressure leading to system failure. Additionally, not properly prepared samples can cause accumulation of debris in the ESI source, leading to increase background of the MS detector. With NMR, debris and cloudy solutions can cause increase in background and poor resolutions. Protein precipitation using organic solvents, as well as filtering and centrifugation can solve these issues. However, on occasions it is recommended for the analyst to optimise sample preparation for the needs of the analysed matrix.

Metabolite extraction, including solid-phase extraction methods[124], can offer several of the requirements for sample preparation, such as protein precipitation. Additionally, it can be used to increase metabolite recovery used as a pre-concentration technique. However, the selectivity observed for extraction solvent systems can sometimes be a disadvantage. Liquid-liquid extraction is mandatory when metabolite extraction from tissue samples is required. In the following paragraph more information on metabolite extraction from tissue samples is presented.

### 2.4.1 Tissue metabolite extraction for untargeted metabolic profiling studies

Metabolite extraction applied on tissue samples can be a key step for a robust untargeted metabolic profiling analysis, and especially when total metabolome coverage is needed. The extraction conditions applied, including solvent polarity, temperature, pH, as well as tissue lysis procedures[17, 104, 119, 125-127]. When it comes to options for tissue lysis, it appears that the high-throughputness of bead-beating is preferred[17, 126], but also appears to be the method of choice as compared to other tissue lysis methods[104]. When it comes to solvent systems modifications of the Folch method[128] and later the simpler and less exhaustive Bligh-Dyer method[129] are currently applied. A simpler approach attempting to recover a range of metabolites using one solvent system in a single step extraction, for analysis in a cross-platform approach, has been described by Geier et al[104]. They report that for untargeted metabolic profiling analyses using GC-MS, UPLC-MS and NMR an 80% methanol solution is more fit for the purpose of their study. Masson et al[17] described a two-step, consecutive extraction protocol, as being more reproducible than bilayer extraction protocols. The authors split the extraction into a polar metabolite extraction using a water/methanol solvent system, followed by a dichloromethane/methanol extraction for non-polar metabolites. Nonetheless, it is

apparent that the analyst should choose an extraction protocol based on platforms to be applied, and if literature findings are not conclusive or fulfilling, should optimise for the specific tissue of interest. For the experimental purposes of this thesis modifications of the protocol from Masson et al[17] is applied.

# 2.5 Multivariate Data Analysis (MVDA)

## 2.5.1 Statistical Analysis and MVDA

Statistical analysis is important in order to allow one to make an inference of a measurement, to the wider population, after sampling a smaller group. Statistics can be used to define, especially with large datasets, the measured variables that make groups different, and (semi)quantify that difference. These tasks are difficult when datasets are large, and typically cannot be performed manually. Usually powerful computational algorithms are employed to cope with the calculations. The outcome is generally a number or figure that can provide a result within a measure of confidence to the scientist.

MVDA is used in cases where a vast number of variables are obtained from an experiment with multiple samples, but are particularly relevant where the number of variables exceeds the number of samples, as typically encountered in 'omic' datasets. MVDA is employed in order to provide a more comprehensive way (usually in terms of plots and model summary statistics) to extract information from complex datasets. It allows interacting variables to mutually define a class, making the analysis more robust than univariate statistical analyses.

## 2.5.2 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a multivariate data analysis (MVDA) method and functions as an orthogonal (linear) transformation of the original variables. PCA can be used as an unsupervised pattern recognition method (meaning that no *a priori* knowledge about the dataset is used) employed to present (project) multivariate data reducing the number of dimensions in a dataset whilst retaining the important information, since visualisation starts to get difficult for more than three dimensions[130, 131].

For a dataset matrix (**X**), with each row representing a sample and each column a variable (Figure 2. 34), PCA will assign every variable to one dimension in the Euclidean space (meaning the physical space/distances, as can be understood by the human mind, but also its generalisation to higher dimensions). Therefore, each sample will be represented by one point in as many dimensions as there are variables, and placed according to their values (Figure 2. 34 and Figure 2. 35). In PCA the first principal component, in this multidimensional system, will be the vector (direction) that explains the greatest variation of the data, or in other words provide the minimum of residual values (**E**) from the PComp to the observation (Figure 2. 35). This first PComp passes through the mean (Figure 2. 35). Usually for biological data, just one PComp cannot explain the variation of the

model sufficiently. Thus, subsequent PComps are employed to represent the remaining variation[130, 131] and are calculated orthogonally to the previous PComps and pass through the mean. It is important to note that every new PComp should be orthogonal to the previously calculated in order to avoid re-including variation already incorporated in the model.

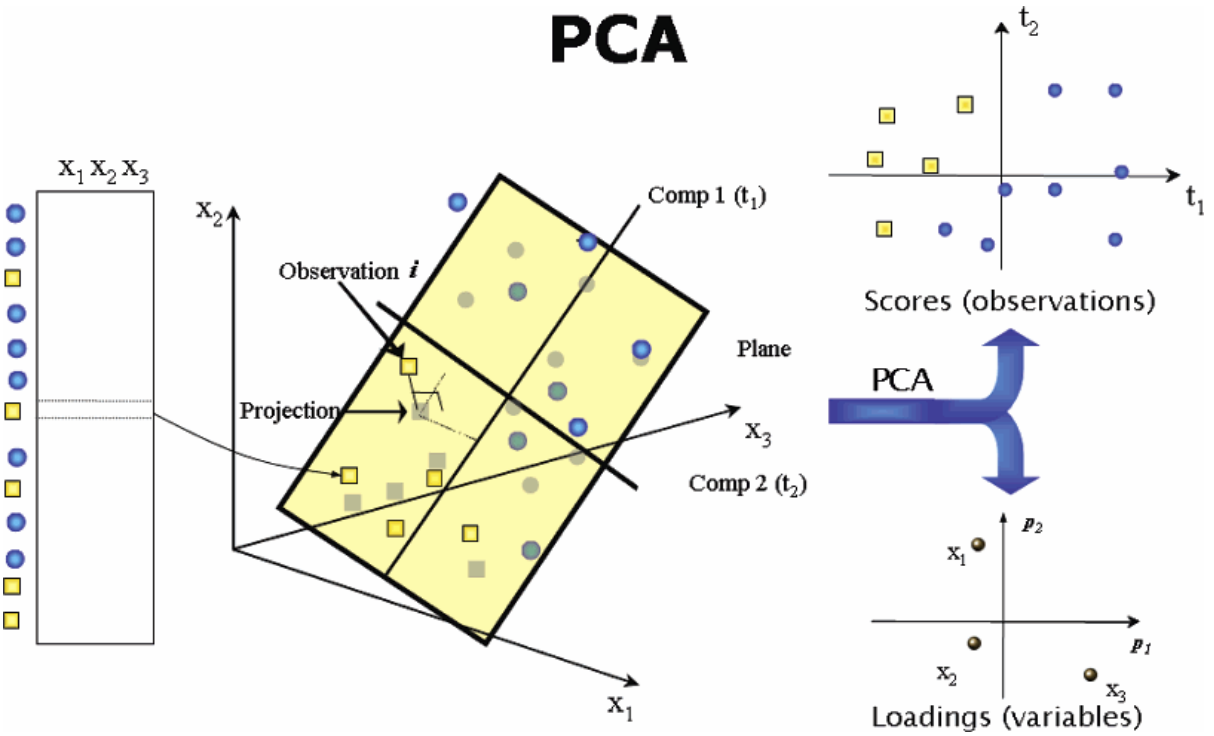


*Figure 2. 34: A schematic representation of the basic steps of PCA adapted from Trygg et al [131]. Three variables ($X_1$, $X_2$, $X_3$) for each of the 13 samples are placed on three dimensions. Each dimension represents one variable. By applying PComp analysis, the variation extracted is displayed in terms of scores plots and loadings plots. Scores plot will provide differences and similarities between the samples after projecting each sample on the vector of the PComp. Loadings plots will provide the contribution of each variable to the model for each of the PComps.[131]*

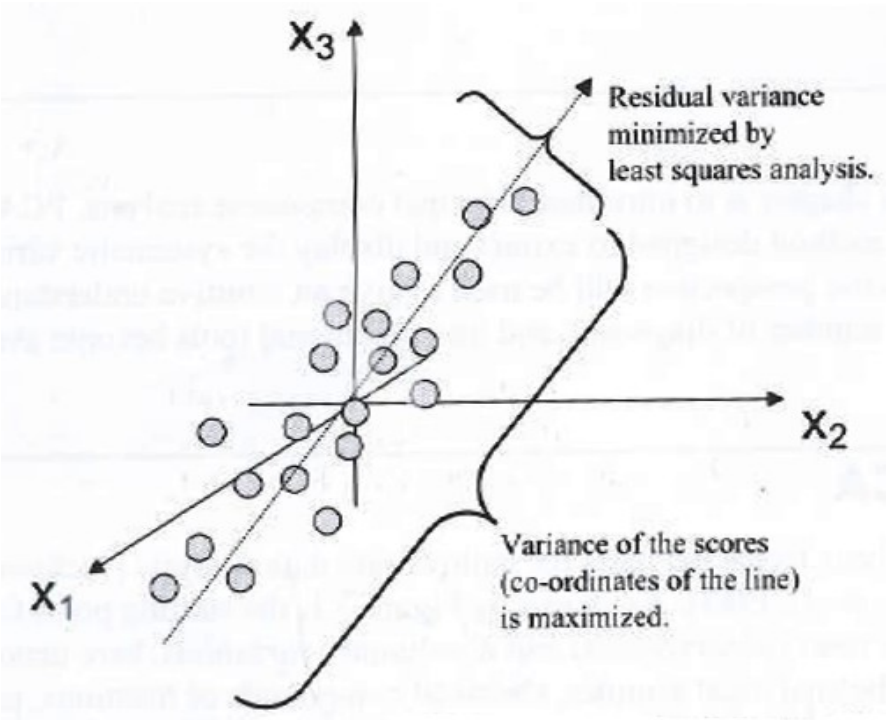


*Figure 2. 35: A representation of the first PComp as generated in a three dimensional system adapted from Eriksson et al[130]. The PComp will capture the maximum variance of the samples, or in other words a minimum of the least squares.[130]*

Results from PCA are comprehensively expressed using scores and loadings plots (Figure 2. 34) to aid data interpretation. Every PComp will provide its respective scores and loadings. The scores (**T**) are the projections of the observations on the respective PComp (Figure 2. 36). The value of this projection of a PComp constitutes the score value (**t**), and when plotted represents the scores plot (Figure 2. 34). From scores plots one can, according to the variation explained by each PComp, visualise the extent of similarity or difference between samples and detect outliers or anomalies in the sample set. On the other hand, loadings (**P**) will provide the magnitude, the contribution, of each variable (in the case of spectral data the variable represents a signal and hence a metabolite) to the PComp. This is generally represented as **p**. Geometrically, they express the orientation of the corresponding PComp in the multi-dimensional system. This is given by the angles from the PComp to each variable, keeping in mind that each variable is represented by a dimension (Figure 2. 37)[130, 131].

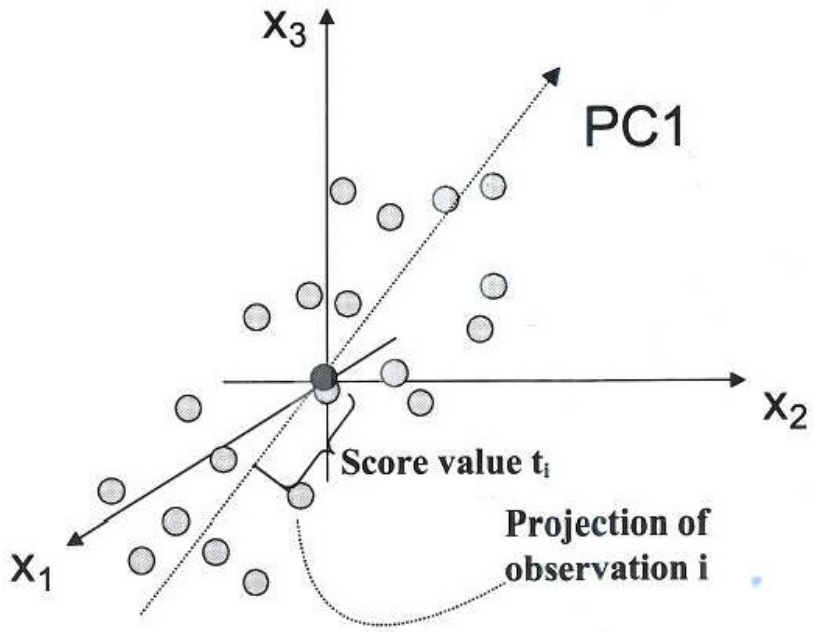


*Figure 2. 36: Scores for each principal component are calculated from the projection of each point on the corresponding principal component. Adapted from Eriksson et al [130].*

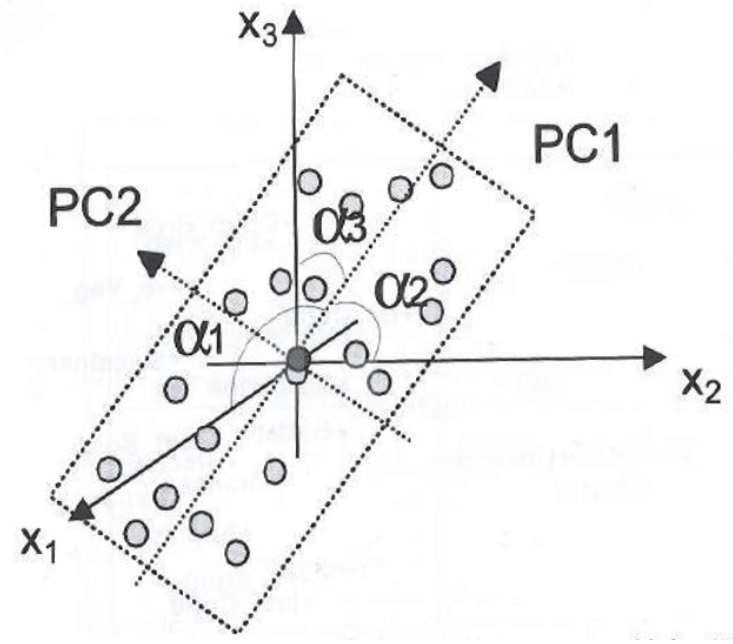


*Figure 2. 37: Loadings for each principal component are calculated from the cosine of the angle of the corresponding principal component, to the axes of each variable. Adapted from Eriksson et al[130].*

Algebraically, the matrix X can be modelled and described by the equation:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E}$$

Where **X** is the data, **T** is the scores, **P´** represents the loadings (transposed), and **E** represents the residuals matrix**.**

PCA displays the greatest variation of the model's variables. Thus for unscaled data a variable with a greater value range will dominate the variation and consequently the PCA model. In order to adjust the importance of each of the variables, one can apply scaling to the data. A scaling method that can attribute the same importance to all the variables is unit variance scaling (Figure 2. 38). For unit-variance scaling each value is divided by the standard deviation of the respective variable. Therefore, no variable can dominate over any other because of a larger range of its values. However, with unit variance scaling noise is given the same importance, and this could affect the interpretability of the model.

A scaling method that can minimise the differences in range of values, but at the same time keep the order of the initial variance of each variable, is Pareto scaling (Figure 2. 38). Pareto scaling is conducted by dividing each value by the square root of the standard deviation. This way, placing this in the context of metabolic profiling analysis (e.g. when using UPLC-MS), data with high intensities are providing more impact to the model than low intensity and noise.

Along with scaling procedures, regularly mean-centering is applied. As the name states, mean-centering is a pre-processing method where the mean of each variable is centred, i.e. becoming the origin of axes. This is simply done by calculating the mean of each variable and subtracting it from each of the variable's values [130, 132]. This procedure makes comparison between variables easier. The described preprocessing methods are shown schematically in Figure 2. 38.

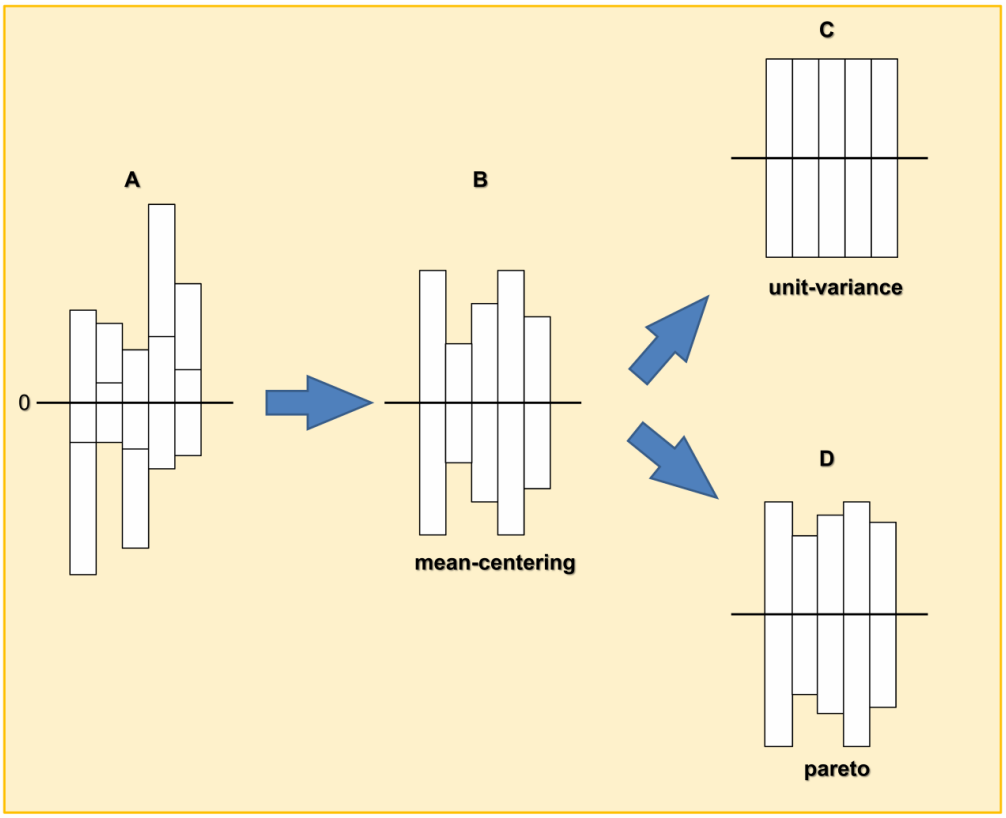


*Figure 2. 38: A schematic representation of a dataset (A) prior to processing, (B) after mean-centering only, (C) after unit-variance scaling and mean-centering, and (D) after Pareto scaling and mean-centering.*

PCA is a robust MVDA method to acquire information from a dataset in an unsupervised manner. However, the information given will only represent the greatest variation in the dataset. Although sometimes the variation explained by the PCA model can explain the variance induced by, for example, a disease, or treatment, this is not always the case. When the main source of variation is not related to disease or biological class, for example sources of physiological variation such as gender or age may dominate a PCA model, supervised methods are employed to reveal this hidden response-related variance, and provide the hidden (latent) variables that could function as biomarkers.

### *2.5.3 Partial Least Squares (PLS)*

PLS is a supervised MVDA method. PLS acronym is interpreted as Partial Least Squares; however it can also be interpreted as Projection to Latent Structures. The PLS is a regression technique used in order to model the association between two data matrices, X and Y. It is generally used in metabolic profiling to obtain a relationship between the dataset matrix X, and the response matrix Y.

When Y is a vector (matrix of one dimension), then the first PComp is going to be the direction, in the X-space, that, at the same time, describes the maximum variation for X (similarly to PCA), and best correlates to the Y-vector. Similarly to PCA, scores are obtained by projecting the position of a sample on this component. If the first PComp is inadequate to describe the variation of the Y vector, then a second PComp is employed. This, again similarly to PCA, will be the direction in the X-space that describes the greatest (remaining) variation, passes through the origin, and is orthogonal to the first PComp. However, it will again try to correlate the PComp with the remaining variation of the Y vector (Figure 2. 39).

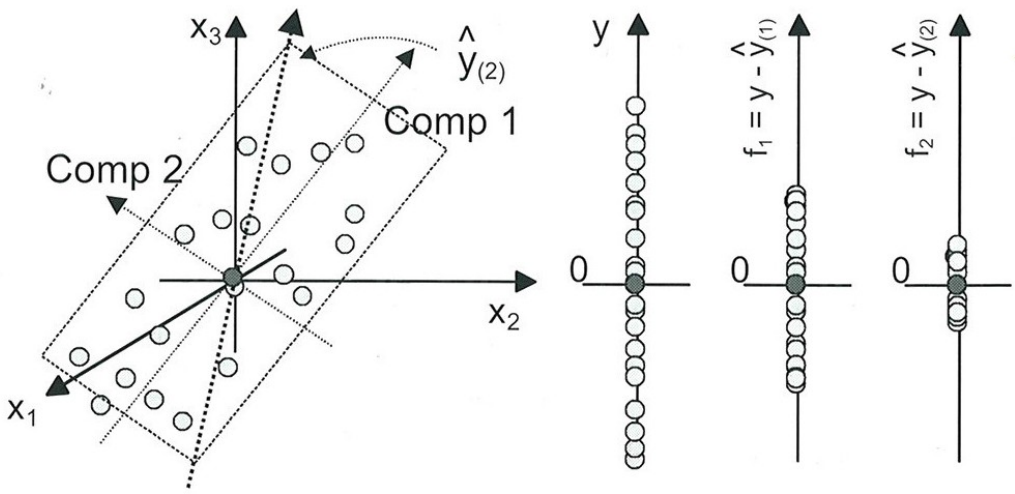


*Figure 2. 39: A schematic representation of the calculation of the first two PComp of a PLS model. Y is a vector and after addition of each PComp the extent of the explained response is shown. Adapted from Eriksson et al [130].*

PLS functions in the same way whether Y is a continuous response matrix (for example level of ALT corresponding to extent of liver damage), or whether it describes a discrete class (discriminant analysis). When Y is a matrix then what is correlated between X and Y are the principal components describing the greatest variation.

### *2.5.4 Orthogonal-PLS (O-PLS)*

Orthogonal-PLS (O-PLS) is a variant of PLS designed to filter extraneous variation, not relevant to the biological class and to enhance interpretation of the variable contributions in defining a class. In this case the model is built initially with one PComp as described for PLS. This PComp is called predictive and is correlated as much as possible to the Y matrix. However, O-PLS goes a step forward and adds an orthogonal PComp to the model. This orthogonal component should be *uncorrelated* to Y as much as possible. Therefore, it is actually describing the unwanted systematic variation, not related to the Y matrix, which induce noise in the model. If a single orthogonal

component cannot explain all the noise in the dataset, more orthogonal components can be employed.

## 2.5.5 Model Diagnostics and Validation

### 2.5.5.1 $R^2$ and $Q^2$

The $R^2$ represents the variation explained by the model, and is also referred to as the 'goodness of fit'[130]. As it can be inferred by the terminology it represents the ability of the model to fully represent the dataset. This implies that for a high $R^2$ value, the residual values are low. On the other hand, $Q^2$ is the predictive ability of the model, also referred to as the 'goodness of prediction'[130]. It represents the ability of the model to predict an outcome, using a training set, a new set (test set) with data not previously used for building the model. Generally this is done by leaving outside the model a small portion of the data, and repeating this procedure for as many cycles needed to have each observation back-predicted once. For PCA $R^2$ and $Q^2$ refer to the X-data, while for PLS or O-PLS, they refer to the Y-data.

As expected, most of the time $R^2$ will be higher than $Q^2$ as model complexity increases, e.g. when the number of components increases. $R^2$ will inevitably keep increasing, along with complexity, as more variation will be explained by the model (Figure 2. 40). However, $Q^2$ will reach a peak whereby the number of PComps it takes to accurately predict a sample to a given class is optimal, and then starts declining. This is due to the fact that when more variation in the form of new PComps enter the model, more noisy uncorrelated variables are also added, which can actually lead to a reduction of class prediction (Figure 2. 40). This is why more PComps should not be added to a model when the $Q^2$ value starts declining. Lastly, while high $Q^2$ is always required, the $R^2$ should not exceed $Q^2$ by more than 0.2-0.3. A disproportionally high $R^2$ would actually mean that the model is overfitted. This would be translated to an incorporation of noise variants in the model, making it unable to explain/predict test sets from another experiment/dataset and giving false confidence in the model.

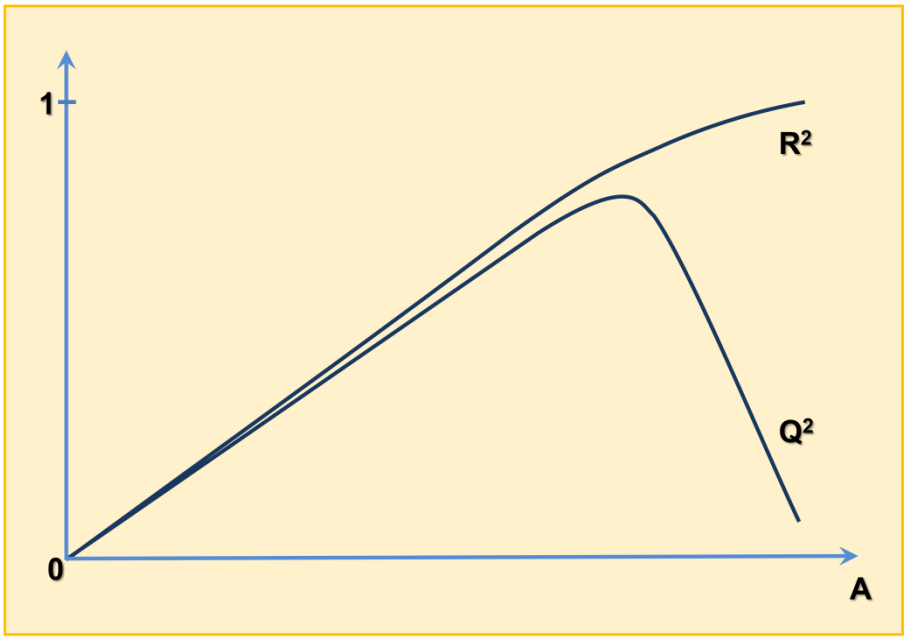


*Figure 2. 40: Plot describing the effect of increased complexity (A; x-axis) towards $R^2$ and $Q^2$.*[130]

### 2.5.5.2 Permutation testing and CV-ANOVA

Permutation testing is a way of cross-validation of the model. It is conducted by shuffling the class values of the Y-matrix, and essentially deliberately mislabelling the samples. This means that when fitting any randomly created dataset, this randomly labelled model should have a lower predictive value $Q^2$ and lower explained variation $R^2$ than the model of the actual dataset. If the values obtained from the series of random dataset are plotted against the correlation of the random dataset to the real dataset, then a plot as in Figure 2. 41 would be obtained. Intercepts of $Q^2$ and $R^2$ to Y-axes need to be as close to point of origin as possible. It is suggested that $Q^2$ should intercept the y-axis lower than 0.05 and $R^2$ lower than 0.3-0.4[130]. Nonetheless, permutation testing can be used for diagnosis of non-valid models. Additionally, it can assist in decisions concerning the number of components to be used.

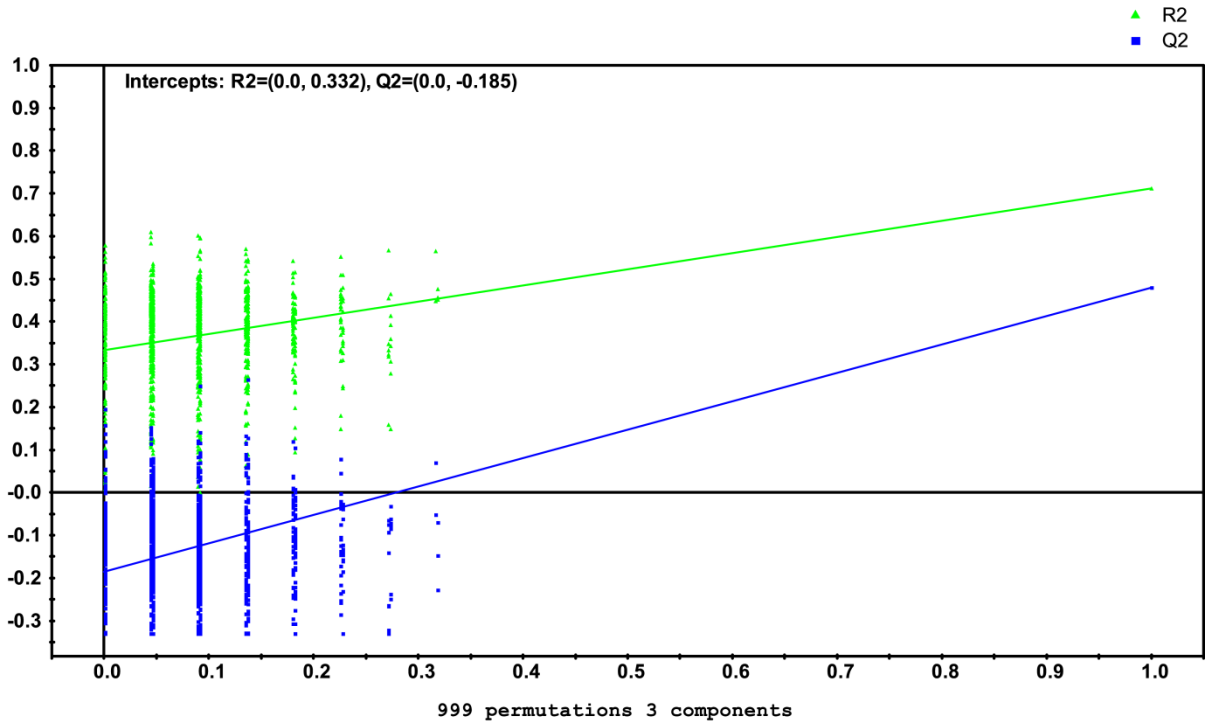


*Figure 2. 41: Permutation plot showing the $R^2$ and $Q^2$ of randomly 'mislabelled' samples as a function of their correlation to the actual data. Optimally, all $R^2$ and $Q^2$ values should be lower than the real response, while intercept should be lower than 0.4 for $R^2$, and 0.05 for $Q^2$.*

Cross validation – analysis of variance (CV-ANOVA)[133] is a significance testing of PLS and O-PLS models with one-dimension Y-vectors. To calculate the significance of the model, ANOVA testing is performed comparing the residuals obtained after fitting 'cross-validated' models, to the residuals of the complete dataset model[133]. Cross-validation is described as the procedure of obtaining the $Q^2$, where, as previously explained, part of the data are left out of the fitted model, and then back-predicted to this model.

### *2.5.6 S-plot*

The S-plot is used to assist when candidate biomarkers need to be obtained rather than a multivariate model. The model covariance is plotted on the x-axis against the model correlation on the y-axis. Therefore, order of magnitude and order of reliability can be visualised in a plot. This is very helpful in the field of metabolic profiling. Instrumentation can provide concentration detection of up to six orders of magnitude which will be fully exploited by a biological sample, where concentrations can range up to twelve orders of magnitude. As a result features with lower intensities can be missed, although they could act as good disease predictors (Figure 2. 42). On the other hand noise features, with very low concentrations should also be avoided as they could introduce false positives[134]. The optimal biomarker would be the feature that has both high covariance and high correlation, indicating less chance of a noise derived feature and with high

confidence. Nonetheless, the S-plot has become a very informative tool in statistical analysis and when determination of highly significant features to serve as candidate biomarkers is needed.

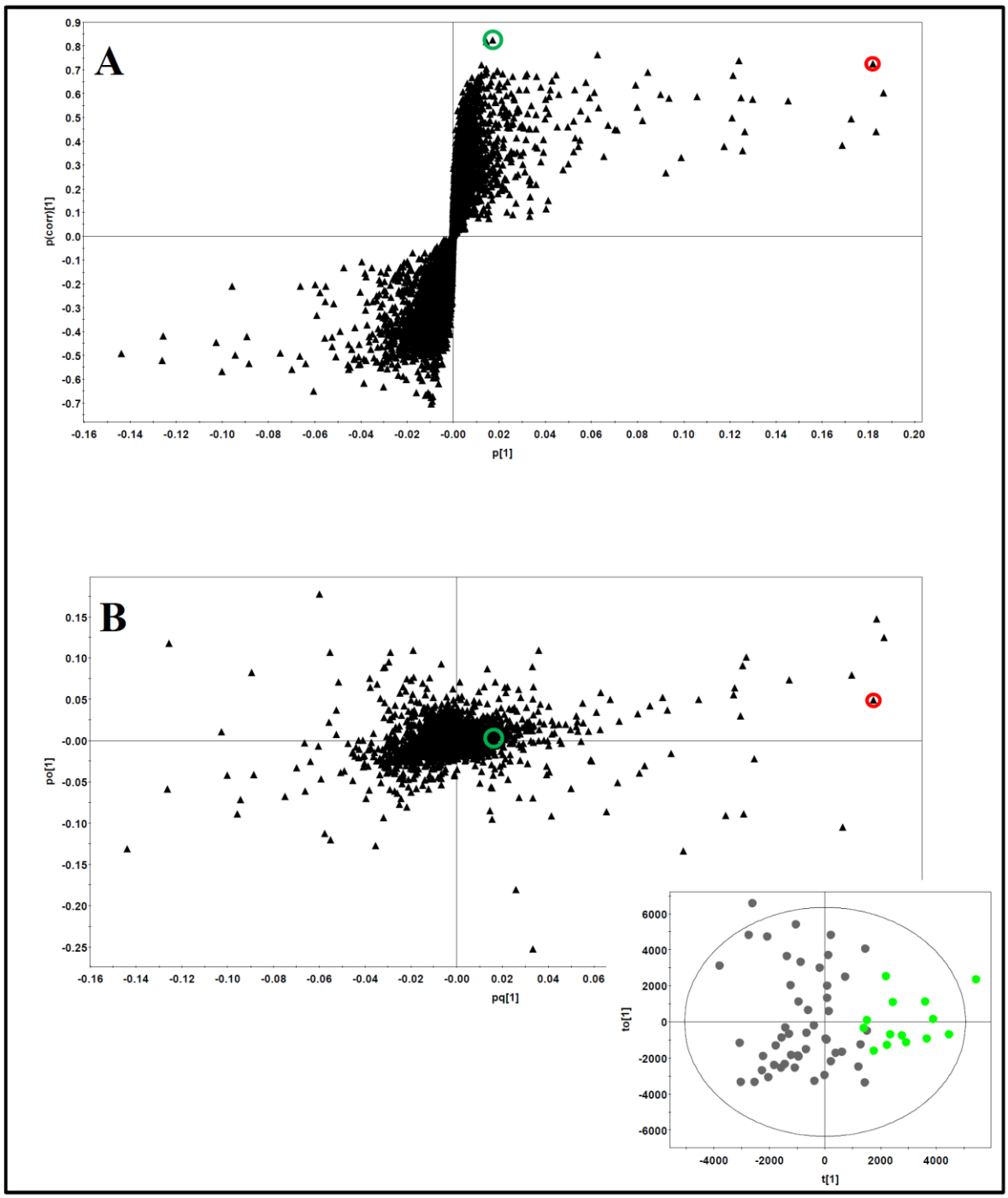


*Figure 2. 42: An example of how S-plot can assist in OPLS-DA analysis for the determination of candidate biomarkers. (A) S-plot where two features are highlighted having a high model correlation. (B) Loadings plot of the same OPLS-DA model where one feature can easily be located driving one of the groups while the other is hidden in noisy variables (Insert: Scores-plot of OPLS-DA model).*

# Chapter 3 Exploratory metabolic profiling methodologies for stratification of high risk for rupture carotid plaques

## 3.1 Introduction

Stroke is one of the major causes of death worldwide[45], and the leading cause of disability in UK adults[135]. 15,000 strokes per year in the UK are caused by thromboembolism originating at a carotid plaque. This thrombus is caused due to the coagulation reaction of platelets, after the rupture of the fibrous cap of the atheromatous plaque[54]. It can become detached and travel along with blood flow, eventually causing blockage to smaller luminal calibre vessels (thromboembolism). This will reduce the oxygen supply to parts of the brain affected with the possibility of causing disabilities and death.

Three-quarters of stroke patients will have been previously asymptomatic[136]. It is therefore of great importance to stratify the patient risk in order to monitor, predict and prevent the rupture of the plaque. In order to accomplish this, the experimental design will be based on the fact that patients with stroke symptoms are at high risk of an imminent life-threatening stroke. This risk declines with time after the symptoms which are nothing other than what can be characterised as a mini-stroke. On the contrary, asymptomatic patients are less likely to suffer plaque rupture and stroke[137].

This is a pilot study with the objective to develop and identify methods that can provide the most of information from carotid plaque samples. The capability of metabolic profiling to risk stratify patients is evaluated, in order to assess whether larger scale studies are worth pursuing. The study will be based on the comparison of recently symptomatic patients (high risk group; imminent life threatening stroke), and asymptomatic patients as the control group which are less likely to have a stroke.

## 3.2 Methods

### 3.2.1 Patient characteristics

The carotid plaque samples used in this study were obtained from 10 patients from the Department of Vascular Surgery, Imperial College London. From these patients, 5 were asymptomatic

(Asympt), and 5 were recently symptomatic (Sympt) of cerebrovascular symptoms occurring in the territory of the ipsilateral carotid circulation. At carotid endarterectomy surgery, diseased intimal arterial segments were retrieved, snap frozen in liquid nitrogen and stored at -80°C. Patients provided informed consent and the study was approved by the ethics committee. Patients clinical characteristics can be found in Table 3. 1.

*Table 3. 1: Patients demographics from symptomatic and asymptomatic groups for stroke.*

| | Age | Male | BMI | DM | HTN | PAD | Renal failure | ASA | A2RA/ ACEI | Statins | Smoker/Ex |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Symptomatic group | | | | | | | | | | | |
| | 67 | Yes | ~ | No | Yes | No | No | Yes | Yes | Yes | Yes |
| | 78 | Yes | 34.3 | No | Yes | No | No | Yes | Yes | Yes | No |
| | 69 | Yes | 30.5 | No | Yes | No | No | Yes | Yes | Yes | Ex |
| | 91 | Yes | 26.6 | No | No | No | No | No | No | No | No |
| | 74 | Yes | 32.3 | Yes | Yes | Yes | No | Yes | Yes | Yes | Ex |
| Average / % | 75.8 | 100% | | 20% | 80% | 20% | 0% | 80% | 80% | 80% | 20%/40% |
| Asymptomatic group | | | | | | | | | | | |
| | 72 | Yes | ~ | No | Yes | No | No | Yes | Yes | Yes | No |
| | 78 | Yes | ~ | No | Yes | Yes | No | Yes | No | Yes | Ex |
| | 59 | Yes | ~ | No | Yes | No | No | No | No | Yes | Yes |
| | 78 | No | 29.3 | No | Yes | No | No | Yes | Yes | Yes | No |
| | 75 | Yes | ~ | No | Yes | Yes | Yes | Yes | Yes | Yes | Ex |
| Average / % | 72.4 | 80% | | 0% | 100% | 40% | 20% | 80% | 60% | 100% | 20%/40% |

Key: BMI: Body mass index; DM: Diabetes Mellitus; HTN: Hypertension; PAD: Peripheral artery disease; ASA: Acetyl salicylic acid; A2RA/ACEI: Angiontensin receptor antagonist/Angiotensin-converting enzyme inhibitors

### *3.2.2 Sample Preparation*

Three segments of arterial tissue were obtained from each patient (Figure 3. 1). The central slice was obtained and stored for future MS-imaging analysis. The other two slices, weighing 120mg±5% each, were placed into separate bead beating tubes (VWR, USA), which were preloaded with 1mm zirconium beads (Percellys 24, Germany). This was followed by a modified extraction protocol from *Masson P et al*[17].

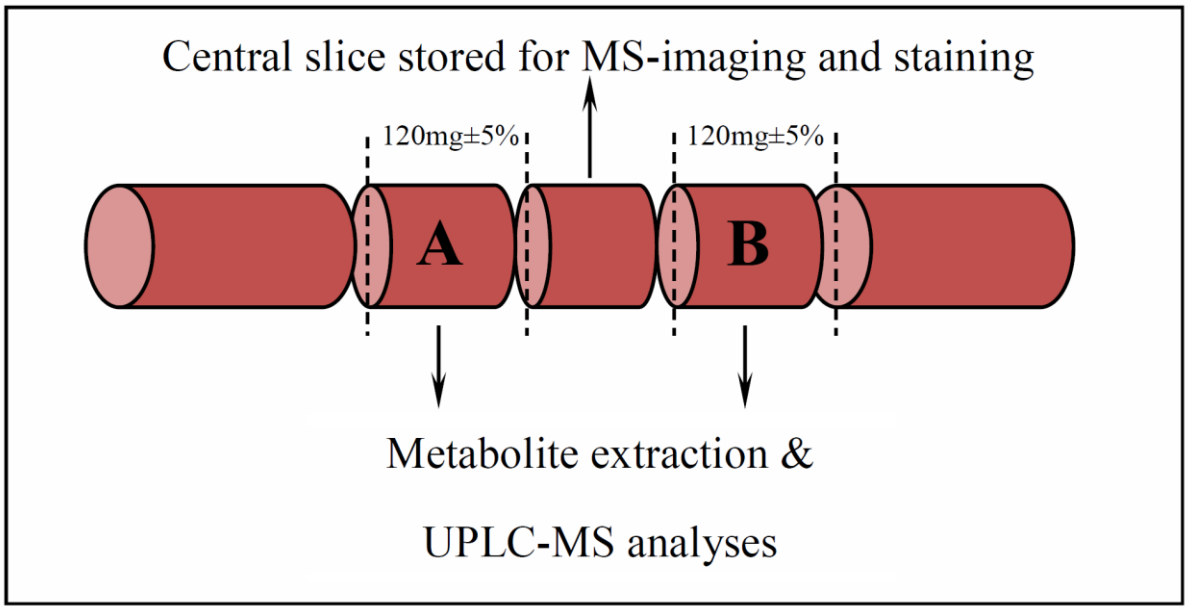


*Figure 3. 1. A schematic representation of the arterial tissue segmentation procedure applied for analysis.*

To obtain the aqueous extract, 1.5mL of chilled methanol (MeOH)/water (1:1) solution (methanol: HPLC gradient grade, Fisher; water: LC-MS grade, Fluka) were added to each tube after randomisation of the samples. Randomisation was conducted using free online software (http://www.random.org/sequences/). Samples were frozen on dry ice and loaded onto a bead beater (Bertin Technologies, Precellys 24), for 2 + 2 cycles, separated by freezing on dry ice. For each cycle, the beater was vibrating for 40 seconds at 6500Hz. Centrifugation (Biofuge Pico, Heraeus) was conducted at 13,000 rcf for 10 minutes, followed by aliquoting of supernatant into Eppendorf tubes (300µL x4). An aliquot of 70µL of each aqueous sample was combined to generate a pooled sample, which was mixed and split into four aliquots of 300µL each. This pooled sample is used for pre-run conditioning of the column and to assess instrumental performance. Additionally, the pooled sample can provide a good idea of the reconstitution volume prior to analyses, as this is the first time such tissue type has been assessed by the current instruments in our department. Samples

were then spun on a vacuum concentrator for 160 minutes (Eppendorf Concentrator Plus, 45°C, V-AQ mode). Extracts were stored at -40°C until analysis.

To obtain the organic extract, chilled 1.5mL solution of dichloromethane (DCM)/MeOH solution in proportions of 3:1 (DCM: HPLC grade, Sigma-Aldrich) were added to the precipitate of the sample from the aqueous extraction. They were then frozen and reloaded onto the bead beater for 2 cycles, vibrating at 6500Hz for 40 seconds per cycle. This was followed by centrifugation at 13,000 rcf for 20 minutes. Extracts were then aliquoted into glass tubes (200µL x4). An aliquot of 50µL of each sample was combined to generate a pool sample, which was mixed and split into four aliquots of 200µL. Samples were allowed to evaporate overnight at room temperature in a fume hood. Extracts were stored at -40°C until analysis.

### *3.2.3 RP-LC-MS lipid analysis*

Tissue organic phase extracts were reconstituted in 250µL of MeOH (LC-MS grade, Fluka). Samples were vortexed for 30s, sonicated for 5min and vortexed again for 30s. This was followed by a 5min centrifugation (Eppendorf, Centrifuge 5417R, Germany) at 13,000rpm, 4°C. Samples were then transferred into LC-MS glass vials (Total Recovery Vials, Waters Corp, USA).

Samples were re-randomised prior to loading on the autosampler of the UPLC system, and preserved at a temperature of 4°C. An aliquot of 50µL from each sample was combined to form a quality control sample (QC), and a QC-format run was conducted[114]. The purpose of the QC sample is to provide confidence of the analytical stability of the run for both UPLC and MS. It is later used during univariate and MVDA to assess reproducibility.

First the column was conditioned by injecting the pooled sample, obtained during the extraction, several times, until data showed adequate stability. The QC sample was then injected thrice, followed by further injections every 4-5 samples. The run was completed by samples of dilutions of the QC sample at 1:2, 1:4 and 1:8, as well as extraction and solvent blank samples. Dilutions are used to assess how responsive our methodology is to fold changes. They are also used at a later stage in the analysis pipeline, during statistical analysis, to remove features where their intensities do not respond to dilutions. Similarly, features that were present in high intensity in blanks were removed from all samples.

Chromatographic separation was conducted using an Acquity UPLC system (Waters Ltd, UK) and a HSS T3 C18 column (Waters Corp, USA), with dimensions of 2.1mm x 100mm, and particle size of 1.8µm. The protocol followed was described by *Shockcor et al* in an application note from

Waters Corporation[138]. The UPLC column was maintained at 65°C with a flow rate of 0.5ml/min. Mobile phase A was constituted of ACN/water (40:60) and 10mM of Ammonium Acetate (AmAc). The mobile phase B constituted of Isopropanol/ACN (90:10) and 10mM AmAc. The UPLC gradient program of the system is presented on Table 3. 2. An aliquot of 10µL of the reconstituted sample was injected with a full loop setting with needle overfill. Samples were preserved at 4°C throughout the run.

*Table 3. 2: Gradient program of the chromatography of the lipid profiling UPLC-MS methodology.*

| Time (min) | %A | %B | Curve |
|---|---|---|---|
| 0.0 | 60 | 40 | - |
| 0.5 | 60 | 40 | 6 |
| 10.50 | 0 | 100 | 6 |
| 12.5 | 0 | 100 | 6 |
| 12.6 | 60 | 40 | 6 |
| 20.0 | 60 | 40 | 6 |

A Q-TOF Premier mass analyser (Waters MS Technologies Ltd., UK) was employed and coupled with the UPLC system using an ESI interface. Data were acquired in both positive and negative modes. The Mass Spectrometer was set at the following parameters: Capillary Voltage +3kV(positive mode) and -2.2kV(negative mode), source temperature 120$^{o}$C, sampling cone 30V, desolvation gas temperature 400$^{o}$C, desolvation gas flow 900L/h, collision energy 5eV, scan speed every 0.3s and mass range 50-1000m/z. Leucine Enkephalin (Sigma, HPLC grade) was used for lock-mass correction. Lock mass scans were acquired every 30s. For $MS^{E}$ experiments, same parameters were used and with collision energy high voltage of 20eV.

### *3.2.4 HILIC-MS analysis*

Tissue aqueous phase extracts (in random order) were reconstituted in 200µL of 1:1 solution of Acetonitrile (ACN)/water (ROMIL). Samples were vortexed for 30s, sonicated for 5min and vortexed again for 30s. This was followed by a 5min centrifugation at 13,000rcf; 4$^{O}$C. Samples were then transferred into appropriate LC-MS glass vials (Total Recovery Vials, Waters Corp, USA). An aliquot of 50µL from each sample was combined to form a quality control sample (QC). Run setting was as described on the preceding paragraph.

The UPLC setting was assembled using an Acquity UPLC system and a BEH HILIC column, 1.7 $\mu$m, 2.1x100mm (Waters Corp, USA). Mobile phase A constituted of ACN/water 95:5, 10mM of

AmAc and 0.1% of formic acid (FA). The UPLC gradient program of the system is presented on Table 3. 3. An aliquot of 5μL of the reconstituted sample was injected using a 10μL loop, in a partial loop setting, with 4μL flashing.

*Table 3. 3: Gradient program of the chromatography of the HILIC-UPLC-MS methodology.*

| Time (min) | %A | %B | Curve |
|---|---|---|---|
| 0.0 | 99 | 1 | - |
| 2.0 | 99 | 1 | - |
| 14.0 | 45 | 55 | 6 |
| 14.1 | 1 | 99 | 6 |
| 18.0 | 1 | 99 | - |
| 18.1 | 99 | 1 | 6 |
| 25.0 | 99 | 1 | - |

An LCT Premier (Waters MS Technologies, Ltd., Manchester, U.K.) mass analyser was employed and coupled with the UPLC system using an ESI interface. Data were acquired in both positive and negative polarities. The Mass Spectrometer was set at the following parameters: Capillary Voltage +3200V(positive mode) and -2400V(negative mode), source temperature 120$^{o}$C, sampling cone 35V, desolvation gas temperature 350$^{o}$C, desolvation gas flow 800L/h, scan speed every 0.3s and mass range 50-1000m/z. Leucine Enkephalin was used for lock mass correction, with scans acquired every 30s.

For MS/MS experiments that were conducted for acyl-carnitines the same UPLC parameters were used. For MS analysis the Q-TOF Premier mass analyser was used with parameters set as with the previous section. The collision energy was ramping between 20 – 40 eV.

### *3.2.5 Data processing and statistical analysis*

Collected data from both analyses and both modes were analysed using the XCMS package version 1.34.0 [112] in R programming software version 2.15.2. The data processing workflow consisted of peak-picking, grouping, retention time (RT) correction, re-grouping (after RT correction), setting of zero values to background intensity, and normalisation to total intensity. Command lines used are summarised in Appendix 2. The resulting normalised intensity was multiplied by $10^9$. This multiplication was done in order to avoid the loss of significant digits, since the MVDA software (described in the next paragraph) does not handle more than 4 decimal places.

Data were imported in SIMCA-P+ 12.0.1. The QC sample dilutions were used to filter out of the resulting models features that were not responding to these dilutions. Similarly, the extraction blanks were used to filter features resulting from the extraction procedure or solvent impurities.

Principal component analysis (PCA) was used as an unsupervised multivariate data analysis (MVDA) method to visualise data and detect variation induced to the model by the studied groups (Symptomatic and Asymptomatic). Features driving the variance in the model were identified on loadings plots. All features were subjected to a 2-tailed t-test, assuming unequal variance. Fold change was calculated using the median of each group.

Metabolite identification was conducted firstly by matching mass measurements to theoretical value from on-line databases. Databases used for this study were LipidMaps[139], Metlin[111] and HMDB[116]. In order to further elucidate the structure of the statistically significant metabolites the isotopic pattern of the features, $MS^E$ spectra, in-house developed libraries of standards, and matching MS/MS spectra to MS/MS spectra from the Metlin database[111] were used.

# 3.3 Results

## 3.3.1 Assessment of analytical reproducibility

For both analytical methodologies conducted for this study, an assessment of reproducibility was carried out using the QC samples. In the PCA scores plots, very close grouping of QCs was observed indicating that analytical variability was acceptably low (Figure 3.2). In Figure 3.2 the QC samples are represented along with the studied samples. This is a good indication of the reproducibility of the method, instrumental stability, and an indication of good quality data. When univariate statistics were applied on metabolites, the reproducibility was also assessed by the coefficient of variation (CV% < 20; < 30 for low intensities[119]).

## 3.3.2 Characterisation of the metabolic profiling of the symptomatic atherosclerotic plaque

### 3.3.2.1 Lipid Profiling

Lipid profiling was conducted using a reversed phase (RP-) UPLC-MS based methodology, of the organic extracts of the atherosclerotic tissue. For the two groups studied, PCA scores plots showed good discrimination and grouping of the studied samples, and in both ESI polarity modes (Figure 3. 5). Representative chromatograms of the method used can be viewed in Figure 3. 3 and Figure 3. 4.

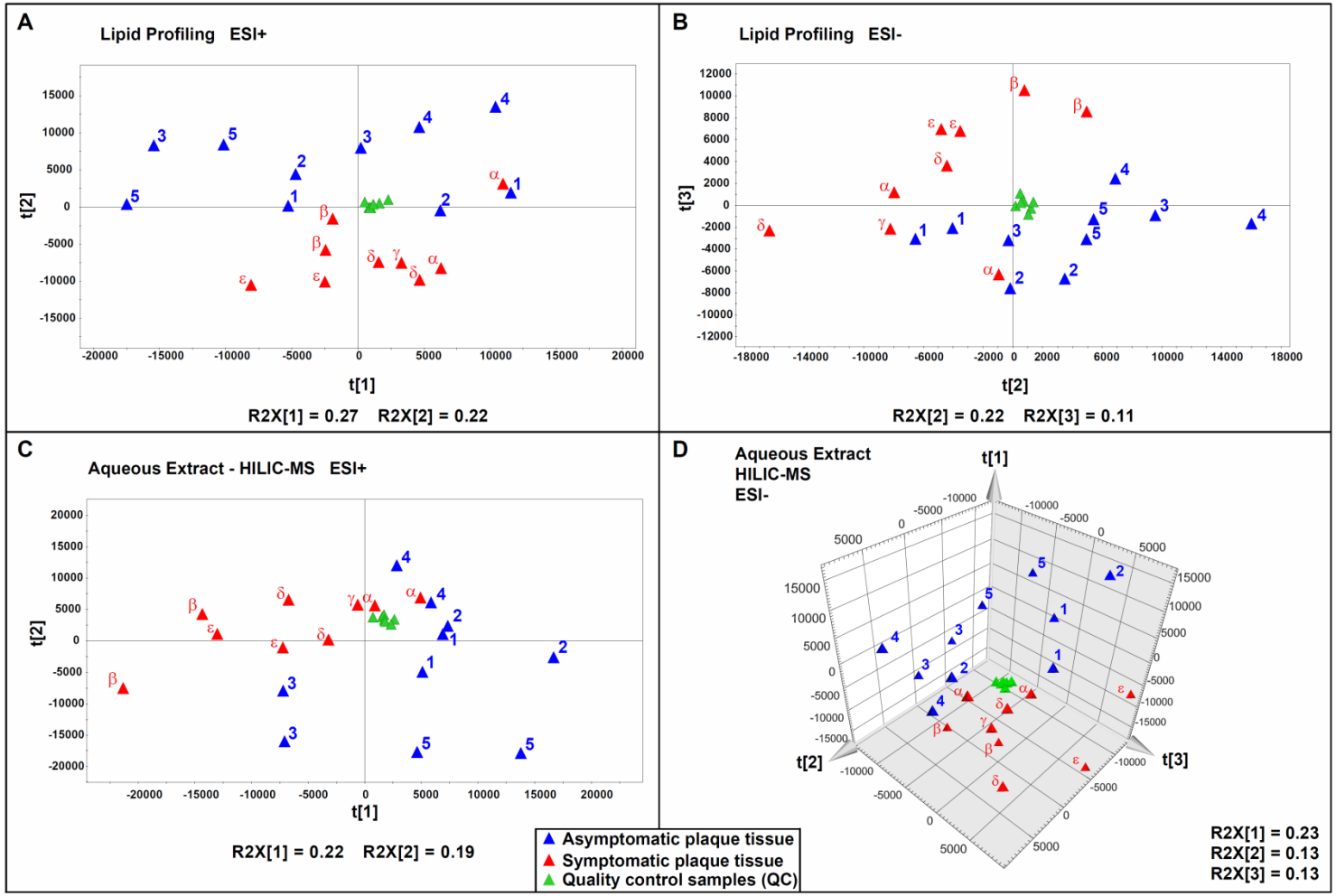


*Figure 3.2: Principal component analysis scores plots of atherosclerotic plaque tissue extracts. Lipid profiling of organic metabolite extracts in (A) positive ionization mode and (B) negative ionization mode. Hydrophilic interaction liquid chromatography coupled to mass spectrometry (HILIC-MS) profiling of aqueous extracts in (C) positive ionization mode and (D) negative ionization mode. The quality control (QC) samples are denoted in green and present a good indication of the reproducibility of the methodology and stability of the specific run. Samples obtained from the same plaque tissue are denoted by the same alphanumeric. Sample γ is represented by only one biological replicate. ESI: Electrospray Ionization.*

The positive mode features showed group discrimination of Sympt and Asympt samples in scores plots of the 1st and 2nd PComp (Figure 3. 5.A), with separation mainly in the second component. These two components together constitute the 50% of the variation ($R^2X$), with the 2nd contributing 23% to the variation. The negative mode data also presented good grouping with best results on the 2nd and 3rd PComp (Figure 3. 5.B), with cumulative $R^2X$ of 35%, and the 2nd PComp contributing 23% of the variance.

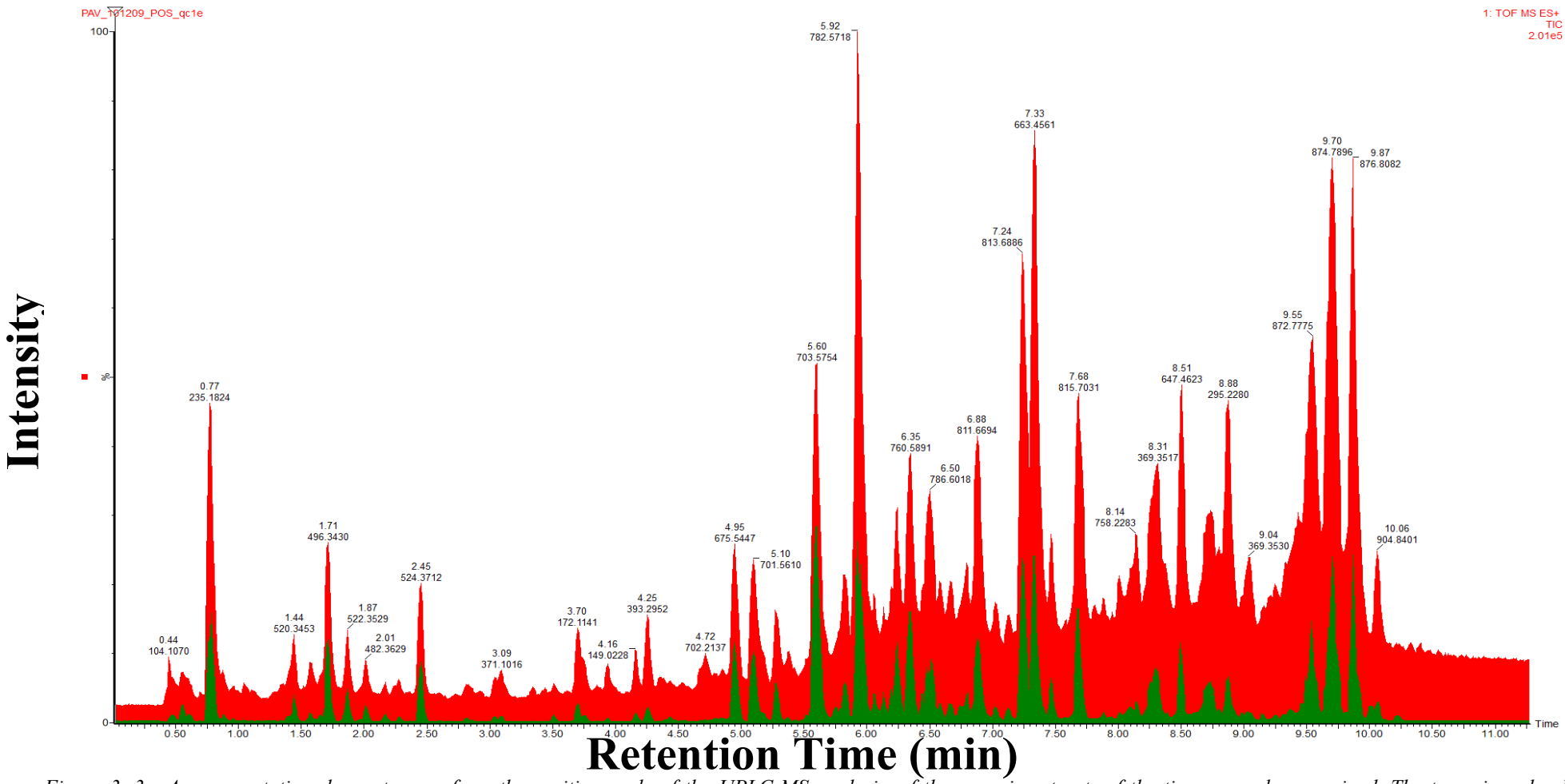


*Figure 3. 3: A representative chromatogram from the positive mode of the UPLC-MS analysis, of the organic extracts of the tissue samples examined. The trace in red colour represents the total ion current, while in green are the base peak intensities (highest peaks within a retention time window).*

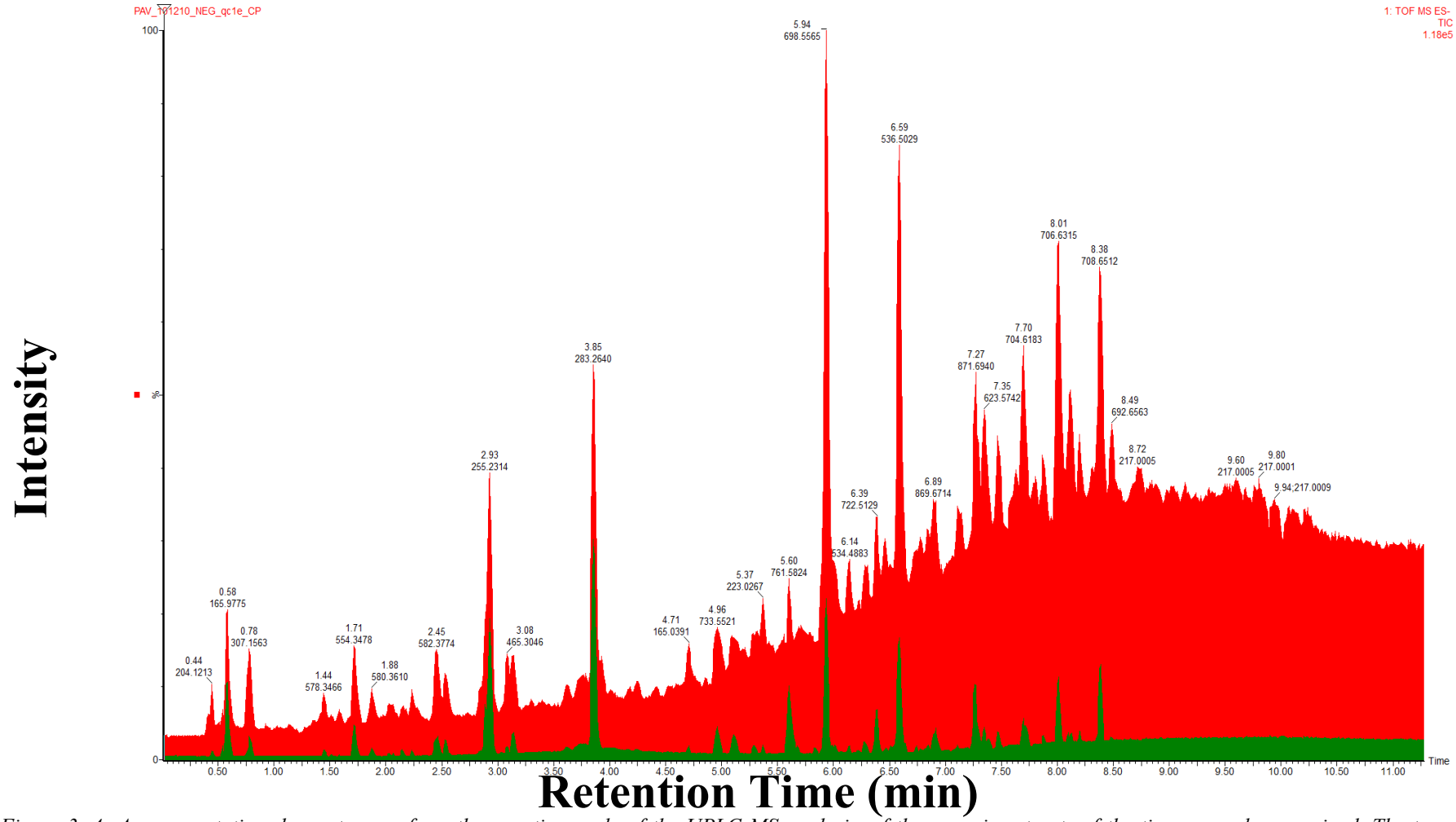


*Figure 3. 4: A representative chromatogram from the negative mode of the UPLC-MS analysis, of the organic extracts of the tissue samples examined. The trace in red colour represents the total ion current, while in green are the base peak intensities (highest peaks within a retention time window).*

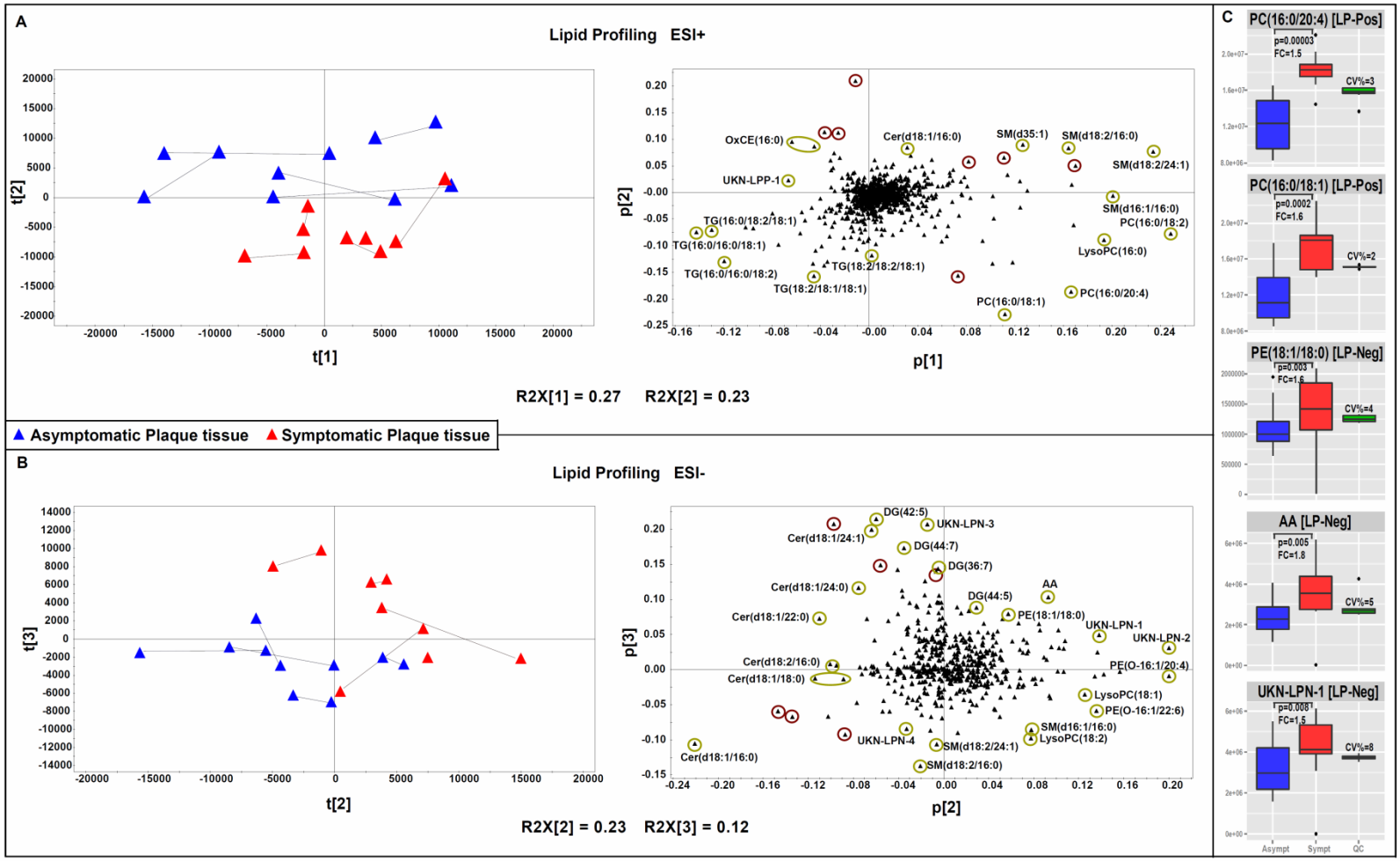


*Figure 3. 5: A and B represent scores and loadings plots from the PCA model of the results obtained from the lipid profiling analysis of the organic extracts (A positive and B negative mode). Points represent samples. Samples connected with a line are biological replicates. C) Box-plots of selected metabolites showing high statistical significance. Colour code: green: QC; blue: Asymptomatic; red: Symptomatic; dark yellow circles: metabolites assigned; dark red circles: features identified to represent isotopes, adducts, or fragments of identified metabolites.*

Using the loadings plots (Figure 3. 5), features that were driving the separation between Sympt and Asympt, were identified (Figure 3. 5 and Table 3. 4. These included a number of phosphatidylcholines (PCs), lysophosphatidylcholines (lysoPCs), phosphatidylethanolamine (PEs), diglycerides (DGs), triglycerides (TGs), sphingomyelins (SMs), ceramides (Cers), cholesterol-esters (CEs), oxidised Cholesterol-esters (oxCEs), and arachidonic acid (AA). Features that could not be assigned were marked as UKN with a suffix indicating the analysis and mode. Table 3. 4 presents the metabolites identified and unassigned features, along with UPLC-MS analysis feature characteristics, and results of univariate statistics.

*Table 3. 4: List of assigned metabolites found to be driving the PCA, obtained from data of the tissue organic extracts.*

| Met Name | LoA | Mol Formula [complex adduct] | RT (min) | m/z (found) | m/z (theor) | ΔDa | Δppm | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | |
| **Lipid Profiling Positive Mode** | | | | | | | | | | |
| Cer(d18:1/16:0) | 4 | C34H66NO2+ [M-H2O+H]+ | 6.85 | 520.5099 | 520.5094 | -0.0005 | -1 | 0.0784 | -1.5 | 2 |
| LysoPC(16:0) | 4 | C24H51NO7P+ | 1.77 | 496.3401 | 496.3403 | 0.0002 | 0 | 0.0735 | 1.5 | 4 |
| oxCE(16:0(OH)) | 4 | C43H76O3Na+ [M+Na]+ | 9.24 | 663.5738 | 663.5692 | -0.0046 | -7 | 0.6583 | 1.0 | 4 |
| PC(16:0/18:1) | 4 | C42H83NO8P+ | 6.61 | 760.5855 | 760.5856 | 0.0001 | 0 | 0.0002 | 1.6 | 2 |
| PC(16:0/18:2) | 4 | C42H81NO8P+ | 6.20 | 758.5718 | 758.5700 | -0.0018 | -2 | 0.2391 | 1.2 | 3 |
| PC(16:0/20:4) | 4 | C44H81NO8P+ | 6.16 | 782.5714 | 782.5700 | -0.0014 | -2 | 0.0000 | 1.5 | 3 |
| SM(d16:1/16:0) | 4 | C37H76N2O6P+ | 5.14 | 675.5452 | 675.5441 | -0.0011 | -2 | 0.2845 | 1.1 | 2 |
| SM(d18:2/16:0) | 4 | C39H78N2O6P+ | 5.30 | 701.5619 | 701.5598 | -0.0021 | -3 | 0.5010 | -1.2 | 2 |
| SM(d18:2/24:1) | 4 | C47H92N2O6P+ | 7.16 | 811.6715 | 811.6693 | -0.0022 | -3 | 0.7966 | -1.1 | 2 |
| SM(d35:1) | 4 | C40H82N2O6P+ | 6.14 | 717.5929 | 717.5911 | -0.0018 | -3 | 0.5623 | -1.3 | 5 |
| TG(16:0/16:0/18:1) | 4 | C53H104NO6+ [M+NH4]+ | 10.24 | 850.7866 | 850.7864 | -0.0002 | 0 | 0.5008 | 1.3 | 3 |
| TG(16:0/16:0/18:2) | 4 | C53H102NO6+ [M+NH4]+ | 10.07 | 848.7731 | 848.7707 | -0.0024 | -3 | 0.1056 | 1.4 | 2 |
| TG(16:0/18:2/18:1) * | 4 | C55H104NO6+ [M+NH4]+ | 10.10 | 875.7933 | 875.7897 | -0.0036 | -4 | 0.6636 | 1.1 | 3 |
| TG(18:2/18:1/18:1) | 4 | C57H106NO6+ [M+NH4]+ | 10.13 | 900.8044 | 900.8020 | -0.0024 | -3 | 0.0359 | 1.5 | 4 |
| TG(18:2/18:2/18:1) | 4 | C57H104NO6+ [M+NH4]+ | 9.96 | 898.7900 | 898.7864 | -0.0036 | -4 | 0.1328 | 1.4 | 5 |
| UKN-LPP-1 | n/a | n/a | 4.41 | 218.1037 | n/a | n/a | n/a | 0.0813 | -2.4 | 2 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula [complex adduct] | RT (min) | m/z (found) | m/z (theor) | ΔDa | Δppm | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | |
| Arachidonic Acid | 2 | C20H31O2- | 2.53 | 303.2311 | 303.2324 | 0.0013 | 4 | 0.00479 | 1.8 | 5 |
| Cer(d18:1/16:0) | 4 | C34H66NO3- | 6.85 | 536.5053 | 536.5043 | -0.0010 | -2 | 0.16158 | -1.2 | 4 |
| Cer(d18:1/18:0) | 4 | C38H74NO5-<br>[M+OAc-H]- | 7.39 | 624.5573 | 624.5567 | -0.0006 | -1 | 0.19005 | -1.2 | 2 |
| Cer(d18:1/22:0) | 4 | C42H82NO5-<br>[M+OAc-H]- | 8.33 | 680.6210 | 680.6193 | -0.0017 | -2 | 0.91956 | 1.1 | 1 |
| Cer(d18:1/24:0) * | 4 | C44H86NO5-<br>[M+OAc-H]- | 8.73 | 709.6559 | 709.6540 | -0.0019 | -3 | 0.66147 | 1.1 | 11 |
| Cer(d18:1/24:1) | 4 | C44H84NO5-<br>[M+OAc-H]- | 8.34 | 706.6355 | 706.6350 | -0.0005 | -1 | 0.27655 | 1.2 | 2 |
| Cer(d18:2/16:0) | 4 | C34H64NO3- | 6.38 | 534.4891 | 534.4886 | -0.0005 | -1 | 0.45521 | -1.1 | 1 |
| DG(36:7) | 1 | C39H61O5- | 5.49 | 609.4528 | 609.4519 | -0.0009 | -1 | 0.08409 | 3.3 | 3 |
| DG(42:5) | 1 | C47H81O7-<br>[M+OAc-H]- | 7.67 | 757.5996 | 757.5982 | -0.0014 | -2 | 0.27600 | 1.3 | 41 |
| DG(44:5) | 1 | C49H85O7-<br>[M+OAc-H]- | 8.05 | 785.6302 | 785.6295 | -0.0007 | -1 | 0.01075 | 4.5 | 4 |
| DG(44:7) | 1 | C49H81O7-<br>[M+OAc-H]- | 7.69 | 781.6006 | 781.5982 | -0.0024 | -3 | 0.09550 | 2.8 | 15 |
| lysoPC(18:1) | 4 | C28H55NO9P-<br>[M+OAc-H]- | 1.94 | 580.3619 | 580.3614 | -0.0005 | -1 | 0.03784 | 1.8 | 2 |
| lysoPC(18:2) | 4 | C28H53NO9P-<br>[M+OAc-H]- | 1.50 | 578.3465 | 578.3458 | -0.0007 | -1 | 0.36053 | 1.5 | 3 |
| PE(18:1/18:0) | 4 | C41H79NO8P- | 7.31 | 744.5552 | 744.5543 | -0.0009 | -1 | 0.00294 | 1.6 | 4 |
| PE(O-16:1/20:4) | 4 | C41H73NO7P- | 6.65 | 722.5131 | 722.5125 | -0.0006 | -1 | 0.01293 | 1.5 | 3 |
| PE(O-16:1/22:6) | 4 | C43H73NO7P- | 6.54 | 746.5135 | 746.5125 | -0.0010 | -1 | 0.25238 | 1.7 | 10 |
| SM(d16:1/16:0) | 4 | C39H78N2O8P-<br>[M+OAc-H]- | 5.15 | 733.5498 | 733.5496 | -0.0002 | 0 | 0.25122 | 1.1 | 2 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula [complex adduct] | RT (min) | m/z (found) | m/z (theor) | ΔDa | Δppm | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | |
| SM(d18:2/16:0) | 4 | C41H80N2O8P-[M+OAc-H]- | 5.31 | 759.5652 | 759.5652 | 0.0000 | 0 | 0.30897 | -1.3 | 6 |
| SM(d18:2/24:1) | 4 | C49H94N2O8P- | 7.18 | 869.6743 | 869.6748 | 0.0005 | 1 | 0.69214 | -1.1 | 4 |
| UKN-LPN-1 | n/a | n/a | 6.55 | 792.5764 | n/a | n/a | n/a | 0.00799 | 1.5 | 8 |
| UKN-LPN-2 | n/a | n/a | 8.46 | 690.6415 | n/a | n/a | n/a | 0.03247 | 2.2 | 3 |
| UKN-LPN-3 | n/a | n/a | 8.30 | 693.5559 | n/a | n/a | n/a | 0.09041 | 3.5 | 46 |
| UKN-LPN-4 | n/a | n/a | 8.00 | 690.6050 | n/a | n/a | n/a | 0.12380 | -1.4 | 3 |

*Statistics and assignment on second isotopic m/z

The positions of double bonds and fatty acyl chains cannot be determined in DGs, PCs, PEs, lysoPCs and TGs; fatty acyl chains are presented from lowest to highest molecular weight (lower number of carbons to higher, and higher number of double bonds, to lower).

Two-tailed t-tests were conducted, assuming unequal variance.

RT: retention time; oxCE: oxidised cholesterol ester, Cer: ceramide, DG: diglyceride; PC: phosphatidylcholine; PE: phosphatidylethanolamine; SM: sphingomyelin; TG: triglyceride; UKN: unknown; OAc: acetate; n/a: not applicable

LoA: level of assignment; 1: m/z matched to online databases; 2: m/z matched to online databases and prior knowledge for RT (in-house database of standards); 3: in-source fragmentation; 4: assignment based on $MS^E$ mode spectrum.

Some metabolites appeared to drive the primary separation of the two groups in the model. From these, some showed increased statistical significance when compared between the two groups Box-plots of the normalised intensities of these metabolites can be found in Figure 3. 5 C. These include PC(16:0/20:4), PC(16:0/18:1) and arachidonic acid (AA), which had a probability of p=0.00003, p=0.0002, and p=0.005, respectively indicating significant differences between the asymptomatic and symptomatic disease groups.

Irrespective of MVDA, univariate statistics, using the t-test and fold-change, were applied to all the features. A number of features showed high statistical significance and therefore were further structurally assigned (Table 3. 5). These include an acyl-carnitine (AcC) the palmitoylcarnitine, and a TG, TG(58:6), whose normalised intensities were higher in Sympt, with p=0.00001 and p=0.00007, respectively, and with a fold-change of 2.5 and 3.1. Figure 3. 6 demonstrates the box-plots of all the detected statistically significant metabolites of the lipid profiling analysis.

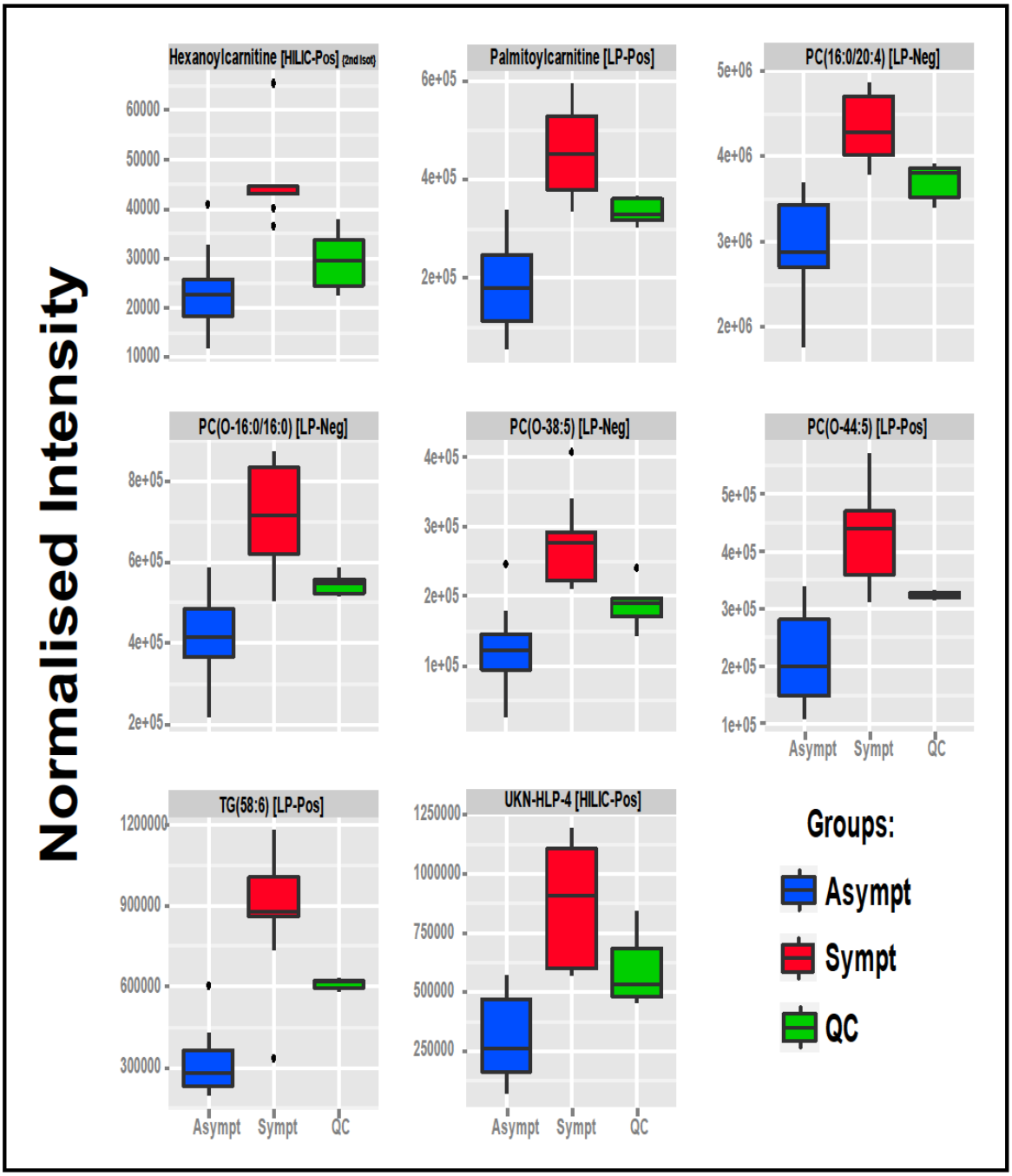


*Figure 3. 6: Box-plots of significant metabolites obtained from univariate statistics of data from both organic and aqueous extracts. Two-tailed t-tests were conducted, assuming unequal variance. PC: phosphatidylcholine; TG: triglyceride; UKN: unknown.*

*Table 3. 5: List of assigned significant metabolites obtained from univariate statistics of data from both organic and aqueous extracts.*

| Analysis - Mode | Met Name | LoA | Mol Formula [complex adduct] | RT (min) | m/z (found) | m/z (theor) | ΔDa | Δppm | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **HILIC Positive** | Hexanoylcarnitine* | 3 | C13H26NO4+ | 7.30 | 261.187 | 261.1894 | 0.0024 | 9 | 0.00003 | 1.9 | 20 |
| **Lipid Prof Positive** | Palmitoylcarnitine | 4 | C23H46NO4+ | 1.63 | 400.3426 | 400.3427 | 1E-04 | 0 | 0.00001 | 2.5 | 8 |
| **Lipid Prof Negative** | PC(16:0/20:4) | 4 | C46H83NO10P-[M+OAc-H]- | 6.19 | 840.5783 | 840.5755 | -0.0028 | -3 | 0.00002 | 1.5 | 6 |
| **Lipid Prof Negative** | PC(O-16:0/16:0)* | 4 | C42H85NO9P-[M+OAc-H]- | 6.91 | 779.6022 | 779.5996 | -0.0026 | -3 | 0.00007 | 1.7 | 5 |
| **Lipid Prof Negative** | PC(O-38:5) | 4 | C48H89NO9P-[M+OAc-H]- | 6.97 | 854.6302 | 854.6275 | -0.0027 | -3 | 0.00006 | 2.3 | 16 |
| **Lipid Prof Positive** | PC(O-44:5) | 4 | C52H97NO7P+ | 8.01 | 878.7051 | 878.7003 | -0.0048 | -5 | 0.00003 | 2.2 | 2 |
| **Lipid Prof Positive** | TG(58:6) | 2 | C61H110NO6+ [M+NH4]+ | 10.22 | 952.8394 | 952.8333 | -0.0061 | -6 | 0.00007 | 3.1 | 3 |
| **HILIC Positive** | UKN-HLP-4 | n/a | n/a | 8.14 | 645.3829 | n/a | n/a | n/a | 0.00004 | 3.5 | 25 |

*Statistics and assignment on second isotopic m/z

The positions of double bonds and fatty acyl chains cannot be determined; fatty acyl chains are presented from lowest to highest molecular weight (lower number of carbons to higher and higher number of double bonds, to lower).

Two-tailed t-tests were conducted, assuming unequal variance.

RT: retention time; PC: phosphatidylcholine; TG: triglyceride; UKN: unknown; OAc: acetate.

LoA: level of assignment; 1: m/z matched to online databases; 2: m/z matched to online databases and prior knowledge for RT (in-house database of standards); 3: in-source fragmentation; 4: assignment based on $MS^E$ mode spectrum.

#### *3.3.2.2 Aqueous Extracts*

The aqueous extracts of the atherosclerotic tissue were run using a HILIC-MS method. Similarly to the lipid profiling methodology, the two diseased groups studied also showed discrimination in the PCA scores plots (Figure 3. 5). Representative chromatograms of the method can be viewed in Figure 3. 7 and Figure 3. 8. The positive mode spectra demonstrated good group discrimination in scores plots of the 1$^{st}$ and 2$^{nd}$ PComps (Figure 3. 5.A). These two components together constituted 42% of the variation ($R^2X$), with the 1$^{st}$ PComp contributing 23% to the variation. The negative mode also presented good grouping in 3 PComps, 1$^{st}$, 2$^{nd}$, and 3$^{rd}$ (Figure 3. 5.B). They explain a total variation of 50%, with the 1$^{st}$ and 2$^{nd}$ PComps contributing with 23% and 14%, respectively indicating that the difference between the symptomatic and asymptomatic lipid profiles were responsible for a large part of the variance in the metabolic profiles.

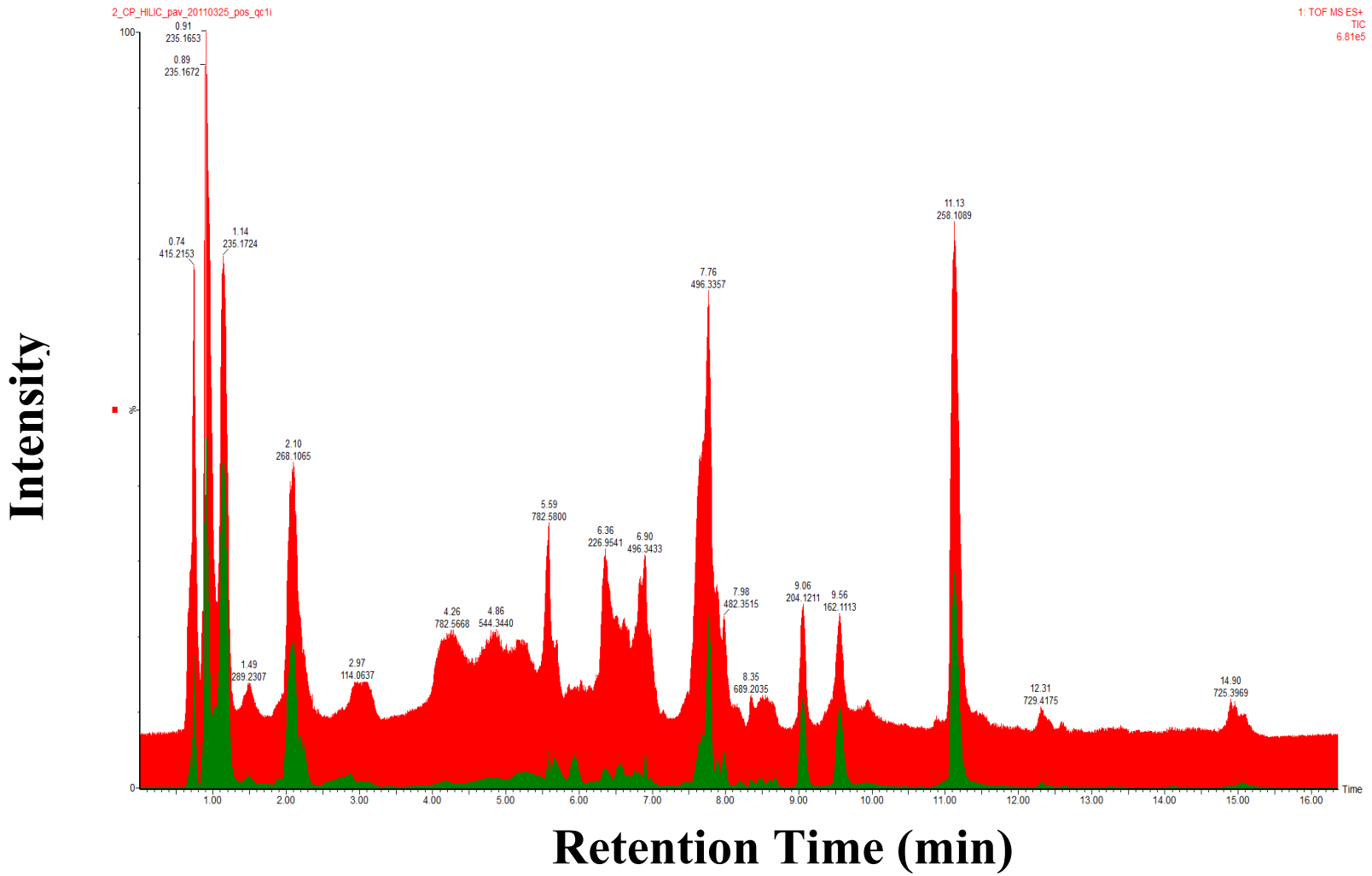


*Figure 3. 7: A representative chromatogram from the positive mode of the UPLC-MS analysis, of the aqueous extracts of the tissue samples examined. The trace in red colour represents the total ion current, while in green are the base peak intensities (highest peaks within a retention time window)*

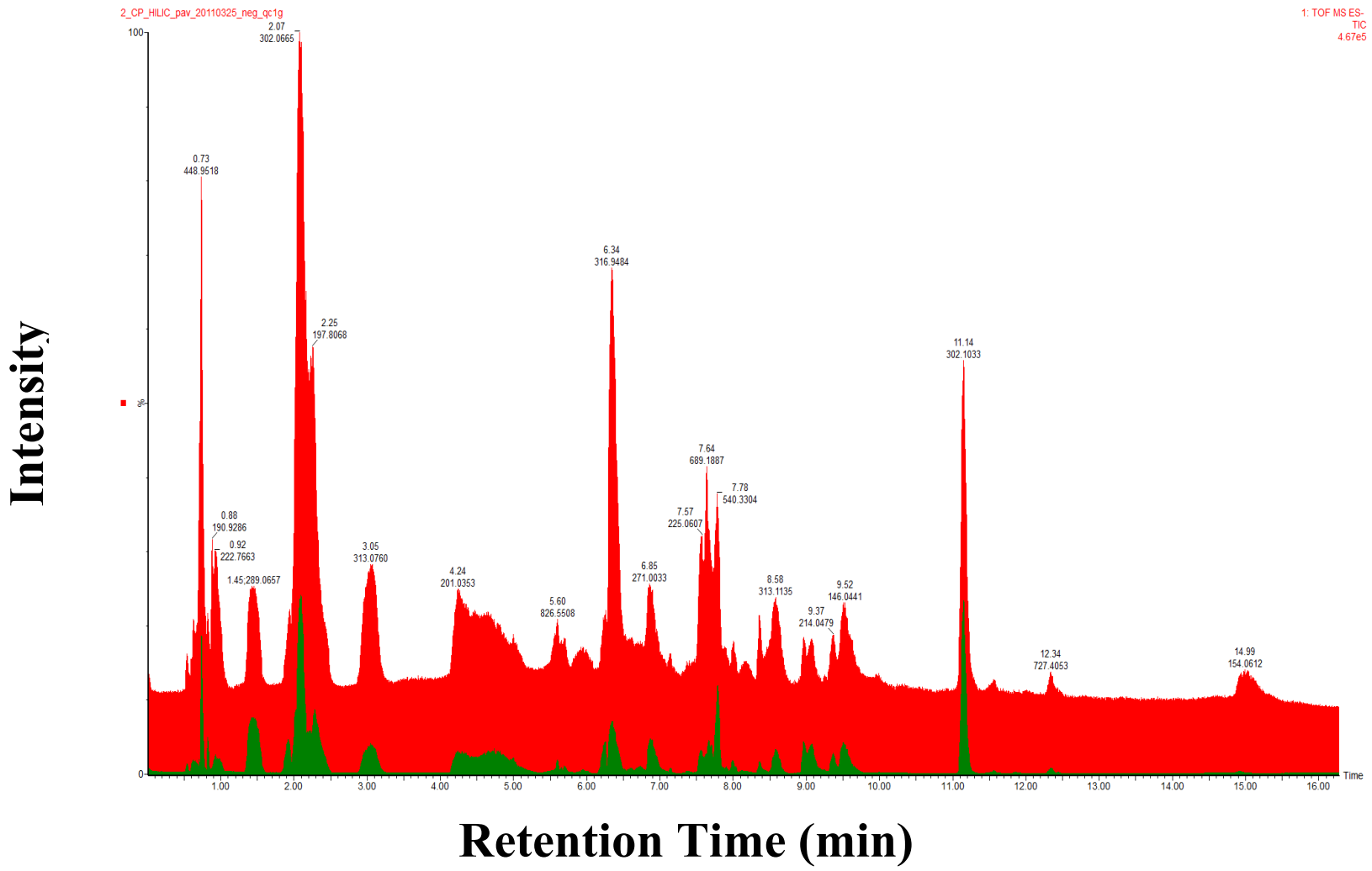


*Figure 3. 8: A representative chromatogram from the negative mode of the UPLC-MS analysis, of the aqueous extracts of the tissue samples examined. The trace in red colour represents the total ion current, while in green are the base peak intensities (highest peaks within a retention time window).*

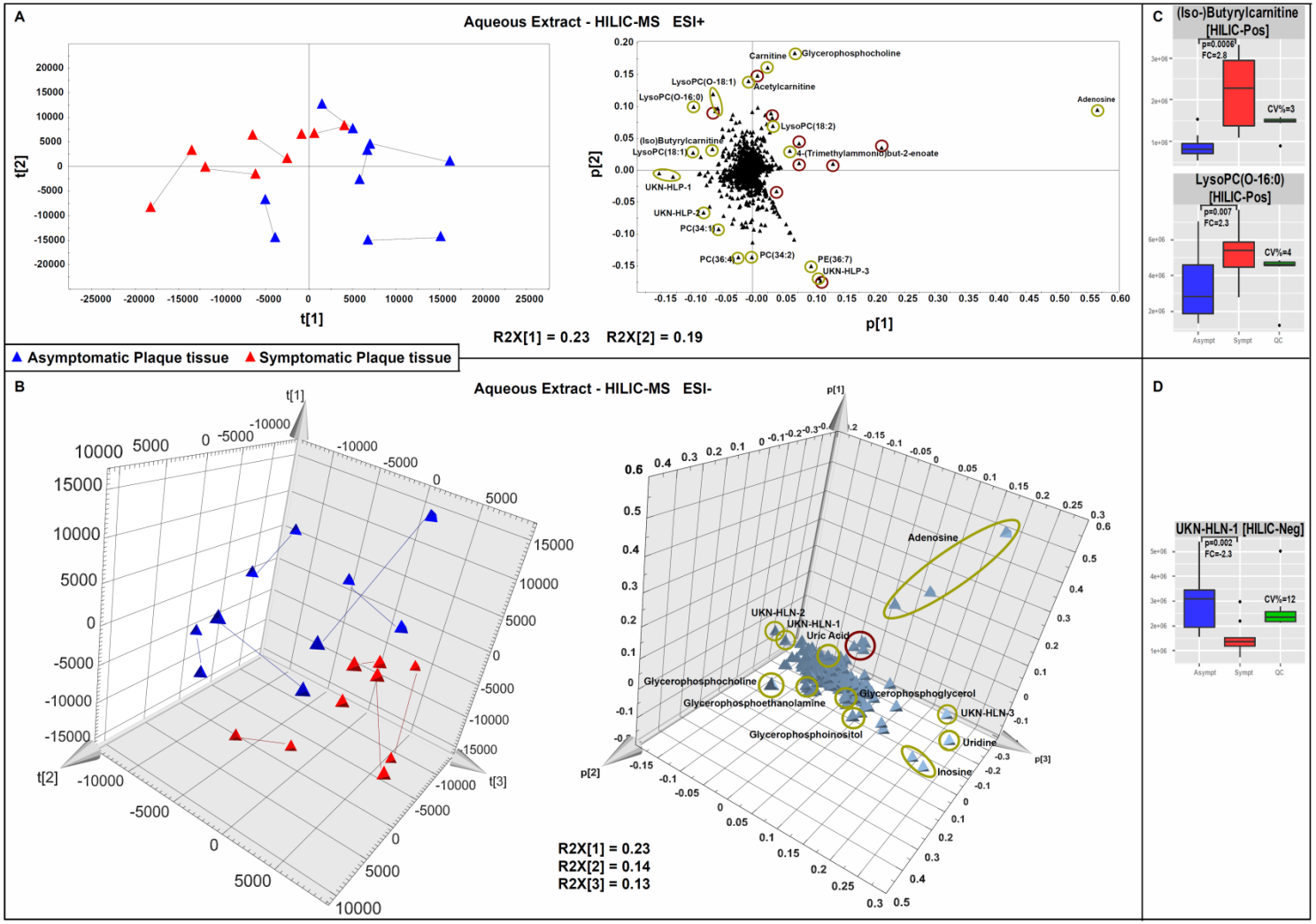


*Figure 3. 9: A and B represent scores and loadings plots from the PCA of the results obtained from the lipid profiling analysis of the organic extracts (A positive and B negative mode). C) and D) Box-plots of selected metabolites showing high statistical significance. Colour code: green: QC; blue: Asymptomatic; red: Symptomatic; dark yellow circles: metabolites assigned; dark red circles: features identified to represent isotopes, adducts, or fragments of identified metabolites.*

*Table 3. 6: List of assigned metabolites driving the separation of symptomatic and asymptomatic samples in the PCA loadings plots obtained from data of aqueous extracts.*

| Met Name | LoA | Mol Formula [complex adduct] | RT (min) | m/z (found) | m/z (theor) | ΔDa | Δppm | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | |
| **HILIC Positive** | | | | | | | | | | |
| (iso)Butyrylcarnitine | 3 | C11H22NO4+ | 8.36 | 232.1522 | 232.1549 | 0.0027 | 12 | 0.00059 | 2.8 | 3 |
| Acetylcarnitine | 3 | C9H18NO4+ | 9.23 | 204.1245 | 204.1236 | -0.0009 | -4 | 0.16980 | 1.1 | 4 |
| Adenosine | 3 | C10H14N5O4+ | 2.11 | 268.1068 | 268.1046 | -0.0022 | -8 | 0.05133 | -1.4 | 9 |
| Carnitine | 2 | C7H16NO3+ | 9.75 | 162.1123 | 162.1130 | 0.0007 | 4 | 0.64011 | 1.1 | 4 |
| Glycerophosphocholine | 3 | C16H41N2O12P2+ [2M+H]+ | 11.36 | 515.2134 | 515.2135 | 0.0001 | 0 | 0.49445 | 1.0 | 7 |
| lysoPC(18:1) | 3 | C26H53NO7P+ | 6.97 | 522.3514 | 522.3560 | 0.0046 | 9 | 0.21956 | 1.4 | 95 |
| lysoPC(18:2) | 3 | C26H51NO7P+ | 6.80 | 520.3374 | 520.3403 | 0.0029 | 6 | 0.85518 | 1.0 | 74 |
| lysoPC(O-16:0) | 3 | C24H52NO6PNa+ [M+Na]+ | 8.13 | 504.3422 | 504.3430 | 0.0008 | 2 | 0.00663 | 2.3 | 4 |
| lysoPC(O-18:1) | 3 | C26H55NO6P+ | 8.04 | 508.3755 | 508.3767 | 0.0012 | 2 | 0.04835 | 1.8 | 6 |
| PC(34:1) | 3 | C42H83NO8P+ | 0.98 | 760.5934 | 760.5856 | -0.0078 | -10 | 0.83316 | 1.0 | 13 |
| PC(34:2) | 3 | C42H81NO8P+ | 0.98 | 758.5769 | 758.5700 | -0.0069 | -9 | 0.16903 | -1.4 | 13 |
| PC(36:4) | 3 | C44H81NO8P+ | 0.97 | 782.5764 | 782.5700 | -0.0064 | -8 | 0.55840 | -1.1 | 13 |
| PE(36:7) | 3 | C41H69NO8P+ | 1.54 | 734.4712 | 734.4761 | 0.0049 | 7 | 0.17197 | 1.4 | 12 |
| 4-(Trimethylammonio)but-2-enoate | 2 | C14H26N2O4Na+ [2M+Na]+ | 7.90 | 309.1791 | 309.1790 | -0.0001 | 0 | 0.34349 | 3.1 | 4 |
| UKN-HLP-1 | n/a | n/a | 10.10 | 616.1784 | n/a | n/a | n/a | 0.13546 | 2.2 | 17 |
| UKN-HLP-2 * | n/a | n/a | 12.58 | 729.4158 | n/a | n/a | n/a | 0.15598 | 2.4 | 13 |
| UKN-HLP-3 | n/a | n/a | 0.74 | 324.1407 | n/a | n/a | n/a | 0.17574 | -3.3 | 9 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula [complex adduct] | RT (min) | m/z (found) | m/z (theor) | ΔDa | Δppm | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | |
| **HILIC Negative** | | | | | | | | | | |
| Adenosine | 3 | C10H13N5O4Cl- [M+Cl]- | 2.13 | 302.0673 | 302.0656 | -0.0017 | -6 | 0.17377 | -1.3 | 6 |
| Glycerophosphocholine | 3 | C9H21NO8P- [M+FA-H]- | 11.37 | 302.1007 | 302.1005 | -0.0002 | -1 | 0.26149 | -1.1 | 3 |
| Glycerophosphoethanolamine | 3 | C5H13NO6P- | 9.55 | 214.0477 | 214.0480 | 0.0003 | 1 | 0.86287 | 1.0 | 3 |
| Glycerophosphoglycerol | 3 | C6H14O8P- | 7.99 | 245.0434 | 245.0426 | -0.0008 | -3 | 0.07132 | 1.4 | 11 |
| Glycerophosphoinositol | 3 | C9H18O11P- | 9.22 | 333.0596 | 333.0587 | -0.0009 | -3 | 0.67020 | -1.1 | 4 |
| Inosine | 3 | C11H13N4O7- [M+FA-H]- | 3.12 | 313.0789 | 313.0784 | -0.0005 | -2 | 0.03082 | 1.4 | 4 |
| UKN-HLN-1 | n/a | n/a | 8.53 | 728.2135 | n/a | n/a | n/a | 0.00217 | -2.3 | 12 |
| UKN-HLN-2 | n/a | n/a | 0.75 | 448.9498 | n/a | n/a | n/a | 0.02408 | -1.4 | 4 |
| UKN-HLN-3 | n/a | n/a | 7.03 | 271.0026 | n/a | n/a | n/a | 0.11986 | 1.3 | 10 |
| Uric Acid | 1 | C5H3N4O3- | 5.19 | 167.0196 | 167.0205 | 0.0009 | 5 | 0.11678 | -1.1 | 2 |
| Uridine | 3 | C10H13N2O8- [M+FA-H]- | 1.47 | 289.0681 | 289.0672 | -0.0009 | -3 | 0.04614 | 1.1 | 1 |

* Assigned as the [2M+Na]+ adduct of 365.2049

PCs and lysoPCs the positions of double bonds and fatty acyl chains cannot be determined; fatty acyl chains are presented from lowest to highest molecular weight (lower number of carbons to higher and higher number of double bonds, to lower).

Two-tailed t-tests were conducted, assuming unequal variance.

RT: retention time; PC: phosphatidylcholine; PE: phosphatidylethanolamine; UKN: unknown; FA: formate.

LoA: level of assignment; 1: m/z matched to online databases; 2: m/z matched to online databases and prior knowledge for RT (in-house database of standards); 3: in-source fragmentation; 4: assignment based on $MS^E$ mode spectrum.

Features that were driving this separation can be seen in the loadings plots (Figure 3. 9) and included short-chain AcCs, carnitine, lysoPCs, PCs, glycerophosphocholine, glycerophosphoethanolamine, glycerophosphoinositol, adenosine, inosine and uridine. Features that could not be assigned are marked as UKN (Figure 3. 9 and Table 3. 6). Table 3. 6 presents the metabolites identified, and unassigned features, along with UPLC-MS analysis feature characteristics, and results of univariate statistics. Metabolites that appeared to drive the separation of the two groups are presented with box-plots on Figure 3. 9C and D. These include (iso-)butyrylcarnitine and lysoPC(O-16:0) with p=0.0006 and p=0.007, respectively.

Univariate statistics (Table 3. 5) demonstrated two highly significant features. One unidentified (UNK-HLP-4) with p=0.00004, and hexanoylcarnitine (detected as the second isotope) with p=0.00003, and FC of 1.9. Of note, the first isotope of hexanoylcarnitine was detected with p=0.0003 and FC of 1.7. The box-plots of all the detected statistically significant metabolites of the aqueous extract HILIC-MS analysis are provided in Figure 3. 6.

## 3.4 Discussion

Patients with symptoms of mini-stroke (symptomatic) have a higher risk of a life threatening stroke in the immediate future after the presenting mini-stroke symptoms[137]. Thus, it is of supreme importance for risk related markers to be provided to the clinic. In this study, a metabolic profiling discrimination of symptomatic and asymptomatic carotid plaque tissue is shown for the first time using multivariate statistical analysis. Furthermore, highly significant individual markers of symptomatic plaques are also reported. A number of them showed $p<0.0001$ in the two-tailed t-test and greater than 2-fold change in normalised intensity (Table 3. 4, Table 3. 5 and Table 3. 6). From literature, previous efforts using a shotgun lipidomics approach to accomplish multivariate statistical discrimination between symptomatic and asymptomatic carotid plaque tissue proved unsuccessful[140].

Metabolic profiles generated from NMR analyses of aqueous and organic extracts of the same samples did not show significant separation of Sympt and Asympt. The same for aqueous extracts ran with an RP-UPLC-MS method. It could be hypothesised that platforms, which were unable to deliver disease related profiles, may be more sensitive, or otherwise appropriate, for metabolites that do not show high statistical significance between the groups studied here. Therefore, the two methodologies presented herein (RP-UPLC-MS on organic extracts and HILIC UPLC-MS on aqueous extracts) will be the methods of choice for further analyses of larger sample sets. Using the combined power of both these analysis methods, these two methods can provide detection of a

range of compounds with wide physicochemical properties, wherein it was possible to structurally assign up to 50 metabolites. These methods were proven to be robust, as evidenced by the tight grouping of QC samples in MVDA (Figure 3.2). Additionally, the coefficient of variation calculated by the 7 injections of the QC sample through the run, gave further confidence in the analytical quality of the model as most of the CV% were typically <10%.

By applying the lipid profiling methodology to analyse the tissue organic extracts, a number of lipid moieties, from 5 different lipid classes, were structurally assigned. These classes are: fatty acids, glycerolipids, glycerophospholipids, sphingolipids and sterol lipids. The identified lipids represent a wide range of lipid subclasses. These include a number of PCs and lysoPCs, SMs and Cers, CEs, TGs and DGs, PEs, AA, and also a long-chain AcC.

On the other hand, due to the properties of HILIC columns, good interaction with the polar structural groups of metabolites could be achieved. Polar metabolites are retained and subjected to good chromatographic separation. Good retention would be an issue with conventional RP-UPLC methods for these metabolites[92]. Metabolites detected driving the models and/or being statistically significant were structurally identified. These included three purines, adenosine, inosine and uric acid, as well as uridine, a pyrimidine. Additionally, carnitine and a number of AcCs were detected, namely acetylcarnitine, (iso-)butyrylcarnitine (note that there is difficulty in the determination between butyryl- and isobutyryl- isomers by MS) and hexanoylcarnitine. Three major glycerophosphates were detected, glycerophosphocholine, glycerophosphoethanolamine and glycerophosphoinositol. A number of glycerophospholipids was detected, which is not surprising provided that the aqueous extraction was conducted using a 50% MeOH solution and therefore acted as a solvent for some organic molecules. Lastly, features that could not be assigned are reported in MVDA and univariate statistics.

An important part of the experimental design was to explore if the metabolic fingerprint from different locations but within the same plaque was comparable. However, biological reproducibility was very difficult to assess as it is hard to obtain multiple biological replicates of the tissue, and in good amounts, in order to perform such a study. It was however possible to obtain two tissue specimens from almost all of the carotid plaque samples. In PCA Figure 3. 5 and Figure 3. 9, duplicates are shown connected with a line. Some of the duplicates indeed displayed low variation. However, the variation of some of the duplicates was high. Nevertheless, despite these differences in the composition of samples from the same plaque, only on one occasion did the variation observed from a biological replicate affect the disease-related grouping of the samples in PCA. This

observation of intra-plaque variability was to some extent expected, as different pathophysiological conditions can coexist within the same plaque tissue giving rise to structural heterogeneity. Therefore, the potentially high variation within the same tissue sample should be considered in further studies.

The current study functions as an exploratory study on a small sample set (n=10; five patients per group). This study was designed with the objective of obtaining pilot data demonstrating feasibility and ability to provide disease-related information. The small sample set and the use of biological duplicates is the reason the statistical analysis was limited to only unsupervised MVDA as supervised methods of MVDA are susceptible to overfitting on small sample sets. Additionally, the use of t-test and fold-change comparisons between metabolites should be considered indicative of the potential and not definitive. Further analyses of larger sample groups should validate these findings.

As discussed, because of the small sample group size it would be risky to try to biologically interpret some of the detected statistically significant metabolites. However, some of differentially detected metabolites are members of the same biological pathways. This provides some extra confidence to these findings, and should also guide future experimentation employing targeted approaches that could validate the presented findings.

One of the most important differences between samples from symptomatic and asymptomatic patients was the finding of higher intensities of AA in Sympt. Higher intensities of this 20:4 free fatty acid and a 20:4 fatty acyl-chain bearing PC, PC(16:0/20:4), were detected in this study and are likely to indicate differences in the same metabolic pathway. PC(16:0/20:4) was detected with statistical significance in both ESI modes. PC(16:0/20:4) can release AA after being hydrolysed by the enzyme Phospholipase A2 (PLA2). At the same time AA is the precursor molecule of a wide spectrum of inflammation-related compounds, the eicosanoids. This is in concordance with literature findings. It has been reviewed by *Libby et al* that plaques in high risk of rupture experience higher inflammation[52, 54]. This is very intriguing and although the results are not conclusive on this matter, it provides a good hypothesis for future experimentation.

Another interesting trend is the higher intensities of short, medium, and long-chain AcC. The (iso-)butyrylcarnitine, hexanoylcarnitine and palmitoylcarnitine were all detected in higher normalised intensities in Sympt. However, acetylcarnitine and carnitine were unaffected. The fact that carnitine remains unaffected provides more evidence pointing towards dysregulation of the pathway of β-oxidation. The β-oxidation takes place in the mitochondria, and consists of the catabolism of free

fatty acids after penetrating the mitochondrial membrane as acyl-carnitines. However, if this process is somehow down-regulated in Sympt, it still cannot explain why all AcC are higher while acetylcarnitine, the final product of this reaction remain unaffected. It could be possible that other catabolic pathways are able to counterbalance for this reduction as they can produce acetyl groups in the form of acetyl-CoA, or these are the pathways that are upregulated, and as a result fatty acid catabolism remains unused.

In summary, the findings demonstrated here provide good proof of the ability of these metabolic profiling technologies to risk stratify patients with carotid atherosclerotic plaques, in danger of rupture and thromboembolism. Such findings could assist towards identifying candidate biomarkers that could be detected via *in-vivo* imaging, or targeted in bodily fluids such as blood. The potential of indentifying pathways that can be targeted for therapeutic purposes is also possible and should be explored. Taken together, the results from this analysis provide good evidence to further pursue a larger sample study of symptomatic and asymptomatic patients. The untargeted methodologies that were applied and succeeded on delivering results towards the discrimination of symptomatic patients should form the core of this study. Additionally, targeted approaches should be employed to investigate systemic evidence of involvement of pathways such as the eicosanoid and β-oxidation pathways.

# *Chapter 4 Application of a metabolic profiling approach to identify latent metabolites and interactions of atherogenesis. Distinct profiles detected between carotid and femoral plaques.*

## *4.1 Introduction*

Atherosclerosis is the number one cause of death in the western world[45]. Most adverse health events associated with atherosclerosis develop alongside plaque formation. Existing pharmaceutical schemes appear to fail in providing a positive outcome in primary prevention, while the scientific community still struggles to find appropriate dose or drug regimes[141].

Atherosclerosis is a multicentric, multistage and systemic disease. Cell populations such as vascular smooth muscle cells and leukocytes are known for their involvement in the manifestation of the disease. These cells participate in microenvironmental interactions in the arterial tissue. To understand the role of these interactions in disease aetiology and progression, they need to be positioned in a holistic framework in order to assist towards the generation of much needed novel hypotheses.

Metabolic profiling can be an effective systems biology approach. By using multivariate data analysis to analyse complex spectral metabolite profiles, global information for the disease of interest can be obtained through identification of dysregulated metabolites and can lead to identification of dysregulated pathways. In this study, the metabolic changes of the progression from intimal thickening to plaque formation are explored. Tissue extracts from carotid and femoral plaque samples were compared to intimal thickening extracts using two different ultra performance liquid chromatography coupled to mass spectrometry (UPLC-MS) methodologies, covering a wide range of metabolites, and with substantial diversity in physicochemical properties. Both the global metabolic profile of plaque formation, as well as metabolic differences attributed to anatomical location, have been characterized. Previously unassociated, as well as recognized, metabolites and

biological pathways related to atherogenesis, are demonstrated, both globally and distinctively to the two tested locations. Finally, significant statistical interactions between detected biomarkers and biological pathways in relation to the disease are assessed.

Additionally a novel experimental design is followed for this study. The concept of the experimental design is to explore the metabolic alterations of the metabasis to the lipid-laden lesion, within the arterial wall. The atherosclerotic lesion is the 'culprit' of the adverse health events. For this, intimal thickening tissue (INT) (n=17) was used as control. Using INT comes with several advantages: 1) INT, rather than normal tissue, is detected at plaque prone sites from the early years of a person's life, 2) it is the immediate stage prior to progression to the lesion, thus the experimental model can provide a more realistic course of the disease, and 3) normal tissue is difficult to obtain for obvious ethical reasons. Further, plaque lesions from two different anatomical locations in the arterial tree, constituting the major sites of atherogenesis, were used: carotid and femoral. Establishing the metabolic phenotype for these sites addresses the necessity of correlating patients' outcome to pharmaceutical schemes. The efficacy of drugs, such as statins could be evaluated, in the context of differences in metabolic phenotype and lipid texture of the lesions, e.g. lipophilic properties of drugs. Additionally, metabolic profiling should be assessed based on the different hemodynamic properties of these anatomical locations. Lastly, comparisons between these anatomical locations are rare in literature and non-existent in the case of metabolic profiling studies.

## 4.2 Methods

### 4.2.1 Patients

The tissue plaque samples used in this study were obtained from a total of 78 patients from the Department of Vascular Surgery, Imperial College London. Segments were retrieved, snap frozen in liquid nitrogen and stored at -80°C at surgery. These segments were obtained from 52 patients who underwent carotid endarterectomy (CAR), and 26 who underwent femoral endarterectomy (FEM). From these samples, tissues without atheroma, but with obvious evidence of thickening of the intima vessel layer (INT), were harvested (9 samples from carotid and 7 samples from femoral tissue). Intima thickening tissue was found at the shoulder of forming atheromas, and in this study serves as control tissue. Patients' clinical characteristics can be found in Table 4. 1.

*Table 4. 1: Patients demographics of carotid, femoral, and intimal thickening groups.*

| | Carotid | Femoral | Intima thickening |
|---|---|---|---|
| Number of Samples | 52 | 26 | 16 |
| Age, Median (range) | 69 (44-87) | 74 (60-91) | 69 (58-91) |
| Gender, Male (%) | 40 (77) | 17 (65) | 13 (81) |
| Statin (%) | 40 (82) {3} | 19 (73) | 13 (87) {1} |
| Antiplatelet (%) | 43 (86) {3} | 22 (85) | 14 (93) {1} |
| HT (%) | 32 (68) {5} | 13 (50) | 7 (47) {1} |
| DM (%) | 11 (22) {3} | 7 (27) | 6 (40) {1} |
| Ever smoker (%) | 25 (52) {4} | 18 (69) | 11 (73) {1} |
| BMI, Median, Range | 25.5 (17.4–41.2) {9} | 23.0 (21.6–33.3) {13} | 25.4 (21.0-47.7) {6} |

Key: HT: Hypertension; DM: Diabetes mellitus; BMI: Body Mass Index; {} number of missing patients

## 4.2.2 Tissue Extraction

Parts of plaque tissue, and intima tissue adjacent to plaque, were dissected and harvested for metabolite extraction. Effort was made to include as much plaque tissue as possible. Tissue weights were in the range of 416-153mg. Tissue samples were loaded into appropriate bead beating tubes (Percellys Steel-Kit, Germany) along with steel beads, and subjected to metabolite extraction by tissue lyses.

### 4.2.2.1 Aqueous extraction

Pre-chilled methanol/water solution (1:1) (methanol HPLC gradient grade, Fisher; water LC-MS grade, Fluka), was added to the tissue samples. The volume of the solution was adjusted according to weight of the sample starting at a maximum weight with 1.5mL, and reducing proportionally to sample weight. Samples were frozen on dry ice and loaded onto a bead beater (Bertin Technologies) vibrating at 6500Hz, 40 seconds, 4 cycles separated by freezing on dry ice. Centrifugation (Eppendorf, Centrifuge 5417R, Germany) followed at 13,000 rcf for 20 min; at $4^{O}$C. Aliquots of 100µL of supernatant were pipetted into Eppendorf tubes. Samples were spun on a vacuum concentrator for 2 hours at 45°C (Eppendorf Concentrator Plus, V-AQ mode). Samples were stored at -40°C until analysis.

### 4.2.2.2 Organic extraction

Following decanting of the supernatant from centrifuged aqueous samples, a solution of pre-chilled dichloromethane (DCM, $CH_2Cl_2$)/methanol (3:1) (methanol HPLC gradient grade, Fisher; DCM HPLC grade, Sigma-Aldrich) was added. The volume of the solution was proportional to the sample

weight (as described in previous paragraph; aqueous extraction). Samples were frozen on dry ice and re-loaded into the bead beater (2 cycles, 6500Hz, 40 seconds). Samples were centrifugation at 13,000 rcf for 20 minutes, followed by aliquoting of organic phase supernatant into glass vials. Samples were allowed to evaporate at room temperature in an extractor hood overnight and stored at -40°C for analysis.

### 4.2.3 HILIC-UPLC-MS analysis of Aqueous Extracts

#### 4.2.3.1 UPLC-MS Analysis

Samples were reconstituted in 200uL of solvent mixture of $H_2O$/acetonitrile (5:95) (LC-MS grade, Fisher Scientific, USA), and transferred into Total Recovery vials (Waters, USA), after centrifugation for 20min at 13000 rcf, 4°C.

UPLC separation was conducted using an Acquity UPLC System (Waters Corp, USA). An Acquity UPLC BEH HILIC 2.1x100mm, 1.8um, column (Waters Corp, USA) was used. Column temperature was set at 35°C. An injection volume of 10uL was used for both positive and negative ionization polarity modes. The auto-sampler was set at 4°C. Mobile phase A consisted of acetonitrile/water (95:5) and mobile phase B acetonitrile/water (50:50). In both solutions ammonium formate (98%, Fluka, USA) was diluted to 10mM and formic acid (MS grade, Fluka, USA) to 0.1%. The chromatographic gradient program is summarized in Table 4. 2.

*Table 4. 2: Gradient program of chromatography of HILIC UPLC-MS method of Aqueous Extracts*

| Time (min) | Flow rate ml/min | %A | %B | Curve |
|---|---|---|---|---|
| 0.0 | 0.4 | 99 | 1 | - |
| 2.0 | 0.4 | 99 | 1 | - |
| 8.0 | 0.4 | 45 | 55 | 6 |
| 9.0 | 0.4 | 1 | 99 | 6 |
| 9.1 | 0.8 | 1 | 99 | 6 |
| 11.0 | 0.8 | 1 | 99 | - |
| 11.1 | 0.8 | 99 | 1 | 6 |
| 19.0 | 0.8 | 99 | 1 | - |
| 19.1 | 0.4 | 99 | 1 | 6 |
| 23.0 | 0.4 | 99 | 1 | - |

Detection of eluting UPLC fractions was achieved using a Premier Q-TOF (Waters MS Technologies Ltd., UK). Scans were acquired with scan time of 0.20 s and interscan time of 0.02 s. Leucine Enkephalin was used for lock mass correction. Lock mass data were collected with scan time of 0.20 s and scan frequency of 30 s. Conditions in positive mode: m/z range:50-1000, cone voltage 30V, capillary voltage 3kV, Source Temperature 120$^{O}$C, Desolvation Temperature 400$^{O}$C, Desolvation Gas Flow 800L/h. Conditions in negative mode: m/z range: 50-1000, cone voltage 30V, capillary voltage 2.5kV, Source Temperature 120$^{O}$C, Desolvation Temperature 350$^{O}$C, Desolvation Gas 800L/h.

A QC format [114] was used for the UPLC-MS analysis. Briefly, a pooled sample (referred to as Quality Control Sample, QC) of the reconstituted extracts was prepared. This sample was re-injected 10 times before initiating the run to condition the column. Then the sample was re-injected once at the beginning, every 10 injections of samples, and at the end of the run (total of 13 injections).

#### *4.2.3.2 Data Processing*

Collected data were subjected to peak-picking and grouping using MarkerLynx XS (Waters Inc, v4.1) software, using the following parameters:

Positive mode: Function: 1, Analysis Type: Peak Detection, Initial Retention Time: 0.50, Final Retention Time: 10.00, Low Mass: 50.00, High Mass: 1000.00, XIC Window 0.10 Da, Peak Width at 5% Height: 25s, Marker Intensity Threshold: 60counts, Mass Window: 0.10Da, Retention Time Window: 0.50

Negative mode: Function: 1, Analysis Type: Peak Detection, Initial Retention Time: 0.40, Final Retention Time: 9.10, Low Mass: 50.00, High Mass: 1000.00, XIC Window 0.10 Da, Peak Width at 5% Height: 25s, Marker Intensity Threshold: 30counts, Mass Window: 0.10Da, Retention Time Window: 0.50

For Both Modes: Noise Elimination Level: 6.00, Peak-to-peak Baseline Noise: Auto, Replicate % Minimum: 0.00, No relative retention time, No Smoothing, No Deisotoping.

Values were reported as height of intensity peaks. Samples were normalized to total intensity. Values were multiplied by 10 000 prior to statistical analyses. The dilution series was used here to remove peaks that were not responding to dilution. This was done by applying multivariate statistics

on the QC samples and dilutions and removing features (variables) that were not responding to the dilutions. Additionally, features attributed to lidocaine and hydroxyl-lidocaine were removed. Lidocaine is administered locally prior to carotid endarterectomy.

## 4.2.4 Lipid Profiling

### 4.2.4.1 UPLC-MS analysis

Samples were reconstituted in 500uL of solvent mixture of $H_2O$ / isopropanol / acetonitrile (1:2:1) (Optima, LC-MS grade, Fisher Scientific, USA), and transferred into Total Recovery vials (Waters, USA), after centrifugation for 10min at 5g and 4°C.

UPLC separation was conducted using an Acquity UPLC System (Waters Corp, USA). An Acquity UPLC CSH C18 2.1x100mm, 1.7um, column (Waters Corp, USA) was used. Column temperature was set at 55$^{o}$C, flow rate of 0.4mL/min. Injection volume of 3uL and 7uL were used for positive and negative ionisation modes respectively. The auto-sampler was set at 4°C. Mobile phase A consisted of acetonitrile/water (60:40) and mobile phase B Isopropanol/acetonitrile (90:10). In both solutions ammonium formate (LC-MS grade, Fluka, USA) was diluted to 10mM and formic acid (MS grade, Fluka, USA) to 0.1%. The chromatographic gradient program is summarized in Table 4. 3.

*Table 4. 3: Gradient program of chromatography of the lipid profiling UPLC-MS methodology.*

| Time (min) | %A | %B | Curve |
|---|---|---|---|
| 0.0 | 60 | 40 | - |
| 2.0 | 57 | 43 | 6 |
| 2.1 | 50 | 50 | 1 |
| 12.0 | 46 | 54 | 6 |
| 12.1 | 30 | 70 | 1 |
| 18.0 | 1 | 99 | 6 |
| 18.1 | 60 | 40 | 6 |
| 20.0 | 60 | 40 | - |

Detection of eluting UPLC fractions was achieved using a Xevo G2 QTof (Waters MS Technologies, UK). Both MS and $MS^E$ data scans were acquired for 0.200 s every 0.214 s. $MS^E$ data were collected after ramping the collision energy from 30 to 40V. Leucine

Enkephalin was used for lock mass correction. Lock mass data were collected every 0.3 s for 0.2 s. Conditions in positive mode: m/z range:150-1200, cone voltage 30V, capillary voltage 2kV, Source Temperature 120$^{O}$C, Desolvation Temperature 550$^{O}$C, Desolvation Gas 900L/h. Conditions in negative mode: m/z range: 50-1200, cone voltage 30V, capillary voltage 1kV, Source Temperature 120$^{O}$C, Desolvation Temperature 550$^{O}$C, Desolvation Gas 900L/h.

As with HILIC-UPLC-MS analysis, the QC format was used, as described in the previous paragraph.

#### *4.2.4.2 Data Processing*

After acquisition, data were centroided (m/z spectra peaks are automatically detected and their centroid is calculated based on the average m/z value and weighted by the intensity). This was followed by peak-picking and grouping using MarkerLynx XS (Waters Inc, v4.1) software, using the following parameters:

Positive mode: Function: 1, Analysis Type: Peak Detection, Initial Retention Time: 0.40, Final Retention Time: 17.00, Low Mass: 150.00, High Mass: 1200.00, XIC Window 0.10 Da, Peak Width at 5% Height: 20s, Marker Intensity Threshold: 1000counts, Mass Window: 0.10Da, Retention Time Window: 0.50

Negative mode: Function: 1, Analysis Type: Peak Detection, Initial Retention Time: 0.45, Final Retention Time: 17.00, Low Mass: 50.00, High Mass: 1200.00, XIC Window 0.10 Da, Peak Width at 5% Height: 20s, Marker Intensity Threshold: 400counts, Mass Window: 0.10Da, Retention Time Window: 0.50

For Both Modes: Noise Elimination Level: 6.00, Peak-to-peak Baseline Noise: Auto, Replicate % Minimum: 0.00, No relative retention time, No Smoothing, No Deisotoping.

Values were reported as area of intensity peaks. Saturated peaks were removed, prior to total area normalisation. Values were multiplied by 10 000 prior to statistical analyses.

### *4.2.5 Statistical analysis*

Multivariate data analysis (MVDA) for UPLC-MS data was conducted using the SIMCA-P+ (v. 12.0.1.0.; Umetrics) package. Principal Component Analysis (PCA) and Orthogonal Projection to Latent Structures – Discriminant Analysis (OPLS-DA) were applied to the processed Pareto-scaled data. Model validation was carried out using CV-ANOVA testing [142].

In order to extract putative biomarkers from UPLC-MS, features (corresponding to metabolites) with correlation coefficient (Note: correlation coefficient that refers to correlation of samples to disease classes will be refer to as: p(corr)) greater than 0.5 in absolute value were initially chosen. This cut-off provides confidence greater than 99% (based on the number of samples). These features were further subjected to two-tailed t-test assuming unequal variance with a threshold of $p < 0.05$[143, 144] and fold-change comparison. Features were reported as significant and structurally identified if: 1) both p(corr) and t-test p-value met the thresholds , 2) were reproducible through the run, with a coefficient of variation (CV%) of the QCs less than 30%, and 3) pass the chromatographic peak shape assessment. A brief flowchart of the conditions a feature has to pass in order to be considered statistically significant is summarized in Figure 4. 1.

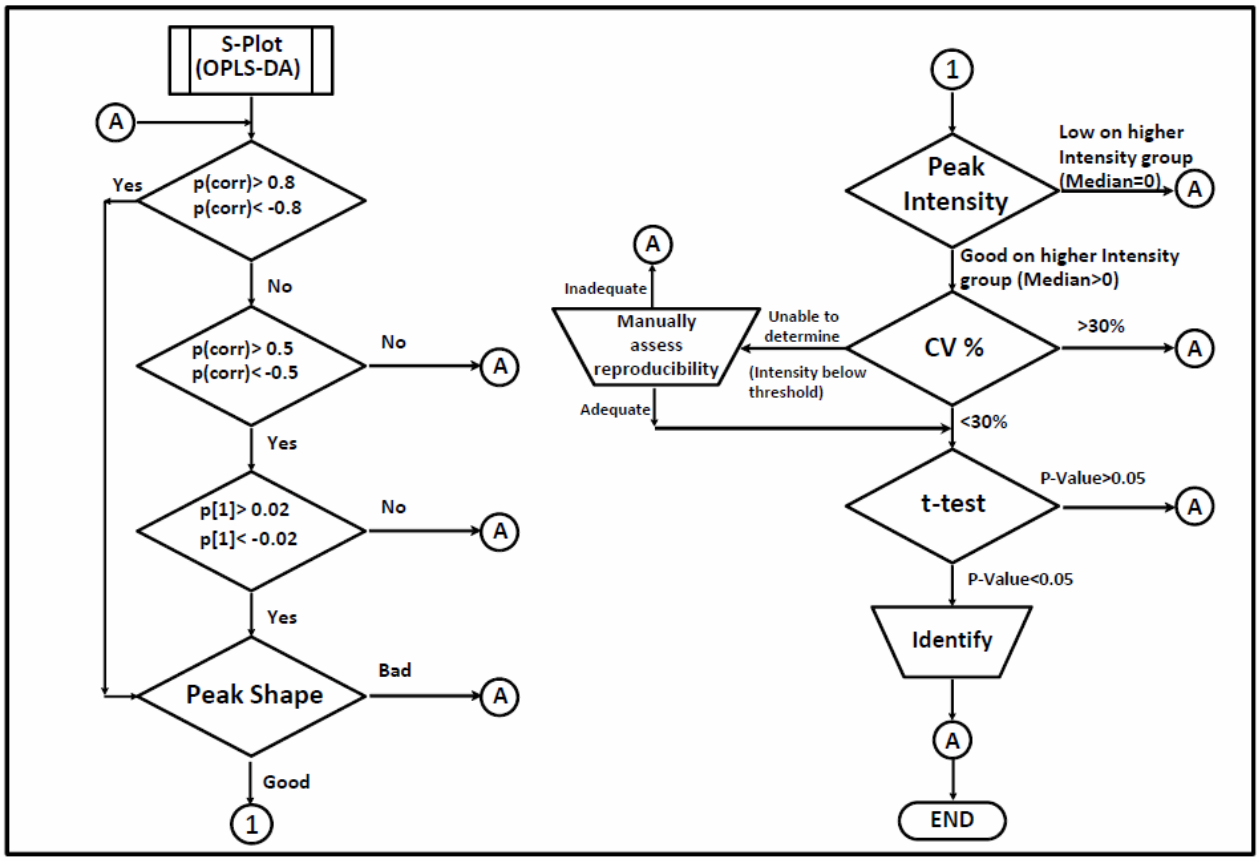


*Figure 4. 1: The workflow of conditions a feature has to fulfil in order to be considered statistically significant and forwarded to structural assignment.*

### 4.2.6 Metabolite Identification of Candidate biomarkers

For structural elucidation of significant features UPLC-$MS^E$ and UPLC-MS/MS data were used. For the lipid profiling $MS^E$ data were collected through the run as described for the UPLC-MS analysis, while for the HILIC-UPLC-MS analysis $MS^E$ data were collected on pooled samples (QCs) at the end of the run.

The same conditions described for each run were the ones used for MS/MS analysis that was conducted using data dependent acquisition (DDA) or by targeting specific ions. MS/MS data were collected with collision energy ramping from 30 – 50eV for lipid profiling and 20 - 40eV for HILIC analysis. Apart from fragmentation patterns, structural elucidation was assisted by matching accurate m/z measurements to metabolites from online available databases[111, 116, 139]. In some cases isotopic patterns also proved useful as well as an in-house library[92].

Where stated (Table 4. 6 - Table 4. 8) authentic standards were used to validate metabolite structural assignment. This included retention time matching of the authentic standard to the analyte, as well as matching of ms/ms spectra. This is accordingly stated by the level of assignment (LoA).

#### *4.2.6.1 Analytical Standards*

Standards were analytically grade with typical purity of >99%. Acetyl-DL-carnitine hydrochloride, adenosine, arachidonic acid sodium salt, benzoic acid, cholesterol sulphate, 2-deoxyuridine, guanosine, hypoxanthine, inosine and 1-methylnicotinamide, were purchased from Sigma. Linoleic acid, 5-methyluridine, and L-proline were purchased from Aldrich. Oleic acid, palmitic acid sodium salt, and stearic acid were purchased from Sigma-Aldrich. Creatine, 6-methylnicotinamide, niacinamide, sphingosine, uracil, and uridine were purchased from Fluka. Butyryl-L-carnitine chloride, decanoyl-L-carnitine chloride, isobutyryl-L-carnitine chloride, oleoyl-L-carnitine chloride, and valerylcarnitine-L-carnitine chloride were purchased from Larodan Fine Chemicals. Lauroylcarnitine chloride and propionylcarnitine chloride were purchased from TOCRIS bioscience, elaidic acid and N-acetyl-L-methionine from Alfa Aesar, and linoelaidic acid from Cayman Chemical Company. Free cholesterol was purchased from Nu-Chek Prep, Inc.

### *4.2.7 Pathway Analysis*

#### *4.2.7.1 Correlation network analysis*

Correlation coefficients (Spearman) between pairs of candidate biomarkers were calculated using R (2.13.2) programming language (Note: Correlation coefficients that describe the Spearman correlation between two metabolites will be referred to as: r). For network visualization, of significant metabolites characterizing either of CAR or FEM plaques, CytoScape (v.3.0.0-beta1) software was used. The cut-off for correlation was at all circumstances higher that 99% confidence.

#### *4.2.7.2 Metabolite Mapping*

Initial metabolite mapping was conducted by importing statistically significant metabolites in KEGG database[24]. A number of metabolites were mapped based on literature findings described.

#### *4.2.7.3 Venn Diagrams*

Venn diagrams were constructed using online available software (http://bioinfogp.cnb.csic.es/tools/venny/).

# *4.3 Results and Discussion*

## *4.3.1 Metabolic profiles of plaque formation*

For the first time the metabolic phenotype of plaque formation is assessed, from tissue of human patients, by comparison to INT. Metabolic profiles for the plaque tissues were determined using multivariate statistics. PCA shows, in an unsupervised fashion, differences in the profiles between tissue of INT and plaque lesions from the CAR and FEM locations (Figure 4. 2). It was however obvious that the metabolic profile of CAR tissue was more similar to the profile of INT, than FEM, forming distinct groups.

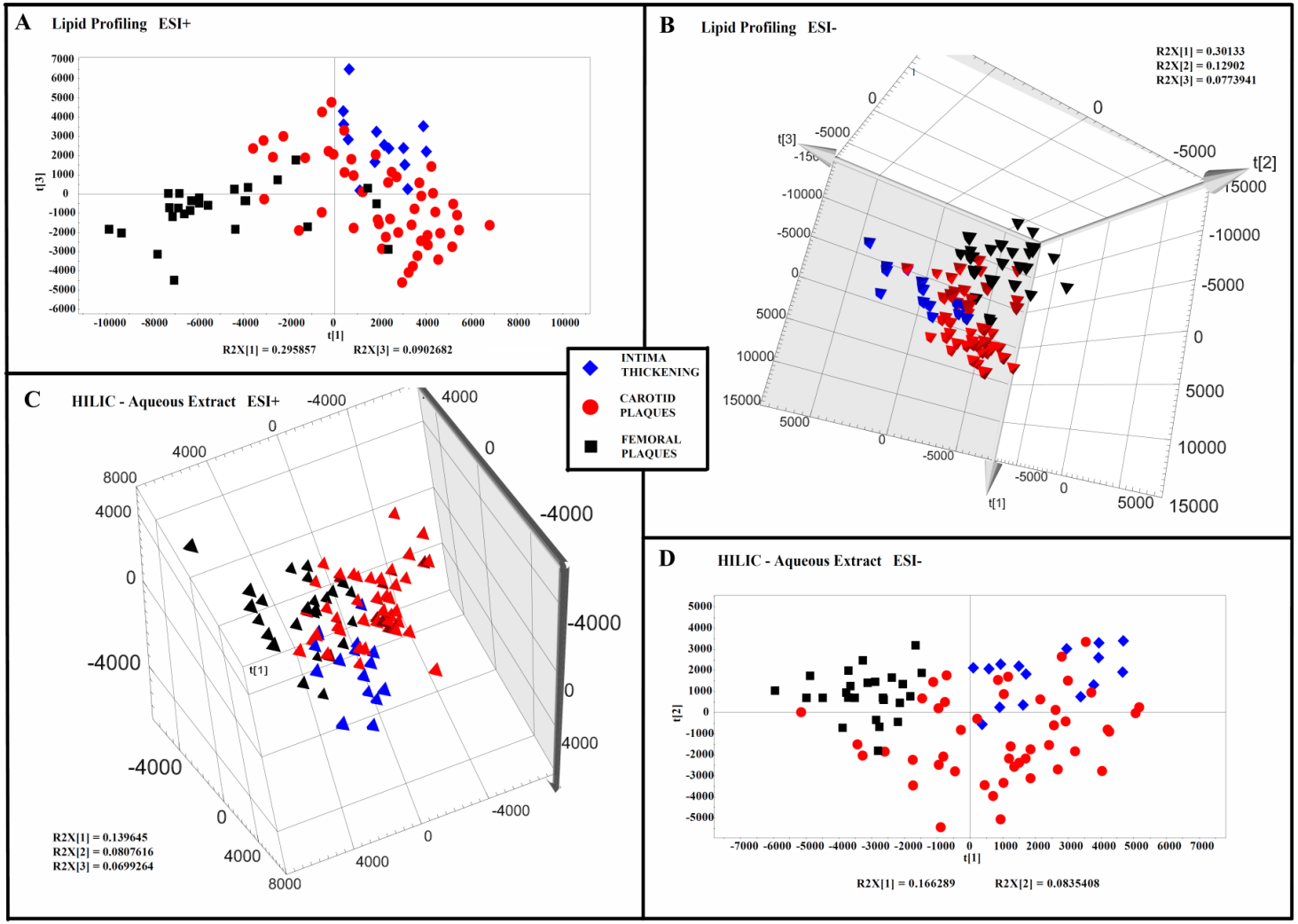


*Figure 4. 2: Scores plots of principal component analysis (PCA) of the tissue extracts. (A) Positive mode PCA of lipid profiling, (B) negative mode PCA of lipid profiling, (C) Positive mode PCA of aqueous extracts using HILIC-MS analysis, (D) negative mode PCA of aqueous extracts using HILIC-MS analysis. (Each point represents a sample; Blue rhombus: Intima thickening samples, Red dot: Carotid plaque samples, Black square: Femoral plaque samples.*

Further supervised MVDA modelling was employed for comparison of each of the plaque tissues separately. Orthogonal projection to latent structure – discriminant analysis (OPLS-DA) fitted models displayed high predictive values indicating distinctive metabolic profiles for each class, and diagnostics showed well-fitted and valid models, as assessed by model characteristics (Table 4. 4). Cross-validated scores plots of the samples included in the analysis are shown in Figure 4. 3. Using the constructed models, and in conjunction with univariate statistical analysis, significant features were structurally assigned to metabolites. Features were considered significant if their statistical characteristics met two criteria. Firstly, a good correlation, in absolute value, to the disease group vector in MVDA (|p(corr)|>0.5), and secondly from univariate statistics a p-value of t-test of < 0.05. In this study, structural assignments of > 150 unique metabolite identities were made on features that were found statistically significant in differentiating two or more groups (Table 4. 6, Table 4. 7, Table 4. 8; end of chapter). A large number of the metabolites have been detected as statistically significant in more than one polarity modes, or UPLC-MS analyses (Figure 4. 4 , Figure 4. 5 and Figure 4. 6). Metabolites detected include 5 different lipid classes and 17 subclasses[38]. Metabolites detected with statistically significance are presented in Table 4. 6, Table 4. 7 and Table 4. 8, at the end of this chapter.

*Table 4. 4: Summary of model characteristics from OPLS-DA multivariate statistical analyses of data obtained from all analysed and polarity modes.*

| Groups | CAR vs INT | | | | FEM vs INT | | | | CAR vs FEM | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Analysis | Lipid Profiling | | HILIC Aqueous | | Lipid Profiling | | HILIC Aqueous | | Lipid Profiling | | HILIC Aqueous | |
| Polarity Mode | Positive | Negative | Positive | Negative | Positive | Negative | Positive | Negative | Positive | Negative | Positive | Negative |
| Comp (pred+orthog) | 1 + 1 | 1 + 2 | 1 + 0 | 1 + 1 | 1 + 1 | 1 + 1 | 1 + 1 | 1 + 0 | 1 + 1 | 1 + 1 | 1 + 1 | 1 + 0 |
| $R^2X$ | 0.280 | 0.412 | 0.131 | 0.215 | 0.423 | 0.383 | 0.209 | 0.193 | 0.377 | 0.441 | 0.205 | 0.173 |
| $R^2Y$ | 0.609 | 0.786 | 0.496 | 0.721 | 0.861 | 0.834 | 0.895 | 0.846 | 0.78 | 0.668 | 0.823 | 0.638 |
| $Q^2Y$ | 0.432 | 0.502 | 0.356 | 0.467 | 0.755 | 0.746 | 0.703 | 0.797 | 0.658 | 0.576 | 0.648 | 0.575 |
| CV-ANOVA | $1x10^{-6}$ | $4x10^{-7}$ | $1x10^{-6}$ | $2x10^{-7}$ | $3x10^{-10}$ | $6x10^{-10}$ | $5x10^{-9}$ | $2x10^{-13}$ | $7x10^{-16}$ | $2x10^{-12}$ | $3x10^{-15}$ | $2x10^{-13}$ |

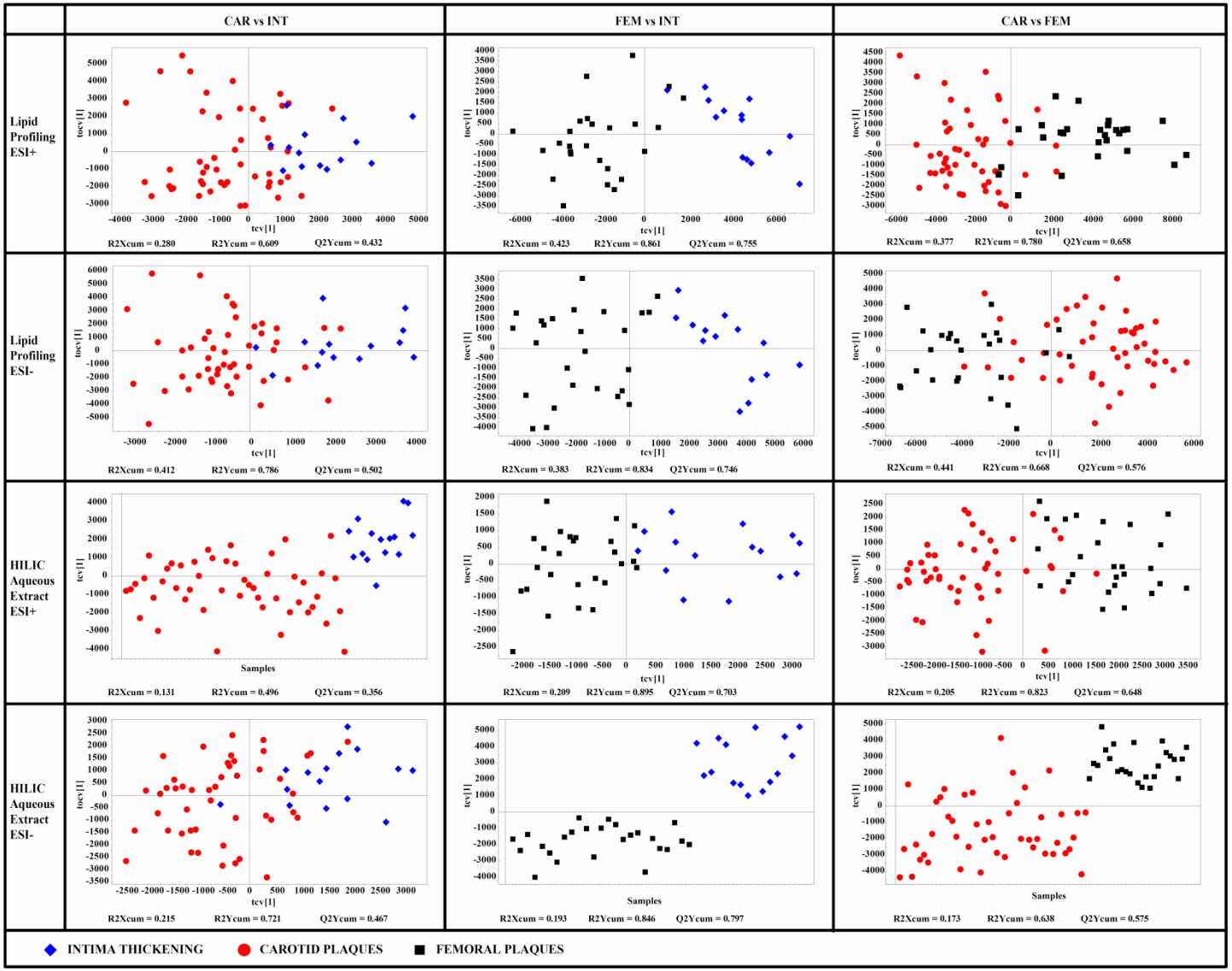


*Figure 4. 3: OPLS-DA cross-validated scores plots of tissue extracts. Columns represent group comparisons, and rows different UPLC-MS analyses and polarity modes.*

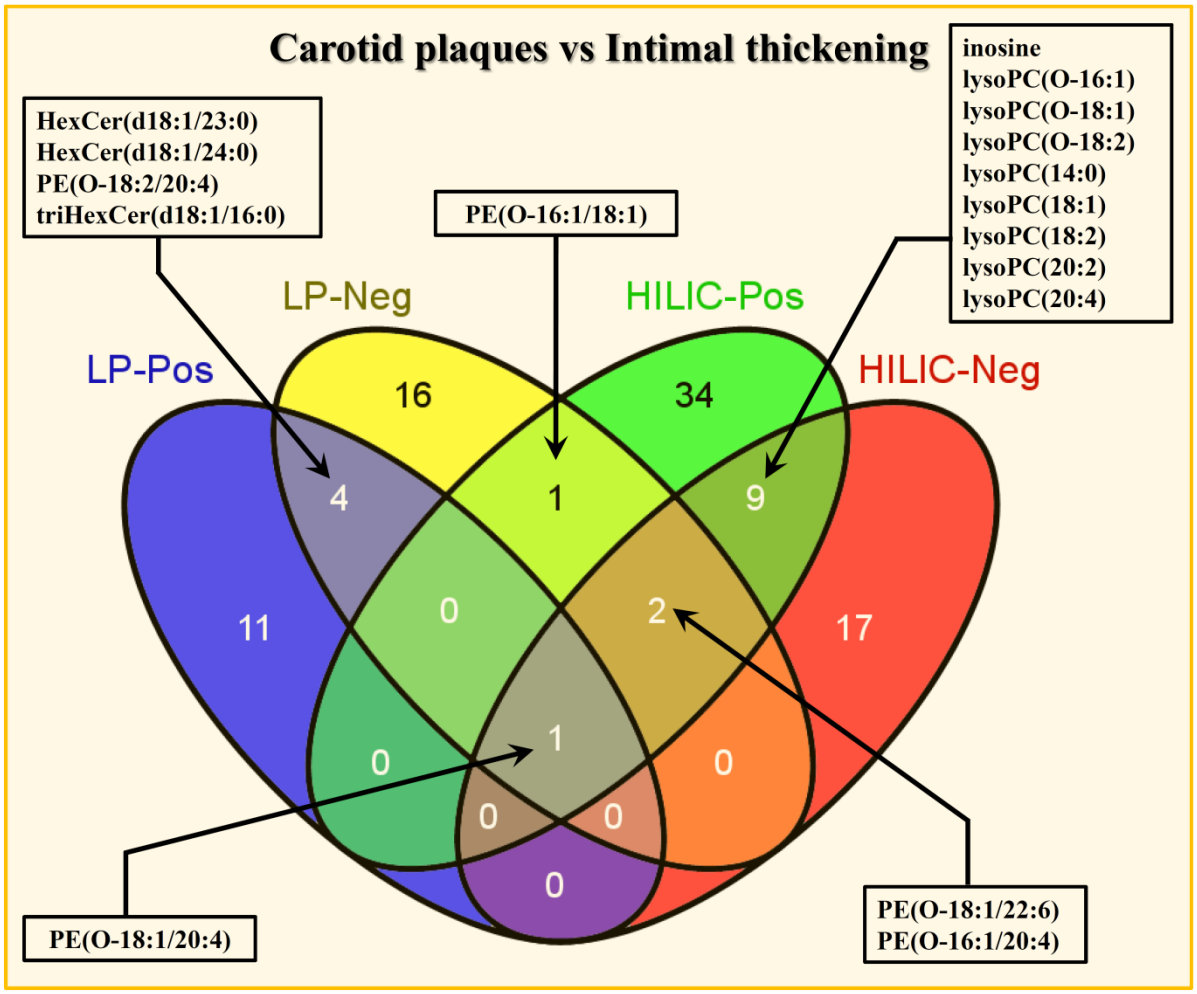


*Figure 4. 4: Venn diagram indicating identical metabolites detected statistically significant in the comparisons between carotid plaques and intimal thickening samples. Metabolites detected with the same trend in more than one analysis and/or polarity mode are shown.*

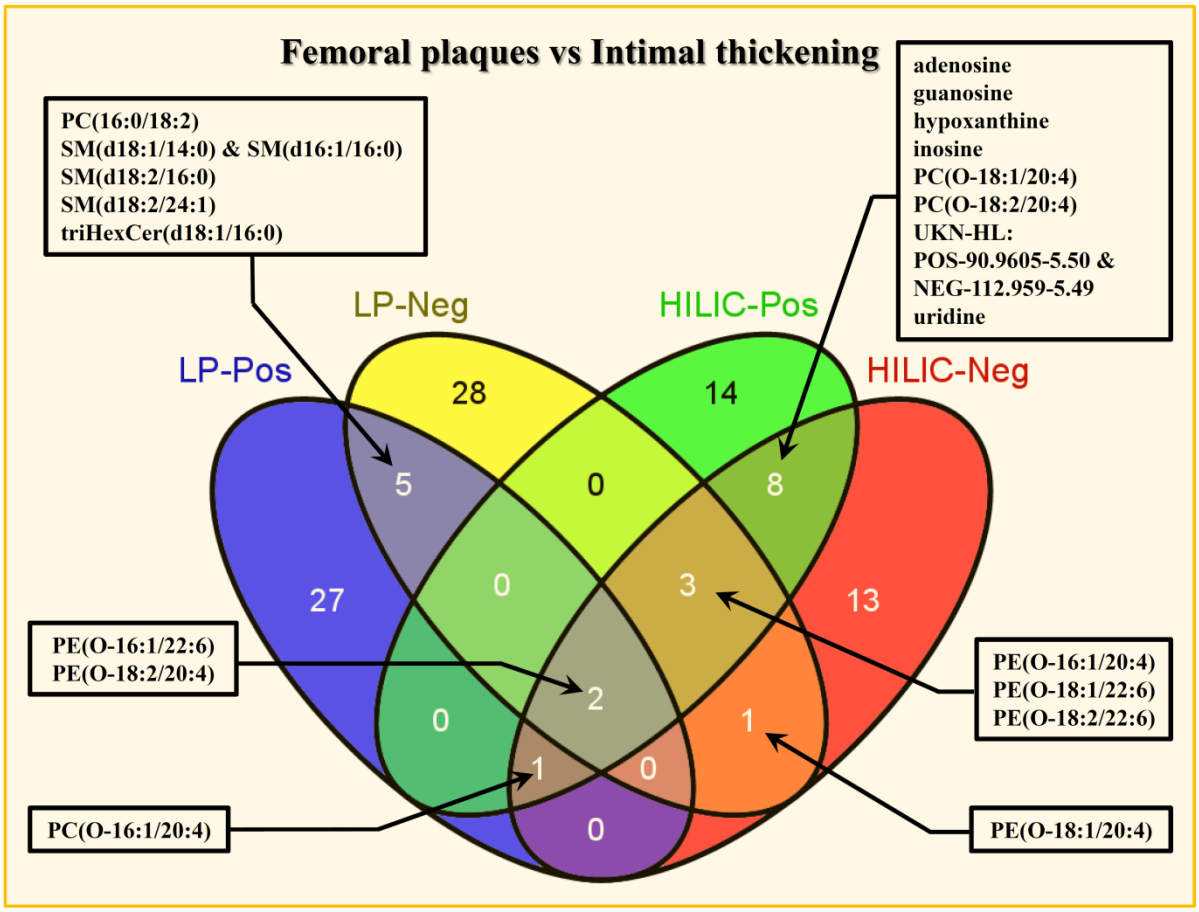


*Figure 4. 5: Venn diagram indicating identical metabolites detected statistically significant in the comparisons between femoral plaques and intimal thickening samples. Metabolites detected with the same trend in more than one analysis and/or polarity mode are shown. Tentatively assigned metabolites are indicated with an asterisk.*

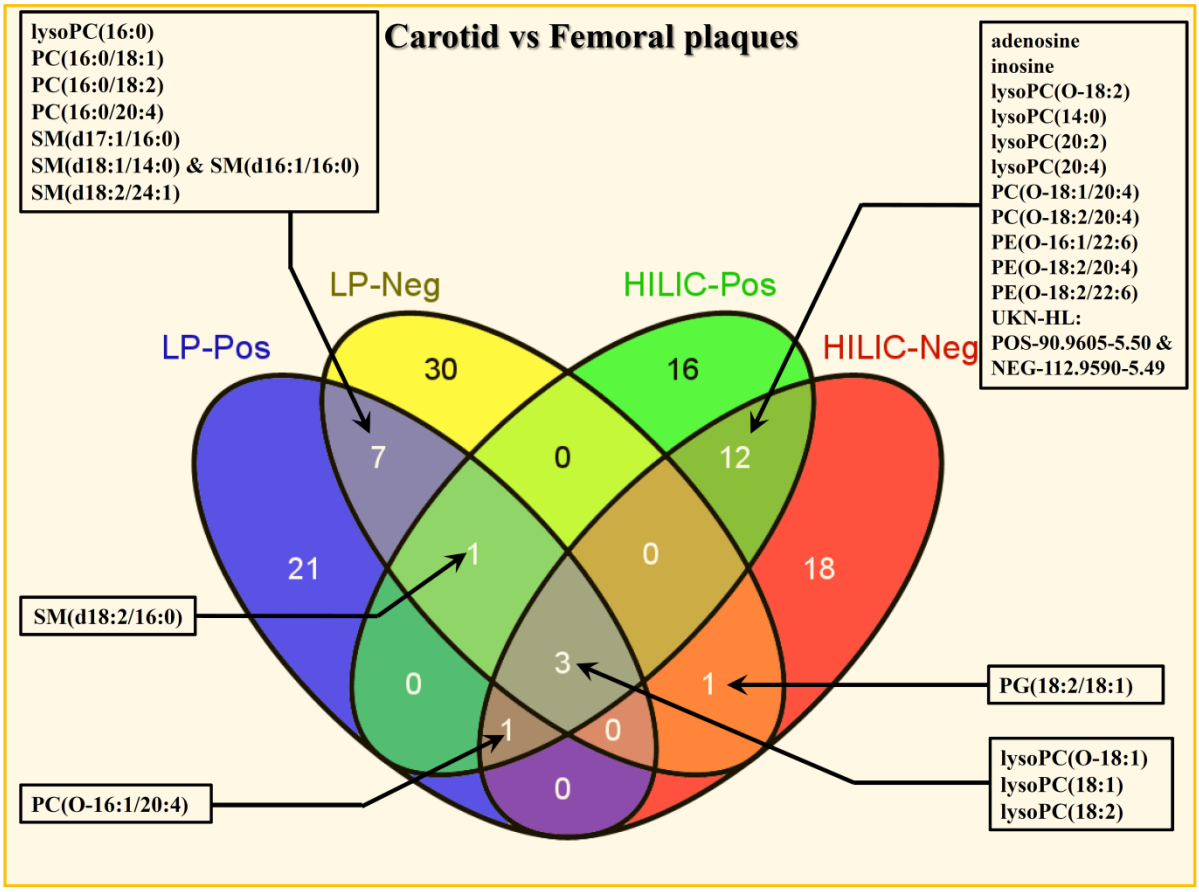


*Figure 4. 6: Venn diagram indicating identical metabolites detected statistically significant in the comparisons between carotid plaques and femoral plaque samples. Metabolites detected with the same trend in more than one analysis and/or polarity mode are shown. Tentatively assigned metabolites are indicated with an asterisk.*

Statistically significant metabolites detected in both CAR and FEM after individual comparisons to control samples are shown in Figure 4. 7. These metabolites were detected following the same trends in both disease locations. Common perturbed metabolites can provide a good overview of the dysregulated metabolic procedures of plaque formation and will be the main focus of analysis and discussion in the following paragraphs.

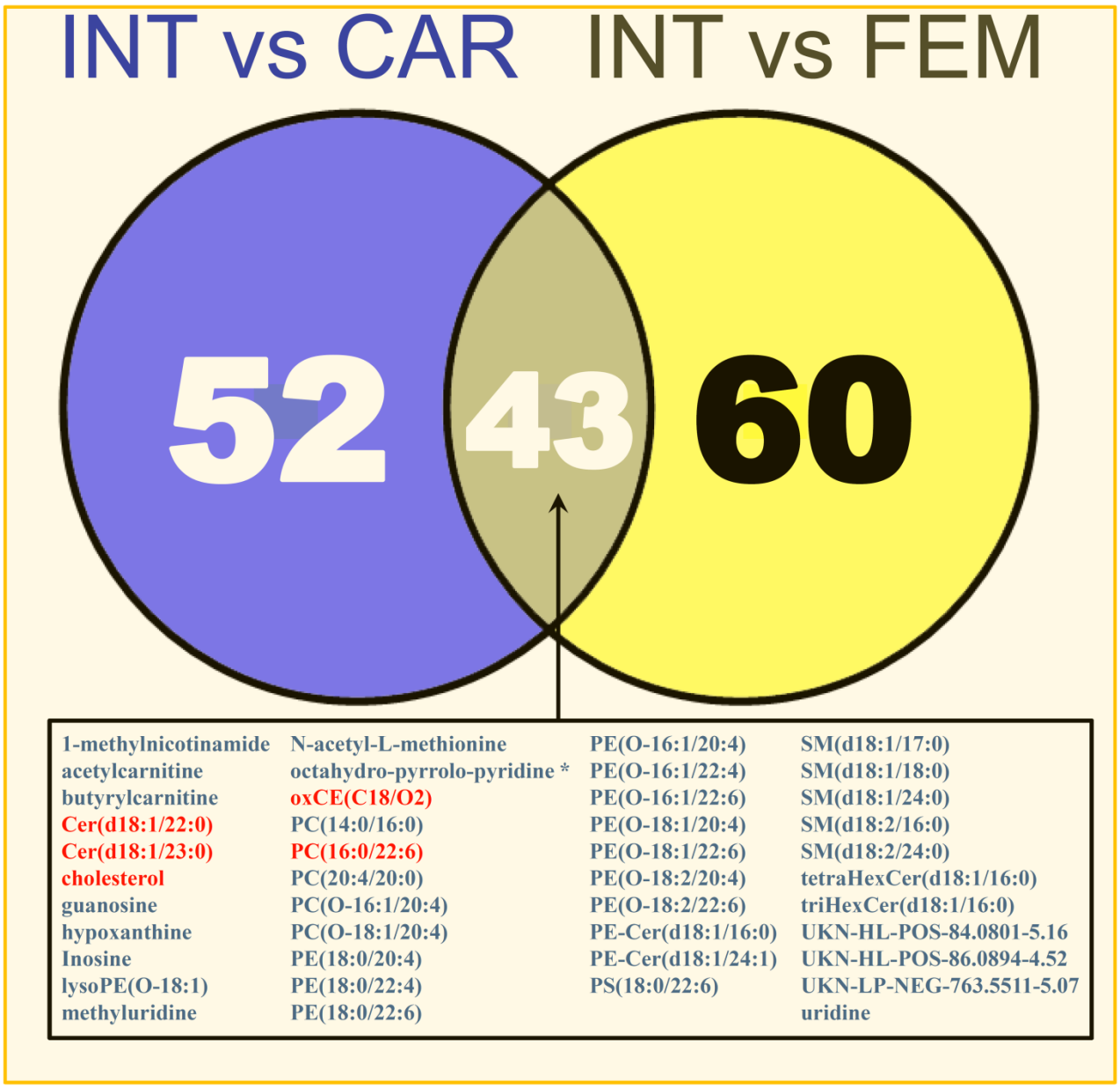


*Figure 4. 7: Venn diagram indicating identical metabolites detected statistically significant in both diseased anatomical locations (carotid and femoral) after separate comparisons to intimal thickening controls. Metabolites were detected with the same trends. Metabolites in red are the ones detected in higher intensities in disease and in blue are detected in lower. Tentatively assigned metabolites are indicated with an asterisk.*

### *4.3.2 Distinct Metabolic Profiles Detected Between Plaques but Not Between Intima Thickening from Different Anatomical Locations*

MVDA detected distinct metabolic profiles between CAR and FEM. OPLS-DA (Figure 4. 3) showed high predictive values in all analyses and polarities (Table 4. 4). Some of the most comprehensive findings were the several TG moieties, mostly incorporating 16 and 18 carbon chains, being significantly higher in FEM plaques. A number of lysoPCs also showed differential

levels with high statistical significance, reaching t-test p-values of 3.5E-16, with higher intensity levels in CAR (Table 4. 6). Other highly significant metabolites include FFAs, adenosine and PGs.

In order for these findings to be attributed to differences of disease progression, rather than anatomical location, a comparison was conducted to determine whether INT from these locations displayed metabolically related differences, prior to plaque manifestation. Efforts to fit multivariate statistical models using the two groups (INT groups of CAR and FEM origin) fell short, as models were unable to be statistically validated. This gave an initial proof of the two groups bearing no differences in the control stage. Additionally, for every univariate comparison between features showing statistical significance of the three groups, a t-test between the two anatomical sites of intimal thickening was also conducted. Out of 292 t-tests conducted in all analyses with p-values <0.01 only 6 t-tests showed p-values of <0.01 (but always larger than 0.001) when comparing the corresponding INT from different anatomical locations. Moreover, for the analysis comparing between FEM and CAR groups, where metabolic differences between the two locations could directly be detected, only 3 metabolites showed a p-value of <0.01 (but larger than 0.001) (Figure 4. 8 and Table 4. 5), as compared with the p-values of t-test comparisons of plaques from the two anatomical locations.

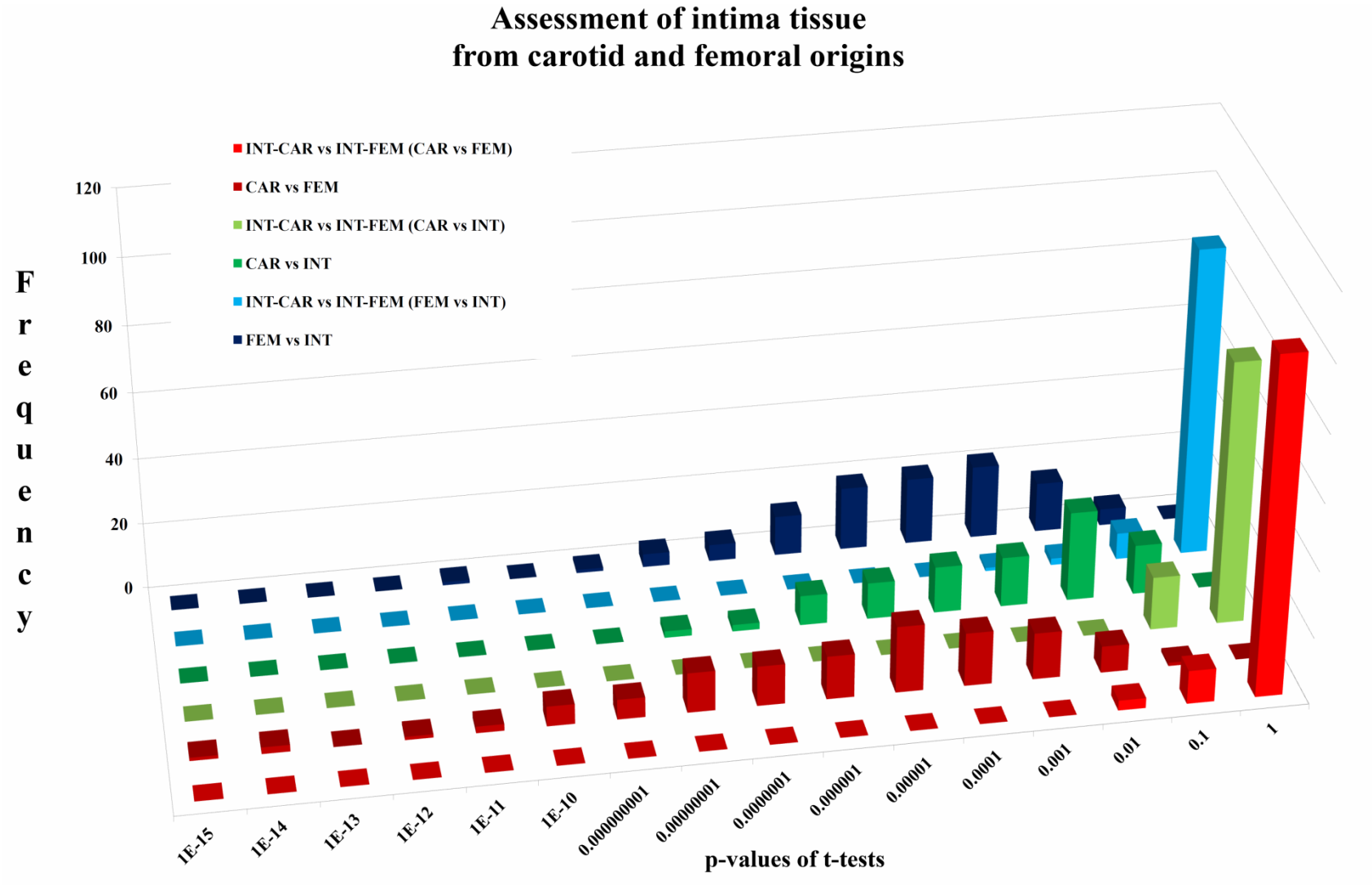


*Figure 4. 8: Assessment of the difference of intimal thickening tissue from the carotid and femoral locations. For every metabolite detected statistically significant, a corresponding comparison was conducted between intimal thickening tissues from the two anatomical locations using the t-test. P-values for all comparisons are indicated.*

*Table 4. 5: Assessment of the difference of intimal thickening tissue from the carotid and femoral locations. For every metabolite detected statistically significant, a corresponding comparison was conducted between intimal thickening tissues from the two anatomical locations using the t-test. Frequencies of p-values for all t-test comparisons are indicated.*

| T-test p-value frequencies | INT-CAR vs INT-FEM (CAR vs FEM comparison) | CAR vs FEM comparison | INT-CAR vs INT-FEM (CAR vs INT comparison) | CAR vs INT comparison | INT-CAR vs INT-FEM (FEM vs INT comparison) | FEM vs INT comparison |
|---|---|---|---|---|---|---|
| 1E-15 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1E-14 | 0 | 2 | 0 | 0 | 0 | 0 |
| 1E-13 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1E-12 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1E-11 | 0 | 2 | 0 | 0 | 0 | 1 |
| 1E-10 | 0 | 6 | 0 | 0 | 0 | 0 |
| 1E-09 | 0 | 6 | 0 | 0 | 0 | 1 |
| 1E-08 | 0 | 12 | 0 | 2 | 0 | 4 |
| 1E-07 | 0 | 12 | 0 | 2 | 0 | 5 |
| 0.000001 | 0 | 13 | 0 | 9 | 0 | 12 |
| 0.00001 | 0 | 20 | 0 | 11 | 0 | 19 |
| 0.0001 | 0 | 16 | 0 | 14 | 0 | 20 |
| 0.001 | 0 | 14 | 0 | 15 | 1 | 22 |
| 0.01 | 3 | 8 | 0 | 27 | 2 | 15 |
| 0.1 | 10 | 1 | 16 | 15 | 8 | 5 |
| 1 | 101 | 0 | 79 | 0 | 93 | 0 |

Studies describing differences between CAR and FEM are limited in literature. A study by *Bianda et al* describes differences on plaque remodelling[145]. Additionally, a study by *Herisson et al.* demonstrates differences between CAR and FEM plaques, with FEM described as being more prevalent to ectopic calcification[146]. Plaque calcification is considered a progression of atherosclerosis[147] and this may also be the explanation of the distinct and distant metabolic profiles of FEM from CAR and INT. Evidence of $Ca^{2+}$ dysregulation was supported by findings of the present study. For FEM vs INT the levels of PI(18:0/20:4) were detected being significantly lower in plaques with a fold change of 2.5, while in the comparison of CAR vs FEM, PI(18:0/22:6) was also lower with 2.9 fold-change. Phosphatidylinositols (PIs) are precursor molecules of inositol-triphosphate ($IP_3$), known for its role in calcium homeostasis. Increase of $IP_3$ can induce export of $Ca^{2+}$ from the endoplasmic reticulum [43]. Haemodynamic differences are more likely to be responsible for this disease variation observed along atheroma manifestation. In the following paragraphs discussion on the significantly dysregulated pathways is expanded, and specific metabolite differences between groups are discussed further.

### 4.3.3 Presence of cholesterol and oxidised cholesterol esters in atheromas

Cholesterol and cholesterol derivatives are nowadays considered well-established risk factors of atheromatogenesis[148]. In this study free cholesterol (Cho) was detected elevated in both CAR and FEM as compared to INT. Moreover several oxidized cholesterol ester (oxCE) moieties were detected in higher levels. The oxCEs were detected with higher statistical significance and fold-change than Cho. Additional involvement of the cholesterol pathway is shown in FEM, where cholesterol sulphate was also found to be higher by 2-fold. Although the detection of higher levels of Cho and derivatives does not constitute a novel finding, it does provide good validation of the experimental design and findings.

All oxCEs detected in higher levels were esterified only with 18C fatty acyl chains. These chains were detected with 1-3 oxygen molecules. Free radicals have been suggested to promote the production of oxCE[149, 150]. The oxCEs are also known to contribute to foam cell formation[151], which are the cells forming the lipid-laden lesion. Lipid peroxidation is caused by ROS and affects unsaturated fatty acids as well as cholesterol esters (CE) and Cho itself[150]. It can eventually lead to a chain reaction and result to cell membrane damage[149]. Lipid oxidation is also recognized to play a role in the progression of atherogenesis. Structures of detected oxCEs moieties are in concordance with literature[152].

### 4.3.4 Purine and pyrimidine pathway dysregulation

Purines and pyrimidines have been known for their involvement in atherosclerosis for many years[153]. Triphosphates of adenosine (ATP) and uridine (UTP) are released from endothelial cells in response to sheer stress[154]. They are controlling vascular tone[154,155], a process related to hypoxia. A number of them, such as adenosine and inosine, are known for their anti-inflammatory effects[156, 157]. Purine and pyrimidine inhibition can lead to apoptosis[158], while on the other hand they are essential for cell proliferation[158]. Metabolites of purine and pyrimidine pathways are components or precursors of RNA and DNA. Herein, a large number of molecules of these two pathways were detected with lower intensities in both disease groups. These include uridine, inosine, hypoxanthine, guanosine and methyluridine (Table 4. 6 and Table 4. 7). These results provide yet another form of validation to the current study, since they are well-established to be involved in atherosclerosis.

The elevation of adenosine levels only in FEM becomes rather intriguing as it was not detected with statistical significance in CAR. Adenosine, apart from its anti-inflammatory abilities, is acting as a highly potent vasodilator and vasoconstrictor[24, 155]. This process is also connected to $Ca^{2+}$ channels, and could explain the high correlation (r=-0.57) of adenosine to PI(18:0/20:4), also known for their

involvement in $Ca^{2+}$ homeostasis. The differential amount of shear stress and stretch applied to these different anatomical locations may provide insights into this detected difference. It can be also claimed that the lower and higher levels of inosine and adenosine in FEM, respectively, are to some extent independent to their directly connecting enzyme, adenosine deaminase, known to be present in vessels. This claim is based on the inverse correlation between these two metabolites (r=-0.56), which reflects some sort of association, but the correlation is not as strong as would be expected for a direct reaction. Adenosine was detected with high fold change in FEM in both positive and negative polarity modes of the HILIC analysis.

### *4.3.5 Ceramide pathway suffers complete homeostatic loss*

The ceramide (Cer) pathway is known for its involvement in atherosclerosis [159]. This can be expected due to the role of the Cer pathway in apoptosis[160], and connections to (pro-) inflammatory factors, reactive oxygen species (ROS) and nitric oxide (NO)[159]. In the present study, a global reduction of the levels of a number of sphingomyelin (SM) moieties was detected. This observation along with the detected elevated levels of Cers may indicate the involvement of the Sphingomyelinase (SMase) enzyme. SMase catalyses the production of Cer by hydrolysis of the phosphoesteric bond of the phosphocholine head-group of SMs.

Glycosphingolipids, another sphingolipid subclass were also found to be dysregulated. From literature review, studies detecting glycosphingolipid involvement in atherogenesis go 40 years back[161]. They have been implicated as signalling molecules in cell proliferation after oxidative conditions in vascular cells[162] and platelet activation and adhesion to the vessel wall[163]. In the present study, tetra- and tri- hexosylceramide (HexCer) were detected in reduced levels, both in the form of d18:1/16:0. Specifically, tetraHexCer was detected with high p(corr) values in MVDA in both diseases (p(corr)= -0.82 CAR; p(corr)= -0.89 FEM) (Figure 4. 9). Findings indicate a more intense involvement of tetraHexCers, and these being highly correlated to triHexCer (r=0.81 CAR; r=0.84 FEM). On the other hand, two monoHexCers were detected with higher intensities and only in CAR. The dysregulation of monoHexCers with an opposite trend and only in CAR indicates involvement of yet another biochemical reaction independent to tri- or tetraHexCer. This is also designated by lack of correlation to tetra- and tri-HexCer, while strong correlations were detected to Cers.

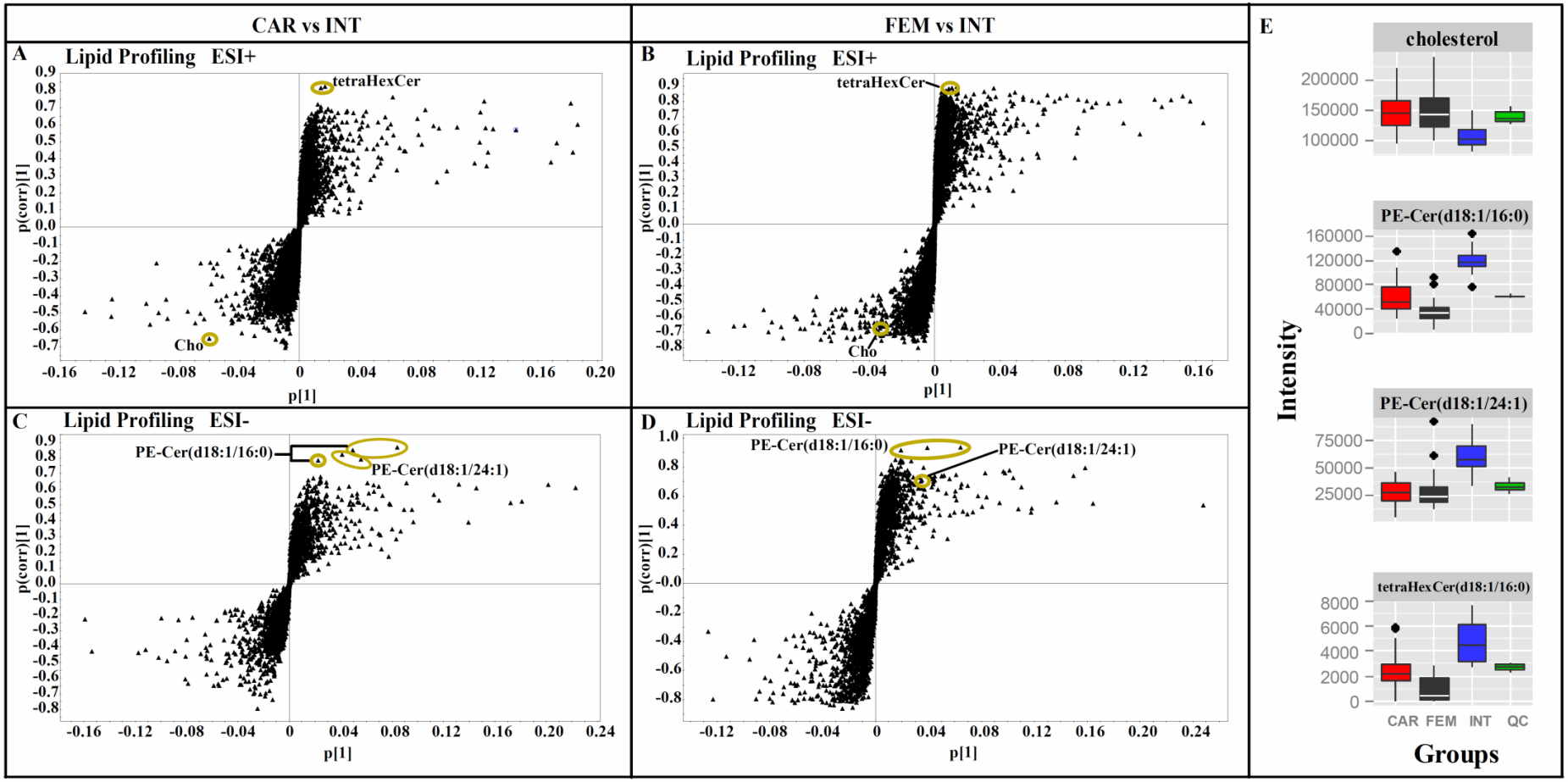


*Figure 4. 9: A-D S-plots of the OPLS-DA analyses of lipid profiling analysis of tissue extracts of (A) positive mode of carotid against intimal thickening, (B) positive mode of femoral against intimal thickening, (C) negative mode of carotid against intimal thickening, and (D) negative mode of femoral against intimal thickening tissue extracts. (E) box-plots of metabolites indicated on S-plots that demonstrated high statistical significance in both statistical comparisons.*

Other differential sphingolipids were sphingosine and C16-sphingosine, showing a 3.5- and 7.6-fold change, only in CAR, and Phosphatidylethanolamine-Ceramides (PE-Cer) with lower levels in both anatomical locations, compared to INT. PE-Cer involvement is considered to be a novel finding of this study, crucial to the pathological process, and it will be further discussed in the following paragraphs.

### *4.3.6 PE-Cer a new candidate biomarker in atherogenesis*

Despite the vast amount of research focused on atherosclerosis PE-Cers have not been investigated in the context of plaque formation. Specifically, apart from p-values as low as $9.8x10^{-12}$ in t-tests (Table 4. 6 and Table 4. 7), PE-Cers showed indications of high correlation to the disease (p(corr)≥-0.94) (Figure 4. 9), using well-validated multivariate models, with high predictive values. They were detected in the form of PE-Cer(d18:1/16:0) and PE-Cer(d18:1/24:1). PE-Cer involvement in the progression of plaque formation was evident in both anatomical locations (CAR and FEM), and is a significant finding as it demonstrates common basis of disease manifestation. PE-Cers are found only in trace concentrations in mammalian cells[164]. However, this lipid has been reported to be essential to ceramide homeostasis in humans[165].

A selective synthase called sphingomyelinase synthase1 – related (SMSr), is the enzyme responsible for PE-Cer synthesis by transferring the phosphoethanolamine group from PE to Cer. Therefore, the inverse correlation of PE-Cer to Cer is in concordance with the corresponding biochemical reaction[24]. However, PEs did not share the same trend as Cer. On the contrary, PEs had a high positive correlation to PE-Cers. This correlation should be further investigated in the context of the possibility of a positive feedback loop with additional biological effects. For CAR the highest correlation occurred with 20:4 FAC bearing PEs, while for FEM with 22:6. PE-Cers also showed high correlations with members of the pyridine and pyrimidine pathways, such as inosine, uridine and guanosine.

As expected, correlation networks revealed connections between PE-Cers and metabolites, members, of the ceramide pathway. These included high positive correlations to tetraHexCer, triHexCer, and SMs and inverse correlations to Cers and monoHexCers. An interesting observation is that correlations to tetraHexCer were generally higher than triHexCer, and in the same level as Cers. This association should be further explored since Cers and triHexCer would represent intermediate products of the reaction from PE-Cer to tetraHexCer, according to reference pathways.

#### *4.3.6.1 Structural assignment of PE-Cers*

Despite the fact that PE-Cers are known to exist in the human lipidome, they are not ubiquitously found in MS databases[111, 116]. PE-Cers were detected as statistically significant with the negative mode of the lipid profiling methodology. A search for the PE-Cer(d18:1/24:1) (m/z on neg=769.6202) returned no results while for the PE-Cer(d18:1/16:0) ) (m/z on neg=659.5125) two PE-Cers were returned (as [M-H]-) with however different fatty acyl chains (Figure 4. 10) and reported as previously detected only in *Drosophila melanogaster*[166]. However, due to the high mass accuracy (Δppm=0) to the detected ion this possibility was further explored.

A first indication was the low intensity of the metabolite. As mentioned, PE-Cers are detected in trace concentrations in humans. Positive mode was used and an ion representing the [M+H]+ was detected at the same retention time. The $MS^E$ mode was then utilised to confirm characteristic fragments of PE-Cers (these ions were not picked from DDA for MS/MS due to their low intensities), in combination with in source fragments. These would be a neutral loss of 141 (phosphoethanolamine loss), and a 264 (backbone of d18:1 ceramide)[166]. These ions were detected and matched relatively well to the chromatographic peak profile, considering the low intensity of the ions (Figure 4. 11).

Additional MS/MS experiments were conducted on positive mode, using samples with high intensities of this signal and further verified the metabolites to be as such based on their fragmentation profile (Figure 4. 12 and Figure 4. 13). Characteristic fragments that led to the structural assignment of these PE-Cers were: 1) the neutral loss of phosphoethanolamine (141), 2) the backbone fragment of d18:1 (m/z = 264) and 3) the fatty amides corresponding to the specific metabolites (m/z = 280 and 390).

LIPID Metabolites and Pathways Strategy (LIPID MAPS)

# Lipidomics Gateway

Search the Lipidomics Gateway | Search

Home | Lipidomics Update | Resources | Meetings | Tutorials | Protocols | About

Overview | Publications | Classification | Standards | Data | Databases | Pathways | Tools | Services | Links | Downloads

Possible Lipid Structures

**Note**: Chains containing double bonds and/or functional groups with defined regiochemistry, geometry and stereochemistry are meant to serve as examples for structure-drawing purposes. In many cases there may be alternative isobaric structures.

| LM_ID | Name | Systematic Name | Input m/z | Exact m/z | Formula | Ion |
|---|---|---|---|---|---|---|
| LMSP03020005 | PE-Cer(d14:1(4E)/20:0) | N-(eicosanoyl)-tetradecasphing-4-enine-1-phosphoethanolamine | 659.5125 | 659.5134 | $C_{36}H_{72}N_2O_6P$ | M-H |
| LMSP03020020 | PE-Cer(d16:1(4E)/18:0) | N-(octadecanoyl)-hexadecasphing-4-enine-1-phosphoethanolamine | 659.5125 | 659.5134 | $C_{36}H_{72}N_2O_6P$ | M-H |

LIPID MAPS Lipidomics Gateway | Terms of use | Site map | Contact us | LIPID MAPS Personnel login

NIGMS

Follow @lipidmaps

*Figure 4. 10: Results of accurate mass search for 659.5125 m/z, detected on negative mode of the lipid profiling analysis, in the lipid maps[139] database. (Search conducted March 2013)*

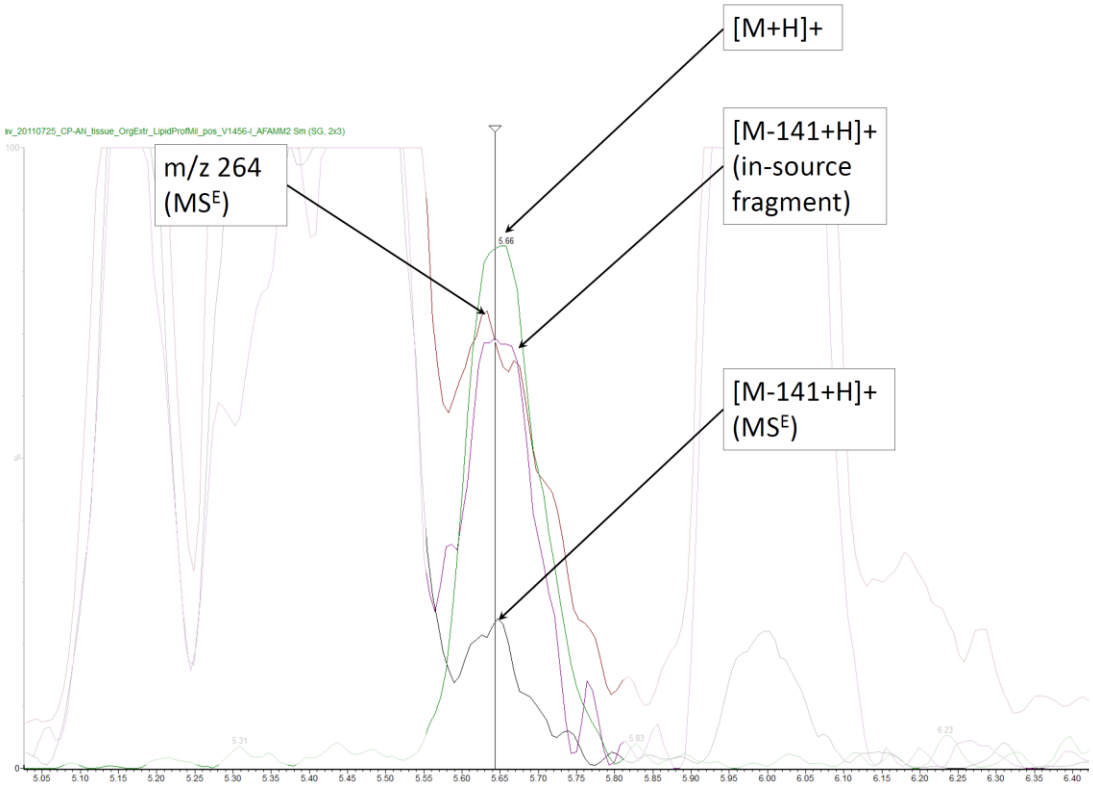


*Figure 4. 11: Extracted ion chromatograms of characteristic fragments of metabolite that was initially tentatively assigned as PE-Cer(d18:1/16:0), and further verified as such by MS/MS experiments. Extracted ion chromatograms were smoothed using the Savitzky-Golay method.*

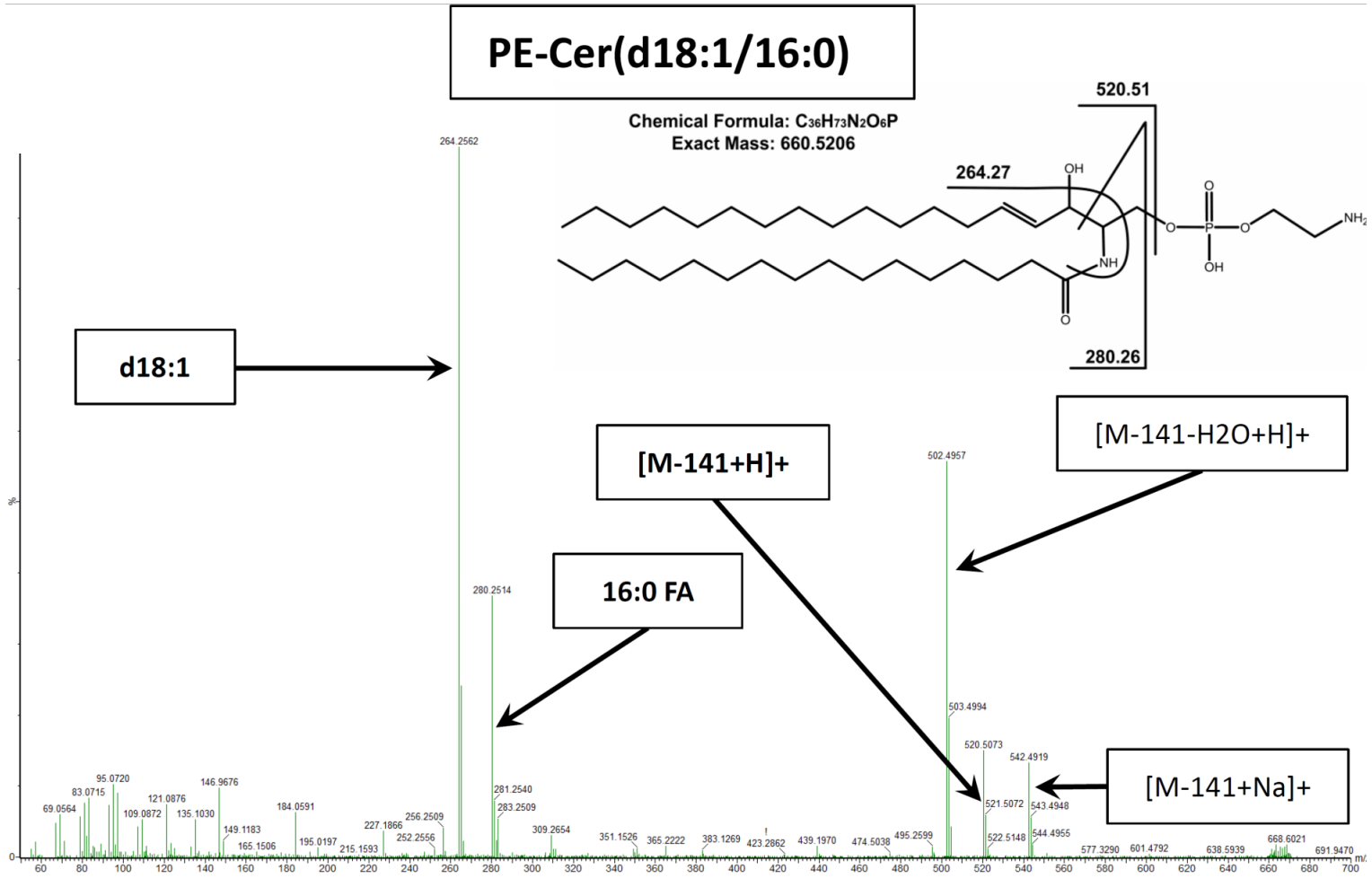


*Figure 4. 12: MS/MS spectrum of the PE-Cer(d18:1/16:0) indicating the characteristic fragments used for structural assignment.*

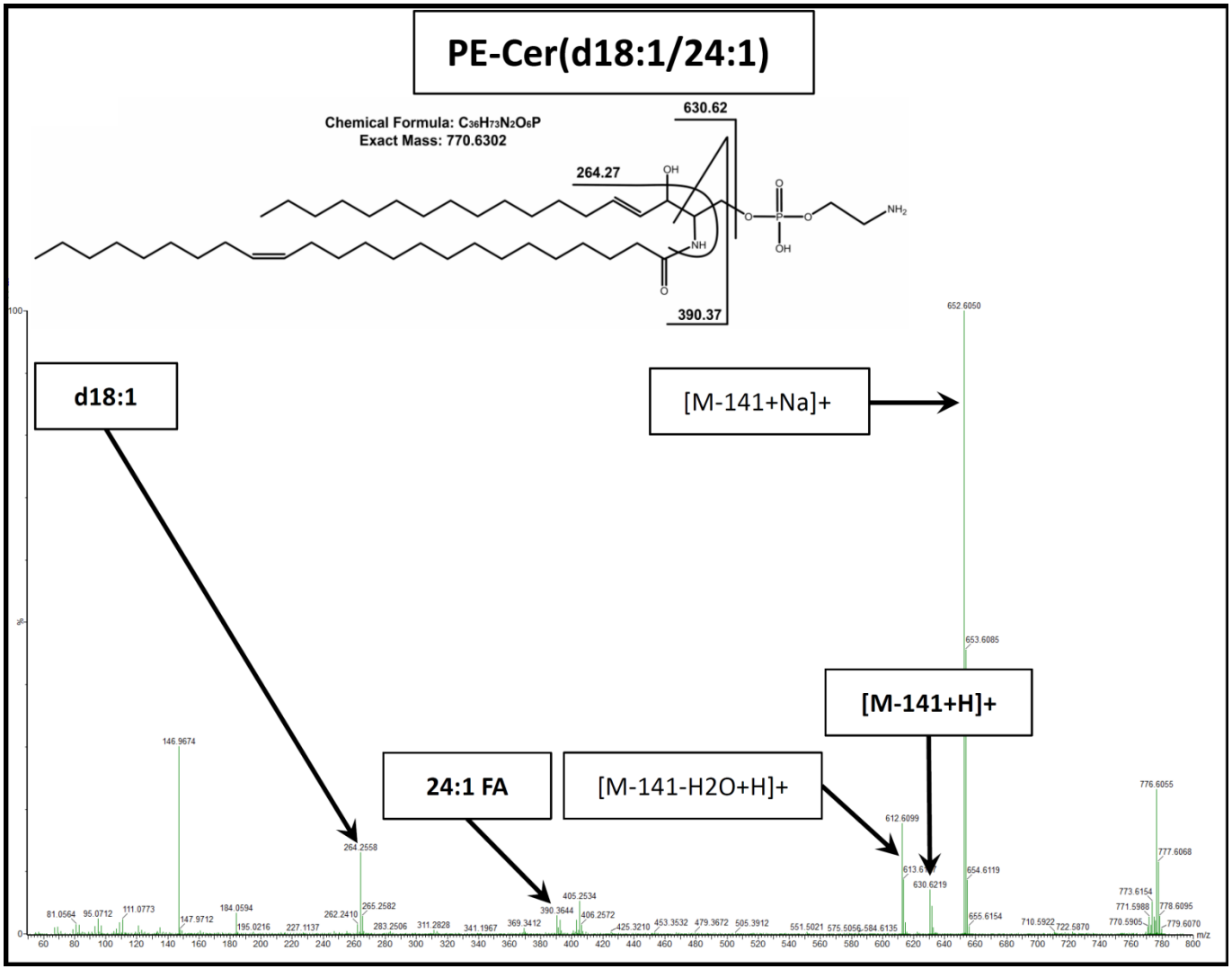


*Figure 4. 13: MS/MS spectrum of the PE-Cer(d18:1/24:1) indicating the characteristic fragments used for structural assignment.*

### *4.3.7 PE-Cer levels are inversely correlated to cholesterol*

Probably the most biologically important observation is the high inverse correlation of PE-Cers to Cho and subsequently to the detected oxCEs. PE-Cer(d18:1/16:0) for the CAR group achieved correlation values of r= -0.84 and for the FEM group r= -0.86, while PE-Cer(d18:1/24:1) r=-0.56 and -0.81, respectively (Figure 4. 14, Figure 4. 15, Figure 4. 16, Figure 4. 17). Although Cho did appear to have high correlation to some of the members of the Cer pathway, they were lower compared to PE-Cer indicating more of an effect rather than a cause. This was further verified by correlation analysis for metabolites found dysregulated in both groups (n=94) (Figure 4. 18). As can be seen by the correlation network (Figure 4. 19), cholesterol was highly correlated only to PE-Cer(d18:1/16:0). From these findings it is hypothesised that PE-Cers and most likely only PE-Cer(d18:1/16:0) is the factor of the Cer pathway having the closest interaction to Cho.

It has previously shown that SMs can form Cho-enriched domains in the cell membrane, while PE-Cers do not favour formation of such domains[167, 168]. This observation may not provide an adequate explanation for the high absolute value of correlation between PE-Cers and Cho, but it may explain the inverse nature of the correlation. Additionally, the fact that PE-Cers have been found to be responsible for controlling Cer levels[165] can add confidence to the hypothesis.

As mentioned, the Cer pathway is known for its involvement in atherosclerosis[159]. Thus, efforts to connect the Cer pathway to Cho have previously been conducted[169, 170] since Cho represents a traditional risk-factor for atherosclerosis, and associations of metabolites of the Cer pathway to Cho have previously been shown[171]. Connections between the Cer pathway and Cho have been the centre of attention in the manifestation of diseases such as Alzheimer[172] and cancer[173]. Therefore, PE-Cers carry the potential of being the missing link, and thereby bringing the quest of integrating Cho and Cer pathways, and consequently the elucidation of the pathophysiology of plaque formation, a step closer to a resolution.

### *4.3.8 Dysregulation of metabolic oxidation is detected via truncation of β-oxidation and unsaturated lipid consumption*

In both CAR and FEM tissue, acyl-carnitines (AcC) with short acyl chains were detected with lower intensities, while, on the contrary, AcCs with medium or long chain AcC levels were increased. AcC are involved in the translocation of fatty acyl chains through the mitochondrial membrane. This finding implicates mitochondrial metabolism and specifically the β-oxidation. From these

results, it could be hypothesised that β-oxidation is somewhat truncated reducing the production of short-chain AcC, while at the same time medium- and long-chain AcCs are accumulated. Additionally, a number of highly unsaturated lipids were detected in lower intensities in the disease groups. As discussed in preceding paragraphs unsaturated fatty acyl chains can react under oxidative stress with their oxidation occurring in the peroxisome rather than the mitochondria. Highly unsaturated lipids were detected with inverse correlations to oxCEs (CAR r>-0.70; FEM r>-0.78) (Figure 4. 14 and Figure 4. 16). *Yanes et al* demonstrated that high concentrations of long-chain AcCs along with low concentrations of unsaturated lipids are present after maturation of embryonic stem cells[174]. In this context, processes imitating embryonic differentiation, such as ectopic ossification[147], are well known to take place in advanced atherosclerotic lesions. Findings also support this further, since FEM showed more intense dysregulation of β-oxidation. FEM have been previously shown to experience more intense calcification than CAR[146]. From literature, β-oxidation and oxidative conditions, in general, are linked to inflammation[175]. Although oxidative stress is known to be involved in the manifestation of atherosclerosis[148], to our knowledge, β-oxidation and acyl-carnitines have not been previously associated with atherogenesis.

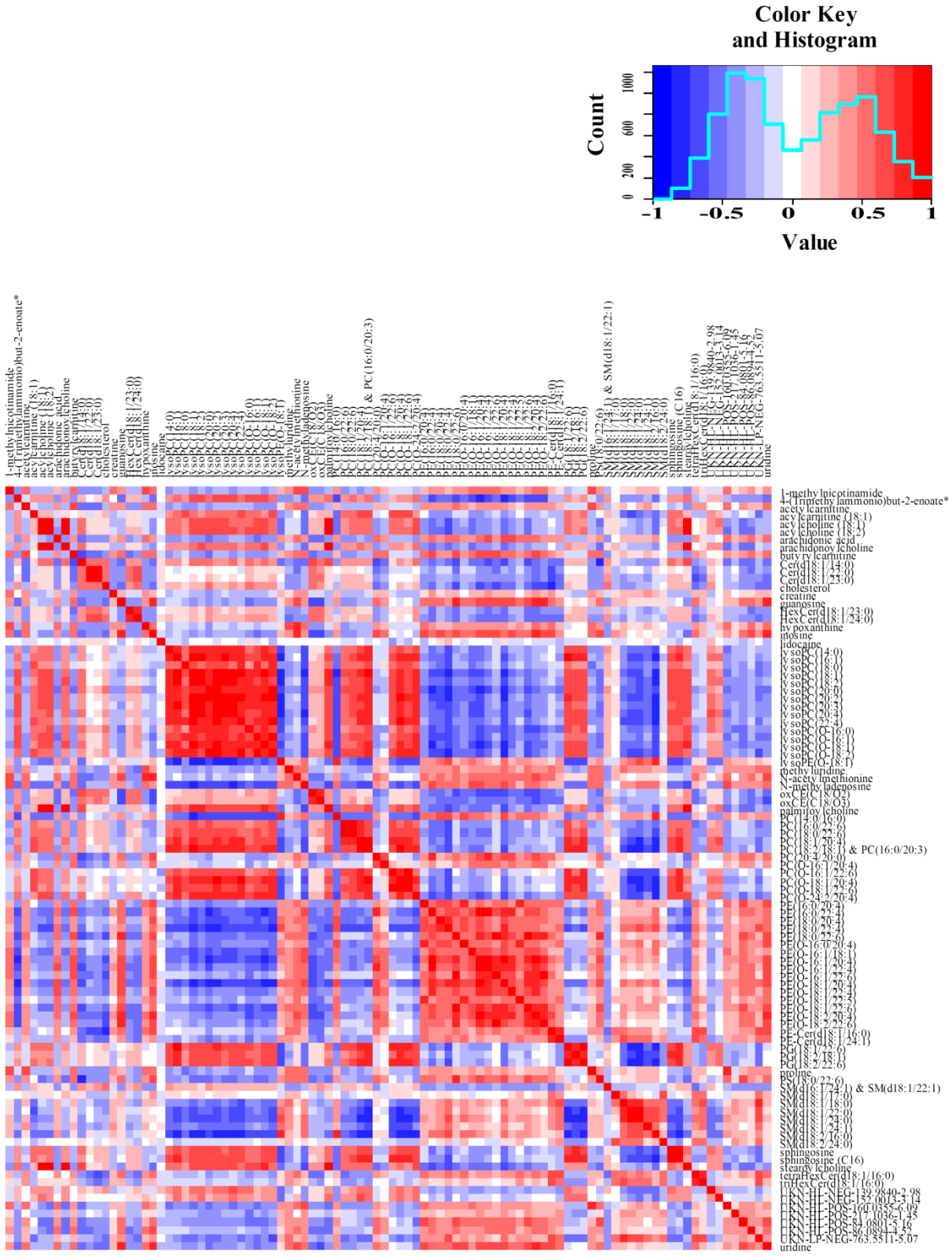


*Figure 4. 14: A heatmap representation of the correlation matrix obtained from Spearman correlations of metabolites detected statistically significant in the comparison of carotid plaques to intimal thickening. Intensity of red colour represents positive correlation and intensity of blue represents negative correlation of the biomarker pair.*

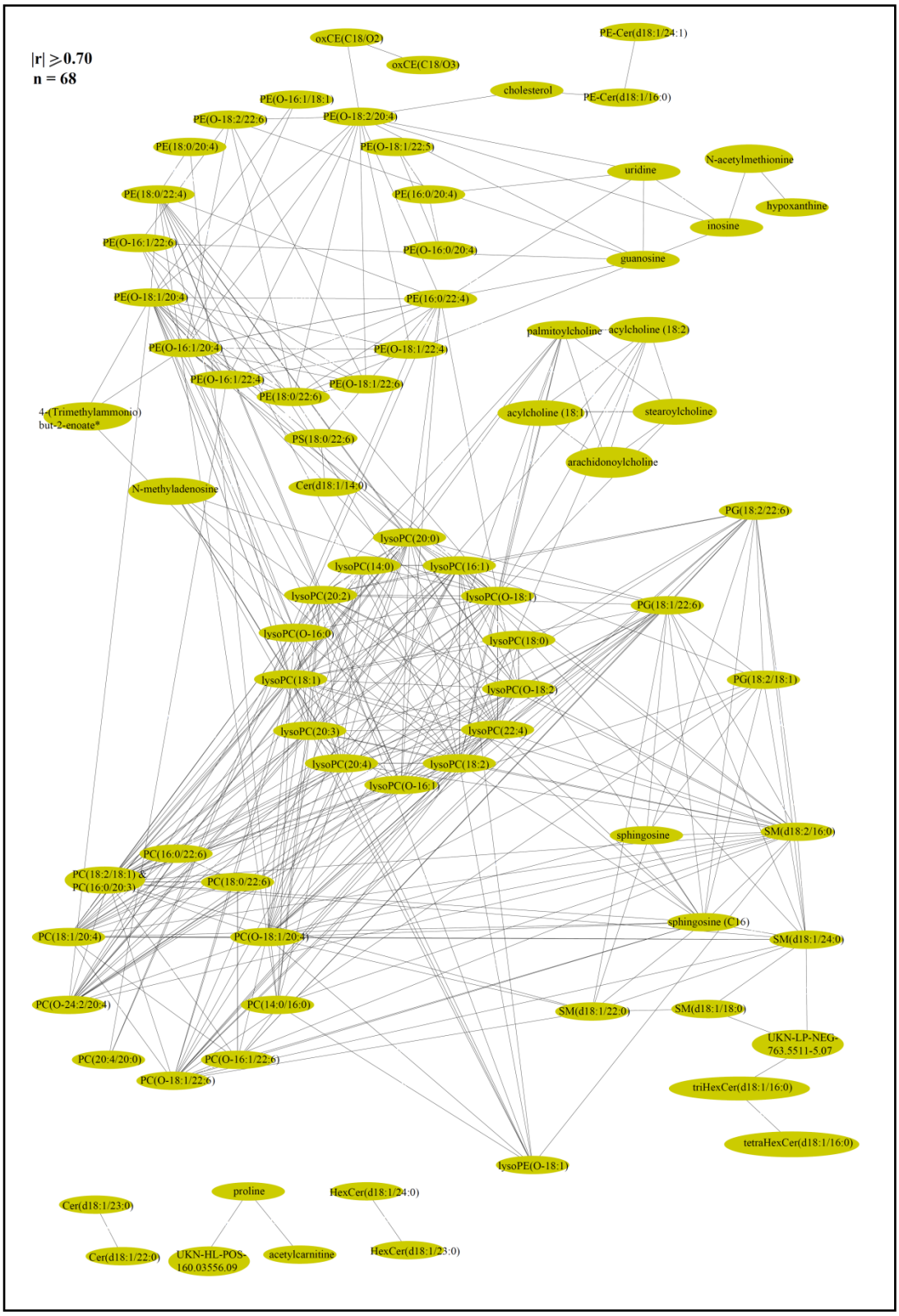


*Figure 4. 15: Correlation network of metabolite pairs found to have a Spearman correlation value of more than 0.70 in absolute value. These metabolites were found to be statistically significant in the comparison of carotid to intimal thickening tissue.*

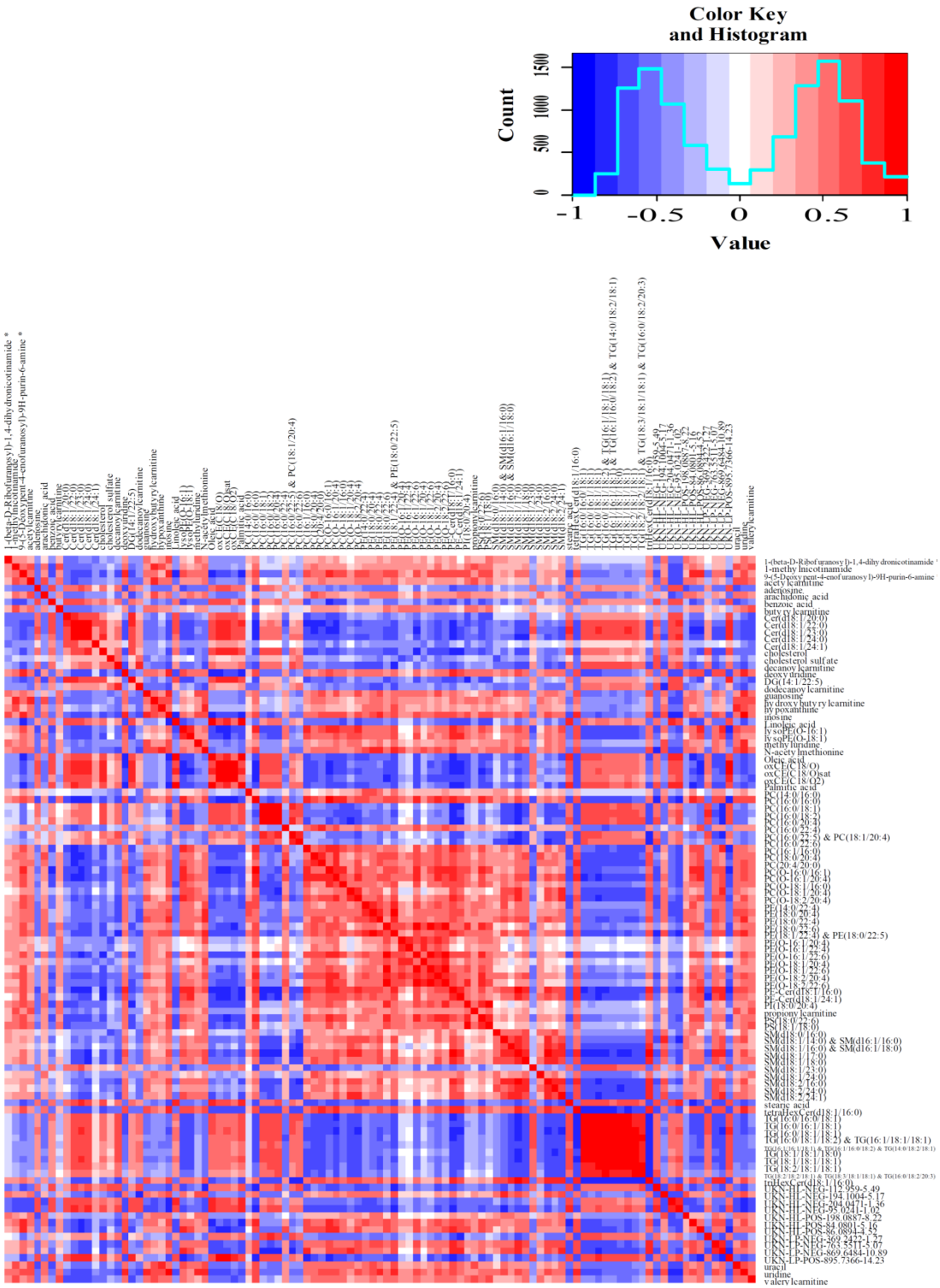


*Figure 4. 16: A heatmap representation of the correlation matrix obtained from Spearman correlations of metabolites detected statistically significant in the comparison of femoral plaques to intimal thickening. Intensity of red colour represents positive correlation and intensity of blue represents negative correlation of the biomarker pair.*

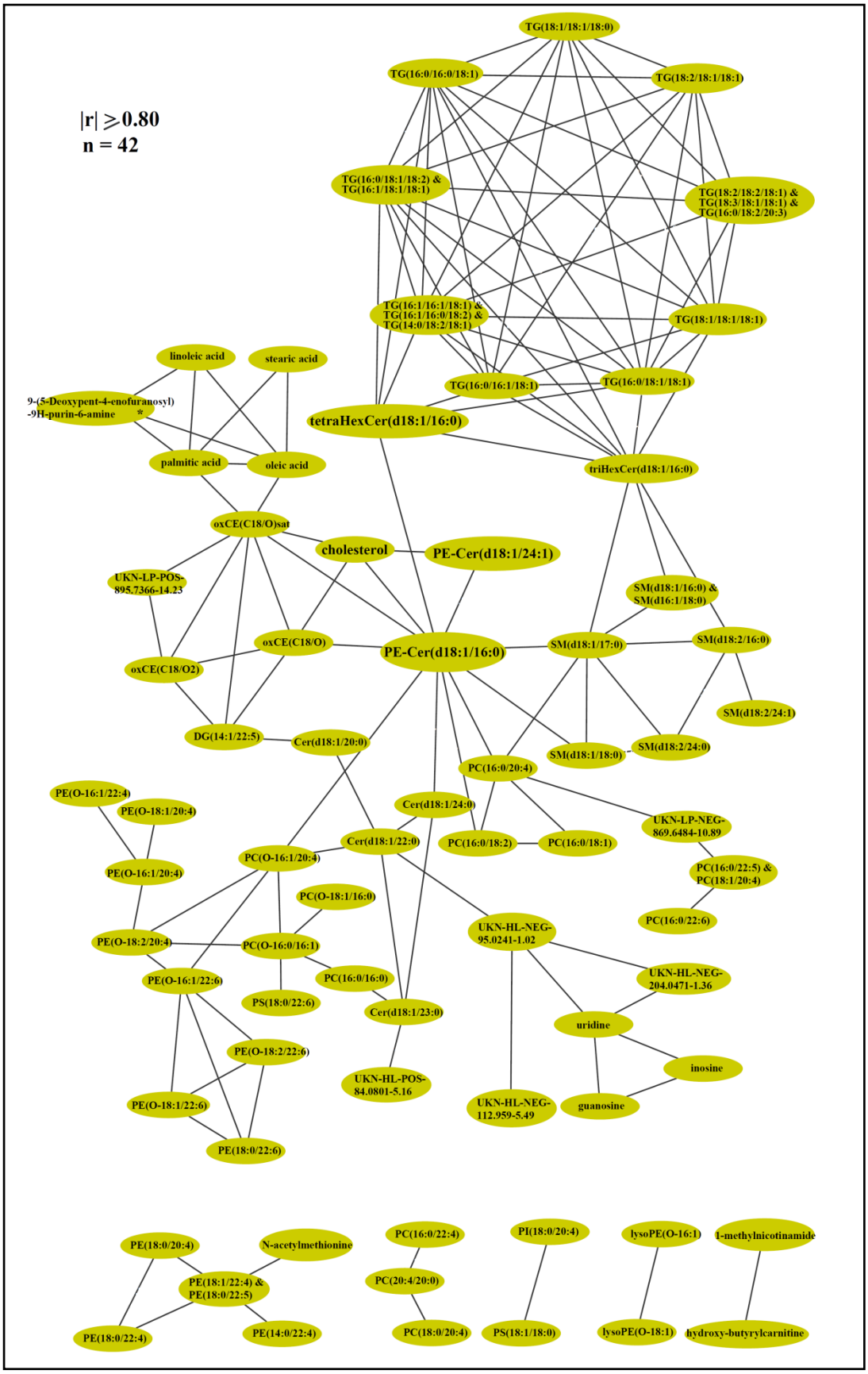


*Figure 4. 17: Correlation network of metabolite pairs found to have a Spearman correlation value of more than 0.80 in absolute value. These metabolites were found statistically significant in the comparison of femoral to intimal thickening tissue.*

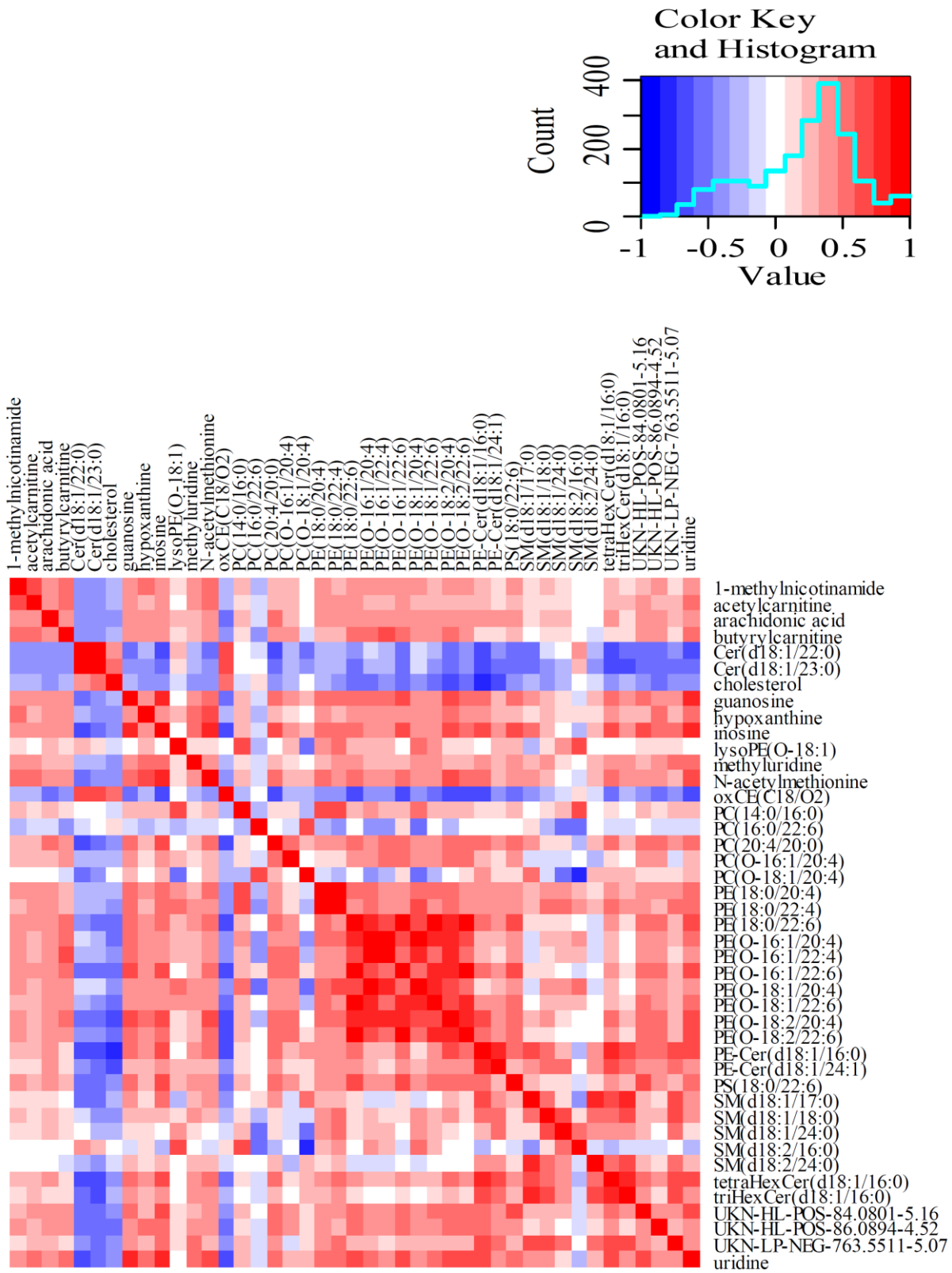


*Figure 4. 18: A heatmap representation of the correlation matrix obtained from Spearman correlations of the common metabolites detected statistically significant in separate comparisons of carotid and femoral plaques to intimal thickening. Intensity of red colour represents positive correlation and intensity of blue represents negative correlation of the biomarker pair.*

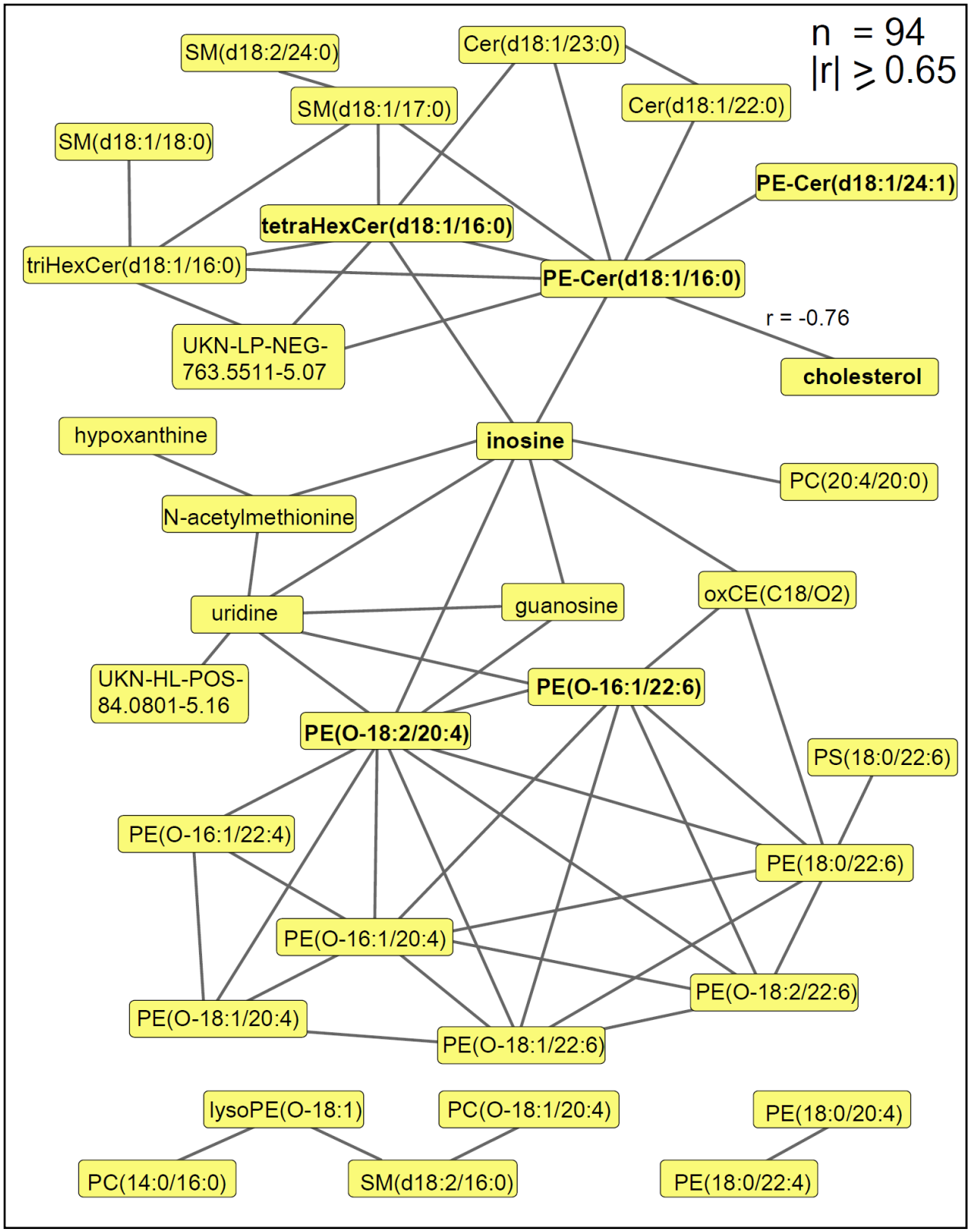


*Figure 4. 19: (A) Correlation network of metabolite pairs found to have a Spearman correlation value of more than 0.65 in absolute value. These metabolites were commonly found as statistically significant in separate comparisons of carotid and femoral plaque tissue to intimal thickening. (B) Box plots of highlighted metabolites from the correlation network. These metabolites are considered important due to the central role (hubs) to the network or prior knowledge of importance in the studied disease, according to literature.*

In FEM, along with the increase of AcCs, accumulation of free fatty acids (FFA) is also detected and indicates low FFA oxidation as well as presence of oxidative conditions. Detected FFAs were in the range of 16-18 carbon chains, and can be only moderately associated with detected dysregulation of β-oxidation (r~ 0.4 - 0.55 to medium-chain AcCs; r~ -0.50 - -0.55 to short-chain AcCs). The accumulation of 16 and 18 carbon chains explains the high abundance of 16 and 18 carbon chain TGs (r<0.69). Correlations of FFAs in the range of 0.61 – 0.80 with the oxCE bearing one oxygen, and in the range of 0.51 – 0.59 to the oxCE bearing two oxygens demonstrates further their association to oxidative stress. Another observation is the very high inverse correlation of these 16 and 18C TGs to tetra- and tri-HexCer (r > -0.86). TGs also showed high correlations to Cers and high negative correlations to SMs and PE-Cers, with, however, lower absolute values. Accumulation of FFA has been shown to lead to toxic pathways and apoptosis in pancreatic β-cells[176]. The reasons of the FFA accumulating only in FEM could be sought to the different haemodynamics between FEM and CAR, since muscle contraction has been shown to induce FFA uptake[177].

### *4.3.9 Presence of acyl-cholines in CAR*

In the CAR higher intensities of acyl-cholines (choline esters) were observed. This was based on accurate m/z measurements and presence of an *N, N, N*-trimethyl loss on MS/MS analysis, which is characteristic of acetylcholine[178]. These esters incorporated fatty acyl chains of 16:0, 18:2, 18:1, 18:0 and 20:4. High pairwise correlations between these molecules (typically r>0.9) is a good indication of a direct connection, which may be a common synthesizing enzyme. They also manifested high correlation to lysoPCs and high inverse correlation to some of the PEs and lysoPEs.

Such a finding is reported for the first time. However, due to the fact that it represents an observation in only one of the groups it was further assessed. The major concern for this matter is that a local anaesthetic, lidocaine, is locally administered only during carotid endarterectomies. However, the detected molecules were not correlated to lidocaine (Figure 4. 14). Still, cholinesterases are known for their involvement in metabolising anaesthetics and therefore this finding should be further evaluated and validated[179].

## *4.4 Conclusions*

The choice of the two UPLC-MS methods used (Reversed-phase – UPLC for organic extracts and HILIC-UPLC for aqueous extracts) provided a wide coverage of metabolites, from lipids and extremely lipophilic molecules, such as TGs and CEs, to polar molecules such as carnitine and

creatine. A number of metabolites were detected as statistically significant from more than one analysis and polarity modes (Figure 4. 4). Good QC grouping in multivariate statistics, translated to low coefficient of variation (CV%), indicated a high reproducibility through the run, ensuring good quality data. The detection of metabolites known to exist in trace concentrations in human is another evidence of the impact new technological achievements have in the area of metabolic profiling.

The experimental design and especially the choice of intimal thickening tissue as the control group gives the confidence of detecting significant metabolites from pathological processes involved in plaque formation. The only drawback of the study was the usage of lidocaine, during surgery, only for carotid endarterectomies. This does not affect findings, as discussed, as the major focusing of the study is on metabolites common to both groups. Additionally, the drug, all the features structurally related to the drug, and a metabolite of the drug, were removed prior to statistical analysis. Moreover, lidocaine intensities were included in correlation analysis, as a further check, and showed no significant correlations to any of the statistically significant metabolites (Figure 4. 14). Keeping in mind that the drug is injected only a few minutes prior to excising the tissue, it is unlikely to cause any severe effect to the tissue biology. Specifically, no correlation between the bioavailability of the drug and tissue profile in CAR samples was observed in unsupervised (Figure 4. 20) or supervised MVDA. Lastly, the fact that the INT samples from the two locations did not show any differences further indicates that lidocaine administration does not induce major biological differences.

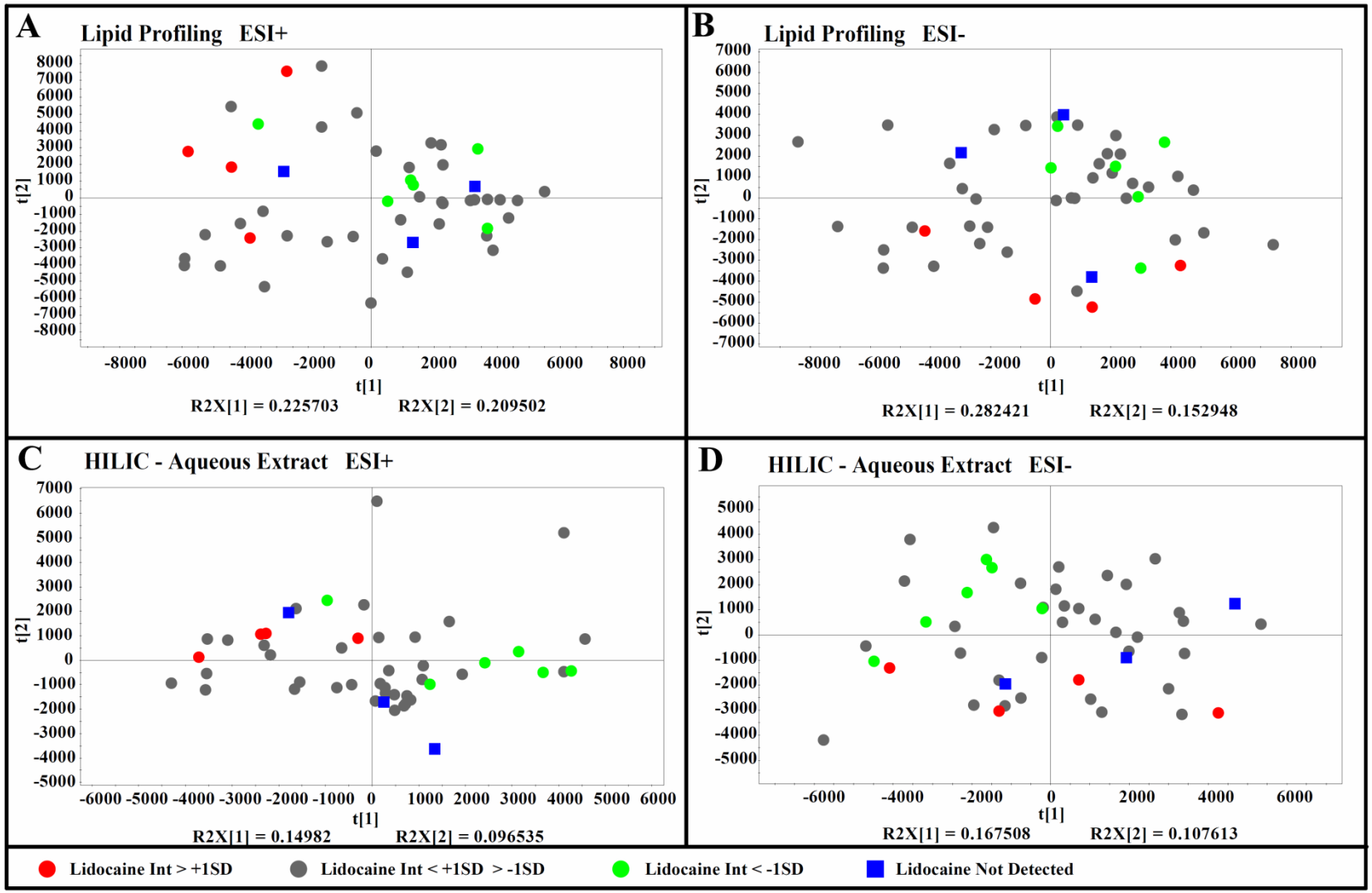


*Figure 4. 20: PCA scores plots of only the carotid samples in all analysis and polarity modes with colouring indicating the bioavailability of the local administered anaesthetic drug lidocaine in the plaque tissue. Colour coding: Red: Intensities over the +1SD, Grey: Intensities within the ±1SD, Green: Intensities lower than -1SD, and Blue: Samples with undetectable levels of lidocaine.*

For the current study, an initial validation was performed by demonstrating concordance with literature and well-established risk factors. Specifically, Cho and oxCEs which were detected with higher intensities in plaques are known to be in elevated concentrations in atheromas[148, 152]. The involvement of purine and pyrimidine pathways provide further confidence to our results and experimental design, as they are also known to be dysregulated in atherosclerosis.

Distinct metabolic profiles were detected from plaque tissue harvested from the carotid and femoral locations. However, the pathological origins of plaque formation, are more likely to be based on the common dysregulated biological pathways of CAR and FEM. Therefore, common and to some extent novel, findings are further elucidated. The involvement of PE-Cers and their strong statistical connections to plaque formation, Cho and CEs, represent a novel finding and common in both anatomical locations. At this point it is difficult to provide a hypothesis from the current information. It could be claimed that the reduction of PE-Cers in the cell membrane (CM) gives rise to Cho, as the CM can incorporate more Cho and Cho production in the endoplasmic reticulum continues. However, the trace amounts of PE-Cers in the human cell may be inconsistent with this hypothesis, and show some sort of signalling function to the molecule. PE-Cers are connected to the Cer pathway, and they may have a role in apoptosis of the cell, thus depositing Cho within the intima layer of the vessel. Nonetheless, this association to cholesterol can provide further insights into manifestation of disease but also might be able to explain ineffectiveness of cholesterol-lowering drugs.

Tetra- and tri-HexCers, are also universally dysregulated and may provide additional insides to the disease. Tri- and tetra-HexCers showed strong statistical power and high association to PE-Cers. The dysregulation of the Cer pathway, in general, was also common supporting the ubiquitous involvement of apoptosis in atherosclerosis. Additionally, the interruption of β-oxidation as inferred by the differential intensities of acyl-carnitines, and along this the recruitment of highly unsaturated fatty acids, reveal previously undemonstrated findings.

Differences between the two plaque tissues should also be taken into account. For FEM, the most intense differential findings were the more obvious dysregulation of β-oxidation, involving AcCs, FFAs and TGs, and the elevation of adenosine. For CAR it was the elevation of lysoPCs and members of the ceramide pathway, such as mono-HexCers, and sphingosines, and the elevation acyl-cholines. These are findings that should not be neglected as they may convey significant mechanistic and translational information by providing the biological rationale for cases of drug resistance/inefficiency. Reasons for these differences could be sought in the differential amount of

shear stress and stretch applied in these regions. Another possibility could be the different embryonic origins of vascular smooth muscle cells in the two anatomical locations which has been shown to phenotypically effect different cardiovascular diseases[180].

An important and, at the same time, difficult part of the present study was to provide rational interactions between pathways detected being dysregulated. Correlation analysis, pathway mapping and literature review provided information that could be used towards deconvolution of the complex pathology of this disease. In this top-down systems biology approach a number of interactive pathways were detected, with the connection of the Cho and Cer pathways being the highlight of the analysis.

Further experiments will seek to validate results and identify connections between pathways using bottom-up approaches and in a larger number of samples. Absolute quantification of metabolites involved, with a special interest in d18:1/16:0 sphingolipids, along with transcript expression and activity of involved enzymes, will also be part of future work.

### *4.4.1 Closing remarks*

Atherosclerosis is a multicentric and multistage disease. Health threatening events can occur as a result of plaque formation. These are generally caused by blood flow limitation, stroke and heart attack. Current therapeutic agents function towards cholesterol lowering. Cholesterol is known to be in the centre of advancement of the disease, but still prescribed drugs appear inadequate to mitigate stenosis or plaque rupture. For these reasons, there is an urgent unmet need for the scientific community to identify novel biomarkers/pharmaceutical targets to increase the potential of therapeutic treatment. This is potentially a rather difficult task, since atherosclerosis has been in the centre of scientific attention for many years now. However, in this study, new biochemical pathways have been identified as being involved in the progression to plaque formation from intimal thickening. This could be credited to the novelty of methodologies, modern instrumentation and experimental design. Systems biology approaches, by using multivariate statistics, can provide global information for the disease of interest and can identify or infer dysregulated pathways. Although the detected candidate biomarkers and inferred interactions from this study need further validation, they definitely constitute a step forward for the elucidation of the mechanisms of plaque formation in the arterial wall.

*Table 4. 6: List of structurally assigned statistically significant metabolites obtained from the analysis of tissue extracts, after comparison of the carotid to the intimal thickening group.*

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **Lipid Profiling Positive Mode** | | | | | | | | | | | |
| SM(d18:1/18:0) | 2 | C41H84N2O6P+ | 7.12 | 731.6073 | 731.6067 | 0.0006 | 1 | -0.73 | 1.5E-05 | -1.48 | 7 |
| SM(d18:1/24:0) | 2 | C47H96N2O6P+ | 13.33 | 815.7010 | 815.7006 | 0.0004 | 0 | -0.60 | 1.8E-05 | -1.42 | 6 |
| tetraHexCer(d18:1/16:0) | 2 | C58H108NO21+ | 9.25 | 1154.7557 | 1154.7414 | 0.0143 | 12 | -0.82 | 1.3E-04 | -2.09 | 10 |
| SM(d18:1/22:0) | 2 | C45H92N2O6P+ | 12.38 | 787.6700 | 787.6693 | 0.0007 | 1 | -0.57 | 2.0E-04 | -1.39 | 5 |
| triHexCer(d18:1/16:0) | 2 | C52H97NO18Na+ | 5.16 | 1046.6612 | 1046.6603 | 0.0009 | 1 | -0.61 | 3.7E-04 | -1.95 | 9 |
| PE(O-18:1/20:4) | 2 | C43H79NO7P+ | 9.24 | 752.5596 | 752.5594 | 0.0002 | 0 | -0.68 | 5.4E-03 | -2.33 | 2 |
| PE(O-18:2/20:4) | 2 | C43H77NO7P+ | 7.15 | 750.5447 | 750.5438 | 0.0009 | 1 | -0.64 | 5.7E-03 | -1.42 | 17 |
| SM(d18:1/17:0) | 2 | C40H82N2O6P+ | 6.15 | 717.5917 | 717.5911 | 0.0006 | 1 | -0.54 | 1.1E-02 | -1.41 | 3 |
| SM(d18:2/24:0) | 2 | C47H94N2O6P+ | 12.82 | 813.6861 | 813.6850 | 0.0011 | 1 | -0.58 | 1.6E-02 | -1.37 | 8 |
| PC(20:4/20:0) | 2 | C48H89NO8P+ | 9.22 | 838.6332 | 838.6326 | 0.0006 | 1 | -0.65 | 2.7E-02 | -1.40 | 3 |
| oxCE(C18/O3) | 2 | C45H74O5Na+ | 13.07 | 717.5424 | 717.5434 | -0.0010 | -1 | 0.62 | 5.8E-08 | 52.05 | 30 |
| HexCer(d18:1/23:0) | 2 | C47H90NO7+ [M-H2O+H]+ | 13.29 | 780.6745 | 780.6717 | 0.0028 | 4 | 0.58 | 3.9E-07 | 1.47 | 16 |
| cholesterol | 3 | C27H45+ [M-(H2O)+(H)]+ | 5.34 | 369.3522 | 369.3521 | 0.0001 | 0 | 0.65 | 5.1E-07 | 1.44 | 7 |
| oxCE(C18/O2) | 2 | C45H76O4Na+ | 13.37 | 703.5646 | 703.5641 | 0.0005 | 1 | 0.56 | 1.0E-04 | 1.87 | 4 |
| HexCer(d18:1/24:0) | 2 | C48H92NO7+ [M-H2O+H]+ | 13.53 | 794.6878 | 794.6874 | 0.0004 | 1 | 0.55 | 8.2E-04 | 1.81 | 29 |
| SM(d16:1/24:1) & SM(d18:1/22:1) | 2 | C45H89N2O6PNa+ [M+Na]+ | 9.37 | 807.6360 | 807.6356 | 0.0004 | 0 | 0.51 | 1.7E-02 | 1.58 | 14 |
| | | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | | |
| PE-Cer(d18:1/16:0) | 2 | C36H72N2O6P- | 5.65 | 659.5125 | 659.5128 | -0.0003 | 0 | -0.87 | 7.0E-09 | -2.24 | 3 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | | |
| SM(d18:1/24:1) | 2 | C48H94N2O8P- [M+FA-H]- | 12.24 | 857.6748 | 857.6748 | 0.0000 | 0 | -0.52 | 1.6E-07 | -1.29 | 3 |
| PE-Cer(d18:1/24:1) | 2 | C44H86N2O6P- | 12.82 | 769.6202 | 769.6223 | -0.0021 | -3 | -0.79 | 3.7E-06 | -2.05 | 14 |
| triHexCer(d18:1/16:0) | 2 | C53H98NO20- [M+FA-H]- | 5.18 | 1068.6684 | 1068.6682 | 0.0002 | 0 | -0.57 | 9.5E-05 | -2.07 | 24 |
| UKN-LP-NEG-763.5511-5.07 | na | ud | 5.07 | 763.5511 | ud | ud | ud | -0.56 | 3.8E-04 | -1.76 | 4 |
| PE(O-18:2/20:4) | 2 | C43H75NO7P- | 7.16 | 748.5284 | 748.5281 | 0.0003 | 0 | -0.63 | 1.3E-03 | -1.56 | 2 |
| PE(16:0/22:4) | 2 | C43H77NO8P- | 7.51 | 766.5381 | 766.5387 | -0.0006 | -1 | -0.53 | 2.1E-03 | -1.56 | 6 |
| PE(O-18:1/20:4) | 2 | C43H77NO7P- | 9.27 | 750.5438 | 750.5438 | 0.0000 | 0 | -0.61 | 2.2E-03 | -1.89 | 1 |
| PE(O-18:1/22:6) | 2 | C45H77NO7P- | 8.53 | 774.5443 | 774.5438 | 0.0005 | 1 | -0.64 | 2.4E-03 | -1.82 | 3 |
| PE(O-18:2/22:6) | 2 | C45H75NO7P- | 6.65 | 772.5288 | 772.5281 | 0.0007 | 1 | -0.58 | 3.0E-03 | -1.58 | 18 |
| PE(O-16:1/22:6) | 2 | C43H73NO7P- | 6.48 | 746.5134 | 746.5125 | 0.0009 | 1 | -0.57 | 6.1E-03 | -1.58 | 3 |
| PE(O-18:1/22:4) | 2 | C45H81NO7P- | 11.25 | 778.5752 | 778.5751 | 0.0001 | 0 | -0.59 | 6.8E-03 | -2.34 | 14 |
| PE(O-16:1/20:4) | 2 | C41H73NO7P- | 7.01 | 722.5123 | 722.5125 | -0.0002 | 0 | -0.53 | 6.8E-03 | -1.45 | 1 |
| PS(18:0/22:6) | 2 | C46H77NO10P- | 5.63 | 834.5305 | 834.5285 | 0.0020 | 2 | -0.52 | 9.8E-03 | -3.02 | 4 |
| PE(O-16:1/18:1) | 2 | C39H75NO7P- | 9.13 | 700.5286 | 700.5281 | 0.0005 | 1 | -0.50 | 1.5E-02 | -1.64 | 5 |
| PE(18:0/22:6) | 2 | C45H77NO8P- | 7.58 | 790.5394 | 790.5387 | 0.0007 | 1 | -0.50 | 1.8E-02 | -1.59 | 4 |
| PE(O-16:1/22:4) | 2 | C43H77NO7P- | 8.53 | 750.5442 | 750.5438 | 0.0004 | 1 | -0.51 | 2.5E-02 | -1.48 | 4 |
| PC(O-16:1/20:4) | 2 | C45H81NO9P- [M+FA-H]- | 6.36 | 810.5563 | 810.5649 | -0.0086 | -11 | -0.52 | 3.8E-02 | -1.19 | 2 |
| lysoPC(O-16:0) | 2 | C25H53NO8P- [M+FA-H]- | 1.39 | 526.3508 | 526.3509 | -0.0001 | 0 | 0.50 | 6.1E-06 | 1.61 | 3 |
| Cer(d18:1/22:0) | 2 | C41H80NO5- [M+FA-H]- | 13.46 | 666.6030 | 666.6037 | -0.0007 | -1 | 0.61 | 4.9E-06 | 1.82 | 10 |
| Cer(d18:1/23:0) | 2 | C42H82NO5- [M+FA-H]- | 13.76 | 680.6195 | 680.6193 | 0.0002 | 0 | 0.65 | 4.3E-07 | 2.13 | 17 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | | |
| HexCer(d18:1/24:0) | 2 | C49H94NO10- [M+FA-H]- | 13.53 | 856.6875 | 856.6878 | -0.0003 | 0 | 0.63 | 4.1E-06 | 1.96 | 12 |
| Cer(d18:1/14:0) | 2 | C33H64NO5- [M+FA-H]- | 5.72 | 554.4790 | 554.4784 | 0.0006 | 1 | 0.58 | 4.2E-07 | 2.64 | 3 |
| HexCer(d18:1/23:0) | 2 | C48H92NO10- [M+FA-H]- | 13.27 | 842.6726 | 842.6721 | 0.0005 | 1 | 0.52 | 3.7E-05 | 1.33 | 6 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| **HILIC Aqueous Extract Positive Mode** | | | | | | | | | | | |
| PE(O-18:1/20:4) | 2 | C43H79NO7P+ | 4.31 | 752.5586 | 752.5594 | -0.0008 | -1 | -0.78 | 2.1E-06 | -1.90 | 22 |
| PE(16:0/20:4) | 2 | C41H75NO8P+ | 4.40 | 740.5224 | 740.5230 | -0.0006 | -1 | -0.65 | 7.9E-06 | -1.61 | 15 |
| uridine | 4 | C18H23N4O12Na2+ (2M+2Na-H]+ | 1.49 | 533.1104 | 533.1108 | -0.0004 | -1 | -0.69 | 1.3E-04 | -1.39 | 9 |
| acetylcarnitine | 4 | C9H18NO4+ | 6.99 | 204.1225 | 204.1236 | -0.0011 | -5 | -0.60 | 1.4E-04 | -1.15 | 9 |
| PE(O-18:1/22:6) | 2 | C45H79NO7P+ | 4.29 | 776.5578 | 776.5594 | -0.0016 | -2 | -0.62 | 2.1E-04 | -1.79 | 12 |
| inosine | 4 | C5H5N4O+ | 3.07 | 137.0513 | 137.0463 | 0.0050 | 36 | -0.69 | 2.5E-04 | -1.94 | 5 |
| PE(O-18:1/22:5) | 2 | C45H81NO7P+ | 4.30 | 778.5767 | 778.5751 | 0.0016 | 2 | -0.60 | 6.3E-04 | -1.44 | 5 |
| PE(O-16:0/20:4) | 2 | C41H77NO7P+ | 4.34 | 726.5373 | 726.5438 | -0.0065 | -9 | -0.66 | 6.7E-04 | -1.44 | 21 |
| PE(O-16:1/20:4) | 2 | C41H75NO7P+ | 4.34 | 724.5275 | 724.5281 | -0.0006 | -1 | -0.65 | 1.2E-03 | -1.24 | 9 |
| SM(d18:2/16:0) | 2 | C39H78N2O6P+ | 5.28 | 701.5591 | 701.5598 | -0.0007 | -1 | -0.63 | 1.2E-03 | -4.59 | 10 |
| lysoPE(O-18:1) | 2 | C23H49NO6P+ | 5.10 | 466.3294 | 466.3297 | -0.0003 | -1 | -0.60 | 1.2E-03 | -2.32 | 9 |
| PC(14:0/16:0) | 2 | C38H77NO8P+ | 4.85 | 706.5375 | 706.5387 | -0.0012 | -2 | -0.61 | 1.7E-03 | -1.68 | 8 |
| UKN-HL-POS-86.0894-4.52 | na | C5H12N+ | 4.52 | 86.0894 | ud | ud | ud | -0.61 | 1.7E-03 | -2.60 | 8 |
| PE(O-16:1/18:1) | 2 | C39H77NO7P+ | 4.40 | 702.5441 | 702.5438 | 0.0003 | 0 | -0.75 | 1.8E-03 | -2.96 | 15 |
| N-methyladenosine | 2 | C11H16N5O4+ | 6.01 | 282.1197 | 282.1202 | -0.0005 | 2 | -0.80 | 2.5E-03 | -1.60 | 7 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **HILIC Aqueous Extract Positive Mode** | | | | | | | | | | | |
| hypoxanthine | 4 | C5H5N4O+ | 2.20 | 137.0491 | 137.0463 | 0.0028 | 20 | -0.58 | 3.8E-03 | -1.80 | 12 |
| UKN-HL-POS-160.0355-6.09 | na | C9H6NO2+ | 6.09 | 160.0355 | ud | ud | ud | -0.56 | 8.3E-03 | -1.61 | 20 |
| butyrylcarnitine * | 4 | C11H22NO4+ | 6.46 | 232.1537 | 232.1549 | -0.0012 | -5 | -0.61 | 1.0E-02 | -2.19 | 6 |
| 1-methylnicotinamide † | 4 | C7H9N2O+ | 5.65 | 137.0772 | 137.0715 | 0.0057 | 42 | -0.62 | 1.3E-02 | -1.29 | 10 |
| guanosine | 4 | C5H6N5O+ | 4.36 | 152.0564 | 152.0572 | -0.0008 | -5 | -0.61 | 1.4E-02 | -1.64 | 4 |
| creatine | 4 | C4H10N3O2+ | 6.52 | 132.0764 | 132.0773 | -0.0009 | -7 | -0.50 | 1.4E-02 | -1.40 | 7 |
| proline | 4 | C5H10NO2+ | 6.11 | 116.0646 | 116.0712 | -0.0066 | -57 | -0.55 | 1.5E-02 | -1.37 | 19 |
| UKN-HL-POS-84.0801-5.16 | na | C5H10N+ ‡ | 5.16 | 84.0801 | ud | ud | ud | -0.51 | 1.9E-02 | -3.31 | 20 |
| UKN-HL-POS-217.1036-1.45 | na | ud | 1.45 | 217.1036 | ud | ud | ud | -0.56 | 3.2E-02 | -1.29 | 5 |
| sphingosine | 4 | C18H36NO+ [M-H2O+H]+ | 2.73 | 282.2786 | 282.2797 | -0.0011 | -4 | 0.53 | 4.1E-09 | 3.49 | 6 |
| sphingosine (C16) | 2 | C16H32NO+ [M-H2O+H]+ | 2.94 | 254.2477 | 254.2484 | -0.0007 | -3 | 0.57 | 2.6E-08 | 7.57 | 8 |
| lysoPC(O-18:2) | 2 | C26H53NO6P+ | 5.39 | 506.3602 | 506.3610 | -0.0008 | -2 | 0.69 | 9.3E-07 | 2.24 | 5 |
| lysoPC(20:4) | 2 | C28H51NO7P+ | 5.39 | 544.3373 | 544.3403 | -0.0030 | -6 | 0.68 | 1.9E-06 | ud | ud |
| lysoPC(20:2) | 2 | C28H55NO7P+ | 5.89 | 548.3694 | 548.3716 | -0.0022 | -4 | 0.71 | 4.1E-06 | 1.98 | 12 |
| lysoPC(22:4) | 2 | C30H55NO7P+ | 5.80 | 572.3710 | 572.3716 | -0.0006 | -1 | 0.67 | 4.3E-06 | 2.75 | 28 |
| lysoPC(O-18:1) | 2 | C26H55NO6P+ | 6.16 | 508.3759 | 508.3767 | -0.0008 | -2 | 0.74 | 4.7E-06 | 1.83 | 12 |
| lysoPC(18:2) | 2 | C26H51NO7P+ | 5.44 | 520.3387 | 520.3403 | -0.0016 | -3 | 0.67 | 5.3E-06 | ud | ud |
| palmitoylcholine | 2 | C21H44NO2+ | 2.10 | 342.3368 | 342.3372 | -0.0004 | -1 | 0.57 | 6.2E-06 | 2.90 | 12 |
| lysoPC(20:0) | 2 | C28H59NO7P+ | 5.41 | 552.4064 | 552.4029 | 0.0035 | 6 | 0.72 | 8.1E-06 | 2.46 | 15 |
| acylcholine (18:2) | 2 | C23H44NO2+ | 2.09 | 366.3361 | 366.3672 | -0.0311 | -85 | 0.54 | 1.2E-05 | 2.91 | 9 |
| lysoPC(16:1) | 2 | C24H49NO7P+ | 6.13 | 494.3233 | 494.3247 | -0.0014 | -3 | 0.56 | 2.0E-05 | 1.75 | 4 |
| lysoPC(O-16:1) | 2 | C24H51NO6P+ | 5.42 | 480.3437 | 480.3454 | -0.0017 | -4 | 0.65 | 3.1E-05 | 1.66 | 6 |
| arachidonoylcholine | 2 | C25H44NO2+ | 2.03 | 390.3354 | 390.3372 | -0.0018 | -5 | 0.51 | 3.3E-05 | 3.18 | 8 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **HILIC Aqueous Extract Positive Mode** | | | | | | | | | | | |
| stearoylcholine | 2 | C23H48NO2+ | 2.05 | 370.3668 | 370.3685 | -0.0017 | -5 | 0.50 | 3.9E-05 | 3.94 | 8 |
| 4-(Trimethylammonio)but-2-enoate | 1 | C7H14NO2+ | 7.70 | 144.1016 | 144.1025 | -0.0009 | -6 | 0.55 | 2.1E-04 | 1.62 | ud |
| lysoPC(18:1) | 2 | C26H53NO7P+ | 5.97 | 522.3558 | 522.3560 | -0.0002 | 0 | 0.77 | 2.7E-04 | 1.57 | 7 |
| PC(O-18:1/22:6) | 2 | C48H85NO7P+ | 4.68 | 818.6045 | 818.6064 | -0.0019 | -2 | 0.50 | 1.4E-03 | 1.43 | 6 |
| acylcholine (18:1) | 2 | C23H46NO2+ | 1.97 | 368.3724 | 368.3529 | 0.0195 | 53 | 0.51 | 1.8E-03 | 2.07 | 9 |
| acylcarnitine (18:1) | 4 | C25H48NO4+ | 4.97 | 426.3562 | 426.3583 | -0.0021 | -5 | 0.50 | 2.1E-03 | 1.90 | 18 |
| lysoPC(14:0) | 2 | C22H47NO7P+ | 6.19 | 468.3083 | 468.3090 | -0.0007 | -1 | 0.65 | 3.1E-03 | 1.18 | 8 |
| lysoPC(18:0) | 2 | C26H54NO7PNa+ [M+Na]+ | 5.91 | 546.3543 | 546.3536 | 0.0007 | 1 | 0.62 | 3.2E-03 | 1.22 | 14 |
| PC(O-16:1/22:6) | 2 | C46H81NO7P+ | 4.67 | 790.5873 | 790.5751 | 0.0122 | 15 | 0.51 | 4.5E-03 | 1.49 | 6 |
| | | | | | | | | | | | |
| **HILIC Aqueous Extract Negative Mode** | | | | | | | | | | | |
| PE(18:0/22:4) | 2 | C45H81NO8P- | 4.35 | 794.5699 | 794.5700 | -0.0001 | 0 | -0.76 | 3.3E-07 | -2.00 | 11 |
| PE(O-18:1/20:4) | 2 | C43H77NO7P- | 4.28 | 750.5438 | 750.5438 | 0.0000 | 0 | -0.80 | 3.6E-07 | -2.27 | 8 |
| PE(O-16:1/20:4) | 2 | C41H73NO7P- | 4.31 | 722.5135 | 722.5125 | 0.0010 | 1 | -0.65 | 1.6E-05 | -1.53 | 7 |
| PE(18:0/20:4) | 2 | C43H77NO8P- | 4.35 | 766.5390 | 766.5387 | 0.0003 | 0 | -0.54 | 2.0E-05 | -1.38 | 9 |
| inosine | 4 | C10H11N4O5- | 3.05 | 267.0725 | 267.0729 | -0.0004 | -2 | -0.48 | 9.2E-05 | -1.42 | 7 |
| PE(O-18:1/22:6) | 2 | C45H77NO7P- | 4.27 | 774.5441 | 774.5438 | 0.0003 | 0 | -0.57 | 2.9E-04 | -1.77 | 5 |
| methyluridine | 4 | C10H13N2O6- | 1.36 | 257.0800 | 257.0774 | 0.0026 | 10 | -0.55 | 1.4E-03 | -1.61 | 10 |
| arachidonic Acid | 3 | C20H31O2- | 0.69 | 303.2324 | 303.2324 | 0.0000 | 0 | -0.50 | 1.9E-03 | -1.74 | 21 |
| N-acetylmethionine | 4 | C7H12NO3S- | 4.33 | 190.0536 | 190.0538 | -0.0002 | -1 | -0.50 | 2.5E-03 | -1.85 | 9 |
| PG(18:1/22:6) | 2 | C46H76O10P- | 1.96 | 819.5178 | 819.5176 | 0.0002 | 0 | 0.59 | 1.6E-07 | 3.07 | 9 |
| PG(18:2/18:1) | 2 | C42H76O10P- | 2.13 | 771.5177 | 771.5176 | 0.0001 | 0 | 0.54 | 4.2E-07 | 3.69 | 12 |
| lysoPC(20:2) | 2 | C29H55NO9P- [M+FA-H]- | 5.80 | 592.3600 | 592.3614 | -0.0014 | -2 | 0.66 | 4.9E-07 | 2.21 | 9 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **HILIC Aqueous Extract Negative Mode** | | | | | | | | | | | |
| UKN-HL-NEG-139.9840-2.98 | na | ud | 2.98 | 139.9840 | ud | ud | ud | 0.53 | 6.3E-07 | 1.88 | 7 |
| lysoPC(O-18:2) | 2 | C27H53NO8P-[M+FA-H]- | 5.38 | 550.3514 | 550.3509 | 0.0005 | 1 | 0.56 | 1.4E-06 | 3.29 | 17 |
| lysoPC(20:4) | 2 | C29H51NO9P-[M+FA-H] | 5.78 | 588.3310 | 588.3301 | 0.0009 | 2 | 0.62 | 1.7E-06 | 1.64 | 6 |
| PG(18:2/22:6) | 2 | C46H74O10P- | 1.96 | 817.5025 | 817.5020 | 0.0005 | 1 | 0.57 | 2.4E-06 | 3.52 | 17 |
| lysoPC(O-18:1) | 2 | C27H55NO8P-[M+FA-H]- | 6.13 | 552.3670 | 552.3665 | 0.0005 | 1 | 0.70 | 1.1E-05 | 2.57 | 7 |
| lysoPC(20:3) | 2 | C29H53NO9P-[M+FA-H]- | 5.82 | 590.3460 | 590.3458 | 0.0002 | 0 | 0.62 | 1.5E-05 | 1.64 | 6 |
| PC(O-18:1/20:4) | 2 | C47H85NO9P-[M+FA-H] | 4.68 | 838.5983 | 838.5962 | 0.0021 | 3 | 0.57 | 1.7E-05 | 1.70 | 7 |
| PC(18:2/18:1) & PC(16:0/20:3) | 2 | C45H83NO10P-[M+FA-H]- | 4.75 | 828.5759 | 828.5755 | 0.0004 | 0 | 0.66 | 6.4E-05 | 1.50 | 6 |
| lysoPC(18:2) | 2 | C27H51NO9P-[M+FA-H]- | 5.97 | 564.3307 | 564.3301 | 0.0006 | 1 | 0.68 | 9.8E-05 | 1.94 | 6 |
| lysoPC(18:1) | 2 | C27H53NO9P-[M+FA-H]- | 5.88 | 566.3464 | 566.3458 | 0.0006 | 1 | 0.66 | 1.4E-04 | 2.18 | 4 |
| PC(18:1/20:4) | 2 | C47H83NO10P-[M+FA-H]- | 4.70 | 852.5767 | 852.5755 | 0.0012 | 1 | 0.69 | 1.5E-04 | 1.54 | 9 |
| lysoPC(O-16:1) | 2 | C25H51NO8P-[M+FA-H]- | 5.39 | 524.3315 | 524.3352 | -0.0037 | -7 | 0.53 | 1.9E-04 | 1.78 | 5 |
| PC(O-24:2/20:4) | 2 | C53H95NO9P-[M+FA-H]- | 4.64 | 920.6780 | 920.6744 | 0.0036 | 4 | 0.64 | 3.6E-04 | 1.80 | 13 |
| UKN-HL-NEG-152.0013-3.14 | na | ud | 3.14 | 152.0013 | ud | ud | ud | 0.55 | 9.2E-04 | 1.96 | ud |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **HILIC Aqueous Extract Negative Mode** | | | | | | | | | | | |
| PC(18:0/22:6) | 2 | C49H85NO10P-[M+FA-H]- | 4.68 | 878.5882 | 878.5911 | -0.0029 | -3 | 0.54 | 1.7E-03 | 1.21 | 6 |
| lysoPC(14:0) | 2 | C23H47NO9P-[M+FA-H]- | 6.15 | 512.2993 | 512.2988 | 0.0005 | 1 | 0.59 | 2.2E-03 | 1.50 | 12 |
| PC(16:0/22:6) | 2 | C47H81NO10P-[M+FA-H]- | 4.70 | 850.5602 | 850.5598 | 0.0004 | 0 | 0.50 | 6.0E-03 | 1.31 | 10 |

* Standard compound of isobutyrylcarnitine exhibited different RT
† Standard compound of 6-methylnicotinamide exhibited different RT
‡ Molecular formula as calculated using the isotopic pattern (tentative)

LoA: Level of Assignment; 1: Tentative assignment; 2: Tandem MS spectrum matched to database or literature; 3: RT matched to standard compound; 4: MS/MS spectrum and RT, matched to standard compound.
RT: Retention time.
HILIC: Hydrophilic Interaction (Liquid) Chromatography.
P(corr) refers to the correlation coefficient of metabolites to disease classes.
Two-tailed t-tests were conducted, assuming unequal variance.
Coefficient of variation (CV%) is calculated based on thirteen injections of the same pooled quality control (QC) sample, acquired throughout the run.
The position of double bonds and the position of fatty acyl chain cannot be determined in lipid moieties (Cer, HexCer, PC, PE, PG, and SM). Fatty acyl chains are presented from lowest to highest molecular weight (lower number of carbons to higher, and higher number of double bonds to lower).

Cer: ceramide; FA: formate; Hex: hexosyl; na: not applicable; oxCE: oxidised cholesterol ester; PC: phosphatidylcholine; PE: phosphatidylethanolamine; PG: phosphatidylglycerol; SM: sphingomyelin; ud: unable to determine; UKN: unknown.

*Table 4. 7: List of structurally assigned significant metabolites obtained from analysis of tissue extracts, after comparison of the femoral to the intimal thickening group.*

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **Lipid Profiling Positive Mode** | | | | | | | | | | | |
| SM(d18:2/24:1) | 2 | C47H92N2O6P+ | 9.74 | 811.6693 | 811.6693 | 0.0000 | 0 | -0.80 | 3.07E-08 | -2.20 | 7 |
| triHexCer(d18:1/16:0) | 2 | C52H97NO18Na+ [M+Na]+ | 5.16 | 1046.6612 | 1046.6603 | 0.0009 | 1 | -0.85 | 5.25E-07 | -3.39 | 9 |
| tetraHexCer(d18:1/16:0) | 2 | C58H108NO21+ | 9.25 | 1154.7557 | 1154.7414 | 0.0143 | 12 | -0.89 | 5.95E-07 | -10.12 | 10 |
| SM(d18:1/17:0) | 2 | C40H82N2O6P+ | 6.15 | 717.5917 | 717.5911 | 0.0006 | 1 | -0.84 | 6.25E-07 | -3.13 | 3 |
| SM(d18:2/16:0) | 2 | C39H78N2O6P+ | 4.31 | 701.5600 | 701.5598 | 0.0002 | 0 | -0.81 | 8.11E-07 | -2.96 | 7 |
| SM(d18:1/14:0) & SM(d16:1/16:0) | 2 | C37H76N2O6P+ | 4.15 | 675.5445 | 675.5441 | 0.0004 | 1 | -0.79 | 1.12E-06 | -2.50 | 8 |
| PC(16:0/16:0) | 2 | C40H81NO8P+ | 7.17 | 734.5706 | 734.5700 | 0.0006 | 1 | -0.81 | 6.15E-06 | -2.18 | 7 |
| SM(d18:1/18:0) | 2 | C41H84N2O6P+ | 7.12 | 731.6073 | 731.6067 | 0.0006 | 1 | -0.79 | 8.98E-06 | -1.71 | 7 |
| PC(16:1/16:0) | 2 | C40H79NO8P+ | 5.66 | 732.5551 | 732.5543 | 0.0008 | 1 | -0.75 | 2.44E-05 | -2.42 | 3 |
| PE(O-16:1/22:6) | 2 | C43H75NO7P+ | 6.47 | 748.5296 | 748.5281 | 0.0015 | 2 | -0.82 | 4.65E-05 | -2.29 | 8 |
| PC(O-16:0/16:1) | 2 | C40H81NO7P+ | 8.12 | 718.5694 | 718.5751 | -0.0057 | -8 | -0.67 | 5.08E-05 | -1.94 | 2 |
| PC(O-16:1/20:4) | 2 | C44H81NO7P+ | 6.45 | 766.5751 | 766.5751 | 0.0000 | 0 | -0.82 | 9.72E-05 | -1.96 | 4 |
| SM(d18:2/24:0) | 2 | C47H94N2O6P+ | 12.82 | 813.6861 | 813.6850 | 0.0011 | 1 | -0.74 | 1.33E-04 | -1.98 | 8 |
| PC(18:0/20:4) | 2 | C46H85NO8P+ | 7.58 | 810.6017 | 810.6013 | 0.0004 | 1 | -0.77 | 4.05E-04 | -1.44 | 7 |
| SM(d18:1/24:0) | 2 | C47H96N2O6P+ | 13.33 | 815.7010 | 815.7006 | 0.0004 | 0 | -0.63 | 4.95E-04 | -1.23 | 6 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **Lipid Profiling Positive Mode** | | | | | | | | | | | |
| PE(O-18:2/20:4) | 2 | C43H77NO7P+ | 7.15 | 750.5447 | 750.5438 | 0.0009 | 1 | -0.68 | 1.85E-03 | -1.50 | 17 |
| PC(20:4/20:0) | 2 | C48H89NO8P+ | 9.22 | 838.6332 | 838.6326 | 0.0006 | 1 | -0.72 | 2.16E-03 | -2.26 | 3 |
| PC(16:0/22:4) | 2 | C46H85NO8P+ | 6.96 | 810.6012 | 810.6013 | -0.0001 | 0 | -0.63 | 3.74E-03 | -1.84 | 3 |
| PC(16:0/18:2) | 2 | C42H80NO8PNa+ [M+Na]+ | 6.01 | 780.5527 | 780.5519 | 0.0008 | 1 | 0.73 | 3.26E-08 | 1.72 | 2 |
| oxCE(C18/O) | 2 | C45H76O3Na+ [M+Na]+ | 13.94 | 687.5696 | 687.5692 | 0.0004 | 1 | 0.74 | 3.91E-08 | 3.34 | 7 |
| TG(16:0/18:1/18:1) | 2 | C55H106NO6+ [M+NH4]+ | 15.69 | 876.8013 | 876.8020 | -0.0007 | -1 | 0.69 | 5.40E-08 | 3.36 | 4 |
| oxCE(C18/O) | 2 | C45H74O3Na+ [M+Na]+ | 13.99 | 685.5541 | 685.5536 | 0.0005 | 1 | 0.70 | 1.56E-07 | 3.22 | 4 |
| UKN-LP-POS-895.7366-14.23 | na | ud | 14.23 | 895.7366 | ud | ud | ud | 0.75 | 2.11E-07 | 2.90 | 9 |
| TG(16:0/16:0/18:1) | 2 | C53H104NO6+ [M+NH4]+ | 15.65 | 850.7849 | 850.7864 | -0.0015 | -2 | 0.68 | 2.15E-07 | 3.44 | 2 |
| TG(16:0/18:1/18:2) & TG(16:1/18:1/18:1) | 2 | C55H104NO6+ [M+NH4]+ | 15.47 | 874.7869 | 874.7864 | 0.0005 | 1 | 0.66 | 2.58E-07 | 3.26 | 4 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **Lipid Profiling Positive Mode** | | | | | | | | | | | |
| oxCE(C18/O2) | 2 | C45H76O4Na+ [M+Na]+ | 13.37 | 703.5646 | 703.5641 | 0.0005 | 1 | 0.68 | 5.57E-07 | 3.46 | 4 |
| TG(16:0/16:1/18:1) | 2 | C53H102NO6+ [M+NH4]+ | 15.44 | 848.7717 | 848.7707 | 0.0010 | 1 | 0.63 | 6.04E-07 | 3.53 | 4 |
| TG(18:1/18:1/18:1) | 2 | C57H108O6N+ [M+NH4]+ | 15.69 | 902.8174 | 902.8177 | -0.0003 | 0 | 0.61 | 1.82E-06 | 3.32 | 3 |
| PC(16:0/18:1) | 2 | C42H82NO8PNa+ [M+Na]+ | 7.39 | 782.5682 | 782.5676 | 0.0006 | 1 | 0.64 | 2.24E-06 | 1.53 | 4 |
| TG(18:1/18:1/18:0) | 2 | C57H110O6N+ [M+NH4]+ | 15.89 | 904.8300 | 904.8333 | -0.0033 | -4 | 0.59 | 5.72E-06 | 3.62 | 6 |
| TG(18:2/18:1/18:1) | 2 | C57H106O6N+ [M+NH4]+ | 15.46 | 900.8016 | 900.8020 | -0.0004 | 0 | 0.59 | 5.81E-06 | 2.88 | 2 |
| PC(16:0/20:4) | 2 | C44H80NO8PNa+ [M+Na]+ | 5.75 | 804.5520 | 804.5519 | 0.0001 | 0 | 0.63 | 6.54E-06 | 1.65 | 14 |
| TG(16:1/16:1/18:1) & TG(16:1/16:0/18:2) & TG(14:0/18:2/18:1) | 2 | C53H100NO6+ [M+NH4]+ | 15.16 | 846.7553 | 846.7551 | 0.0002 | 0 | 0.55 | 1.02E-05 | 3.09 | 3 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **Lipid Profiling Positive Mode** | | | | | | | | | | | |
| TG(18:2/18:2/18:1) & TG(18:3/18:1/18:1) & TG(16:0/18:2/20:3) | 2 | C57H104NO6+ [M+NH4]+ | 15.22 | 898.7860 | 898.7864 | -0.0004 | 0 | 0.56 | 1.58E-05 | 2.36 | 2 |
| Cholesterol | 3 | C27H45+ [M-H2O+H]+ | 5.34 | 369.3522 | 369.3521 | 0.0001 | 0 | 0.67 | 2.22E-05 | 1.41 | 7 |
| PC(16:0/22:6) | 2 | C46H81NO8P+ | 5.34 | 806.5708 | 806.5700 | 0.0008 | 1 | 0.55 | 1.03E-04 | 1.81 | 6 |
| | | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | | |
| PE-Cer(d18:1/16:0) | 2 | C36H72N2O6P- | 5.65 | 659.5125 | 659.5128 | -0.0003 | 0 | -0.94 | 9.80E-12 | -3.54 | 3 |
| triHexCer(d18:1/16:0) | 2 | C53H98NO20- [M+FA-H]- | 5.18 | 1068.6684 | 1068.6682 | 0.0002 | 0 | -0.84 | 7.37E-07 | -3.34 | 24 |
| SM(d18:2/24:1) | 2 | C48H92N2O8P- [M+FA-H]- | 9.74 | 855.6589 | 855.6591 | -0.0002 | 0 | -0.73 | 2.39E-06 | -1.42 | 3 |
| PE-Cer(d18:1/24:1) | 2 | C44H86N2O6P- | 12.82 | 769.6202 | 769.6223 | -0.0021 | -3 | -0.70 | 3.76E-06 | -2.41 | 14 |
| UKN-LP-NEG-763.5511-5.07 | | ud | 5.07 | 763.5511 | ud | ud | ud | -0.74 | 3.87E-06 | -2.88 | 4 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | | |
| SM(d18:2/16:0) | 2 | C40H78N2O8P- [M+FA-H]- | 4.32 | 745.5492 | 745.5496 | -0.0004 | -1 | -0.74 | 3.98E-06 | -1.94 | 4 |
| UKN-LP-NEG-869.6484-10.89 | na | ud | 10.89 | 869.6484 | ud | ud | ud | -0.72 | 5.52E-06 | ud | 5 |
| SM(d18:1/14:0) & SM(d16:1/16:0) | 2 | C38H76N2O8P- [M+FA-H]- | 4.15 | 719.5339 | 719.5339 | 0.0000 | 0 | -0.66 | 1.19E-05 | -1.55 | 3 |
| SM(d18:1/16:0) & SM(d16:1/18:0) | 2 | C40H80N2O8P- [M+FA-H]- | 5.39 | 747.5648 | 747.5652 | -0.0004 | -1 | -0.54 | 9.47E-05 | -1.28 | 5 |
| lysoPE(O-18:1) | 2 | C23H47NO6P- | 1.98 | 464.3139 | 464.3141 | -0.0002 | 0 | -0.67 | 2.36E-04 | -1.86 | 5 |
| lysoPE(O-16:1) | 2 | C21H43NO6P- | 1.42 | 436.2825 | 436.2828 | -0.0003 | -1 | -0.67 | 2.37E-04 | -2.08 | 19 |
| PE(O-16:1/22:6) | 2 | C43H73NO7P- | 6.48 | 746.5134 | 746.5125 | 0.0009 | 1 | -0.77 | 3.04E-04 | -2.29 | 3 |
| SM(d18:0/16:0) | 2 | C40H82N2O8P- [M+FA-H]- | 5.96 | 749.5806 | 749.5809 | -0.0003 | 0 | -0.57 | 3.71E-04 | -1.24 | 2 |
| PE(O-18:2/22:6) | 2 | C45H75NO7P- | 6.65 | 772.5288 | 772.5281 | 0.0007 | 1 | -0.72 | 4.02E-04 | -2.01 | 18 |
| PE(O-18:2/20:4) | 2 | C43H75NO7P- | 7.16 | 748.5284 | 748.5281 | 0.0003 | 0 | -0.79 | 4.58E-04 | -1.67 | 2 |
| PC(O-18:1/16:0) | 2 | C43H85NO9P- [M+FA-H]- | 8.65 | 790.5972 | 790.5962 | 0.0010 | 1 | -0.65 | 5.96E-04 | -2.94 | 11 |
| PS(18:0/22:6) | 2 | C46H77NO10P- | 5.63 | 834.5305 | 834.5285 | 0.0020 | 2 | -0.69 | 2.54E-03 | -16.05 | 4 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | | |
| PI(18:0/20:4) | 2 | C47H82O13P- | 5.82 | 885.5495 | 885.5493 | 0.0002 | 0 | -0.56 | 5.26E-03 | -2.49 | 3 |
| PE(18:0/22:6) | 2 | C45H77NO8P- | 7.58 | 790.5394 | 790.5387 | 0.0007 | 1 | -0.60 | 6.95E-03 | -1.95 | 4 |
| PE(O-18:1/22:6) | 2 | C45H77NO7P- | 8.53 | 774.5443 | 774.5438 | 0.0005 | 1 | -0.64 | 7.10E-03 | -1.91 | 3 |
| PS(18:1/18:0) | 2 | C42H79NO10P- | 7.68 | 788.5443 | 788.5442 | 0.0001 | 0 | -0.60 | 2.17E-02 | -5.75 | 5 |
| PE(O-18:1/20:4) | 2 | C43H77NO7P- | 9.27 | 750.5438 | 750.5438 | 0.0000 | 0 | -0.52 | 3.68E-02 | -1.39 | 1 |
| PE(O-16:1/20:4) | 2 | C41H73NO7P- | 7.01 | 722.5123 | 722.5125 | -0.0002 | 0 | -0.51 | 3.75E-02 | -1.32 | 1 |
| PE(O-16:1/22:4) | 2 | C43H77NO7P- | 8.53 | 750.5442 | 750.5438 | 0.0004 | 1 | -0.52 | 4.97E-02 | -1.31 | 4 |
| Cer(d18:1/23:0) | 2 | C42H82NO5- [M+FA-H]- | 13.76 | 680.6195 | 680.6193 | 0.0002 | 0 | 0.82 | 6.13E-10 | 3.97 | 17 |
| Cer(d18:1/24:0) | 2 | C43H84NO5- [M+FA-H]- | 13.99 | 694.6346 | 694.6350 | -0.0004 | -1 | 0.80 | 1.56E-09 | 5.91 | 25 |
| Cer(d18:1/22:0) | 2 | C41H80NO5- [M+FA-H]- | 13.46 | 666.6030 | 666.6037 | -0.0007 | -1 | 0.79 | 3.36E-09 | 2.98 | 10 |
| oleic acid * | 3 | C18H33O2- | 2.63 | 281.2474 | 281.2481 | -0.0007 | -2 | 0.75 | 9.22E-07 | 3.68 | 7 |
| palmitic acid | 3 | C16H31O2- | 2.51 | 255.2317 | 255.2324 | -0.0007 | -3 | 0.75 | 1.30E-06 | 3.02 | 6 |
| stearic acid | 3 | C18H35O2- | 3.34 | 283.2630 | 283.2637 | -0.0007 | -2 | 0.69 | 1.43E-06 | 2.19 | 7 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | | |
| Cer(d18:1/20:0) | 2 | C39H76NO5- [M+FA-H]- | 12.90 | 638.5727 | 638.5724 | 0.0003 | 0 | 0.59 | 2.87E-05 | 1.78 | 4 |
| DG(14:1/22:5) | 2 | C39H63O5- | 7.60 | 611.4670 | 611.4675 | -0.0005 | -1 | 0.62 | 6.18E-05 | 17.46 | 16 |
| PC(16:0/22:5) & PC(18:1/20:4) | 2 | C47H83NO10P- [M+FA-H]- | 5.89 | 852.5758 | 852.5755 | 0.0003 | 0 | 0.57 | 8.69E-05 | 1.52 | 4 |
| UKN-LP-NEG-369.2422-1.27 | na | ud | 1.27 | 369.2422 | ud | ud | ud | 0.51 | 9.62E-05 | 1.22 | 2 |
| SM(d18:1/23:0) | 2 | C47H94N2O8P- [M+FA-H]- | 13.03 | 845.6696 | 845.6748 | -0.0052 | -6 | 0.59 | 3.49E-04 | 1.34 | 2 |
| linoleic acid † | 3 | C18H31O2- | 2.00 | 279.2316 | 279.2324 | -0.0008 | -3 | 0.54 | 5.58E-04 | 4.63 | 9 |
| Cer(d18:1/24:1) | 2 | C43H82NO5- [M+FA-H]- | 13.49 | 692.6188 | 692.6193 | -0.0005 | -1 | 0.52 | 6.35E-04 | 1.44 | 3 |
| PC(16:0/18:2) | 2 | C43H81NO10P- [M+FA-H]- | 6.01 | 802.5594 | 802.5598 | -0.0004 | 0 | 0.50 | 6.95E-04 | 1.38 | 4 |
| cholesterol sulphate | 4 | C27H45O4S- | 2.27 | 465.3038 | 465.3039 | -0.0001 | 0 | 0.51 | 1.75E-03 | 1.83 | 2 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **HILIC - Aqueous Extracts Positive Mode** | | | | | | | | | | | |
| uridine | 4 | C18H23N4O12Na2+ [2M+2Na-H]+ | 1.49 | 533.1104 | 533.1108 | -0.0004 | -1 | -0.82 | 8.80E-09 | -2.22 | 9 |
| PE(O-18:2/22:6) | 2 | C45H77NO7P+ | 4.29 | 774.5422 | 774.5438 | -0.0016 | -2 | -0.79 | 7.33E-07 | -2.06 | 10 |
| inosine ‡ | 4 | C5H5N4O+ | 3.07 | 137.0513 | 137.0463 | 0.0050 | 36 | -0.86 | 2.01E-06 | -3.68 | 5 |
| PE(O-18:1/22:6) | 2 | C45H79NO7P+ | 4.29 | 776.5578 | 776.5594 | -0.0016 | -2 | -0.76 | 1.64E-05 | -2.39 | 12 |
| PE(O-16:1/22:6) | 2 | C43H75NO7P+ | 4.32 | 748.5262 | 748.5281 | -0.0019 | -3 | -0.57 | 2.32E-05 | -1.55 | 8 |
| PC(O-16:1/20:4) | 2 | C44H81NO7P+ | 4.68 | 766.5750 | 766.5751 | -0.0001 | 0 | -0.66 | 3.53E-05 | -1.59 | 7 |
| 9-(5-Deoxypent-4-enofuranosyl)-9H-purin-6-amine | 1 | C10H12N5O3+ | 1.29 | 250.0924 | 250.0940 | -0.0016 | -6 | -0.70 | 5.42E-05 | -2.73 | ud |
| guanosine ‡ | 4 | C5H6N5O+ | 4.36 | 152.0564 | 152.0572 | -0.0008 | -5 | -0.71 | 8.19E-05 | -2.13 | 4 |
| acetylcarnitine | 4 | C9H17NO4Na+ [M+Na]+ | 6.99 | 226.1078 | 226.1055 | 0.0023 | 10 | -0.74 | 9.55E-05 | -3.30 | 17 |
| PE(14:0/22:4) | 2 | C41H75NO8P+ | 4.40 | 740.5224 | 740.5230 | -0.0006 | -1 | -0.55 | 1.27E-04 | -1.77 | 15 |
| PE(O-18:2/20:4) | 2 | C43H77NO7P+ | 4.31 | 750.5426 | 750.5438 | -0.0012 | -2 | -0.64 | 1.28E-04 | -1.50 | 6 |
| hydroxybutyrylcarnitine | 2 | C11H22NO5+ | 7.26 | 248.1488 | 248.1498 | -0.0010 | -4 | -0.76 | 3.47E-04 | -2.49 | 8 |
| PC(O-18:2/20:4) | 2 | C46H83NO7P+ | 4.61 | 792.5836 | 792.5907 | -0.0071 | -9 | -0.71 | 3.60E-04 | -1.87 | 7 |
| UKN-HL-POS-84.0801-5.16 | na | ud | 5.16 | 84.0801 | ud | ud | ud | -0.71 | 8.19E-04 | ud | 20 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **HILIC - Aqueous Extracts Positive Mode** | | | | | | | | | | | |
| UKN-HL-POS-86.0894-4.52 | na | ud | 4.52 | 86.0894 | ud | ud | ud | -0.56 | 9.14E-04 | -3.07 | 8 |
| PE(O-16:1/20:4) | 2 | C41H75NO7P+ | 4.34 | 724.5275 | 724.5281 | -0.0006 | -1 | -0.51 | 1.15E-03 | -1.49 | 9 |
| PC(O-18:1/20:4) | 2 | C46H85NO7P+ | 4.70 | 794.6056 | 794.6064 | -0.0008 | -1 | -0.58 | 1.20E-03 | -1.64 | 7 |
| UKN-HL-POS-198.0887-8.22 | na | ud | 8.22 | 198.0887 | ud | ud | ud | -0.58 | 2.24E-03 | ud | 11 |
| hypoxanthine | 4 | C5H5N4O+ | 2.20 | 137.0491 | 137.0463 | 0.0028 | 20 | -0.60 | 3.54E-03 | -1.83 | 12 |
| PC(14:0/16:0) | 2 | C38H77NO8P+ | 4.85 | 706.5375 | 706.5387 | -0.0012 | -2 | -0.58 | 4.01E-03 | -1.45 | 8 |
| propionylcarnitine | 4 | C10H20NO4+ | 6.72 | 218.1379 | 218.1392 | -0.0013 | -6 | -0.60 | 8.28E-03 | -1.25 | 8 |
| valerylcarnitine | 4 | C12H24NO4+ | 6.23 | 246.1693 | 246.1705 | -0.0012 | -5 | -0.62 | 8.58E-03 | -1.42 | 8 |
| 1-methylnicotinamide § | 4 | C7H9N2O+ | 5.65 | 137.0772 | 137.0715 | 0.0057 | 42 | -0.63 | 9.89E-03 | -1.45 | 10 |
| butyrylcarnitine | 4 | C11H22NO4+ | 6.46 | 232.1537 | 232.1549 | -0.0012 | -5 | -0.59 | 1.00E-02 | -1.92 | 6 |
| dodecanoylcarnitine | 4 | C19H38NO4+ | 5.16 | 344.2801 | 344.2801 | 0.0000 | 0 | 0.67 | 1.33E-06 | 2.95 | 9 |
| decanoylcarnitine | 4 | C17H34NO4+ | 5.28 | 316.2485 | 316.2488 | -0.0003 | -1 | 0.64 | 3.58E-06 | 3.40 | 18 |
| adenosine | 4 | C10H14N5O4+ | 2.09 | 268.1034 | 268.1046 | -0.0012 | -4 | 0.64 | 4.63E-06 | 41.98 | 26 |
| UKN-HL-POS-90.9605-5.50 | na | ud | 5.50 | 90.9605 | ud | ud | ud | 0.59 | 1.25E-05 | 1.32 | 6 |
| | | | | | | | | | | | |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **HILIC - Aqueous Extracts Negative Mode** | | | | | | | | | | | |
| uridine | 4 | C9H11N2O6- | 1.49 | 243.0613 | 243.0617 | -0.0004 | -2 | -0.93 | 1.80E-09 | -1.99 | 5 |
| inosine | 4 | C10H11N4O5- | 3.05 | 267.0725 | 267.0729 | -0.0004 | -2 | -0.94 | 2.49E-09 | -3.68 | 7 |
| PE(O-16:1/20:4) | 2 | C41H73NO7P- | 4.31 | 722.5135 | 722.5125 | 0.0010 | 1 | -0.88 | 3.95E-09 | -2.11 | 7 |
| PE(O-18:2/20:4) | 2 | C43H75NO7P- | 4.29 | 748.5289 | 748.5271 | 0.0018 | 2 | -0.90 | 8.29E-09 | -2.76 | 7 |
| PE(18:0/22:4) | 2 | C45H81NO8P- | 4.35 | 794.5699 | 794.5700 | -0.0001 | 0 | -0.83 | 1.24E-08 | -2.27 | 11 |
| deoxyuridine | 3 | C9H11N2O5- | 1.22 | 227.0716 | 227.0668 | 0.0048 | 21 | -0.76 | 5.57E-08 | -2.92 | 11 |
| PE(18:1/22:4) & PE(18:0/22:5) | 2 | C45H79NO8P- | 4.35 | 792.5514 | 792.5543 | -0.0029 | -4 | -0.84 | 1.27E-07 | -1.76 | 12 |
| PE(O-18:2/22:6) | 2 | C45H75NO7P- | 4.27 | 772.5295 | 772.5281 | 0.0014 | 2 | -0.91 | 1.52E-07 | -3.20 | 9 |
| PE(O-16:1/22:6) | 2 | C43H73NO7P- | 4.29 | 746.5135 | 746.5125 | 0.0010 | 1 | -0.86 | 1.86E-07 | -2.94 | 9 |
| PE(O-18:1/20:4) | 2 | C43H77NO7P- | 4.28 | 750.5438 | 750.5438 | 0.0000 | 0 | -0.84 | 2.08E-07 | -1.70 | 8 |
| PE(18:0/20:4) | 2 | C43H77NO8P- | 4.35 | 766.5390 | 766.5387 | 0.0003 | 0 | -0.75 | 2.42E-07 | -1.56 | 9 |
| PE(O-18:1/22:6) | 2 | C45H77NO7P- | 4.27 | 774.5441 | 774.5438 | 0.0003 | 0 | -0.88 | 1.77E-06 | -2.69 | 5 |
| PC(O-16:1/20:4) | 2 | C45H81NO9P- [M+FA-H]- | 4.67 | 810.5660 | 810.5649 | 0.0011 | 1 | -0.71 | 6.91E-06 | -1.91 | 5 |
| PC(O-18:2/20:4) | 2 | C47H83NO9P- | 4.65 | 836.5820 | 836.5805 | 0.0015 | 2 | -0.73 | 7.13E-06 | -1.87 | 9 |
| guanosine | 4 | C10H12N5O5- | 4.33 | 282.0835 | 282.0838 | -0.0003 | -1 | -0.76 | 4.07E-05 | -1.81 | 4 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | RT (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | Fold Change | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **HILIC - Aqueous Extracts Negative Mode** | | | | | | | | | | | |
| arachidonic acid | 3 | C20H31O2- | 0.69 | 303.2324 | 303.2324 | 0.0000 | 0 | -0.64 | 6.97E-05 | -2.16 | 21 |
| UKN-HL-NEG-194.1004-5.17 | na | ud | 5.17 | 194.1004 | ud | ud | ud | -0.67 | 7.20E-05 | -3.35 | 9 |
| N-acetylmethionine | 4 | C7H12NO3S- | 4.33 | 190.0536 | 190.0538 | -0.0002 | -1 | -0.82 | 7.66E-05 | -2.88 | 9 |
| methyluridine | 4 | C10H13N2O6- | 1.36 | 257.0800 | 257.0774 | 0.0026 | 10 | -0.76 | 3.45E-04 | -1.74 | 10 |
| PC(O-18:1/20:4) | 2 | C47H85NO9P- [M+FA-H]- | 4.68 | 838.5983 | 838.5962 | 0.0021 | 3 | -0.59 | 4.89E-04 | -1.59 | 7 |
| uracil | 4 | C4H3N2O2- | 1.00 | 111.0061 | 111.0195 | -0.0134 | -121 | -0.69 | 1.88E-03 | -2.71 | 4 |
| 1-(beta-D-Ribofuranosyl)-1,4-dihydronicotinamide | 1 | C11H15N2O5- | 1.96 | 255.0979 | 255.0981 | -0.0002 | -1 | -0.62 | 5.75E-03 | ud | 20 |
| hypoxanthine | 4 | C5H4N4OCl- [M+Cl]- | 2.11 | 171.0134 | 171.0074 | 0.0060 | 35 | -0.59 | 6.31E-03 | -1.45 | 4 |
| UKN-HL-NEG-112.959-5.49 | na | ud | 5.49 | 112.9590 | ud | ud | ud | 0.78 | 1.12E-06 | 1.24 | 4 |
| UKN-HL-NEG-95.0241-1.02 | na | ud | 1.02 | 95.0241 | ud | ud | ud | 0.70 | 2.13E-06 | 3.71 | 11 |
| adenosine | 4 | C10H13N5O4Cl- [M+Cl]- | 2.08 | 302.0650 | 302.0656 | -0.0006 | -2 | 0.55 | 2.49E-05 | 31.69 | 10 |
| UKN-HL-NEG-204.0471-1.36 | na | ud | 1.36 | 204.0471 | ud | ud | ud | 0.65 | 2.67E-05 | ud | 8 |
| benzoic acid | 4 | C7H5O2- | 1.03 | 121.0288 | 121.0290 | -0.0002 | -2 | 0.58 | 7.74E-04 | 1.40 | 23 |

* Standard compound of elaidic acid (stereoisomer of oleic acid) exhibited different RT

† Standard compound of linoelaidic acid (stereoisomer of linoleic acid) exhibited different RT

‡ Detected as a fragment of parent ion
§ Standard compound of 6-methylnicotinamide exhibited different RT

LoA: Level of Assignment; 1: Tentative assignment; 2: Tandem MS spectrum matched to database or literature; 3: RT matched to standard compound; 4: MS/MS spectrum matched to standard compound.
RT: Retention time.
HILIC: Hydrophilic Interaction (Liquid) Chromatography.
P(corr) refers to the correlation coefficient of metabolites to disease classes.
Two-tailed t-tests were conducted, assuming unequal variance.
Coefficient of variation (CV%) is calculated based on thirteen injections of the same pooled quality control (QC) sample, acquired throughout the run.
The position of double bonds and the position of fatty acyl chain cannot be determined in lipid moieties (Cer, DG, HexCer, PC, PE, PG, PI, SM, and TG). Fatty acyl chains are presented from lowest to highest molecular weight (lower number of carbons to higher, and higher number of double bonds to lower).
Cer: Ceramide; DG: diglyceride; FA: formate; Hex: Hexosyl; na: not applicable; oxCE: oxidised cholesterol ester; PC: phosphatidylcholine; PE: Phosphatidylethanolamine; PG: Phosphatidylglycerol; PI: phosphatidylinositol; SM: sphingomyelin; TG: triglyceride; ud: unable to determine; UKN: unknown.

*Table 4. 8: List of structurally assigned significant metabolites obtained from analysis of tissue extracts, after comparison of the carotid to the femoral plaque group.*

| Met Name | LoA | Mol Formula (as detected) [adduct] | Ret Time (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | p-value [t-test] (INT-CAR vs INT-FEM) | Fold Change (CAR/FEM) | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | |
| **Lipid Profiling Positive Mode** | | | | | | | | | | | | |
| lysoPC(O-18:1) | 2 | C26H55NO6P+ | 1.46 | 508.3779 | 508.3767 | 0.0012 | 2 | 0.73 | 3.56E-16 | 0.02 | 3.79 | 6 |
| SM(d18:1/14:0) & SM(d16:1/16:0) | 2 | C37H76N2O6P+ | 4.15 | 675.5445 | 675.5441 | 0.0004 | 1 | 0.78 | 8.00E-15 | 0.39 | 2.91 | 8 |
| lysoPC(18:2) | 2 | C26H51NO7P+ | 1.02 | 520.3419 | 520.3403 | 0.0016 | 3 | 0.71 | 9.55E-13 | 0.24 | 3.14 | 5 |
| lysoPC(18:1) | 2 | C26H53NO7P+ | 1.25 | 522.3580 | 522.3560 | 0.0020 | 4 | 0.71 | 1.38E-12 | 0.10 | 2.49 | 5 |
| SM(d17:1/16:0) | 2 | C38H78N2O6P+ | 4.71 | 689.5601 | 689.5598 | 0.0003 | 0 | 0.75 | 1.07E-11 | 0.98 | 2.44 | 6 |
| SM(d18:2/24:1) | 2 | C47H92N2O6P+ | 9.74 | 811.6693 | 811.6693 | 0.0000 | 0 | 0.72 | 1.36E-11 | 0.50 | 2.37 | 7 |
| SM(d18:2/16:0) | 2 | C39H78N2O6P+ | 4.31 | 701.5600 | 701.5598 | 0.0002 | 0 | 0.79 | 4.53E-11 | 0.48 | 2.55 | 7 |
| PC(16:0/16:0) | 2 | C40H81NO8P+ | 7.17 | 734.5706 | 734.5700 | 0.0006 | 1 | 0.62 | 7.46E-11 | 0.73 | 2.14 | 7 |
| SM(d18:0/16:0)* | 2 | C39H82N2O6P+ | 5.95 | 706.5937 | 706.5944 | -0.0007 | -1 | 0.79 | 1.16E-10 | 0.69 | 2.35 | 6 |
| SM(d18:1/17:0) | 2 | C40H82N2O6P+ | 6.15 | 717.5917 | 717.5911 | 0.0006 | 1 | 0.75 | 2.20E-09 | 0.85 | 2.21 | 3 |
| PC(16:0/16:1) | 2 | C40H79NO8P+ | 5.66 | 732.5551 | 732.5543 | 0.0008 | 1 | 0.58 | 6.04E-09 | 0.39 | 2.20 | 3 |
| PC(O-16:1/20:4) | 2 | C44H81NO7P+ | 6.45 | 766.5751 | 766.5751 | 0.0000 | 0 | 0.57 | 8.40E-08 | 0.10 | 1.64 | 4 |
| lysoPC(16:0) | 2 | C24H51NO7P+ | 1.20 | 496.3424 | 496.3403 | 0.0021 | 4 | 0.57 | 5.84E-07 | 0.25 | 2.12 | 8 |
| tetraHexCer(d18:1/16:0) | 2 | C58H108NO21+ | 9.25 | 1154.7557 | 1154.7414 | 0.0143 | 12 | 0.65 | 2.24E-06 | 0.52 | 4.85 | 10 |
| triHexCer(d18:1/16:0) | 2 | C52H97NO18Na+ [M+Na]+ | 5.16 | 1046.6612 | 1046.6603 | 0.0009 | 1 | 0.52 | 8.01E-05 | 0.33 | 1.74 | 9 |
| SM(d18:2/24:0) | 2 | C47H94N2O6P+ | 12.82 | 813.6861 | 813.6850 | 0.0011 | 1 | 0.55 | 3.21E-04 | 0.80 | 1.45 | 8 |
| SM(d18:1/24:1) | 2 | C47H94N2O6P+ | 12.23 | 813.6857 | 813.6850 | 0.0007 | 1 | 0.52 | 2.64E-03 | 0.26 | 1.26 | 13 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | Ret Time (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | p-value [t-test] (INT-CAR vs INT-FEM) | Fold Change (CAR/FEM) | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | |
| **Lipid Profiling Positive Mode** | | | | | | | | | | | | |
| TG(16:0/18:1/18:1) | 2 | C55H106NO6+ [M+NH4]+ | 15.69 | 876.8013 | 876.8020 | -0.0007 | -1 | -0.88 | 5.11E-08 | 0.60 | -3.75 | 4 |
| TG(16:0/18:1/18:2) & TG(16:1/18:1/18:1) | 2 | C55H104NO6+ [M+NH4]+ | 15.47 | 874.7869 | 874.7864 | 0.0005 | 1 | -0.87 | 1.88E-07 | 0.29 | -3.57 | 4 |
| TG(16:0/16:0/18:1) | 2 | C53H104NO6+ [M+NH4]+ | 15.65 | 850.7849 | 850.7864 | -0.0015 | -2 | -0.85 | 2.86E-07 | 0.85 | -3.42 | 2 |
| TG(16:0/16:1/18:1) | 2 | C53H102NO6+ [M+NH4]+ | 15.44 | 848.7717 | 848.7707 | 0.0010 | 1 | -0.85 | 3.86E-07 | 0.82 | -3.94 | 4 |
| PC(16:0/18:2) | 2 | C42H80NO8PNa+ [M+Na]+ | 6.01 | 780.5527 | 780.5519 | 0.0008 | 1 | -0.60 | 1.09E-06 | 0.89 | -1.89 | 2 |
| oxCE(C18/O) | 2 | C45H76O3Na+ [M+Na]+ | 13.94 | 687.5696 | 687.5692 | 0.0004 | 1 | -0.66 | 1.32E-06 | 0.08 | -3.52 | 7 |
| PC(16:0/20:4) | 2 | C44H80NO8PNa+ [M+Na]+ | 5.75 | 804.5520 | 804.5519 | 0.0001 | 0 | -0.59 | 1.39E-06 | 0.61 | -1.89 | 14 |
| UKN-LP-POS-895.7366-14.23 | na | ud | 14.23 | 895.7366 | ud | ud | ud | -0.75 | 2.11E-06 | 0.01 | -2.42 | 9 |
| TG(18:1/18:1/18:1) | 2 | C57H108NO6+ [M+NH4]+ | 15.69 | 902.8174 | 902.8177 | -0.0003 | 0 | -0.82 | 2.11E-06 | 0.27 | -3.68 | 3 |
| PC(16:0/18:1) | 2 | C42H82NO8PNa+ [M+Na]+ | 7.39 | 782.5682 | 782.5676 | 0.0006 | 1 | -0.54 | 4.57E-06 | 0.85 | -1.72 | 4 |
| TG(16:1/16:1/18:1) & TG(16:0/16:1/18:2) & TG(14:0/18:1/18:2) | 2 | C53H100NO6+ [M+NH4]+ | 15.16 | 846.7553 | 846.7551 | 0.0002 | 0 | -0.80 | 6.51E-06 | 0.57 | -3.63 | 3 |
| TG(18:0/18:1/18:1) | 2 | C57H110NO6+ [M+NH4]+ | 15.89 | 904.8300 | 904.8333 | -0.0033 | -4 | -0.80 | 7.91E-06 | 0.42 | -4.13 | 6 |
| UKN-LP-POS-621.3115-2.88 | na | C26H41N10O8+ † | 2.88 | 621.3115 | ud | ud | ud | -0.61 | 9.09E-06 | 0.27 | -5.13 | 2 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | Ret Time (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | p-value [t-test] (INT-CAR vs INT-FEM) | Fold Change (CAR/FEM) | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | |
| **Lipid Profiling Positive Mode** | | | | | | | | | | | | |
| TG(18:1/18:1/18:2) | 2 | C57H106NO6+ [M+NH4]+ | 15.46 | 900.8016 | 900.8020 | -0.0004 | 0 | -0.78 | 1.17E-05 | 0.18 | -2.82 | 2 |
| oxCE(C18/O) | 2 | C45H74O3Na+ [M+Na]+ | 13.99 | 685.5541 | 685.5536 | 0.0005 | 1 | -0.62 | 1.57E-05 | 0.15 | -2.29 | 4 |
| TG(14:0/16:1/18:1) & TG(14:1/16:0/18:1) & TG(14:0/16:0/18:2) & TG(12:0/18:0/18:2) & TG(12:0/18:1/18:1) | 2 | C51H98NO6+ [M+NH4]+ | 15.12 | 820.7395 | 820.7394 | 0.0001 | 0 | -0.73 | 6.72E-05 | 0.50 | -3.88 | 2 |
| TG(18:1/18:2/18:2) & TG(18:1/18:1/18:3) & TG(16:0/18:2/20:3) | 2 | C57H104NO6+ [M+NH4]+ | 15.22 | 898.7860 | 898.7864 | -0.0004 | 0 | -0.67 | 1.25E-04 | 0.41 | -1.99 | 2 |
| | | | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | | | |
| lysoPC(O-18:1) | 2 | C27H55NO8P- [M+FA-H]- | 1.47 | 552.3669 | 552.3665 | 0.0004 | 1 | 0.80 | 6.63E-16 | 0.01 | 2.50 | 4 |
| UKN-LP-NEG-869.6484-10.89 | na | ud | 10.89 | 869.6484 | ud | ud | ud | 0.86 | 2.74E-15 | 0.74 | 32.57 | 5 |
| lysoPC(O-16:0) | 2 | C25H53NO8P- [M+FA-H]- | 1.39 | 526.3508 | 526.3509 | -0.0001 | 0 | 0.72 | 6.88E-15 | 0.02 | 2.08 | 3 |
| UKN-LP-NEG-853.6442-7.74 | na | ud | 7.74 | 853.6442 | ud | ud | ud | 0.77 | 1.64E-13 | 0.69 | 4.29 | 18 |
| SM(d18:1/14:0) & SM(d16:1/16:0) | 2 | C38H76N2O8P- [M+FA-H]- | 4.15 | 719.5339 | 719.5339 | 0.0000 | 0 | 0.75 | 8.04E-11 | 0.45 | 1.54 | 3 |
| lysoPE(O-16:1) | 2 | C21H43NO6P- | 1.42 | 436.2825 | 436.2828 | -0.0003 | -1 | 0.77 | 1.41E-10 | 0.56 | 2.76 | 19 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | Ret Time (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | p-value [t-test] (INT-CAR vs INT-FEM) | Fold Change (CAR/FEM) | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | | | |
| HexCer(d18:1/24:1) | 2 | C49H92NO10-[M+FA-H]- | 12.97 | 854.6717 | 854.6721 | -0.0004 | 0 | 0.76 | 3.66E-10 | 0.56 | 1.98 | 4 |
| lysoPC(18:2) | 2 | C27H51NO9P-[M+FA-H]- | 1.03 | 564.3299 | 564.3301 | -0.0002 | 0 | 0.70 | 1.08E-09 | 0.26 | 1.96 | 3 |
| lysoPC(18:1) | 2 | C27H53NO9P-[M+FA-H]- | 1.26 | 566.3455 | 566.3458 | -0.0003 | -1 | 0.75 | 8.32E-09 | 0.12 | 1.75 | 3 |
| SM(d18:2/24:1) | 2 | C48H92N2O8P-[M+FA+H]- | 9.74 | 855.6589 | 855.6591 | -0.0002 | 0 | 0.61 | 1.81E-08 | 0.60 | 1.28 | 3 |
| SM(d18:1/16:0) & SM(d16:1/18:0) | 2 | C40H80N2O8P-[M+FA-H]- | 5.39 | 747.5648 | 747.5652 | -0.0004 | -1 | 0.72 | 2.63E-08 | 0.06 | 1.29 | 5 |
| PC(O-18:1/16:0) | 2 | C43H85NO9P-[M+FA-H]- | 8.65 | 790.5972 | 790.5962 | 0.0010 | 1 | 0.75 | 7.56E-08 | 0.77 | 1.64 | 11 |
| SM(d18:2/16:0) | 2 | C40H78N2O8P-[M+FA-H]- | 4.32 | 745.5492 | 745.5496 | -0.0004 | -1 | 0.69 | 1.76E-07 | 0.52 | 1.42 | 4 |
| SM(d17:1/16:0) | 2 | C39H78N2O8P-[M+FA-H]- | 4.73 | 733.5492 | 733.5496 | -0.0004 | -1 | 0.57 | 1.32E-06 | 0.79 | 1.44 | 2 |
| lysoPE(O-18:1) | 2 | C23H47NO6P- | 1.98 | 464.3139 | 464.3141 | -0.0002 | 0 | 0.57 | 9.01E-06 | 0.60 | 1.64 | 5 |
| PG(18:2/18:1) | 2 | C42H76O10P- | 4.35 | 771.5186 | 771.5176 | 0.0010 | 1 | 0.60 | 9.52E-06 | 0.53 | 3.02 | 13 |
| PE-Cer(d18:1/16:0) | 2 | C36H72N2O6P- | 5.65 | 659.5125 | 659.5128 | -0.0003 | 0 | 0.61 | 2.67E-04 | 0.94 | 1.58 | 3 |
| lysoPC(16:0) | 2 | C25H51NO9P-[M+FA-H]- | 1.20 | 540.3299 | 540.3301 | -0.0002 | 0 | 0.52 | 1.55E-03 | 0.33 | 1.57 | 5 |
| lysoPE(18:0) | 2 | C23H47NO7P- | 1.73 | 480.3092 | 480.3090 | 0.0002 | 0 | 0.55 | 1.89E-03 | 0.79 | 1.39 | 3 |
| PI(18:0/22:6) | 2 | C49H82O13P- | 5.43 | 909.5505 | 909.5493 | 0.0012 | 1 | 0.51 | 1.35E-02 | 0.44 | 2.93 | 6 |
| SM(d18:1/22:0) | 2 | C46H92N2O8P-[M+FA-H]- | 12.42 | 831.6592 | 831.6591 | 0.0001 | 0 | -0.86 | 2.19E-09 | 0.19 | -1.75 | 27 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | Ret Time (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | p-value [t-test] (INT-CAR vs INT-FEM) | Fold Change (CAR/FEM) | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | | | |
| UKN-LP-NEG-815.5537-5.39 | na | ud | 5.39 | 815.5537 | ud | ud | ud | -0.85 | 4.77E-08 | 0.44 | -1.79 | 8 |
| SM(d18:1/24:0) | 2 | C48H96N2O8P- [M+FA-H]- | 13.33 | 859.6903 | 859.6904 | -0.0001 | 0 | -0.81 | 6.01E-08 | 0.83 | -1.46 | 6 |
| SM(d16:1/22:0) | 2 | C44H88N2O8P- [M+FA-H]- | 9.47 | 803.6281 | 803.6278 | 0.0003 | 0 | -0.75 | 2.33E-07 | 0.14 | -1.54 | 13 |
| SM(d18:1/23:0) | 2 | C47H94N2O8P- [M+FA-H]- | 13.03 | 845.6696 | 845.6748 | -0.0052 | -6 | -0.81 | 6.70E-07 | 0.22 | -1.44 | 2 |
| SM(d18:1/18:0) | 2 | C42H84N2O8P- [M+FA-H]- | 7.12 | 775.5965 | 775.5965 | 0.0000 | 0 | -0.71 | 1.20E-06 | 0.15 | -1.49 | 3 |
| oleic acid ‡ | 3 | C18H33O2- | 2.63 | 281.2474 | 281.2481 | -0.0007 | -2 | -0.81 | 1.24E-06 | 0.15 | -3.89 | 7 |
| stearic acid | 3 | C18H35O2- | 3.34 | 283.2630 | 283.2637 | -0.0007 | -2 | -0.80 | 1.58E-06 | 0.75 | -2.22 | 7 |
| palmitic acid | 3 | C16H31O2- | 2.51 | 255.2317 | 255.2324 | -0.0007 | -3 | -0.82 | 1.73E-06 | 0.31 | -2.99 | 6 |
| Cer(d18:1/24:0) | 2 | C43H84NO5- [M+FA-H]- | 13.99 | 694.6346 | 694.6350 | -0.0004 | -1 | -0.70 | 3.81E-06 | 0.58 | -1.70 | 25 |
| UKN-LP-NEG-369.2422-1.27 | na | ud | 1.27 | 369.2422 | ud | ud | ud | -0.65 | 7.87E-06 | 0.01 | -1.30 | 2 |
| Cer(d18:1/22:0) | 2 | C41H80NO5- [M+FA-H]- | 13.46 | 666.6030 | 666.6037 | -0.0007 | -1 | -0.72 | 1.94E-05 | 0.58 | -1.47 | 10 |
| PC(18:1/18:0) | 2 | C45H87NO10P- [M+FA-H]- | 9.77 | 832.6065 | 832.6068 | -0.0003 | 0 | -0.55 | 2.19E-05 | 0.38 | -1.50 | 30 |
| Cer(d18:1/23:0) | 2 | C42H82NO5- [M+FA-H]- | 13.76 | 680.6195 | 680.6193 | 0.0002 | 0 | -0.68 | 2.93E-05 | 0.86 | -1.58 | 17 |
| PC(16:0/20:4) | 2 | C45H81NO10P- [M+FA-H]- | 5.77 | 826.5596 | 826.5598 | -0.0002 | 0 | -0.67 | 5.15E-05 | 0.54 | -1.54 | 3 |
| PC(16:0/22:5) & PC(18:1/20:4) | 2 | C47H83NO10P- [M+FA-H]- | 5.89 | 852.5758 | 852.5755 | 0.0003 | 0 | -0.61 | 1.17E-04 | 0.36 | -1.49 | 4 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | Ret Time (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | p-value [t-test] (INT-CAR vs INT-FEM) | Fold Change (CAR/FEM) | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | |
| **Lipid Profiling Negative Mode** | | | | | | | | | | | | |
| PC(16:0/18:1)* | 2 | C43H83NO10P- [M+FA-H]- | 7.41 | 805.5787 | 805.5788 | -0.0001 | 0 | -0.50 | 1.76E-04 | 0.46 | -1.31 | 2 |
| arachidonic acid | 3 | C20H31O2- | 1.87 | 303.2317 | 303.2324 | -0.0007 | -2 | -0.65 | 2.03E-04 | 0.35 | -1.76 | 3 |
| PC(16:0/18:2) | 2 | C43H81NO10P- [M+FA-H]- | 6.01 | 802.5594 | 802.5598 | -0.0004 | 0 | -0.50 | 2.79E-04 | 0.13 | -1.33 | 4 |
| PC(18:0/20:4) | 2 | C47H85NO10P- [M+FA-H]- | 7.59 | 854.5937 | 854.5911 | 0.0026 | 3 | -0.56 | 5.17E-04 | 0.97 | -1.28 | 2 |
| linoleic acid § | 3 | C18H31O2- | 2.00 | 279.2316 | 279.2324 | -0.0008 | -3 | -0.62 | 1.39E-03 | 0.21 | -3.83 | 9 |
| Cer(d18:1/20:0) | 2 | C39H76NO5- [M+FA-H]- | 12.90 | 638.5727 | 638.5724 | 0.0003 | 0 | -0.56 | 1.60E-03 | 0.77 | -1.49 | 4 |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| **HILIC Aqueous Positive Mode** | | | | | | | | | | | | |
| lysoPC(O-18:2) | 2 | C26H53NO6P+ | 6.21 | 506.3612 | 506.3610 | 0.0002 | 0 | 0.73 | 1.57E-12 | 0.18 | Ud | 7 |
| lysoPC(20:2) | 2 | C28H54NO7PNa+ [M+Na]+ | 5.84 | 570.3557 | 570.3536 | 0.0021 | 4 | 0.67 | 1.11E-11 | 0.48 | 2.94 | 19 |
| lysoPC(20:4) | 2 | C28H50NO7PNa+ [M+Na]+ | 5.86 | 566.3230 | 566.3223 | 0.0007 | 1 | 0.67 | 1.95E-10 | 0.25 | 2.55 | 22 |
| lysoPC(O-18:1) | 2 | C26H55NO6P+ | 6.16 | 508.3759 | 508.3767 | -0.0008 | -2 | 0.58 | 3.47E-10 | 0.48 | 3.00 | 12 |
| lysoPC(18:1) | 2 | C26H53NO7P+ | 5.97 | 522.3558 | 522.3560 | -0.0002 | 0 | 0.63 | 1.35E-09 | 0.21 | 3.44 | 7 |
| PC(O-18:2/20:4) | 2 | C46H83NO7P+ | 4.61 | 792.5836 | 792.5907 | -0.0071 | -9 | 0.66 | 1.81E-09 | 0.09 | 3.10 | 7 |
| UKN-HL-POS-876.6866-4.66 | na | ud | 4.66 | 876.6866 | ud | ud | ud | 0.69 | 3.66E-09 | 0.79 | 2.91 | 7 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | Ret Time (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | p-value [t-test] (INT-CAR vs INT-FEM) | Fold Change (CAR/FEM) | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | |
| **HILIC Aqueous Positive Mode** | | | | | | | | | | | | |
| lysoPC(16:1) | 2 | C24H49NO7P+ | 6.13 | 494.3233 | 494.3247 | -0.0014 | -3 | 0.64 | 4.06E-09 | 0.18 | 2.92 | 4 |
| lysoPC(18:2) | 2 | C26H50NO7PK+ [M+K]+ | 5.44 | 558.2979 | 558.2962 | 0.0017 | 3 | 0.59 | 5.14E-09 | 0.61 | 3.28 | 13 |
| PC(16:2/16:0) | 2 | C40H77NO8P+ | 4.86 | 730.5435 | 730.5387 | 0.0048 | 7 | 0.64 | 3.15E-08 | 0.83 | 2.02 | 4 |
| sphingosine | 4 | C18H36NO+ [M-H2O+H]+ | 2.73 | 282.2786 | 282.2797 | -0.0011 | -4 | 0.56 | 4.05E-08 | 0.91 | 2.82 | 6 |
| PC(O-18:1/20:4) | 2 | C46H85NO7P+ | 4.70 | 794.6056 | 794.6064 | -0.0008 | -1 | 0.62 | 8.30E-08 | 0.48 | 1.95 | 7 |
| lysoPC(14:0) | 2 | C22H47NO7P+ | 6.19 | 468.3083 | 468.3090 | -0.0007 | -1 | 0.53 | 9.51E-08 | 0.32 | 1.96 | 8 |
| PE(O-18:2/22:6) | 2 | C45H77NO7P+ | 4.29 | 774.5422 | 774.5438 | -0.0016 | -2 | 0.63 | 1.61E-07 | 0.27 | 1.89 | 10 |
| PC(O-18:2/22:5) | 2 | C48H85NO7P+ | 4.68 | 818.6045 | 818.6064 | -0.0019 | -2 | 0.59 | 1.63E-07 | 0.72 | 2.76 | 6 |
| PC(O-16:1/22:6) | 2 | C46H81NO7P+ | 4.67 | 790.5873 | 790.5751 | 0.0122 | 15 | 0.55 | 3.12E-07 | 0.33 | 1.80 | 6 |
| PC(O-16:1/20:4) | 2 | C44H81NO7P+ | 4.68 | 766.5750 | 766.5751 | -0.0001 | 0 | 0.63 | 4.13E-07 | 0.03 | 1.50 | 7 |
| PC(14:0/16:1) | 2 | C38H75NO8P+ | 4.84 | 704.5253 | 704.5230 | 0.0023 | 3 | 0.60 | 4.48E-07 | 0.89 | 1.82 | 10 |
| PC(O-18:2/22:4) | 2 | C48H87NO7P+ | 4.69 | 820.6253 | 820.6220 | 0.0033 | 4 | 0.59 | 9.77E-07 | 0.65 | 1.94 | 6 |
| PE(O-18:2/20:4) | 2 | C43H77NO7P+ | 4.31 | 750.5426 | 750.5438 | -0.0012 | -2 | 0.58 | 6.50E-06 | 0.11 | 1.46 | 6 |
| sphinganine | 4 | C18H38NO+ [M-H2O+H]+ | 2.48 | 284.2946 | 284.2953 | -0.0007 | -2 | 0.52 | 1.61E-05 | 0.61 | Ud | 5 |
| PC(O-18:2/22:6) | 2 | C48H83NO7P+ | 4.66 | 816.5887 | 816.5907 | -0.0020 | -2 | 0.55 | 2.63E-05 | 0.19 | 1.89 | 9 |
| PE(O-16:1/22:6) | 2 | C43H75NO7P+ | 4.32 | 748.5262 | 748.5281 | -0.0019 | -3 | 0.51 | 3.55E-05 | 0.03 | 1.58 | 8 |
| inosine ¶ | 4 | C5H5N4O+ | 3.07 | 137.0513 | 137.0463 | 0.0050 | 36 | 0.52 | 1.28E-04 | 0.10 | 1.90 | 5 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | Ret Time (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | p-value [t-test] (INT-CAR vs INT-FEM) | Fold Change (CAR/FEM) | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | |
| **HILIC Aqueous Positive Mode** | | | | | | | | | | | | |
| UKN-HL-POS-90.9605-5.50 | na | ud | 5.50 | 90.9605 | ud | ud | ud | -0.79 | 1.56E-09 | 0.10 | -1.43 | 6 |
| creatine | 4 | C4H10N3O2+ | 6.52 | 132.0764 | 132.0773 | -0.0009 | -7 | -0.59 | 8.65E-06 | 0.13 | -1.82 | 7 |
| SM(d18:2/16:0) | 2 | C39H78N2O6P+ | 5.28 | 701.5591 | 701.5598 | -0.0007 | -1 | -0.71 | 1.45E-05 | 0.20 | -7.17 | 10 |
| adenosine | 4 | C10H14N5O4+ | 2.09 | 268.1034 | 268.1046 | -0.0012 | -4 | -0.64 | 1.89E-05 | 0.21 | -9.05 | 26 |
| UKN-HL-POS-226.0682-1.42 | na | ud | 1.42 | 226.0682 | ud | ud | ud | -0.70 | 8.87E-05 | 0.01 | -6.98 | Ud |
| UKN-HL-POS-162.0509-6.09 | na | ud | 6.09 | 162.0509 | ud | ud | ud | -0.58 | 1.07E-04 | 0.12 | -1.84 | 14 |
| betaine | 4 | C5H12NO2+ | 6.18 | 118.0889 | 118.0868 | 0.0021 | 18 | -0.49 | 1.69E-04 | 0.26 | -1.28 | 12 |
| UKN-HL-POS-254.1603-5.64 | na | ud | 5.64 | 254.1603 | 254.1617 | -0.0014 | -6 | -0.57 | 1.04E-03 | 0.35 | -3.02 | 7 |
| niacinamide ¶ | 4 | C6H4NO+ | 1.23 | 106.0305 | 106.0293 | 0.0011 | 10 | -0.50 | 7.41E-03 | 0.37 | -3.18 | 11 |
| | | | | | | | | | | | | |
| **HILIC Aqueous Negative Mode** | | | | | | | | | | | | |
| lysoPC(18:2) | 2 | C27H51NO9P-[M+FA-H]- | 5.97 | 564.3307 | 564.3301 | 0.0006 | 1 | 0.73 | 7.53E-14 | 0.22 | 3.47 | 6 |
| lysoPC(O-18:1) | 2 | C27H55NO8P-[M+FA-H]- | 6.13 | 552.3670 | 552.3665 | 0.0005 | 1 | 0.67 | 1.31E-13 | 0.03 | 3.68 | 7 |
| PE(O-16:1/22:6) | 2 | C43H73NO7P- | 4.29 | 746.5135 | 746.5125 | 0.0010 | 1 | 0.84 | 7.03E-13 | 0.70 | 2.34 | 9 |
| PE(O-18:2/20:4) | 2 | C43H75NO7P- | 4.29 | 748.5289 | 748.5271 | 0.0018 | 2 | 0.89 | 4.61E-12 | 0.65 | 2.30 | 7 |
| PE(O-18:2/22:6) | 2 | C45H75NO7P- | 4.27 | 772.5295 | 772.5281 | 0.0014 | 2 | 0.86 | 7.15E-12 | 0.99 | 3.07 | 9 |
| lysoPC(20:3) | 2 | C29H53NO9P-[M+FA-H]- | 5.82 | 590.3460 | 590.3458 | 0.0002 | 0 | 0.62 | 1.20E-11 | 0.12 | 3.25 | 6 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | Ret Time (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | p-value [t-test] (INT-CAR vs INT-FEM) | Fold Change (CAR/FEM) | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | |
| **HILIC Aqueous Negative Mode** | | | | | | | | | | | | |
| lysoPC(20:2) | 2 | C29H55NO9P- [M+FA-H]- | 5.80 | 592.3600 | 592.3614 | -0.0014 | -2 | 0.68 | 2.52E-11 | 0.38 | 3.45 | 9 |
| lysoPC(20:4) | 2 | C29H51NO9P- [M+FA-H] | 5.78 | 588.3310 | 588.3301 | 0.0009 | 2 | 0.62 | 5.74E-11 | 0.03 | 3.02 | 6 |
| PC(O-16:1/20:4) | 2 | C45H81NO9P- [M+FA-H]- | 4.67 | 810.5660 | 810.5649 | 0.0011 | 1 | 0.74 | 9.09E-11 | 0.23 | 2.00 | 5 |
| PC(18:1/20:4) | 2 | C47H83NO10P- [M+FA-H]- | 4.70 | 852.5767 | 852.5755 | 0.0012 | 1 | 0.72 | 9.98E-11 | 0.12 | 1.91 | 9 |
| lysoPC(18:1) | 2 | C27H53NO9P- [M+FA-H]- | 5.88 | 566.3464 | 566.3458 | 0.0006 | 1 | 0.63 | 1.33E-10 | 0.06 | 2.83 | 4 |
| PG(18:1/22:6) | 2 | C46H76O10P- | 1.96 | 819.5178 | 819.5176 | 0.0002 | 0 | 0.73 | 3.27E-10 | 0.37 | 7.79 | 9 |
| PC(O-18:2/20:4) | 2 | C47H83NO9P- [M+FA-H]- | 4.65 | 836.5820 | 836.5805 | 0.0015 | 2 | 0.74 | 4.83E-10 | 0.22 | 2.31 | 9 |
| PC(O-18:1/20:4) | 2 | C47H85NO9P- [M+FA-H] | 4.68 | 838.5983 | 838.5962 | 0.0021 | 3 | 0.72 | 5.22E-10 | 0.76 | 2.69 | 7 |
| lysoPC(14:0) | 2 | C23H47NO9P- [M+FA-H]- | 6.15 | 512.2993 | 512.2988 | 0.0005 | 1 | 0.62 | 6.18E-10 | 0.04 | 2.49 | 12 |
| PC(18:2/18:1) & PC(16:0/20:3) | 2 | C45H83NO10P- [M+FA-H]- | 4.75 | 828.5759 | 828.5755 | 0.0004 | 0 | 0.70 | 1.89E-09 | 0.58 | 1.93 | 6 |
| PG(18:2/22:6) | 2 | C46H74O10P- | 1.96 | 817.5025 | 817.5020 | 0.0005 | 1 | 0.71 | 3.74E-09 | 0.37 | 11.68 | 17 |
| inosine | 4 | C10H11N4O5- | 3.05 | 267.0725 | 267.0729 | -0.0004 | -2 | 0.81 | 8.39E-09 | 0.46 | 2.60 | 7 |
| PG(18:2/18:1) | | C42H76O10P- | 2.13 | 771.5177 | 771.5176 | 0.0001 | 0 | 0.76 | 9.23E-09 | 0.99 | 4.92 | 12 |
| uridine | 4 | C9H11N2O6- | 1.49 | 243.0613 | 243.0617 | -0.0004 | -2 | 0.79 | 4.00E-08 | 0.62 | 1.60 | 5 |
| UKN-HL-NEG-194.1004-5.17 | na | C7H16NO5- † [M+FA-H]- | 5.17 | 194.1004 | 194.1028 | -0.0024 | -12 | 0.65 | 1.06E-07 | 0.13 | 3.36 | 9 |

(Continued on next page)

(Continued from previous page)

| Met Name | LoA | Mol Formula (as detected) [adduct] | Ret Time (min) | m/z (found) | m/z (theoretical) | ΔDa | ppm | p(corr) [OPLS-DA] | p-value [t-test] | p-value [t-test] (INT-CAR vs INT-FEM) | Fold Change (CAR/FEM) | CV% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | |
| **HILIC Aqueous Negative Mode** | | | | | | | | | | | | |
| PC(18:0/22:6) | 2 | C49H85NO10P-[M+FA-H]- | 4.68 | 878.5882 | 878.5911 | -0.0029 | -3 | 0.53 | 3.02E-07 | 0.35 | 1.50 | 6 |
| lysoPC(O-18:2) | 2 | C27H53NO8P-[M+FA-H]- | 5.38 | 550.3514 | 550.3509 | 0.0005 | 1 | 0.54 | 8.20E-07 | 0.15 | 5.09 | 17 |
| PE(18:1/22:4) & PE(18:0/22:5) | 2 | C45H79NO8P- | 4.35 | 792.5514 | 792.5543 | -0.0029 | -4 | 0.69 | 9.45E-06 | 0.03 | 1.41 | 12 |
| UKN-HL-NEG-139.9840-2.98 | na | ud | 2.98 | 139.9840 | ud | ud | ud | 0.53 | 4.01E-05 | 0.04 | 1.72 | 7 |
| PE(O-18:1/22:6) | 2 | C45H77NO7P- | 4.27 | 774.5441 | 774.5438 | 0.0003 | 0 | 0.57 | 4.47E-05 | 0.51 | 1.52 | 5 |
| PE(O-16:1/20:4) | 2 | C41H73NO7P- | 4.31 | 722.5135 | 722.5125 | 0.0010 | 1 | 0.55 | 1.79E-04 | 0.25 | 1.38 | 7 |
| UKN-HL-NEG-126.0015-3.77 | na | ud | 3.77 | 126.0015 | ud | ud | ud | 0.58 | 3.06E-04 | 0.03 | 1.86 | 11 |
| SM(d18:2/18:0) | 2 | C42H82N2O8P-[M+FA-H]- | 5.23 | 773.5826 | 773.5809 | 0.0017 | 2 | -0.74 | 4.89E-08 | 0.44 | -7.97 | 22 |
| UKN-HL-NEG-112.9590-5.49 | na | ud | 5.49 | 112.9590 | ud | ud | ud | -0.81 | 6.48E-07 | 0.97 | -1.23 | 4 |
| UKN-HL-NEG-95.0241-1.02 | na | ud | 1.02 | 95.0241 | 95.0245 | -0.0004 | -4 | -0.72 | 2.74E-05 | 0.29 | -3.03 | 11 |
| benzoic acid | 4 | C7H5O2- | 1.03 | 121.0288 | 121.0290 | -0.0002 | -2 | -0.63 | 2.21E-04 | 0.48 | -1.68 | 23 |
| adenosine | 4 | C10H13N5O4Cl-[M+Cl-] | 2.08 | 302.0650 | 302.0656 | -0.0006 | -2 | -0.59 | 5.21E-04 | 0.20 | -8.87 | 10 |
| UKN-HL-NEG-263.1027-5.20 | na | ud | 5.20 | 263.1027 | ud | ud | ud | -0.52 | 5.25E-04 | 0.22 | -2.53 | 7 |
| UKN-HL-NEG-204.0471-1.36 | na | ud | 1.36 | 204.0471 | ud | ud | ud | -0.60 | 1.47E-03 | 0.41 | ud | 8 |

* Statistics and assignment on second isotope
† Molecular formula as calculated using the isotopic pattern (tentative)
‡ Standard compound of elaidic acid (stereoisomer of oleic acid) exhibited different RT
§ Standard compound of linoelaidic acid (stereoisomer of linoleic acid) exhibited different RT
¶ Detected as a fragment of parent ion

LoA: Level of Assignment; 1: Tentative assignment; 2: Tandem MS spectrum matched to databases or literature; 3: RT matched to standard compound; 4: MS/MS spectrum matched to standard compound.
RT: Retention time.
HILIC: Hydrophilic Interaction (Liquid) Chromatography.
P(corr) refers to the correlation coefficient of metabolites to disease classes.
Two-tailed t-tests were conducted, assuming unequal variance.
Coefficient of variation (CV%) is calculated based on thirteen injections of the same pooled quality control (QC) sample, acquired throughout the run.
The position of double bonds and the position of fatty acyl chain cannot be determined in lipid moieties (Cer, HexCer, PC, PE, PG, PI, SM, and TG). Fatty acyl chains are presented from lowest to highest molecular weight (lower number of carbons to higher, and higher number of double bonds to lower).
Cer: Ceramide; FA: formate; Hex: Hexosyl; na: not applicable; oxCE: oxidised cholesterol ester; PC: phosphatidylcholine; PE: phosphatidylethanolamine; PG: phosphatidylglycerol; PI: phosphatidylinositol; SM: sphingomyelin; TG: triglyceride; ud: unable to determine; UKN: unknown.

# *Chapter 5 A metabolic profiling approach to explore candidate biomarkers and pathways of cardiovascular calcification*

## *5.1 Introduction*

Vascular and valvural calcifications are pathological conditions of ectopic bone formation. Vascular calcification can coexist with atherosclerosis. However, there is no clear knowledge of whether it can actually be the cause, effect, or a co-progression[57, 60]. Nonetheless, evidence of independent mechanisms of these two conditions also exist[57]. Elucidation of the pathology of vascular calcification, as well as biomarker discovery studies, become more important due to contradictory reports of prognostic significance of calcification in cardiovascular events[57, 59, 60]. Aortic valve calcification is the second most common indication for cardiac surgery[61], and the first for valve replacement in the United States[62]. Calcific aortic valve disease is also associated with atherosclerosis[62]. The failure of conventional risk factors and drugs, the absence of easy-to-apply screening-biomarkers, along with the severity of the disease, declare the critical need for alternative technologies.

Calcification is no longer regarded as a passive calcium deposition, degenerative disease. It is now widely considered to be an active biological process of ossification[60, 62, 181, 182]. Bone morphogenetic proteins (BMPs), matrix Gla protein (MGP), osteoprotegerin, osteopontin, osteonectin are some of the bone-formation related factors known to be involved[57, 60, 181, 182]. For this reason, calcification is now believed to recapitulate orthotopic embryogenic osteogenesis[57, 60, 181]. Inflammation also plays a crucial role in calcification. Cells such as macrophages and lymphocytes can penetrate epithelial tissue (and plaque) and release inflammatory factors such as cytokines[57, 60, 62]. Apoptosis is yet another process suggested to be involved in vascular calcification[60, 181, 182]. Dysregulation of levels of inorganic pyrophosphate and phosphate, vitamin D, alkaline phosphatase and reactive oxygen species (ROS), are also considered to be some, but definitely not all, of the mediators of calcification[57, 60, 181, 182].

The aim of this study is to apply cross-platform untargeted techniques to profile the metabolome of a cohort of patients with exertional symptoms presenting various degrees of coronary calcification, but no flow limiting lesions. These patients were compared to patients with no indications of calcified lesions. Additionally, a comparison to a parallel class of patients, with aortic stenosis due to valve calcification, was also conducted in order to ascertain whether there is a different metabolic signature for the two calcific diseases. Ultimate objectives of this study are 1) to obtain candidate biomarkers in order to deliver on early and accurate diagnostics, and risk stratification to guide for appropriate intervention, and 2) to generate hypotheses and follow through with an elucidation of disease pathways in large population studies (targeted or untargeted) and bottom-up approaches. To the best of our knowledge, this is the first metabonomic application on calcific cardiovascular disease, in any matrix (blood, tissue or cells).

## 5.2 Methods

### 5.2.1 Patients

Serum samples were collected from patients that visited the clinic (Heart Centre and Department of Public Health and Clinical Medicine, Umea University, Umea, Sweden) with exertional angina pain. No patient had prior myocardial infarction or coronary artery intervention (PCI). No patient had valve disease, heart failure or renal dysfunction. A parallel class of aortic stenosis patients were recruited before aortic valve replacement surgery. The two groups of patients underwent a multislice CT scan of the chest from which coronary Calcium Score (CS) was measured using the Agatston score and Hounsfield units. Patients for whom no calcium was detected served as the control group. Patients completed a clinical questionnaire investigating their lifestyle, prior clinical conditions, and medications.

A research nurse explained the program to the patients who signed an informed consent to participate in the study. The study protocol was approved by the Regional Ethics Committee of Umea. Subsequently patients had a needle inserted in one of the veins in the cubital fossa and an experienced nurse took a blood samples. The sample left to clot for 30min, centrifuged, and separated serum was stored at -40$^{O}$C. Patients demographics are summarized in Table 5. 1.

.

*Table 5. 1: Patients demographics for all four patient groups included in present studies. CCAD: Calcific coronary artery disease, CAVD: Calcific aortic valve disease.*

| **Class** | **No Calcification** | **Mild Calcification (CCAD)** | **Severe Calcification (CCAD)** | **CAVD** |
|---|---|---|---|---|
| **Number of patients** | 26 | 27 | 17 | 9 |
| **Calcium Score Median (range) [No. of missing]** | 0 (n/a) [0] | 50 (4-219) [0] | 801 (254-1840) [0] | 694 (248-2454) [4] |
| **Median Age (range)** | 61 (50-74) | 67 (50-80) | 67 (48-83) | 79 (68-83) |
| **Male gender (%)** | 7 (27) | 11 (41) | 11 (65) | 4 (44) |
| **Median BMI (range)** | 26.4 (21.5-37) | 26.4 (19.2-41.4) | 27.8 (23.3-37.3) | 27.8 (20.7-31.0) |
| **Median of blood pressure, mm Hg (range)** | | | | |
| **Systolic** | 138 (110-180) | 140 (104-172) | 138 (120-178) | 139 (116-184) |
| **Diastolic** | 84 (62-100) | 80 (66-96) | 82 (62-100) | 82 (64-88) |
| **Hypertension (%)** | 16 (62) | 20 (74) | 15 (88) | 7 (78) |
| **Smoker** | | | | |
| **yes (%),** | 1 (4) | 3 (11) | 2 (12) | 0 (0) |
| **former (%)** | 12 (46) | 13 (48) | 10 (59) | 5 (56) |
| **Diabetes Mellitus (%)** | 4 (15) | 3 (11) | 5 (29) | 1 (11) |
| **Hereditary (%)** | 17 (65) | 18 (67) | 14 (82) | 1 (11) |
| **Osteoporosis (%)** | 2 (8) | 2 (7) | 1 (6) | 1 (11) |
| **Hypercholesterolemia (%)** | 13 (50) | 23 (85) | 17 (100) | 5 (56) |
| **Statins (%)** | 10 (38) | 21 (78) | 15 (88) | 4 (44) |
| **Aspirin (%)** | 11 (42) | 21 (78) | 13 (76) | 0 (0) |
| **Metoprolol (%)** | 6 (23) | 11 (41) | 9 (53) | 5 (56) |
| **Nitrates (%)** | 10 (38) | 9 (33) | 9 (53) | 2 (22) |
| **Amlodipin (%)** | 3 (12) | 4 (15) | 5 (29) | 1 (11) |

## 5.2.2 Lipid Profiling using Ultra-Performance Liquid Chromatography – Mass Spectrometry (UPLC-MS) Analysis

### 5.2.2.1 Serum Liquid-Liquid Extraction

An aliquot of 100μL of serum was mixed with 600uL of organic solvent mixture in an Eppendorf tube. The organic solvent mixture consisted of 3:1 dichloromethane (Chromasolv, LC-MS grade, Fluka, Germany) and methanol (Chromasolv, LC-MS grade, Fluka). After intense vortexing for 30s, samples were centrifuged for 10min at 12g and $4^{O}C$ . A volume of 500uL of the organic layer was then transferred to a glass vial and left to evaporate overnight in a fume hood. Extracts were stored at $-40^{O}C$ until analysis.

### 5.2.2.2 UPLC-MS analysis

Samples were reconstituted in 400uL of solvent mixture of $H_2O$/ isopropanol/acetonitrile (1:2:1) (Optima, LC-MS grade, Fisher Scientific, USA), and transferred into Total Recovery vials (Waters, USA), after centrifugation for 10min at 5g and $4^{O}C$.

UPLC separation was conducted using an Acquity UPLC System (Waters Corp, USA). An Acquity UPLC CSH C18 2.1x100mm, 1.7um, column (Waters Corp, USA) was used. Column temperature was set at $55^{O}C$, flow rate of 0.4mL/min. Injection volume of 3uL and 7uL were used for positive and negative ionisation mode respectively. Mobile phase A consisted of acetonitrile/water (60:40) and mobile phase B Isopropanol/acetonitrile (90:10). In both solutions ammonium formate (LC-MS grade, Fluka, USA) was diluted to 10mM and formic acid (MS grade, Fluka, USA) to 0.1%. The chromatographic gradient program is summarized in Table 5. 2.

*Table 5. 2: Gradient program of chromatography of the lipid profiling UPLC-MS methodology.*

| Time (min) | %A | %B | Curve |
|---|---|---|---|
| 0.0 | 60 | 40 | |
| 2.0 | 57 | 43 | 6 |
| 2.1 | 50 | 50 | 1 |
| 12.0 | 46 | 54 | 6 |
| 12.1 | 30 | 70 | 1 |
| 18.0 | 1 | 99 | 6 |
| 18.1 | 60 | 40 | 6 |
| 20.0 | 60 | 40 | - |

Detection of eluting UPLC fractions was achieved using a Xevo G2 QTof (Waters MS Technologies, UK). Both MS and $MS^E$ data scans were acquired for 0.200 s every 0.214 s. $MS^E$ data were collected after ramping the collision energy from 30 to 40V. Leucine Enkephalin was used for lock mass correction. Lock mass data were collected every 0.3 s for 0.2 s. Conditions in positive mode: m/z range:150-1200, cone voltage 30V, capillary voltage 2kV, Source Temperature $120^{O}C$, Desolvation Temperature $550^{O}C$, Desolvation Gas 900L/h. Conditions in negative mode: m/z range: 50-1200, cone voltage 30V, capillary voltage 1kV, Source Temperature $120^{O}C$, Desolvation Temperature $550^{O}C$, Desolvation Gas 900L/h.

A QC format [114] was used for the UPLC-MS analysis. Briefly, a pooled sample (referred to as Quality Control Sample, QC) of the reconstituted extracts was prepared. This sample was re-injected 10 times before initiating the run to condition the column of 10 injections. Then the sample was re-injected once at the beginning, every 10 injections of samples, and at the end of the run (total of 9 injections). A tight grouping of the QC samples in MVDA, and a Coefficient of Variation percentage (CV%) of less than 30% for candidate biomarkers [17, 92] formed the quality control criteria.

#### *5.2.2.3 UPLC-MS Data Processing*

After acquisition, data were centroided (m/z spectra peaks are automatically detected and their centroid is calculated based on the average m/z value and weighted by the intensity). A post-acquisition calibration was conducted on negative mode data, to adjust for mass accuracy. Calibration was conducted using m/z of peaks with known m/z.

This was followed by peak-picking and grouping using MarkerLynx XS (Waters Inc, v4.1) software, using the following parameters:

<u>Positive mode:</u> Function: 1, Analysis Type: Peak Detection, Initial Retention Time: 0.40, Final Retention Time: 17.00, Low Mass: 150.00, High Mass: 1200.00, XIC Window 0.10 Da, Peak Width at 5% Height: 20s, Marker Intensity Threshold: 1000counts, Mass Window: 0.10Da, Retention Time Window: 0.50

<u>Negative mode:</u> Function: 1, Analysis Type: Peak Detection, Initial Retention Time: 0.45, Final Retention Time: 17.00, Low Mass: 50.00, High Mass: 1200.00, XIC Window 0.10 Da, Peak Width at 5% Height: 20s, Marker Intensity Threshold: 400counts, Mass Window: 0.10Da, Retention Time Window: 0.50

For Both Modes: Noise Elimination Level: 6.00, Peak-to-peak Baseline Noise: Auto, Replicate % Minimum: 0.00, No relative retention time, No Smoothing, No Deisotoping.

Values were reported as area of intensity peaks. Saturated peaks were removed, prior to total area normalisation. Values were multiplied by 10 000 prior to statistical analyses.

#### *5.2.2.4 Statistical analysis and validation*

Multivariate data analysis (MVDA) for UPLC-MS data was conducted using the SIMCA-P+ (v. 12.0.1.0.; Umetrics) package. Principal Component Analysis (PCA) and Orthogonal Projection to Latent Structures – Discriminant Analysis (OPLS-DA) were applied to the processed Pareto-scaled data. Model validation was carried out using CV-ANOVA testing [142] and permutation (n=999) testing. Permutation testing is the process of deliberately randomizing the sample labelling such that the class descriptor has no meaning and recalculating their predictive values. Randomly classified models should demonstrate predictive values that are lower than those derived from the actual model using correct classifiers.

In order to extract putative biomarkers from UPLC-MS, features (corresponding to metabolites) with correlation coefficient (Note: correlation coefficients that refer to correlation of samples to disease classes will be refer to as: p(corr)) greater than 0.5 in absolute value were initially chosen. This cut-off provides confidence greater than 99%. These features were further subjected to two-tailed t-test assuming unequal variance with a threshold of $p<0.05$[143, 144] and fold-change comparison. Features were reported as significant and structurally identified if: 1) both p(corr) and t-test p-value met the thresholds , 2) were reproducible through the run, with a coefficient of variation (CV%) of the QCs less than 30%, and 3) pass the chromatographic peak shape assessment.

For structural elucidation of significant features UPLC-$MS^E$ and UPLC-MS/MS data were used. $MS^E$ data were collected through the run as described for the UPLC-MS analysis. These conditions were the ones used for MS/MS data additional analysis that was conducted using data dependent acquisition (DDA) or by targeting specific ions. MS/MS data were collected with collision energy ramping from 30 - 50V. Apart from fragmentation patterns, structural elucidation was assisted by matching accurate m/z measurements to metabolites from online available databases[111, 116, 139]. In some cases isotopic patterns also proved useful.

#### *5.2.2.5 Structural Identification of signals discriminating controls from case detected in UPLC-MS*

For structural elucidation of significant features UPLC-$MS^E$ and UPLC-MS/MS data were used. $MS^E$ data were collected through the run as described for the UPLC-MS analysis. These conditions were the ones used for MS/MS data additional analysis that was conducted using data dependent acquisition (DDA) or by targeting specific ions. MS/MS data were collected with collision energy ramping from 30 - 50V. Apart from fragmentation patterns, structural elucidation was assisted by matching accurate m/z measurements to metabolites from online available databases[111, 116, 139]. In some cases isotopic patterns also proved useful.

### *5.2.3 $^1$H NMR spectral acquisition*

#### *5.2.3.1 Sample preparation*

An aliquot of 350μL of serum was diluted in 350ul of phosphate buffer. Phosphate buffer (pH=7.4) was prepared by dissolving 4.56g of $NaH_2PO_4$ and 43.46g of $Na_2HPO_4$ per litre of $D_2O$. A total of 600μL of the diluted serum was transferred into a 5mm outer diameter NMR tubes.

#### *5.2.3.2 Standard one-dimensional spectra acquisition*

$^1$H NMR spectra of serum samples were acquired on a 600MHz Bruker Avance spectrometer (Bruker Biospin, Rheinstetten, Germany) operating at 600.13 MHz and a temperature of 300K, using a standard 1D pulse sequence (recycle delay (RD)-$90^o$-$t_1$-$90^o$-$t_m$-$90^o$-acquire free induction decay(FID)) with water suppression applied during RD of 2 s and mixing time ($t_m$) of 100 ms. The $90^o$ pulse was set at 16.75 μs, with a power level of 1.0 dB. Spectra were acquired using 256 scans into 64K data points with a spectral width of 20 ppm. The FIDs were multiplied by an exponential function corresponding to 0.3 Hz line broadening before applying Fourier transformation.

#### *5.2.3.3 NMR Data processing*

Spectra were subjected to automatic phasing, calibration to the β-anomeric glucose peak (δ 5.233), and background correction, using an in-house developed software. Bin size was set at δ 0.0003364. Spectra from δ -0.6 - 10.0 excluding the region containing the residual water signal (δ 4.700 - 4.895) were subjected to global alignment and median fold change normalisation[183] using an in-house algorithm.

#### *5.2.3.4 Statistical analysis and cross-validation of NMR data*

MVDA for NMR data was conducted using Matlab programming language (R2011a) and in-house developed code. PCA and OPLS-DA were applied on the processed data. In all MVDA scaling to

unit variance was applied. For model validation permutation testing was applied (n=10000). Features with p(corr) greater than 0.5 in absolute value (>99% confidence) were considered significant. These features were subjected to further analysis using univariate statistics, which were conducted on the corresponding integrated peaks, instead of bin intensity. Univariate statistics consisted of two-tailed t-test (assuming unequal variance) with a threshold of $p<0.05$[143, 144], and fold-change comparison. Features were reported as significant if both p(corr) and t-test p-value met the thresholds and forwarded for identification. Identification of statistically significant features was based on two-dimensional (2D) NMR correlation experiments of a pooled sample[14], internal [184] and online databases[116].

### 5.2.4 Pathway Analysis

#### 5.2.4.1 Correlation and Network analysis

Correlation coefficients (Spearman) between pairs of candidate biomarkers were calculated using R (2.13.2) programming language (Note: Correlation coefficients that express the Spearman correlation between two metabolites will be referred to as: r). For network visualization, of significant metabolites characterizing disease, CytoScape v.2.8.3 software was used. A correlation of r=0.5 and 0.4 in absolute values was used for calcific coronary artery disease (CCAD) and calcific aortic valve disease (CAVD), respectively, when constructing the inference pathways. In both cases, the cut-off for correlation was higher that 99% confidence. However, for CCAD the cut-off was raised as the network was too complex to interpret based on the number of significant features detected.

#### 5.2.4.2 Metabolite Mapping to Key Pathways

Metabolite mapping to pathways was manually conducted. Pathway information was retrieved from KEGG database [23, 24], and relevant literature (see Discussion).

## 5.3 Results

### 5.3.1 Metabolic characterization of calcific coronary artery disease (CCAD)

The NMR analysis did not provide strong biomarkers through either MVDA or univariate analyses. Preliminary MVDA of the UPLC-MS data suggested there were three basic groups within the data. Accordingly CCAD samples were divided into three classes of CS of zero (No Calcification, NC), 1-250 (Mild Calcification, MC), and over 250 (Severe Calcification, SC). A UPLC-MS

chromatogram and NMR spectrum of pooled sample can be found at Figure 5. 1 and Figure 5. 2, respectively.

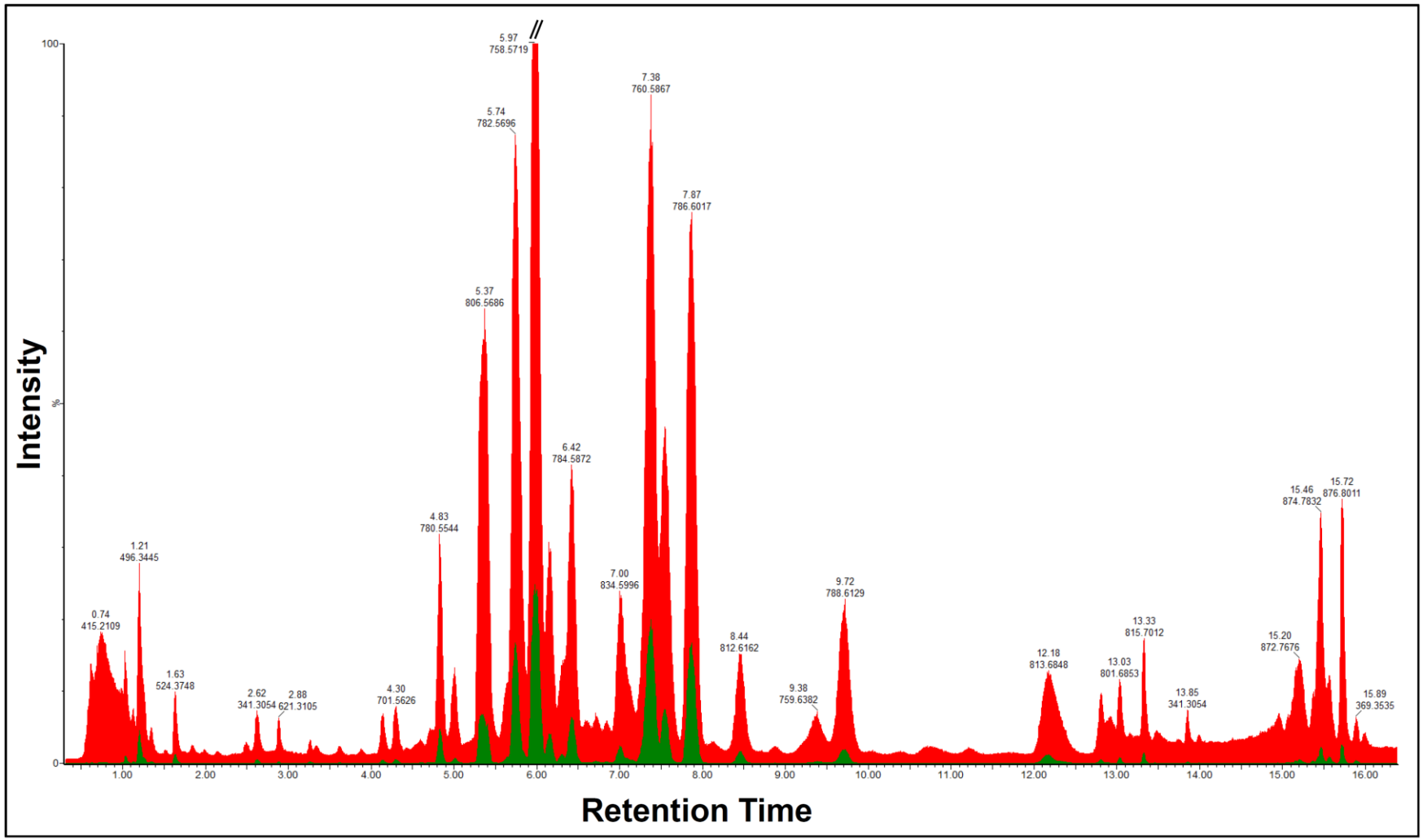


*Figure 5. 1: A representative chromatogram obtained from the positive mode analysis. The chromatogram was generated from a quality control (pooled) sample of organic extracts of blood serum. Red colour represents the total ion current, while the green colour represents base peak intensities, showing the most intense peaks in each retention time window.*

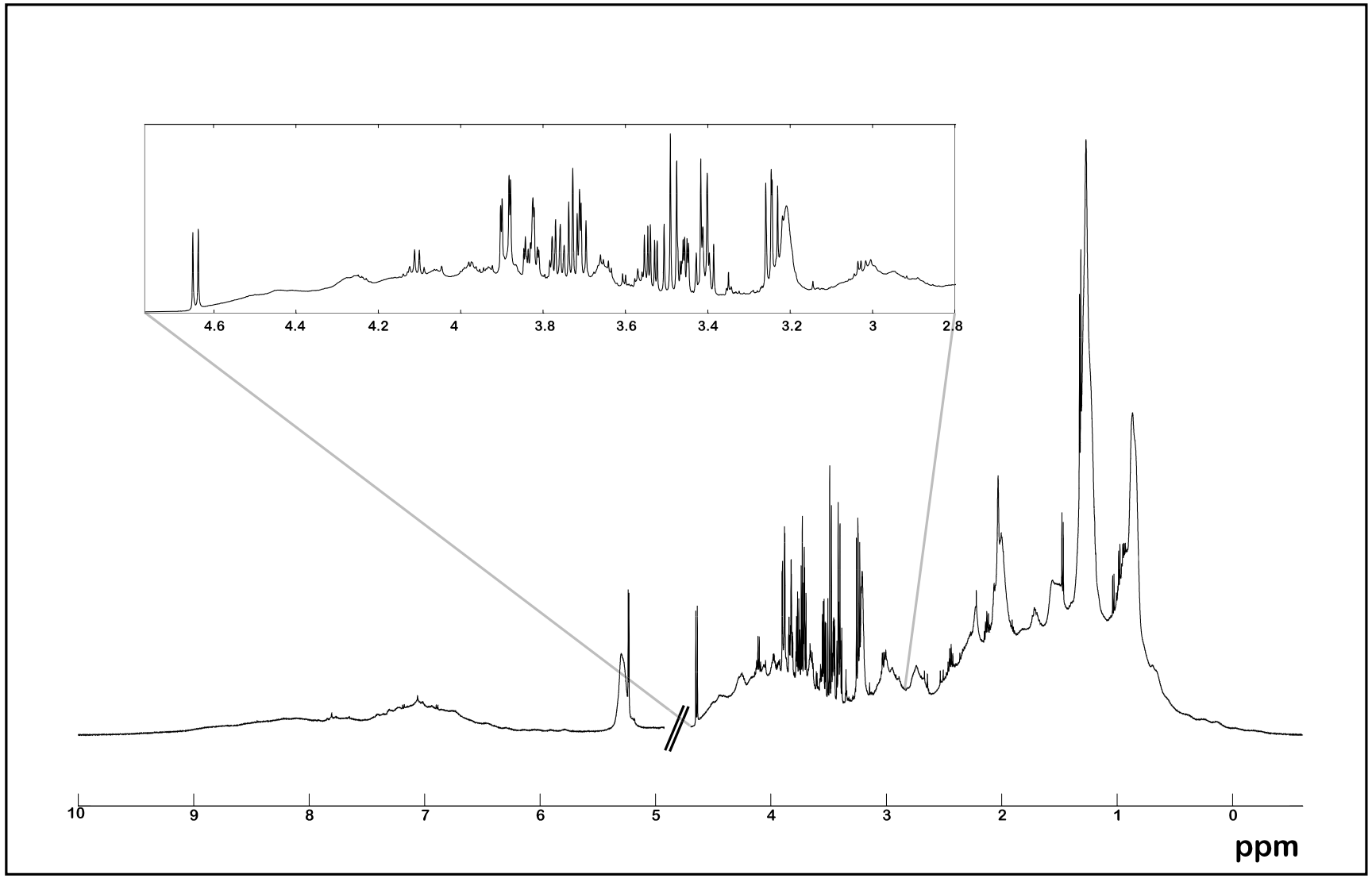


*Figure 5. 2: A representative spectrum obtained from the $^{1}H$ NMR analysis of a pooled serum sample. Insert represents an expansion of 2.8 – 4.8 ppm. H2O peak at 4.700 - 4.895 ppm has been removed from the spectrum and subsequent analysis.*

For UPLC-MS analysis of the organic extracts of the serum samples, it can be easily observed with PCA analysis (unsupervised) scores plots, that better grouping could be accomplished when the two extreme classes (NC and SC) were compared (Figure 5. 3B). The intermediate class (MC) was scattered throughout both extreme groups (Figure 5. 3A). Additionally, in OPLS-DA analysis (supervised) valid and validated models were only observed when comparing samples from the NC class to the SC class (Figure 5. 4 A, B and Figure 5. 5A, B). A model using one predictive component was found to be significant. Model characteristics are summarized in Table 2. The dataset acquired in negative ionization mode gave models with better predictive value ($Q^2Y$). Multivariate Statistics found no statistically significant differences between NC and MC classes.


*Figure 5. 3: Scores plots of principal component analysis (PCA). Data for the analysis were derived from UPLC-MS of organic extracts, of serum samples. (A) Positive mode PCA of NC, MC and SC classes, (B) positive mode PCA conducted only on NC and SC classes. (C) Negative Mode PCA of NC, MC and SC classes, (B) negative mode PCA applied only with NC and SC classes. (Each point represents a sample; Blue rhombus: No Calcification (NC), Black square: Mild calcification - calcific coronary artery disease (MC), Red dot: Severe calcification - calcific coronary artery disease SC).*

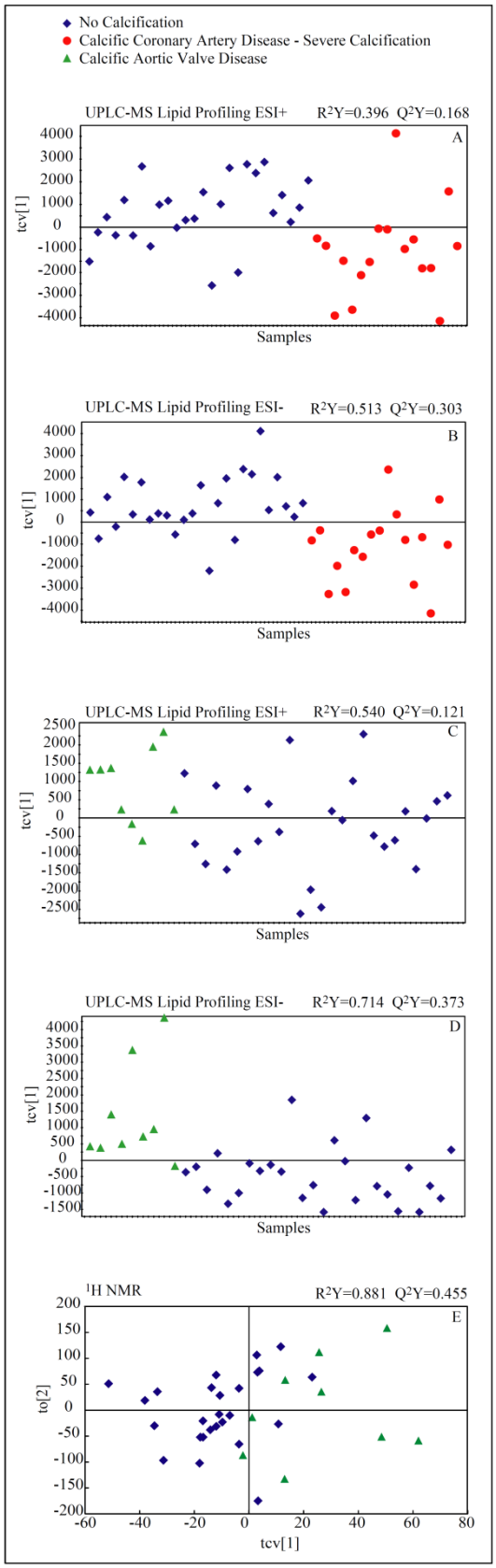


*Figure 5. 4: OPLS-DA scores plots; (A-D) derived from UPLS-MS of organic extracts, and (E) from $^{1}$H NMR analyses of serum samples. UPLC-MS positive mode (A), and negative mode (B) of no calcification versus severe calcification (calcific coronary artery disease) class. UPLC-MS positive mode (C), and negative mode (D) of no calcification versus calcific aortic valve disease class. Cross-validates scores (tcv) plots can be found in Supplementary data Figure S3. (Each point represents a sample; Blue rhombus: No Calcification, Red dot: Severe calcification - calcific coronary artery disease, Green triangle: Calcific aortic valve disease).*

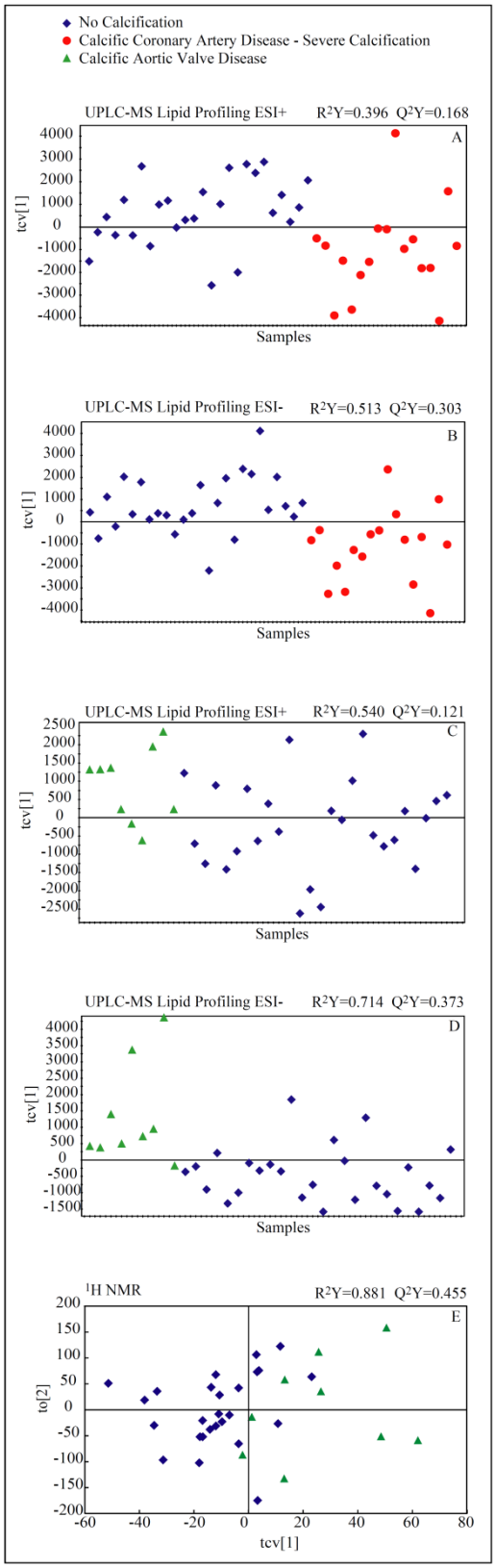


*Figure 5. 5: Cross-validated OPLS-DA scores plots. (A-D) were derived from UPLS-MS of organic extracts, and (E) from $^{1}H$ NMR analyses of serum samples. UPLC-MS positive mode (A), and negative mode (B) of no calcification versus severe calcification (calcific coronary artery disease) class. UPLC-MS positive mode (C), and negative mode (D) of no calcification versus calcific aortic valve disease class. (Each point represents a sample; Blue rhombus: No Calcification, Red dot: Severe calcification - calcific coronary artery disease, Green triangle: Calcific aortic valve disease).*

*Table 5. 3: Summary of model characteristics from (unsupervised; OPLS-DA and PLS-DA) multivariate statistical analyses of data obtained from serum samples.*

| | **OPLS-DA** | | | | | | | **PLS-DA** | |
|---|---|---|---|---|---|---|---|---|---|
| **Group comparison** | NC vs SC | NC vs SC | NC vs CAVD | NC vs CAVD | NC vs CAVD | SC vs CAVD | SC vs CAVD | NC vs SC vs CAVD | NC vs SC vs CAVD |
| **Mode, No of Comp** | POS, 1 Comp | NEG, 1 Comp | POS, 1 Comp | NEG, 1 Comp | $^{1}$H NMR, 1+2 Comp | POS, 1 Comp | NEG, 1 Comp | Pos, 2 Comp | Neg, 2 Comp |
| **$R^2X$(cum)** | 0.175 | 0.140 | 0.106 | 0.107 | 0.393 | 0.183 | 0.106 | 0.251 | 0.190 |
| **$R^2Y$(cum)** | 0.396 | 0.513 | 0.540 | 0.714 | 0.881 | 0.509 | 0.760 | 0.395 | 0.565 |
| **$Q^2Y$(cum)** | 0.168 | 0.303 | 0.121 | 0.373 | 0.455 | 0.222 | 0.480 | 0.116 | 0.321 |
| **CV-ANOVA** | 0.025139 | 0.000736 | 0.126912 | 0.000572 | n/a | 0.055402 | 0.000546 | 0.297803 | 0.000283 |
| **Permutation Testing** | 0.005005 | < 0.001001 | 0.031031 | < 0.001001 | 0.000500 | 0.0190190 | < 0.001001 | 0.005005 | < 0.001001 |

Robustness of the models was characterized using the following model parameters: $R^2X$, variation of the X-matrix (data) explained by the model; $R^2Y$, predicted percentage; and, $Q^2Y$ cross-validated predicted percentage.

Cross-validation parameters are also demonstrated. These parameters show the validity of the models and the likelihood of overfitting.

Pos and Neg refer to positive and negative polarity. Comp is the number of validated principal components.

A global reduction of the sphingomyelin (SM) lipids in the diseased class was observed with high statistical significance (NC vs SC). This class of lipids did not show statistical significance, from t-test comparison, of NC against MC. In contrast, a comparison between MC and SC showed that SMs were significantly different in these two classes (Table 5. 4 and Figure 5. 6).

A number of phosphatidylcholine (PC) moieties appeared to be dysregulated. In contrast to the observation of SM behaviour, PCs did not follow a global trend. PC(16:0/20:4) was detected in higher intensities in the SC diseased class, while PC(18:2/18:0), PC (O-16:1/18:2), PC(15:0/18:2), PC(18:2/18:2) and PC(16:0/18:3) showed a reduction in the SC class compared to the NC group with t-test statistical significance of $p< 0.05$. PCs were the only lipid class that showed a statistical trend through univariate statistics for NC vs MC classes. Although not highly significant, the fact that such a trend is observed only with PCs, and follows the same pattern (Figure 5. 6) for each individual PC as with the SC class is noteworthy. The t-test p-values were as follows: PC(16:0/20:4): $p=0.0238$; PC(18:0/20:4): $p=0.0902$; PC(16:0/18:3): $p=0.0566$; PC(18:2/18:2): $p=0.0743$.

More biomarker lipid classes found to be discriminatory between NC and SC included triglycerides (TGs), with TG(16:0/18:1/22:5) and TG(18:1/18:1/20:4) being higher. These two lipids coelute and have the same m/z value. Therefore, further analysis is needed to assess if one is more significant over the other. TG(16:0/18:1/18:1) was also found in higher intensities in but barely met significance criteria (Table 5. 4). Lastly, lysoPC(20:4) and serum concentrations of phosphatidylinositol(18:2/18:0) (phosphatidylinositol, PI) were also found to be higher and lower, respectively, in the disease class.

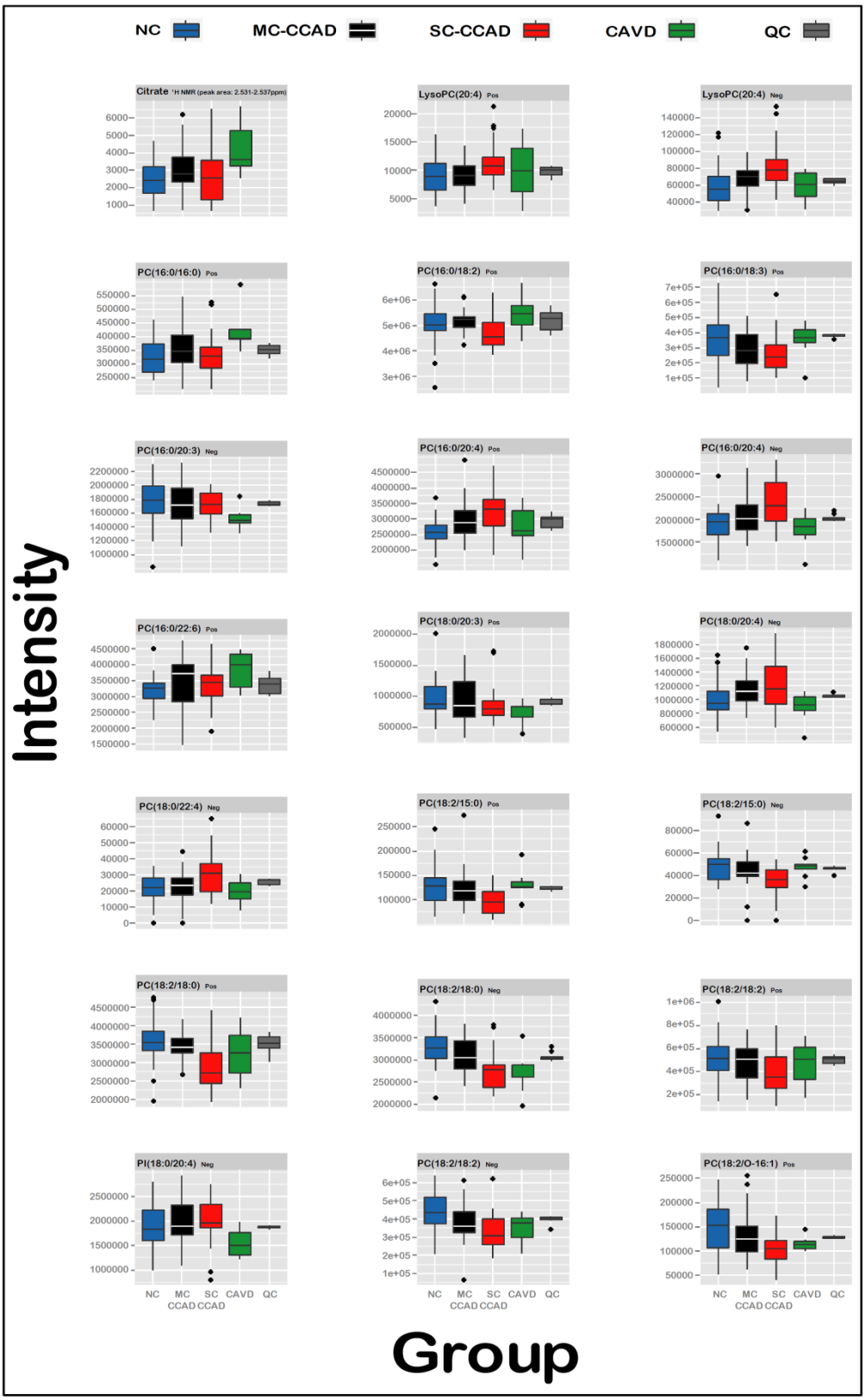


*Figure 5. 6 (part 1): Box-plots of all metabolites found to be statistically different in any of the comparisons between all classes, in both UPLC-MS and NMR analysis. Metabolites are presented in alphabetical order. The analytical platform from where the metabolite was detected is designated (Pos: Positive mode UPLC-MS; Neg: Negative mode UPLC-MS). Each box represents one class: Blue: No Calcification (NC), Black: Mild calcification - calcific coronary artery disease (MC-CCAD), Red: Severe calcification - calcific coronary artery disease SC-CCAD, Green: Calcific aortic valve disease (CAVD), Grey: Quality controls (QC).*

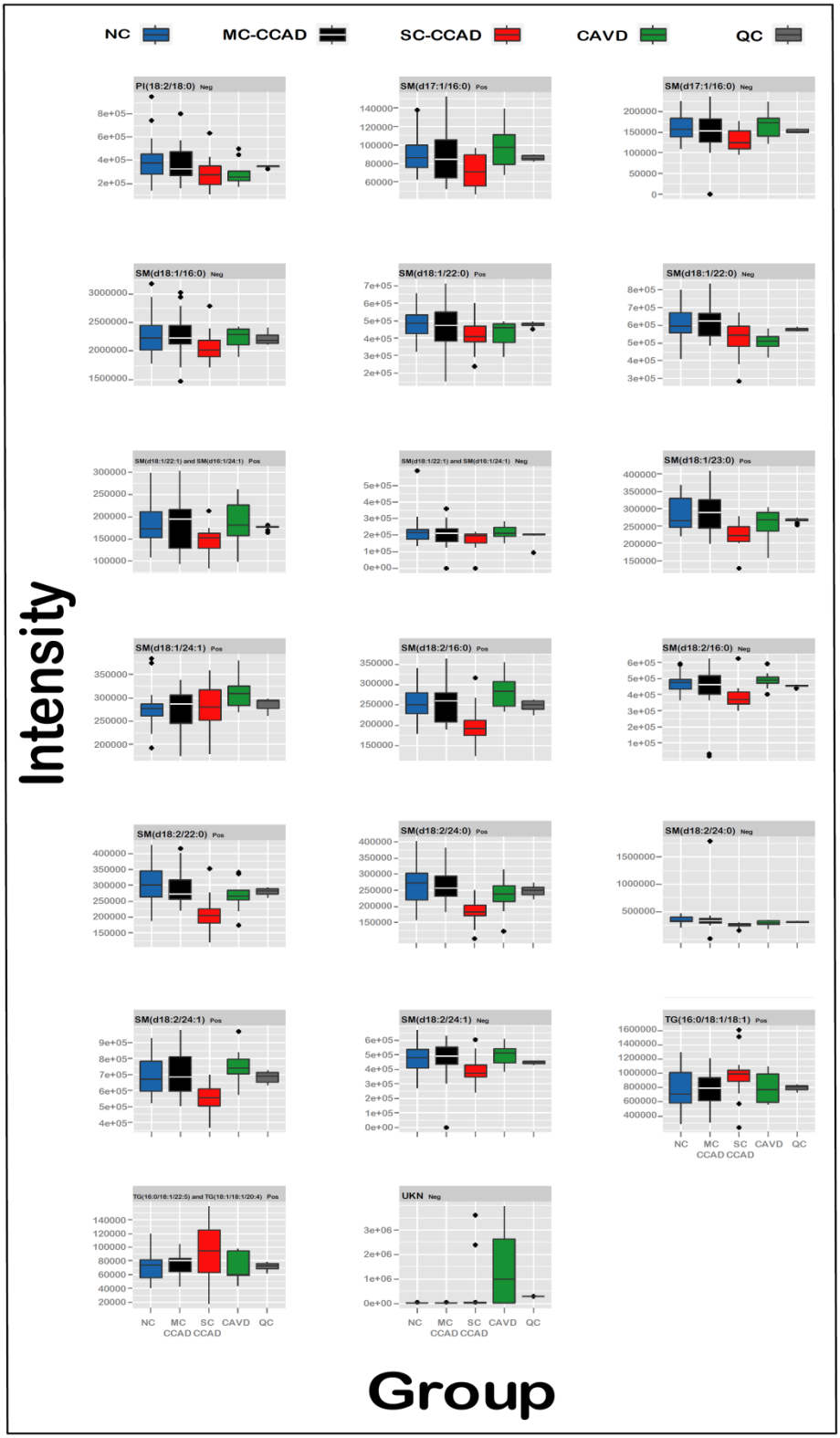


*Figure 5. 6 (part 2): Box-plots of all metabolites found to be statistically different in any of the comparisons between classes, in both UPLC-MS and NMR analysis. Metabolites are presented in alphabetical order. The analytical platform from where the metabolite was detected is designated (Pos: Positive mode UPLC-MS; Neg: Negative mode UPLC-MS). Each box represents one class: Blue: No Calcification (NC), Black: Mild calcification - calcific coronary artery disease (MC-CCAD), Red: Severe calcification - calcific coronary artery disease SC-CCAD, Green: Calcific aortic valve disease (CAVD), Grey: Quality controls (QC).*

### *5.3.2 Metabolic characterization of calcific aortic valve disease (CAVD)*

For the CAVD class, patients with a CS known to be <200, were excluded. OPLS-DA generated good models for the NMR and UPLC-MS negative mode data sets (Table 5. 3). Similarly to the comparison between SC and NC (CCAD), positive polarity UPLC-MS analysis did not show as high predictive value, although models were valid. In UPLC-MS analysis, OPLS-DA of both positive and negative mode showed only one component to be significant with no orthogonal components indicating that the disease class accounted for the main source of variation in the data. OPLS-DA models can be found in Figure 5. 4 C-E and Figure 5. 5 C-E for cross-validated models.

As with CCAD, PCs showed a mixed trend of dysregulation in the two classes with some PCs in higher and others with lower intensities in the disease group(Table 5. 5). Only two SM moieties were found to be dysregulated, and with a mixed trend. The intensity of the peak corresponding to SM(d18:1/24:1) was higher with a t-test p-value of 0.0299, while SM(d18:1/22:0) was found to be lower in the disease class with p-value of 0.0004. Additionally, a metabolite that could not be assigned (UKN) and citrate were higher in the disease class. Citrate was detected from NMR analysis, where a good model was observed with OPLS-DA from the comparison between NC and CAVD. The model possesses good predictive value and was well validated using permutation testing (Table 5. 3). Citrate peaks had the highest p(corr) and above the threshold Table 5. 5 and Figure 5. 7). Citrate was detected from a characteristic chemical shift of the two groups of $CH_2$-hydrogens in the standard one dimensional NMR spectra[184] and the COSY experiment (Figure 5. 8). T-test was applied on all four peaks of the citrate signal after integration (Table 5. 5).

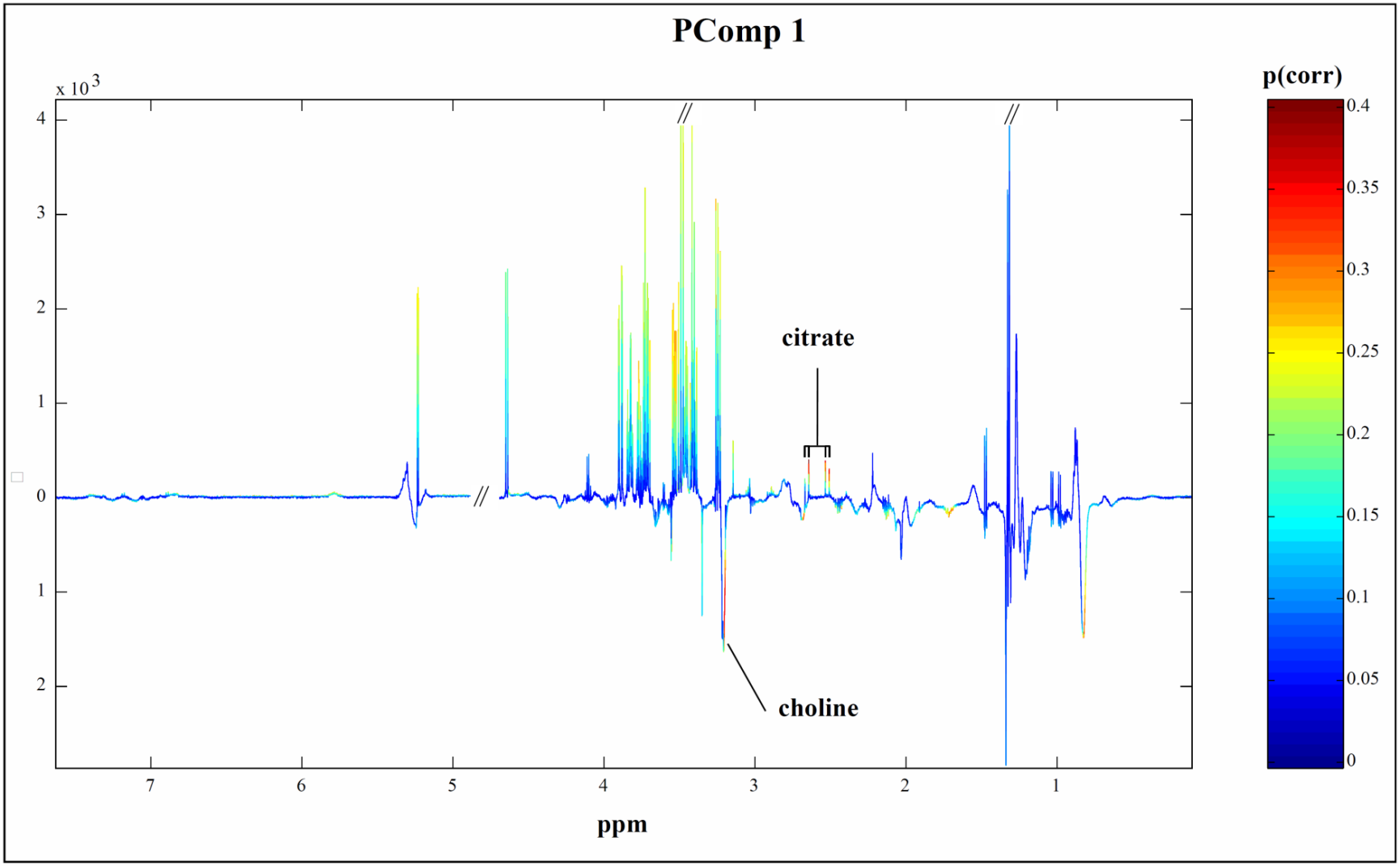


*Figure 5. 7:Plot derived from OPLS-DA of NMR analysis of serum samples of the comparison of CAVD to the control group. Colouring of NMR peaks indicates correlation to the disease p(corr). Apart from citrate, choline is also demonstrating a correlation to class but did not pass the threshold criteria.*

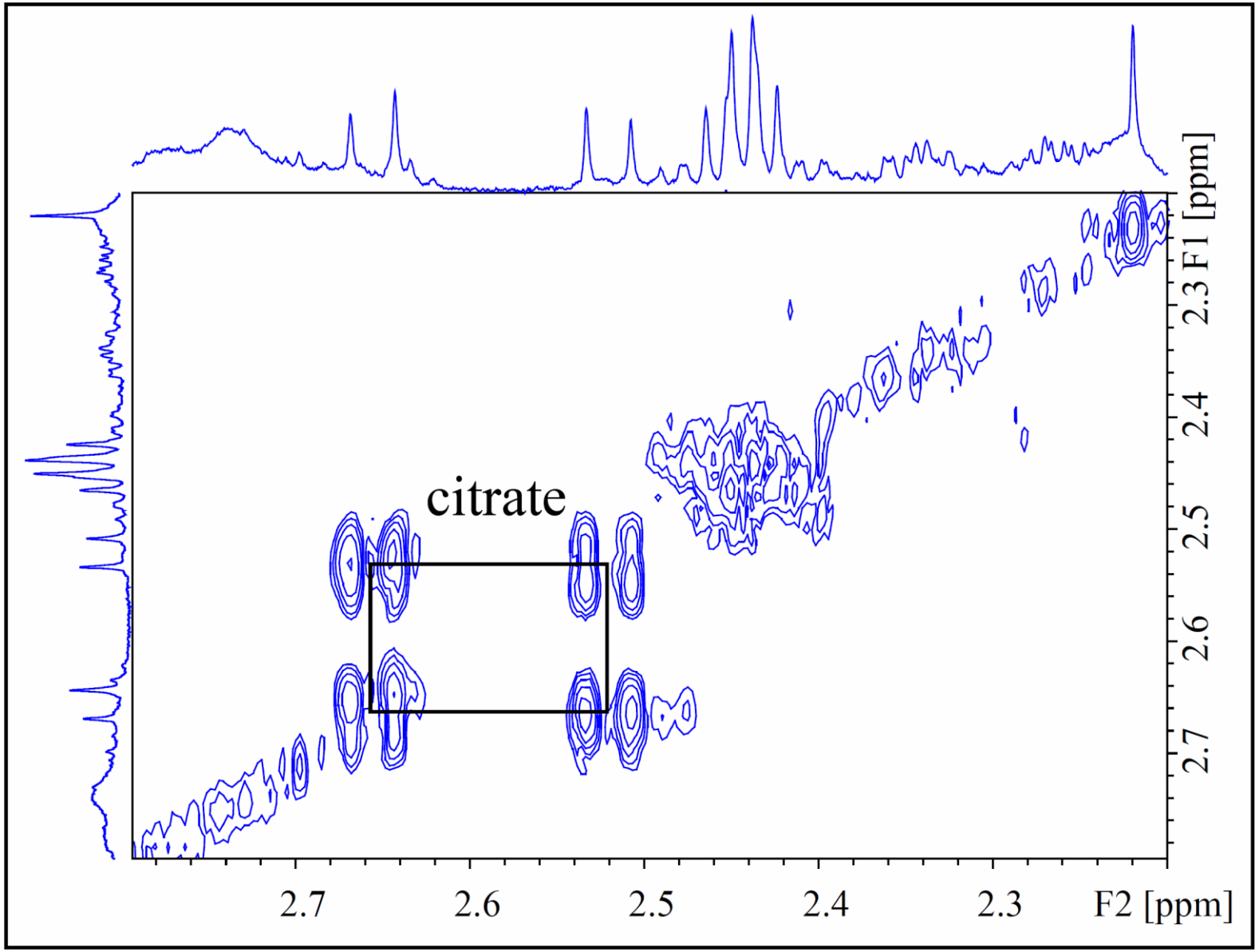


*Figure 5. 8: 2D $^{1}H$-$^{1}H$ COSY NMR spectra demonstrating the correlated peaks of citrate.*

### *5.3.3 Comparison of serum metabolite profiles from calcific coronary artery disease and calcific aortic valve disease patients.*

In order to establish whether the markers of coronary disease are general or whether CAVD and CCAD indeed have different mechanisms, a direct comparison between the two disease classes was carried out. Features can become statistically significant when two classes are compared directly. MVDA comparing between the two disease classes (SC vs CAVD) revealed different profiles for each disease (Figure 5. 9 and Table 5. 3).

Table 5. 6 summarises metabolites found to be significantly different between the two disease groups. An interesting addition to the candidate biomarkers identified from the comparison of a single disease group with control is a differential PI moiety PI(18:0/20:4), which was detected in lower intensities in CAVD. PC(16:0/20:4), PC(18:0/20:4) and PC(18:0/22:4) were also lower in CAVD, as compared to the SC class. On the other hand, six SMs (with p-values ranging between 0.0001 to 0.04), and PC(15:0/18:2), PC(16:0/16:0) and PC(16:0/18:2) were lower in the SC class. The only noticeable similarity between the two diseases was two PCs, PC(18:2/18:0) and PC(18:2/18:2), and SM(d18:1/22:0).

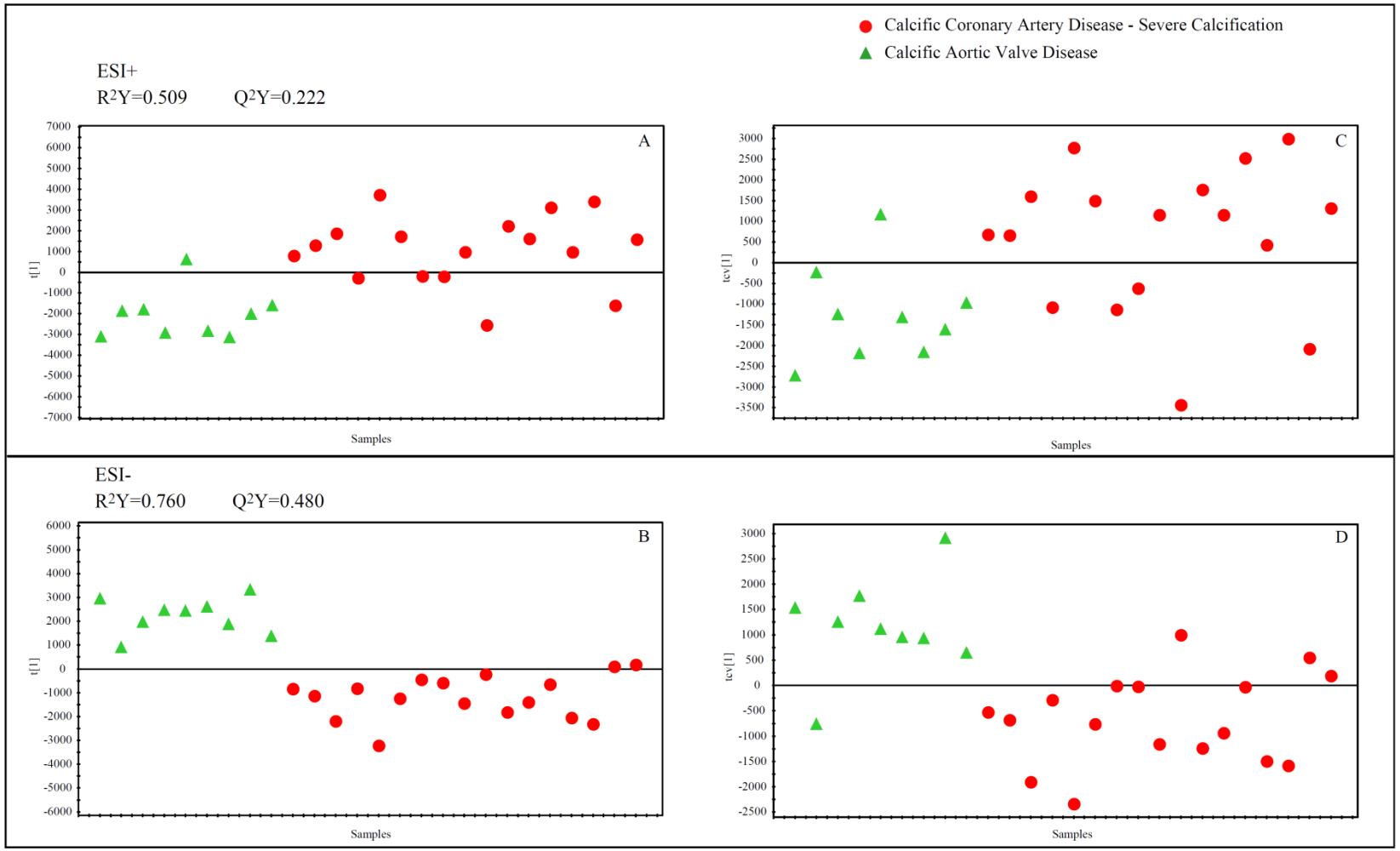


*Figure 5. 9: OPLS-DA scores plots from comparison of calcific coronary artery disease and calcific aortic valve disease. (A and B) Plots derived from UPLS-MS of organic extracts in positive and negative mode, respectively. (C and D) The corresponding cross-validated scores plots from positive and negative mode, respectively. (Each point represents a sample: Green triangle: Calcific aortic valve disease; Red dot: Severe calcification - calcific coronary artery disease).*

Additionally, valid models via PLS-DA analysis showed good grouping of the three classes tested (NC, SC, CAVD) (Figure 5. 10 A and B) with the UPLC-MS data. Model characteristics can be found in Table 5. 3. Again, the negative mode models showed superior predictive value. Loadings plots (Figure 5. 10 C and D) can informatively show the metabolites responsible for the separation of each of the three classes (Table 5. 7). It was also noticed that metabolites that were not statistically significant when comparing either of the disease classes, SC or CAVD, to the NC class, appeared driving the models. A good example is the three TGs, and the two PEs, which can be seen in positive and negative mode loadings plots respectively (Figure 5. 10 C and D).

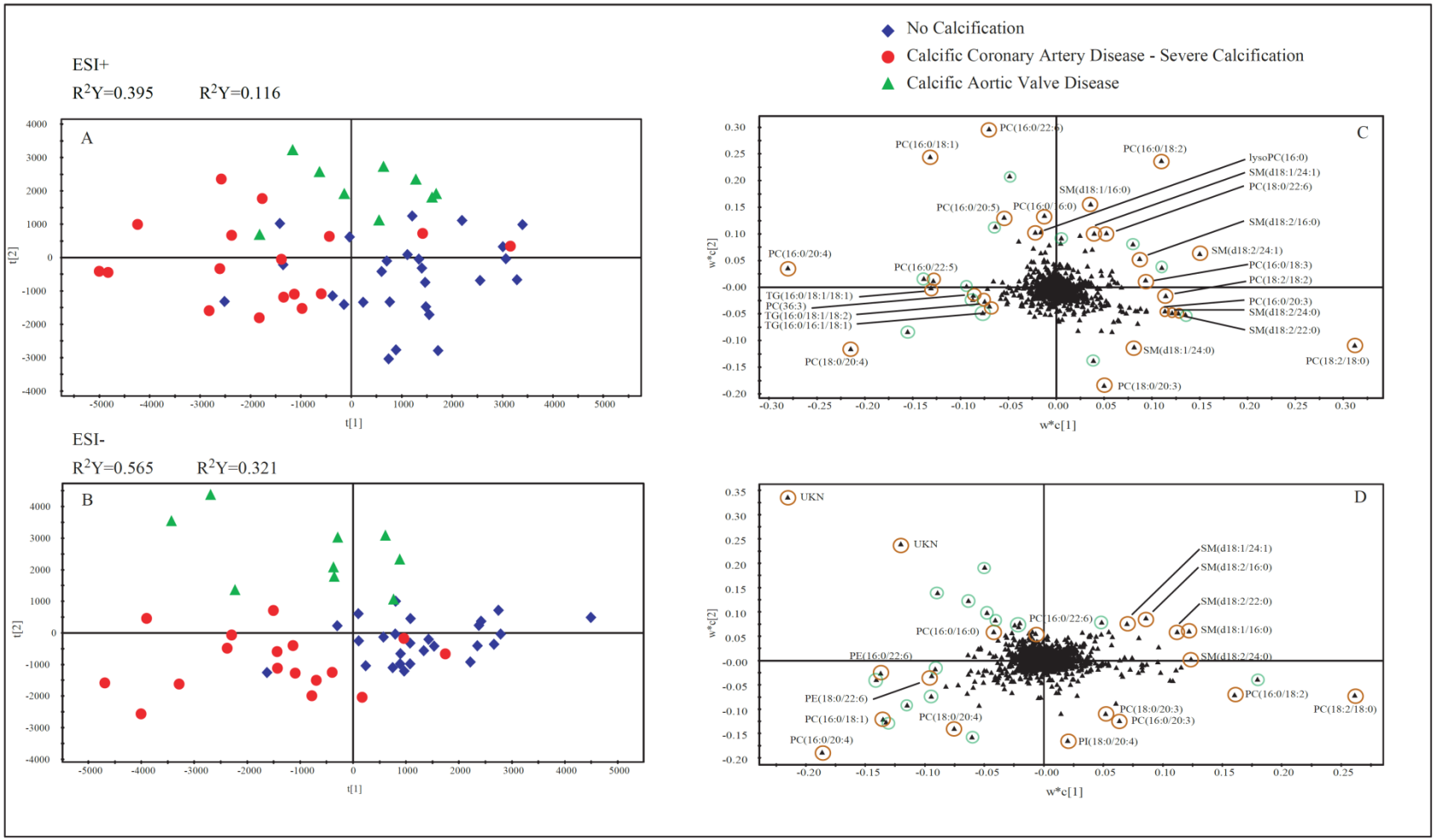


*Figure 5. 10: PLS-DA analysis of UPLS-MS of organic extracts of serum samples. (A) Scores plots and (C) loadings plots from positive mode. (B) Scores plots and (D) loadings plots from negative mode analysis. For A and B, each point represents a sample; Blue rhombus: No Calcification, Red dot: Severe calcification - calcific coronary artery disease, Green triangle: Calcific aortic valve disease. For C and D, each point represents a feature included in the analysis, whereas a feature is circled in brown colour the identity of the metabolite is demonstrated, while green circle represents features that have been identified to be isotopes, adducts, or fragments of metabolites already assigned in the figure, or artefacts.*

## 5.4 Discussion

This is the first metabonomic approach to be applied in calcific disease. Most of the detected metabolites and lipid classes are presented here for the first time for cardiovascular calcification. Metabolite dysregulation was specific to particular FACs. Although specific FACs in the form of fatty acids have been implicated before in calcification[185, 186], this is the first study to provide evidence of extensive dysregulation in a range of FACs. Additionally, a number of FAC are reported to be involved in the mechanism of disease for the first time. More importantly, the panels of metabolites differentiating CCAD and CAVD from control samples were different in each of the two diseases studied. Evidence supporting apoptosis, such as SM reduction, was only present in CCAD. Lastly, from present findings $Ca^{2+}$ channel and TCA cycle intermediates also appear to be dysregulated.

### 5.4.1 Metabolic characterization of calcific coronary artery disease (CCAD)

From all the lipid classes found to be discriminatory, SMs were the most statistically significant. SMs are found in the epithelial cell (EC) membrane in high concentrations. The lower level of all significant discriminatory SMs in the SC class, along with their apparent high r between each other, is consistent with dysregulation of their common enzyme Sphingomyelinase. This potentially implicates the ceramide pathway, a pathway well known for involvement in signalling pathways of inflammation and apoptosis[43]. It has been suggested that apoptosis is a potential mechanism of calcification[181, 182]. However, these consistent lower levels of SMs were observed in the SC class only. In fact, a comparison of the levels of SMs between the SC and MC class, proved to be statistically different (Table 5. 4). This suggests that their use as biomarkers may only apply to later stage disease. This could mean that apoptosis contributes to the mechanism of the calcification process at a later stage, or otherwise affects severely calcified lesions. Although the SMs may implicate the ceramide pathway in CCAD, ceramides themselves were not detected as candidate marker. This is not surprising given the low abundance of ceramides in the blood. Ceramides do not have a polar head-group and are therefore not found in the blood in concentrations of the magnitude of SMs. Molecular concentration of ceramides, compared to SMs, has been reported to be 30 times less[40]. From studies conducted in atherosclerotic tissue (Chapter 4), it was observed that the reduction of SMs was accompanied with an elevation in ceramides, validating the assumption that Sphingomyelinase is the likely enzyme responsible for this biological finding.

One of the most interesting perturbed lipid classes in the CCAD is that of PCs. PCs are major constituents of the cell membrane, and are involved in metabolism and signalling. Although several lipids were statistically different between CCAD and controls, PCs were the only molecules to

show some sort of trend towards differentiating the intermediate (MC) class from controls, and therefore may have potential relevance as biomarkers at an early diagnosis. Based on the group size in the current study (n=70), this observation would require further validation in future population studies to establish the utility of this set of PCs as early calcification markers. Nonetheless, this observation could provide important information when it comes to dysregulated biological mechanisms of disease manifestation. The PCs did not follow a global trend, rather than a trend being affected by the characteristics of the FAC. As will be discussed in the following paragraphs, PCs essentially function as fatty acid carriers, thus it could be inferred that FAC, and fatty acids, play an important role in CCAD.

Prior to establishing a mechanistic explanation for disease manifestation, the possibility that some of the discriminating molecules may reflect confounding factors such as diet or lifestyle rather than the direct aetiology of the underlying disease must be considered. Additionally, trends in the plasma lipidome have been previously shown to be due to BMI, diabetes or hypertension status[39]. As can be seen in the patients' characteristics table (Table 5. 1), an effort to match the groups in these aspects was also conducted, although perfect matching of the disease and control groups was not possible.

Initial focus was given to explaining the dysregulation of the PCs. A schematic representation of a hypothesis derived from the results can be found in Figure 5. 11. Levels of PC(16:0/20:4) were higher in the disease class, in combination with a reduction in 18-carbon fatty acyl chain (FAC) PCs. These molecules can release arachidonic acid (AA) through hydrolysis of their ester bonds. Arachidonic acid can be released directly, in the case of PC(16:0/20:4)[43], while for the rest of the PCs, AA can be produced after hydrolysis and free fatty acid (FFA) elongation [23, 24]. A well known enzyme that can hydrolyse the sn-2 bond of PCs is Phospholipase A(2) ($PLA_2$)[187]. A group of IVA $PLA_2$ enzymes is specific for only AA hydrolysis from ester bonds[187] which can explain the different trend to rest of PCs. Of note, PIs also function as donors of FACs (again through $PLA_2$ enzyme catalysis), and could also be mechanistically implicated as such[187]. Downstream of AA lays the network of eicosanoids, well known for their involvement in inflammation.

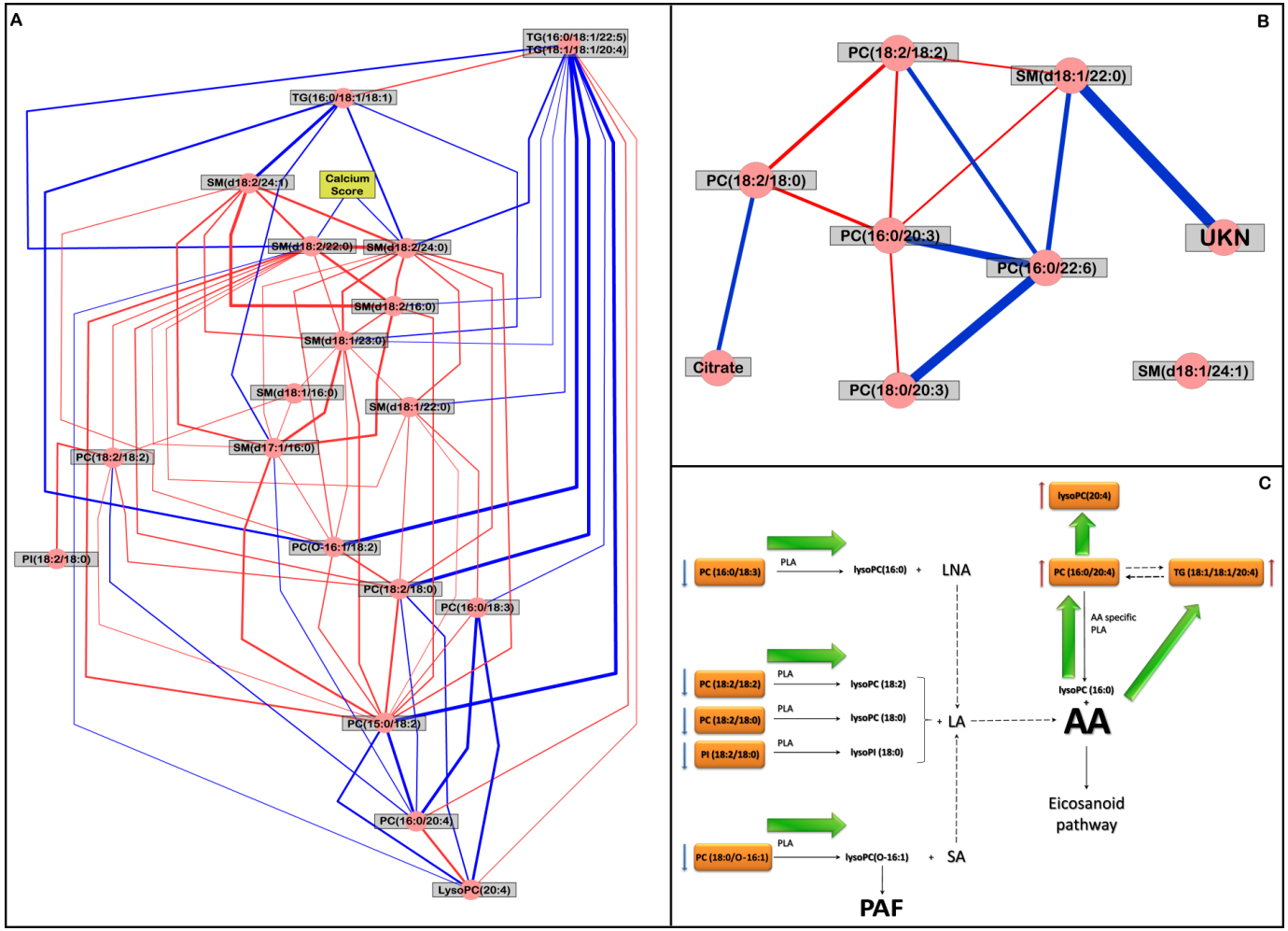


*Figure 5. 11: Pathway analysis networks. Schemes (A) and (B) represent networks constructed using Spearman correlation coefficients (r) of candidate biomarkers of calcific coronary artery and calcific aortic valve disease respectively. Circles (nodes) are labelled according to the metabolite they represent. Lines' (edges) colours represent positive (red) or negative (blue) correlation. Thickness of the edge represents increase in absolute value. (C) Pathway assembled using candidate biomarkers and literature information. Black arrows represent direction of canonical metabolism; green arrows represent direction of equilibrium shifting due to disease. Blue and red arrows represent reduction or increase of metabolite level. Solid arrows represent direct reactions, and dashed indirect. Orange boxes represent detected metabolites. PC: Phosphatidylcholine, SM: Sphingomyelin, TG: Triglyceride, PI: Phosphatidylinositol, AA: Arachidonic Acid, LA: Linoleic Acid, LNA: Linolenic Acid, SA: Stearic Acid, PAF: Platelet-Activating Factor and PLA: Phospholipase A. Note that position of double bonds cannot be determined.*

According to this hypothesis PCs and PIs containing 18-carbon chains should be found in lower intensities in the disease class as they are transformed into AA through the fatty acid elongation pathway. There are actually four 18-carbon containing PCs and PI(18:2/18:0) that were found to be present in lower intensities in the SC class. This hypothesis is consistent with the fact that there are enzymes specific to the hydrolysis of only linoleic acyl-containing ester bonds (such as the group IVD of $PLA_2$)[187]. A trend was also observed with PC(16:0/18:2), which was detected with lower intensity in the disease group. This PC presented a p(corr) of -0.63, but did not meet the t-test threshold, with p=0.16.

On the other hand, an otherwise normal pathway of AA release, by hydrolysis of the 20:4 FAC of the corresponding PC, will move the equilibrium to compensate with this increase in AA. This will cause an increase in the corresponding PC. As mentioned levels of PC(16:0/20:4) were higher. The PC(18:0/20:4) also presented the same trend but again did not meet the t-test threshold of 0.05 (p(corr)=0.68, p-value=0.08). The mechanistic rationale can also be supported by the high negative r of PC(16:0/20:4) to other PCs (Figure 5. 11A and Figure 5. 12). Moreover, lysoPC(20:4) and 20:4 FAC containing TG were also detected in higher intensities in the disease group.

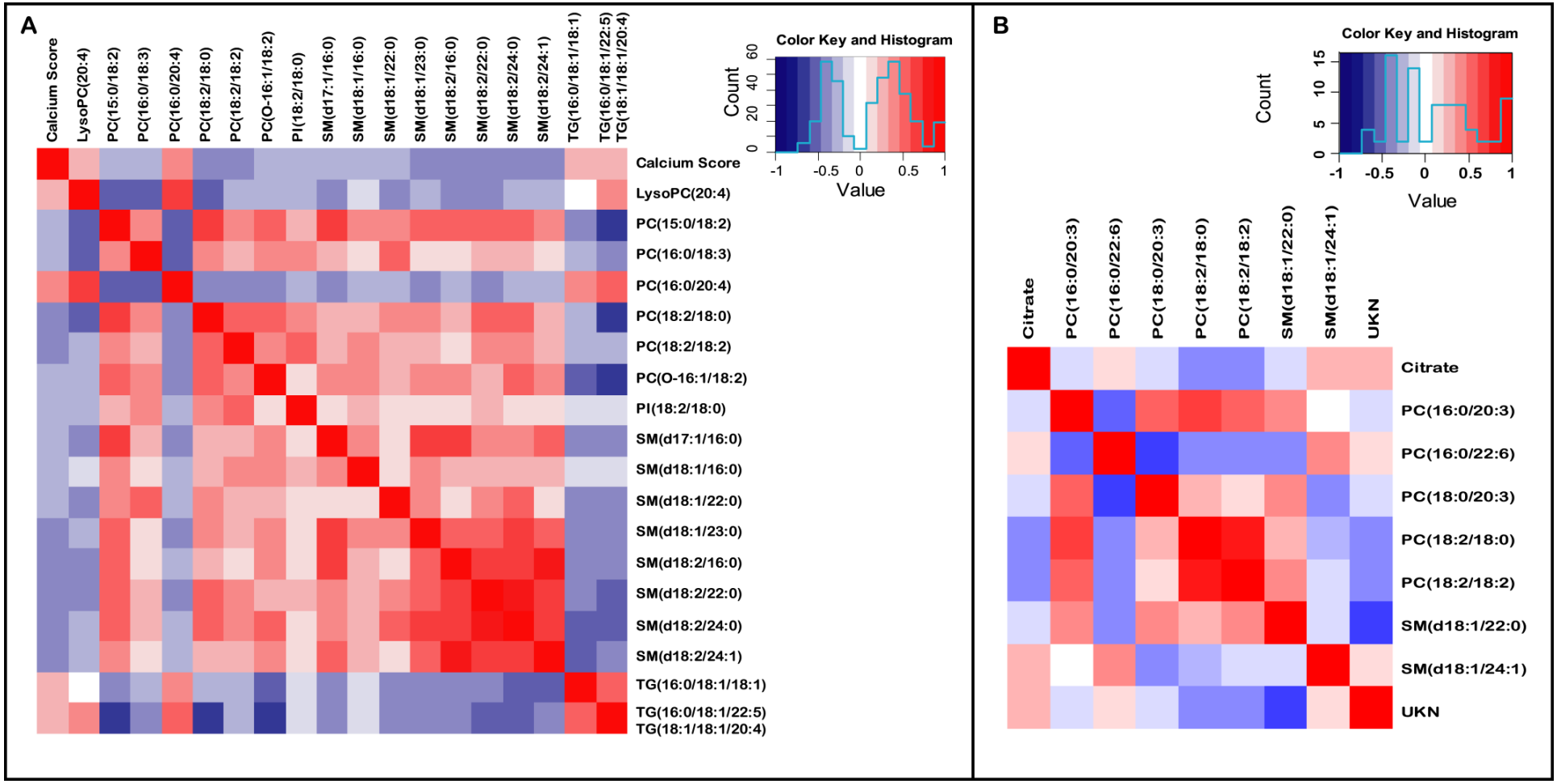


*Figure 5. 12: A heatmap representation of the correlation matrices obtained from Spearman correlations. (A) The calcific coronary artery disease candidate biomarkers correlation matrix, with all three groups included in the analysis, and (B) the calcific aortic valve disease candidate biomarkers correlation matrix. Intensity of red colour represents positive correlation and intensity of blue represents negative correlation of the biomarker pair.*

A product of PC hydrolysis is a lysoPC with the remaining FAC as a FFA. It has been previously shown by Chaudhuri et al, that an increase of intracellular $Ca^{2+}$ can be induced by lysoPCs (lysoPC(16:0) was used for those experiments)[188]. This mechanism also appears to cause upregulated expression of Calpain [188]. Calpain is involved in apoptotic pathways[139]. Additionally, authors show inhibition of EC migration, a mechanism of repair of the vessel wall, with lysoPC increase.

Another important molecule implicated in inflammation is the Platelet Activating Factor (PAF). PAF is produced by acetylation of the free position of ether-linked lysoPCs. One of the PCs found to be present in lower concentrations in the SC class was PC(18:2/O-16:1). This reduction infers a production of lysoPC(O-16:1) which is the direct precursor of PAF or PC(P-16:0/2:0).

Ideally, in addition to deriving mechanistic information from this study, it would be essential to explore the utility of the CCAD discriminating metabolites as biomarkers. Present findings come from blood samples, which is a minimal invasive and highly accessible matrix. Additionally, individuals at risk of cardiovascular diseases are subjected to routine blood test for related markers. Therefore, collection of samples could be tied into existing protocols. Overall statistical models from MVDA gave good predictive values (Table 5. 3). The high statistical power of SMs in differentiating CCAD from control is obvious, with p-values as low as 0.000001. Some of the PCs were also significant in discriminating case from control with p-values starting at 0.001. Although the PCs are not such strong candidate biomarkers as the SMs for differentiating CCAD, the statistical trend found for PCs, to discriminate between the NC and MC class is important as it carries the potential of early diagnosis and currently further samples in order to validate this observation are collected.

### *5.4.2 Metabolic characterization of calcific aortic valve disease*

A smaller number of statistically significant molecules were detected for CAVD, which made the elucidation of disease-specific mechanisms difficult, as information was limited. A mixed trend in levels of PCs was observed. PCs with 18:0 and 18:2 FACs were found in lower levels in the serum of CAVD patients. Interestingly, two PCs with a 20:3 FAC were also found in lower intensities in the CAVD class. The higher levels of the PC(16:0/22:6) may have mechanistic significance and maybe able to assist in connecting some of the other disease related findings. The mechanism could be similar to the one proposed for CCAD (Figure 5. 11C), i.e. in the disease state, enzymes release FACs (in this case 20:3, 18:2 and 18:0), which are elongated causing elevation FAC of healthy pathways (in this case 22:6 bearing PC) as a response to maintain equilibrium. This is further

supported by the correlation analysis and network (Figure 5. 11C and Figure 5. 12), where PC(16:0/22:6) is found to be highly and inversely correlated to the metabolites found to be in lower levels in the CAVD class. Additionally, a low r=0.20 was found between PC(16:0/16:0) and PC(16:0/22:6). This further excludes the possibility of the 16:0 FAC being involved rather than 22:6. The class of inflammation-related molecules of docosanoids is derived from the 22:6 fatty acid and thus higher levels of lipid moieties with a 22:6 FAC could implicate this group of molecules in the pathophysiology of the disease.

From the preceding paragraph it becomes obvious that in CAVD fatty acid metabolism is, as with CCAD, dysregulated. However, different FACs are involved in this case. The significance of the different involvement of FACs may lie in the metabolism of unsaturated fatty acids. In the case of the metabolism of the n-3 family of unsaturated fatty acids, enzymes that can metabolise chains that have up to six double bonds are commonly involved. On the contrary, in the metabolism of the n-6 family of unsaturated FFAs, only production of FFAs bearing up to five double-bonds is possible[23, 24]. Unfortunately, this analytical MS format is inadequate in providing information concerning the location of the double bonds.

The elevation of citric acid, a product of the tricarboxylic acid cycle (TCA cycle), is also mechanistically intriguing. It is produced after donation of an acetyl group from acetyl-CoA to oxaloacetic acid. The TCA cycle is an integral step in the degradation of a number of metabolites and processes, such as glycolysis, gluconeogenesis, fatty acid/ phenylpropanoid/ terpenoid/ steroid/ BCAA biosynthesis and degradation. From the correlation analysis of the candidate biomarkers, the metabolite most highly correlated with citrate was PC(18:2/18:0) (r= -0.426), which may indicate that the altered levels of citrate in CAVD may be associated with perturbation in FFA metabolism. Reduction of inorganic pyrophosphate (PPi)[60, 181] and elevation of the Reactive Oxygen Species (ROS)[181] are known to be inducing osteoblastic phenotype, and are affected by dysregulations in the TCA cycle.

Despite the small numbers of patients in the CAVD disease group, the MVDA models were valid, as evidenced by model diagnostics (Table 5. 3) and generated a series of potential biomarkers with good t-test p-values, which would be good candidates for further validation (Table 5. 5). Again, a number of PCs showed statistical significance, with PC(16:0/22:6) correlation to the disease group with p(corr)=0.8 and a p=0.009. Although the most significant metabolite was found to be SM(d18:1/22:0) (p-value=0.0004), the fact that it is the only SM with a lower levels in the disease

class, and the presence of another SM with the opposite trend, is worrying. This could mean that apoptotic pathways are not implicated in the calcification process of the aortic valve.

### *5.4.3 Comparison of serum metabolite profiles from calcific coronary artery disease and calcific aortic valve disease patients.*

A comparison between the two groups is very interesting as it can provide evidence of differences between seemingly analogous disease phenotypes and help in answering the question as to whether these two conditions are mechanistically distinct. When directly compared, the two different classes of patients demonstrated systematic metabolic differences. The two disease groups essentially showed a different profile (). Lower serum levels of SM were apparent in CCAD patients but not in samples from CAVD patients, and this constitutes a major difference between the two groups. Additionally, two PCs and a PI constituted of a 20:4 chain were amongst the lipids detected to be higher in the CCAD, but not CAVD, further supporting the involvement of eicosanoids only in CCAD. The two PCs, PC(18:2/18:0) and PC(18:2/18:2), were the only similarities between the two groups (Figure 5. 13), and as stated the release of 18-carbon FACs is considered by present findings to play a key role to the initiation of both diseases, by activation of their Phospholipase A enzyme.

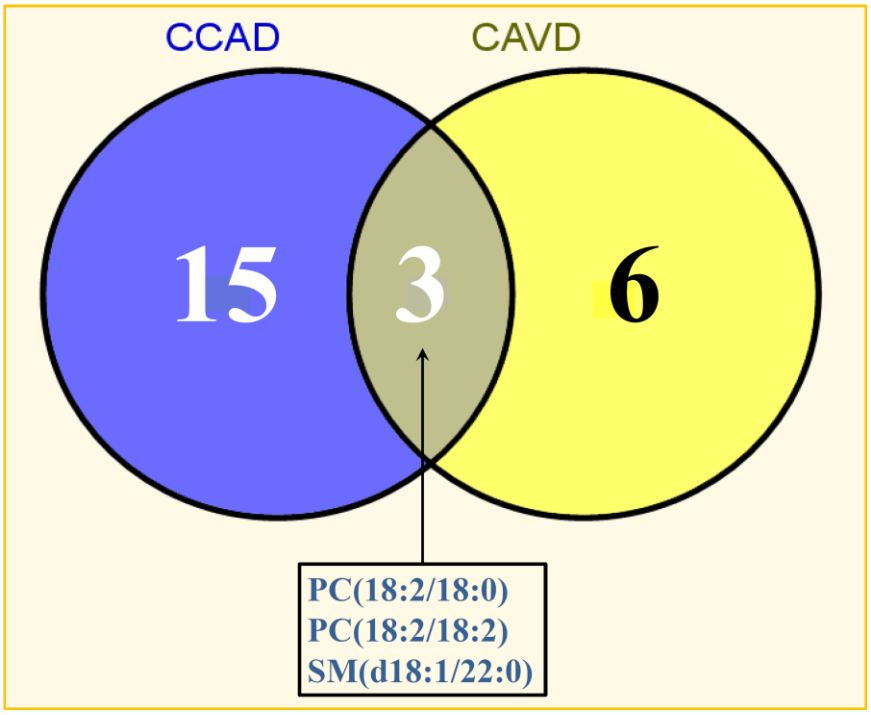


*Figure 5. 13: Venn diagram indicating identical metabolites detected statistically significant in separate comparisons between CCAD serum samples and CAVD serum samples to controls. Metabolites detected with the same trend in both analyses are indicated.*

PLS-DA analysis of the two disease groups and the control class in the same model demonstrated some very interesting results. All assigned metabolites can be found in Table 5. 7. Starting with the PC lipid class, which showed mixed trends, the loadings plots of the PLS-DA analysis showed that PCs bearing the same acyl chain behave similarly. This becomes evident by their proximity in the loadings plots (Figure 5. 10 C and D). Some examples of PCs bearing the same FACs are, the PC(18:2/18:0) and PC(18:2/18:2), PC(16:0/20:4) and PC(18:0/20:4), PC(16:0/20:3) and PC(18:0/20:3). At the same time, their mixed trends are also demonstrated by their appearance in multiple and opposite directions of the plots.

Similarities in FAC behaviour were observed in other lipid classes, namely PE(16:0/22:6) and PE(18:0/22:6), and TG(16:0/18:1/18:1), TG(16:0/18:1/18:2) and TG(16:0/16:1/18:1). These observations suggest that FACs are important in characterising calcific disease and that these FACs are differentially dysregulated between CCAD and CAVD.

### *5.4.4 Evidence of Calcium involvement*

$Ca^{2+}$ signalling is crucial in normal and pathological conditions. For both calcification diseases studied here, evidence is consistent with $Ca^{2+}$ signalling dysfunction. Phosphatidylinositols (PIs) are intermediate products towards the production of inositol-triphosphate ($IP_3$), known for its role in calcium channels, as increase of $IP_3$ can induce the export of $Ca^{2+}$ from endoplasmic reticulum[43]. In this study, PI(18:2/18:0) was detected in lower levels, in the SC class compared to the control (NC) class.

As previously mentioned, lysoPCs can increase the amount of $Ca^{2+}$ in the cell. One part of this hypothesis is that increase amounts of $Ca^{2+}$ in the cell might be causing a process known as mitochondrial calcium overload, and the effect of this is cell death. Signals of cell death do not point to the same pathway of cell death for both CCAD and CAVD disease. CCAD cell death is suggested via apoptosis inferred by the lower levels of SM in the diseased class. In the case of CAVD evidence of apoptosis are not present, but other means of cell death may be involved, such as necrosis. Apoptotic or necrotic cells can release amongst other factors, $Ca^{2+}$ and inorganic phosphate. The increase of these factors in the microenvironment of the cells in the area could be the signal initialising the osteoblastic phenotype[57, 60, 181] inducing the ectopic calcification phenotype. Matrix vesicles from apoptotic or necrotic cells are known for their ability to function as the initial site of calcification[182].

Citrate has the ability to bind and form a complex with $Ca^{2+}$. In this context, the elevation of citrate may have an effect on calcium homeostasis. If this is indeed the case, citrate could be produced as a response in order to reduce the free active form of $Ca^{2+}$.

### *5.4.5 Correlation and Network analysis*

Technologies that can produce a tremendous amount of data can assist in the effort of the scientific community to escape the reductionist approach to attain a more global, systemic level of understanding of physiological and pathological processes. However, such technologies typically generate vast amounts of data that require systematic mathematical analysis with subsequent linkage to biological networks to facilitate interpretation, and the amount of data can be overwhelming. Network analysis can provide information on a system as a whole and represent the molecular complexity of disease[189]. In the current study, correlation analysis was applied to metabolites that were found to be statistically different between case and control for each of the two diseases or between CCAD and CAVD in order to infer connections between metabolites in the network, and subsequently to the analysed system. High correlation values (r) between detected candidate biomarkers that are known to operate in the same or associated metabolic pathways can also provide more confidence to the analysis.

Lipid networks can interact significantly due to their common downstream targets and enzymes [43]. Network analysis from this study extracted associations between several biologically connected molecules and, despite the overwhelming appearance of such a network, was useful in identifying differential pathways implicated in CCAD and CAVD. Firstly, the global trend of the SMs was verified with high correlations between them (Figure 5. 11A and Figure 5. 12). Also, known biological interactions, from reference pathways, verified results, e.g. the connection between TGs, lyso-PCs and PCs. Other interesting findings are the inverse correlation between 20:3 and 22:6 FA PCs, and between 20:4 and 18 carbon FAC PCs. This might be suggestive of competing pathways. Lastly, although it is know that PIs and PCs can share common enzymes for ester bond hydrolysis, the correlation between them in this dataset was quite low. As can be seen from Figure 5. 11A PI(18:2/18:0) is quite excluded from the PC sub-network. This may suggest, as previously mentioned, involvement in other pathways, such as $Ca^{2+}$ signalling[43]. A closer look can also show that the highest r (r=0.58) of PI(18:2/18:0) is with PC(18:2/18:2) and second highest (r=0.40) is with PC(18:2/18:0), most likely due to enzyme specificity to the FAC. These are just some examples of enhanced capacity for data interpretation by use of correlation networks when it comes

to rationalising findings from lipid profiling studies, as reference biochemical pathways can run short on providing adequate mapping of these metabolites.

Correlation analyses have also assisted in validating the structural assignments of metabolites. One example is the connection of SM(d17:1/16:0) to PC(15:0/18:2) with high correlation value (r=0.63). SMs with 17:1 backbone are found in trace concentrations in humans. However, evidence provided by tandem MS supports this finding. The high r to a 15:0-acyl PC could signify the *de novo* synthesis of a 17:1 backbone SM from a 15:0 FAC and serine, to the corresponding ceramide. Ceramides are the direct precursors and SM.

For the CCAD, the calcium score of all patients was available. In correlation analysis CS was included in an effort to evaluate the best biomarker to correlate with this information. However, CS was not highly correlated with any of the discriminating candidate markers, in an analysis including a total of 70 samples. The highest correlation values were found with SMs. Specifically, SM(d18:2/22:0) and SM(d18:2/24:0) with correlation values of r=-0.45 and -0.44, respectively. On one hand, it could be argued that the CS is an imperfect metric and the purpose of the current studies was to identify superior diagnostics for CCAD and CAVD. There is an unmet need to discover better diagnostic biomarker, which may contribute to a framework for stratification of patients in a concise and clinically meaningful way.

## 5.5 Conclusions

In the current study both similarities and differences in the metabolic profiles of the two calcific diseases (coronary artery and aortic valve) are demonstrated leading to the development of a hypothesis that the diseases are mechanistically different. To our knowledge, this has not been shown before at the level of the metabolic profile. Herein, a cross-platform metabonomic study was applied for the first time in cardiovascular calcific diseases. It is obvious that metabolic profiling approaches can deliver disease related biomarkers in blood serum from CCAD and CAVD patients, and combined with correlation and network analysis, can be useful in generating hypotheses relating to mechanisms of pathology. The lipid profiling, UPLC-MS method, provided more information in this case than the NMR analysis, which is unsurprising given that most of the disease signature manifested in the composition of serum lipids. The UPLC-MS data were highly reproducible, with all candidate markers, except one, having a coefficient of variation of < 9%. Thus, this will be the method of choice for future untargeted large population studies, which will follow to validate findings.

CCAD was characterized by lower serum levels of SMs, higher concentrations of TGs and a PC profile that was modulated differentially across the severe calcification versus control group. The trend in dysregulated PC profile in CCAD was noted in individuals with a mild to moderate disease stage and may have utility as an early indicator of disease. The metabolic signature of CAVD was dominated by changes in the PC profile and higher serum citrate levels. It appears that, while disease pathways are different between CCAD and CAVD, leading to a pathological process in which only the end-stage is osteoblastic/osteogenic phenotype and calcification, a number of processes co-exist. Dysregulation of specific and previously unreported FAC provides novel hypotheses that could shed light to the pathophysiology of vascular and valvural calcification.

The results from this study beg further investigation into the fatty acid metabolism, the eicosanoid pathway, as well as other inflammation-related metabolites, such as docosanoids and inflammation-related proteins. The specificity of the PLA2 enzyme towards the 20:4 and 22:6 FAC PCs also needs to be addressed. Another obvious lead is the suggestion of apoptotic pathways in CCAD emerging from findings related to the ceramide and mitochondrial $Ca^{2+}$ overload signalling pathways. All of these along with established markers of osteoblastic/osteogenic phenotype will form the core of experimental design for future studies.

*Table 5. 4: List of assigned significant metabolites obtained from analysis of serum samples, after comparison of calcific coronary artery disease class to the control group.*

| **Met Name** | **Mol Formula [complex adduct]** | **Ret Time (min)** | **m/z (found)** | **m/z (theor)** | **Δppm** | **p(corr)** | **p(1)** | **p [t-test, NC Vs SC]** | **Fold Change** | **p [t-test, NC Vs MC]** | **p [t-test, MC Vs SC]** | **CV % QCs (9)** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **UPLC-MS Positive Mode** | | | | | | | | | | | | |
| **Higher in SC** | | | | | | | | | | | | |
| PC(16:0/20:4) *,† | C44H81NO8P+ | 5.75 | 782.5704 | 782.5700 | 1 | | | | | | | |
| | C44H81NO8P+ (C13 Isotope) | 5.75 | 783.5745 | 783.5733 | 2 | 0.85 | 0.32 | 0.005036 | 1.29 | 0.023786 | 0.251396 | 7 |
| LysoPC(20:4) * | C28H50NO7PNa+ [C28H50NO7P+Na]+ | 1.00 | 566.3286 | 566.3223 | 11 | 0.70 | 0.02 | 0.011478 | 1.22 | 0.681547 | 0.017397 | 8 |
| TG(16:0/18:1/22:5) TG(18:1/18:1/20:4) | C59H106NO6+ [C59H102O6+NH4]+ | 15.36 | 924.7982 | 924.8020 | -4 | 0.73 | 0.06 | 0.019369 | 1.29 | 0.461709 | 0.039684 | 8 |
| TG(16:0/18:1/18:1) | C55H106NO6+ [C55H102O6+NH4]+ | 15.68 | 876.8015 | 876.8020 | -1 | 0.66 | 0.17 | 0.044971 | 1.40 | 0.743774 | 0.058982 | 5 |
| **Lower in SC** | | | | | | | | | | | | |
| SM(d18:2/24:0) * | C47H94N2O6P+ | 12.81 | 813.6858 | 813.6850 | 1 | -0.74 | -0.09 | 0.000001 | 1.50 | 0.894457 | 0.000001 | 7 |
| SM(d18:2/22:0) | C45H90N2O6P+ | 9.77 | 785.6542 | 785.6537 | 1 | -0.74 | -0.10 | 0.000006 | 1.47 | 0.449461 | 0.000020 | 4 |
| SM(d18:2/24:1) | C47H92N2O6P+ | 9.71 | 811.6691 | 811.6693 | 0 | -0.72 | -0.12 | 0.000041 | 1.21 | 0.797471 | 0.000031 | 5 |
| SM(d18:1/23:0) | C46H94N2O6P+ | 13.03 | 801.6847 | 801.6850 | 0 | -0.63 | -0.07 | 0.000041 | 1.19 | 0.718762 | 0.000044 | 3 |
| SM(d18:2/16:0) | C39H78N2O6P+ | 4.30 | 701.5604 | 701.5598 | 1 | -0.68 | -0.07 | 0.000347 | 1.31 | 0.631770 | 0.000308 | 6 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PC(18:2/18:0) *,† | C44H85NO8P+ | 7.87 | 786.6010 | 786.6013 | 0 | | | | | | | |
| | C44H85NO8P+ (C13 Isotope) | 7.87 | 787.6046 | 787.6047 | 0 | -0.77 | -0.31 | 0.001371 | 1.30 | 0.259191 | 0.005971 | 8 |
| PC(O-16:1/18:2) | C42H81NO7P+ | 6.73 | 742.5748 | 742.5751 | 0 | -0.72 | -0.08 | 0.001941 | 1.46 | 0.301605 | 0.032098 | 2 |
| SM(d17:1/16:0) | C38H78N2O6P+ | 4.70 | 689.5597 | 689.5598 | 0 | -0.67 | -0.05 | 0.002220 | 1.22 | 0.814786 | 0.010492 | 3 |
| PC(18:2/15:0) * | C41H79NO8P+ | 5.27 | 744.5545 | 744.5543 | 0 | -0.78 | -0.07 | 0.002464 | 1.36 | 0.612282 | 0.011036 | 3 |
| SM(d18:1/22:0) | C45H92N2O6P+ | 12.32 | 787.6696 | 787.6693 | 0 | -0.65 | -0.09 | 0.012648 | 1.19 | 0.502127 | 0.115613 | 3 |
| PC(18:2/18:2) * | C44H81NO8P+ | 4.96 | 782.5682 | 782.5700 | -2 | -0.75 | -0.16 | 0.025146 | 1.47 | 0.357176 | 0.104597 | 6 |
| PC(16:0/18:3) | C42H79NO8P+ | 5.00 | 756.5546 | 756.5543 | 0 | -0.68 | -0.13 | 0.038234 | 1.55 | 0.056644 | 0.613973 | 3 |
| **UPLC-MS Negative Mode** | | | | | | | | | | | | |
| **Higher in SC** | | | | | | | | | | | | |
| PC(16:0/20:4) * | C45H81NO10P- [C44H80NO8P+FA]- | 5.76 | 826.5625 | 826.5598 | 3 | 0.73 | 0.26 | 0.005061 | 1.18 | 0.087226 | 0.090059 | 4 |
| LysoPC(20:4) *,‡ | C27H47NO7P- | 1.00 | 528.2996 | 528.3090 | -18 | 0.66 | 0.03 | 0.012770 | 1.41 | 0.243202 | 0.059168 | 5 |
| lower in SC | | | | | | | | | | | | |
| SM(d18:2/24:0) * | C48H94N2O8P- [C47H93N2O6P+FA]- | 12.82 | 857.6757 | 857.6748 | 1 | -0.74 | -0.10 | 0.000001 | 1.35 | 0.583687 | 0.035079 | 3 |
| PC(18:2/15:0) * | C42H79NO10P- [C41H78NO8P+FA]- | 5.25 | 788.5475 | 788.5442 | 4 | -0.84 | -0.06 | 0.002136 | 1.37 | 0.158770 | 0.048678 | 5 |
| PC(18:2/18:0) * | C45H85NO10P- [C44H84NO8P+FA]- | 7.90 | 830.5934 | 830.5911 | 3 | -0.84 | -0.30 | 0.002328 | 1.17 | 0.059537 | 0.070687 | 3 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PC(18:2/18:2) * | C45H81NO10P-[C44H80NO8P+FA]- | 4.99 | 826.5604 | 826.5598 | 1 | -0.78 | -0.13 | 0.003234 | 1.41 | 0.074336 | 0.126333 | 5 |
| PI(18:2/18:0) | C45H82O13P- | 6.01 | 861.5509 | 861.5493 | 2 | -0.66 | -0.13 | 0.013215 | 1.39 | 0.378844 | 0.068470 | 3 |
| SM(d18:1/16:0) | C40H80N2O8P-[C39H79N2O6P+FA]- | 5.37 | 747.5662 | 747.5662 | 0 | -0.69 | -0.20 | 0.024765 | 1.10 | 0.903099 | 0.018255 | 5 |

(*) Found with statistical significance in both polarities.

(†) Metabolite removed due to detector saturation. Used only for assignment purposes.

(‡) Metabolite detected as a fragment

Coefficient of variation (CV%) is calculated based on nine injections of the same pooled sample, acquired through the run.

P(corr) and p(1) refer to the correlation coefficient and covariance of the metabolites to disease classes, respectively.

The position of double bonds cannot be determined. Position of fatty acyl chain cannot be determined; fatty acyl chains are presented from lowest to highest molecular weight (lower number of carbons to higher, and higher number of double bonds, to lower).

Two-tailed t-tests were conducted, assuming unequal variance.

NC: No calcification; MC: Mild calcification – calcific coronary artery disease; SC: Severe calcification – calcific coronary artery disease.

PC: phosphatidylcholine; TG: triglyceride; SM: sphingomyelin; PI: phosphatidylinositol; FA: formate.

*Table 5. 5: List of assigned significant metabolites obtained from analysis of serum samples, after comparison of calcific aortic valve disease class to the control group.*

| Met Name | Mol Formula [non hydrogen adduct] | Ret Time (min) | m/z (found) | m/z (theoretical) | Δppm | p(corr) | p(1) | p [t-test, NC vs CAVD] | Fold Change | CV% QCs (9) |
|---|---|---|---|---|---|---|---|---|---|---|
| **UPLC-MS** | **Positive Mode** | | | | | | | | | |
| **Higher in CAVD** | | | | | | | | | | |
| PC(16:0/22:6) | C46H81NO8P+ | 5.34 | 806.5699 | 806.5700 | 0 | 0.78 | 0.38 | 0.0089 | 1.23 | 8 |
| SM(d18:1/24:1) | C47H93N2O6PNa+ [C47H93N2O6P+Na]+ | 12.16 | 835.6667 | 835.6669 | 0 | 0.62 | 0.08 | 0.0299 | 1.12 | 4 |
| **Lower in CAVD** | | | | | | | | | | |
| PC(18:0/20:3) | C46H87NO8P+ | 8.47 | 812.6178 | 812.6169 | 1 | -0.70 | -0.25 | 0.0079 | 1.30 | 6 |
| **UPLC-MS; Negative Mode** | | | | | | | | | | |
| **Higher in CAVD** | | | | | | | | | | |
| UKN | | 0.66 | 307.0981 | - | - | 0.79 | 0.43 | 0.0336 | 53.79 | 1 |
| **Lower in CAVD** | | | | | | | | | | |
| SM(d18:1/22:0) | C46H92N2O8P- [C45H91N2O6P+FA]- | 12.42 | 831.6618 | 831.6591 | 3 | -0.59 | -0.10 | 0.0004 | 1.16 | 1 |
| PC(16:0/20:3) | C45H83NO10P- [C44H82NO8P+FA]- | 6.46 | 828.5782 | 828.5755 | 3 | -0.52 | -0.16 | 0.0045 | 1.20 | 2 |
| PC(18:2/18:0) | C45H85NO10P- [C44H84NO8P+FA]- | 7.90 | 830.5934 | 830.5911 | 3 | -0.74 | -0.29 | 0.0073 | 1.13 | 3 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| PC(18:2/18:2) | C45H81NO10P-[C44H80NO8P+FA]- | 4.99 | 826.5604 | 826.5598 | 1 | -0.63 | -0.11 | 0.0141 | 1.15 | 5 |
| **$^{1}$H NMR** | | | | | | | | | | |
| **Higher in CAVD** | | | **δ $^{1}$H (ppm)** | | | | | | | |
| Citrate | | | 2.657 (d) ½ CH2,<br>2.522 (d) ½ CH2 | | | 0.6334 (max) | 0.0004 | 0.004-0.03 | 1.30-1.48 | n/a |

Coefficient of variation (CV%) is calculated based on nine injections of the same pooled sample, acquired through the run.

P(corr) and p(1) refer to the correlation coefficient and covariance of the metabolites to disease classes, respectively.

The position of double bonds cannot be determined. Position of fatty acyl chain cannot be determined; fatty acyl chains are presented from lowest to highest molecular weight (lower number of carbons to higher, and higher number of double bonds, to lower).

Two-tailed t-tests were conducted, assuming unequal variance. NC: No calcification; CAVD: Calcific aortic valve disease.

PC: phosphatidylcholine; SM: sphingomyelin; UKN: unknown; FA: formate.

*Table 5. 6: List of assigned significant metabolites obtained from analysis of serum samples, after comparison of calcific coronary artery disease to calcific aortic valve disease class.*

| **Met Name** | **Mol Formula [non hydrogen adduct]** | **Ret Time (min)** | **m/z (found)** | **m/z (theoretical)** | **Δppm** | **p(corr)** | **p(1)** | **p [t-test, SC Vs CAVD]** | **Fold Change** | **CV% QCs (9)** |
|---|---|---|---|---|---|---|---|---|---|---|
| **UPLC-MS Positive Mode** | | | | | | | | | | |
| **Lower in CCAD** | | | | | | | | | | |
| SM(d18:2/16:0) * | C39H78N2O6P+ | 4.30 | 701.5604 | 701.5598 | 1 | 0.68 | 0.08 | 0.0001 | 1.48 | 6 |
| SM(d18:2/24:1) * | C47H92N2O6P+ | 9.71 | 811.6691 | 811.6693 | 0 | 0.77 | 0.13 | 0.0003 | 1.34 | 5 |
| SM(d17:1/16:0) * | C38H78N2O6P+ | 4.70 | 689.5597 | 689.5598 | 0 | 0.67 | 0.05 | 0.0147 | 1.38 | 3 |
| PC(15:0/18:2) | C41H79NO8P+ | 5.27 | 744.5545 | 744.5543 | 0 | 0.72 | 0.06 | 0.0168 | 1.34 | 3 |
| PC(16:0/16:0) | C40H81NO8P+ | 7.15 | 734.5705 | 734.5700 | 1 | 0.59 | 0.08 | 0.0231 | 1.20 | 5 |
| PC(16:0/18:2) † | C42H81NO8P+ | 5.99 | 758.5724 | 758.5700 | 3 | | | | | |
| | C42H81NO8P+ (C13 Isotope) | 5.99 | 759.5754 | 759.5733 | 3 | 0.79 | 0.32 | 0.0258 | 1.20 | 8 |
| SM(d18:2/24:0) | C47H94N2O6P+ | 12.81 | 813.6858 | 813.6850 | 1 | 0.61 | 0.06 | 0.0339 | 1.31 | 7 |
| **UPLC-MS Negative Mode** | | | | | | | | | | |
| **Lower in CCAD** | | | | | | | | | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SM(d18:2/16:0) * | C40H78N2O8P-<br>[C39H77N2O6P+FA]- | 4.32 | 745.5505 | 745.5496 | 1 | 0.71 | 0.11 | 0.0005 | 1.32 | 2 |
| PC(15:0/18:2) | C42H79NO10P-<br>[C41H78NO8P+FA]- | 5.25 | 788.5475 | 788.5442 | 4 | 0.57 | 0.04 | 0.0065 | 1.33 | 5 |
| SM(d18:2/24:1) * | C48H92N2O8P-<br>[C47H91N2O6P+FA]- | 9.73 | 855.6601 | 855.6591 | 1 | 0.50 | 0.09 | 0.0075 | 1.36 | 2 |
| SM(d17:1/16:0) * | C39H78N2O8P-<br>[C38H77N2O6P+FA]- | 4.72 | 733.5496 | 733.5496 | 0 | 0.72 | 0.07 | 0.0094 | 1.40 | 3 |
| SM(d18:1/22:1)<br>SM(d16:1/24:1) | C46H90N2O8P-<br>[C45H89N2O6P+FA]- | 9.38 | 829.6466 | 829.6435 | 4 | 0.58 | 0.07 | 0.0395 | 1.06 | 19 |
| **Higher in CCAD** | | | | | | | | | | |
| PC(16:0/20:4) | C45H81NO10P-<br>[C44H80NO8P+FA]- | 5.76 | 826.5625 | 826.5598 | 3 | -0.69 | -0.28 | 0.0051 | 1.25 | 4 |
| PC(18:0/22:4) | C49H89NO10P-<br>[C48H88NO8P+FA]- | 9.21 | 882.6222 | 882.6224 | 0 | -0.65 | -0.04 | 0.0133 | 1.59 | 7 |
| PC(18:0/20:4) | C47H85NO10P-<br>[C46H84NO8P+FA]- | 7.58 | 854.5922 | 854.5911 | 1 | -0.60 | -0.20 | 0.0141 | 1.26 | 2 |
| PI(18:0/20:4) | C47H82O13P- | 5.82 | 885.5476 | 885.5493 | -2 | -0.56 | -0.21 | 0.0256 | 1.30 | 2 |

(*) Found with statistical significance in both polarities.

(†) Metabolite removed due to detector saturation. Used only for assignment purposes.

Coefficient of variation (CV%) is calculated based on nine injections of the same pooled sample, acquired through the run.

P(corr) and p(1) refer to the correlation coefficient and covariance of the metabolites to disease classes, respectively.

The position of double bonds cannot be determined.  Position of fatty acyl chain cannot be determined; fatty acyl chains are presented from lowest to highest molecular weight (lower number of carbons to higher, and higher number of double bonds, to lower).

Two-tailed t-tests were conducted, assuming unequal variance.

SC: Severe calcification – calcific coronary artery disease; CAVD: Calcific aortic valve disease.

PC: phosphatidylcholine; SM: sphingomyelin; FA: formate.

*Table 5. 7: List of assigned metabolites obtained from PLS-DA of analysed serum samples, after comparison of control, calcific coronary artery and calcific aortic valve disease groups.*

| **Metabolite Name** | **Mol Formula [non hydrogen adduct]** | **Ret Time (min)** | **m/z (found)** | **m/z (theoretical)** | **Δppm** |
|---|---|---|---|---|---|
| **UPLC-MS**<br>**Positive Mode** | | | | | |
| lysoPC(16:0) | C24H51NO7P+ | 1.19 | 496.3459 | 496.3403 | 11 |
| PC(16:0/16:0) * | C40H81NO8P+ | 7.15 | 734.5705 | 734.5700 | 1 |
| PC(16:0/18:1) * | C42H83NO8P+ | 7.38 | 760.5877 | 760.5856 | 3 |
| PC(16:0/18:2) * | C42H81NO8P+ | 5.99 | 758.5724 | 758.5700 | 3 |
| PC(16:0/18:3) | C42H79NO8P+ | 5.00 | 756.5546 | 756.5543 | 0 |
| PC(16:0/20:3) * | C44H83NO8P+ | 6.40 | 784.5858 | 784.5856 | 0 |
| PC(16:0/20:4) * | C44H81NO8P+ | 5.75 | 782.5704 | 782.5700 | 1 |
| PC(16:0/20:5) | C44H79NO8P+ | 4.83 | 780.5550 | 780.5543 | 1 |
| PC(16:0/22:5) | C46H83NO8P+ | 5.82 | 808.5865 | 808.5856 | 1 |
| PC(16:0/22:6) * | C46H81NO8P+ | 5.34 | 806.5699 | 806.5700 | 0 |
| PC(18:0/20:3) * | C46H87NO8P+ | 8.47 | 812.6178 | 812.6169 | 1 |
| PC(18:0/20:4) * | C46H85NO8P+ | 7.55 | 810.6019 | 810.6013 | 1 |
| PC(18:0/22:6) | C48H85NO8P+ | 7.01 | 834.6015 | 834.6013 | 0 |
| PC(18:2/18:0) * | C44H85NO8P+ | 7.87 | 786.6010 | 786.6013 | 0 |
| PC(18:2/18:2) | C44H81NO8P+ | 4.96 | 782.5682 | 782.5700 | -2 |
| PC(36:3) (†) | C44H83NO8P+ | 6.95 | 784.5856 | 784.5856 | 0 |
| SM(d18:1/16:0) * | C39H80N2O6P+ | 5.40 | 703.5761 | 703.5754 | 1 |
| SM(d18:1/24:0) | C47H96N2O6P+ | 13.32 | 815.7007 | 815.7006 | 0 |
| SM(d18:1/24:1) | C47H94N2O6P+ | 12.18 | 813.6863 | 813.6850 | 2 |
| SM(d18:2/16:0) * | C39H78N2O6P+ | 4.30 | 701.5604 | 701.5598 | 1 |
| SM(d18:2/22:0) * | C45H90N2O6P+ | 9.77 | 785.6542 | 785.6537 | 1 |
| SM(d18:2/24:0) * | C47H94N2O6P+ | 12.81 | 813.6858 | 813.6850 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| SM(d18:2/24:1) * | C47H92N2O6P+ | 9.71 | 811.6691 | 811.6693 | 0 |
| TG(16:0/16:1/18:1) | C53H102NO6 [C53H98O6+NH4]+ | 15.43 | 848.7715 | 848.7707 | 1 |
| TG(16:0/18:1/18:1) | C55H106NO6 [C55H102O6+NH4]+ | 15.68 | 876.8015 | 876.8020 | -1 |
| TG(16:0/18:2/18:1) | C55H104NO6 [C55H100O6+NH4]+ | 15.46 | 874.7870 | 874.7864 | 1 |
| | | | | | |
| **UPLC-MS**<br>**Negative Mode** | | | | | |
| PC(16:0/16:0) * | C41H81NO10P- [C40H80NO8P+FA]- | 7.16 | 778.5644 | 778.5598 | 6 |
| PC(16:0/18:1) * | C43H83NO10P- [C42H82NO8P+FA]- | 7.41 | 804.5785 | 804.5755 | 4 |
| PC(16:0/18:2) * | C43H81NO10P- [C42H80NO8P+FA]- | 6.00 | 802.5622 | 802.5598 | 3 |
| PC(16:0/20:3) * | C45H83NO10P- [C44H82NO8P+FA]- | 6.46 | 828.5782 | 828.5755 | 3 |
| PC(16:0/20:4) * | C45H81NO10P- [C44H80NO8P+FA]- | 5.76 | 826.5625 | 826.5598 | 3 |
| PC(16:0/22:6) * | C47H81NO10P- [C46H80NO8P+FA]- | 5.36 | 850.5614 | 850.5598 | 2 |
| PC(18:0/20:3) * | C47H87NO10P- [C46H86NO8P+FA]- | 8.50 | 856.6086 | 856.6068 | 2 |
| PC(18:0/20:4) * | C47H85NO10P- [C46H84NO8P+FA]- | 7.58 | 854.5922 | 854.5911 | 1 |
| PC(18:2/18:0) * | C45H85NO10P- [C44H84NO8P+FA]- | 7.90 | 830.5934 | 830.5911 | 3 |
| PE(16:0/22:6) | C43H73NO8P- | 5.76 | 762.5108 | 762.5074 | 4 |
| PE(18:0/22:6) | C45H77NO8P- | 7.57 | 790.5418 | 790.5387 | 4 |
| PI(18:0/20:4) | C47H82O13P- | 5.82 | 885.5476 | 885.5493 | -2 |
| SM(d18:1/16:0) * | C40H80N2O8P- [C39H79N2O6P+FA]- | 5.37 | 747.5662 | 747.5662 | 0 |
| SM(d18:2/16:0) * | C40H78N2O8P- [C39H77N2O6P+FA]- | 4.32 | 745.5505 | 745.5496 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| SM(d18:2/22:0) * | C46H90N2O8P-[C45H89N2O6P+FA]- | 9.85 | 829.6462 | 829.6435 | 3 |
| SM(d18:2/24:0) * | C48H94N2O8P-[C47H93N2O6P+FA]- | 12.82 | 857.6757 | 857.6748 | 1 |
| SM(d18:2/24:1) * | C48H92N2O8P-[C47H91N2O6P+FA]- | 9.73 | 855.6601 | 855.6591 | 1 |
| UKN | | 0.63 | 309.1149 | | |
| UKN | | 0.66 | 307.0981 | | |

(*) Found in both ionization modes.

(†)Fatty acyl chains could not be determined.

PC: phosphatidylcholine; TG: triglyceride; SM: sphingomyelin; PI: phosphatidylinositol; PE: phosphoethanolamine; UKN: unknown; FA: formate.

# Chapter 6 General Discussion

Cardiovascular disease (CVD) is one of the most studied clinical conditions. Most of the pathological events associated with CVD are directly or indirectly related to atherosclerosis[45]. In the context of novel technological advancements in instrumentation utilised in bioanalysis, atherosclerosis could still have 'virgin' grounds for science to explore. The holistic approach taken with metabolic profiling facilitates a systems biology framework, delivering interacting metabolites and pathways. Since 'omic' approaches generate vast quantities of data, especially when it comes to the field of metabolic profiling, CVD needs further investigation. Although numerous metabolic profiling studies have been conducted on biofluids, mostly focusing on diagnostic biomarker discovery, applications interrogating the pathophysiology of the disease using tissue samples are rare. In this chapter, combined analysis of both biofluids and plaque tissue to achieve a deep exploration of selected conditions associated with CVD, is presented.

Apart from elucidation of the pathology of the disease, experiments in tissue may serve as biomarkers if they are appropriate for *in vivo* imaging. However, there are cases where tissue cannot be available to the researcher, either due to the format of the surgical procedure or due to the difficulty obtaining appropriate controls. In such cases interacting, or otherwise related biofluids, can serve as matrices to explore and assist towards making inferences of the manifestation of disease. Biofluids, such as blood or urine, can be obtained by minimally-invasive or non-invasive procedures, and would be the optimal biological matrix to provide biomarkers of disease.

## 6.1 Overview of results

In Chapter 3 an exploratory study was conducted in order to evaluate the performance of metabolic profiling methodologies on carotid plaque tissue, and the ability to discriminate between recently stroke-symptomatic patients (in high risk for a life-threatening stroke) and asymptomatic patients (low risk; control group). A range of metabolites were detected by using a combination of NMR and UPLC-MS platforms. Further assessment using multivariate data analysis (MVDA) and univariate statistics demonstrated that these two methodologies could discriminate the symptomatic from asymptomatic groups by using their metabolic profiles, and several compounds were characteristic with strong statistical significance for one of the groups. Although the sample number was low (n=10) some pathways could be cautiously inferred in the dysregulation associated with symptomatic stroke. Intensities of precursor metabolites (arachidonic acid and PC(16:0/20:4)) of the

eicosanoid pathway, were detected in higher concentrations in tissue samples from symptomatic patients (Figure 6. 1). Three acyl-carnitines (AcC), intermediates of β-oxidation, were detected in higher intensities in the stroke symptomatic group. This finding is also very important, depicting involvement of mitochondrial dysregulation. Lastly, but with less statistical significance, inosine and uridine were detected higher, again directly correlating with the symptomatic group.

In Chapter 4 the metabolic basis of atherogenesis, and specifically plaque formation were sought. This was achieved by comparison of atherosclerotic plaques to intimal thickening tissue. Intima thickening tissue can be considered the immediate pre-plaque stage. It is therefore the optimal matrix to be used as control, when studying plaque formation and it is more easily acquired than normal tissue. Results from this study both validate well-established risk factors and literature findings, but also provide novel insights into plaque formation. Higher concentrations of cholesterol and oxidised cholesterol esters in plaque tissue, the involvement of purines and pyrimides, and implication of the ceramide (Cer) pathway, are know and well-established findings[148, 152-154, 159].

Novel findings included:

1. The suggested truncation of β-oxidation (with decrease in short-chain AcCs, such as butyrylcarnitine), along with a reduction of polyunsaturated lipids. Although it is well-known that mitochondria manifest metabolic abnormalities in atherosclerosis [190], there is only one report of acyl-carnitine dysregulation[191].
2. The detection of a previously unassociated to atherogenesis lipid, namely phosphoethanolamine-ceramide (PE-Cer), with higher levels in plaques, as compared to intimal thickening tissue, and with strong statistical significance. PE-Cers, a subgroup of sphingolipids and member of the Cer pathway, were detected as two moieties, PE-Cer(d18:1/16:0) and PE-Cer(d18:1/24:1). They generated t-test p-values as low as $9.8 \times 10^{-12}$. Most importantly they presented a high inverse Spearman correlation (r) to cholesterol, higher in absolute value than any other member of the Cer pathway. This infers that PE-Cers may be the link between the Cer pathway and cholesterol, that the scientific community has been trying to pinpoint[169, 170]. This finding could be of interest not only to explain the manifestation of disease, and provide a novel pharmaceutical target, but also to potentially explain ineffectiveness of current cholesterol lowering treatment.

Chapter 5 describes a metabolic profiling study of cardiovascular calcification (CVC). Serum samples from patients with calcific coronary artery (CCAD) and calcific aortic valve disease (CAVD) were compared to patients with no traces of calcification in the heart. The two diseases

manifested distinct profiles with similarities limited to only two 18-carbon phosphatidylcholines (PC) and a sphingomyelin (SM) (Figure 6. 1). One highlighted difference was the global dysregulation of SMs in CCAD disease not observed in CAVD, and not detected in mildly calcified patients with CCAD, either. PCs did not appear to follow the same trend across the two diseases. Their trends were more associated to the fatty acyl-chains incorporated in each PC. Hypotheses were generated stating the possible biochemical mechanism of metabolic dysregulation of PCs for CCAD and CAVD manifestation. Several FACs found to be dysregulated in this study have not been previously implicated in CVC and require further validation.

All novel and significant findings uncovered in these studies should be further validated in independent sample groups, via metabolite, or enzyme activity assays. Additionally, bottom-up approaches are required to verify key interactions detected between metabolites.

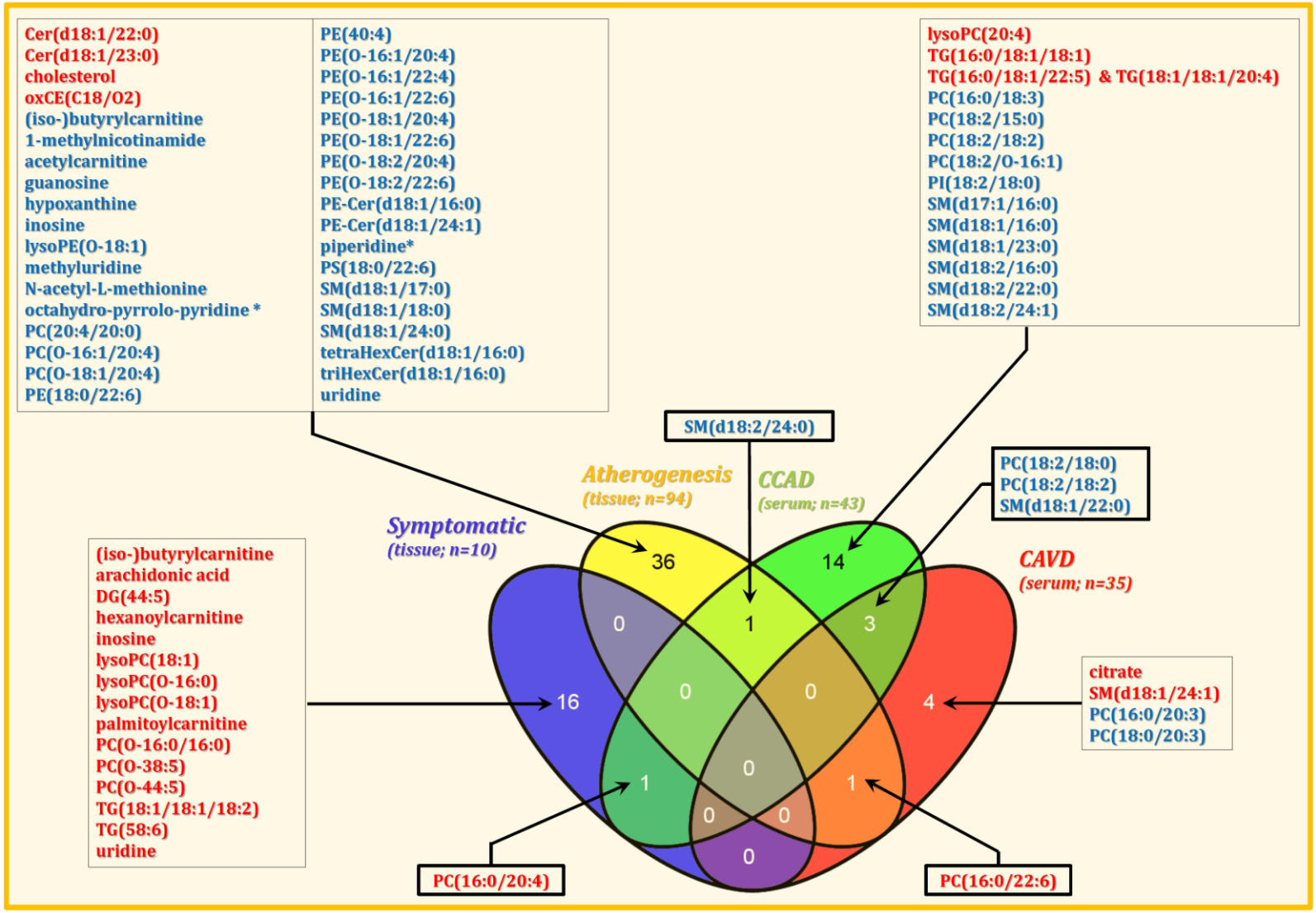


*Figure 6. 1: Venn diagram demonstrating all the metabolites detected dysregulated in this thesis, for the studied diseases. Metabolites were coloured according to the manifested trends as presented in disease (red: metabolites detected with higher intensities in disease, blue: metabolites detected with lower intensities in disease).*

## 6.2 Pathway mapping of cardiovascular disease

The major causes of CVD are related to atherosclerosis[45]. This may mislead one to believe that metabolic profiles of CVD would be identical or, to an extent, similar. If the totality of metabolites found to be discriminatory for the aspects of CVD studied in this thesis is combined, similarities between the three studies, i.e. i) symptomatic versus asymptomatic, ii) plaque formation (atherosclerosis), iii) CCAD and CAVD; metabolites demonstrating analogous dysregulated pattern are limited. This can be observed from the Venn diagram (Figure 6. 1) demonstrating the common irregularities between the diseases studied. This comparison should be cautiously interpreted due to the different matrices the studies are conducted on (tissue and serum), and the number of samples in each study.

The most obvious similarity, but only between plaque formation and CCAD, is the lower intensities of SMs in the disease groups. Low intensities of SMs are connected to Cer reduction and activation of the sphingomyelinase enzyme. The Cer pathway is connected to apoptosis[160]. However, how different FACs and Cer backbone lengths affect the biological activity of SMs should be addressed in future studies.

Two PCs, namely PC(16:0/20:4) and PC(16:0/22:6), were found to follow the same trend in Symptomatic and CCAD, and Atherosclerosis and CAVD groups, respectively. PCs incorporating the 20:4 and 22:6 FACs have been associated to the downstream pathways of eicosanoids and docosanoids, respectively. These PCs can function as donors of the immediate precursor molecules of these pathways, arachidonic acid and docosahexanoic acid. Phospholipase A(2) ($PLA_2$) is the enzyme hydrolysing the sn-2 ester bond releasing the free fatty acid. $PLA_2$ has become a recognised biomarker in CVD risk assessment, in the form of lipoprotein-associated $PLA_2$[192].

Inosine and uridine, two members of the purine and pyrimidine pathways, and (iso-)butyrylcarnitine a downstream product of β-oxidation were found to be dysregulated in both the atherosclerosis and symptomatic study. However, this dysregulation manifested opposite trends for the two conditions and is another indication of the distinct biological procedures occurring even in the direct progression and advancement of atherosclerosis, and, therefore, should be perceived as such.

## 6.3 Metabolic profiling techniques and methodologies in cardiovascular disease

During the course of this thesis, constant optimisation was conducted to applied methodologies, introduced by experimental development or from literature and/or collaborators. Significant improvements were made, especially in chromatographic separation, when comparing the HILIC methods and lipid profiling reversed-phase methods. Nonetheless, it became apparent, especially with the number of compounds detected in all three results chapters (Chapters 3-5), but predominately in chapter 4, that the right choice of UPLC-MS methodologies can maximise the amount and range of physicochemical properties of the compounds detected, but also minimise analysis time needed to obtain such a wealth of information. The number of compounds detected in a UPLC-MS assay can be expected to increase with the technological advancements focusing a lot on increasing sensitivity and dynamic range of concentration.

Current data and statistical analysis tools for untargeted UPLC-MS experiments provide the automation needed to process such data. Thus, disease profiles and candidate biomarkers can efficiently filtered and forwarded for further validation.

## 6.4 Closing remarks

From the whole of the experimental results presented in this thesis, it becomes apparent that metabolic profiling not only can identify disease profiles and interactions, but it can also contribute to uncovering novel candidate biomarkers. This has been achieved even in a well studied disease such as cardiovascular disease. Finally, metabolic profiling in combination with pathway mapping can fulfil its potential as a hypothesis generating discipline. Present findings are more of an end to a beginning, rather than a beginning of an end, in the quest for elucidating the pathology of cardiovascular disease and discovery of novel biomarkers.

# *References*


**1.** Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica.* Nov 1999;29(11):1181-1189.

**2.** Patti GJ, Yanes O, Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol.* Apr 2012;13(4):263-269.

**3.** Holmes E, Kinross J, Gibson GR, et al. Therapeutic modulation of microbiota-host metabolic interactions. *Sci Transl Med.* Jun 6 2012;4(137):137rv136.

**4.** Brindle JT, Antti H, Holmes E, et al. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. *Nat Med.* Dec 2002;8(12):1439-1444.

**5.** Clayton TA, Lindon JC, Cloarec O, et al. Pharmaco-metabonomic phenotyping and personalized drug treatment. *Nature.* Apr 20 2006;440(7087):1073-1077.

**6.** Nicholson JK, Wilson ID. Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat Rev Drug Discov.* Aug 2003;2(8):668-676.

**7.** Nicholson JK, Lindon JC. Systems biology: Metabonomics. *Nature.* Oct 23 2008;455(7216):1054-1056.

**8.** Nicholson JK, Holmes E, Wilson ID. Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol.* May 2005;3(5):431-438.

**9.** Coen M, Holmes E, Lindon JC, Nicholson JK. NMR-based metabolic profiling and metabonomic approaches to problems in molecular toxicology. *Chem Res Toxicol.* Jan 2008;21(1):9-27.

**10.** Nicholson JK, Connelly J, Lindon JC, Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov.* Feb 2002;1(2):153-161.

**11.** Saghatelian A, Trauger SA, Want EJ, Hawkins EG, Siuzdak G, Cravatt BF. Assignment of endogenous substrates to enzymes by global metabolite profiling. *Biochemistry.* Nov 16 2004;43(45):14332-14339.

**12.** Tagore DM, Nolte WM, Neveu JM, et al. Peptidase substrates via global peptide profiling. *Nat Chem Biol.* Jan 2009;5(1):23-25.

**13.** Schmidt C. Metabolomics takes its place as latest up-and-coming "omic" science. *J Natl Cancer Inst.* May 19 2004;96(10):732-734.

**14.** Beckonert O, Keun HC, Ebbels TM, et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc.* 2007;2(11):2692-2703.

**15.** Beckonert O, Coen M, Keun HC, et al. High-resolution magic-angle-spinning NMR spectroscopy for metabolic profiling of intact tissues. *Nat Protoc.* 2010;5(6):1019-1032.

**16.** Want EJ, Wilson ID, Gika H, et al. Global metabolic profiling procedures for urine using UPLC-MS. *Nat Protoc.* 2010;5(6):1005-1018.

**17.** Masson P, Alves AC, Ebbels TM, Nicholson JK, Want EJ. Optimization and evaluation of metabolite extraction protocols for untargeted metabolic profiling of liver samples by UPLC-MS. *Anal Chem.* Sep 15 2010;82(18):7779-7786.

**18.** Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. *Cell.* Sep 5 2008;134(5):714-717.

**19.** Martin FP, Dumas ME, Wang Y, et al. A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model. *Mol Syst Biol.* 2007;3:112.

**20.** De Vos RC, Moco S, Lommen A, Keurentjes JJ, Bino RJ, Hall RD. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat Protoc.* 2007;2(4):778-791.

**21.** Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat Protoc.* 2006;1(1):387-396.
**22.** Hirai MY, Klein M, Fujikawa Y, et al. Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. *J Biol Chem.* Jul 8 2005;280(27):25590-25595.
**23.** Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* Jan 2012;40(Database issue):D109-114.
**24.** Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* Jan 1 2000;28(1):27-30.
**25.** Orešič M. Informatics and computational strategies for the study of lipids. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids.* 2011;1811(11):991-999.
**26.** McDonnell LA, Heeren RM. Imaging mass spectrometry. *Mass Spectrom Rev.* Jul-Aug 2007;26(4):606-643.
**27.** Bunch J, Clench MR, Richards DS. Determination of pharmaceutical compounds in skin by imaging matrix-assisted laser desorption/ionisation mass spectrometry. *Rapid Commun Mass Spectrom.* 2004;18(24):3051-3060.
**28.** Trim PJ, Atkinson SJ, Princivalle AP, Marshall PS, West A, Clench MR. Matrix-assisted laser desorption / ionisation mass spectrometry imaging of lipids in rat brain tissue with integrated unsupervised and supervised multivariant statistical analysis. *Rapid Communications in Mass Spectrometry.* 2008:1503-1509.
**29.** Trim PJ, Henson CM, Avery JL, et al. Matrix-Assisted Laser Desorption / Ionization-Ion Mobility Separation-Mass Spectrometry Imaging of Vinblastine in Whole Body Tissue Sections. *Society.* 2008;80:8628-8634.
**30.** Burrell M, Earnshaw C, Clench M. Imaging Matrix Assisted Laser Desorption Ionization Mass Spectrometry: a technique to map plant metabolites within tissues at high spatial resolution. *J Exp Bot.* 2007;58(4):757-763.
**31.** Takats Z, Wiseman JM, Gologan B, Cooks GR. Mass Spectrometry Sampling Under Ambient Conditions withDesorption Electrospray Ionization. *Science.* 2004;306(5695):471.
**32.** Mirnezami R, Kinross JM, Vorkas PA, et al. Implementation of molecular phenotyping approaches in the personalized surgical patient journey. *Ann Surg.* May 2012;255(5):881-889.
**33.** Nicholson JK, Holmes E. Global systems biology and personalized healthcare solutions. *Discov Med.* Apr 2006;6(32):63-70.
**34.** Swann JR, Want EJ, Geier FM, et al. Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. *Proc Natl Acad Sci U S A.* Mar 15 2011;108 Suppl 1:4523-4530.
**35.** Sreekumar A, Poisson LM, Rajendiran TM, et al. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature.* 2009;457:910-914.
**36.** Patti GJ, Yanes O, Shriver LP, et al. Metabolomics implicates altered sphingolipids in chronic pain of neuropathic origin. *Nat Chem Biol.* Mar 2012;8(3):232-234.
**37.** Edwards LM, Thiele I. Applying systems biology methods to the study of human physiology in extreme environments. *Extrem Physiol Med.* 2013;2(1):8.
**38.** Fahy E, Subramaniam S, Murphy RC, et al. Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res.* Apr 2009;50 Suppl:S9-14.
**39.** Quehenberger O, Dennis EA. The human plasma lipidome. *N Engl J Med.* Nov 10 2011;365(19):1812-1823.
**40.** Quehenberger O, Armando AM, Brown AH, et al. Lipidomics reveals a remarkable diversity of lipids in human plasma. *J Lipid Res.* Nov 2010;51(11):3299-3305.

**41.** Shevchenko A, Simons K. Lipidomics: coming to grips with lipid diversity. *Nat Rev Mol Cell Biol.* Aug 2010;11(8):593-598.
**42.** Wenk MR. The emerging field of lipidomics. *Nat Rev Drug Discov.* Jul 2005;4(7):594-610.
**43.** Wymann MP, Schneiter R. Lipid signalling in disease. *Nat Rev Mol Cell Biol.* Feb 2008;9(2):162-176.
**44.** Dennis EA. Lipidomics joins the omics evolution. *Proc Natl Acad Sci U S A.* Feb 17 2009;106(7):2089-2090.
**45.** WHO WHO. Deaths estimates for 2008 by cause for WHO Member States. *Health statistics and health information systems* http://www.who.int/healthinfo/global_burden_disease/estimates_country/en/index.html. Accessed 02/04/2013, 2013.
**46.** Toth PP. Cardiology patient page. The "good cholesterol": high-density lipoprotein. *Circulation.* Feb 8 2005;111(5):e89-91.
**47.** Berliner JA, Navab M, Fogelman AM, et al. Atherosclerosis: basic mechanisms. Oxidation, inflammation, and genetics. *Circulation.* May 1 1995;91(9):2488-2496.
**48.** Frink RJ. *Inflammatory Atherosclerosis: Characteristics of the Injurious Agent.* Sacramento (CA): Heart Research Foundation; 2002.
**49.** Libby P. Inducible gene expression in vascular wall cells and lesional leukocytes. *J Vasc Surg.* May 1992;15(5):916-917.
**50.** Deshmane SL, Kremlev S, Amini S, Sawaya BE. Monocyte chemoattractant protein-1 (MCP-1): an overview. *J Interferon Cytokine Res.* Jun 2009;29(6):313-326.
**51.** Palinski W, Rosenfeld ME, Yla-Herttuala S, et al. Low density lipoprotein undergoes oxidative modification in vivo. *Proc Natl Acad Sci U S A.* Feb 1989;86(4):1372-1376.
**52.** Libby P, Ridker PM, Maseri A. Inflammation and atherosclerosis. *Circulation.* Mar 5 2002;105(9):1135-1143.
**53.** Malek AM, Alper SL, Izumo S. Hemodynamic shear stress and its role in atherosclerosis. *JAMA.* Dec 1 1999;282(21):2035-2042.
**54.** Libby P. Current concepts of the pathogenesis of the acute coronary syndromes. *Circulation.* Jul 17 2001;104(3):365-372.
**55.** Holmgren A, Rumsby G, Gustafsson S, Naslund U, Henein MY. The nature of cardiac calcification in aortic stenosis. *Int J Cardiol.* Jul 12 2012;158(2):319-321.
**56.** Frink RJ, Achor RW, Brown AL, Jr., Kincaid OW, Brandenburg RO. Significance of calcification of the coronary arteries. *Am J Cardiol.* Sep 1970;26(3):241-247.
**57.** Abedin M, Tintut Y, Demer LL. Vascular calcification: mechanisms and clinical ramifications. *Arterioscler Thromb Vasc Biol.* Jul 2004;24(7):1161-1170.
**58.** Rumberger JA, Sheedy PF, 2nd, Breen JF, Fitzpatrick LA, Schwartz RS. Electron beam computed tomography and coronary artery disease: scanning for coronary artery calcification. *Mayo Clin Proc.* Apr 1996;71(4):369-377.
**59.** Nicoll R, Henein MY. Arterial calcification: Friend or foe? *Int J Cardiol.* Jul 16 2012.
**60.** Sage AP, Tintut Y, Demer LL. Regulatory mechanisms in vascular calcification. *Nat Rev Cardiol.* Sep 2010;7(9):528-536.
**61.** Bridgewater B, Kinsman R, Walton P, Gummert J, Kappetein AP. The 4th European Association for Cardio-Thoracic Surgery adult cardiac surgery database report. *Interact Cardiovasc Thorac Surg.* Jan 2011;12(1):4-5.
**62.** Rajamannan NM. Mechanisms of aortic valve calcification: the LDL-density-radius theory: a translation from cell signaling to physiology. *Am J Physiol Heart Circ Physiol.* Jan 2010;298(1):H5-15.
**63.** David Merryman W. Mechano-potential etiologies of aortic valve disease. *J Biomech.* Jan 5 2010;43(1):87-92.

**64.** Callister TQ, Raggi P, Cooil B, Lippolis NJ, Russo DJ. Effect of HMG-CoA reductase inhibitors on coronary artery disease as assessed by electron-beam computed tomography. *N Engl J Med.* Dec 31 1998;339(27):1972-1978.
**65.** Sabatine MS, Liu E, Morrow DA, et al. Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation.* Dec 20 2005;112(25):3868-3875.
**66.** Mayr M, Chung YL, Mayr U, et al. Proteomic and metabolomic analyses of atherosclerotic vessels from apolipoprotein E-deficient mice reveal alterations in inflammation, oxidative stress, and energy metabolism. *Arterioscler Thromb Vasc Biol.* Oct 2005;25(10):2135-2142.
**67.** Leo GC, Darrow AL. NMR-based metabolomics of urine for the atherosclerotic mouse model using apolipoprotein-E deficient mice. *Magn Reson Chem.* Dec 2009;47 Suppl 1:S20-25.
**68.** Cheng KK, Benson GM, Grimsditch DC, Reid DG, Connor SC, Griffin JL. A metabolomic study of the LDL receptor null mouse fed a high-fat diet reveals profound perturbations in choline metabolism that are shared with ApoE null mice. *Physiol Genomics.* Mar 2 2010.
**69.** Lu Y, Boekschoten MV, Wopereis S, Muller M, Kersten S. Comparative transcriptomic and metabolomic analysis of fenofibrate and fish oil treatments in mice. *Physiol Genomics.* Dec 5 2011;43(23):1307-1318.
**70.** Rossmeisl M, Jilkova ZM, Kuda O, et al. Metabolic effects of n-3 PUFA as phospholipids are superior to triglycerides in mice fed a high-fat diet: possible role of endocannabinoids. *PLoS One.* 2012;7(6):e38834.
**71.** Rull A, Aragones G, Beltran-Debon R, Rodriguez-Gallego E, Camps J, Joven J. Exploring PPAR modulation in experimental mice. *Methods Mol Biol.* 2013;952:253-273.
**72.** Jove M, Ayala V, Ramirez-Nunez O, et al. Lipidomic and metabolomic analyses reveal potential plasma biomarkers of early atheromatous plaque formation in hamsters. *Cardiovasc Res.* Mar 15 2013;97(4):642-652.
**73.** Trupp M, Zhu H, Wikoff WR, et al. Metabolomics reveals amino acids contribute to variation in response to simvastatin treatment. *PLoS One.* 2012;7(7):e38386.
**74.** Yerges-Armstrong LM, Ellero-Simatos S, Georgiades A, et al. Purine pathway implicated in mechanism of resistance to aspirin therapy: pharmacometabolomics-informed pharmacogenomics. *Clin Pharmacol Ther.* Oct 2013;94(4):525-532.
**75.** Llorach R, Urpi-Sarda M, Tulipani S, Garcia-Aloy M, Monagas M, Andres-Lacueva C. Metabolomic fingerprint in patients at high risk of cardiovascular disease by cocoa intervention. *Mol Nutr Food Res.* Jun 2013;57(6):962-973.
**76.** Gurdeniz G, Rago D, Bendsen NT, Savorani F, Astrup A, Dragsted LO. Effect of trans fatty acid intake on LC-MS and NMR plasma profiles. *PLoS One.* 2013;8(7):e69589.
**77.** McCombie G, Browning LM, Titman CM, et al. omega-3 oil intake during weight loss in obese women results in remodelling of plasma triglyceride and fatty acids. *Metabolomics.* Sep 2009;5(3):363-374.
**78.** Yap IK, Brown IJ, Chan Q, et al. Metabolome-wide association study identifies multiple biomarkers that discriminate north and south Chinese populations at differing risks of cardiovascular disease: INTERMAP study. *J Proteome Res.* Dec 3 2010;9(12):6647-6654.
**79.** Yu B, Zheng Y, Alexander D, et al. Genome-wide association study of a heart failure related metabolomic profile among African Americans in the Atherosclerosis Risk in Communities (ARIC) study. *Genet Epidemiol.* Dec 2013;37(8):840-845.
**80.** Zheng Y, Yu B, Alexander D, et al. Metabolomics and incident hypertension among blacks: the atherosclerosis risk in communities study. *Hypertension.* Aug 2013;62(2):398-403.
**81.** Walsh MC, McLoughlin GA, Roche HM, et al. Impact of geographical region on urinary metabolomic and plasma fatty acid profiles in subjects with the metabolic syndrome across Europe: the LIPGENE study. *Br J Nutr.* Feb 2014;111(3):424-431.

**82.** Kim SH, Yang SO, Kim HS, Kim Y, Park T, Choi HK. 1H-nuclear magnetic resonance spectroscopy-based metabolic assessment in a rat model of obesity induced by a high-fat diet. *Anal Bioanal Chem.* Oct 2009;395(4):1117-1124.

**83.** Mercer JR, Cheng KK, Figg N, et al. DNA damage links mitochondrial dysfunction to atherosclerosis and the metabolic syndrome. *Circ Res.* Oct 15 2010;107(8):1021-1031.

**84.** Ciborowski M, Martin-Ventura JL, Meilhac O, et al. Metabolites secreted by human atherothrombotic aneurysms revealed through a metabolomic approach. *J Proteome Res.* Mar 4 2011;10(3):1374-1382.

**85.** Plumb RS, Johnson KA, Rainville P, et al. UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun Mass Spectrom.* 2006;20(13):1989-1994.

**86.** Plumb RS, Granger JH, Stumpf CL, et al. A rapid screening approach to metabonomics using UPLC and oa-TOF mass spectrometry: application to age, gender and diurnal variation in normal/Zucker obese rats and black, white and nude mice. *Analyst.* Jun 2005;130(6):844-849.

**87.** Isella F, Canellas E, Bosetti O, Nerin C. Migration of non intentionally added substances from adhesives by UPLC-Q-TOF/MS and the role of EVOH to avoid migration in multilayer packaging materials. *J Mass Spectrom.* Apr 2013;48(4):430-437.

**88.** Fu J, Chu J, Sun X, Wang J, Yan C. Simple, rapid, and simultaneous assay of multiple carboxyl containing phytohormones in wounded tomatoes by UPLC-MS/MS using single SPE purification and isotope dilution. *Anal Sci.* 2012;28(11):1081-1087.

**89.** Plumb RS, Ayrton J, Dear GJ, Sweatman BC, Ismail IM. The use of preparative high performance liquid chromatography with tandem mass spectrometric directed fraction collection for the isolation and characterisation of drug metabolites in urine by nuclear magnetic resonance spectroscopy and liquid chromatography/sequential mass spectrometry. *Rapid Commun Mass Spectrom.* 1999;13(10):845-854.

**90.** Gika HG, Theodoridou A, Michopoulos F, et al. Determination of two COX-2 inhibitors in serum and synovial fluid of patients with inflammatory arthritis by ultra performance liquid chromatography-inductively coupled plasma mass spectroscopy and quadrupole time-of-flight mass spectrometry. *J Pharm Biomed Anal.* Apr 5 2009;49(3):579-586.

**91.** Yu K, Di L, Kerns E, Li SQ, Alden P, Plumb RS. Ultra-performance liquid chromatography/tandem mass spectrometric quantification of structurally diverse drug mixtures using an ESI-APCI multimode ionization source. *Rapid Commun Mass Spectrom.* 2007;21(6):893-902.

**92.** Spagou K, Wilson ID, Masson P, et al. HILIC-UPLC-MS for exploratory urinary metabolic profiling in toxicological studies. *Anal Chem.* Jan 1 2011;83(1):382-390.

**93.** Snyder LR, Kirkland JJ. *Introduction to modern liquid chromatography*. 2d ed. New York: Wiley; 1979.

**94.** Swartz ME. Ultra performance liquid chromatography (UPLC): An introduction. *Lc Gc N Am.* May 2005:8-14.

**95.** Swartz ME, Murphy B. New frontiers in chromatography. *Am Lab.* Feb 2005;37(3):22-+.

**96.** Wilson ID, Nicholson JK, Castro-Perez J, et al. High resolution "Ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *Journal of Proteome Research.* Mar-Apr 2005;4(2):591-598.

**97.** Waters. HSS (High Strength Silica)Technology. *http://www.waters.com/waters/nav.htm?cid=134618105.*

**98.** Waters. CSH (Charged Surface Hybrid) Technology. *http://www.waters.com/waters/nav.htm?cid=134618101.*

**99.** Waters. Particle Technology. *http://www.waters.com/waters/nav.htm?cid=10009386.*

**100.** Hemstrom P, Irgum K. Hydrophilic interaction chromatography. *J Sep Sci.* Aug 2006;29(12):1784-1821.
**101.** Ivanisevic J, Zhu ZJ, Plate L, et al. Toward 'omic scale metabolite profiling: a dual separation-mass spectrometry approach for coverage of lipid and central carbon metabolism. *Anal Chem.* Jul 16 2013;85(14):6876-6884.
**102.** Spagou K, Tsoukali H, Raikos N, Gika H, Wilson ID, Theodoridis G. Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies. *J Sep Sci.* Mar 2010;33(6-7):716-727.
**103.** Neue UD. Hydrophilic interaction chromatography (HILIC). *http://www.sepscience.com/Techniques/LC/Articles/501-/Hydrophilic-interaction-chromatography-HILIC.*
**104.** Geier FM, Want EJ, Leroi AM, Bundy JG. Cross-platform comparison of Caenorhabditis elegans tissue extraction strategies for comprehensive metabolome coverage. *Anal Chem.* May 15 2011;83(10):3730-3736.
**105.** de Hoffmann E, Stroobant V. *Mass Spectrometry Principles and Applications*. Third ed: WILEY; 2007.
**106.** Siuzdak G. *The Expanding Role of Mass Spectrometry in Biotechnology - 2nd Edition*: MCC Press: San Diego; 2006.
**107.** Waters. *Waters Xevo G2 QTof Operator's Overview and Maintenance Guide*2011.
**108.** Fenn JB. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* 1989;246(4926):64.
**109.** Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* Oct 6 1989;246(4926):64-71.
**110.** Waters. MS - Mass Spectrometry. http://www.waters.com/waters/nav.htm?cid=10073244.
**111.** Smith CA, O'Maille G, Want EJ, et al. METLIN: a metabolite mass spectral database. *Ther Drug Monit.* Dec 2005;27(6):747-751.
**112.** Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem.* Feb 1 2006;78(3):779-787.
**113.** Veselkov KA, Vingara LK, Masson P, et al. Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. *Anal Chem.* Aug 1 2011;83(15):5864-5872.
**114.** Gika HG, Theodoridis GA, Wingate JE, Wilson ID. Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: application to human urine. *J Proteome Res.* Aug 2007;6(8):3291-3303.
**115.** Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics.* Sep 2007;3(3):211-221.
**116.** Wishart DS, Knox C, Guo AC, et al. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* Jan 2009;37(Database issue):D603-610.
**117.** Cotter D, Maer A, Guda C, Saunders B, Subramaniam S. LMPD: LIPID MAPS proteome database. *Nucleic Acids Res.* Jan 1 2006;34(Database issue):D507-510.
**118.** Bictash M, Ebbels TM, Chan Q, et al. Opening up the "Black Box": metabolic phenotyping and metabolome-wide association studies in epidemiology. *J Clin Epidemiol.* Sep 2010;63(9):970-979.
**119.** Masson P, Spagou K, Nicholson JK, Want EJ. Technical and biological variation in UPLC-MS-based untargeted metabolic profiling of liver extracts: application in an experimental toxicity study on galactosamine. *Anal Chem.* Feb 1 2011;83(3):1116-1123.

**120.** Want EJ, Coen M, Masson P, et al. Ultra performance liquid chromatography-mass spectrometry profiling of bile acid metabolites in biofluids: application to experimental toxicology studies. *Anal Chem.* Jun 15 2010;82(12):5282-5289.
**121.** Rainville PD, Stumpf CL, Shockcor JP, Plumb RS, Nicholson JK. Novel application of reversed-phase UPLC-oaTOF-MS for lipid analysis in complex biological mixtures: a new tool for lipidomics. *J Proteome Res.* Feb 2007;6(2):552-558.
**122.** Want EJ, Masson P, Michopoulos F, et al. Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat Protoc.* Jan 2013;8(1):17-32.
**123.** de Graaf RA. *In Vivo NMR Spectroscopy: Principles and Techniques*. 2nd ed: Wiley; 2007.
**124.** Michopoulos F, Lai L, Gika H, Theodoridis G, Wilson I. UPLC-MS-based analysis of human plasma for metabonomics using solvent precipitation or solid phase extraction. *J Proteome Res.* Apr 2009;8(4):2114-2121.
**125.** Canelas AB, ten Pierick A, Ras C, et al. Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics. *Anal Chem.* Sep 1 2009;81(17):7379-7389.
**126.** Wu H, Southam AD, Hines A, Viant MR. High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Anal Biochem.* Jan 15 2008;372(2):204-212.
**127.** Yanes O, Tautenhahn R, Patti GJ, Siuzdak G. Expanding coverage of the metabolome for global metabolite profiling. *Anal Chem.* Mar 15 2011;83(6):2152-2161.
**128.** Folch J, Lees M, Sloane Stanley GH. A simple method for the isolation and purification of total lipides from animal tissues. *J Biol Chem.* May 1957;226(1):497-509.
**129.** Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. *Can J Biochem Physiol.* Aug 1959;37(8):911-917.
**130.** Eriksson L, Johansson E, Kettaneh-Wold N, Wold S. *Multi- and Megavariate Data Analysis*: UMETRICS AB; 2001.
**131.** Trygg J, Holmes E, Lundstedt T. Chemometrics in metabonomics. *J Proteome Res.* Feb 2007;6(2):469-479.
**132.** Geladi P, Kowalski BR. PARTIAL LEAST-SQUARES REGRESSION: A TUTORIAL. *Analytica Chimica Acta.* 1986;185:1-17.
**133.** Eriksson L, Trygg J, Wold S. CV-ANOVA for significance testing of PLS and OPLS® models. *Journal of Chemometrics.* 2008;22(11-12):594-600.
**134.** Wiklund S, Johansson E, Sjostrom L, et al. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal Chem.* Jan 1 2008;80(1):115-122.
**135.** Adamson J, Beswick A, Ebrahim S. Is stroke the most common cause of disability? *J Stroke Cerebrovasc Dis.* Jul-Aug 2004;13(4):171-177.
**136.** Halliday A, Harrison M, Hayter E, et al. 10-year stroke prevention after successful carotid endarterectomy for asymptomatic stenosis (ACST-1): a multicentre randomised trial. *Lancet.* Sep 25 2010;376(9746):1074-1084.
**137.** Chatzikonstantinou A, Wolf ME, Schaefer A, Hennerici MG. Asymptomatic and symptomatic carotid stenosis: an obsolete classification? *Stroke Res Treat.* 2012;2012:340798.
**138.** Shockcor J, Crowe H, Yu K, Shion H. *Analysis of Intact Lipids from Biological Matrices by UPLC/High Definition MS*2010.
**139.** The LIPID MAPS Lipidomics Gateway, http://www.lipidmaps.org/.
**140.** Stegemann C, Drozdov I, Shalhoub J, et al. Comparative lipidomics profiling of human atherosclerotic plaques. *Circ Cardiovasc Genet.* Jun 2011;4(3):232-242.
**141.** Tian J, Gu X, Sun Y, et al. Effect of statin therapy on the progression of coronary atherosclerosis. *BMC Cardiovasc Disord.* 2012;12:70.
**142.** Eriksson L, Trygg J, Wold S. CV-ANOVA for significance testing of PLS and OPLS (R) models. *J Chemometr.* Nov-Dec 2008;22(11-12):594-600.

**143.** Holmes E, Loo RL, Stamler J, et al. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature.* May 15 2008;453(7193):396-400.
**144.** Legido-Quigley C, McDermott L, Vilca-Melendez H, et al. Bile UPLC-MS fingerprinting and bile acid fluxes during human liver transplantation. *Electrophoresis.* Aug 2011;32(15):2063-2070.
**145.** Bianda N, Di Valentino M, Periat D, et al. Progression of human carotid and femoral atherosclerosis: a prospective follow-up study by magnetic resonance vessel wall imaging. *Eur Heart J.* Jan 2012;33(2):230-237.
**146.** Herisson F, Heymann MF, Chetiveaux M, et al. Carotid and femoral atherosclerotic plaques show different morphology. *Atherosclerosis.* Jun 2011;216(2):348-354.
**147.** Demer LL, Tintut Y. Vascular calcification: pathobiology of a multifaceted disease. *Circulation.* Jun 3 2008;117(22):2938-2948.
**148.** Griffin JL, Atherton H, Shockcor J, Atzori L. Metabolomics as a tool for cardiac research. *Nat Rev Cardiol.* Nov 2011;8(11):630-643.
**149.** Catala A. A synopsis of the process of lipid peroxidation since the discovery of the essential fatty acids. *Biochem Biophys Res Commun.* Aug 27 2010;399(3):318-323.
**150.** Mackness MI, Durrington PN. HDL, its enzymes and its potential to influence lipid peroxidation. *Atherosclerosis.* Jun 1995;115(2):243-253.
**151.** Leonarduzzi G, Gamba P, Gargiulo S, Biasi F, Poli G. Inflammation-related gene expression by lipid oxidation-derived products in the progression of atherosclerosis. *Free Radic Biol Med.* Jan 1 2012;52(1):19-34.
**152.** Hutchins PM, Moore EE, Murphy RC. Electrospray MS/MS reveals extensive and nonspecific oxidation of cholesterol esters in human peripheral vascular lesions. *J Lipid Res.* Nov 2011;52(11):2070-2083.
**153.** Fillios LC, Naito C, Andrus SB, Roach AM. The hypercholesteremic and atherogenic properties of various purines and pyrimidines. *Am J Clin Nutr.* Jan-Feb 1959;7(1):70-75.
**154.** Burnstock G. Control of vascular tone by purines and pyrimidines. *Br J Pharmacol.* Oct 2010;161(3):527-529.
**155.** Ralevic V, Burnstock G. Effects of purines and pyrimidines on the rat mesenteric arterial bed. *Circ Res.* Dec 1991;69(6):1583-1590.
**156.** Cronstein BN. Adenosine, an endogenous anti-inflammatory agent. *J Appl Physiol.* Jan 1994;76(1):5-13.
**157.** da Rocha Lapa F, da Silva MD, de Almeida Cabrini D, Santos AR. Anti-inflammatory effects of purine nucleosides, adenosine and inosine, in a mouse model of pleurisy: evidence for the role of adenosine A2 receptors. *Purinergic Signal.* Dec 2012;8(4):693-704.
**158.** Quemeneur L, Gerland LM, Flacher M, Ffrench M, Revillard JP, Genestier L. Differential control of cell cycle, proliferation, and survival of primary T lymphocytes by purine and pyrimidine nucleotides. *J Immunol.* May 15 2003;170(10):4986-4995.
**159.** Bismuth J, Lin P, Yao Q, Chen C. Ceramide: a common pathway for atherosclerosis? *Atherosclerosis.* Feb 2008;196(2):497-504.
**160.** Taha TA, Mullen TD, Obeid LM. A house divided: ceramide, sphingosine, and sphingosine-1-phosphate in programmed cell death. *Biochim Biophys Acta.* Dec 2006;1758(12):2027-2036.
**161.** Portman OW, Alexander M. Metabolism of sphingolipids by normal and atherosclerotic aorta of squirrel monkeys. *J Lipid Res.* Jan 1970;11(1):23-30.
**162.** Chatterjee S. Sphingolipids in atherosclerosis and vascular biology. *Arterioscler Thromb Vasc Biol.* Oct 1998;18(10):1523-1533.
**163.** Prokazova NV, Bergelson LD. Gangliosides and atherosclerosis. *Lipids.* Jan 1994;29(1):1-5.
**164.** Merrill AH, Jr. Sphingolipid and glycosphingolipid metabolic pathways in the era of sphingolipidomics. *Chem Rev.* Oct 12 2011;111(10):6387-6422.

**165.** Vacaru AM, Tafesse FG, Ternes P, et al. Sphingomyelin synthase-related protein SMSr controls ceramide homeostasis in the ER. *J Cell Biol.* Jun 15 2009;185(6):1013-1027.
**166.** Guan XL, Cestra G, Shui G, et al. Biochemical membrane lipidomics during Drosophila development. *Dev Cell.* Jan 14 2013;24(1):98-111.
**167.** Terova B, Petersen G, Hansen HS, Slotte JP. N-acyl phosphatidylethanolamines affect the lateral distribution of cholesterol in membranes. *Biochim Biophys Acta.* Aug 30 2005;1715(1):49-56.
**168.** Ramstedt B, Slotte JP. Sphingolipids and the formation of sterol-enriched ordered membrane domains. *Biochim Biophys Acta.* Dec 2006;1758(12):1945-1956.
**169.** Ichi I, Nakahara K, Miyashita Y, et al. Association of ceramides in human plasma with risk factors of atherosclerosis. *Lipids.* Sep 2006;41(9):859-863.
**170.** Grandl M, Bared SM, Liebisch G, Werner T, Barlage S, Schmitz G. E-LDL and Ox-LDL differentially regulate ceramide and cholesterol raft microdomains in human Macrophages. *Cytometry A.* Mar 2006;69(3):189-191.
**171.** Ridgway ND, Byers DM, Cook HW, Storey MK. Integration of phospholipid and sterol metabolism in mammalian cells. *Prog Lipid Res.* Jul 1999;38(4):337-360.
**172.** Costantini C, Kolasani RM, Puglielli L. Ceramide and cholesterol: possible connections between normal aging of the brain and Alzheimer's disease. Just hypotheses or molecular pathways to be identified? *Alzheimers Dement.* Jul 2005;1(1):43-50.
**173.** Fernandez-Checa JC. Editorial: lipids in cancer cell biology and therapy. *Anticancer Agents Med Chem.* May 2012;12(4):283-284.
**174.** Yanes O, Clark J, Wong DM, et al. Metabolic oxidation regulates embryonic stem cell differentiation. *Nat Chem Biol.* Jun 2010;6(6):411-417.
**175.** Bunn RC, Cockrell GE, Ou Y, Thrailkill KM, Lumpkin CK, Jr., Fowlkes JL. Palmitate and insulin synergistically induce IL-6 expression in human monocytes. *Cardiovasc Diabetol.* 2010;9:73.
**176.** Unger RH, Zhou YT. Lipotoxicity of beta-cells in obesity and in other causes of fatty acid spillover. *Diabetes.* Feb 2001;50 Suppl 1:S118-121.
**177.** Luiken JJ, Coort SL, Willems J, et al. Contraction-induced fatty acid translocase/CD36 translocation in rat cardiac myocytes is mediated through AMP-activated protein kinase signaling. *Diabetes.* Jul 2003;52(7):1627-1634.
**178.** Schebb NH, Fischer D, Hein EM, et al. Fast sample preparation and liquid chromatography-tandem mass spectrometry method for assaying cell lysate acetylcholine. *J Chromatogr A.* Mar 7 2008;1183(1-2):100-107.
**179.** Maiorana A, Roach RB, Jr. Heterozygous pseudocholinesterase deficiency: a case report and review of the literature. *J Oral Maxillofac Surg.* Jul 2003;61(7):845-847.
**180.** Gittenberger-de Groot AC, DeRuiter MC, Bergwerff M, Poelmann RE. Smooth muscle cell origin and its relation to heterogeneity in development and disease. *Arterioscler Thromb Vasc Biol.* Jul 1999;19(7):1589-1594.
**181.** Johnson RC, Leopold JA, Loscalzo J. Vascular calcification: pathobiological mechanisms and clinical implications. *Circ Res.* Nov 10 2006;99(10):1044-1059.
**182.** Wallin R, Wajih N, Greenwood GT, Sane DC. Arterial calcification: a review of mechanisms, animal models, and the prospects for therapy. *Med Res Rev.* Jul 2001;21(4):274-301.
**183.** 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. *Anal Chem.* Jul 1 2006;78(13):4281-4290.
**184.** Nicholson JK, Foxall PJ, Spraul M, Farrant RD, Lindon JC. 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. *Anal Chem.* Mar 1 1995;67(5):793-811.

**185.** Masuda M, Ting TC, Levi M, Saunders SJ, Miyazaki-Anzai S, Miyazaki M. Activating transcription factor 4 regulates stearate-induced vascular calcification. *J Lipid Res.* Aug 2012;53(8):1543-1552.
**186.** Abedin M, Lim J, Tang TB, Park D, Demer LL, Tintut Y. N-3 fatty acids inhibit vascular calcification via the p38-mitogen-activated protein kinase and peroxisome proliferator-activated receptor-gamma pathways. *Circ Res.* Mar 31 2006;98(6):727-729.
**187.** Burke JE, Dennis EA. Phospholipase A2 biochemistry. *Cardiovasc Drugs Ther.* Feb 2009;23(1):49-59.
**188.** Chaudhuri P, Colles SM, Damron DS, Graham LM. Lysophosphatidylcholine inhibits endothelial cell migration by increasing intracellular calcium and activating calpain. *Arterioscler Thromb Vasc Biol.* Feb 1 2003;23(2):218-223.
**189.** Chan SY, Loscalzo J. The emerging paradigm of network medicine in the study of human disease. *Circ Res.* Jul 20 2012;111(3):359-374.
**190.** Dromparis P, Michelakis ED. Mitochondria in vascular health and disease. *Annu Rev Physiol.* 2013;75:95-126.
**191.** Bell FP. Carnitine ester hydrolysis in arteries from normal and cholesterol-fed rabbits and the effects of carnitine esters on arterial microsomal ACAT. *Comp Biochem Physiol B.* 1984;79(2):125-128.
**192.** Dadu RT, Nambi V, Ballantyne CM. Developing and assessing cardiovascular biomarkers. *Transl Res.* Apr 2012;159(4):265-276.

# *Appendices*

# *Appendix 1: Supplementary methodologies*

## *Chapter 3: Command lines used for data processing of LC-MS data, using the XCMS package[112].*

```
library(xcms)

library(snow)

source("http://bioconductor.org/biocLite.R")

biocLite("multtest")

% Data must be in NetCDF format
```

***Lipid profiling positive mode***

```
dir ="directory where data are "

setwd(dir)

% peak width at 5% approx. 3-12s be used

pw1=3

pw2=12


ppm=30

sn=20


pp <- xcmsSet(method="centWave", peakwidth=c(pw1,pw2), lock=TRUE, ppm=ppm, snthresh=sn,
nSlaves=4)
```

```
gpp <- group(pp, method="density", bw=15, mzwid=0.07, minfrac=0.5, minsamp=1, sleep=.001)

gpp

% using missing=13 to account for blanks etc that don't have peaks in all samples

cgpp <- retcor(gpp,  family="s", plottype="m", missing=13)

gcgpp <- group(pp, method="density", bw=15, mzwid=0.07, minfrac=0.5, minsamp=1, sleep=.001)

fgcgpp <- fillPeaks(gcgpp)

report <- diffreport(fgcgpp,class1="sympt",class2="asympt",filebase="20130226_results_liprof_pos", 1000, metlin = 0.02)
```

***Lipid profiling positive mode***

```
dir ="directory where data are "

setwd(dir)

pw1=3

pw2=12

ppm=30

sn=20
```

```
pp <- xcmsSet(method="centWave", peakwidth=c(pw1,pw2), lock=TRUE, ppm=ppm, snthresh=sn, nSlaves=4)

gpp <- group(pp, method="density", bw=15, mzwid=0.07, minfrac=0.5, minsamp=1, sleep=.001)

% using missing=13 to account for blanks etc that don't have peaks in all samples

cgpp <- retcor(gpp, method="linear", family="s", plottype="m", missing=13)

gcgpp <- group(pp, method="density", bw=15, mzwid=0.07, minfrac=0.5, minsamp=1, sleep=.001)

fgcgpp <- fillPeaks(gcgpp)

report <- diffreport(fgcgpp,class1="sympt",class2="asympt",filebase="20130227_results_liprof_neg", 1000, metlin = 0.02)

peak width at 5% approx. 0.10-0.20min

6-20s
```

***HILIC UPLC-MS of aqueous extracts – Positive mode***

```
dir ="directory where data are "

setwd(dir)

% peak width at 5% approx. 0.10-0.20min (6-20ss will be used)
```

```
pw1=6

pw2=20


ppm=30

sn=20


pp <- xcmsSet(method="centWave", peakwidth=c(pw1,pw2), lock=TRUE, ppm=ppm, snthresh=sn,
nSlaves=4)


gpp <- group(pp, method="density", bw=30, mzwid=0.07, minfrac=0.5, minsamp=1, sleep=.001)


cgpp <- retcor(gpp,  method="linear", family="s", plottype="m")



gcgpp <- group(pp, method="density", bw=30, mzwid=0.07, minfrac=0.5, minsamp=1, sleep=.001)


fgcgpp <- fillPeaks(gcgpp)


report <- diffreport(fgcgpp,class1="sympt",class2="asympt",filebase="results", 1000, metlin =
0.02)
```

***HILIC UPLC-MS of aqueous extracts – Negative mode***

```
dir ="directory where data are "

setwd(dir)
```

```
pw1=6

pw2=20


ppm=30

sn=20


pp <- xcmsSet(method="centWave", peakwidth=c(pw1,pw2), lock=TRUE, ppm=ppm, snthresh=sn,
nSlaves=4)


gpp <- group(pp, method="density", bw=30, mzwid=0.07, minfrac=0.5, minsamp=1, sleep=.001)


cgpp <- retcor(gpp,  method="linear", family="s", plottype="m")


gcgpp <- group(pp, method="density", bw=30, mzwid=0.07, minfrac=0.5, minsamp=1, sleep=.001)



fgcgpp <- fillPeaks(gcgpp)


report <- diffreport(fgcgpp,class1="sympt",class2="asympt",filebase="results", 1000, metlin =
0.02)
```

# *Appendix 2: Awarded travel grants to present parts of this thesis at conferences*

Royal Society of Chemistry Travel award (2013): For poster presentation at the ASMS (American Society of Mass Spectrometry) annual conference.

MSACL Young Investigator Travel Award (2013): For poster presentation at the MSACL (Mass Spectrometry Applications to the Clinical Lab) Conference.

Travel Grant from the ESC Council (2012): For poster presentation at the ESC (European Society of Cardiology) Congress.

Royal Society of Chemistry and British Mass Spectrometry Society Travel awards (2012): For poster presentation at the ASMS (American Society of Mass Spectrometry) annual conference.

Chromatographic Society "John Dolphin Fellowship" (2011): For poster presentation at the ASMS annual conference