NMR and UPLC-HRMS based Metabolomics : application in Non-alcoholic fatty liver disease and Prostate Cancer Biomarker Discovery

Xiangping Lin

THESE DE DOCTORAT

Université Sorbonne Paris Nord

École doctorale 146 « Sciences, Technologies, Santé - Galilée »

Spécialité : Chimie

Par

Xiangping LIN

NMR and UPLC-HRMS based Metabolomics: Application in Non-alcoholic fatty liver disease and Prostate Cancer Biomarker Discovery

Thèse présentée et soutenue publiquement, le 12 mars 2020

Composition du Jury :

Dr Christophe Junot, Université Paris-Saclay, Rapporteur

Pr François Mesnard, Université de Picardie, Rapporteur

Dr Carina Prip-Buus, Université de Paris, Examinatrice

Dr Mohamed Nawfal Triba, Université Sorbonne Paris Nord, Examinateur

Pr Philippe Savarin, Université Sorbonne Paris Nord, Directeur de thèse

ACKNOWLEDGMENTS

First, I would like to thank my PhD. supervisor, Pr. Philippe SAVARIN, the director of CSPBAT (Chimie, Structures, Propriétés de Biomatériaux et d'Agents Thérapeutiques) laboratory, for offering me the possibility of the present PhD. thesis opportunity, and for his guidance and advice both scientifically and personally throughout my thesis work.

Next, I would like to thank Pr. Guowang XU, the director of Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, who offer me a chance to perform part of my thesis work in his laboratory, where I can access the cutting edge MS laboratory equipment and very enthusiastic research environment, thanks also for his invaluable scientific advice.

Then, I would like also to express my deepest appreciation to Dr. Claudine MANACH, who offer me the opportunity to begin in metabolomics. I was really lucky to work with great people in our small metabolomics group: Dr. Mohamed N. TRIBA, Pr. Edith HANTZ, Dr. Nadia BOUCHEMAL, Dr. Zhicheng LIU, Dr. Tony PALAMA, Dr. Roland AMATHIEU and Msc. Agnes VICTOR BALA, and all staffs in Pr. XU's laboratory, for all of their help, accompaniment and advice during my thesis work. They suggestions have always enlightened me and helped me to think in different perspectives. I would like to thank separately Dr. Xinyu LIU, for her constructive suggestions to my thesis work and for all her help during my stay in Dalian.

Thanks also to all my thesis committee members, Dr Christophe Junot, Pr François Mesnard, Dr Carina Prip-Buus, for their thoughtful input and they time spend in the reviewing of my Thesis. Here, I also would like to thank the Doctoral school 146 « Sciences, Technologies, Santé - Galilée » and the French government for offering me opportunity to perform a PhD thesis in France, and for they financial support.

Lastly, I would like to thank my parents, my family especially my elder sister for all they support and patience over the years. They bring sunshine to my life. Without them, I may never have gotten to where I am today.

Metabolomics

A Textbook

To my parents

Changmei LIU, Jialiang LIN and my elder sister, Xiaojun LIN

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Full Name
ALT	Alanine aminotransferase
ANOVA	Analysis of Variance
APCI	Atmospheric Pressure Chemical Ionization
AST	Aspartate aminotransferase
AUROC	Area under the receiver operating characteristic curve
BMI	Body mass index
CE-MS	Capillary Electrophoresis-Mass Spectrometry
CI	Chemical Ionization
CID	Collision Induced Dissociation
CT	Computed Tomography
Da	Dalton
DC	Direct Current
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
EI	Electron Ionization
EIC	Extracted-ion chromatogram
ESI	Electrospray Ionization
FAB	Fast Atom Bombardment
FDR	False discovery rate
FID	Free Induction Decay
FT-ICR	Fourier Transform-Ion Cyclotron Resonance
FWHM	Full Width at High Maximum
GC-MS	Gas Chromatography-Mass Spectrometry
HCA	Hierarchical Clustering Analysis
HCC	Hepatocellular Carcinoma
HMDB	Human Metabolome DataBase
HPLC	High Pressure Liquid Chromatography
HR-MAS	High Resolution Magic Angle Spinning Probes
HRMS	High-resolution mass spectrometry
IS	Internal Standard
IT	Ion Trap
KEGG	Kyoto Encyclopedia of Genes and Genomes
KMTB	α-Keto-γ-methylthiobutyric acid
LB	Line broadening
LC	Liquid chromatography
LC-MS	Liquid Chromatography-Mass Spectrometry
LPC	Lysophosphatidylcholines
LPE	Lysophosphatidylethanolamine
MALDI	Matrix Assisted Laser Desorption Ionisation
MRE	Magnetic resonance elastography
MRI	Magnetic Resonance Imaging
MRI-PDFF	Magnetic resonance imaging proton density fat fraction
MS	Mass Spectrometry

NAFL	Nonalcoholic fatty liver or simple hepatic steatosis
NAFLD	Nonalcoholic fatty liver disease
NaN	Not a Number
NAS	NAFLD activity score
NASH	Nonalcoholic steatohepatitis
NIST	National Institute of standards and technology
NL	Normal liver
NMR	Nuclear Magnetic Resonance
NPV	Negative predictive value
OPLS-DA	Orthogonal Partial Least Squares Discriminant Analysis
OR	Odds Ratio
PC	Phosphatidylcholine
PCA	Principal Component Analysis
PCa	Prostate cancer
PE	Phosphatidylethanolamine
PLS-DA	Partial Least Squares Discriminant Analysis
PPV	Positive predictive value
Q	Quadrupole
Q-TOF	Quadrupole - Time Of Flight
QC	Quality Control
Rf	Radiofrequency
RF	Radio Frequency
ROC	Receiver operating characteristic
RSD	Relative Standard Divergence
S/N	Signal/Noise
SIM	Single Ion Monitoring
SM	Sphingomyelin
SOP	Standard Operating Procedure
TG	Triglycerides
TIC	Total ion current chromatogram
TMAO	Trimethylamine N-oxide
TMS	Tetramethylsilane
TMS	Tetramethylsilane
TOF	Time of Flight
TSP	Trimethylsilylpropanoic acid
UPLC	Ultra-performance liquid chromatography
VCTE	Vibration-controlled transient elastography
VIP	Variable Importance in the Projection

LIST OF FIGURES

Figure 36. Scheme for the identification and validation of putative Prostate cancer biomarkers.

LIST OF TABLES

Part One GENERAL INTRODUCTION

1.1 Metabolites, metabolome and metabolomics

Metabolites are low molecular weight biochemicals (chemically defined as small molecules, typically MW < 1500 Da) [1], such as sugars, fatty acids, amino acids (Tryptophan, Phenylalanine …), but also some peptides (Glutathione …), organic acids, vitamins, steroids, xenobiotics and other exogenous molecules, which are intermediates and products of metabolic reactions (metabolism) catalyzed by various enzymes that naturally occur within cells [2].

They cover a wide range of chemical formulas, for example, the Human Metabolome Database (HMDB) (version 4.0) contains 114,100 metabolite entries including both water-soluble and lipid soluble metabolites as well as metabolites that would be regarded as either abundant (> 1 μM) or relatively rare (< 1 nM). In Plant Kingdom, it has been estimated that there is at least 200,000 different metabolites, and between 7000 and 15,000 within an individual species [3], [4].

Depending on their origin, metabolites can be distinguished as endogenous metabolites that are naturally produced by an organism (such as amino acids, organic acids, nucleic acids, fatty acids, amines, sugars, vitamins, co-factors, pigments, antibiotics, etc.) as well as exogenous metabolites (such as drugs, environmental contaminants, food additives, toxins and other xenobiotics) that are not naturally produced by an organism.

Endogenous metabolites can be further classified as primary and secondary metabolites. A primary metabolite is directly involved in the normal growth, development, and reproduction (such as sugars, organic acids, amino acids, phosphorylated sugars). A secondary metabolite is not directly involved in those processes, but usually has other functions (such as terpenes, flavonoids, alkaloids, drugs, toxins, xenobiotics, etc). Primary metabolites are ubiquitous (bacterial, plant and animal kingdoms), the term of “secondary metabolites” is particularly used for plant or microbial metabolites. Secondary metabolites may include pigments, antibiotics or waste products derived from partially metabolized xenobiotics [5].

The term of metabolome was initially proposed in the literature in 1998 by Oliver et al [1], and Tweeddale et al [6]. In parallel to the terms of genome, transcriptome and proteome, the term of metabolome represents all metabolites contained in a biofluids (such as urine, blood plasma…), tissue or cell of a living organism (Figure 1).

Gene	**mRNA**	**Proteins**	**Metabolites**
Genome	(Retro-)Transcriptome	Proteome	Metabolome
Genomics	(Retro-)Transcriptomics	Proteomics	Metabolomics

Figure 1. An overview from genomics to metabolomics (inspired by Dettmer et al, [7]).

Compared to genome, transcriptome and proteome studies, the metabolome study has complementary advantages. One of great advantage is that metabolites are endpoint products of interactions between biological systems, genome and environment, thus, compare with genome, transcriptome and proteome, metabolome may better reflect molecular phenotypic behavior of a living organism [7].

Metabolomics is the science designed to comprehensively study the metabolome, which gives the broadest insight into theses chemical fingerprints (metabolites) by identifying and quantifying them, eg. establishing a profile of the metabolites of a studied sample. By identifying and quantifying metabolites, metabolomics gives a comprehensive snapshot of the physiological state of the studied extract or cell [8], [9].

Technically speaking, metabolomics can be also defined as systematic analysis of metabolites in biofluids [10], [11], tissues [12], [13] or cells [14], [15] and investigate metabolites changes (or perturbations) during diseases (eg., cancer) [16]–[18], physiological processes (eg., aging) [19] or external stimulus (eg., drug treatment) [20], [21]. Thus, measuring metabolites by using metabolomics is a very important complementary to genome, transcriptome and proteome studies, which may improve

our understanding of how genetics, environment, the microbiome, disease, drug exposure, diet, and lifestyle all influence the phenotype [2], [22], [23].

1.2 Metabolomic study

The two main approaches frequently used in metabolomics are untargeted and targeted metabolomics, just like all method, each has their own advantages and drawbacks. Other approaches which have been proposed in the literature such as Pseudotargeted metabolomics and Imaging metabolomics are also discussed in the present section. However, in order to facilitate the understanding of the methods used in the following part, I will briefly describe these types of metabolomic studies here, a more exhaustive description of each metabolomics study can be found in these books [24]–[27].

1.2.1 Untargeted metabolomics

Untargeted or global metabolomic analysis usually involves comparing the metabolome of control and test groups (such as disease or treated group) to identify differences between their metabolite profiles [10], [28]–[30].

An untargeted experimental workflow is usually composed of three steps: 1. Profiling, in order to seek the metabolites with statistically significant variations in abundance within a set of experimental and control samples; 2. Identification of metabolite, and elicitation of the chemical structure; 3. Interpretations, which makes connections between the identified metabolites and the biological processes.

Untargeted metabolomics try to cover as many metabolites as possible present in a biological sample, which is a very useful tool during the primary phase in the biomarker discovery as this approach is non-hypothesis driven, with a wide range of metabolite classes. That's why this method was also called discovery metabolomics. A lot of new biomarkers has identified by using this method.

Now, between 200 and 500 metabolites can be detected by untargeted metabolomics, However, it is not yet possible to detect all metabolite classes as uncompleted NMR and

MS database, convolution and ionization problem exist. Furthermore, a lot of compounds detected in this method remain unknown in metabolite databases [31].

1.2.2 Targeted metabolomics

In contrast to untargeted or global metabolomic analysis, targeted metabolomics aim at quantitation of a preselected set of metabolites (targeted metabolites) [32]–[34].

Targeted metabolomics is often used in the confirmation and validation stage in the biomarker discovery, with highly advantage in specificity and in quantitation. To estimate a metabolite concentration, a standard curve for a concentration range of the metabolite of interest by using they chemical standard is established. Thus, to perform targeted metabolomic analysis, the chemical standard for the metabolite of interest should be available or should be easy to synthetized [32], [35].

1.2.3 Pseudotargeted metabolomics

Pseudotargeted metabolomics was initially proposed in the literature by Prof. Xu's group where I performed part of my thesis work. This method is a new approach combining the advantages of both untargeted and targeted methods. Briefly, a pooled (2X or 3X times) concentrated Quality Control (QC) sample was analyzed at first by using untargeted method in order to cover the maximum metabolites features possible, then a targeted quantitative analysis with the detected metabolites in previous steps was performed with real samples in order to estimate they quantity [36], [37].

1.2.4 Imaging metabolomics

Untargeted and targeted metabolomics which represented above involve the extraction of metabolites from a sample, and the homogenization of the samples before measurement. Consequently, all metabolite spatial distribution information is lost. In imaging metabolomics, in contrast, a thin section of sample (basically a small portion of tissue/organ) is measured by using mostly a mass spectrometer while leaving location

information intact on the sample, thereby permitting measurement of metabolite distribution information [31], [38]–[40].

Imaging mass spectrometry techniques, such as matrix-assisted laser desorption ionization (MALDI) [41], nanostructure-imaging mass spectrometry (NIMS) [42], desorption electrospray ionization mass spectrometry (DESI) [43] and secondary ion mass spectrometry (SIMS) [44] are frequently used techniques in Imaging metabolomics, among them, NIMS and DESI are particularly suited to the analysis of small molecules [31]. MS-imaging data can also combine with detailed optical microscope images in order to get more relevant biological information [45], [46].

1.3 Applications of Metabolomics

Unlike well-established approach such as transcriptomics, so far, metabolomics is still in its infancy, however, related studies revealed already its considerable potential applications in various research fields such as Agriculture (e.g. development of new pesticides, improve genetically modified plants) [47], [48], Precision medicine (e.g. newborn screening for inborn errors of metabolism, customize drug treatments) [49], [50], Drug discovery (e.g. identify new pathways / novel drug targets, toxicology test) [51]–[53] and Biomarker discovery (e.g. cancer biomarker, diabetes biomarker, Alzheimer's disease biomarkers …) [21], [31], [54], it is also important to know that its application will not be only limited in these fields. The Figure below list possible metabolomic applications.

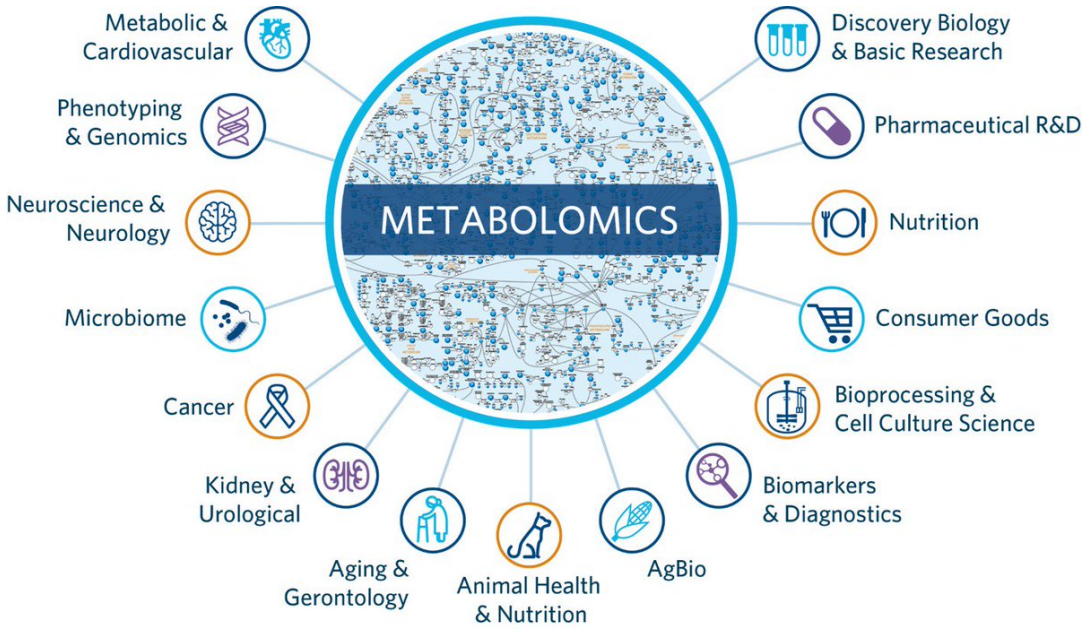


Figure 2. A schematic representation of metabolomic applications (www.metabolon.com).

In health-related field, especially in clinic research, one important application of metabolomics is the discovery for early disease Biomarker [55]. According to Biomarkers Definitions Working Group, a biological marker (biomarker) is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention [56].

My PhD thesis focus on the Biomarker discovery part of metabolomics, especially on (1) using NMR and UPLC-HRMS based metabolomic and lipidomic profiling, to identify novel plasma biomarkers which characterize the different stages: normal liver (NL), Non-alcoholic fatty liver (NAFL) or simple steatosis and nonalcoholic steatohepatitis (NASH) of Non-alcoholic fatty liver disease (NAFLD), by (2) combining UPLC-HRMS based untargeted metabolomics with epidemiology approach, to identify plasma biomarkers which are associated with the risk of developing prostate cancer (PCa) within the following decade, and (3) application of NMR based metabolomics in Sepsis and Septic shock.

My thesis is composed of 4 parts:

The first part is a general introduction about metabolomics, the definition of different term used in metabolomics, the different types of metabolomic study, some applications of metabolomics were also discussed with further focus on the applications of metabolomics in NAFLD, in Prostate Cancer (PCa).

The second part is methodology, in this section, sample preparation, Data acquisition, Analytical technologies, Data preprocessing, Data analysis and Interpretation used in metabolomics study were detailed.

The third part is the presentation of my PhD thesis work, application NMR and UPLC-HRMS based Metabolomics in Non-alcoholic fatty liver disease (NAFLD), in Prostate Cancer Biomarker Discovery and in Sepsis and Septic shock.

Finally, the last part is the general conclusion and outlook of the thesis.

1.3.1 Metabolomics studies in NAFLD: State-of-the-art

The liver is one of the largest organs in our body. It plays many important functions in metabolism, such as carbohydrates, protein and lipids metabolism. In normal condition, it converts the macronutrients (such as carbohydrates, protein and lipids) in our diets into substances that the body can use (glucose, amino acids and fatty acids), stores these substances in form of glycogen or fatty acids, and supplies cells with glycogen or fatty acids when needed [57].

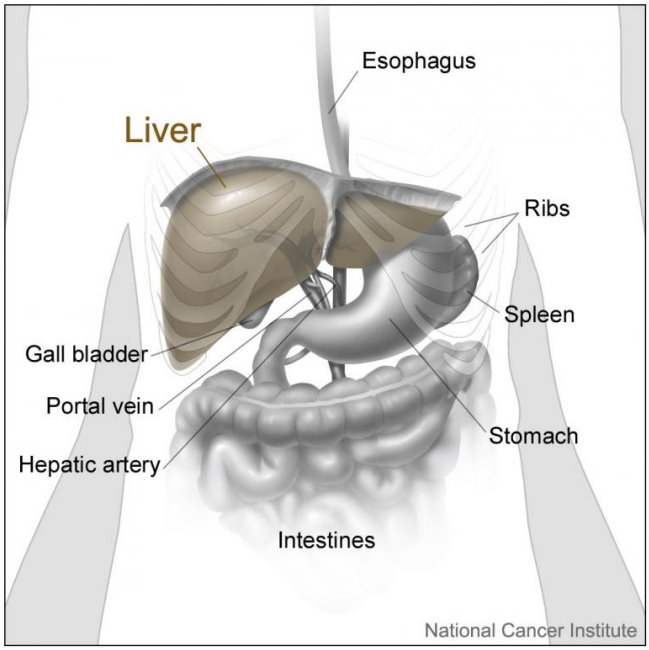


Figure 3. The liver and nearby organs. (Image credit: Don Bliss, National Cancer Institute)

However, in abnormal condition such as perturbation of lipids metabolism, liver may be subject to damage, even develop disease. One of the common chronic liver conditions is Non-alcoholic fatty liver disease (NAFLD) [58], [59]. NAFLD characterized by abnormal accumulation of lipids mainly triacylglycerols (TGs) in the liver, based on clinical-histologic characters, NAFLD spectrum range from simple fatty liver (NAFL) or steatosis to the advanced form termed NASH, without therapeutic intervention, a subset of patient with NASH will subsequently progress towards cirrhosis and, ultimately, hepatocellular carcinoma [60].

So far, liver biopsy is the gold standard for diagnosis, staging and monitoring progression of NAFLD during treatments. However, biopsy has well-known limitations, such as invasiveness, poor acceptability by patients, sampling variability, and financial cost… which limit its application in large-scale population. Moreover, recently developed Noninvasive imaging biomarker assessment method, even the most accurate noninvasive liver elastography based methods, such as vibration-controlled transient elastography (VCTE), magnetic resonance elastography (MRE), shear-wave elastography and acoustic radiation force impulse have other limits including couldn't access inflammation, with very limited guidance (or even unavailable) for how clinicians should anticipate and manage the pitfalls of these tests [61]–[64].

Thus, the development of an alternative noninvasive and familiar for clinicians' strategy such as using non-invasive biomarkers is an urgent need for prognostication, early detection, staging, selection of patients for treatment and monitoring of disease [58]. Metabolomics is the science designed to comprehensively study the metabolome, the repertoire of small molecule metabolites, which has been used to investigate in prognosis, risk estimation, early diagnosis, and identification of novel biomarkers of NAFLD. Recent metabolomic studies in NAFLD and NASH were summarized in tables below [65] (Table 1, Table 2).

Table 1. Summary of recent metabolomics studies in NAFLD.

Reference	Tissue	Platform	Decreased or increased metabolites in the patients compared to the control group
Puri P (2009)	Plasma	HPTLC GCFID LCMS	palmitoleic, oleic acids dihomo gamma-linolenic (20:3n6) acids, Triacylglycerols, palmitoleic acid to palmitic acid (16:0) ratio Linoleic acid
García-Canaveras JC(2011)	liver	UPLCESIQTOFMS	hypoxanthineglutamine, γ-glutamyl-dipeptides creatinine, GCDCA , TCDCA
Kalhan SC(2011)	plasma	UPLCESITQMS GCMS	L-glutamyl-L-lysine, Lleucyl-L-proline, glutamate, carnitine, butyrylcarnitine (C5), glutamyl dipeptides, glutamyl valine, glutamyl leucine, glutamyl phenylalanine and glutamyl tyrosine, methylbutyrylcarnitine, glycocholate, taurocholate, glycochenodeoxycholate, Mannose and lactate glutathione, long-chain fatty acids and cysteine-glutathione
JonathanBarr(2010)	serum	UPLC/MS	lysophosphatidylcholine (LPC),deoxycholic acid bile acids (BAs), sphingomyelin: (SM 36:3), (d18:2/16:0), (d18:2/14:0), (d18:1/18:0), (d18:1/16:0), (d18:1/12:0), and (d18:0/16:0) Creatine

Red color: increased metabolite
Blue color: decreased metabolite

Adapted from Safaei A, et al. [65].

Table 2. Summary of recent metabolomics studies in NASH.

Subject	Tissue	Platform	Decreased or increased metabolites in the patients compared to the control group
Puri P etal(2009)	plasma	HPTLC GCFID	11-HETE(5-HETE), 8-HETE and 15-HETE
		LCMS	docosahexanoic acid (22:6 n3) to docosapentenoic acid (22:5n3) ratio
Kalhan SC(2011)	plasma	UPLCESITQMS GCMS	aspartate, glutamate, phenylalanine, tyrosine, lactate,isoleucine, leucine, valine, acylcarnitines (C3, C4 and C5), γ-glutamyl-tyrosine glucose, xanthine pyruvate, carnitine, butylcarnitin,free carnitin,methylbutylcarnitin, GCA, TCA, GCDCA,glutamyl dipeptide phenylacetate, indolepropionate, cysteine-glutathione disulfide, glycerate, glycerophosphocholine, LPC(18:1), LPC(18:2), LPC(20:4)
Jonathan Barr(2010)	serum	UPLCESIQTOFMS	NASH vs. NAFLD: PC(14:0/20:4), LPC(18:1) NASH vs. NAFLD: LPC(24:0)

Red color: increased metabolite
Blue color: decreased metabolite

Adapted from Safaei A, et al. [65].

NAFLD is a heterogeneous and complex disease [66], Alonso C. et al., identified 2 major subtypes of NAFLD, M-subtype and non-M-subtype [67], characteristics of each subtype were detailed in the table below.

Table 3. Nonalcoholic fatty liver disease subtype classification, Mato et al. [66].

NAFLD subtype	Characteristics	Mouse model-based classification	Treatments tested
M-Subtype	Increased fatty acid uptake. Low liver glutathione and SAMe content. Reduced synthesis of PC-PUFA. Abnormal VLDL-TG assembly and export.	*Mat1a* KO[57] 0.1MCD[59]	SAMe Aramchol
Non-M-Subtype	Increased DNL. Normal hepatic SAMe levels. Normal VLDL-TG secretion. High serum levels of cholesterol and TG.	*Ldlr* KO/HFD[68]	Obeticholic acid

NAFLD: Nonalcoholic fatty liver disease; SAMe: S-adenosylmethionine; PC-PUFA: Phosphatidylcholines containing polyunsaturated fatty acids; VLDLTG: Very low-density lipoprotein-triglycerides; DNL: de novo lipogenesis; KO: Knockout; MCD: Methionine and choline deficient.

In the further research, investigation subtypes of this heterogeneous and complex disease could be a novel perspective direction. Moreover, combine with other omics research such as transcriptomics, proteomics, and also clinical characteristics may improve novel subtyping approach of NAFLD patients, allowing further more precisely classification and staging of patients, in order to correctly interpret the biochemical processes behind the disease, which could contribute to the development of appropriate therapy and precision medicine-based management of patients.

1.3.2 Metabolomics studies in PCa: State-of-the-art

Prostate cancer (PCa) is the second most commonly diagnosed cancer and the second leading cause of cancer death (7.1% for incidence) among males [68]. Currently, there is no single definitive test to identify prostate cancer in men [69]. Prostate-Specific Antigen (PSA) test and digital rectal examination are screening methods used for PCa, for the definitive diagnosis, prostate biopsy and supplementary imaging are required [70]. The PSA test is a relatively easy to perform test and applicable for population in large scale, however, it has well known limits such as sensitivity, specificity, and can lead to false-positive and false-negative results [69].

Although extensive efforts in biomarker discovery during the last decades, including the genome and transcriptome approach, which has contributed to the identification of predictive biomarkers, more sensitive and specific biomarkers are still very demanding in early detection, prognosis, monitoring, and clinical management of PCa patients [71]–[74]. Metabolomics, defined as systematic analysis of metabolites in biofluids [10], [11], tissues [12], [13] or cells [14], [15] and investigate metabolites changes (or perturbations) during diseases (eg., cancer) [16]–[18], physiological processes (eg., aging) [19] or external stimulus (eg., drug treatment) [20], [21], has shown to be a promising and powerful tool to identify novel PCa biomarkers [75]–[79]. The figure below displayed major metabolic pathways changes in the tissues (Figure 4).

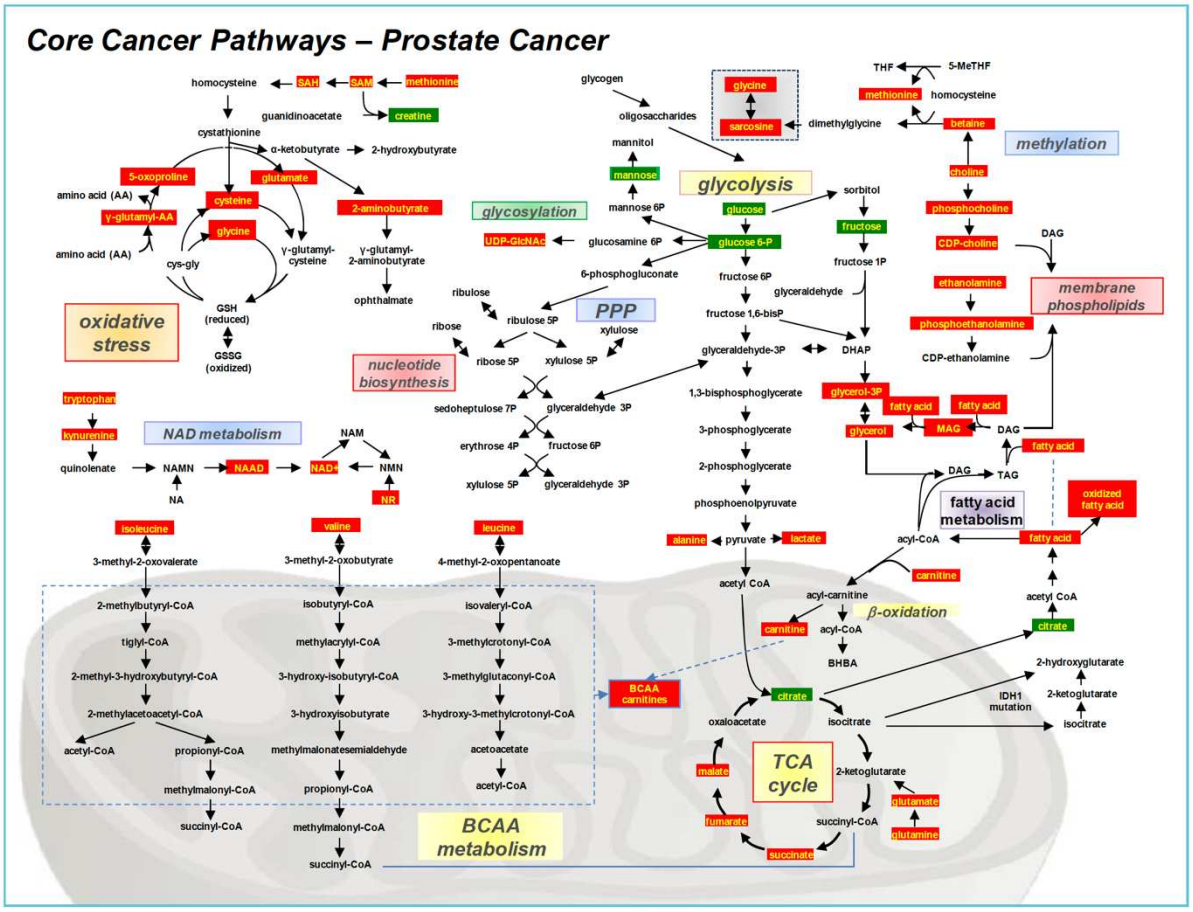


Figure 4. Changes in core metabolic pathways in prostate cancer [77]. Major metabolic pathways changes in the malignant tissues compared with normal prostate tissues are displayed. Metabolites in red boxes have been observed to be increased in prostate tumors relative to cancer-free prostate tissue, while green boxes indicate a decrease [80]. Elevated metabolites are seen in pathways related to membrane phospholipid synthesis, methylation and oxidative stress. Increases in branched chain amino acid (BCAA) metabolism are suggested by an increase in BCAA related carnitines and an increase in the three BCAAs. TCA cycle intermediates were elevated along with glutamine and glutamate which can feed the TCA cycle through 2-ketoglutarate. Citrate, which acts as an intermediate in the TCA cycle and is also utilized in fatty acid synthesis, was observed to be lower in prostate cancers. Unlike many tumor types, prostate cancer tissue did not display a large increase in glycolysis intermediates typical of a shift in energy metabolism away from mitochondrial oxidative phosphorylation and toward aerobic glycolysis – although lactate and alanine which can be markers of increased aerobic glycolysis were elevated. The inter-conversion of glycine and sarcosine is highlighted in the dashed box (Legend of figure adapted from E. McDunn et al. [77]).

For studies with biofluid, one of remarkable example is sarcosine, an N-methyl derivative of glycine, Sreekumar et al. found that sarcosine is a differential metabolite that was highly increased during prostate cancer progression to metastasis, also, sarcosine can be detected non-invasively in urine [81], which could be a very good biomarker candidate. However, whether sarcosine could use as a reliable biomarker for prostate cancer is still in discussion in the scientific committee [82], some study confirm this finding [83], but other not [84]. Potentially however, serum PSA concentrations in relation to serum sarcosine concentrations might have additional diagnostic value [84].

Apart from metabolism, there is other direction as well, which could promote application of metabolomics to prostate cancer, such as data processing. C. Pérez-Rambla et al. [85] show that variable selection such as the regression coefficient (b-coefficient) based method [86], [87] improved the classification predictiveness of model.

Accumulated evidence suggests that metabolic alterations specific to prostate carcinogenesis and progression may represent potential metabolic biomarkers. In the further research, validation of promising biomarkers should be a priori, and a number of approach such as transcriptomics, proteomics should be used as complement to promote and validate metabolomic findings in the study of prostate cancer [79].

Part Two METHODOLOGY

2.1 General workflow of a Metabolomic study

Several very informative and well detailed protocols and reviews are available in the literature which described the different operations in a typical Metabolomic study [88]–[90]. Essentially, a metabolomic study consists of several steps, the main steps (partly represented in the Figure 5, which is an analysis workflow of an untargeted metabolomic) include experimental design (didn't represented in the Figure 5), sampling and storage, samples preparation, data acquisition, data processing and analyses, metabolites feature identification, and biological interpretation.

Almost any metabolomic approach start by one or more biological or clinical questions to which we wish to answer. Whether searching for early disease biomarkers, monitoring the effects of treatment or study effect of targeted gene in the regulation of metabolism … the experimental design must be carefully thought out to reduce as much as possible bias and avoid the introduction of irrelevant variables [91].

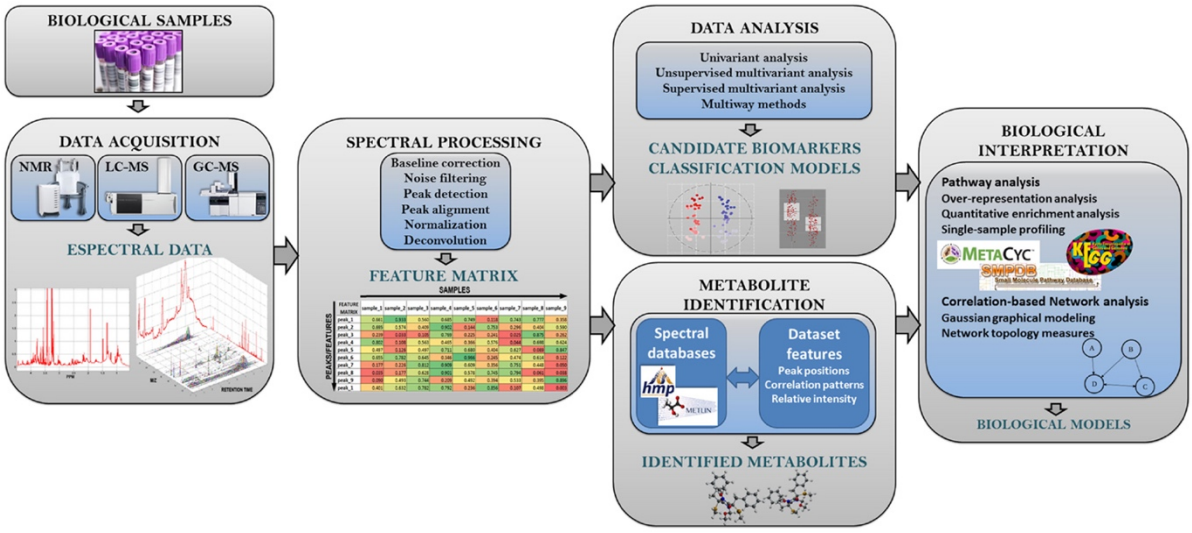


Figure 5. Main steps of the metabolomic analysis [90].

Also, when designing the study, it may be necessary to establish more specific recommendations or specific inclusion criteria in order to have the most homogeneous population possible.

The steps for samples preparation, data acquisition, data processing and analyses, metabolites feature identification, and biological interpretation will be described in the following section of the present part.

2.2 ^{1}H NMR Spectroscopy based metabolomics

In order to study small molecules (metabolites) in biological samples (such as urine, blood plasma, blood serum, saliva, …), a number of different high throughput, sophisticated analytical instruments can be used, such as Nuclear Magnetic Resonance (NMR) spectrometer and Mass Spectrometer (MS).

Ongoing advances in analytical techniques including NMR Spectroscopy and Mass Spectrometry (MS) will certainly lead to a continuous improvement of the breadth and throughput of metabolomic analysis.

The NMR spectroscopy is one of the two most used techniques in metabolomic analysis. Here, in order to facilitate the understanding of the results presented in the following part, I will shortly describe the basic principles of NMR. A more exhaustive explanation of principles and physics related to NMR can be found in books, for example, see [92].

2.2.1 The basic principle of NMR

NMR is based on the magnetic properties of certain atomic nuclei (e.g. ^{1}H, ^{13}C, ^{31}P, ^{19}F). In quantum mechanics and particle physics, spin is an intrinsic physical property of the atomic nucleus. It is characterized in particular by an intrinsic magnetic moment, if the moment is different from zero, the spin will give magnetic properties to the nuclei which will be exploitable in the NMR spectroscopy [92].

Theoretically, without an external magnetic field, spins of a given atomic nucleus (for example the nuclei of hydrogen) have the same energy level, when an external magnetic field (B_0) is applied, in the case of the spin 1/2 nucleus, nuclei spins will split into two energy levels, higher energy (-1/2 or β) and base energy (+1/2 or α) level (called The Zeeman effect). The Boltzmann distribution describes the population of nuclei in each spin state, equation: (Eq. 0.1).

$$N\beta/N\alpha = e^{-\left(\frac{\Delta E}{kT}\right)} \qquad \text{(Eq. 0.1)}$$

where N_α and N_β represent the number of spins expect to measure in the α and β states, *ΔE* is the difference in energy between the two states, *k* is the Boltzmann constant (1.381 x 10^{-23} joules/°K), and *T* is the absolute temperature in Kelvin degrees.

And the difference in energy between the two spin states can be described as (Eq. 0.2):

$$\Delta E_{\alpha \rightarrow \beta} = E_\beta - E_\alpha = (h/2\pi)\gamma B_0 \qquad \text{(Eq. 0.2)}$$

where ΔE is the difference in energy between the two states, *h* is the Planck constant, γ is the gyromagnetic ratio.

Between the β and the α energy states, an energy transfer is possible when an external magnetic field (B_1) is applied. The energy transfer happens by the way of a wavelength that corresponds to the energy of radio frequencies received and when the spin returns to its base level, energy is emitted at the same frequency. The signal, called Free Induction Decay (FID) that correspond to this transfer is measured in time domain and processed into frequency domain (Fourier Transform) in order to yield an NMR spectrum for the nucleus studied [92]. The schema below (Figure X) simplify resumes the principle of the NMR spectroscopy.

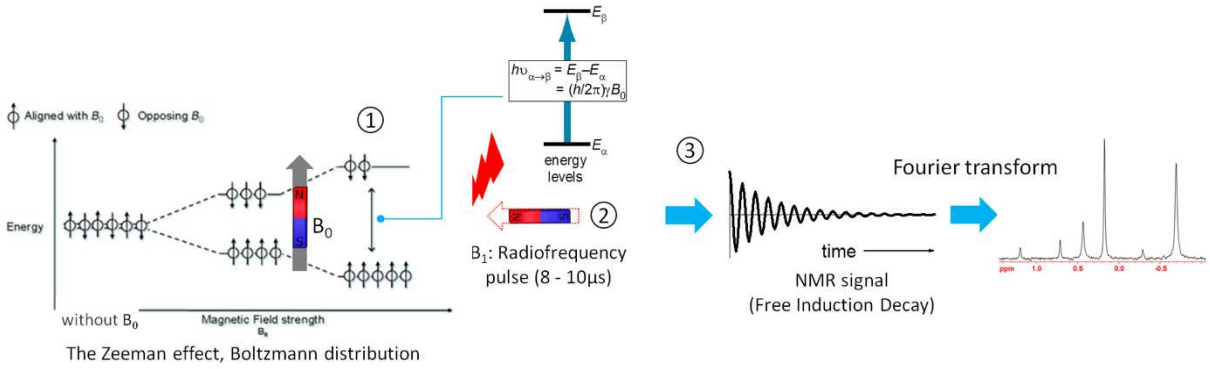


Figure 6. A simplified explanation of the principle behind NMR spectroscopy [92].

1. Without external magnetic field (B_0), nucleus spins have the same energy state. When an external magnetic field (B_0) is applied, nuclei spins will be split into two energy states, higher energy (β) and base energy (α) states.

2. Between the β and the α energy states, energy transfer is possible by application of B_1 in form of a radio frequencies pulse, the energy corresponds to the energy of radio frequencies pulse received. When the spin returns to its base level, a magnetization can be recorded at the same frequency.

3. The signal, called free induction decay (FID) that correspond to this transfer is measured in time domain and processed into frequency domain (called "Fourier Transform") in order to generate an NMR spectrum for the nucleus studied.

2.2.2 Chemical shift

For a given atomic nucleus, the effective magnetic field strength (B_{eff}) at the nucleus is affected by the magnetic field generated by the movement of electron surround-by (electron shielding), which is dependent on the chemical environment of the nucleus. The relationship between the external field strength (B_0) and the effective field strength (B_{eff}) for the nuclei can be described as:

$$B_{eff} = B_0 (1 - \sigma) \qquad \text{(Eq. 0.3)}$$

where σ is the screening constant.

The chemical shift (δ) is measured with respect to a reference signal depending on the frequency of the spectrometer by the following relation (Eq. 0.4). This value defines the position of the signal on the frequency axis.

$$\delta = (\nu_{\text{compound}} - \nu_{\text{ref}})/\nu_{\text{ref}} \qquad \text{(Eq. 0.4)}$$

where "ν_{compound}" is the absolute resonance frequency of a test compound and "ν_{ref}" is the absolute resonance frequency of a standard reference compound, measured in the same applied magnetic field B_0. Usually, the numerator "$\nu_{\text{compound}} - \nu_{\text{ref}}$" in this equation is expressed in "hertz", and the denominator "ν_{ref}" in "megahertz", thus, the chemical shift "δ", using in this equation, is not dependent on the magnetic field and it is expressed in "parts per million" (ppm) by frequency.

It is customary to adopt tetramethylsilane (TMS) as the proton reference frequency, because the precise resonance frequency shift of each nucleus depends on the magnetic field used. For TMS, which is set to "ν_{ref}" in this equation, its chemical shift (δ) is "0".

2.2.3 NMR Instrumentation

An NMR spectrometer consists essentially of the superconducting magnet (composed of niobium titanium), a transmitter and a high radio frequency receiver (Figure 7). The sample to be analyzed is introduced into the measurement cell called probe, itself placed in the magnetic field B_0. The probe excites the nucleus in sample with high radio frequency radiation and also receives the signal from the relaxed nucleus. The spectrum is then recorded after amplification and processing.

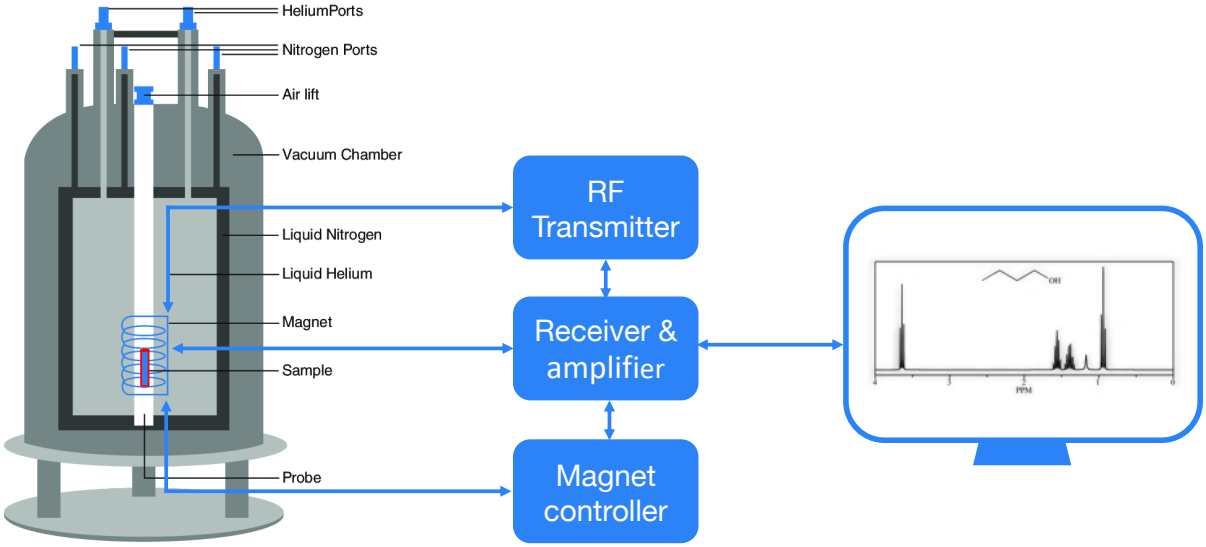


Figure 7. Simplified illustration of an NMR spectrometer [93].

2.2.4 High Resolution Magic Angle Spinning Probes (HR-MAS Probes)

One of the advantages for NMR is that analysis of intact biological tissue samples is possible, HR-MAS is an established technique for analyzing intact biological tissue samples. By spinning at the magic angle ($\theta = 54.7°$), line broadening effects due to dipolar interactions and susceptibility differences within the sample are removed resulting in high resolution quality spectra, a more detailed description about HR-MAS for metabolomic analysis can be found in the review of Beckonert et al 2010 [13], [94].

2.2.5 Several important parameters for NMR data acquisition

In order to obtain comparable spectra, during NMR data acquisition, the experimental conditions must be comparable and therefore optimized for the analysis of all the samples. The main parameters to consider when acquiring a spectrum are listed below [95], [96], setting these parameters will determine the nature and quality of the NMR spectrum.

- Spectral Window (SW). It defines the range of frequencies observed.

- Acquisition time of the FID (AQ). It must be long enough to maximize the amount of signal that is contained in the FID and ensure good digital resolution.

- Number of points acquired on the FID (TD, time domain data size) in other words, the NMR data size, in general, a larger the number of points defining the FID will correspond to a higher spectral resolution.

The relationship between acquisition time (AQ), the spectral window (SW) and the number of points (TD) can be represented by the following equation:

$$\mathrm{AQ} = \mathrm{TD}/2\mathrm{SW} \qquad \text{(Eq. 0.5)}$$

- Relaxation time (d1) correspond to the return to equilibrium of the magnetization vector before each new sequence of pulses. The FID is decreasing exponentially with the relaxation times, it is usually recommended to wait 5 times the value of the highest T1 in order to obtain the complete return of the magnetization vector after a 90 ° pulse.

- Number of scans accumulations (NS). In order to improve the sensitivity, a series of n pulses can be applied immediately after the recording of the first signal allowing the recording of n signals which will accumulate before obtaining the spectrum. The signal-to-noise ratio (S/N) being proportional to $\sqrt{n}$, a greater the number of accumulations will correspond to a better S/N.

- Receiver Gain (RG). The receive gain controls the amplitude of the FID which itself depends on the concentration of the sample. The optimum gain determined corresponds to the maximum value obtained for the intensity of the strongest signal. In order to be able to compare the spectra, it is recommended to have the same value of gain and to fix this value a little below the value that has been optimally determined to avoid saturating in one of the spectra, especially if the samples to be analyzed differ between them.

- The duration of the radiofrequency pulse sent is of the order of a few µs. P1 defines the time required to fully switch the magnetization vector of the Z axis in the XY plane for a 90 ° pulse. This duration is dependent on the intensity of the irradiation.

- Power (PW). PL sets the attenuation parameter on this irradiation intensity.
- Irradiation frequency, SFO1, corresponds to the frequency sent to excite the desired nucleus. It is composed of two terms: BF1, which corresponds to the base frequency recorded for the chosen kernel (500 MHz for 1H corresponding to a B0 of 11.7 tesla) and O1 ("offset value"), which makes it possible to adjust the exact value of the reference frequency. Setting of O1 makes it possible to focus on the spectral region of acquisition.

2.3 Mass Spectrometry based metabolomics

The other most used Analytical technologies in metabolomics is Mass Spectrometry (MS). The principle of MS based metabolomics is measuring the mass-to-charge ratio (m/z) of ions to identify and quantify molecules. As in the NMR section, here, I will briefly represent the basic principles of MS in order to facilitate the understanding of the results presented in the following part. A more exhaustive explanation of principles and physics related to MS can be found in "Mass Spectrometry: A Textbook" [97].

2.3.1 The basic principle of Mass Spectrometry

The principle of the mass spectrometry is fully described in reviews, briefly, "the basic principle of MS is to generate ions from the sample molecules by thermally, by electric fields or by impacting energetic electrons, ions or photons, to separate these ions by their m/z and to detect them qualitatively and quantitatively by their respective m/z and abundance. The ions can be single ionized atoms, clusters, molecules or their fragments or associates. Ion separation is effected by static or dynamic electric or magnetic fields." [97], [98].

2.3.2 MS Instrumentation

Fundamentally, a mass spectrometer contains an ion source, a mass analyzer and an ion detector (Figure 8). The analyzer, detector, and often the ionization source too, are maintained under high vacuum to ensure the ions travelling through the instrument without any impact from air molecules (such as N_2, O_2). Samples are introduced into the mass spectrometer in liquid or gas form and then vaporized and ionized by the ion source [97].

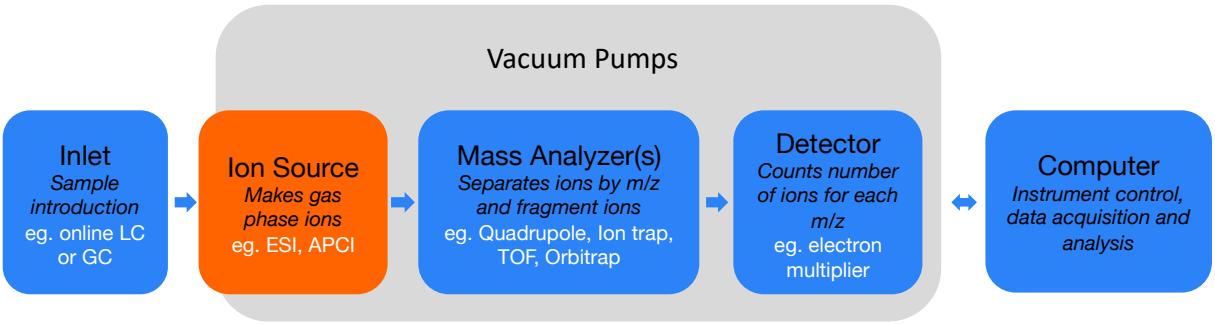


Figure 8. Simplified illustration of a Mass spectrometer [97].

2.3.3 Ionization

Ionization is a process to generate charged ions from the sample molecules. Several ionization methods are proposed in the literature, such as Atmospheric Pressure Chemical Ionization (APCI), Chemical Ionisation (CI), Electron Impact (EI) and Matrix Assisted Laser Desorption Ionisation (MALDI). The detailed description for each ionization methods are available in these publications [99]–[103]. The choice of the method should depend on the nature of molecules in samples to be studied.

During my thesis work, Electrospray Ionization (ESI) is the main method we used, ESI is one of the Atmospheric Pressure Ionization (API) techniques and is well-suited to the analysis of polar molecules with molecular mass ranging from less than 100 Da to more than 1,000,000 Da [104], [105]. In ESI, the ionization mode can be positive or negative. The principle of ESI is described in the figure 9.

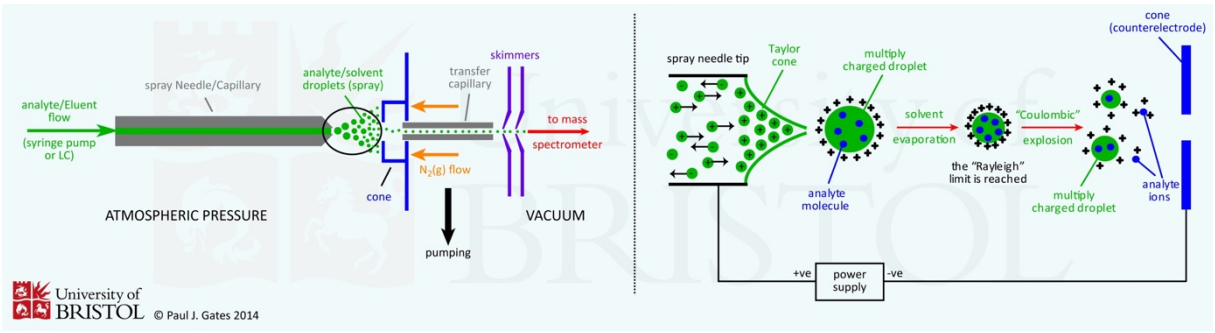


Figure 9. Simplified explanation of Electrospray Ionization (ESI). (by Paul J. Gates 2014)

The sample to be analyzed such as blood plasma (after preparation) is introduced into the ion source with a polar, volatile solvent (such as methanol or acetonitrile) and pumped (from a syringe pump or as the eluent flow from liquid chromatography) through a narrow, stainless steel capillary (75 - 150 micrometers i.d.) with a flow rate between 1 µL/min and 1 mL/min.

A high voltage (from 2.5 to 5 kV) is applied to the tip of the capillary (right part in the figure), as a consequence, the sample emerging from the tip is dispersed into an aerosol of highly charged droplets (right part in the figure), a process that is aided by a co-axially introduced nebulizing gas flowing around the outside of the capillary (left part in the figure). This gas, usually nitrogen, helps to direct the spray emerging from the capillary tip towards the mass spectrometer.

The charged droplets diminish in size by solvent evaporation (right part in the figure), assisted by a warm flow of nitrogen known as the drying gas which passes across the front of the ionization source.

Finally, charged sample ions, free from solvent, are released from the droplets, some of which pass through a sampling cone or orifice into an intermediate vacuum region, and from there through a small aperture into the analyzer of the mass spectrometer, which is held under high vacuum. The lens voltages are optimized individually for each sample.

2.3.4 Mass Analyzer

The main function of the mass analyzer is to separate, or resolve, the ions formed in the ion source of the mass spectrometer according to their mass-to-charge (m/z) ratios. There are a number of mass analyzers currently available, such as quadrupoles (Q) [106], time-of-flight (TOF) [107] and ion trap mass analyzers [108]. These mass analyzers have different features, including the m/z range that can be covered, the mass accuracy, and the achievable resolution [97], [109], [110]. During my thesis work, the Quadrupoles, TOF, and Orbitrap are the mass analyzers we used.

2.3.4.1 Quadrupoles mass analyzer

As presented in Fig. X, a quadrupole mass analyzer consists of four parallel rods that have fixed Direct Current (DC) and alternating Radio Frequency (RF) potentials applied to them. The two opposite rods in the quadrupole have a potential of + (U + Vcos(ωt)) ('+' in the figure) and the other two - (U + Vcos(ωt)) ('-' in the figure) where 'U' is the fixed potential and Vcos(ωt) is the applied RF of potential 'V' and frequency 'ω'.

The applied potentials on the opposed pairs of rods varies sinusoidally as cos(ωt) cycles with time 't'. This results in ions being able to traverse the field free region along the central axis of the rods but with oscillations amongst the poles themselves. These oscillations result in complex ion trajectories dependent on the m/z of the ions.

Specific combinations of the potentials 'U' and 'V' and frequency 'ω' will result in specific ions being in resonance creating a stable trajectory through the quadrupole to the detector. All other m/z values will be non-resonant and will hit the quadrupoles and not be detected (Figure 10). The mass range and resolution of the instrument is determined by the length and diameter of the rods.

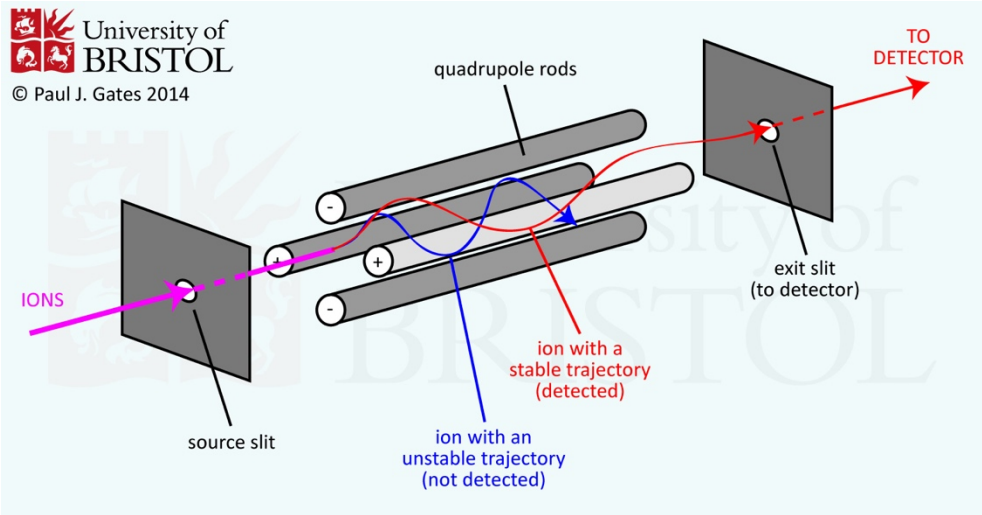


Figure 10. Simplified schematic of a Quadrupoles mass analyzer. (by Paul J. Gates 2014)

Quadrupoles mass analyzers are very commonly used in combination with either gas-chromatography (GC/MS) or liquid-chromatography (LC/MS) as a simple high throughput

screening system. Quadrupoles can also be placed in tandem to enable them to perform fragmentation studies - the most common set-up is the triple quadrupole (Q1qQ3) [111] mass spectrometer, where Q1 and Q3 are mass filters, q, is the collision cell, which enables basic ion fragmentation studies (tandem mass spectrometry MS/MS) to be performed.

2.3.4.2 TOF mass analyzer

The principle of TOF mass analyzer is shown in the Figure 11, which is a linear TOF in this figure. The ions are introduced either directly from the ion source of the instrument or from a previous analyzer (eg. Q-TOF) as a pulse. This results in all the ions receiving the same initial kinetic energy. As they pass along the field free drift zone, they are separated by their masses, lighter ions travel faster. This enables the instrument to record all ions as they arrive at the detector and so accounts for the technique's high sensitivity.

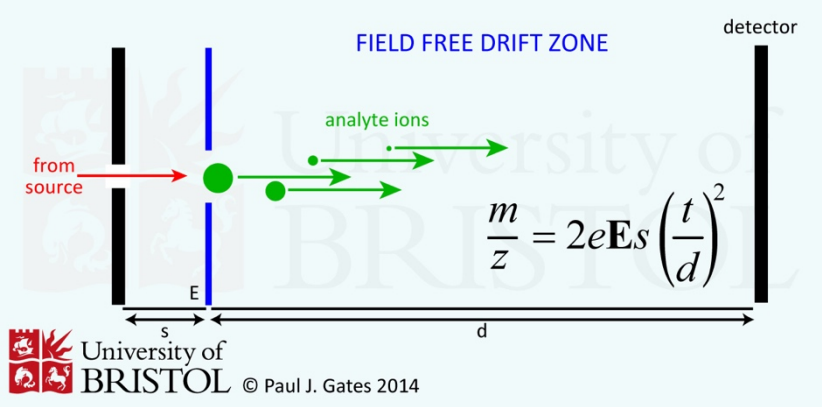


Figure 11. Simplified schematic of a TOF mass analyzer. (by Paul J. Gates 2014)

The principle of TOF mass analyzer separation ions by m/z can be described by equation included in the figure, where 'E' is the extraction potential, 's' is the length of the source, 'd' is the length of the flight tube and 't' is the time-of-flight for that particular m/z which is what is measured by the instrument.

2.3.4.3 Orbitrap mass analyzer

The Orbitrap is an ion trap mass analyzer that consists of two outer electrodes and a central electrode, which enable it to act as both an analyzer and detector.

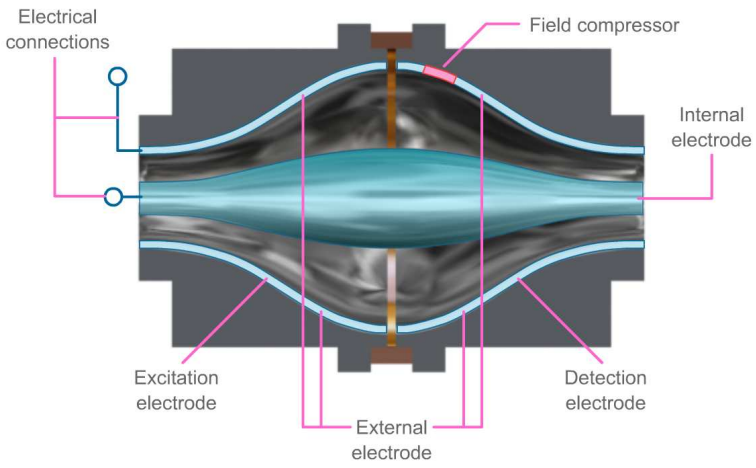


Figure 12. Simplified schematic of an Orbitrap (https://www.chromacademy.com).

Ion introduction into Orbitrap can be performed after modification of the electric field at the injection port. This can be achieved by using a field compressor which is a small portion of the outer electrode (Figure 12). Ions entering the Orbitrap are captured through "electrodynamic squeezing" after which they oscillate around the central electrode and between the two outer electrodes. Different ions oscillate at different frequencies, resulting in their separation [112].

By measuring the oscillation frequencies induced by ions on the outer electrodes, the mass spectra of the ions are acquired using image current detection [113].

The combination between these mass analyzers are also available, which is usually the case for modern mass spectrometer, such as Q-TOF, triple quadrupole-TOF, Q-Orbitrap…

2.3.5 Mass detector

The function of the detector is to respond to ions passing through the mass analyzer. It consists mainly of two parts: a high-energy dynode and an electron multiplier. Among them, the role of the high-energy dynode is to convert the charged ions into electrons, and the electron multiplier will amplify the generated electrons into electrical signals that the software can recognize. Then we have the m/z values of the ions are plotted against their intensities to show the number of components in the sample, the molecular mass of each component, and the relative abundance of the various components in the sample.

The most common types of ion detector used in modern instruments are the photomultiplier, the electron multiplier and the Faraday Cup detector [114].

2.3.6 Separation techniques coupled with mass spectrometry

Mass spectrometry is not only used to analyze pure compounds, but also used to analyze mixture compounds, the latter is even more common in metabolomics as biological samples are usually complex mixture. Although it is possible (and sometime desirable) to do total analysis of mixtures by direct injection, it is often preferable to combine on-line separation and/or chromatography with the mass spectrometry.

There are a number of combining techniques available and Gas Chromatography Mass Spectrometry (GC/MS) and Liquid Chromatography Mass Spectrometry (LC-MS) are the two methods we used during my thesis works.

2.3.6.1 Gas Chromatography Mass Spectrometry (GC/MS)

Gas chromatography mass spectrometry (GC/MS) comprising a gas chromatograph (GC) coupled to a mass spectrometer, by which complex mixtures of chemicals may be separated before MS analysis. The sample solution is injected into the GC inlet where it is vaporized and swept onto a chromatographic column by the carrier gas (usually helium).

The sample flows through the column and the compounds comprising the mixture of interest are separated by their relative interaction with the coating of the column (stationary phase) and the carrier gas (mobile phase). The latter part of the column passes through a heated transfer line and ends at the entrance to ion source where compounds eluting from the column are converted to ions.

In order for a compound to be analyzed by GC/MS, it must be sufficiently volatile and thermally stable. In addition, functionalized compounds may require chemical modification (derivatization), prior to analysis, to eliminate undesirable adsorption effects that would otherwise affect the quality of the data obtained.

2.3.6.2 Liquid chromatography coupled to mass spectrometry (LC-MS)

Liquid chromatography (LC) is a widely used method of sample separation prior to analysis and is frequently coupled with mass spectrometry. With LC-MS, solubilized compounds (the mobile phase) are passed through a column packed with a stationary (solid) phase. This effectively separates the compounds based on their weight and affinity for the mobile and stationary phases of the column. This also leads to fragmentation of the sample and its anionization through loss of H^+ ions.
During my thesis work, Ultra-Performance Liquid Chromatography [UPLC® Technology] was used, UPLC is a combination of a 1.7µm reverse-phase packing material and a chromatographic system that can operate at pressures in the 6000-15000psi range, these configuration allow dramatic increases in resolution, speed and sensitivity compared to a conventional liquid chromatography [115].

2.4 Comparison between NMR and MS based metabolomic analysis

In comparison with NMR and MS, each technique provides broad coverage of many classes of organic compounds, including lipids, amino acids, sugars, biogenic amines and organic acids. A detailed comparison of different analytical technologies used in metabolomics is presented in the Table 4.

Table 4. A comparison of different metabolomic analytical technologies.

Technology	Advantages	Disadvantages
NMR spectroscopy	Quantitative	Not sensitive (LOD = 5 μM)
	Non-destructive	High start-up cost (>US$1 million)
	Fast (2-3 min per sample)	Large instrument footprint
	Requires no derivatization	Cannot detect or identify salts and inorganic ions
	Requires no separation	Cannot detect non-protonated compounds
	Detects most organic classes	Requires larger sample volumes (0.1-0.5 mL)
	Allows identification of novel chemicals	
	Most spectral features are identifiable	
	Robust, mature technology	
	Can be used for metabolite imaging (fMRI or MRS)	
	Can be fully automated	
	Compatible with liquids and solids	
	Long instrument lifetime (over 20 years)	
GC-MS	Robust, mature technology	Destructive (sample not recoverable)
	Modest start-up cost (~$150,000)	Requires sample derivatization
	Quantitative (with calibration)	Requires separation
	Modest sample volume (0.1-0.2 mL)	Slow (20-40 min per sample)
	Good sensitivity (LOD = 0.5 μM)	Cannot be used in imaging
	Large body of software and databases for metabolite identification	Not compatible with solids
	Detects most organic and some inorganic molecules	Novel compound identification is difficult
	Excellent separation reproducibility	
	Many spectral features are identifiable	
	Can be mostly automated	
	Compatible with gases and liquids	
LC-MS	Superb sensitivity (LOD = 0.5 nM)	Destructive (sample not recoverable)
	Very flexible technology	Not very quantitative
	Detects most organic and some inorganic molecules	Higher start-up cost (>$300,000)
	Small sample volumes (10-100 μL)	Slow (15-40 min per sample)
	Can be used in metabolite imaging (MALDI or DESI)	Usually requires separation
	Can be done without separation (direct injection)	Poor separation resolution and lower reproducibility versus GC-MS
	Has the potential to detect the largest portion of metabolome	Less-robust instrumentation than NMR or GC-MS
	Can be mostly automated	Most spectral features are not yet identifiable
	Compatible with solids and liquids	Novel compound identification is difficult
		Short instrument lifetime (<9 years)

Adapted from [53], [116]–[122].

2.5 Sample preparation

2.5.1 Sample types

A large range of biofluids, including blood serum, blood plasma, urine, saliva and cerebrospinal fluid are commonly used samples in metabolomic analysis. Apart biofluids, cell, tissue or even organ could be samples used in metabolomic analysis. In our laboratory, blood serum, blood plasma, urine, saliva, cell, and also liver biopsy are commonly used samples. During my thesis work, blood plasma is the main sample we used.

2.5.2 Quality Control (QC) sample preparation

Several good publications are available in the literature which detailed the QC sample preparation [88], [91], [123]. Using the QC samples is a way to assess the quality of the data when performing a Principal Component Analysis (PCA, this will be described in the section 2.7.2.1 Principal Component Analysis). After transformation, scaling and normalization, QC samples should be a tightly clustered group that should be located in the middle of the PCA plot. Apart this, QC samples can also be used to monitor drift, equilibrate the analytical platform, correct the drift of the signal and allow the integration of multiple analytical experiments (Adapted from University of Birmingham and Birmingham Metabolomics Training Centre.).

The ideal QC sample is a pooled QC sample of each biological sample in the study. However, sometimes, due to limited sample amounts or if the study involves large number of samples (such as thousands), then an alternative QC sample should be used. In case of a large sample size (such as more than 500 samples) the QC may be prepared from the first batch of samples collected. In this situation, the recruitment of subjects should be randomized and the samples should be representative of the entire study group. Alternatively, a commercially available QC sample could be used, for example human serum purchased from commercial suppliers. If neither a pooled QC nor a commercial alternative is available for example in samples with low volumes such as tears or bile then a synthetic substitute may be used [91].

Also, preparation of the QCs should follow the same sample procedure used in the preparation of the study biological samples, and the number of freeze-thaw cycles should be concordant between the QC and study biological samples.

The preparation of QC samples can be done by pool samples before or after extraction, depending on the objective to verify. If the objective is trying to investigate all the variation during preparation procedure, then it may be better to pool before extraction, in this case, the reproducibility of the sample preparation technique is the main variation, and this variation can be corrected by adding internal standards in the extraction solvent.

If the purpose is to use the pooled QC to correct for the behavior of metabolites in analytical system, that are not able to be corrected by the internal standard (e.g. metabolites that behave in an opposite pattern to the internal standard), then it is better to pool after extraction and then split into individual aliquots that are run throughout batch sequence at regular interval (such as 1 QC for every 10 samples). In this case, a homogenous identical mixture has created that theoretically should give identical chromatograms, but in fact will reflect any variations in the analytical system (this paragraph is inspired by exchange with Dr. David P. De Souza, Metabolomics Australia, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne).

2.5.3 Sample preparation for NMR based metabolomic analysis

During my thesis work, for NMR based metabolomic analysis, the blood plasma samples were prepared as follow: plasma samples were stocked in the freezer at – 80°C, before analysis, samples were at first thawed on the ice. Then, 250 µL of each plasma sample was added in a new clean 1.5 Eppendorf tube, and mixed with 350 µL of D_2O which contain 10mM Phosphate-Buffered Saline (PBS), the pH of D_2O was adjust to 7.48 at 21.1°C. After that, the mixture was centrifuged at 12 000 RCF, for 10 min at 4°C. Finally, 550 µL of supernatant was transferred into a clean 5mm NMR tube for analysis.

2.5.4 Sample preparation for LC/GC-MS based analysis

Basically, the sample preparation for LC-MS based analysis consists the following steps: remove proteins, metabolites extraction, lyophilize and reconstitution.
Methanol and acetonitrile are commonly used in LC-MS based analysis to remove proteins and for metabolites extraction. During my thesis work, methanol / plasma 4:1 (volume / volume) was used. The mixture (methanol / plasma) was then centrifugated, the supernatant is drawn and lyophilized in order to concentrate extracted metabolites.

Before analysis with LC-MS, the lyophilized supernatant was reconstituted with usually 1:4 (volume / volume) methanol (or acetonitrile) / water mixed solution. For GC-MS studies, a derivatization step is needed by using derivatization reagent to protect active function group.

The detailed sample preparation method for LC-MS based analysis that we used during my thesis work will be described in the **Part Three** PRESENTATION OF THE THESIS WORK, section 3.1 “Protocol for blood plasma sample extraction for metabolomics” and section 3.2 “Protocol for blood plasma sample extraction for lipidomics”.

2.6 Data processing: From raw data to data matrix

In this section, the properties of raw NMR and raw MS data, general preprocessing steps from raw NMR and raw LC-MS data to metabolomic data matrix will be described.

2.6.1 NMR data preprocessing

Raw NMR data, which are a series of intensity values collected as a function of time, thus, it is time domain data, usually with 16,000 or 32,000 entries. The data values are composed of two types, real and imaginary, which reflect the two channels of the NMR receiver. For each time point in the FID, there is a pair of data values and in the order real, imaginary, real, imaginary …, start with FID, the following steps are needed to be performed, during my thesis work, NMRPipe software [124] was used to perform these steps:

2.6.1.1 Apodization

The goal of Apodization is to emphasize the early data (mostly signal) in the FID and de-emphasize the later data (mostly noise). Which is usually achieved by multiply the FID by an exponential decay function, such as: , where LB is line-broadening, which is the additional line-width in Hertz, and LB is usually set to 0.3 (value used during my thesis work) for proton spectra and 1.0 for carbon spectra [125].

2.6.1.2 Zeros filling

Zeros filling consists add zeros to the end of the FID, this operation has no effect on the peak positions, intensities, or linewidths of spectrum, but increase the digital resolution in the spectrum.

2.6.1.3 Fourier transform

The Fourier transform is a mathematical function which converts the time domain data (FID) into a frequency domain spectrum. The Fourier transform extracts from the FID

about the different frequencies of the signal, their intensities, the rate at which they decay, which determines the linewidth of each peak in the spectrum. The signals which decay quickly are transformed into board peaks, while signals which decay during a long time will be transformed into sharp peaks.

2.6.1.4 Phase correction

After Fourier transform FID into frequency domain spectrum, due to imperfections in the RF electronics and variability of samples, it is impossible to start the FID at 0° for all acquisition, so phase correction (or Phasing) is necessary to correct phase errors in order to get absorptive peak shape. Two steps of phase correction may be necessary. The first one is order 0 phase correction. It applies the same phase correction to the entire spectrum and aims to account for any phase shift that may occur independently of the signal frequency. The second one is order 1 phase correction. This time, it applies a phase correction depending on the frequency of the signal. The order 0 phase correction could be sufficient for metabolomic study, it depends on the sequence that is used.

2.6.1.5 Setting the Reference

This step consists of selecting a reference peak (eg. Tetramethylsilane, TMS) and giving a chemical shift value to this reference, without this reference, the chemical shift scale of the spectrum will be approximative.

However, references such as Trimethylsilylpropanoic acid (TSP) or 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) are known to have a certain affinity with proteins such as human serum albumin and therefore cause a variation in the chemical shift of the signals [126].

In absence of added reference, the signal of a metabolite whose chemical shift is known and not sensitive to experimental conditions can be used as reference (eg. the anomeric proton doublet of α-glucose at 5.23 ppm) [127].

Another problem, appearing during the statistical processing of the NMR data, is the absolute and the relative position of an NMR signal can be affected by several chemical and physical factors, for example, changes in the magnetic field, changes in pH, in temperature, a different saline concentration, or different relative concentrations of specific ions, and it is not always possible or desirable to eliminate these effects.

To remove misalignment of NMR signals, several algorithms have been proposed in the literature, such as interval-correlation-shifting (or icoshift) program [128]. The icoshift algorithm derives its name from the basic coshift algorithm [129], [130]. The basic idea is: independently aligns each NMR signal to a target (which can optionally be an actual signal or a synthetic one like the average, or the median) by maximizing the cross-correlation between user-defined intervals. The figure below (Figure 13) illustrates an overview of icoshift results when applied to a misaligned set of human urine NMR spectra zoomed into a strongly misaligned region [128].

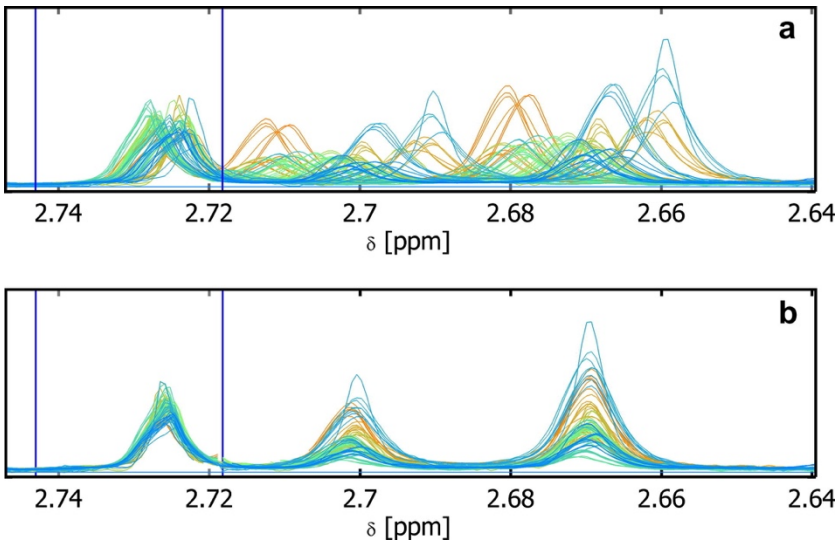


Figure 13. A misaligned set of human urine NMR spectra before (a) and after(b) icoshift [128].

2.6.1.6 Baseline Correction

The baseline is the average of the noise part of the spectrum, ideally, this would be a straight, horizontal line representing zero intensity. In real experiment, it can drift, roll, and wiggle. These errors result for example from erroneous data which are collected at

the very beginning of the FID, when the electronics is still recovering from the shock of the exciting RF pulse. This distortion of the baseline can be corrected by subtracting a polynomial function (here an order 1 polynomial function, which is a straight line) [131].

2.6.1.7 Binning or Bucketing

Binning or Bucketing is an operation to reduce the NMR data dimension (named variables afterwards). In binning, the spectra are divided into bins (called buckets) and the total area within each bin is calculated to represent the original spectrum. The approach consists usually to divide all the spectra with uniform areas width (such as 10^{-3} ppm used frequently for ^{1}H spectrum).

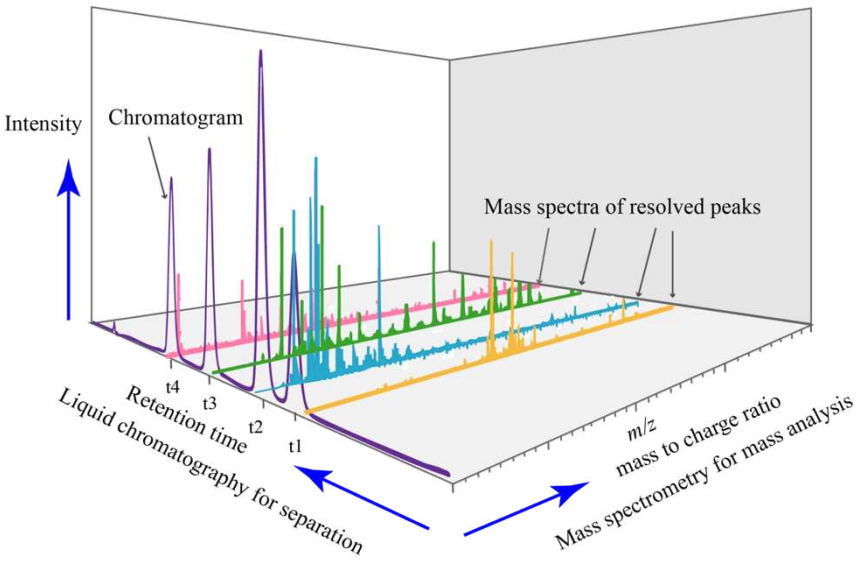
Due to the arbitrary division of peaks, one bin may contain pieces from two or more peaks which may affect the data analysis. Intelligent Binning method [132] was then proposed, these methods attempt to split the spectra so that each area common to all spectra contains the same resonance, eg. belonging to the same metabolite. In such methods, the width of each area is then determined by the maximum difference of chemical shift among all spectra.

2.6.1.8 Exclusion of spectral regions

Spectrum regions that do not contain information on metabolites and are likely to introduce artifacts for statistical analysis are preferably excluded. Thus, spectral regions outside of 0 to 10 ppm are generally removed in metabolomics studies. Another part of the spectrum corresponding to the resonances of solvent signals such as those of water between 4.6 ppm and 5 ppm can also be excluded, because water signal is very strong in NMR spectrum, can lead to a very important variability, and have impact in further statistical analysis.

2.6.2 LC-MS data preprocessing

In LC-MS, the raw data consists of a set of chromatograms (1 for each sample). Each chromatogram representing the intensity of the total ion current as a function of the retention time, there is in fact a third dimension which corresponds to the resolution in mass to charge ration (m/z) of ions detected for each spectrum (Figure 14). Compared to the NMR, the data processing step is even more complex by the presence of a separation technique that brings an extra dimension, a large amount of background noise, artifacts and redundancy information.



© Daniel Norena-Caro, Wikipedia

Figure 14. Schematic representation in 3 dimensions of a LC-MS chromatogram.

The processing of LC-MS based metabolomics data consists therefore in extracting the analytically relevant signals (m/z-retention time: tr, intensity) from the raw data for each sample, then realigning in the temporal (or chromatographic) and spectral domains. The detailed MS data processing process has been described in these publications [133]–[138].

Fundamentally, the main steps are: detect masses from mass spectra, construct extracted-ion chromatogram (EICs), detect chromatographic peaks, their alignment and integration of peaks. Specific software and free workflow are available to complete these

operations, which may cover the process from raw LC-MS data into data matrix, such as Thermo Scientific™ Compound Discoverer™ software (also preform ion feature identification), Workflow4Metabolomics [139].

2.6.3 Common steps for NMR data and LC-MS data preprocessing

After preprocessing, raw data (NMR data or LC-MS data) were converted into the data matrix, which is represented in the Table below (Table 5).

Table 5. Schematic representation of a typical metabolomic data matrix.

	Group	Variable1	Variable2	Variable3	Variable4	Variable5	Variable6	Variable7	Variable8	Variable9	Variable10	...
Sample1	0	4.19E+08	9.96E+08	6.19E+08	4.53E+07	3.50E+08	1.20E+07	9.46E+08	3.33E+08	4.74E+08	2.87E+08	...
Sample2	1	8.82E+08	6.84E+08	1.14E+08	5.81E+08	6.29E+08	2.75E+08	9.88E+07	4.36E+08	3.33E+08	5.39E+08	...
Sample3	0	6.39E+08	4.62E+06	3.64E+08	3.86E+08	6.29E+08	9.87E+08	7.38E+08	1.28E+08	NaN	1.69E+07	...
Sample4	1	5.32E+08	7.29E+08	6.55E+07	6.43E+08	9.81E+08	7.65E+08	7.01E+08	4.21E+08	5.23E+08	7.20E+08	...
Sample5	0	4.24E+08	2.55E+08	9.43E+08	7.43E+08	3.10E+08	1.72E+08	6.40E+08	9.79E+08	1.24E+08	5.14E+07	...
Sample6	1	5.75E+08	6.59E+08	6.53E+08	5.20E+08	4.86E+08	5.76E+08	9.70E+08	6.82E+08	6.69E+08	7.13E+08	...
Sample7	1	1.13E+08	5.26E+08	7.13E+08	4.05E+08	9.22E+08	8.38E+08	6.63E+08	4.20E+08	7.07E+08	5.33E+07	...
Sample8	0	7.61E+08	NaN	5.88E+08	7.42E+08	1.65E+08	9.09E+08	1.14E+08	7.31E+08	8.80E+07	3.61E+08	...
Sample9	1	4.08E+08	9.54E+08	6.60E+08	3.05E+08	5.90E+08	7.75E+07	3.49E+08	7.75E+08	7.73E+08	2.36E+08	...
Sample10	0	9.55E+08	2.91E+08	4.79E+08	1.16E+06	5.36E+08	4.75E+08	4.09E+07	2.88E+08	4.19E+08	5.43E+08	...
Sample11	1	9.72E+08	3.95E+08	3.41E+08	4.29E+08	2.41E+08	6.11E+08	1.28E+08	6.88E+07	3.84E+08	3.00E+08	...
Sample12	0	3.60E+08	7.87E+08	1.89E+08	9.76E+07	7.13E+08	9.44E+08	7.54E+08	5.58E+08	5.43E+08	7.40E+08	...
Sample13	1	9.84E+08	8.53E+08	8.52E+08	1.49E+08	5.51E+08	1.01E+08	9.38E+08	7.34E+08	3.46E+08	3.20E+08	...
Sample14	0	7.19E+08	7.15E+08	3.37E+06	8.71E+08	NaN	1.55E+08	5.06E+08	6.53E+07	4.74E+08	7.21E+08	...
Sample15	0	4.38E+08	7.18E+08	9.81E+08	2.20E+08	7.41E+08	3.80E+08	3.60E+08	7.13E+08	3.84E+08	8.81E+08	...
Sample16	1	9.04E+08	6.77E+08	5.07E+08	1.17E+08	7.19E+08	4.74E+08	4.19E+08	2.29E+08	5.36E+08	5.00E+08	...
Sample17	0	2.46E+07	2.66E+08	9.33E+08	5.98E+08	1.11E+08	9.76E+06	6.75E+08	3.23E+08	3.10E+08	7.33E+08	...
Sample18	1	1.27E+08	1.94E+08	7.75E+08	5.54E+08	4.36E+08	7.94E+08	1.51E+08	2.09E+08	3.59E+08	4.85E+08	...
Sample19	0	5.04E+08	6.71E+08	3.43E+08	9.79E+08	9.99E+08	5.92E+08	6.91E+08	5.56E+08	3.15E+08	2.09E+08	...
Sample20	1	1.89E+07	8.39E+08	1.36E+08	2.63E+08	8.03E+06	6.04E+08	2.41E+08	8.88E+08	2.50E+08	4.53E+08	...
...	...	...	...	...	...	...	...	...	...	...	...	...

"Variables" in column correspond to detected mass-TR for LC-MS data or spectral bin for NMR data, observation in row correspond to each analyzed samples and responses, the value in the matrix represents detected LC-MS peak area or NMR spectral bin intensities, here, missing value in response were replaced by "NaN" (Not a Number). Group information was coded by 0 (control group) or 1 (disease group).

Basically, before further statistical analysis, the following operations (such as Missing value imputation, Normalization, Transformation and Scaling) on the metabolomic data matrix are needed to be performed.

2.6.3.1 Missing value imputation

The presence of missing values in metabolomics data occur widely and can originate from a number of sources, including for both technical and biological reasons: (1) metabolite is detected in one sample but is not present at any concentration in another sample; (2) metabolite is present in a sample but at a concentration less than the

analytical method's limit of detection, and (3) metabolite is present in a sample at a concentration greater than the analytical method's limit of detection but the data processing software has not detected it and has not reported the metabolite [140], [141].

Fundamentally, a feature (variable) with more than 20% of missing value will be excluded for further analysis, also, a lot of software such as SIMCA® (Umetrics®) tolerate portion of missing value, thus, missing value is not a major concern in metabolomic data preprocessing.

During my thesis works, for LC-MS data matrix, K-nearest neighbour imputation (KNN) was used as the missing value imputation method. Briefly, the missing values are replaced by the average of the corresponding (feature specific) non-missing values in the k (here k = 10) closest features in terms of Euclidean distance of the responses across all the samples. Therefore, a unique value is imputed for every missing value in a feature instead of using the same value multiple times [140], [142]. A detailed Missing value imputation method was described in the publication Di Guida et al. 2016 [141].

2.6.3.2 Normalization: Integral, Quotient, Quantile

The objective of normalization is to conserve the maximum biologically variation and minimums errors during sample preparation and data acquisition. Normalization is not absolutely necessary, but in some case, it is crucial, especially for biological fluids.

Normalization is in particular important for urine samples, as metabolites concentration in urine are basically negative correlated with water intake quantity, animal drinking different quantity of water will result different concentration in metabolite among different individual, normalization try to reduce the variation of metabolites concentration which is not biological interesting (this is named the dilution effect).

Several metabolomic data normalization methods have been proposed in the literature. In our laboratory, Integral, Quotient and Quantile Normalization are the most frequently methods used.

2.6.3.2.1 Integral Normalization

Each response (peak intensity or peak area) in a sample is divided by the total sum of the sample and multiplied by a constant (the appropriate constant defined by the response) to restore the original response form (called “smooth”) [143].

2.6.3.2.2 Probabilistic Quotient Normalization

The probabilistic quotient normalization, which is introduced by Dieterle et al. [144]. This method is based on the calculation of a most probable dilution factor (such as median of quotients) by looking at the distribution of the quotients of the amplitudes of a test spectrum by those of a reference spectrum. The reference spectrum can be a QC sample or the median spectrum.

2.6.3.2.3 Quantile Normalization

It was initially developed for gene expression microarrays [145], [146] but today it is applied in a wide-range of data types. Quantile normalization is a nonlinear transformation that replaces each feature response (row) with the mean of the features across all the samples with the same rank or quantile. A schematic of quantile normalization is shown in Figure 15.

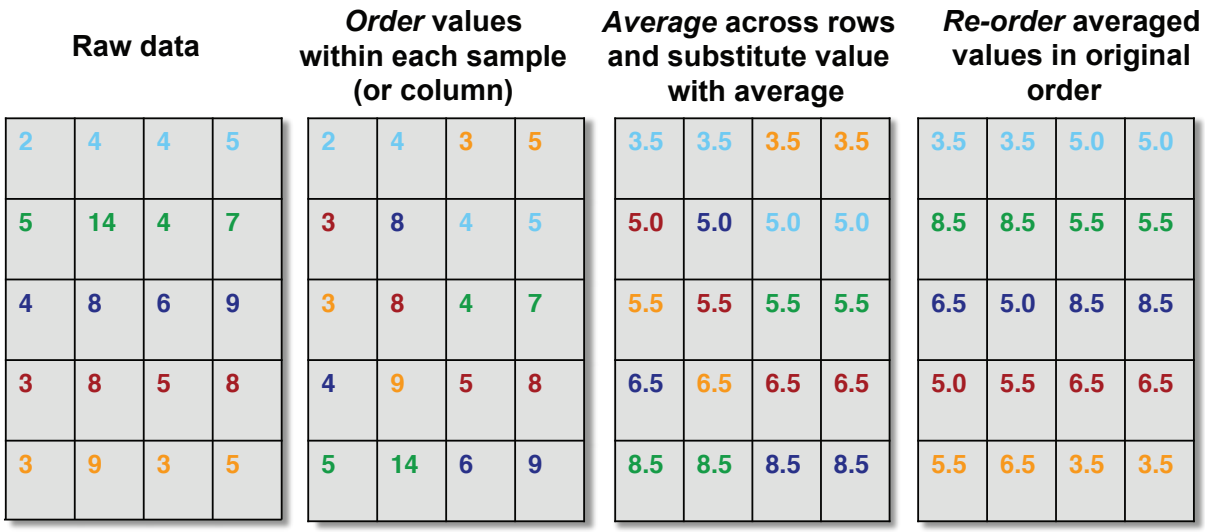


Figure 15. A schematic of quantile normalization (Stephanie C. et al. 2014) [147].

(1) order the feature values within each sample (2) for each feature, average across the rows (3) substitute the raw feature value with the average (4) re-order the transformed values by placing in the original order.

2.6.3.3 Transformation

Transformations are generally applied to correct for heteroscedasticity, which is the situation that the variability of a variable is unequal across the range of values of the variable that predicts it [148], to convert multiplicative relations into additive relations, and to make skewed distributions more symmetric [149]. The log transformation is the commonly used method; however, it is unable to deal with the zero values. Thus, generalized logarithm (glog) transformation is proposed, which is a simple variation of ordinary log in order to deal with zero or negative values in the data set [150]. Its formula is:

$$\mathrm{glog}_2(x) = \log_2 \frac{x + \sqrt{x^2 + a^2}}{2} \qquad \text{(Eq. 0.6)}$$

where *a* is a constant with a default value of 1.

2.6.3.4 Scaling: Mean-centering, UV scaling, Pareto scaling

The objective of “scaling” or “weighting” was to give all variables more reasonable weight (importance) in the modelling. This is especially important in cases where the variables being compared have different response units, such as measure of body weight and body height [151].

For NMR and mass spectrometric data, variables correspond to peak intensities are areas and hence have the same units, so scaling is not absolutely essential, but is still usually useful. If the original data is not scaled, the variables with the largest variance will tend to dominate the Principal Component Analysis (PCA). For example, a very large variable which is approximately constant for all samples may dominate the first component of PCA (PC1).

Several data scaling methods have been proposed in the literature, the Unit Variance (UV) scaling and the Pareto scaling (Par) are commonly used methods in our laboratory.

Table 6. Overview of common centering, scaling and transformation methods. In the Unit column, *O* represents the original Unit, and (-) presents dimensionless data.

Class	Method	Formula	Unit	Goal	Advantages	Disadvantages
I	Centering	$\tilde{x}_{ij} = x_{ij} - \overline{x}_i$	*O*	Focus on the differences and not the similarities in the data	Remove the offset from the data	When data is heteroscedastic, the effect of this pretreatment method is not always sufficient
II	Autoscaling	$\tilde{x}_{ij} = \frac{x_{ij} - \overline{x}_i}{s_i}$	(-)	Compare metabolites based on correlations	All metabolites become equally important	Inflation of the measurement errors
	Pareto scaling	$\tilde{x}_{ij} = \frac{x_{ij} - \overline{x}_i}{\sqrt{s_i}}$	*O*	Reduce the relative importance of large values, but keep data structure partially intact	Stays closer to the original measurement than autoscaling	Sensitive to large fold changes
III	Log transformation	$\tilde{x}_{ij} = {}^{10}\log\left(x_{ij}\right)$ $\hat{x}_{ij} = \tilde{x}_{ij} - \overline{\tilde{x}}_i$	Log *O*	Correct for heteroscedasticity, pseudo scaling. Make multiplicative models additive	Reduce heteroscedasticity, multiplicative effects become additive	Difficulties with values with large relative standard deviation and zeros

Adapted from van den Berg et al. 2006 [149].

Comparing the UV scaling with the Pareto scaling, UV scaling gives all variables (signal or noise) equal weight (or importance) in the modeling, however, Pareto scaling gives important variables (mostly signal) more weight than noise, as a result, QC samples will be more closely in PCA scores plot after Pareto scaling than UV scaling, that is why the Pareto scaling is the preferred method in most of cases.

2.7 Statistical analysis and Interpretation

2.7.1 Univariate analysis

For two groups study (e.g. health control and disease), to compare each variable in the data, the student's *t*-test which relies on the comparison of the two-sample means is commonly used. It requires:
1, each sample population have a normal distribution;
2, there should be an equal variance of the two populations;
3, the data is independently sampled.
In case where two sample populations have unequal variances, Welch's *t*-test, or unequal variances *t*-test can be applied. For data contains more than two groups, one-way analysis of variance (ANOVA) can be utilized to test the difference between group means of each variable.

When the assumption of normal distribution is not met for the sample population, nonparametric analysis such as Mann-Whitney U test (also called Wilcoxon rank-sum test) can be used to test the difference between two independent samples by comparing their medians. Kruskal-Wallis test is the nonparametric equivalent of ANOVA for comparing data with multiple groups.

2.7.2 Multivariate analysis

To account for the impact of multiple variables (combination of two or more variables) on the outcome of measurement (e.g. health control or disease), or study several variables at one time, multivariate analysis is required. Several multivariate analysis methods have been proposed in the literature, Principal Component Analysis (PCA) and Partial Least Squares Projection to Latent Structures (PLS) are commonly used multivariate methods in metabolomics, as these models have better interpretability than other multivariate methods.

2.7.2.1 Principal Component Analysis (PCA)

Principal Component Analysis, or PCA, is a dimensionality-reduction method, the idea of PCA is to reduce the number of variables of a data set, while preserving as much information as possible. This is performed by transforming a large set of variables (highly dimension) into a new smaller set of uncorrelated variables (lower dimension), also called Principal Components (PC), that still contains most of the information in the large set. Principal components are new variables that are constructed as linear combinations or mixtures of the initial variables.

Mathematically, PCA model shows the correlation structure of the data matrix X, approximating it by a matrix product of lower dimension (TP'), called the Principal Components plus a matrix of residuals (E).

$$X = X_{bar} + TP' + E \qquad \text{(Eq. 0.7)}$$

Where

X_{bar} contains X average.

T is a matrix of scores that summarizes the X-variables.

P is a matrix of loadings showing the influence of the variables.

E is a matrix of residuals, the deviations between the original values and the projections.

This geometrically corresponds to fitting a line, plane or hyper plane to the data in the multidimensional space with the variables as axes [152]. The scaling of the variables specifies the length and also the direction of the axes of this space.

Scores plot is generated to assess the clustering of different samples, with the corresponding loadings plot demonstrating the variables accounting for the most variation in the specified principal component.

As an unsupervised analysis method, that means without the prior knowledge of the sample classification in the model building, PCA is particularly useful in the first step in metabolomic data analysis to identify how one sample is different from another, which variables contribute most to this difference and whether those variables contribute in the same way (e.g. are correlated) or independently (e.g. uncorrelated) from each other.

A more detailed description of PCA can be found in the book of Jackson, J.E. (1991) [153].

2.7.2.2 Partial Least Squares Projection to Latent Structures (PLS)

Partial Least Squares Projection to Latent Structures, or PLS, is a supervised analysis method, in contrast with unsupervised method, information of sample class labels (e.g. health control or disease) are also used in the statistic models building. PLS finds the linear relationship between a matrix Y (dependent variables) and a matrix X (predictor variables), expressed as:

$$Y = f(X) + E \qquad \text{(Eq. 0.8)}$$

Where “E” is a matrix of residuals, the deviations between the original values and the projections [152].

The PLS used in metabolomics is usually its discriminant version, called Partial Least Square-Discriminant Analysis (PLS-DA), which is a variant of the PLS regression that allows to build a model that maximizes the separation between the classes to which the samples belong. PLS-DA has the advantages of PLS: it can manage a large number of X variables, manage multicollinearities and missing data [154]. The difference between the PLS and the PLS-DA is based on the nature of the Y variables, for a classical PLS the Y are quantitative variables, for the PLS-DA the Y variables may be qualitative or categorical.

PLS, as well as PLS-DA allows the construction of an explanatory model. This model thus makes it possible to highlight metabolites (variables) whose intensity is characteristic of a given biological state and which contribute to the separation of the different groups. PLS uses variable importance to projection (VIP) scores to demonstrate the contribution of each variable in the PLS model, a metabolite with VIP score > 1 considered an important variable in classification as the average VIP scores is 1. A more detailed description of PLS can be found in Wold et al, 2001 [155].

2.7.2.3 Orthogonal PLS modeling (OPLS)

The OPLS is a modification of PLS model, the difference between PLS and OPLS is in their handling of the variance of the X matrix. PLS separates the variability in X into two parts (Figure 16), the systematic and residual parts. The systematic part is the sum of the variability in X that is linearly related to Y (predictive part) and the variability in X that is uncorrelated to Y (orthogonal part). Only the variation related to Y is used to model Y [151]. The OPLS can, like PLS-DA, be used for discrimination purposes (OPLS-DA).

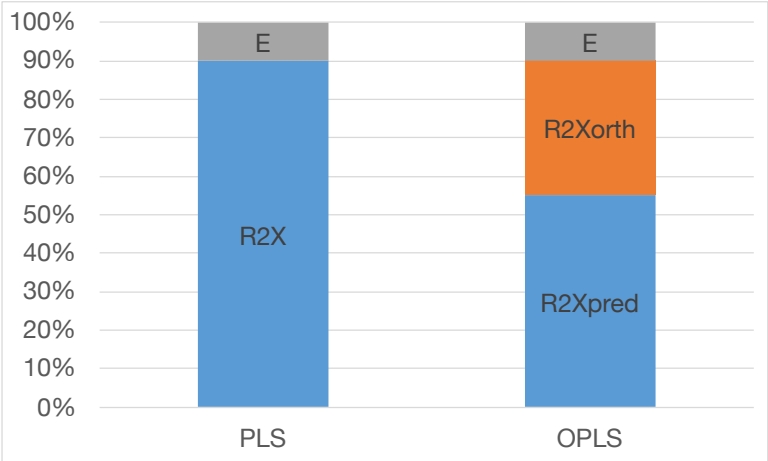


Figure 16. PLS (left) separates the variability in the X matrix in two parts, the systematic variability (R2X) and the residual variability (E). OPLS (right) further splits the systematic variability, R2X, in two parts, the part that is linearly related to Y (predictive, R2Xpred) and the part that is uncorrelated to Y (orthogonal, R2Xorth) [151].

PLS divides the sum of squares of X in two parts, OPLS divides it in three parts. Also, within group and between group variations are separated on both components in OPLS-DA while they were not in PLS-DA (Figure 16), thus, those greatly facilitates the interpretation of the OPLS-DA model. However, OPLS-DA provides no predictive advantage over PLS-DA [156]–[160].

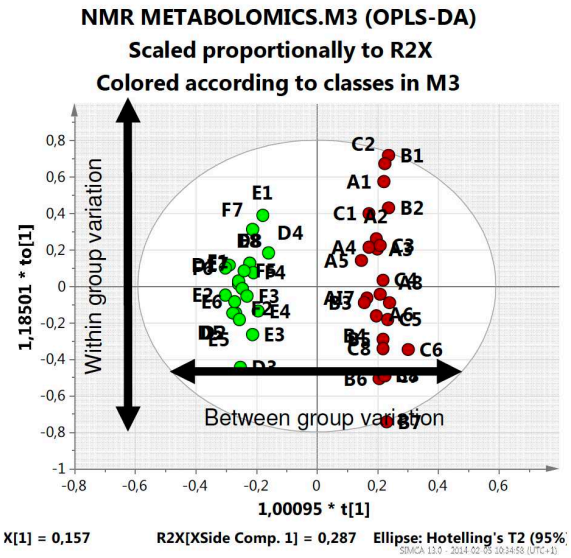


Figure 17. Score plot of an OPLS-DA model, within group and between group variations are separated on both components [151].

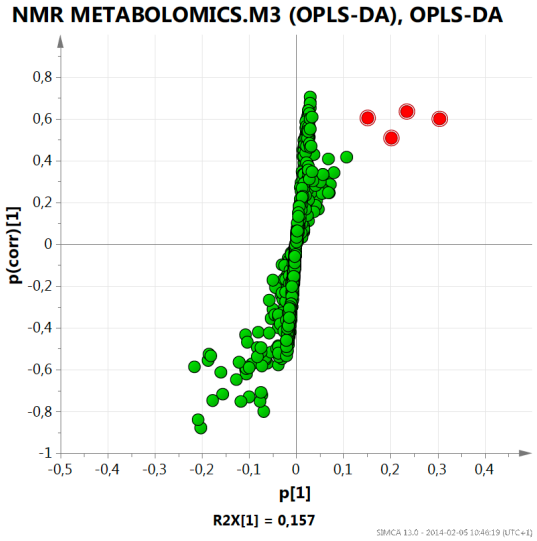


Figure 18. S-plots of an OPLS-DA model, p(corr) indicates the reliability of a variable as a marker while the loading, p, indicates the influence of the variables in the model [151].

The loading and S-plots [161] are usually used to identify what is different between classes. The S-plot (Figure 18) is one of methods to highlight putative biomarkers from a two group OPLS-DA model with NMR, MS based metabolomics data. Compare with loading plots, S-plot adds another dimension to the loading plot by also providing the

p(corr) value. This value indicates the reliability of a variable as a marker while the loading, p, indicates the influence of the variables in the model [151].

It should be noted that in the single-Y case (e.g. 1 health control and 1 disease groups), by theory, the OPLS model can only have one predictive component [162]. However, with multiple Y-variables there can be more than one predictive OPLS component (the case of O2PLS).

2.7.2.4 Model diagnosis and validation

Several diagnostic statistics approaches are currently employed in the optimization and the assessment of performance of PLS-DA models in metabolomics data analyses, such as cross-validated explained variation Q2 and the Area Under the Curve of a Receiver Operating Characteristic (AU-ROC) analysis.

The Q2 estimates the predictability of the model, which means the ability of model to correctly class a new set of data, the more the value of Q2 close to 1 the better the predictability of the model [163].

The AU-ROC is equal to the probability that the classifier will score a randomly drawn positive sample higher than a randomly drawn negative sample. In fact, AU-ROC and Mann-Whitney U test are closely related. From the perspective of Mann–Whitney U statistic, AU-ROC can be explicated in this way, given 2 classes (0 and 1), randomly select one sample from class 1, randomly select the other sample from class 0, and then predict these two random samples with the classification model. The probability of predicting 1 as class 1 is p1, the probability of predicting 0 as class 1 is p0, and the AU-ROC is equal to the probability of $p1 > p0$.

So, the AUC reflects the sorting ability of the classification model for the sample. According to this explanation, if we classify the samples completely randomly, then the AUC should be close to 0.5. It is also worth noting that AUC is not sensitive to the consistency of sample categories, which is the reason why AU-ROC is usually use to evaluate classifier performance for unbalanced samples [163]–[166]. Also, for a two groups classification problem, the Area Under the Receiver Operating characteristic

Curve (AU-ROC) will be a better estimator than Q2 to access OPLS-DA model performance.

To validate outcomes of the multivariate models such as PLS-DA model, Confusion Matrix, Cross Validation (CV) and permutation test are commonly used Internal Validation techniques.

The confusion matrix is a table that lists the correct and false predictions versus the actual observations, which is usually used to describe the performance of a classification model (classifier) on a new set of test data for which the true values (such as groups information: health, disease…) are known previous. It allows easily identify confusion or error (like one class is misclassed as the other) of classification model when it makes predictions.

Also, it allows to calculate several insightful performance metrics, such as Accuracy: how often is the classifier correct, defined as: "correct predictions / total predictions", calculated by "(TP + TN) / (TP + TN + FN + FP)", with:

True Positive (TP): Observation is positive, and is predicted as positive (eg. they do have the disease, and predicted as disease).
False Negative (FN): Observation is positive, but is predicted as negative.
True Negative (TN): Observation is negative, and is predicted as negative.
False Positive (FP): Observation is negative, but is predicted as positive.

A more detailed description of confusion matrix can be found in the book "Fundamentals of Clinical Data Science, Chapter 8 Prediction Modeling Methodology" [167].

The basic of Cross Validation (CV) involves splitting the data into a training set and a test set. The training set is used to build the classification models (e.g. PLS-DA model), and the model is then applied to predict the outcome of the test set, the process will repeat several times until all subjects will predicted. Leave-one-out CV and k-fold CV are commonly used method for Cross Validation (7-fold CV is the default method used in SIMCA® Umetrics®).

A permutation test can assess whether the classification based on true sample class is significantly better than classification based on randomly assigned sample class. The principle of permutation test is to compare model outcomes between the classification based on true sample class (really Y) and the classification based on randomly assigned sample class (randomly assigned Y).

2.7.3 Biological interpretation

Biological interpretation increases the information generated by metabolomic, and exploit the relational properties present in metabolomic data by analyzing metabolite patterns from an integrative point of view [90].

In general, after identified important variable or metabolite of interest, the next step is trying to integrate the metabolite of interest into biological network especially metabolic pathway, which reveals metabolites changes (or perturbations) in biological network during diseases, physiological processes or external stimulus. And the possible enzymes controlling the metabolite levels in the cell could be then investigated, by testing its impact further on the metabolite level may promote understanding of biological mechanisms associated with the specific disease. Thus, it may improve our understanding of biological etiology of specific diseases, and providing insight further in the development of targeted treatment methods [9], [168], [169].

Metabolic pathway can be assessed using biological databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [170], Small Molecule Pathway DataBase (SMPDB) [171], EHMN [172], WikiPathways [173], and MetaCyc [174], these databases provide exhaustive information of a large number of metabolic pathways.

Part Three PRESENTATION OF THE THESIS WORK

In order to prepare blood plasma samples for LC-MS analysis, two protocols, with one for blood plasma sample extraction for metabolomics and another for blood plasma sample extraction for lipidomics, were adapted in consideration of the equipment, reagents and chemicals availability in the CSPBAT laboratory.

3.1 Protocol for blood plasma sample extraction for metabolomics

The present Standard Operating Procedure (SOP) applies to blood plasma sample extraction and protein precipitation for LC-MS based global metabolomics analysis. This SOP was edited by Xiangping LIN, reviewed by Zhicheng LIU, Philippe SAVARIN and Xinyu LIU (Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences) in 10 Oct. 2017.

I) Protein Precipitation & Metabolites extraction

MATERIELS:

- Reagents: Blood Plasma, LC-MS grade Methanol, Internal Standard as listed
- Equipment: Pipettes and Pipette Tips (100, 200, 1000 μL), Eppendorf (tube1.5, 2mL), Centrifuge tube (15mL, 50mL), 500mL glass bottle, gloves, protective goggles, timer, Fume hood, centrifuge and lyophilizer.

Table 7. List and concentration of used internal standard (IS).

Stable isotope labeled IS	final.Con(ug/mL)	v methanol (mL)	Total (mg)
Carnitine C8-d3	0,1	**44,4** of 465 mL	0,047
LPC 19:0	0,75		0,349
Carnitine C16:0-d3	0,15		0,070
FFA C16:0-d3	2,5		1,180
FFA C18:0-d3	2,5		1,180
CA-d4	1,854		0,862
CDCA-d4	1,485		0,691
Phe-d5	3,6125		1,680
Trp-d5	4,25		1,976

PROCEDURE:

1) Preparation of methanol solvent containing internal standard:

- prepare stock solution of all internal standard as in the Table 1 with pure LC-MS grade Methanol to 1mg/mL or other concentration and stock in 4°C.
- add adequate quantity of IS stock solution into pure LC-MS grade Methanol to have desired concentration and volume as in the table1.

2) Preparation of plasma before freeze-drying, NAFLD samples:

Divided operation into 2 days, with 55 samples/day QC included, calculate the total number of QC samples n (n >= N/10 + 20, N: the total number of all plasma samples, 82, 1QC per 10 samples, about 30 QC)

3 a) Preparation of QC and plasma for NAFLD samples:

3 a1)- Day 1, randomize sample order and balance case and control then thaw all samples at room temperature, mix on ice 50 µL of each (81x50µL= 4050 then add excluded samples to yield 5000 µL) into a 15mL Centrifuge tube, vortex mix 60s to ensure mixing

3 a2) - Aliquot of 100 µL mix into 30 new Labeled 1.5 mL Ep tubes (aliquot of 40 µL mix into 30 new Labeled 2mL Ep tubes for lipidomics)

3 a3)- (take 40 samples and 15 QC, then stock the rest into -80°C) For day 2, take 41 samples and 15 QC, thaw at room temperature, transfer 100 µL of plasma or QC sample into new labeled 1.5 mL Ep tubes and add 400 µL methanol solvent. Vortex mix 60s to ensure mixing

3 a4)- Centrifuge @ 16 000xg at 4°C for 10 min, then transfer 2 times 200 µL of the supernatant into 2 new Labeled 1.5 mL Ep tubes (one for +, one for -)

3 b) Lyophilize QC and plasma samples

- Lyophilize QC and plasma samples then stock into -80°C

REFERENCES: [91], [175]–[177]

3.2 Protocol for blood plasma sample extraction for lipidomics

The present SOP applies to blood plasma sample extraction and protein precipitation for LC-MS based lipidomic analysis.

I) Protein Precipitation & Metabolites extraction

MATERIELS:

- Reagents: Blood Plasma, LC-MS grade Methanol, milli-Q water, MTBE, Internal Standard (IS)
- Equipment: Pipettes and Pipette Tips (100, 200, 1000 µL), Eppendorf (tube 2mL), Centrifuge tube (15mL, 50mL), 500mL glass bottle, gloves, protective goggles, timer, Fume hood, centrifuge and lyophilizer.

Table 8. list and concentration of internal standard (IS).

internal standard (ID)	MeOH(ml) 300ul/sample	final conc. ug/ml in MeOH	Total ug	Source
PC(19:0/19:0) chlor	36	0,67	24,12	Avanti Lipids Polar
LPC19:0 (lipidom) chlor	36	0,33	11,88	Avanti Lipids Polar
SM12:0 (15,9ug) etha	36	0,17	6,12	Avanti Lipids Polar
Cer17:0	36	0,17	6,12	Avanti Lipids Polar
FFA C16:0-d3 (for lipidom)	36	0,67	24,12	cdnisotopes
FFA C18:0-d3 (lipidom)	36	0,67	24,12	Cambridge Isotope
TG45:0	36	0,53	19,08	Sigma-Aldrich

PROCEDURE:

1) Preparation of methanol solvent containing internal standard:

- prepare stock solution of all internal standard as in the table1 with pure LC-MS grade Methanol to 1mg/mL or other concentration and stock in 4°C.
- add adequate quantity of IS stock solution into pure LC-MS grade Methanol to have desired concentration and volume as in the table1.

2) Preparation of plasma before freeze-drying for NAFLD samples:

Divided operation into 2 days, with 55 samples/day QC included, calculate the total number of QC samples n (n >= N/10 + 20, N: the total number of all plasma samples, about 81, 1QC per 10 samples, about 30 QC)

3 a) Preparation of QC and plasma, NAFLD samples:

3 a1)- Day 1, randomize sample order and balance case and control then thaw all samples at room temperature, mix on ice 50 µL of each (81x50µL= 4050 then add excluded samples to yield 5000 µL) into a 15mL Centrifuge tube, vortex mix 60s to ensure mixing

3 a2) - Aliquot of 100 µL mix into 30 new Labeled 1.5 mL Ep tubes (aliquot of 40 µL mix into 30 new Labeled 2mL Ep tubes for lipidomics)

3 a3)- (take 40 samples and 15 QC, then stock the rest into -80°C) For day 2, take 41 samples and 15 QC, thaw at room temperature, transfer 40 µL of plasma or QC sample into new labeled 2mL Ep tubes and add 300 µL methanol solvent. Vortex mix 30s to ensure mixing

Add under Fume hood 1mL of MTBE into each tube, and shaken at room temperature with an oscillator for 1 hour, then add 250 µL milli-Q water

3 a4)- Centrifuge @ 10 621xg at 4°C for 10 min, then transfer 2 times 400 µL of the supernatant into 2 new Labeled 2mL Ep tubes (one for +, one for -)

3 b) Lyophilize QC and plasma samples
- Lyophilize QC and plasma samples then stock into -80°C

REFERENCES: [91], [175]–[177]

3.3 NMR and UPLC-HRMS based plasma metabolomic and lipidomic profiling in NonAlcoholic Fatty Liver Disease

Nota bene: this document is in preparation for submitting as a research article, the name listed below is a primary list for people who mainly involved in this work.

Xiangping LIN[1,2,3], Xinyu LIU[3], Mohamed N. TRIBA[1,2], Nadia BOUCHEMAL[1,2], Zhicheng LIU[4], Tony PALAMA[1,2], Laurence LE MOYEC[5], Marianne ZIOL[8], Nada HELMY[7], Corinne VONS[7], Guowang XU[3,*], Carina PRIP-BUUS[6,*], Philippe SAVARIN[1,2,*]

1. Paris 13 University, Sorbonne Paris Cité University, Paris, France.
2. Chemistry Structures Properties of Biomaterials and Therapeutic Agents (CSPBAT Laboratory), The National Center for Scientific Research (CNRS) 7244, 93017 Bobigny Cedex, France.
3. CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China.
4. School of pharmacy, Anhui medical university, Hefei, China.
5. University Paris Saclay, Univ Evry, UBIAE, EA7362, Evry, France.
6. Institut Cochin, INSERM U1016, CNRS UMR8104, Université Paris Descartes, 75014 Paris, France.
7. Department of Digestive and Metabolic Surgery, Jean Verdier Hospital, Paris XIII University-University Hospitals of Paris Seine Saint-Denis, 93140, Bondy, Paris, France.
8. Department of pathology, University Hospital Jean Verdier, Assistance Publique-Hôpitaux de Paris, Bondy, France
* To whom correspondence should be addressed.

3.3.1 INTRODUCTION

Non-Alcoholic Fatty Liver Disease (NAFLD) is defined as the presence of steatosis in at least 5% of hepatocytes on liver biopsy assessment or on imaging in patients who have a history of little alcohol consumption (limited daily alcohol intake < 20 g for women and < 30 g for men) or no alcohol consumption at all and have no other cause of hepatic steatosis [178]–[181].

With increasing metabolic diseases (obesity, diabetes …) rates, NAFLD has emerged as a leading global cause of chronic liver disease with the prevalence of more than 25% in the global adult population in the past few decades [58], [59].

Despite growing prevalence, the factors involved in NAFLD development and subsequent progression to non-alcoholic steatohepatitis (NASH), liver fibrosis, cirrhosis and hepatocellular carcinoma are poorly understood, however, it is well considered that the pathogenesis of NAFLD is multifactorial, and the main risk factors are such as genetic predisposition (e.g., polymorphisms of patatine-like phospholipase domain-containing protein 3 (PNPLA3) gene) [182], dietary factors (e.g., fructose), Insulin Resistance (IR) [183], obesity, type II diabetes mellitus, hyperlipidemia, endocrine disruptors [184] and the gut microbiota dysbiosis [185], [186].

NAFLD is characterized by an abnormal accumulation of lipids mainly triacylglycerols (TGs) in the liver, based on clinical-histologic characters, NAFLD spectrum range from simple fatty liver (NAFL) or steatosis to the advanced form termed NASH, without therapeutic intervention, a subset of patient with NASH will subsequently progress towards cirrhosis and, ultimately, hepatocellular carcinoma [60].

So far, liver biopsy is the gold standard for diagnosis, staging and monitoring progression of NAFLD during treatments. However, biopsy has well-known limitations, such as invasiveness, poor acceptability by patients, sampling variability, and financial cost… which limit its application in a large population. Moreover, recently developed Noninvasive imaging biomarker assessment method, even the most accurate noninvasive liver elastography based methods, such as vibration-controlled transient elastography (VCTE), magnetic resonance elastography (MRE), shear-wave elastography

and acoustic radiation force impulse have other limits including couldn't access inflammation, with very limited guidance (or even unavailable) for how clinicians should anticipate and manage the pitfalls of these tests [61].

Thus, the development of an alternative noninvasive and familiar for clinicians' strategy such as using non-invasive biomarkers is an urgent need [58] for prognostication, staging, selection of patients for treatment and monitoring of the disease.

Previous studies found that plasma metabolome was a better predictor for steatosis (80%) than noninvasive basal clinical data (predictive power of 58%) [187], [188]. Moreover, in our primary NMR study, lipids may be major patterns which could discriminate NASH patients from NL patients.

The objectives of our study were (1) describe the relative plasma metabolome and lipidome changes in Nonalcoholic fatty liver (NAFL) and in Non-Alcoholic SteatoHepatitis (NASH) compared with Normal Liver (NL) obese patients, (2) investigated whether Ultra-Performance Liquid Chromatography (UPLC) coupled with High-resolution mass spectrometry (HRMS) based plasma metabolomics and lipidomics analysis could help to identify potentials biomarkers, if any, associated with different stages of NAFLD (NAFL, NASH), and (3) identify metabolomic or lipidomic patterns which could discriminate NAFL and NASH from NL obese patients, by using appropriate statistical models.

3.3.2 MATERIAL AND METHODS

3.3.2.1 Reagents and chemicals

For metabolomics, HPLC grade solvent (acetonitrile, methanol) and formic acid were purchased from Merck KGaA (Darmstadt, Germany). Ultrapure water (H_2O) was collected from a Milli-Q system (Millipore, Billerica, MA). Internal standards (ISs) Carnitine C8:0-d3, Carnitine C16:0-d3, FFA C18:0-d3, CA-d4, CDCA-d5, Phe-d5 and Trp-d5 were purchased from Cambridge Isotope (Tewksbury, MA), FFA C16:0-d3 was purchased from C/D/N Isotopes Inc. (Pointe-Claire, Québec) and LPC 19:0 was supplied by Avanti Polar Lipids (Alabaster, AL).

For lipidomics, HPLC grade solvent (acetonitrile, methanol and isopropanol), ammonium acetate and tert-butyl methyl ether (MTBE) were purchased from Merck KGaA (Darmstadt, Germany). Ultrapure water (H2O) was collected from a Milli-Q system (Millipore, Billerica, MA). Internal standards (ISs) phosphatidylcholine PC(19:0/19:0), lysophosphatidylcholine LPC(19:0), sphingomyelin SM(d18:1/12:0), ceramide Cer(d18:1/17:0) were purchased from Avanti Polar Lipids (Alabaster, AL). Free fatty acid d3-FFA (C16:0), free fatty acid d3-FFA (C18:0) and triacylglycerol TAG(15:0/15:0/15:0) were purchased from C/D/N Isotopes Inc. (Pointe-Claire, Québec), Cambridge Isotope (Tewksbury, MA), and Merck KGaA (Darmstadt, Germany), respectively.

ISs were prepared in methanol as stock solution with the follow concentrations and stored in -20°C before use: PC (19:0/19:0) 0.67 µg/mL, LPC (19:0) 0.33 µg/mL, SM (d18:1/12:0) 0.17 µg/mL, Cer(d18:1/17:0) 0.17 µg/mL, TAG (15:0/15:0/15:0) 0.53 µg/mL, d3-FFA (C16:0) 0.67 µg/mL, and d3-FFA(C18:0) 0.67 µg/mL.

3.3.2.2 Biological samples

Between June 2011 and May 2015, 82 obese patients were recruited into the digestive and metabolic surgery service in the Jean Verdier University Hospital, Bondy, France. These patients were candidates for a bariatric surgery (Laparoscopic adjustable gastric banding, longitudinal sleeve gastrectomy or Roux-en-Y Gastric Bypass). A total of 82

blood plasma were collected from 82 patients. Biopsies were performed intraoperatively by laparotomy during bariatric surgery. All patients included in the cohort signed a consent covering intraoperative liver biopsy and subsequent use of frozen specimens and blood plasma.

Inclusion criteria were: (1) age 18 and over with morbid obesity: BMI ≥ 40 or BMI ≥ 35 with at least one associated comorbidity (hypertension arterial disease, type 2 diabetes, dyslipidemia, obstructive sleep apnea syndrome, osteoarthritis and / or NAFLD); (2) absence of other hepatic infection (exclusion criteria): autoimmune or infectious hepatic disease (chronic viral hepatitis), hemochromatosis, history of chemotherapy or hepatotoxic drugs, alcohol consumption over 20 g / day for women and over 30g / day for men ; (3) Physical and psychological eligibility for a bariatric surgical procedure (decided at the multidisciplinary consultation meeting in the surgery service).

Liver biopsy was performed on the outer edge of the left lobe of the liver with a depth of 2 cm minimum in the parenchyma. This biopsy was immediately divided into three portions: 1) a part for metabolic flow analyzes; 2) a part for histological analysis; 3) a part for direct freezing at -80° C. Histological analysis based on an algorithm described by Bedossa et al. [189], which allowed the classification of liver biopsies patients in three categories: Normal Liver (NL; steatosis, inflammation and ballooning = 0); steatosis (NAFL; steatosis 1-3 + either inflammation 1-3 or ballooning 1-2) and NASH (steatosis 1-3 with 1-3 inflammation and 1-2 ballooning + 1-4 fibrosis).

3.3.2.3 Samples Preparation

For metabolomics, plasma samples were randomized, thaw on ice, then aliquots of each plasma sample were pooled as QC samples. For deproteinization and metabolites extraction, 150 µL of plasma was mixed with 600 µL methanol containing ISs, after vortexed for 2 min, centrifuged for 10 min at 16 000 g at 4 °C. The supernatant was lyophilized and then stock in -80°C before analysis, for quality control during sample preparation, a QC sample was prepared with every 5 plasma samples.

For lipidomics, plasma samples were randomized, then thaw on ice, 20 µL aliquot of each samples were pooled as QC samples. For deproteinization, 40 µL plasma was mixed with

300 μL ice-cold methanol containing ISs, after vortexed 30s, 1 mL MTBE was added in the mixture and vibrated at room temperature for 1 h for lipids extraction. Then, 300 μL water was added followed by vortexing 30 s and stay at 4°C for 10 min. After centrifugation at 14000 g for 15 min at 4 °C, 2 times 400 μL supernatants were transferred into two new Eppendorf tubes, supernatants were lyophilized, then stock in -80°C before analysis, for quality control during sample preparation, a QC sample was prepared for every 5 plasma samples.

The detailed sample preparation was descripted in **Part Two** METHODOLOGY, section 2.5.3 Sample preparation for NMR based metabolomic analysis, 2.5.4 Sample preparation for LC/GC-MS based analysis, section 3.1 Protocol for blood plasma sample extraction for metabolomics and section 3.2 Protocol for blood plasma sample extraction for lipidomics.

3.3.2.4 Data Acquisition

3.3.2.4.1 Analysis Equipment

For NMR study, samples were analyzed by using a 500 MHz Bruker AVANCE III ^{1}H NMR spectrometer (Advance III, Bruker, Germany) with automatic sample changer.

For UPLC-HRMS based analysis, an ACQUITY Ultra Performance Liquid Chromatography (UPLC, Waters Corporation, Manchester, U.K.) was coupled with a Q Exactive HF (Thermo Fisher Scientific, Rockford, IL) system and an AB SCIEX TripleTOF™ 5600 plus mass spectrometer system (AB SCIEX™, Framingham, MA) for UPLC-HRMS based Metabolomic analysis and UPLC-HRMS based Lipidomic analysis, respectively.

3.3.2.4.2 Analysis by NMR and LC/MS

3.3.2.4.2.1 NMR Analysis

For NMR analysis, 250 μL thawed plasma was prepared with 350 μL deuterated PBS (contain NaN_3), which permits the deuterium frequency-field lock, after vortexed mix for 1 min, then centrifuged for 10 min at 12 000 g at 4 °C, 550 μL supernatant was transferred into a clean 5 mm NMR tube.

3.3.2.4.2.2 UPLC-HRMS based Metabolomic analysis

The supernatant was re-dissolved in methanol/water (1:4, v/v) solvent before analysis. 5 μL re-dissolved supernatant was used for Metabolomic analysis, which was performed on a Q Exactive HF (Thermo Fisher Scientific, Rockford, IL) system coupled with an ACQUITY Ultra Performance Liquid Chromatography (UPLC, Waters Corporation, Manchester, U.K.). Column temperature and automatic sampler temperature were set at 60°C and 6°C, respectively. In order to cover as many types of compounds as possible, different columns were used in the positive and negative ionization mode.

For electrospray positive ion (ESI+) mode, BEH C8 (2.1 × 50 mm, Waters) column was used to ensure the separation of weakly polar compounds such as carnitine and lipids, the mobile phases were water (A) and acetonitrile (B) with 0.1% formic acid. The flow rate was 0.40 mL/min and the total run time was 12 min. The elution program started with 5% B and was held for 0.5 min, then linearly increased to 40% B at 2 min and increased to 100% B in 8 min, maintained 2 min, then went back to 5% B in 0.1 min and kept for 1.9 min for post equilibrium.

For electrospray negative ion (ESI-) mode, HSS T3 (2.1 × 50 mm, Waters) column was used to ensure the retention and separation of polar compounds in reverse phase, the mobile phases were water (A) and methanol/water (95:5, v/v) (B) containing 6.5 mM Ammonium bicarbonate. The flow rate was 0.40 mL/min and the total run time was 12 min. The elution program started with 2% B and was held for 0.5 min, then linearly increased to 40% B at 2 min and increased to 100% B in 8 min, maintained 2 min, then went back to 2% B in 0.1 min and kept for 1.9 min for post equilibrium.

3.3.2.4.2.3 UPLC-HRMS based Lipidomic analysis

Before analysis, the supernatant was re-dissolved in the mix of 20 μL solution A (chloroform : methanol, 2 : 1 (v/v)) and 80 μL solution B (water : isopropanol : acetonitrile, 5 : 30 : 65 (v/v/v) containing 5mmol/L ammonium acetate), after vortexed for 1 min and centrifuged at 14 000 g for 5 min at 4°C, lipidomic analysis was performed on the AB SCIEX TripleTOF™ 5600 plus mass spectrometer system (AB SCIEX™, Framingham, MA) coupled with an Ultra-high-performance liquid chromatography (Waters, Milford, MA), equipped a reversed-phase UPLC ACQUITY C8 BEH column (2.1 mm × 100 mm × 1.7 μm, Waters, Milford, USA), the column temperature was 55°C in electrospray positive and negative ionization (ESI+ and ESI-) modes. Acetonitrile : water, 6 : 4 (v/v) containing 10 mM ammonium acetate was used as mobile phase A. Isopropanol : acetonitrile, 9 : 1 (v/v) containing 10 mM ammonium acetate was used as mobile phase B. The flow rate was 0.26 mL/min, with the elution gradient as follows: 32% B was firstly maintained for 1.5 min, then linearly increased to 85% B in 14 min, linearly increased to 97% B from 15.5 min to 15.6 min, finally maintained for 2.4 min and followed by equilibration with 32% B in next 2 min.

3.3.2.4.3 Analysis Sequences

For each plasma sample, NMR spectra were acquired from two complementary experiments: One-dimensional ^{1}H Nuclear Overhauser Effect Spectroscopy NOESY1D presat (NOESY1dgppr sequence) [190] and Carr-Purcell-Meiboom-Gill (CPMG presat) [191]. The spectral width was 10 kHz.

For MS based metabolomic analysis, the resolutions of full scan MS and $ddMS_2$ were set at 120 000 and 60 000, respectively. The automatic gain control (AGC) target and maximum injection time in full scan MS settings were 1×10^6 and 200 ms, while their values were 1×10^5 and 50 ms in $ddMS_2$ settings. The TopN (N, the number of top most abundant ions for fragmentation) was set to 10, and collision energy was set to 15 eV, 30 eV and 45 eV. A heated ESI source was used at positive and negative ion mode. The spray voltage was set as 3.5 kV. The capillary temperature and aux gas heater temperature were set as 300 and 350 °C, respectively. Sheath gas and aux gas flow rate were set at 45 and 10 (in arbitrary units), respectively. The S-lens rf level was 50.

For MS based lipidomic analysis, data acquisition was performed both in full scan (with mass range from 200 to 1000m/z for ESI+, 90 to 1000m/z for ESI-) and IDA mode (with mass range from 100 to 1000m/z for ESI+, 50 to 1000m/z for ESI-). Mass spectrometry parameters were as follows: ion spray voltage, 5500V for ESI+ and -4500V for ESI-; curtain gas was 35 psi; declustering potential, full scan mode: 100V for ESI+ and -100V for ESI-, IDA mode: 80V for ESI+ and -100V for ESI-; collision energy, full scan mode: 10V for ESI+ and -10V for ESI-, IDA mode: 35V for ESI+ and -35V for ESI-, collision energy spread was 15 in ESI+ and ESI- mode; interface heater temperature, 500°C for ESI+ and 550°C for ESI-.

3.3.2.5 Data processing

3.3.2.5.1 NMR Data processing

The detailed processing steps were descripted in the section 2.6.1 NMR data preprocessing. Briefly, a Fourier transformation was applied on NMR data with line-broadening (LB) at 0.3 Hz. Spectra were phased and baseline corrected. All spectra were aligned on glucose doublet at 5.23 ppm. After processing and calibration, each 1D NMR spectrum was sliced into buckets of 0.001 ppm, containing NMR signals. These steps were performed by using NMRPipe software [124]. Spectra bin were further normalized to the median of intensities before statistical analysis.

3.3.2.5.2 UPLC-HRMS Data processing

For MS based metabolomic analysis, raw MS data were collected and processed on TraceFinder software (version 4.1, Thermo Fisher Scientific, Rockford, IL), for peak extraction and integration.

For MS based lipidomic analysis, raw MS data were collected and processed on PeakView® Software (version 2.2, AB SCIEX™, Framingham, MA) and MultiQuant™ Software (version 3.0.3, AB SCIEX™, Framingham, MA), for peak extraction and integration, respectively.

The missing values are replaced by the average of the corresponding (feature specific) non-missing values in the k (here k = 10) closest features in terms of Euclidean distance of the responses across all the samples. After processing, MS features with relative standard deviation (RSD) < 30 were used for further statistical analysis, before the flow statistical analysis, MS data were Probabilistic Quotient normalized.

3.3.2.6 Statistical Analysis

Dunn's Multiple Comparison Test [177] was performed with IBM® SPSS® Statistics V25.0 for macOS (Armonk, NY: IBM Corp.), Heatmap and Kruskal-Wallis test with false discovery rate (FDR) correction were performed with Multiple Experiment Viewer (V_4_8_1_r2727_mac) for macOS [192], metabolites were selected with p-value < 0.05, the regrouping of metabolites was performed with pearson correlation based hierarchical clustering.

Principal Component Analysis (PCA), Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) were performed using in house MATLAB OPLS script based on Trygg and Wold method [162], analyses were performed with MATLAB® (R2016b) for macOS (Mathworks, Natick, Massachusetts, USA). Quality parameters of the models, the explained variance (R2Y) and the predictability of the model (Q2Y) were calculated. Q2Y was calculated by a 7-fold cross validation and confirmed by exploring the impact of permutations in the dataset rows [193]. To evaluate the discriminatory power of the model, the area under the receiver operating curve during the cross validation (CV-AUROC) and Confusion Matrix were calculated. The Confusion matrix was generated by a logistic regression, components in OPLS-DA model were used as variables, after a 7-fold cross validation, all samples have a predicted probability, and probability >= 0.5 consider as class 1, probability < 0.5 consider as class 0. Models were validated by permutation tests (n=200). S-plot [161] for OPLS-DA model were used to identify potentials biomarkers.

3.3.2.7 Identification

For NMR study, identification was performed with the help of 2D experience, Chenomex software and Human Metabolome Database (HMDB) (version 4.0).

For MS based metabolomic analysis, identification was performed with MS/MS experience, in-house MS database, TraceFinder software (version 4.1, Thermo Fisher Scientific, Rockford, IL) and Human Metabolome Database (HMDB) (version 4.0).

For MS based lipidomic analysis, identification was performed with MS/MS experience, in-house MS database, and LIPID MAPS [194]–[198].

3.3.3 RESULTS

3.3.3.1 Characteristics of NL, NAFL and NASH Patients

There were 66 females (80%) and 16 males (20%) involved in the present study, the diagnoses of NL, NAFL and NASH were established histologically in liver biopsy specimens. Patient's characteristics and clinical laboratory data are represented in the Table 9. Compared with NL obese patients, there were no significant differences in terms of BMI, Cholesterol and Phospholipid (in the liver) in patients with NAFL or NASH. Patients with NASH had significant high level of ALT, ASAT, GGT, TG and Fasting blood glucose compared with both Normal Liver obese controls and NAFL patients.

As anticipated, it can be seen that patients with NAFLD (NAFL and NASH) are Insulin Resistance (HOMA IR) [199] and have significantly higher levels of triglycerides in liver. Because the diagnoses of NL, NAFL and NASH patients was mainly based on these parameters.

Table 9. Characteristics of NL, NAFL and NASH Patients.

Characteristics	NL (n=19)	NAFL (n=39)	NASH(n=24)	p-value
Age (year)	32.2 (8.7)	36.2 (10.2)	41.7 (10)†	< 0.01
Gender (♂/♀)	(1/18)	(5/34)	(10/14)	-
BMI (kg/m2)	42.9 (4.6)	45.3 (5.1)	44.5 (5.3)	NS
ALT (IU/L)	18.7 (7.3)	28.5 (20.3)‡	50.0 (28.3)†	< 0.001
AST (IU/L)	20.1 (5.1)	25.3 (11.7)‡	34.7 (15.5)†	< 0.002
GGT (IU/L)	19.9 (5.5)	35.2 (38.5)‡	47.2 (31.9)†	< 0.001
Phosphate alcaline	68.9 (16.3)	74.1 (18.9)	69.0 (24.4)	NS
TG (mM)	1.0 (0.5)	1.3 (0.6)‡	2.3 (1.5)†	< 0.001
Cholesterol (mM)	4.8 (0.9)	4.7 (1.0)	5.0 (1.1)	NS
Fasting blood Glucose (mM)	5.0 (0.6)	5.3 (1.2)‡	6.7 (2.6)†	< 0.002
Fasting Insulin (pM)	71.5 (39.3)	142.2 (109.0)*	171.8 (150.6)†	< 0.02
HOMA IR	2.4 (1.3)	5.2 (4.5)*	7.2 (5.6)†	< 0.004
Triglycerides (mg/g of liver)	6.1 (2.9)	16.0 (12.4)*‡	25.5 (14.6)†	< 0.001
Phospholipids (mg/g of liver)	0.8 (0.7)	0.8 (0.6)	0.8 (0.4)	NS
Diacylglycerols (mg/g of liver)	0.6 (0.4)	0.7 (0.3)	0.9 (0.5)†	< 0.04

Data are expressed as “mean (SD)”. p-value were calculated with Dunn’s Multiple Comparison Test [200], which is a post-hoc test for Kruskal-Wallis. This will compare the pairs of groups, but is statistically more sensible than doing pairwise Mann-Whitney tests, significant differences between groups ($p < 0.05$) are represented by *: NL - NAFL, †: NL - NASH, and ‡ : NAFL - NASH.

BMI: Body Mass Index, TG: Triglycerides, AST: Aspartate transaminase, ALT: Alanine transaminase, GGT: Gamma-Glutamyl Transferase, HOMA: Homeostatic model assessment [199], is a method for assessing β-cell function and Insulin Resistance (IR), NS: Non Significant, NL: Normal Liver, NAFL: Nonalcoholic fatty liver or steatosis, NASH: Non-Alcoholic SteatoHepatitis, SD: Standard deviation.

3.3.3.2 Metabolomic and lipidomic profiles of NL, NAFL and NASH patients

3.3.3.2.1 NMR Analysis

The first part of metabolomics analysis was performed with an NMR spectrometer, as NMR analysis requires minimal sample preparation and there is NMR spectrometer available in our laboratory. After processing as procedure described in Material and Method, 1D NOESY NMR data were used to build multivariate statistical models for classification of NL, NAFL and NASH patients. In our study, blood samples were collected with tube containing Sodium Citrate which was used as anticoagulant during the preparation of blood plasma from blood, consequently, data of NMR spectra region of Citrate (between 2.5 - 2.7ppm) were excluded for multivariate statistical analysis. Also, 1 NL and 8 NASH patients were excluded for multivariate statistical analysis due to presence of strong sugar signal. The presence of this abnormal sugar signal lead to a very large variability on the spectra and have impact in statistical analysis.

OPLS-DA models for classification of NL, NAFL and NASH patients were investigated, the model for classification of NL and NASH patients was tested firstly as these 2 groups patients represent normal and the advanced stage of NAFLD, respectively, which means they have the maximum difference in the stage of NAFLD. An OPLS-DA model for classification of NL and NASH patients was obtained with 2 components (1 predictive and 1 orthogonal) as represented in Figure 1. The score plot of the OPLS-DA model for classification of NL and NASH patients was represented in Figure 1. A.

To estimate the model performance, OPLS-DA models were evaluated by 200 times permutation tests (as represented in Figure 19. B), and repeated (n = 200) 7-fold CV-AUC, with R2Xcum: 0.64, R2Ycum: 0.48, R2 is the indicator of how model fit the data, the more closely to 1 the better the fit, Q2cum: 0.39, Q2 is the capacity of model to correctly class a new dataset, the more closely to 1 the better the model predictability. CV-AUC: 0.896, which means the probability that the OPLS-DA model will score a randomly chosen NASH classed as NASH patient higher than a randomly chosen NASH classed as NL patient is 89.6%. All of which indicate that our OPLS-DA model for classification of NL and NASH patients is fairly stable.

The variables in the model that contributed importantly in the classification of NL and NASH patients are represented in figure 19.C. The covariance plot of OPLS-DA was generated, the covariance plot restores the form of NMR spectra, and colored it with correlation score with NASH, the red spectra regions were having a correlation > 0.5 with NASH.

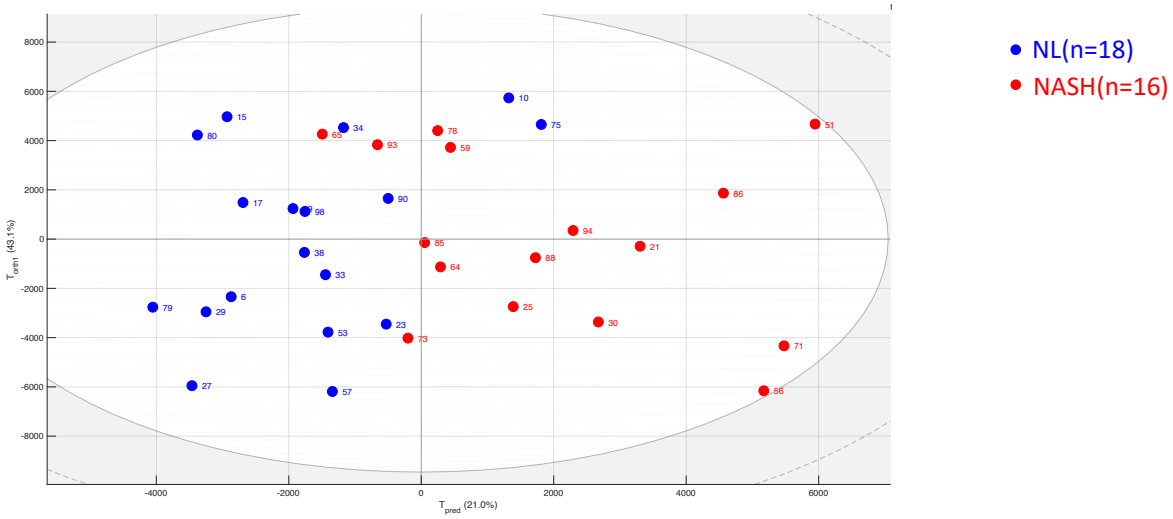
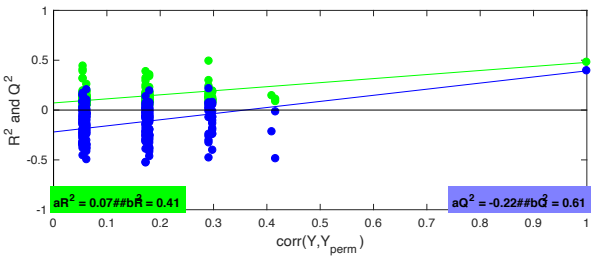


Figure 19. A, Score plot of OPLS-DA (NL vs NASH, 1D NOESY Data)



200 times permutations, 2 components

Figure 19. B, 200 times permutation test for OPLS-DA (NL vs NASH, 1D NOESY Data)

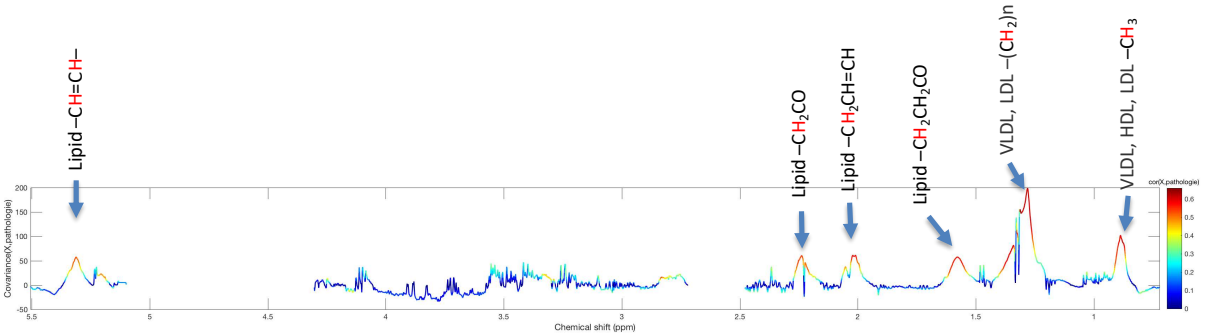


Figure 19. C, Covariance plot of OPLS-DA (NL vs NASH, 1D NOESY Data)

As represented in the Figure 19. C, it looks like that lipids such as VLDL, LDL and HDL were major discriminants compounds to differentiate NL patients from NASH patients.

OPLS-DA models for classification of NL and NAFL; NAFL and NASH patients were also investigated. However, after several tests, we didn't find significant models neither for classification of NL and NAFL patients nor for classification of NAFL and NASH patients.

In summary, this first part of metabolomics analysis with NMR demonstrated that lipids such as VLDL, LDL and HDL were particularly important to discriminate between NL and NASH patients. To go father, we want to determine lipids or the exact lipid classes which contribute the most in the discrimination. As we know that these metabolites were important in the disease, it is logical to suggest that some lipids may help in discriminating of NL, NAFL and NASH patients. Thus, we performed further UPLC-HRMS based metabolomics and lipidomics analysis to cover as much as possible lipids and metabolites.

3.3.3.2.2 UPLC-HRMS based Metabolomics analysis

3.3.3.2.2.1 Global metabolites changes between NL, NAFL and NASH patients

There were 19 NL, 37 NAFL (lost 2 samples during reconstitution before MS analysis) and 24 NASH plasma samples analyzed with an UPLC-HRMS (Orbitrap) based metabolomics

according to the procedure described in Material and Method, 198 distinct metabolites (positive and negative mode) were identified in the plasma samples. The changes in metabolites between NL, NAFL and NASH patients were compared by using Kruskal-Wallis test, there were 36 metabolites with $p < 0.05$, none of them were significative after false discovery rate (FDR) correction (FDR limit 0.05), however, when FDR limit set at 0.25 they were significative. These 36 metabolites were used to generate the Heatmap below (Figure 20). The complete list for detected mass with retention time used for metabolites identification is displayed in Annexe 1.

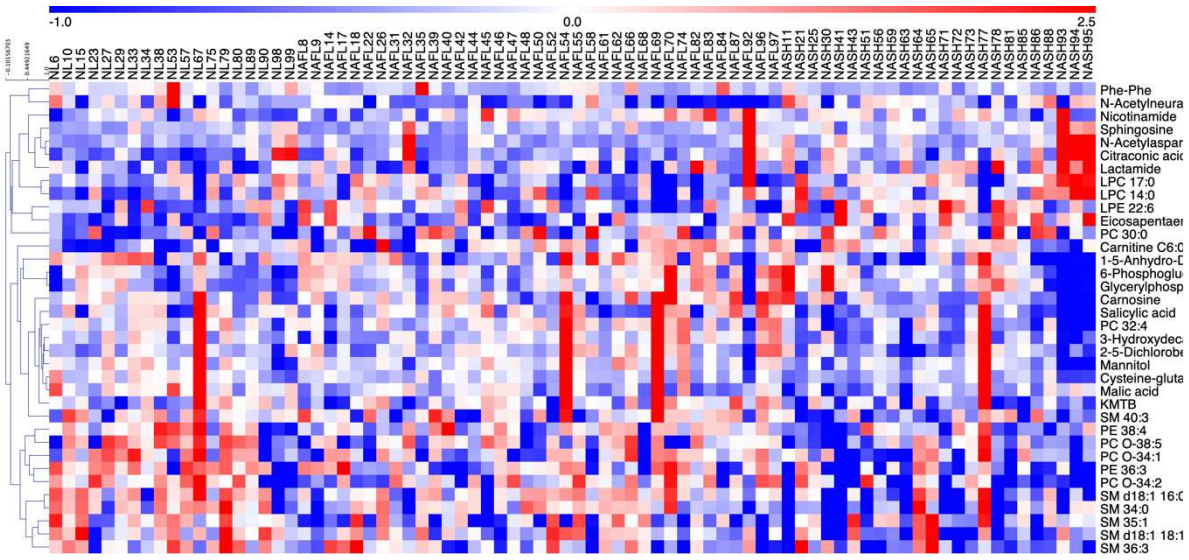


Figure 20. Heatmap shows metabolites changes between NL, NAFL and NASH patients. Metabolites were selected with p-value < 0.05, the regrouping of metabolites was performed with hierarchical clustering (pearson correlation).

Compared with NL obese patients, Sphingomyelin (SM), Phosphatidylethanolamine (PE) and Phosphatidylcholine (PC) tend to have a lower relative concentration in NAFL and NASH obese patients. In contrast, Lysophosphatidylcholines (LPC), lysophosphatidylethanolamine (LPE), Sphingosine, lactamide tend to have a higher relative concentration in NASH obese patients than in NAFL and in NL obese patients (Figure 20).

In view of these results, it can be seen that certain metabolites may have a relative different concentration in different stage of NAFLD. Thus, we performed further multivariate statistical analysis to see if combination of several metabolites could be useful to class different stage of disease.

3.3.3.2.2.2 OPLS-DA models for comparison between NL, NAFL and NASH patients

After processing as procedure described in Material and Method, UPLC-HRMS data were used for multivariate statistical analysis, the PCA Score Plot of all analyzed plasma samples was represented in figure below (Figure 21).

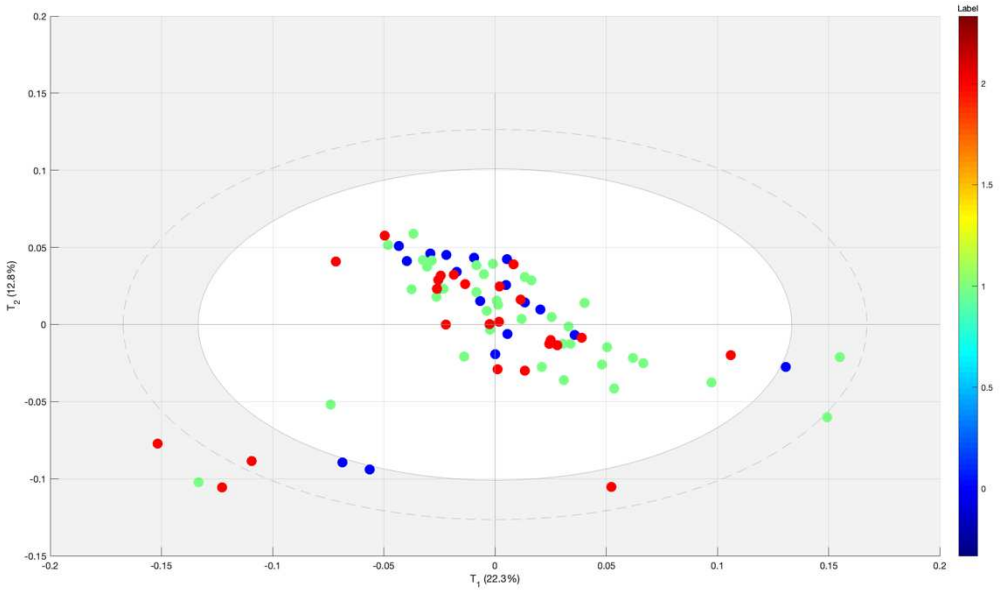


Figure 21. PCA Score Plot for all analyzed plasma samples. NL (bleu, n=19), NAFL (green, n=37) and NASH (red, n=24).

For classification of NL, NAFL and NASH patients, several OPLS-DA models have been investigated. The outcomes of different models were summarized in the table below (Table 10).

Table 10. Summary of OPLS-DA models for metabolomics analysis.

Groups	**Model**	**A**	**Q2**	**CV-AUC (median, n=200)**
NL vs NASH (16 vs 19)	OPLS-DA	1+1	0.378	0.84 (0.83-0.86)
NL vs NAFL (16 vs 33)	OPLS-DA	1+0	0.108	0.71 (0.69-0.72)
NAFL vs NASH (34 vs 20)	OPLS-DA	(1)*+0	-0.033	0.61 (0.58-0.65)

A, number of significant components given by cross-validation, *, non-significant component. CV-AUC, median (1st - 3rd quartile) with 200 times cross-validation. Outliers identified by Principal Component Analysis (PCA) were excluded in building OPLS-DA models.

As represented in the table above (Table X), the classification of NL and NASH patients gives better result in term of model quality (Q2: 0.378; AU-ROC: 0.85) than the classification of NL, NAFL (Q2: 0.108; AU-ROC: 0.69) and NAFL, NASH (Q2: -0.033; AU-ROC: 0.62) patients. The Score Plot of OPLS-DA model for classification between NL (bleu) and NASH (red) patients was represented below (Figure 22).

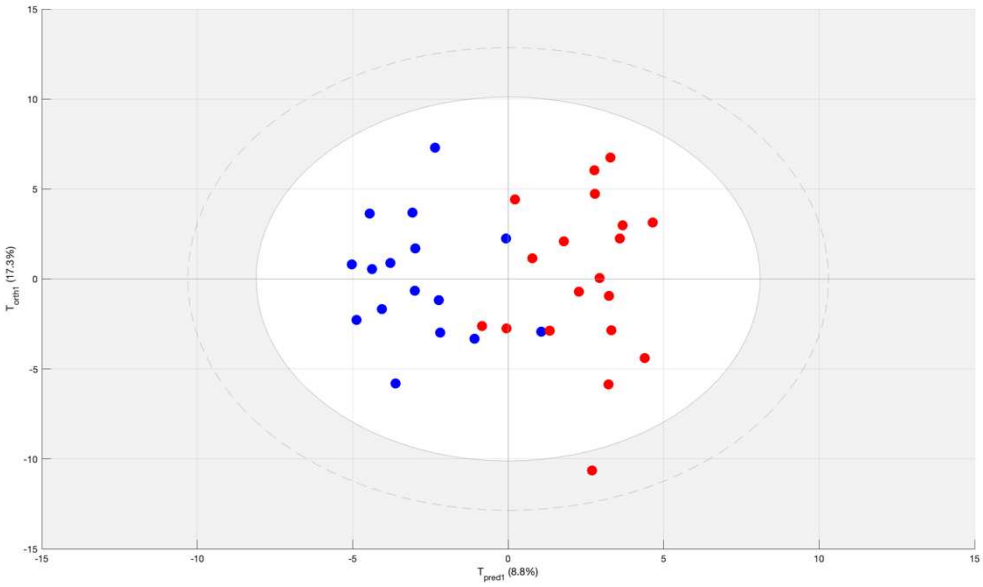


Figure 22. Score Plot of OPLS-DA model for classification between NL (bleu) and NASH (red) patients (Metabolomics analysis).

OPLS-DA models classification performance were further evaluated by confusion matrix, which is a table that lists the correct and false predictions versus the actual observations, it is a good way to describe the performance of a classification model (classifier) when presented with new data. It allows easily identify confusion or error (like one class is misclassed as the other) of classification model when it makes predictions.

As represented in Table 11, OPLS-DA model has difficulty to distinguish NL patients from NAFL patients (the model is more likely confused in the classification of NL patients), and to distinguish NAFL patients from NASH patients (the model looks like straggled in the classification of NASH patients).

Table 11. Confusion matrix of the sample by OPLS-DA models.

	NL	NASH	Accuracy	Class error
NL	**13**	3	0.83	0.19
NASH	3	**16**		0.16
	NL	NAFL	Accuracy	Class error
NL	**4**	12	0.67	0.75
NAFL	5	**28**		0.15
	NAFL	NASH	Accuracy	Class error
NAFL	**30**	4	0.65	0.12
NASH	15	**5**		0.75

Rows represent the real groups; columns list the predicted groupings by metabolomic analysis. The Confusion matrix was generated by using a logistic regression, the predictive component in OPLS-DA model was used as variable, after a 7-fold cross validation (7-fold CV), all samples have a predicted probability, and probability >= 0.5 consider as class 1, probability < 0.5 consider as class 0. Accuracy: Overall, how often is the classifier correct, calculated by (TP+TN)/total, with True Positives (TP): These are cases in which we predicted yes (they have the disease), and they do have the disease. True Negatives (TN): We predicted no, and they don't have the disease.

To highlight important metabolites which mainly responsible for the classification of NL, NAFL and NASH patients in OPLS-DA models, the S-plot for OPLS-DA model was employed. S-plot is one of methods used to highlight putative biomarkers from a two group OPLS-DA model. S-plot combines the model covariance (Variable Contribution, in X-axis) and model correlation (Variable Confidence, in Y-axis) from the OPLS-DA model, and project on a scatter plot, which allow to highlight discriminants variables in the OPLS-

DA model. The S-plot of OPLS-DA model for classification between NL and NASH patients with highlighted important metabolites (red) was represented below (Figure 23), and the list of highlighted important metabolites were given in the table below (Table 12).

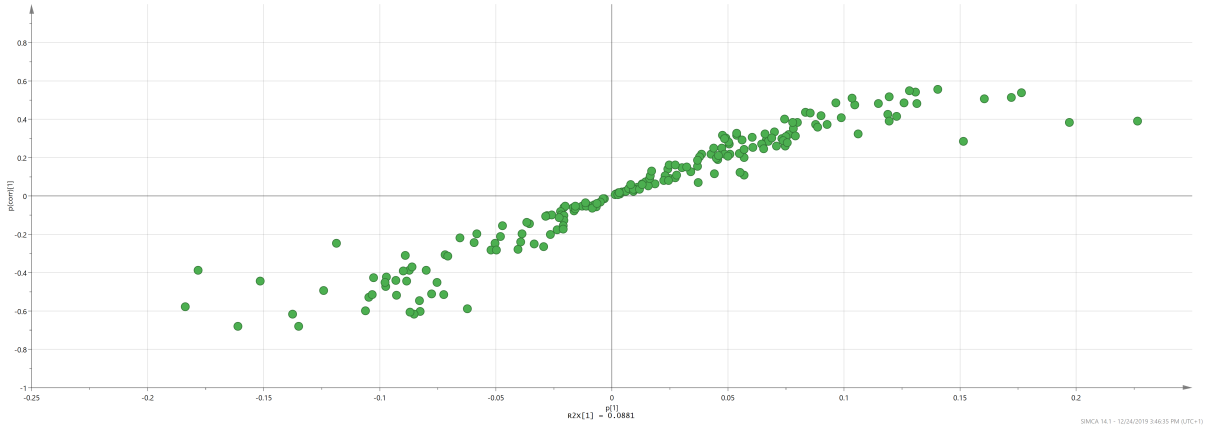


Figure 23. S-plot for OPLS-DA models highlight important metabolites (red) in the classification between NL and NASH patients (Metabolomics analysis).

Table 12. List of metabolites highlighted by S-plot for OPLS-DA models.

	Metabolites high in NL	**Metabolites high in NASH**
NL vs NASH	Cysteine-glutathione disulfide, Malic acid, Melezitose	Phenyl sulfate, Eicosapentaenoic acid, Glycodeoxycholic acid, LPC 14:0, LPC 16:1, PC 30:0

3.3.3.2.3 UPLC-HRMS based lipidomics analysis

3.3.3.2.3.1 Global lipids changes between NL, NAFL and NASH patients

19 NL, 39 NAFL and 24 NASH plasma samples were analyzed with UPLC-HRMS (TripleTOF) based metabolomics according to the procedure described in Material and Method. 419 distinct lipids were identified in patient's plasma (positive and negative mode). The changes in lipids between NL, NAFL and NASH patients were compared by using Kruskal-Wallis test, there were 97 lipids with $p < 0.05$, and all of them were significative after FDR correction (FDR limit: 0.05). The first 45 lipids with lowest p-value were used to generate the Heatmap below (Figure 24). The complete list for detected mass with retention time used for lipid ions identification is displayed in Annexe 2.

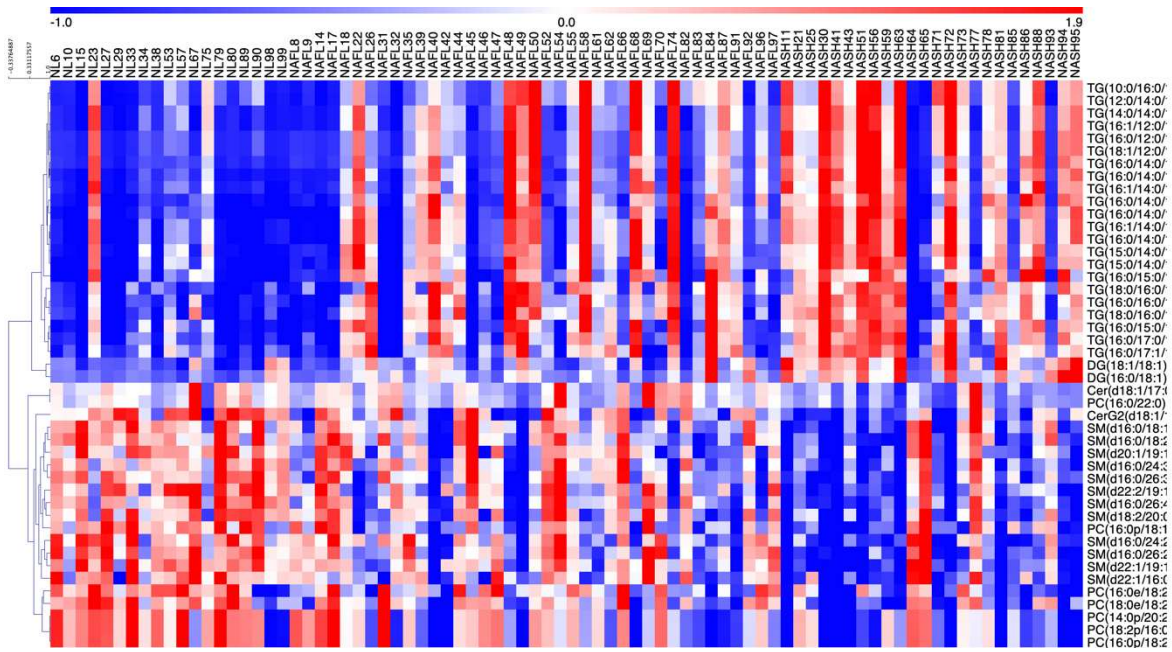


Figure 24. Heatmap shows lipids changes between NL, NAFL and NASH patients. Lipids were selected with p-value < 0.05, the regrouping of metabolites was performed with hierarchical clustering (pearson correlation).

Compared with NL obese patients, Sphingomyelin (SM), Phosphatidylcholine (PC) and Ceramides (Cer) were tend to have a lower relative concentration in NAFL and NASH obese patients. In contrast, Triglycerides (TG), Diglycerides (DG) were tend to have a

higher relative concentration in NAFL and NASH obese patients than NL obese patients (Figure 24).

3.3.3.2.3.2 OPLS-DA models for comparison between NL, NAFL and NASH patients

After processing as procedure described in Material and Method, UPLC-HRMS data were used to multivariate statistical analysis, the PCA Score Plot of all analyzed plasma samples was represented in figure below (Figure 25).

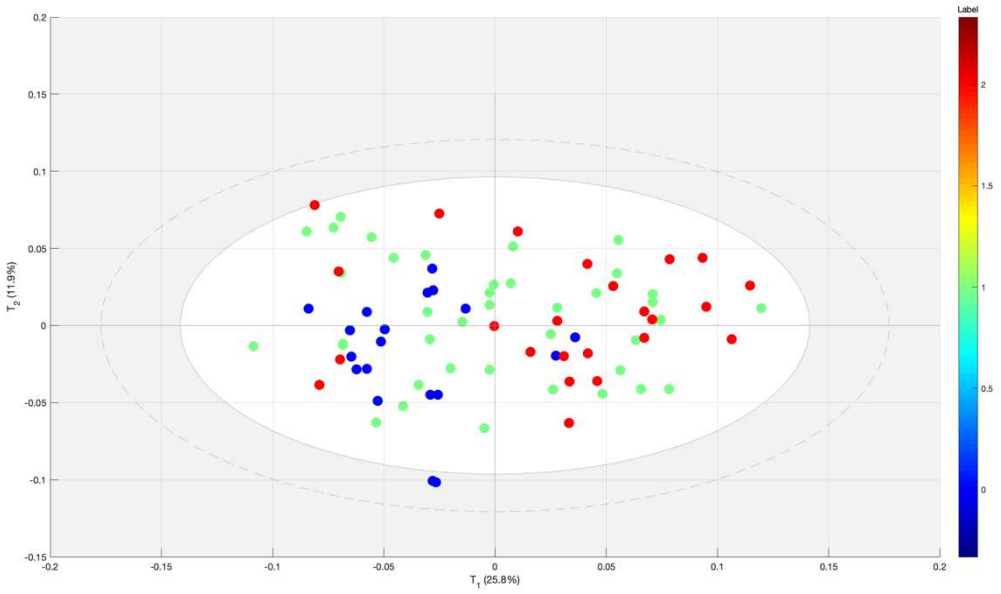


Figure 25. PCA Score Plot for all analyzed plasma samples. NL (bleu, n=19), NAFL (green, n=39) and NASH (red, n=24).

Several OPLS-DA models for classification of NL, NAFL and NASH patients have been investigated. A table which summarize OPLS-DA model's outcome is presented below (Table 13).

Table 13. Summary of OPLS-DA models for lipidomics analysis.

Groups	Model	A	Q2	CV-AUC (median, n=200)
NL vs NASH (18 vs 24)	OPLS-DA	1+(1)*	0.482	0.87 (0.85-0.88)
NL vs NAFL (18 vs 39)	OPLS-DA	1	0.168	0.76 (0.75-0.77)
NAFL vs NASH (39 vs 24)	OPLS-DA	1	0.066	0.68 (0.67-0.69)

A, number of significant components given by cross-validation, *, non-significant component. CV-AUC, median (1st - 3rd quartile) with 200 times cross-validation. Outliers identified by Principal Component Analysis (PCA) were excluded in building OPLS-DA models.

As represented in the table, the classification of NL and NASH patients gives better result again in term of model quality (Q2: 0.482; AU-ROC: 0.87) than the classification of NL, NAFL (Q2: 0.168; AU-ROC: 0.76) and NAFL, NASH (Q2: 0.066; AU-ROC: 0.68) patients. The OPLS-DA models (Score Plot, S-plot) for classification of NL, NASH patients and NL, NAFL patients were represented below (Figure 26).

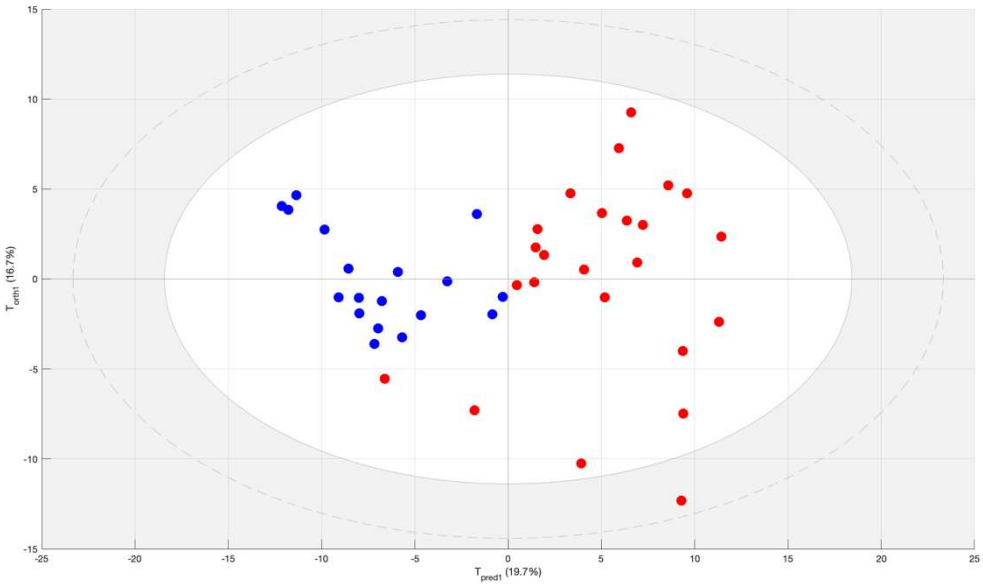


Figure 26. Score Plot of OPLS-DA models for classification between NL (bleu) and NASH (red) patients (Lipidomics analysis).

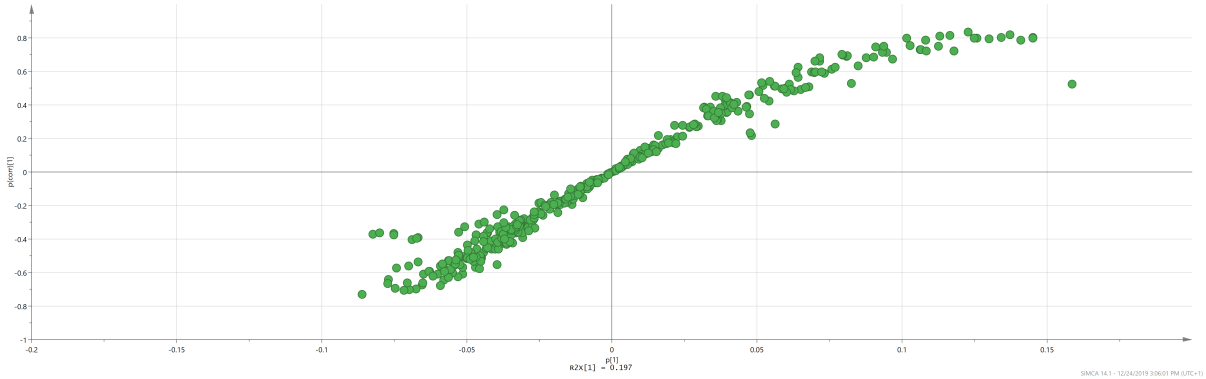


Figure 27. S-plot for OPLS-DA models highlight important lipids (red) in the classification between NL and NASH patients (Lipidomics analysis).

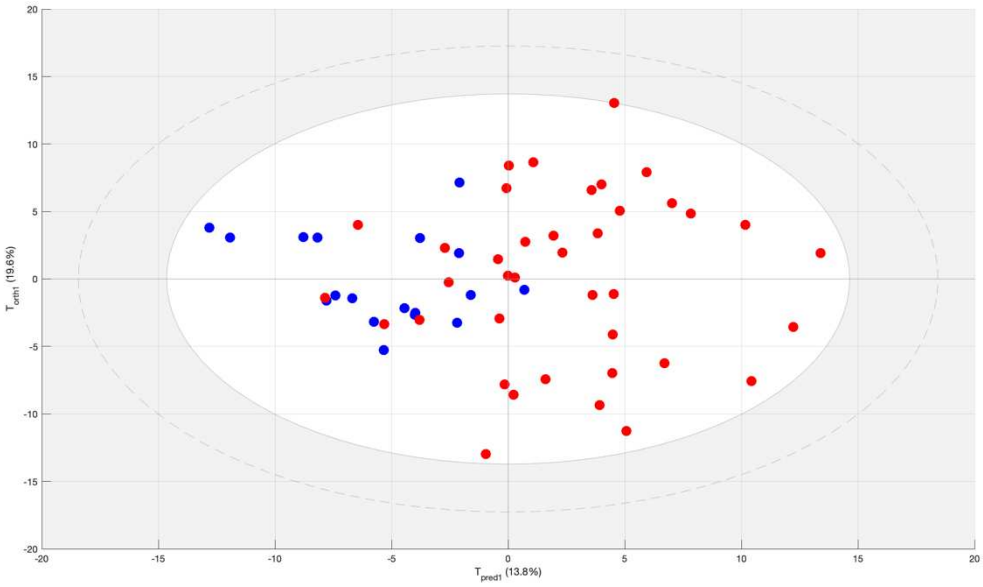


Figure 28. Score Plot of OPLS-DA models for classification between NL (bleu) and NASH (red) patients (Lipidomics analysis).

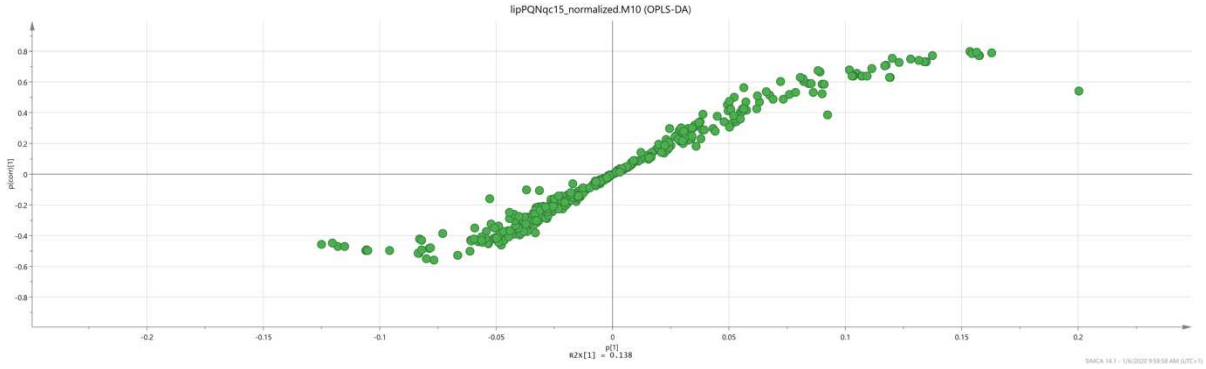


Figure 29. S-plot for OPLS-DA models highlight important lipids (red) in the classification between NL and NASH patients (Lipidomics analysis).

The S-plot for OPLS-DA model (Figure 29) was used to highlight important lipids (red) which mainly responsible for the classification of NL and NASH patients. The list of highlighted important lipids was given in the table below (Table 14).

Table 14. List of lipids highlighted by S-plot for OPLS-DA models.

	Lipids high in NL	**Lipids high in NASH**
NL vs NASH	SM(d16:0/26:4), SM(d16:0/24:3), SM(d22:2/19:1), SM(d22:1/19:1)	TG(18:1/12:0/14:0), TG(16:0/14:0/16:0), TG(16:0/14:0/16:1), TG(16:0/14:0/14:0), TG(16:1/14:0/14:0)

OPLS-DA models classification performance were further evaluated by confusion matrix. As represented in Table 15, OPLS-DA model has difficulty to distinguish NL patients from NAFL patients (the model looks like straggled in the classification of NL patients), and to distinguish NAFL patients from NASH patients (the model is more likely confused in the classification of NASH patients). The results are concordant (Table 11 and Table 15) with our UPLC-HRMS based metabolomics analysis results.

Table 15. Confusion matrix of the sample by OPLS-DA models.

	NL	NASH	Accuracy	Class error
NL	**15**	3	0.86	0.17
NASH	3	**21**		0.13
	NL	NAFL	Accuracy	Class error
NL	**8**	10	0.70	0.56
NAFL	7	**32**		0.18
	NAFL	NASH	Accuracy	Class error
NAFL	**34**	5	0.68	0.13
NASH	15	**9**		0.63

Rows represent the real groups; columns list the predicted groupings, 7-fold CV.

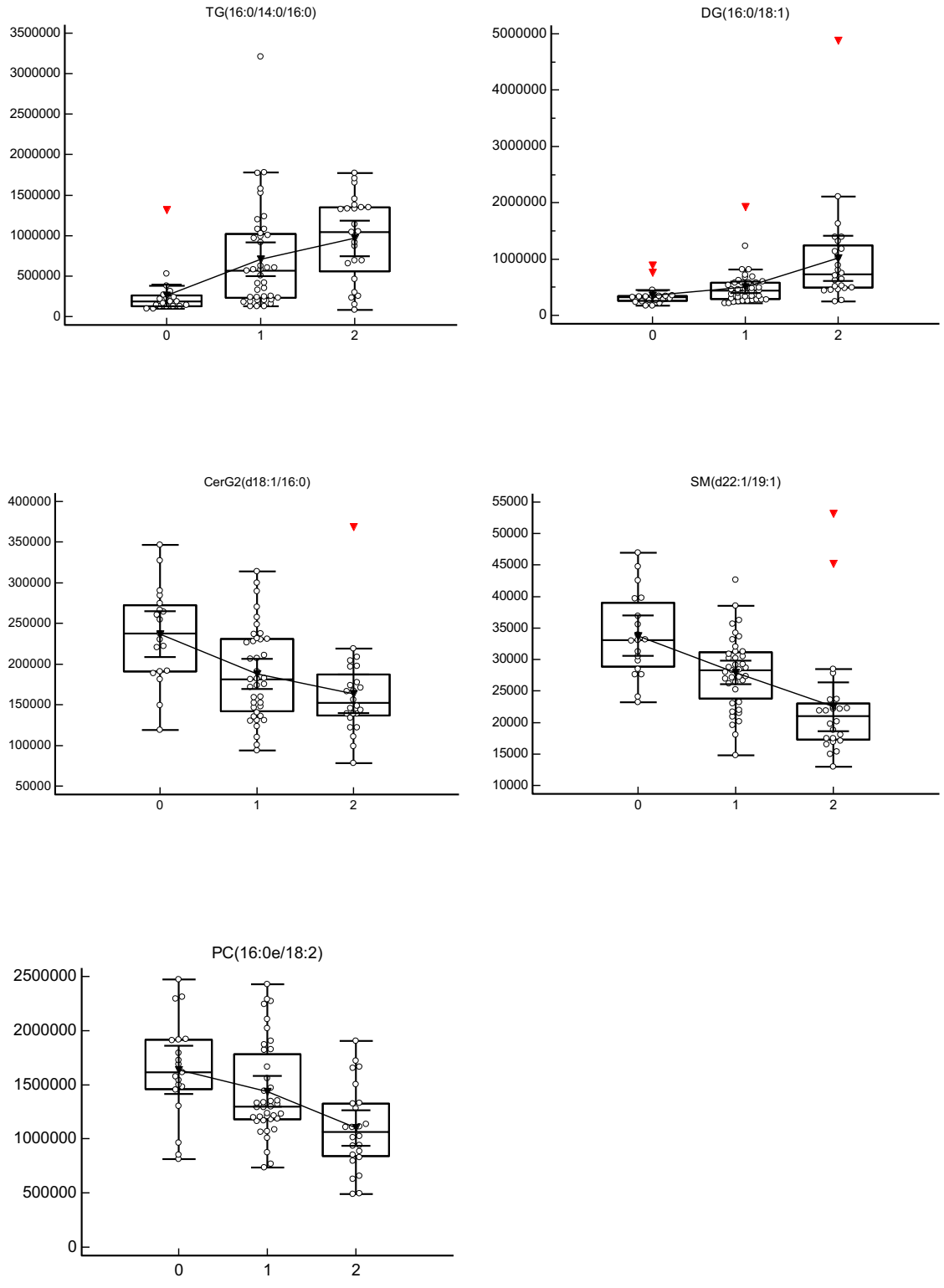


Figure 30. Box plots of plasma levels of selected metabolites or lipids in NL (0), NAFL (1) and NASH (2) patients. Probabilistic Quotient normalized areas are presented on the y-axis. The circle represents a sample, the red triangle represents a “far out” value.

3.3.4 CONCLUSION AND DISCUSSION

NAFLD is becoming the most common chronic liver condition in the world, however the diagnosis of NAFLD remains challenge currently, and there is an urgent need of non-invasive biomarkers for prognostication, selection of patients for treatment and monitoring of the disease [58]. The objectives of our study were (1) describe the relative plasma metabolome and lipidome changes in NAFL and NASH compared with NL obese patients, (2) investigated whether Ultra-Performance Liquid Chromatography (UPLC) coupled with High-resolution mass spectrometry (HRMS) based plasma metabolomics and lipidomics analysis could help to identify potentials biomarkers, if any, associated with different stages of NAFLD (NAFL, NASH), and (3) identify metabolomic or lipidomic patterns which could discriminate NAFL and NASH from NL patients, by using appropriate statistical models.

To the best of our knowledge, this study was the first using NMR, UPLC-HRMS based metabolomics and lipidomics analysis with biopsy confirmed patients in NAFLD. UPLC-HRMS based plasma metabolomics and lipidomics analysis were performed to obtain at first a global view of metabolites and lipids changes in NAFLD patients, then to identify disease-related patterns and to identify further biochemical perturbations.

Our results revealed significant relative changes in certain metabolites and lipids, especially for lipids metabolism, Sphingomyelin (SM), Phosphatidylcholine (PC) and Ceramides (Cer), Phosphatidylethanolamine (PE) were tend to have a lower relative concentration in NAFL and NASH obese patients. In contrast, Triglycerides (TG), Diglycerides (DG), Lysophosphatidylcholines (LPC), lysophosphatidylethanolamine (LPE), Sphingosine, lactamide were tend to have a higher relative concentration in NAFL and NASH obese patients than NL obese patients (Figure 20, Figure 24, Figure 30). Compared with NL obese patients, the changes in the plasma metabolome and lipidome were more distinct in NASH patients than in NAFL patients.

Triglycerides (TGs) and Diglycerides (DG) are mainly lipids involved in NAFLD [201], our results confirm the utility of UPLC-HRMS based metabolomics and lipidomics analysis in NAFLD, especially lipidomics, the results from lipidomics (Figure 24) such as changes in TG were concordant with clinic biochemistry analysis (Table 9). For metabolomics

analysis, the method used in this study doesn't cover TGs, this may explain why OPLS-DA models for lipidomics are slightly better than metabolomics in our study.

The classification of NAFL patients from NL patients is difficult in our study, this may due to the subjects are already in advanced stage of obesity with a very high BMI. The discrimination of NASH patients from NAFL patients remains a challenge. In classification of NL, NAFL and NASH patients, UPLC-HRMS based lipidomics and metabolomics analysis results in our study are concordant. Nevertheless, the classification model's outcome such as Q2, AU-ROC and confusion matrix are slightly better in lipidomics analysis than in metabolomics analysis.

The strengths of this study were the combination of robustness NMR, UPLC-HRMS based metabolomics and lipidomics analysis with biopsy confirmed samples. However, our study has limits. First, the subjects are already obese and with a very high BMI, this may have impact in metabolism such as lipids metabolism, thus, complicate interpretation of results. Second, it is a relatively small population, therefore, the number of patients in each group may not enough to achieve more strong significant statistic outcome. Third, the precision of LC-MS peak annotation was limited, especially for peaks which have a retention time less than 0.5 minute, as reverse phase UPLC column was used, these peaks correspond to high polarity metabolites, which not or less retained by stationary phase, thus, they retention time will be approximative, also, our in-house database used for metabolomics analysis may not cover other metabolites class which may have strong association with different stages of NAFLD. Moreover, validation in an independent cohort will be necessary.

In the following work, at first, we will interpret more profoundly metabolites and lipids changes, with integration in metabolism pathway. Then, we will try data filtering [202] and variable selection approach such as sparse PLS-DA (sPLS-DA) [86], [203], [204], to investigate if we could find better model. Besides, validation in an independent cohort was also planned and in preparation.

In conclusion, this study suggested that UPLC-HRMS based metabolomics, especially lipidomics analysis could be promising approach to identify biomarkers in NAFLD. Nevertheless, it should be underscored that NAFLD is a heterogeneous and complex

multi-organ disease [205], further investigation should be particularly focus on lipidomics, as well as investigation subtypes of patients, appropriate data processing and statistic model. Moreover, combine with other omics research such as transcriptomics, proteomics, and also clinical characteristics may improve novel subtyping approach of NAFLD patients, allowing further more precisely classification and staging of patients, in order to correctly interpret the biochemical processes behind the disease, which could contribute to the development of appropriate therapy and precision medicine-based management of patients.

3.4 UPLC-HRMS based untargeted plasma metabolomics in discovery of early biomarkers associated with risk of prostate cancer

Nota bene: this document is in preparation for submitting as a research article, the name listed below is a primary uncompleted list for people who mainly involved in this work.

Xiangping LIN[1,2,4], Lucie LÉCUYER[3], Xinyu LIU[4], Mohamed N. TRIBA[1,2], Tony PALAMA[1,2], Agnès VICTOR BALA[1,2], Zhicheng LIU[5], Nadia BOUCHEMAL[1,2], Philippe SAVARIN[1,2,*], Guowang XU[4,*], Mathilde TOUVIER[3,*]

1. Paris 13 University, Sorbonne Paris Cité University, Paris, France.
2. Chemistry Structures Properties of Biomaterials and Therapeutic Agents (CSPBAT Laboratory), The National Center for Scientific Research (CNRS) 7244, 93017 Bobigny Cedex, France.
3. Sorbonne Paris Cité Epidemiology and Statistics Research Center (CRESS), French National Institute of Health and Medical Research (Inserm) U1153, French National Institute for Agricultural Research (Inra) U1125, French National Conservatory of Arts and Crafts (Cnam), Paris 13 University, Nutritional Epidemiology Research Team (EREN), 93017 Bobigny Cedex, France.
4. CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China.
5. School of pharmacy, Anhui medical university, Hefei, China.
* To whom correspondence should be addressed.

3.4.1 INTRODUCTION

Prostate cancer (PCa) is the second most commonly diagnosed cancer and the second leading cause of cancer death (7.1% for incidence) among males [68]. Currently, there is no single definitive test to identify prostate cancer in men [69]. Prostate-Specific Antigen (PSA) test and digital rectal examination are screening methods used for PCa, for the definitive diagnosis, prostate biopsy and supplementary imaging are required [70]. The PSA test is a relatively easy to perform test and applicable for population in large scale, however, it has well known limits such as sensitivity, specificity, and can lead to false-positive and false-negative results [69].

Although extensive efforts in biomarker discovery during the last decades, including the genome and transcriptome approach, which has contributed to the identification of predictive biomarkers, more sensitive and specific biomarkers are still very demanding in early detection, prognosis, monitoring, and clinical management of PCa patients [71]–[74].

Metabolomics, defined as systematic analysis of metabolites in biofluids [10], [11], tissues [12], [13] or cells [14], [15] and investigate metabolites changes (or perturbations) during diseases (eg., cancer) [16]–[18], physiological processes (eg., aging) [19] or external stimulus (eg., drug treatment) [20], [21], has shown to be a promising and powerful tool to identify novel PCa biomarkers in biofluids [75]–[79].

In this context, combination of UPLC-HRMS (Obitrap) based metabolomics and epidemiological approaches may open new perspectives in PCa research, especially for identifying novel biomarkers, evaluation of risk and investigation the etiology of PCa [206]–[208]. In the present study, a prospective nested case-control study was set up in the *Supplémentation en Vitamines et Minéraux Antioxydants* (SU.VI.MAX) cohort [209], [210], with selected baseline plasma samples from 146 prostate cancer cases and 272 matched controls diagnosed during a 13-year follow-up.

The SU.VI.MAX study (clinicaltrials.gov; NCT00272428) was initially designed as a double-blind placebo-controlled trial with the purpose of assessing the influence of a daily supplementation with nutritional doses of antioxidants on the incidence of

cardiovascular diseases and cancers [209]. The study design and methods have been previously detailed. Briefly, a total of 13 017 participants were enrolled in 1994–95 for an 8-year intervention trial and were followed up for health events until September 2007 [209], [210].

The aim of our present study was to investigate whether UPLC-HRMS (Obitrap) based plasma untargeted metabolomic profiles, established from a simple baseline blood draw from healthy men, and appropriate statistical models, could identify biomarkers, if any, associated with the risk of developing prostate cancer within the following decade. And which may use further to improve our understanding of the aetiology of this complex disease. In order to guide therapy decisions, improve outcomes and reduce overtreatment.

3.4.2 MATERIAL AND METHODS

3.4.2.1 Reagents and chemicals

HPLC grade solvent (acetonitrile, methanol) and formic acid were purchased from Merck KGaA (Darmstadt, Germany). Ultrapure water (H_2O) was collected from a Milli-Q system (Millipore, Billerica, MA). Internal standards (ISs) Carnitine C8:0-d3, Carnitine C16:0-d3, FFA C18:0-d3, CA-d4, CDCA-d5, Phe-d5 and Trp-d5 were purchased from Cambridge Isotope (Tewksbury, MA), FFA C16:0-d3 was purchased from C/D/N Isotopes Inc. (Pointe-Claire, Québec) and LPC 19:0 was supplied by Avanti Polar Lipids (Alabaster, AL).

3.4.2.2 Biological samples

A prospective nested case-control study was set up in the *Supplémentation en Vitamines et Minéraux Antioxydants* (SU.VI.MAX) cohort [209], [210], with selected plasma samples from 146 prostate cancer cases and 272 matched controls diagnosed during a 13-year follow-up. The SU.VI.MAX study (clinicaltrials.gov; NCT00272428) was initially designed as a double-blind placebo-controlled trial with the purpose of assessing the influence of

a daily supplementation with nutritional doses of antioxidants on the incidence of cardiovascular diseases and cancers.

The study design and methods have been previously detailed. Briefly, a total of 13 017 participants were enrolled in 1994-95 for an 8-year intervention trial and were followed up for health events until September 2007. The written informed consent for each participant was obtained. Participants were advised against taking any self-prescribed supplements during the trial. The study was conducted according to the Declaration of Helsinki guidelines and was approved by the Ethics Committee for Studies with Human Subjects of Paris-Cochin Hospital (CCPPRB 706/2364) and the 'Commission Nationale de l'Informatique et des Libertés' (CNIL 334641/907094) [209], [210]. At enrolment, all participants underwent a clinical examination by the study nurses and physicians, with anthropometric measurements and a blood draw, occurring after a 12-hour fasting period. Information on socio-demographics, smoking habits, physical activity, medication use and health status were collected through self-administered questionnaires. Age at menopause was collected from the participants by follow-up questionnaires.

A 35 mL venous blood sample was collected in sodium heparin Vacutainer tubes (Becton Dickinson, Rungis, France) from all fasting participants. After centrifugation at 4°C, plasma aliquots were immediately prepared and stored frozen at –20°C during less than 2 days and then stored in liquid nitrogen.

Health events were self-reported by the participants in regular follow-up questionnaires. Then, all relevant medical information and pathological reports were gathered through participants, physicians and/or hospitals and subsequently reviewed by an independent physician expert committee. Validated cases were classified according to the International Chronic Diseases Classification, 10th Revision (ICD-10) [211].

All participants with a first incident invasive prostate cancer, diagnosed between 1 year after their inclusion in the SU.VI.MAX cohort in 1994–95 and September 2007, were included in this nested case–control study (n=146). Incident prostate cancers diagnosed during the first year of follow-up were excluded to avoid reverse causality bias and guarantee the prospective design. For each case, two controls were randomly selected and matched for baseline age (45–49 years/50–54 years/55–59 years/>60 years), body

mass index (BMI) (underweight, normal weight and overweight/obese), intervention group of the initial SU.VI.MAX trial (placebo/supplemented), smoking status (current smokers and non-smokers) and season of blood draw (a priori-defined periods: October–November/December–January–February/March–April–May). The method for control selection was density sampling, i.e. every time a case was diagnosed, two controls were selected from other members of the cohort who, at that time, did not have diagnosed prostate cancer.

3.4.2.3 Samples Preparation

Plasma samples were randomized, balanced case and control, thaw on ice, then aliquots of each plasma sample were pooled as QC samples. For deproteinization and metabolites extraction, 150 µL of plasma was mixed with 600 µL methanol containing ISs, after vortexed for 2 min, centrifuged for 10 min at 16 000 g at 4 °C. The supernatant was lyophilized and then stock in -80°C before analysis, for quality control during sample preparation, a QC sample was prepared with every 10 plasma samples.

3.4.2.4 Data Acquisition

3.4.2.4.1 Analysis Equipment

An ACQUITY Ultra Performance Liquid Chromatography (UPLC, Waters Corporation, Manchester, U.K.) was coupled with a Q Exactive HF (Thermo Fisher Scientific, Rockford, IL) MS system.

3.4.2.4.2 Analyzes by LC/MS

The supernatant was re-dissolved in methanol/water (1:4, v/v) solvent before analysis. 5 µL re-dissolved supernatant was used for Metabolomic analysis, which was performed on a Q Exactive HF (Thermo Fisher Scientific, Rockford, IL) MS system coupled with an ACQUITY Ultra Performance Liquid Chromatography (UPLC, Waters Corporation, Manchester, U.K.). Column temperature and automatic sampler temperature were set at

60°C and 6°C, respectively. In order to cover as many types of compounds as possible, different columns were used in the positive and negative ionization mode.

For electrospray positive ion (ESI+) mode, BEH C8 (2.1 × 50 mm, Waters) column was used to ensure the separation of weakly polar compounds such as carnitine and lipids, the mobile phases were water (A) and acetonitrile (B) with 0.1% formic acid. The flow rate was 0.40 mL/min and the total run time was 12 min. The elution program started with 5% B and was held for 0.5 min, then linearly increased to 40% B at 2 min and increased to 100% B in 8 min, maintained 2 min, then went back to 5% B in 0.1 min and kept for 1.9 min for post equilibrium.

For electrospray negative ion (ESI-) mode, HSS T3 (2.1 × 50 mm, Waters) column was used to ensure the retention and separation of polar compounds in reverse phase, the mobile phases were water (A) and methanol/water (95:5, v/v) (B) containing 6.5mM Ammonium bicarbonate. The flow rate was 0.40 mL/min and the total run time was 12 min. The elution program started with 2% B and was held for 0.5 min, then linearly increased to 40% B at 2 min and increased to 100% B in 8 min, maintained 2 min, then went back to 2% B in 0.1 min and kept for 1.9 min for post equilibrium.

3.4.2.4.3 Analysis Sequences

The resolutions of full scan MS and $ddMS_2$ were set at 120 000 and 60 000, respectively. The automatic gain control (AGC) target and maximum injection time in full scan MS settings were 1×10^6 and 200 ms, while their values were 1×10^5 and 50 ms in $ddMS_2$ settings. The TopN (N, the number of top most abundant ions for fragmentation) was set to 10, and collision energy was set to 15 eV, 30 eV and 45 eV. A heated ESI source was used at positive and negative ion mode. The spray voltage was set as 3.5 kV. The capillary temperature and aux gas heater temperature were set as 300 and 350 °C, respectively. Sheath gas and aux gas flow rate were set at 45 and 10 (in arbitrary units), respectively. The S-lens rf level was 50.

3.4.2.5 Data Processing

Raw MS data were collected, to optimize MS data extraction, TraceFinder software (version 4.1, Thermo Fisher Scientific, Rockford, IL) and Compound Discoverer software (version 3.0, Thermo Fisher Scientific, Rockford, IL) were used for peak extraction and integration. 2 MS data matrixes were generated, with one from TraceFinder software, the other from Compound Discoverer software.

3.4.2.6 Statistical Analysis

Nonparametric analysis, Mann-Whitney U test, with Benjamini–Hochberg [212] base false discovery rate (FDR) correction were performed with Multiple Experiment Viewer (V_4_8_1_r2727_mac) for macOS [192]. Principal Component Analysis (PCA), Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) were performed using in-house MATLAB OPLS script based on Trygg and Wold method [162], analyses were performed with MATLAB® (R2016b) for macOS (Mathworks, Natick, Massachusetts, USA). Quality parameters of the models, the explained variance (R2Y) and the predictability of the model (Q2Y) were calculated. Q2Y was calculated by a 7-fold cross validation and confirmed by exploring the impact of permutations in the dataset rows [193]. To evaluate the discriminatory power of the model, the area under the receiver operating curve during the cross validation (CV-AUROC) and Confusion Matrix were calculated. The Confusion matrix was generated by a logistic regression, components in OPLS-DA model were used as variables, after a 7-fold cross validation, all samples have a predicted probability, and probability >= 0.5 consider as class 1, probability < 0.5 consider as class 0. Models were validated by permutation tests (n=200). VIP_{pred} of OPLS-DA model were used to identify potentials biomarkers.

Binary logistic regression analysis for biomarker selection was performed with IBM® SPSS® Statistics V25.0 for macOS (Armonk, NY: IBM Corp.). For variable selection, forward conditional selection method was used. MedCalc Statistical Software version 19.1 (MedCalc Software bv, Ostend, Belgium; https://www.medcalc.org; 2019) was used for ROC curve and the box-and-whisker plot (Tukey, 1977), ROC curve use the method of DeLong et al. (1988) [165], the Youden index J, is defined as: J = max (sensitivityc + specificityc - 1) where c ranges over all possible criterion values [213]. Equal weight is given to sensitivity and specificity. For box-and-whisker plot, outside and far out values are according to the original definitions of Tukey (1977) [214].

3.4.2.7 Identification

Identification was performed with Compound Discoverer software (version 3.0, Thermo Fisher Scientific, Rockford, IL), MS/MS experience in QC samples, in-house MS database, TraceFinder software (version 4.1, Thermo Fisher Scientific, Rockford, IL) and Human Metabolome Database (HMDB) (version 4.0).

3.4.3 RESULTS

3.4.3.1 UPLC-HRMS based Metabolomics analysis

There were 146 prostate cancer cases and 272 matched controls plasma samples analyzed with an UPLC-HRMS (Orbitrap) based metabolomics according to the procedure described in Material and Method.

For MS data matrix from TraceFinder software, 259 distinct metabolites (positive mode) were identified in the plasma samples. The changes in metabolites between prostate cancer cases and matched controls samples were calculated by the ratio of their group means (also called “Fold change”). The statistical significance of the changes was analyzed by Mann-Whitney U test, with Benjamini–Hochberg base FDR correction, $p < 0.05$ considered to be significant. After FDR correction, there were 18 metabolites significative different between cancer and control group. The complete list for detected mass with retention time used for metabolites identification is displayed in Annexe 1.

For MS data matrix from Compound Discoverer software, 323 distinct metabolites (positive mode) were identified in the plasma samples. After primary multivariate analysis with 2 MS data matrixes from TraceFinder software and Compound Discoverer software. There is an exploitable model with MS data matrixes from TraceFinder software, however, we haven’t found exploitable multivariate model yet with MS data matrixes from Compound Discoverer software, with the limitation of time during the PhD, for the follow analysis, we focused only on MS data matrixes from TraceFinder software, and we will exploit MS data matrixes from Compound Discoverer software in a later stage.

3.4.3.2 OPLS-DA model for metabolomics analysis

Plasma samples of 418 male participants with prostate cancer cases (n=146), matched control (n=272) from SU.VI.MAX cohort were partitioned randomly into 2 cohorts: estimation (70%, Cases: n=102 / Control: n=190) and validation (30%, Cases: n= 44 / Control: n= 82) cohorts, with an equal proportion of case/control.

After processing as procedure described in Material and Method, UPLC-HRMS (Orbitrap) data were used for multivariate statistical analysis. The PCA Score Plot of all analyzed plasma samples was represented in figure below (Figure 31).

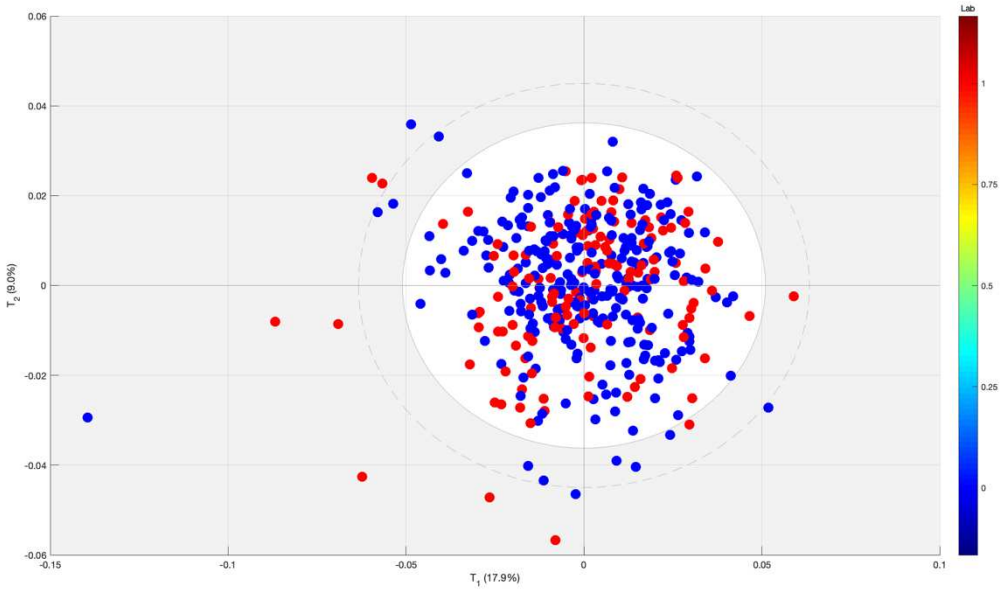


Figure 31. Score Plot of PCA for all analyzed plasma samples. Prostate cancer cases (red, n=146), matched control (bleu, n=272).

For Estimation cohort, OPLS-DA model for classification of prostate cancer cases and matched controls has been investigated. The Score Plot of OPLS-DA models for Estimation cohort was represented below (Figure 32).

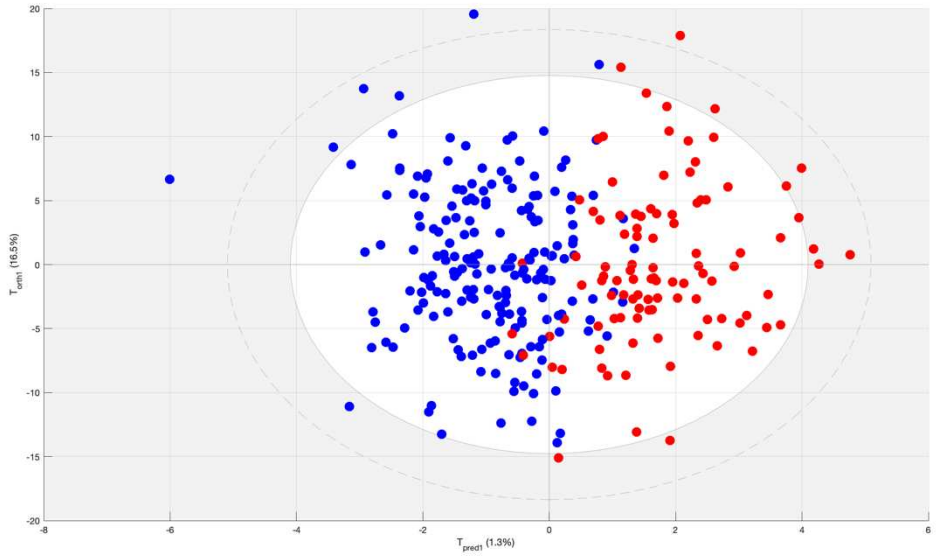


Figure 32. Score Plot of OPLS-DA models for Estimation cohort. Prostate cancer cases (red), matched controls (bleu). Components: 1+3; Q2cum: 0.387; Cases: n=102 / Control: n=190.

The OPLS-DA model classification performance was evaluated firstly by 999 times permutation test (Figure 33). A permutation test can assess whether the classification based on true sample class is significantly better than classification based on randomly assigned sample class. The principle of permutation test is to compare model outcomes between the classification based on true sample class (really Y) and the classification based on randomly assigned sample class (randomly assigned Y). In our study, the classification based on true sample class (really Y) is better than the classification based on randomly assigned sample class (randomly assigned Y), with intercepts: R2= 0.311, Q2 = -0.38, which means the OPLS-DA model is stable and not overfitting.

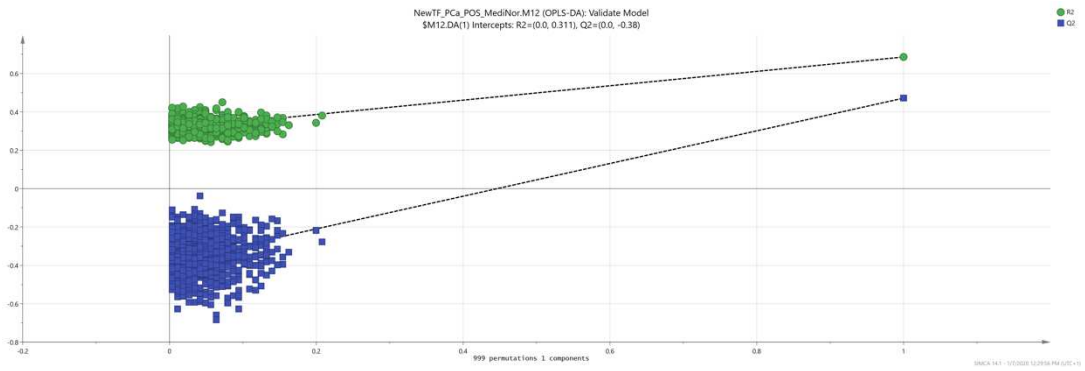


Figure 33. 999 times permutation test of OPLS-DA model for R2 (green) and Q2 (bleu).

OPLS-DA model classification performance was further evaluated by confusion matrix, which is a table that lists the correct and false predictions versus the actual observations, it is a good way to describe the performance of a classification model (classifier) when presented with new data. It allows easily identify confusion or error (like one class is misclassed as the other) of classification model when it makes predictions.

As represented in the Table 16, confusion matrix confirm that our OPLS-DA model classification performance is good, with high Accuracy (0.83) and low Class error (0.11 for matched control, 0.28 for cancer case).

Table 16. Confusion matrix of the sample by OPLS-DA in Estimation cohort.

	Matched control	Cancer cases	Accuracy	Class error
Matched control	**170**	20	0.83	0.11
Cancer cases	29	**73**		0.28

Rows represent the real groups; columns list the predicted groupings. Estimation cohort, Cancer cases: n=102 / Matched Control: n=190. 7-fold CV.

To highlight important metabolites in the OPLS-DA model, the Variable Importance in Projection (VIP) values from the predictive component (VIP_{pred}) was used, VIP value describes a quantitative estimation of the discriminatory power of each individual feature. Here, 13 metabolites ($VIP_{pred} > 2$) were selected as discriminant metabolites (Table 17).

Table 17. List of metabolites selected with VIP_{pred} of OPLS-DA.

Metabolites	**VIP_{pred}**
Sphingosine	3.99
Gly-Tyr	3.89
Sphinganine	3.28
Cis-11,14-Eicosadienoic acid	3.11
2-Hydroxy-6-aminopurine	2.68
Guanine	2.66
4-Acetamidophenol	2.65
Ethyl oleate	2.43
FFA C18:1	2.42
FFAD C18:1	2.36
Glutamic acid	2.23
Cis-8,11,14-Eicosatrienoic acid	2.07
Orthophosphoric acid	2.03

Logistic regression is a common and powerful regression method for binary classification problem, especially in epidemiology, which allow not only analyze multiple explanatory variables simultaneously, but also reducing the effect of confounding factors [215]. With these selected 13 metabolites, a binary logistic regression analysis was performed in our Estimation cohort (70%, Cancer cases: n=102 / Matched Control: n=190). 7 of these selected 13 metabolites were further selected as biomarker candidate (Table 18).

Table 18. List of variables in the logistic regression equation.

	B	S.E.	Sig.	Exp(B)
2-Hydroxy-6-aminopurine	0.000	0.000	0.000	1.000
Sphingosine	-0.001	0.000	0.000	0.999
Gly-Tyr	0.000	0.000	0.004	1.000
Cis-11,14-Eicosadienoic acid	0.001	0.000	0.016	1.001
Ethyl oleate	0.003	0.001	0.000	1.003
Orthophosphoric acid	0.000	0.000	0.000	1.000
Sphinganine	-0.001	0.000	0.003	0.999
Constant	-0.190	1.093	0.862	0.827

B: estimated logit coefficient, S.E.: standard error of the coefficient, Sig: significance level of the coefficient, Exp(B): odds ratio of the individual coefficient.

The logistic regression equation was:

p = 1/(1 + EXP(-y));
y =
-0.189676 + (-0.000029*(2-Hydroxy-6-aminopurine)) + (-0.000057*Gly-Tyr) +
(-0.000659*Sphinganine) + (-0.000546*Sphingosine) +
(0.001166*(Cis-11,14-Eicosadienoic acid)) +
(0.003000*Ethyl oleate) + (0.000037*Orthophosphoric acid);

With p: predicted probability.

To test these selected 7 metabolites in a new data set, the binary logistic regression model was applied in the validation cohort (Cases: n= 44 / Control: n= 82), the predicted probability ("p" in the equation above) for every sample was calculated, and represented in form of area under the ROC curve (Figure 34).

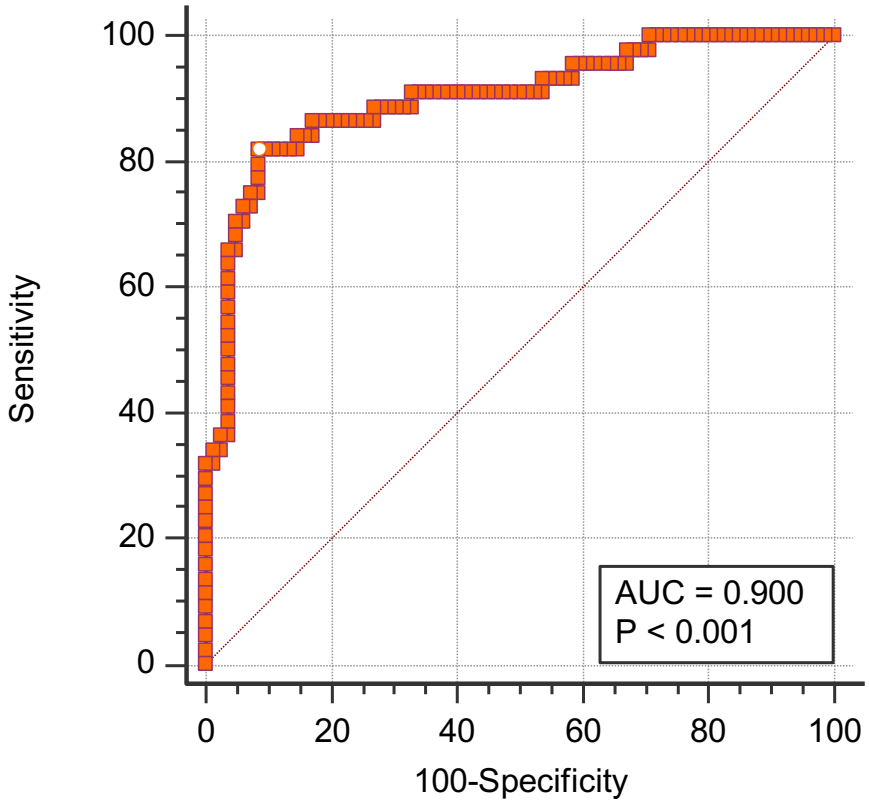


Figure 34. ROC curve for validation cohort. Cases: n= 44 / Control: n= 82, AUC: 0.900 (95% CI: 0.833 to 0.950), Youden index J: 0.733. Associated criterion: > 0.481 (Sensitivity: 81.82%; Specificity: 91.46%).

As represented in the figure above (Figure 34) , the area under the ROC curve is 0.9 (95% CI: 0.833 to 0.950), which means with the selected 7 metabolites, randomly select a cancer sample from all cancer samples, randomly select a control sample from all control samples, and then predict two random samples with our model. The probability of predicting cancer sample as cancer is p1, the probability of predicting control samples as cancer is p0, the probability of p1> p0 is 90%. This confirm the prediction power of our model is fairly good.

Moreover, with the OPLS-DA model from established with estimation cohort, we project samples of our validation cohort in the model. As represented in the (Figure 35), there is a clear discrimination of Cancer and Control groups in the projection, which confirm that our OPLS-DA model can well predict new data.

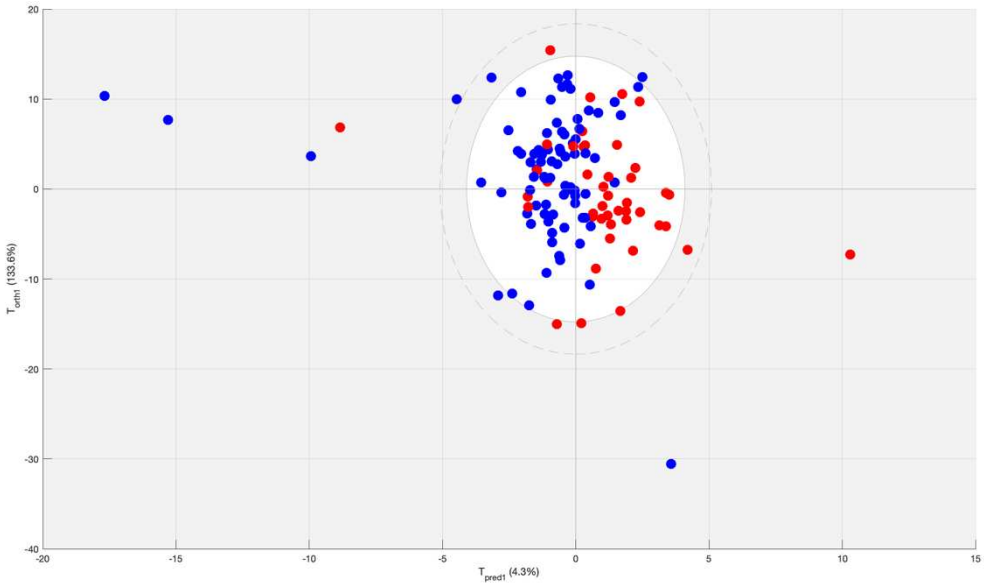


Figure 35. Projection validation cohort samples in estimation cohort OPLS-DA model. Validation cohort, prostate cancer cases: n= 44 (red), matched controls: n= 82 (bleu).

Furthermore, confusion matrix calculated with our OPLS-DA model for prediction of sample in validation cohort was generated (Table 19).

Table 19. Confusion matrix of the sample by OPLS-DA in validation cohort.

	Matched control	Cancer cases	Accuracy	Class error
Matched control	**74**	8	0.80	0.10
Cancer cases	17	**27**		0.39

Rows represent the real groups; columns list the predicted groupings. validation cohort (30%, Cases: n= 44 / Control: n= 82).

To sum-up, a scheme simplified different steps in identification and validation of putative Prostate cancer biomarkers in our study (Figure 36).

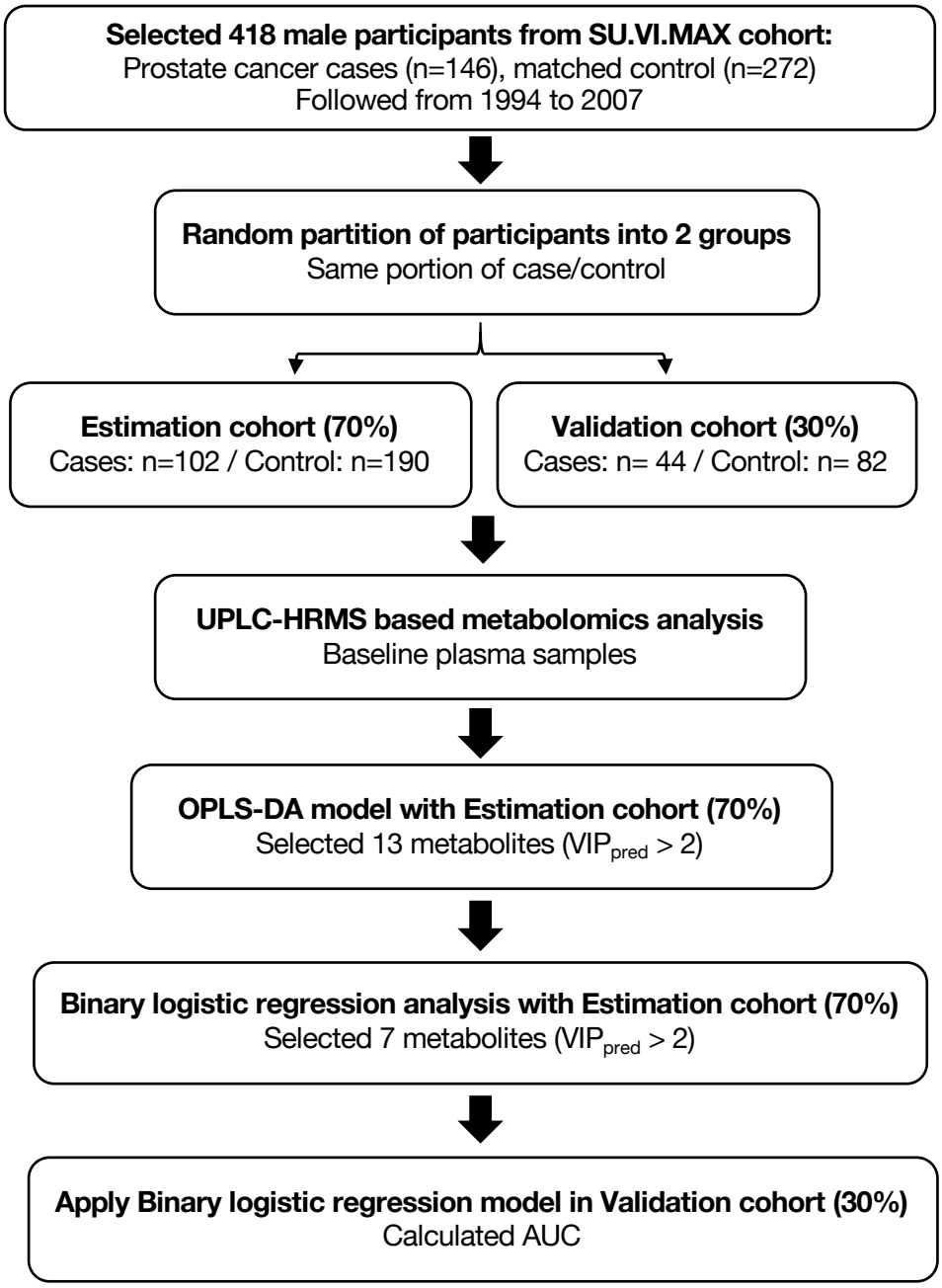
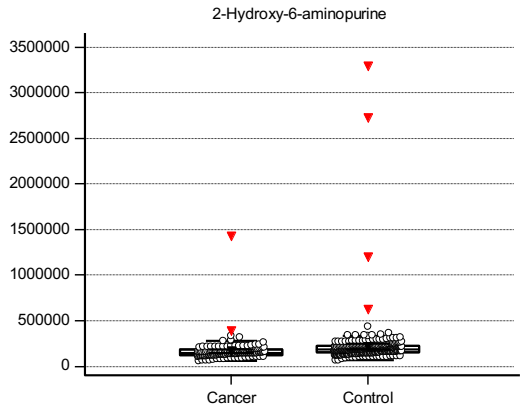
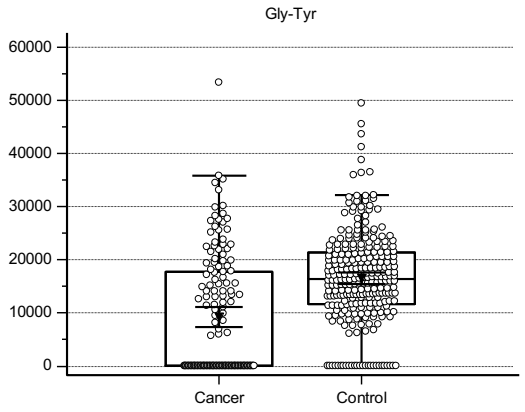


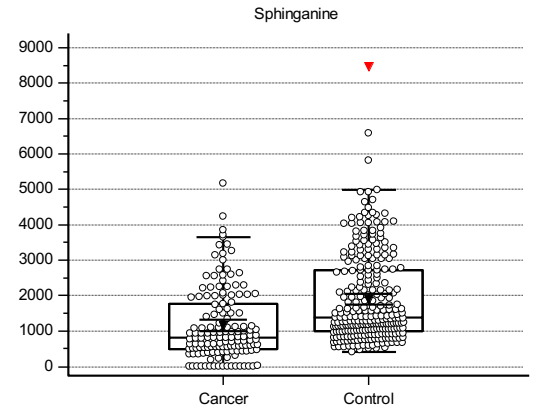
Figure 36. Scheme for the identification and validation of putative Prostate cancer biomarkers. Firstly, plasma samples of 418 male participants from SU.VI.MAX cohort with prostate cancer cases (n=146), matched control (n=272) were partitioned randomly into 2 cohorts: estimation (70%) and validation (30%) cohorts, with an equal proportion of cancer/control. Then, an OPLS-DA model for classification of prostate cancer cases and matched controls was established with Estimation cohort, and 13 metabolites were selected with $VIP_{pred} > 2$. After, a binary logistic regression analysis was performed in our Estimation cohort, 7 of these selected 13 metabolites were further selected as biomarker

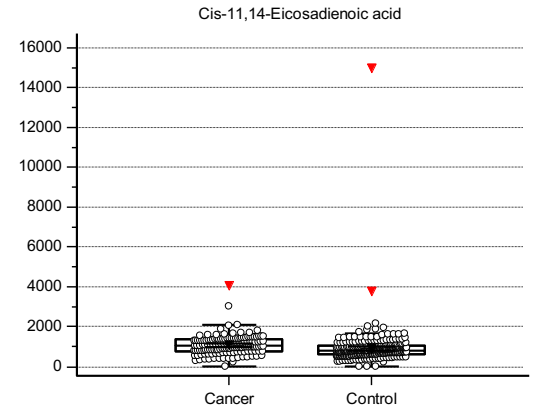
candidate. Finally, to test these selected 7 metabolites in a new data set, the binary logistic regression model was applied in the validation cohort, the probability for every sample was calculated, and represented in form of area under the ROC curve.

For, these selected 7 metabolites, they plasma relative change in the cancer and control group were represented below (Figures 37).









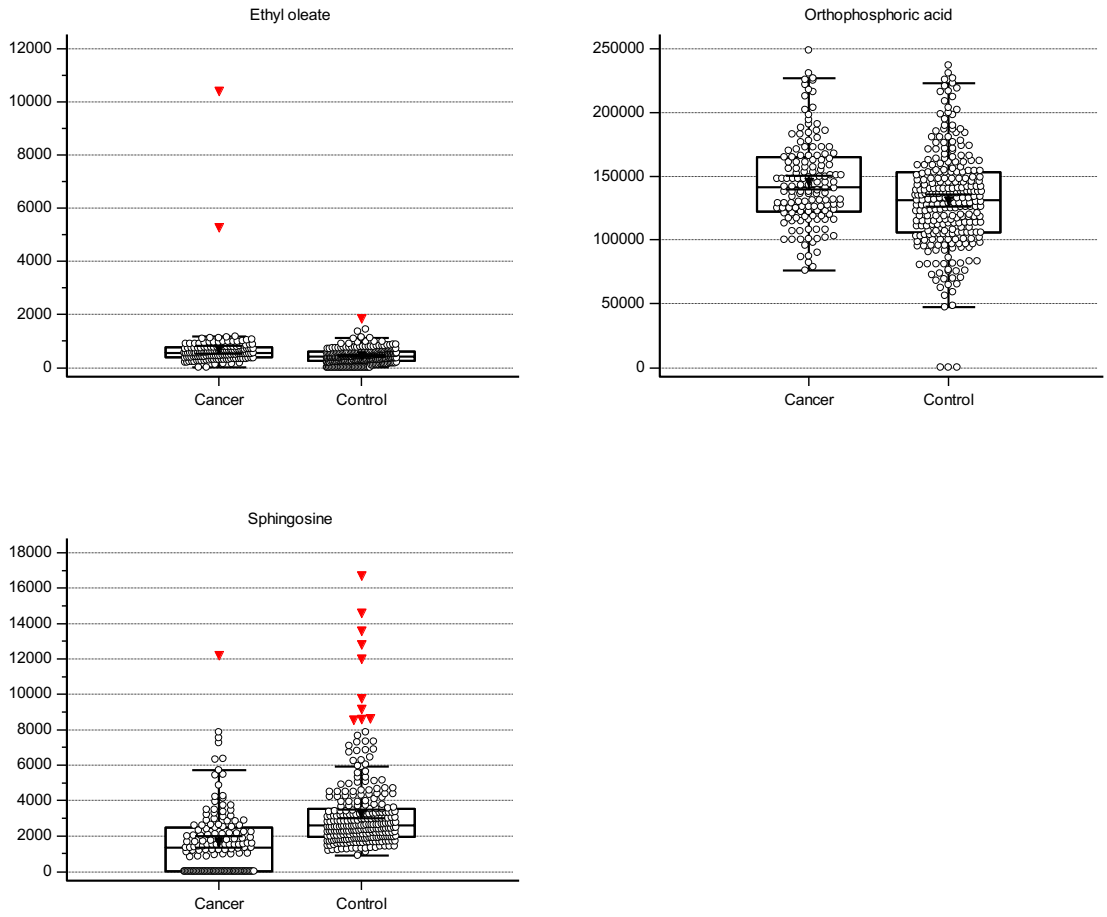


Figure 37. Box plots of plasma levels of selected 7 metabolites in prostate cancer and matched controls subjects. Median normalized areas are presented on the y-axis. The circle represents a sample, the red triangle represents a “far out” value.

3.4.4 CONCLUSION AND DISCUSSION

Prostate cancer (PCa) is the second most commonly diagnosed cancer and the second leading cause of cancer death among males [68]. There is no single definitive test to identify prostate cancer in men currently [69]. The exist tests like PSA test has limits such as sensitivity, specificity, and can lead to false-positive and false-negative results [69]. More sensitive and specific biomarkers are still very demanding in early detection, prognosis, monitoring, and clinical management of PCa patients [71]–[74].

The objective of our present study was to investigate whether UPLC-HRMS (Obitrap) based plasma untargeted metabolomic profiles, established from a simple baseline blood draw from healthy men, could identify biomarkers, if any, associated with the risk of developing prostate cancer within the following decade. UPLC-HRMS (Orbitrap) based plasma metabolomics analysis were performed with plasma samples of 418 male participants from a randomized, placebo-controlled trial SU.VI.MAX cohort [209].

To the best of our knowledge, this study was the first using a robustness 12 minutes UPLC-HRMS (Orbitrap) based metabolomics analysis to investigate the relationship between baseline plasma metabolites profiles and long-term prostate cancer risk in a large prospective male cohort. Our study revealed a panel of 7 metabolites (Table 18) which may useful for prediction the risk of prostate cancer decade before, with AUC: 0.900 (95% CI: 0.833 to 0.950), Sensitivity: 81.82%; Specificity: 91.46% in our validation cohort (Cases: n= 44 / Control: n= 82). Men characterized by higher fasting plasma levels of Cis-11,14-Eicosadienoic acid, Ethyl oleate and phosphoric acid had a higher risk of developing prostate cancer during the 13-year follow-up (Table 18, Figure 37).

Our result show promising advantage compared with currently used PSA testing, with a cutoff of 4.0 ng/mL has a sensitivity of 67.5-80%, the specificity of PSA at levels higher than 4.0 ng/mL is 60-70% [216].

The strengths of the present study were the combination of a robustness 12 minutes UPLC-HRMS (Orbitrap) based metabolomics analysis with a prospective cohort design and long follow-up. Nevertheless, our study has limits. First, males included in this study were age between 45-60 years, which may not fully represent all male population. Second, metabolomic analysis was performed with a single blood draw, the intra-individual variability of metabolomic profile over time was not controlled in our study. Third, the precision of LC-MS peak annotation was limited, especially for peaks which have a retention time less than 0.5 minute, as reverse phase UPLC column was used, these peaks correspond to high polarity metabolites, which not or less retained by stationary phase, thus, they retention time will be approximative, also, our in-house database used for metabolomics analysis may not cover other metabolites class which may have strong association with cancer risk. We will investigate latterly MS data matrix from Compound Discoverer software, which integrated a more complete database.

Moreover, validation of the panel of 7 metabolites will be necessary by using a targeted quantitative analysis in an independent prospective cohort.

In the following work, at first, we will interpret more profoundly these 7 metabolites in metabolism pathway, then by using data filtering [202] and variable selection approach such as sparse PLS-DA (sPLS-DA) [86], [203], [204], to investigate MS data matrix from Compound Discoverer software.

In conclusion, this prospective study suggested that UPLC-HRMS (Obitrap) based plasma untargeted metabolomic profiles, established from a simple baseline blood draw from healthy men, may identify biomarkers, that associated with the risk of developing prostate cancer within the following decade. Still, our preliminary promising findings should be validated in other independent prospective studies, to allow the identification of more robustness biomarkers, that associated with prostate cancer risk. After validated, our study may contribute to (1) develop early screening strategies to predict prostate cancer risk well before symptoms appear, to (2) improve our understanding of the aetiology of this complex disease. In order to guide therapy decisions, improve outcomes and reduce overtreatment.

3.5 Metabolomic studies of sepsis and septic shock

3.5.1 Introduction of sepsis and septic shock

Sepsis is a serious medical condition characterized by an exaggerated, uncontrollable immune response to an infection [217]–[219]. Which occurs in up to 30% of patients in intensive care units (ICUs) [220], [221]. The term “sepsis” is usually used to describe a progression of infection, with spectrum range from systemic inflammatory response syndrome (SIRS) to septic shock [217], which can result in multiple organ dysfunction syndrome (MODS) and death [222]. Currently, standard of care recommends aggressive appropriate antimicrobial therapy, which can led to drug-resistant [223], moreover, for patient management, there are no reliable biomarkers which can predict outcomes, aid clinical decision and direct more precise therapeutic intervention, thus, new approaches to more accurately phenotype sepsis is an urgent need.

Metabolomics is the science designed to comprehensively study the metabolome, the repertoire of small molecule metabolites, which has been used to investigate in prognosis, risk estimation, early diagnosis, and identification of novel biomarkers of sepsis. Metabolomics provides novel level of detail, highlighting specific biochemical pathways, by investigating metabolites changes in the pathophysiology of sepsis [224]–[226]. Accumulated results on metabolomics suggest that it is an important approach in prognosis, diagnosis, pathophysiology, and treatment of sepsis. Also, in complement with other systems biology approaches, such as transcriptomics, proteomics, to aid in defining specific sepsis phenotypes and to find novel predictive and prognostic biomarkers that can lead to more personalized management and therapeutics [223], [224], [226]. Still, independent prospective validation studies are needed to translate metabolomics findings into the clinical applications in sepsis.

3.5.2 Nuclear magnetic resonance-based serum metabolomic analysis reveals different disease evolution profiles between septic shock survivors and non-survivors

Liu *et al. Critical Care* (2019) 23:169
https://doi.org/10.1186/s13054-019-2456-z

Critical Care

RESEARCH Open Access

Nuclear magnetic resonance-based serum metabolomic analysis reveals different disease evolution profiles between septic shock survivors and non-survivors

Zhicheng Liu[1,2], Mohamed N. Triba[2], Roland Amathieu[2,3], Xiangping Lin[2], Nadia Bouchemal[2], Edith Hantz[2], Laurence Le Moyec[4] and Philippe Savarin[2*]

Abstract

Background: Septic shock is the most severe phase of sepsis and is associated with high rates of mortality. However, early stage prediction of septic shock outcomes remains difficult. Metabolomic techniques have emerged as a promising tool for improving prognosis.

Methods: Orthogonal projections to latent structures-discriminant analysis (OPLS-DA) models separating the serum metabolomes of survivors from those of non-survivors were established with samples obtained at the intensive care unit (ICU) admission (H0) and 24 h later (H24). For 51 patients with available H0 and H24 samples, multi-level modeling was performed to provide insight into different metabolic evolutions that occurred between H0 and H24 in the surviving and non-surviving patients. Relative quantification and receiver operational characteristic curves (ROC) were applied to estimate the predictability of key discriminatory metabolites for septic shock mortality.

Results: Metabolites that were involved in energy supply and protein breakdown were primarily responsible for differentiating survivors from non-survivors. This was not only seen in the H0 and H24 discriminatory models, but also in the H0-H24 paired models. Reanalysis of extra H0-H24 paired samples in the established multi-level model demonstrated good performance of the model for the classification of samplings. According to the ROC results, nine discriminatory metabolites defined consistently from the unpaired model and the H0-H24 time-trend change (Δ_{H24-H0}) show good prediction of mortality. These results suggest that NMR-based metabolomic analysis is useful for a better overall assessment of septic shock patients.

Conclusions: Dysregulation of the metabolites identified by this study is associated with poor outcomes for septic shock. Evaluation of these compounds during the first 24 h after ICU admission in the septic shock patient may be helpful for estimating the severity of cases and for predicting outcomes.

Trial registration: All human serum samples were collected and stored, provided by the "center of biologic resources for liver disease", in Jean Verdier Hospital, Bondy, France (BB-0033-00027).

Keywords: ^{1}H nuclear magnetic resonance spectroscopy, Metabolomics, Septic shock, Outcome prediction

* Correspondence: philippe.savarin@univ-paris13.fr
[2]Sorbonne Paris Cité, Laboratoire de Chimie, Structures et Propriétés de Biomateriaux et d'Agents Therapeutiques, UMR 7244, University Paris 13, F-93017 Bobigny, France
Full list of author information is available at the end of the article

Background

Septic shock is the most severe phase of sepsis [1, 2]. It is defined as sepsis complicated either by hypotension that is refractory to fluid resuscitation or by hyperlactacidemia and is often accompanied by acute organ failure. Mortality rates associated with septic shock are 20 to 30% in many series, principally due to multiple organ dysfunction syndrome (MODS) [3]. Common strategies for the treatment of septic shock include prompt initiation of therapy to treat the underlying infection with antibiotics, vasopressor therapy, and support for failing organs. In recent years, early goal-directed therapy (EGDT), which improves curative effect, has been extensively applied to improve rescue outcomes [4, 5]. However, early personalized prognosis and diagnosis remain challenging due to the complicated etiology and pathogenesis of septic shock. Determination of an acute prognosis in the early stage of sepsis is of great importance to improve therapeutic efficacy and will aid in the development of adapted strategies for different cases. In fact, evaluation of existing biomarkers (e.g., TNF-α, IL-6, and PCT) and clinical scores such as the sequential organ failure assessment (SOFA) [6] have been applied prognostically but their performance (sensitivity, specificity) has not proven adequate for all cases [7]. Thus, new methods for reliable early prognosis are still urgently needed.

Metabolomics has been proven to be a promising tool that aid in the prognosis of sepsis. This is because metabolomics allows to provide comprehensive information of personalized metabolome and therefore to enable the prediction of personalized outcome for septic patients. Previous studies have shown that there are considerable differences in the metabolome fingerprints between septic shock survivors and non-survivors. However, notably, most of the previous studies in human septic patients were designed to be performed by analysis of one unique sampling, and no studies have derived dynamic alterations of patient metabolomes during clinical therapy. However, good outcomes for septic shock are associated with a less severe disease course and a positive therapeutic response to treatment. In a previous study, we reported comprehensive differences in the metabolic profiles between septic shock survivors and non-survivors at the admission to the intensive care unit (ICU), based on a liquid chromatography-mass spectrometry (LC-MS) approach [8]. In this current study, samples from the septic shock patients which were obtained 24 h after ICU admission were also included. The aim of the present study was to analyze the discriminatory ability of metabolic profiles between septic shock survivors and non-survivors at the beginning and 24 h after ICU admission and also to describe the evolution of metabolic profiles for septic shock patients during this period by using ^{1}H NMR spectroscopy-based metabolomics.

Materials and methods

Patient inclusion

Between January 2009 and December 2011, all consecutive adults admitted to our intensive care unit were enrolled in this study if they had an indisputable or probable septic shock in the first 24 h after ICU admission [9]. Septic shock was defined as the presence of a clinically or microbiologically documented infection and on-going treatment with vasopressor therapy (norepinephrine or epinephrine at a dose ≥ 0.25 μg per kilogram of body weight per minute or at least equal to 1 mg per hour) for at least 6 h to maintain a systolic blood pressure of at least 90 mmHg or a mean blood pressure of at least 65 mmHg. Non-inclusion criteria were (i) patient younger than 18 years, (ii) patient with solid cancer or blood cancer, and (iii) patient with liver cirrhosis or chronic kidney disease. Patients were treated according to the international guidelines for the management of sepsis and shock septic [5].

Biological parameters, hemodynamic parameters, and the use of catecholamine and mechanical ventilation were recorded at inclusion. Cause of septic shock was recorded. To evaluate the severity of the disease, the Sequential Organ Failure Assessment (SOFA) score was calculated during the first day of admission [10]. ICU and hospital length of stay and mortality were recorded. The survival status of each patient was noted 7 days after the first sample.

Sample collection

All the first samplings (H0) were obtained withdrawn just before or immediately after clinical vasopressor therapy initiation on the patients. The second samples were withdrawn 24 h after the beginning of the vasopressor introduction. Blood samples were collected in serum separator tubes (SST). SST were stored for at least 30 min and not more than 1 h and 30 min. After centrifugation (1000×*g*, 25 °C, 10 min), the serum was stored at −80 °C. All human serum samples were collected and stored, provided by the "center of biologic resources for liver disease", in Jean Verdier Hospital, Bondy, France (BB-0033-00027). Written informed consent was obtained from all subjects or their surrogate decision-maker. The local ethics committee approved the protocol.

Regrouping and matching of samples

As shown in Fig. 1, 122 samples from 70 patients were obtained. Seventy samples were drawn at ICU admission and are noted as H0 samples. 52 samples were obtained 24 h after the first sampling and are noted as H24

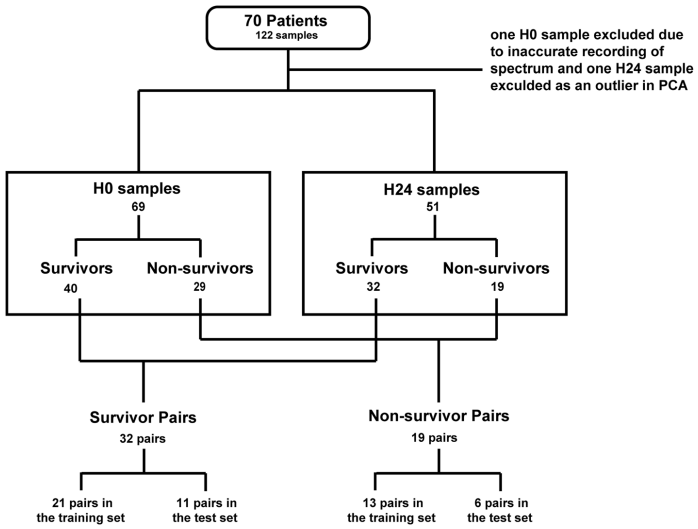


Fig. 1 Regrouping and matching of samples. One H24 sample from a survivor was excluded due to the problem of NMR gain parameter; one H0 sample from a non-survivor was excluded as it was found to be an outlier in PCA. For the paired H0-H24 samples obtained from the same patients, the pairs have been divided into training set and test set. The pairs of survivors and non-survivors were analyzed separately. Samples in the training set were analyzed for establishing discriminatory models between H0 and H24 samples. The pairs in the test set were reanalyzed in the established models

samples. During analysis, one H0 sample from a non-survivor who did not have a matching H24 sample was excluded as a spectral outlier. One H24 sample from a survivor was excluded while the H0 sample that belonged to the same patient was retained. The exclusion of the sample was due to the drastically affected NMR gain parameter. The spectrum of this sample was therefore found to be clearly different from the others. Among the non-survivors, 11 patients died prior to the H24 sampling and their H24 samples were therefore not available. For the other samples, each H24 sample was matched with the H0 sample which was collected from the same patient. In this case, 32 pairs for survivors and 19 pairs for non-survivors were obtained. For both septic shock survivor (SSS) and non-survivor (SSN) H0-H24 pairs, two thirds were randomly taken into the training set while the remaining were put into the test set.

Sample preparation and NMR data acquisition

Samples were defrosted at room temperature. A volume of 450 μL of each sample was diluted with 50 μL of D_2O in an NMR tube of 5 mm diameter. All the samples were then analyzed with a 500-MHz NMR spectrometer (Advance III, Bruker, Germany) at 297 K. The free induction decay (FID) signals were collected onto 64k data points, with a spectral width of 6000 Hz. The 1D ^{1}H NMR spectra were recorded by the Carr-Purcell-Meiboom-Gill (CPMG) sequence [11] with 128 transients for each spectrum. For several samples, 2D NMR experiments (TOCSY and JRES sequences) were achieved to confirm spectral assignments. The mixing time of the TOCSY spectra was 80 ms with 32 transients.

Data processing

After the FIDS were acquired for all the samples, they were processed using the NMRPipe software [12]. All FIDs were multiplied by a 0.3-Hz exponential line broadening factor prior to Fourier transformation. Phasing of each spectrum was manually adjusted, and baselines were corrected using a linear method. All the spectra were divided into 0.001 ppm buckets between −1 and 10 ppm. The residual water signal (4.6 to 5.5 ppm) was excluded, and the spectral region from 3.16 to 4 ppm was also removed since signals observed in this section represented the infusion of hydroxyethyl starch (HES), which was applied in the ICU to heighten blood tension for the patients who suffered from hypotension. The spectra were then normalized using the probabilistic quotient method [13]. All the buckets were centered by the method of auto-scaling. The peaks were adequately assigned using the Human Metabolome Database (HMDB, www.hmdb.ca) NMR library, the Chenomx

software (Chenomx Inc., Canada), and the 2D experiments. As to the annotation of NMR peaks, an exemplar NMR spectrum has been shown in Additional file 1: Figure S1 with some of the assignments.

Statistical analyses

All the multivariate analyses were achieved using an in-house code which is based on the code of Trygg and Wold [14], developed using Matlab software (version 2012b, MA, USA). Prior to the establishment of discriminatory analyses, a principal component analysis (PCA) with H0 samples from all the included non-survivors shows that the main variability among these samples does not correspond to the time of the death, as shown in Additional file 1: Figure S2. Another PCA for all the acquired spectra was performed to ensure that there were no outliers (Additional file 1: Figure S3). Orthogonal projections to latent structures-discriminant analyses (OPLS-DA) were performed for differentiating survivors from non-survivors with H0 and H24 samples, respectively. Samples obtained at H0 and H24 from the same patients were paired and analyzed in multilevel models to study the interindividual variability [15]. The paired samples were divided into survivor and non-survivor groups. Two OPLS-DA multilevel models were applied with the survivors and non-survivors, respectively, separating H0 from H24 samples. The models were all validated by cross-validation with 500 permutations of variable X and Y, where X represents the data matrix and Y represents the discriminatory variable for each model [15, 16]. For the univariate analyses analyzing the significant differences between two groups, the P values were calculated with Student's T test. The false discovery rate (FDR) was calculated by the Multiple Experiment Viewer Toll (version 4.9.0, OriginLab, Northampton, USA); the correction of P value is performed with "adjusted Bonferroni correction" [17]. The threshold of FDR was set at 0.1 for the screening of the discriminatory metabolites, that is, the variables with FDRs superior to 0.1 were not considered as important discriminants. A significant difference between compared groups was defined with an adjusted P value inferior to 0.05.

Results

Baseline characteristics of patients

The baseline biological characteristics of all the included patients are shown in Table 1. Partial pressure of arterial oxygen (PaO_2) and the ratio of PaO_2 to the percentage of inspired oxygen (FiO_2) were significantly different between the survivors and non-survivors at H0. Lactate level in non-survivors was also found to be significantly increased than those in survivors. For the clinical scores, SAPSII and SOFA, SAPSII was able to discriminate between survivors and non-survivors. However, due to the sample size, it was not satisfying to predict mortality with SOFA in this study, according to the results both at H0 and H24.

Discriminatory analyses separating septic survivors from non-survivors with samples drawn before treatment (H0)

For the H0 samples, a total of 69 samples were analyzed using an OPLS-DA model (PCA models separating the survivors from the non-survivors prior to the exclusion of the outlier have been illustrated in Additional file 1: Figure S3). Among these, 40 samples were obtained from survivors and 29 were from non-survivors. As is shown in Fig. 2a for H0, a clear separation between the two groups of patients is demonstrated by the score plot. The Q^2Y, which indicates the predictability of the model, was equal to 0.60 with three components, and the R^2Y, which indicates the fraction of explained variance of the Y variable, was 0.75, where Y corresponded to the survival condition in the model for the SSS vs. SSN comparison. Cross-validation showed that the model was not over-fitted (Additional file 1: Figure S4). In the loading plot (Fig. 2c), the peaks are colored according to the correlation coefficients, which relate to their contribution to the discriminatory model. Corresponding discriminatory metabolites have been listed in Table 2 with their chemical shifts, multiplicity, correlations, variance importance projections (VIPs), and P values. The concentrations of various amino acids such as alanine, glutamate, glutamine, methionine, and aromatic amino acids were increased in the non-survivors as compared to the survivors. Significant variations between the two groups were also found in energy-associated metabolites including two tricarboxylic acid (TCA) cycle intermediates, citrate and fumarate, and lactate and pyruvate. Ketone bodies, 3-hydroxybutyrate, and acetate were also elevated in the non-surviving patients. The only decreased signal was observed for the N-acetyl moieties of glycoproteins. Together, the results showed considerable differences in the metabolic profiles between the survivors and the non-survivors at H0.

Discriminatory analyses separating septic survivors from non-survivors with samples drawn 24 h after ICU admission (H24)

A second OPLS-DA model differentiating metabolic profiles of SSS from those of SSN were performed with 51 H24 samples. Of these samples, 19 non-survivors were compared with 32 survivors. As shown in Fig. 2b, a separation between SSS and SSN was observed. The R^2Y and Q^2Y values in the model were equal to 0.86 and 0.46, respectively, and were calculated with three components. The validation by permutations is shown in Additional file 1: Figure S5. Significant discriminant metabolites were identified referring to the loading plot

Table 1 Baseline characteristics of the patients recorded at admission to the ICU

	Total/average	Survivors	Non-survivors	Adj *P*	FDR
Number of patients	70	40	30		
Male (%)	40 (57%)	27 (67%)	13 (32%)	0.07	0.09
Age	70.1 ± 0.16	68.5 ± 0.29	72.1 ± 0.36	0.12	0.23
Temperature (°C)	37.3 ± 0.02	37.1 ± 0.03	36.9 ± 0.05	0.32	0.35
Mean arterial pressure (mmHg) at admission	72.8 ± 2,44	71,0 ± 3,25	75.2 ± 3.72	0.41	0.35
pH at admission	7.3 ± 0.00	7.31 ± 0.00	7.29 ± 0.00	0.43	0.53
Pa_{O2} $(mmHg)_{H0}$	144.7 ± 1.54	167.5 ± 2.83	113.6 ± 3.12	0.05	0.13
Pa_{CO2} (mmHg)	37.6 ± 0.18	37.6 ± 0.31	37.5 ± 0.45	0.66	0.63
Pa_{O2}/FiO_2 $ratio_{H0}$	212.5 ± 4.10	168.6 ± 3.19	242.2 ± 5.59	0.04	0.07
Lactate $(mmol/L)_{H0}$	5.0 ± 0.07	3.7 ± 0.11	6.6 ± 0.17	0.01	0.03
Creatininemia $(\mu mol/L)_{H0}$	212.2 ± 3.88	200.1 ± 4.34	241.9 ± 3.55	0.07	0.15
Glycemia (mmol/L)	9.8 ± 0.09	9.9 ± 0.20	9.6 ± 0.23	0.91	0.95
Hemoglobin (mmol/L)	10.5 ± 0.03	10.1 ± 0.06	10.9 ± 0.07	0.09	0.06
Albumin (g/L)	24.6 ± 0.13	22.2 ± 0.15	26.8 ± 0.38	0.11	0.21
Platelet (g/L)	156.7 ± 1.53	151.8 ± 2.61	162.2 ± 3.70	0.33	0.13
Total bilirubin (μmol/L)	38.4 ± 0.90	37.8 ± 1.90	39.1 ± 1.53	0.54	0.77
CRP (mg/dL)	162.3 ± 2.15	174.0 ± 3.98	147.4 ± 4.71	0.49	0.67
PCT (mg/dL)	25.2 ± 0.52	28.0 ± 1.15	21.6 ± 1.39	0.74	0.88
SAPSII	59.0 ± 0.24	55.2 ± 0.39	64.3 ± 0.58	0.02	0.07
$SOFA_{H0}$	11.7 ± 0.06	10.9 ± 0.11	12.4 ± 0.12	0.10	0.15
$SOFA_{H24}$	9.4 ± 0.06	8.7 ± 0.11	10.3 ± 0.10	0.07	0.11
ICU LOS (day)	9.16 ± 1.21	15.1 ± 1.36	3.62 ± 0.09	0.05	0.09
Mechanical ventilation (%)	84%	75%	93%		
Hospital-acquired infection (%)	45%	35%	60%		
Sepsis causes (%)					
Pulmonary	54%	55%	53%		
Abdominal	30%	22%	40%		
Urinary tract	7%	7%	6%		
Others	8%	15%	0%		

All the data is represented as mean ± standard error of mean (SEM)
PaO2 partial pressure of arterial oxygen, *FiO2* percentage of inspired oxygen, *SOFA* Sepsis-related Organ Failure Assessment, $SOFA_{H0}$ SOFA measured at H0, $SOFA_{H24}$ SOFA measured at H24, *SAPSII* new simplified acute physiology score, *LOS* length of stay, *FDR* false discovery rate, *Adj P P* value adjusted with Bonferroni correction

(Fig. 2d) and are listed in Table 3. Interestingly, increasing levels of most amino acids and energy-related metabolites, as well as the decreases of *N*-acetyl moieties of glycoproteins, were still detected in SSN, compared with SSS, in line with the findings of the H0 model. Besides, an increase in ketone bodies and diminishing lipid-related signals were only present at H24, but not at H0, in the non-survivors when compared to survivors. Both H0 and H24 unpaired models revealed extensive variations in the metabolic profiles between SSS and SSN at the admission and 24 h after ICU admission.

Discriminatory analysis of the evolution of septic shock from H0 to H24 for septic shock survivors and non-survivors

On the basis of the separation found between SSS and SSN within the above two discriminant models, we hypothesized that the therapeutic response between SSS and SSN could be different. To verify this hypothesis, SSS and SSN groups were compartmentalized and studied by two multi-level OPLS-DA models which focused on the intraindividual variability of the metabolome between H0 and H24. The pairwise distance of metabolome variations between H0 and H24 samples were analyzed for the patients for whom both H0 and H24 samples were collected.

For the SSS group, 21 pairs were randomly included to establish a model separating the H0 sample from the H24 sample, as shown in Fig. 3a. The Q^2Y was 0.78 with 2 components, and R^2Y was 0.94. This model was subsequently applied to 11 other pairs. The predictions for these pairs are shown in Fig. 3b. R^2Y for the prediction was 0.76, showing a prominent prediction of the classification among H0 and H24 samples. For the SSN model,

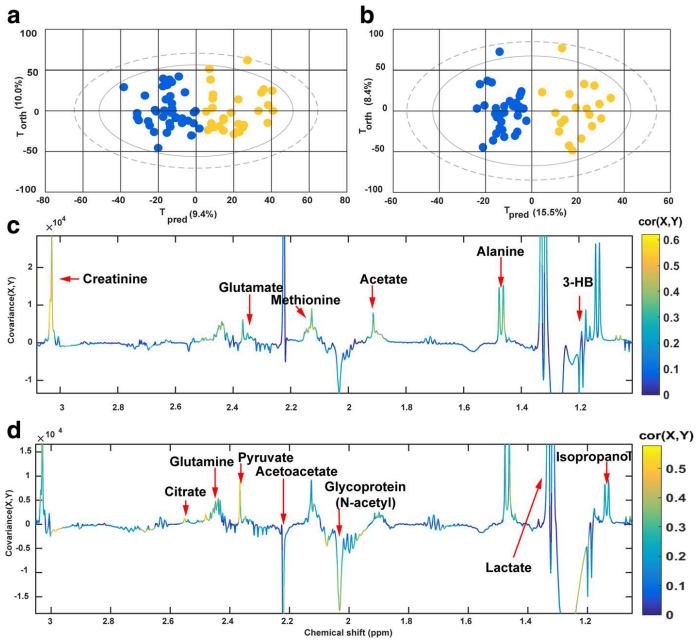


Fig. 2 OPLS-DA between septic shock survivors and non-survivors at H0 and H24. **a**, **b** Score plots for the H0 and H24 models, respectively. Blue dots represent the survivors and yellow dots represent the non-survivors. T_{pred}: The components that predict the differences between the groups; T_{orth}: components that do not predict the differences between the groups; **c**, **d** Loading plot for the H0 and H24 models, respectively. The color of the peaks indicates the correlation between the marked peak and the classification of the sample. Colors that are close to red correspond to a higher correlation. Positive peaks in the loading plot correspond to metabolites which increased in non-survivors; negative peaks correspond to metabolites that decreased in non-survivors

separation between H0 and H24 samples was also observed, as shown in Fig. 3c. Thirteen pairs were used to set up a training model, and 6 pairs were included in the test set. Consequently, R^2Y and Q^2Y were 0.57 and 0.91, respectively, and R^2Y for the reanalysis was 0.33 (as shown Fig. 3d). The loading plots for the two paired models are shown in Additional file 1: Figure S6. Metabolites which are listed in Table 4 exhibited opposite H0-H24 metabolome evolutions between SSN and SSS. Accordingly, increases of amino acids, energy-related metabolites, and creatinine and a decline of glycoprotein could be observed during the evolution from H0 to H24 for the non-survivors. However, this was not the case for survivors.

Discrimination between SSS and SSN based on the relative quantification of key discriminators

Spectral signals corresponding to the metabolites in Table 4 were integrated for the spectra of paired samples. As shown in Table 4, the molecules varied oppositely during the H0-H24 evolution between the survivors and non-survivors. The time-trend change of area, ΔSignal $area_{H24\text{-}H0}$, was calculated for each metabolite. Average values for ΔSignal $area_{H24\text{-}H0}$ of involved metabolites resulting from the SSN model were compared to those from the SSS model, as shown in Fig. 4a. Interestingly, the metabolites were also shown to be discriminant variables in the comparison between SSS and SSN in previously mentioned H0 and H24 unpaired models. We further calculated the area under the ROC curve for the metabolites in order to test their performance in the classification of surviving patients. ROCs for the discriminant metabolites in the H0 and H24 models, as well as for $\Delta_{H24\text{-}H0}$, were performed and are shown in Table 5. Accordingly, based on our data, most ROCs for the metabolites showed slightly better performance in the classification of survival than SAPSII and SOFA, not only within the H0 and H24 models, but also with the value of $\Delta_{H24\text{-}H0}$.

Table 2 Metabolites found to discriminate between SSS and SSN at H0

Peaks	Assignment	VIP	Correlation	Adj *P*	FDR
1.06^{d}	3-Hydroxyisobutyrate	3.06	0.52	0.0001	0.0001
5.79^{s}	Urea	2.64	0.45	0.002	0.003
7.31^{m} 7.36^{m}	Phenylalanine	2.64	0.44	0.01	0.01
2.12^{m}, 2.32^{m}	Glutamate	2.6	0.44	0.02	0.02
2.43^{m}	Glutamine	2.55	0.43	0.03	0.01
3.03^{s}	Creatinine	2.43	0.41	0.03	0.04
1.32^{d} 4.11^{q}	Lactate	2.38	0.4	0.02	0.04
2.14^{s}	Methionine	2.17	0.37	0.06	0.05
1.46^{d}	Alanine	2.12	0.26	0.07	0.08
6.88^{d} 7.18^{d}	Tyrosine	2.02	0.34	0.03	0.04
2.36^{s}	Pyruvate	2.01	0.34	0.03	0.01
2.52^{d} 2.62^{d}	Citrate	1.94	0.33	0.03	0.04
1.7^{m}	Lysine	1.91	0.27	0.09	0.08
6.52^{s}	Fumarate	1.9	0.32	0.04	0.05
7.67^{s}	1-Methylhistidine	1.66	0.28	0.07	0.06
2.03^{s}	Glycoprotein (*N*-acetyl)	1.64	− 0.28	0.08	0.03
1.91^{s}	Acetate	1.56	0.25	0.09	0.10
1.16^{d}	Isopropanol	1.53	0.26	0.09	0.03

Chemical shifts for the assigned metabolites are shown in the peak column. The superscripts for the peaks represent the multiplicity of the peaks. s, singlet; d, doublet; t, triplet; q, quadruplet; m, multiplet. A positive correlation indicates an increased level of the metabolite in the non-survivor while negative correlation indicates a decreased level of the metabolite. The threshold of FDR was set at 0.1. Similar expressions are also applied for Tables 3 and 4

Adj P P values that are calculated by Student's *T* test are adjusted with Bonferroni correction, *FDR* false discovery rate

Discussion

Effective prognosis can help to improve outcomes for septic shock patients. However, septic shock prognosis can be complicated by patient-specific factors that affect responsiveness to therapy. With the use of metabolomic techniques, we have determined the serum metabolome fingerprint of septic shock patients with both H0 and H24 samples. We also investigate the metabolic footprint along with the evolution from H0 to H24. To our knowledge, our study is the first to reveal time-trend metabolic differences using NMR-based metabolomics between septic shock survivors and non-survivors within 24 h after ICU admission.

Metabolic variations for H0 and H24 unpaired models separating SSS from SSN

The H0 and H24 unpaired models reveal the differences of metabolome fingerprint between SSS and SSN at the admission to ICU and at 24 h after the admission. Regarding the common discriminatory metabolites found in both models, consistent increases in energy-related metabolites, creatinine, 1-MH, and several amino acids, as well as decreases in glycoproteins are observed as important signals in the non-survivors. Such variations found in SSN at both H0 and H24 are likely to reflect more severe sepsis-induced inflammatory responses and organ dysfunctions that contribute to poor outcomes.

The deregulation of TCA cycle intermediates, such as more concentrated citrate found in the SSN, is one of the consequences of severe stress induced by sepsis [18]. Stress also results in an unregulated catabolism [19]. Enhanced degradation of glycoproteins indicates an aggravated stress in the non-survivors. Also, increases in various amino acids and ketone bodies at H24 in the

Table 3 Metabolites found to discriminate between SSS and SSN at 24 h after admission to ICU

Peaks	Assignment	Correlation	VIP	Adj *P*	FDR
2.37^{s}	Pyruvate	0.52	3.55	0.0001	0.0001
2.52^{d} 2.62^{d}	Citrate	0.52	3.5	0.0002	0.0003
7.31^{m} 7.36^{m}	Phenylalanine	0.48	3.25	0.001	0.001
6.88^{d}, 7.18^{d}	Tyrosine	0.45	3.03	0.004	0.004
2.72^{m}	Lipids (fatty acid residues)	− 0.44	2.99	0.01	0.01
2.43^{m}	Glutamine	0.44	2.96	0.01	0.01
1.32^{d}, 4.41^{q}	Lactate	0.44	2.9	0.01	0.02
1.06^{d}	2-Hydroxyisovalerate	0.41	2.77	0.02	0.03
3.03^{s}	Creatinine	0.37	2.51	0.03	0.05
2.03^{s}	Glycoprotein (*N*-acetyl)	− 0.37	2.48	0.05	0.05
7.03^{s} 7.67^{s}	1-Methylhistidine	0.35	2.37	0.06	0.07
2.12^{m},2.33^{m}	Glutamate	0.33	2.22	0.07	0.09
1.7^{m}	Lysine	− 0.29	1.95	0.09	0.07
1.46^{d}	Alanine	0.28	1.9	0.13	0.10

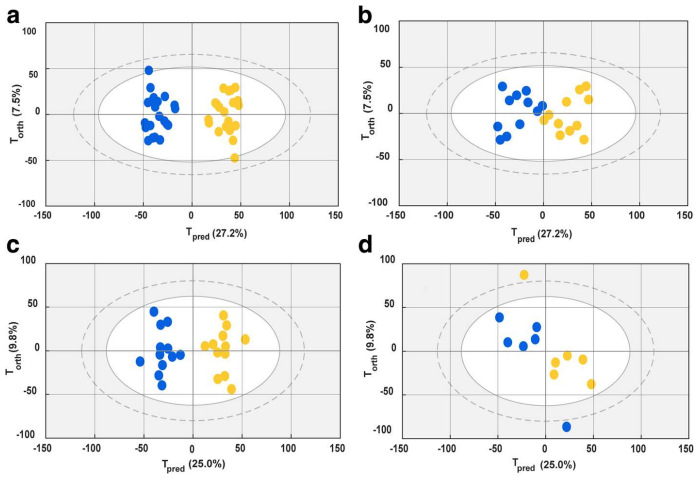


Fig. 3 Score plots of OPLS-DA separating H0 from H24. For the patients whose H0 and H24 are both available, their H0 and H24 samples are matched in the discriminatory models. The pairs from the survivors and non-survivors are analyzed in two separated paired models. Blue dots represent the H24 samples and yellow dots represent the H0 samples. **a** Score plot for the training set separating H0 from H24 for the survivors. **b** Reanalysis of test set samples of survivors in the established H0-H24 multi-level model. **c** Score plot for the training set separating H0 from H24 for the non-survivors. **d** Reanalysis of test set samples of non-survivors in the established H0-H24 multi-level model

non-survivors are known as effects of protein breakdown and enhanced lipid oxidation [20]. Notably, ketone bodies have recently been reported to be immune suppressors [21], and elevations in these metabolites may contribute to a negative response in critical illness [22].

Our results also provide evidence for the metabolic variations that are associated with severe organ dysfunction in non-survivors. As shown in Table 1, significantly lower oxygen pressures with higher blood lactate levels indicate the presence of more severe hypoxia in the non-survivors than in survivors. This may be due to mitochondrial disorder, defective TCA cycle [23], which results in dampened aerobic respiration and abnormal energy supply. Severe disorders in energy supply should be an import factor inducing organ failure [24–26]. Other variations involving organ dysfunction are found in creatinine and 1-MH. Their elevations in the comparison in SSN are also supported by some other previous studies [27, 28].

Table 4 Discriminatory metabolites with different variations along the H0-H24 evolution between the non-survivor group and the survivor group

Peaks	Assignment	C_1	Adj P_1	FDR_1	V_1	C_2	Adj P_2	FDR_2	V_2
2.12^{m} 2.32^{m}	Glutamate	− 0.62	0.0001	0.001	↓	0.49	0.03	0.02	↑
2.52^{d} 2.66^{d}	Citrate	− 0.59	0.0001	0.001	↓	0.59	0.002	0.004	↑
7.32^{d} 7.36^{d}	Phenylalanine	− 0.53	0.0004	0.003	↓	0.4	0.08	0.07	↑
2.07^{m} 2.43^{m}	Glutamine	− 0.49	0.001	0.01	↓	0.44	0.03	0.03	↑
1.47^{d}	Alanine	− 0.42	0.004	0.03	↓	0.42	0.05	0.06	↑
1.32^{d} 4.11^{q}	Lactate	− 0.62	0.0001	0.002	↓	0.2	0.26	0.21	NS
2.37^{s}	Pyruvate	− 0.38	0.04	0.05	↓	0.06	0.42	0.33	NS
2.03^{s}	Glycoprotein (*N*-acetyl)	0.26	0.08	0.09	NS	− 0.48	0.01	0.01	↓
3.02^{s}	Creatinine	− 0.21	0.15	0.13	NS	0.47	0.02	0.03	↑

C_1, correlation of the metabolite to the discriminatory model for the survivors; C_2, correlation of the metabolite to the discriminatory model for the non-survivors. For each listed metabolite, the sign of C_1 is opposite to that of C_2; Adj P_1, adjusted *P* value (with Bonferroni correction) of the metabolite in the comparison between H0 and H24 samples for the survivors; Adj P_2, adjusted *P* value of the metabolite in the comparison between H0 and H24 samples for the non-survivors; FDR_1, false discovery rate for the *P* value calculated with the survivors; FDR_2, false discovery rate for the *P* value calculated with the non-survivors; V_1, variation in concentration for the metabolites from H0 to H24 for the survivors; V_2, variation in concentration for the metabolites from H0 to H24 for the non-survivors; ↑, increased concentration of the metabolite at H24 compared with H0; ↓, decreased concentration of the metabolite at H24 compared with H0. NS, non-significant (Adj $P > 0.05$) variation

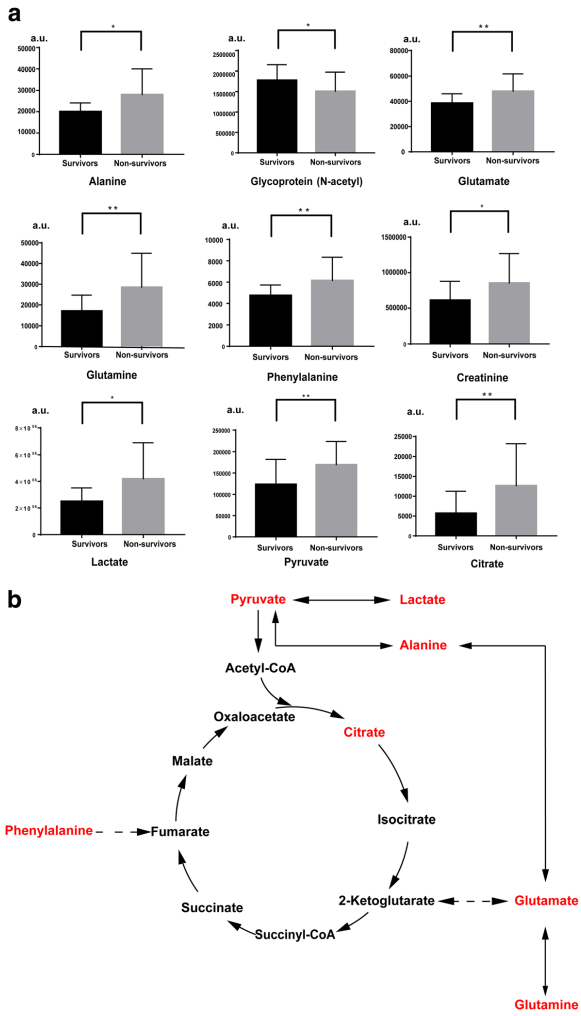


Fig. 4 Levels of key discriminatory metabolites and their relevant metabolic pathway in the comparison between SSS and SSN patients during the H0-H24 evolution. **a** Levels of key discriminatory metabolites in the SSS and in SSN. The levels of the metabolites are calculated with the average of time-trend change ($\Delta_{H24\text{-}H0}$). Averages of $\Delta_{H24\text{-}H0}$ for the survivors and non-survivors have been respectively shown and calibrated by the standard deviation. *: $p < 0.05$. **: $p < 0.01$. a.u.: arbitrary unit; **b** relevant metabolic pathway for energy-related metabolites and amino acids that vary differentially between SSS and SSN in the H0-H24 evolution. Metabolites marked by red color are those that increase in the SSN compared to SSS in all the models. Solid flashes express a direct conversion between two metabolites and dotted lines represent undirect conversions between two metabolites, according to KEGG metabolic pathway database

Different variations for some key discriminators between SSS and SSN H0-H24 multilevel models

As shown in Table 4, different pairwise alterations of relevant metabolites in the comparison between SSS and SSN groups indicate distinct trends in development along with clinical therapy. Interestingly, most of these metabolites related to septic shock evolution are in accordance with the discriminatory molecules found with

Table 5 Area under ROC for key metabolites that separate septic shock survivors from non-survivors

	$AUROC_{H0}$ ($n = 69$)	$AUROC_{H24}$ ($n = 51$)	$AUROC_{H24\text{-}H0}$ ($n = 51$)
Lactate	0.74	0.75	0.73
Alanine	0.78	0.78	0.67
Glycoprotein (*N*-acetyl)	0.71	0.60	0.65
Glutamate	0.61	0.81	0.71
Glutamine	0.80	0.70	0.74
Pyruvate	0.81	0.83	0.79
Citrate	0.82	0.72	0.72
Creatinine	0.79	0.69	0.70
Phenylalanine	0.84	0.73	0.79
SOFA	0.60	0.64	0.61
SAPSII	0.62		

n number of patients

the H0 and H24 unpaired models. Besides the deregulation of energy-related molecules, increases in four amino acids may be associated with severe protein breakdown and muscle wasting for the non-survivors. Notably, serum concentrations in some amino acids, such as alanine, glutamate, and phenylalanine, are otherwise documented to be involved with hemolysis associated with sepsis [29]. Moreover, as shown in Fig. 4b, glutamate is known to be a core amino acid for conversion into TCA cycle intermediates [30]. Its elevation, as well as the elevation of related amino acids such as glutamine and alanine, is associated with increases in citrate. Phenylalanine can be converted into fumarate. Increases in phenylalanine in patients with poor outcomes have been also reported in other studies [31–33]. The conversion from amino acids to TCA cycle intermediates is likely to provide supplementary energy during severe anoxic conditions, however, is detrimental for the outcomes [34]. Creatinine is known to be an important indicator for monitoring renal injury. Decreases in *N*-acetyl glycoproteins may correspond to a breakdown of proteins. Interestingly, it has been demonstrated by DeCoux et al. [35] that inflammation-induced enriched extracellular glycoproteins are associated with an optimal response during septic shock.

The present study not only provides support for the findings in our previous work, which investigated metabolic differences between SSS and SSN at H0 [8], but also reveals that different evolutions in the first 24 h after admission to ICU between septic shock survivors and non-survivors are linked to variations in metabolites identified by this study. The ROC results shown in Table 5 also show that the key metabolite discriminators are good classifiers for separating SSN from SSS during the first 24 h after ICU admission. We have reason to believe that sustained enrichment of energy-related metabolites and amino acids can provide early warning of a bad outcome.

Conclusion

In the present study, we have investigated metabolic differences between the survivors and non-survivors of septic shock with the samples obtained at ICU admission and with those obtained 24 h later. We have provided evidence that the sustained enrichment of energy-supply metabolites and amino acids is predictive of a bad outcome. We suggest that monitoring the relevant metabolites in the first 24 h may help to evaluate early therapeutic response.

Additional files

Additional file 1 Table S1. Assignment of spectra recorded with one exemplar serum sample from a septic shock patient. Figure S1. Assignment of spectra recorded with an example of a representative ^{1}H-NMR spectrum. The assigned peaks corresponding with the key metabolite discriminants have been marked in the figure. Figure S2. A PCA calculated with H0 samples from 11 nonsurvivors who died during the first 24 h (red dots) and those from the other non-survivors who died from the second day to the seventh day after the first sampling (blue dots). Figure S3. PCA model separating survivors from non-survivors with H0 samples before the exclusion of outlier. One sample of a non-survivor was observed as an outlier for the PCA. This outlier has been removed before statistical analyses. Blue dots: survivors, yellow dots: non-survivors. Figure S4 (respectively S5). Cross-validation by 200 times permutation between *X* and *Y* for the OPLS-DA model with H0 samples (respectively H24). The green dots stand for the obtained R^2 value and the blue dots stand for the obtained Q^2 value within the 200 permutations. The *Y*-axis represents R^2 and Q^2 calculated for every model while the *X*-axis represents the correlation coefficient between original and permuted response data. Figure S6. Loading plots for paired OPLS-DA models showing important discriminatory metabolites that contribute to the separation between H0 and H24 samples. The paired models for the survivors and non-survivors are shown separately. The peaks are assigned to corresponding discriminatory metabolites. The correlations between the assigned metabolites and the model have been shown with the colors. a: loading plot for the separation between H0 and H24 for the survivors; b: loading plot for the separation between H0 and H24 for the non-survivors. (DOCX 829 kb)

Abbreviations

AUROC: Area under receiver operation curve; FDR: False discovery rate; HB: Hydroxybutyrate; HES: Hydroxyethyl starch; ICU: Intense care unit; JRES: J-resolved spectroscopy; MH: Methylhistidine; MODS: Multi-organ dysfunction syndrome; NMR: Nuclear magnetic resonance; OPLS-DA: Orthogonal projections to latent structures-discriminant analysis; PCA: Principal component analysis; PCT: Procalcitonin; SAPSII: New Simplified Acute Physiology Score; SOFA: Sequential organ failure assessment; SSN: Septic shock non-survivors; SSS: Septic shock survivors; TCA: Tricarboxylic acid; TOCSY: Total correlation spectroscopy

Acknowledgements

We thank Dr. Pierre Nahon for his help with sample collection and for his constructive suggestions for the study design. We also would like to express our appreciation to Agnes Victor Bala for her help in the assignment of NMR spectra. We thank Sandy Field, PhD, for medical editing assistance. We acknowledge the NMR-PF facility (University Paris 13 France).

Funding
This study is supported by the funding of Anhui Medical University (XJ201729) and by the funding from the University of Paris 13.

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions
This study was designed by PS, RA, and LLM. ZL participated in conducting the NMR experiments, data analysis, and paper writing. MT provided the programming for the in-house program and for the data analysis. RA conducted sample collection and sample storage. XL, NB, and EH helped with the NMR experiments. All authors participated in the review and revision of the paper. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Written informed consent was obtained from all subjects or their surrogate decision-maker. The local ethics committee approved the protocol.

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]School of Pharmacy, Anhui Medical University, Hefei, China. [2]Sorbonne Paris Cité, Laboratoire de Chimie, Structures et Propriétés de Biomateriaux et d'Agents Therapeutiques, UMR 7244, University Paris 13, F-93017 Bobigny, France. [3]Intensive Care Unit, Diaconesse-Croix Saint-Simon Hospital, 125 rue d'Avron, 75020 Paris, France. [4]Université Paris Saclay, University Evry, UBIAE EA 7362, 91025 Evry, France.

Received: 29 November 2018 Accepted: 25 April 2019
Published online: 14 May 2019

References

1. Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, Schein RM, Sibbald WJ. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Chest. 1992;101(6):1644–55.
2. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, Cohen J, Opal SM, Vincent JL, Ramsay G, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Intensive Care Med. 2003;29(4):530–8.
3. Balk RA. Systemic inflammatory response syndrome (SIRS): where did it come from and is it still relevant today? Virulence. 2014;5(1):20–6.
4. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, Peterson E, Tomlanovich M. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med. 2001;345(19):1368–77.
5. Dellinger RL, Levy MM, Carlet JM, Bion J, Parker MM, Jaeschke R, Reinhart K, Angus DC, Brun-Buisson C, Beale R. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. Crit Care Med. 2008;36:296–327.
6. Vincent JL, Moreno R, Takala J, Willatts S, DeMendonca A, Bruining H, Reinhart CK, Suter PM, Thijs LG. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. Intensive Care Med. 1996;22(7):707–10.
7. Lee JW, Yamamoto T, Uchikata T, Matsubara A, Fukusaki E, Bamba T. Development of a polar lipid profiling method by supercritical fluid chromatography/mass spectrometry. J Sep Sci. 2011;34(24):3553–60.
8. Liu Z, Yin P, Amathieu R, Savarin P, Xu G. Application of LC-MS-based metabolomics method in differentiating septic survivors from non-survivors. Anal Bioanal Chem. 2016;408(27):7641–9.
9. Annane D, Bellissant E, Cavaillon J-M. Septic shock. Lancet. 2005;365(9453):63–78.
10. Vincent JL, de Mendonca A, Cantraine F, Moreno R, Takala J, Suter PM, Sprung CL, Colardyn F, Blecher S. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Crit Care Med. 1998;26(11):1793–800.
11. Carr HY, Purcell EM. Effects of diffusion on free precession in nuclear magnetic resonance experiments. Phys Rev. 1954;94(3):630.
12. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995;6(3):277–93.
13. 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. 2006;78(13):4281–90.
14. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). J Chemom. 2002;16(3):119–28.
15. Westerhuis JA, van Velzen EJ, Hoefsloot HC, Smilde AK. Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. Metabolomics. 2010;6(1):119–28.
16. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, van Velzen EJJ, van Duijnhoven JPM, van Dorsten FA. Assessment of PLSDA cross validation. Metabolomics. 2008;4(1):81–9.
17. Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage. 2002;15(4):870–8.
18. Mehta NN, McGillicuddy FC, Anderson PD, Hinkle CC, Shah R, Pruscino L, Tabita-Martinez J, Sellers KF, Rickels MR, Reilly MP. Experimental endotoxemia induces adipose inflammation and insulin resistance in humans. Diabetes. 2010;59(1):172–81.
19. Preiser JC, Ichai C, Orban JC, Groeneveld AB. Metabolic response to the stress of critical illness. Br J Anaesth. 2014;113(6):945–54.
20. Pawlak M, Baugé E, Lalloyer F, Lefebvre P, Staels B. Ketone body therapy protects from lipotoxicity and acute liver failure upon Ppar α deficiency. Mol Endocrinol. 2015;29(8):1134–43.
21. Newman JC, Verdin E. Ketone bodies as signaling metabolites. Trends Endocrinol Metab. 2014;25(1):42–52.
22. Mira JC, Gentile LF, Mathias BJ, Efron PA, Brakenridge SC, Mohr AM, Moore FA, Moldawer LL. Sepsis pathophysiology, chronic critical illness, and persistent inflammation-immunosuppression and catabolism syndrome. Crit Care Med. 2017;45(2):253–62.
23. Rodriguez MC, MacDonald JR, Mahoney DJ, Parise G, Beal MF, Tarnopolsky MA. Beneficial effects of creatine, CoQ10, and lipoic acid in mitochondrial disorders. Muscle Nerve. 2007;35(2):235–42.
24. Brealey D, Brand M, Hargreaves I, Heales S, Land J, Smolenski R, Davies NA, Cooper CE, Singer M. Association between mitochondrial dysfunction and severity and outcome of septic shock. Lancet. 2002;360(9328):219–23.
25. Strassburg CP. Shock liver. Best Pract Res Clin Gastroenterol. 2003;17(3):369–81.
26. Poole RC, Halestrap AP. Transport of lactate and other monocarboxylates across mammalian plasma membranes. Am J Phys Cell Phys. 1993;264(4):C761–82.
27. Lin ZY, Xu PB, Yan SK, Meng HB, Yang GJ, Dai WX, Liu XR, Li JB, Deng XM, Zhang WD. A metabonomic approach to early prognostic evaluation of experimental sepsis by (1) H NMR and pattern recognition. NMR Biomed. 2009;22(6):601–8.
28. Seymour CW, Yende S, Scott MJ, Pribis J, Mohney RP, Bell LN, Chen YF, Zuckerbraun BS, Bigbee WL, Yealy DM, et al. Metabolomics in pneumonia and sepsis: an analysis of the GenIMS cohort study. Intensive Care Med. 2013;39(8):1423–34.
29. Stringer KA, Younger JG, McHugh C, Yeomans L, Finkel MA, Puskarich MA, Jones AE, Trexel J, Karnovsky A. Whole blood reveals more metabolic detail of the human metabolome than serum as measured by 1H-NMR spectroscopy: implications for sepsis metabolomics. Shock. 2015;44(3):200–208.
30. Vaitheesvaran B, Xu J, Yee J, Q-Y L, Go VL, Xiao GG, Lee WN. The Warburg effect: a balance of flux analysis. Metabolomics. 2015;11(4):787–96.
31. Shapiro N, Howell MD, Bates DW, Angus DC, Ngo L, Talmor D. The association of sepsis syndrome and organ dysfunction with mortality in emergency department patients with suspected infection. Ann Emerg Med. 2006;48(5):583–90 590 e581.
32. Cooney RN, Kimball SR, Vary TC. Regulation of skeletal muscle protein turnover during sepsis: mechanisms and mediators. Shock. 1997;7(1):1–16.

33. Schakman O, Kalista S, Barbe C, Loumaye A, Thissen JP. Glucocorticoid-induced skeletal muscle atrophy. Int J Biochem Cell Biol. 2013;45(10):2163–72.
34. Hasselgren PO, Fischer JE. Sepsis: stimulation of energy-dependent protein breakdown resulting in protein loss in skeletal muscle. World J Surg. 1998; 22(2):203–8.
35. DeCoux A, Tian Y, DeLeon-Pennell KY, Nguyen NT, de Castro Bras LE, Flynn ER, Cannon PL, Griswold ME, Jin YF, Puskarich MA, et al. Plasma Glycoproteomics reveals sepsis outcomes linked to distinct proteins in common pathways. Crit Care Med. 2015;43(10):2049–58.

Part Four GENERAL CONCLUSIONS AND OUTLOOK

Metabolomics is the science designed to comprehensively study the metabolome, the repertoire of small molecule metabolites, which gives a comprehensive snapshot of the physiological state of the biofluid, extracts or cells studied. Currently, with remarkable advances in analytical techniques including NMR Spectroscopy and Mass Spectrometry, robustness statistical analysis, as well as improved calculation power for computers, that lead to continued improvement of the breadth and throughput of metabolomic analysis. Measuring metabolites by using metabolomics is a key complementary to genome, transcriptome and proteome studies, which may improve our understanding of how genetics, environment, the microbiome, disease, drug exposure, diet, and lifestyle influence the phenotype.

One of important application of metabolomics in clinical research is the discovery of novel biomarkers. The present PhD thesis focus on biomarkers discovery, especially applying metabolomics in Non-alcoholic fatty liver disease (NAFLD), prostate cancer (PCa) and septic shock.

The objectives for the first part of the study (NAFLD) were: (1) describe the relative plasma metabolome and lipidome changes in Nonalcoholic fatty liver (NAFL) and in Non-Alcoholic SteatoHepatitis (NASH) compared with Normal Liver (NL) obese patients, investigated whether Ultra-Performance Liquid Chromatography (UPLC) coupled with High-resolution mass spectrometry (HRMS) based plasma metabolomics and lipidomics analysis could help to identify potentials biomarkers, if any, associated with different stages of NAFLD (NAFL, NASH), and identify metabolomic or lipidomic patterns which could discriminate NAFL and NASH from NL obese patients, by using appropriate statistical models.

Our results revealed significant relative changes in certain metabolites and lipids, especially for lipids metabolism in different stage of NAFLD, such as Sphingomyelin, Phosphatidylcholine, Ceramides, Phosphatidylethanolamine, Lysophosphatidylcholines, lysophosphatidylethanolamine, Sphingosine and lactamide. Compared with NL obese patients, the changes in the plasma metabolome and lipidome were more distinct in NASH patients than in NAFL patients. The study suggested that UPLC-HRMS based

metabolomics, especially lipidomics analysis could be promising approach in NAFLD. Further investigation should be particularly focus on lipidomics, as well as investigation subtypes, appropriate data processing and statistic model.

The objective for the second part of the study (PCa) was to investigate whether UPLC-HRMS (Obitrap) based plasma untargeted metabolomic profiles, established from a simple baseline blood draw from healthy men, and appropriate statistical models, could identify biomarkers, if any, associated with the risk of developing prostate cancer within the following decade.

Our results revealed a panel of 7 metabolites which may useful for prediction the risk of prostate cancer decade before, with AUC: 0.900 (95% CI: 0.833 to 0.950), Sensitivity: 81.82%; Specificity: 91.46% in our validation cohort (Cases: n= 44 / Control: n= 82). Men characterized by higher fasting plasma levels of Cis-11,14-Eicosadienoic acid, Ethyl oleate and phosphoric acid had a higher risk of developing prostate cancer during the 13-year follow-up. Which suggested that UPLC-HRMS (Obitrap) based plasma untargeted metabolomic profiles, established from a simple baseline blood draw from healthy men, may identify biomarkers, that associated with the risk of developing prostate cancer within the following decade. Nevertheless, interpret more profoundly these 7 metabolites in metabolism pathway, and other data analysis method such as sparse PLS-DA (sPLS-DA) still need to perform.

For the part concerning septic shock, the aim was to analyze the discriminatory ability of metabolic profiles between septic shock survivors and non-survivors at the beginning and 24h after ICU admission and also to describe the evolution of metabolic profiles for septic shock patients during this period by using 1H NMR spectroscopy-based metabolomics. Our results show that the sustained enrichment of energy-supply metabolites and amino acids is predictive of a bad outcome. We suggest that monitoring the relevant metabolites in the first 24h may help to evaluate early therapeutic response.

In conclusion, accumulated evidence suggests the promising perspective for application of metabolomics in clinical research, especially for biomarkers discovery. Nevertheless, independent prospective validation studies are needed to translate metabolomics findings into the clinical applications.

In further perspective, other directions should be taken into account as well, such as subtyping of patients, more robustness data processing and appropriate statistical analysis, sparse PLS-DA for example. Moreover, combine with other omics research such as transcriptomics, proteomics, and also clinical characteristics may improve novel subtyping approach, allowing further more precisely classification and staging of patients, in order to correctly interpret the biochemical processes behind the disease, which could contribute to (1) improve our understanding in the aetiology of disease, to (2) development of appropriate therapy and precision medicine-based management of patients.

ANNEXES

Annexe 1. List of detected mass and retention time used for metabolites identification.

Metabolites	Mode	m/z	RT
Pyrrolidine	ESI+	72.0821	0.39
TMAO	ESI+	76.0762	0.36
Lactamide	ESI+	90.056	0.37
Orthophosphoric acid	ESI+	98.9843	0.41
Choline	ESI+	104.1072	0.36
Proline	ESI+	116.0707	0.4
Valine	ESI+	118.0857	0.38
Indoline	ESI+	120.0802	0.68
Salicylic acid	ESI+	121.0279	4.46
Nicotinamide	ESI+	123.0548	0.47
D-Pipecolinic acid	ESI+	130.0874	0.32
Creatine	ESI+	132.0777	0.37
Leucine	ESI+	132.10176	0.53
Ornithine	ESI+	133.0971	0.3
Hypoxanthine	ESI+	137.04567	0.46
1-Amlnocydohexanecarboxylic acid	ESI+	144.1018	0.4
Acetylchloline	ESI+	146.1162	0.38
L-Glutamine	ESI+	147.0757	0.39
L-Lysine	ESI+	147.1127	0.31
L-Glutamic acid	ESI+	148.0609	0.38
α-Keto-γ-(methylthio)butyric acid	ESI+	149.0226	4.45
D-Methionine	ESI+	150.0582	0.46
Guanine	ESI+	152.0567	0.48
2-Hydroxy-6-aminopurine	ESI+	152.0567	0.39
4-Acetamidophenol	ESI+	152.0693	0.8
Xanthine	ESI+	153.0404	0.47
Histidine	ESI+	156.0755	0.37
L-Carnitine	ESI+	162.1122	0.38
trans-P-Coumaric acid	ESI+	165.0539	0.48
L-Phenylalanine	ESI+	166.08589	0.68
IS L-Phenylalanine-d5	ESI+	171.1173	0.66
L-arginine	ESI+	175.1188	0.33
N-acetylaspartate	ESI+	176.0697	2
Serotonin	ESI+	177.10223	0.48
1,7-Dimethylxanthine	ESI+	181.0729	0.97
L-Tyrosine	ESI+	182.0806	0.48
D-Mannitol	ESI+	183.0843	0.46
Epinephrine	ESI+	184.0943	0.34

Tryptophan	ESI+	188.0699	0.97
N8-Acetylspermidine	ESI+	188.1749	0.31
3-Indolepropionic acid	ESI+	190.0862	2.2
Cotinine N-oxide	ESI+	193.10186	0.94
L-Altrose	ESI+	203.0527	0.36
Symmetric dimethylarginine	ESI+	203.1494	0.36
Carnitine C2:0	ESI+	204.1228	0.39
L-Tryptophan	ESI+	205.0962	0.95
Kynurenine	ESI+	209.0921	0.62
IS Tryptophan-d5	ESI+	210.128	0.91
Carnitine C3:0	ESI+	218.1384	0.53
Gluconate	ESI+	219.0462	0.37
Zeatin	ESI+	220.1181	0.77
L-Carnosine	ESI+	227.1135	0.31
Leu-pro	ESI+	229.1543	0.38
Carnitine C4:0	ESI+	232.1541	0.91
Gly-Tyr	ESI+	239.1067	0.39
Carnitine C5:1	ESI+	244.1539	1.31
Carnitine C5:0	ESI+	246.1696	1.45
N-Acetyl-D-tryptophan	ESI+	247.1076	1.82
Octopine	ESI+	247.14367	1.26
Carnitine C4-OH	ESI+	248.147	1.26
Palmitoleic acid	ESI+	255.2311	5.56
Hexadecanamide	ESI+	256.2651	5.11
Choline glycerophosphate	ESI+	258.1117	0.39
Carnitine C6:0	ESI+	260.1853	1.8
γ-Glu-Leu	ESI+	261.1439	1.28
Phe-pro	ESI+	263.1386	1.55
Phenylacetylglutamine	ESI+	265.1179	1.45
1-Hexadecanol	ESI+	265.2524	5.88
FFAD C18:1	ESI+	282.2786	5.33
FFA C18:1	ESI+	283.2627	5.85
8-hydroxy-2'-deoxyguanosine	ESI+	284.0981	0.48
stearamide	ESI+	284.295	5.73
Carnitine C8:1	ESI+	286.2013	2.06
Carnitine C8:0	ESI+	288.2165	2.28
IS Carnitine C8:0-d3	ESI+	291.23526	2.27
glycated valine	ESI+	296.067	0.36
Palmitoylethanolamide	ESI+	300.29	4.88
Sphingosine	ESI+	300.2902	3.51
Sphinganine	ESI+	302.3046	3.64
Eicosapentaenoic acid	ESI+	303.2305	5.17
Arachidonic acid	ESI+	305.2471	5.5
cis-8,11,14-Eicosatrienoic acid	ESI+	307.264	5.73
Cis-11,14-Eicosadienoic acid	ESI+	309.2776	6.04
fatty amide C20:1	ESI+	310.3103	5.89

Ethyl oleate	ESI+	311.2928	6.97
Fatty amide C20:0	ESI+	312.3263	6.31
Phe-Phe	ESI+	313.1542	1.78
Carnitine C10:1	ESI+	314.2319	2.5
Carnitine C10:0	ESI+	316.2477	2.66
Phytosphingosine	ESI+	318.3004	3.36
Methyl arachidonate	ESI+	319.2627	6.3
Cis-8,11,14-Eicosatrienoic acid Methyl ester	ESI+	321.2719	2.61
N-Oleoylethanolamine	ESI+	326.3051	5.11
Docosahexaenoic acid	ESI+	329.2477	5.45
all-cis-4,7,10,13,16-Docosapentaenoic acid	ESI+	331.2631	5.77
Glycerol 1-hexadecanoate	ESI+	331.2836	5.33
Fatty amide C22:2	ESI+	336.326	6.05
FFA C22:1	ESI+	338.3402	6.42
Erucamide	ESI+	338.341	6.38
Fatty amide C22:0	ESI+	340.3558	6.82
Carnitine C12:1	ESI+	342.2633	2.88
Carnitine C12:0	ESI+	344.2788	3.12
Adenosine 5'-monophosphate	ESI+	348.0688	0.48
Anandamide	ESI+	348.2868	4.79
15-Ketoprostaglandin F2α	ESI+	353.2315	2.29
Phenol red	ESI+	355.0641	1.91
MAG 18:1	ESI+	357.3007	7.26
Hydrocortisone	ESI+	363.2161	2.27
Alpha-Lactose	ESI+	365.1069	0.37
Carnitine C14:2	ESI+	368.2788	3.12
Carnitine C14:1	ESI+	370.2945	3.38
Carnitine C14:0	ESI+	372.3101	3.61
3-Hydroxydecanoic acid	ESI+	377.2982	2.99
Sphingosine-1-phosphate	ESI+	380.2552	3.57
Carnitine C14-OH	ESI+	388.306	3.05
Diisooctyl phthalate	ESI+	391.2839	6.71
Carnitine C16:2	ESI+	396.312	3
Carnitine C16:1	ESI+	398.3258	3.81
Carnitine C16:0	ESI+	400.3416	4.13
7-Ketocholestero	ESI+	401.3407	5.96
20α-Hydroxy Cholesterol	ESI+	403.3562	6.28
c16-d3	ESI+	403.3611	4.06
GCDCA	ESI+	414.3001	2.63
carnitine C16:0-OH	ESI+	416.3368	3.54
carnintine C18:3	ESI+	422.3268	3.72
carnitine C18:2	ESI+	424.3413	4.01
Carnitine C18:1	ESI+	426.357	4.29
L-Cysteine-Glutathione	ESI+	427.0954	0.39
Carnitine C18:0	ESI+	428.3729	4.6
carnitine C18:0	ESI+	428.3749	4.58

Carnitine C18:2-OH	ESI+	440.3362	3.52
Carnitine C18:1-OH	ESI+	442.3532	3.8
Carnitine C18-OH	ESI+	444.3694	4.03
Glycodeoxycholic acid	ESI+	450.3189	3.05
LPE 16:1	ESI+	452.278	3.79
LPE 16:0 sn-1	ESI+	454.2922	4.15
LPE 16:0 sn-2	ESI+	454.2922	4.04
LPC 14:0	ESI+	468.3061	3.61
LPE 18:2 sn-1	ESI+	478.2921	4.01
LPE 18:2 sn-2	ESI+	478.2921	3.91
LPE 18:1	ESI+	480.3114	4.36
LPC O-16:1	ESI+	480.3439	4.36
LPE 18:0 sn-1	ESI+	482.3223	4.74
LPE 18:0 sn-2	ESI+	482.3223	4.62
LPC 15:0 sn-1	ESI+	482.3238	3.89
LPC 15:0 sn-2	ESI+	482.3238	3.79
LPC O-16:0	ESI+	482.3586	4.32
LPE 18:0-2 sn-2	ESI+	482.3596	4.72
LPC 16:1 sn-2	ESI+	494.3233	3.81
LPC 16:1	ESI+	494.3236	3.82
LPC 16:0 sn-1	ESI+	496.3388	4.18
LPC 16:0 sn-2	ESI+	496.3389	4.07
LPE 20:4 sn-1	ESI+	502.2919	4.07
LPE 20:4	ESI+	502.2931	4.07
LPE 20:3	ESI+	504.3062	4.25
LPE 20:2	ESI+	506.3262	4.53
LPE 20:1	ESI+	508.3381	4.88
LPC O-18:1	ESI+	508.3754	4.5
LPE 20:0	ESI+	510.354	5.3
LPC 17:0	ESI+	510.3548	4.4
LPC O-18:0	ESI+	510.3909	4.9
LPC 18:3 sn-2	ESI+	518.3205	3.75
LPC 18:3 sn-1	ESI+	518.3207	4.18
LPC 18:3	ESI+	518.3224	3.75
LPC 18:2 sn-1	ESI+	520.3388	4.04
LPC 18:2 sn-2	ESI+	520.3391	4.04
LPC 18:2	ESI+	520.3391	4.03
LPC 18:1 sn-1	ESI+	522.3543	4.36
LPC 18:1 sn-2	ESI+	522.3547	4.25
LPC 18:0 sn-1	ESI+	524.3701	4.76
LPC 18:0 sn-2	ESI+	524.3703	4.62
LPE 22:6 sn-1	ESI+	526.292	4.07
LPE 22:6 sn-2	ESI+	526.292	4.07
LPE 22:6	ESI+	526.2928	3.99
D(+)-Melezitose	ESI+	527.1579	0.38
LPE 22:4 -1	ESI+	528.30977	4.21

LPE 22:4 -2	ESI+	530.32466	4.51
LPC 19:0	ESI+	538.386	5.02
Cer(d18:1/16:0)-2	ESI+	538.5187	7.68
Cer(d18:0/16:0)-1	ESI+	540.5317	7.76
Cer(d18:0/16:0)-2	ESI+	540.5328	7.05
LPC 20:5 sn-1	ESI+	542.3233	3.8
LPC 20:5 sn-2	ESI+	542.3233	3.8
LPC 20:4 sn-2	ESI+	544.3373	3.99
LPC 20:4 sn-1	ESI+	544.339	4.39
LPC 20:3 sn-1	ESI+	546.3544	4.27
LPC 20:3	ESI+	546.3576	4.28
LPC 20:2 sn-1	ESI+	548.3716	4.56
LPC 20:2 sn-2	ESI+	548.3716	4.45
LPC 20:1	ESI+	550.3861	4.9
LPC 20:0 sn-1	ESI+	552.4014	5.32
LPC 20:0 sn-2	ESI+	552.4014	5.21
LPC 22:6 sn-1	ESI+	568.3391	4.09
LPC 22:6 sn-2	ESI+	568.3391	4.09
LPC 22:6	ESI+	568.3395	4
Cer(d18:0/18:0)-1	ESI+	568.5627	8.04
LPC 22:5	ESI+	570.3537	4.24
LPC 22:5 sn-1	ESI+	570.3542	4.35
LPC 22:5 sn-2	ESI+	570.3542	4.23
LPC 22:4	ESI+	572.3698	4.52
LPC 22:2	ESI+	576.4091	4.35
Biliverdin	ESI+	583.2543	2.86
L-Glutathione	ESI+	613.1594	0.47
SM 30:1	ESI+	647.5105	5.97
SM 32:2	ESI+	673.5269	6.1
SM 32:1 SM(d18:1/14:0)	ESI+	675.5414	6.46
SM 32:1	ESI+	675.5426	6.43
PC(28:0)	ESI+	678.5077	6.82
SM 33:1	ESI+	689.5584	6.65
SM 33:1 SM(d18:1/15:0)	ESI+	689.5588	6.67
PE O-34:3	ESI+	700.5278	7.55
SM 34:2	ESI+	701.5579	6.55
SM 34:2 SM(d18:1/16:1)	ESI+	701.5597	6.57
SM 34:1	ESI+	703.5739	6.85
SM 34:1 SM(d18:1/16:0)	ESI+	703.5748	6.88
SM 34:0	ESI+	705.5802	6.97
PC 30:0	ESI+	706.5371	7.16
SM 35:2	ESI+	715.5737	6.79
PE 34:2	ESI+	716.5234	7.41
SM 35:1	ESI+	717.5889	7.09
PE O-36:5	ESI+	724.5263	7.49
PC(32:4)	ESI+	726.5023	7.01

SM 36:3	ESI+	727.5733	6.7
SM 36:2	ESI+	729.5894	6.97
SM 36:2 SM(d18:1/18:1)	ESI+	729.5894	7.01
PC 32:2	ESI+	730.5371	7
SM 36:1	ESI+	731.6049	7.23
PC 32:1	ESI+	732.5503	7.8
PC 32:0	ESI+	734.5677	7.54
PE 36:4	ESI+	740.5248	7.42
PE 36:3	ESI+	742.5362	7.47
PC 33:2	ESI+	744.5525	7.17
PC O-34:2	ESI+	744.5903	7.52
PC 33:1	ESI+	746.5676	7.42
PE 33:1	ESI+	746.5696	7.95
PC O-34:1	ESI+	746.6047	7.7
PE O-38:7	ESI+	748.5257	7.4
PE O-38:6	ESI+	750.541	7.59
PC 34:4	ESI+	754.537	7
PC 34:3	ESI+	756.5524	7.11
PC 34:2(16:0/18:2)	ESI+	758.5676	7.4
PC 34:2	ESI+	758.5683	7.35
PC 34:1	ESI+	760.5836	7.58
PC 34:0	ESI+	762.6016	7.85
PE 38:6	ESI+	764.5208	7.29
PC O-36:5	ESI+	766.5767	7.5
PC O-36:4	ESI+	768.5504	7.21
PE 38:4	ESI+	768.5518	7.67
PC 35:3	ESI+	770.5673	7.29
PC 36:5	ESI+	780.5526	7.12
PC 36:4	ESI+	782.5679	7.36
SM 40:3	ESI+	783.6359	7.39
PC 36:3	ESI+	784.5845	7.47
PC 36:2	ESI+	786.5996	7.71
PC 36:1	ESI+	788.6178	7.92
PC o-38:5	ESI+	794.6044	7.54
PC 38:7	ESI+	804.5516	7.07
PC 38:6	ESI+	806.5686	7.3
PC 38:5	ESI+	808.5833	7.44
PC 38:4	ESI+	810.5989	7.54
CA-d4	ESI-	411.3022	5.18
FFA 16:0-d3	ESI-	258.2527	7.14
FFA 18:0-d3	ESI-	286.2827	7.67
(19S)-Hydroxyicosatetraenoic acid	ESI-	319.2241	6.69
1,3-dimethyluric acid	ESI-	217.0293	0.36
1,5-Anhydro-D-Glucitol	ESI-	163.0603	0.43
11,12-Epoxy-(5z,8z,14z)-Eicosatrienoic Acid	ESI-	319.2284	7.13
13-cis-acitretin	ESI-	325.1848	6.78

15-Oxoete	ESI-	317.2126	5.74
1-Stearoyl-Sn-Glycerol-3-Phosphocholine	ESI-	522.3562	8.07
2,5-Dichlorobenzoic Acid	ESI-	188.9376	0.38
2-Aminoethylphosphonic acid	ESI-	124.0077	0.34
2-Phosphoenol pyruvate	ESI-	166.9739	0.29
3-(1-Pyrazolyl)-L-alanine	ESI-	154.0609	0.37
3-Indolepropionic acid	ESI-	188.071	2.02
3-Methylhistidine	ESI-	168.0771	0.41
5(S),6(R)-Lipoxin a4	ESI-	351.2172	4.82
5-Oxoete	ESI-	317.2098	6.21
6-Phosphogluconic acid	ESI-	275.0173	0.29
8(R)-Hydroxy-(5Z,9E,11Z,14Z)-eicosatetraenoic acid	ESI-	319.2269	6
9-Oxoode	ESI-	293.211	5.7
Acetanilide	ESI-	134.0601	1.57
Adenosine5-phosphosulfate	ESI-	426.0188	0.32
a-ketoglutaric acid	ESI-	145.0134	0.36
Ala-Gly	ESI-	145.0611	0.34
Carnosine	ESI-	225.0991	0.38
Chenodeoxycholic acid	ESI-	391.2848	5.78
Cholic acid	ESI-	407.2797	5.23
Citraconic acid	ESI-	129.0188	0.31
Citric acid	ESI-	191.0201	0.29
Creatinine	ESI-	112.0505	0.49
Cystathionine	ESI-	221.0583	0.32
Cysteine-glutathione gisulfide	ESI-	425.0767	0.31
Cytidine-3(2)-Monophisphoric acid	ESI-	322.042	0.32
D-3-Phosphoglyceric acid	ESI-	184.9849	0.29
D-Citramalic acid	ESI-	147.029	0.31
Dehydroepiandrosterone Sulfate	ESI-	367.158	4.05
D-Glucose 6-phosphate	ESI-	259.0209	0.31
Dimethyluric acid	ESI-	195.05	0.34
DL-Aspartic acid	ESI-	132.0298	0.32
DL-malic acid	ESI-	133.0133	0.31
D-Mannitol	ESI-	181.0706	0.37
FFA 11:0	ESI-	185.1547	5.1
FFA 18:2	ESI-	279.2331	6.92
FFA 18:4	ESI-	275.2021	6.22
FFA 23:0	ESI-	353.3425	5.4
FFA 25:0	ESI-	381.3732	8.72
FFA 26:0	ESI-	395.3889	8.87
FFA C10:0	ESI-	171.1383	4.65
FFA C12:0	ESI-	199.1697	5.71
FFA C13:0	ESI-	213.1855	6.14
FFA C14:0 (Myristic acid)	ESI-	227.2023	6.51
FFA C15:0 (Pentadecanoic acid)	ESI-	241.2171	6.8
FFA C16:0 (Palmitic acid)	ESI-	255.2338	7.02

FFA C16:1 (Hexadecenoic acid)	ESI-	253.2172	6.7
FFA C16:2	ESI-	251.2014	6.4
FFA C17:0 (Heptadecanoic acid)	ESI-	269.2492	7.44
FFA C17:1 (Heptadecenoic acid)	ESI-	267.2324	6.98
FFA C18:0 (Stearic acid)	ESI-	283.2633	7.67
FFA C18:1 (Vaccenic acid-2)	ESI-	281.2489	7.31
FFA C18:2 (β-Linoleic acid)	ESI-	279.2333	6.99
FFA C18:3 (Linolenic acid)	ESI-	277.2158	6.67
FFA C19:0 (Nonadecanoic acid)	ESI-	297.279	7.81
FFA C19:1(Cis-10-Nonadecenoic acid)	ESI-	295.2626	7.56
FFA C20:0 (Arachidic acid)	ESI-	311.2946	8.06
FFA C20:1 (Eicosenoic acid)	ESI-	309.2798	7.77
FFA C20:2 (Cis-11,14-Eicosadienoic acid)	ESI-	307.2636	7.51
FFA C20:3 (Cis-8,11,14-Eicosatrienoic acid)	ESI-	305.2469	7.24
FFA C20:4 (Arachidonic acid)	ESI-	303.2334	7.03
FFA C20:5 (Eicosapentaenoic acid)	ESI-	301.2164	6.73
FFA C21:0 (Heneicosanoic acid)	ESI-	325.3091	8.21
FFA C22:0 (Behenic acid)	ESI-	339.3272	8.4
FFA C22:1 (Erucic acid)	ESI-	337.3119	8.13
FFA C22:2	ESI-	335.294	7.81
FFA C22:5 (All-cis-4,7,10,13,16-docosapentaenoic acid)	ESI-	329.247	7.09
FFA C22:6 (Docosahexaenoic acid)	ESI-	327.2317	7.03
FFA C22:7	ESI-	325.2145	6.88
FFA C24:0 (Tetracosanoic acid)	ESI-	367.3565	8.73
FFA C24:1 (Nervonic acid)	ESI-	365.3418	8.46
FFA C24:2	ESI-	363.3253	8.19
FFA C24:5	ESI-	357.2785	7.64
FFA C24:6	ESI-	355.2623	7.44
FFA C8:0	ESI-	143.1076	3.21
FFA C9:0	ESI-	157.1225	3.99
Fructose 1,6-bisphosphate	ESI-	338.9879	0.27
Glutathione	ESI-	611.1467	0.31
Glycerylphosphorylethanolamine	ESI-	214.0481	0.34
Glycocholic acid	ESI-	464.3011	5.16
Glycodeoxycholic acid	ESI-	448.3052	5.66
Glycoursodeoxycholic acid	ESI-	448.3046	4.7
Hexadecatrienoic acid/FFA 16:3	ESI-	249.1846	6.13
Hippuric acid	ESI-	178.0511	1.56
Hypoxanthine	ESI-	135.032	0.47
Hypoxanthine-9-beta-D-arabinofuranosine	ESI-	267.0729	1.21
Indolelactic acid	ESI-	204.0664	1.67
Indoxyl sulfate potassium salt	ESI-	212.0029	1.6
L(+)-Ornithine	ESI-	131.0829	0.43
Lactamide	ESI-	88.0415	0.38
Lactic acid	ESI-	89.0233	0.36
L-Arginine	ESI-	173.1049	0.42

L-Asparagine	ESI-	131.034	0.55
Leucine/Isoleucine	ESI-	130.0863	0.76
L-Glutamic acid	ESI-	146.045	0.33
L-Glutamine	ESI-	145.061	0.36
L-Homoserine	ESI-	118.0499	0.38
Lithocholic acid	ESI-	375.2911	6.33
LPC16:0	ESI-	494.3245	7.53
L-Tryptophan	ESI-	203.0819	1.66
Mannose	ESI-	215.0325	0.37
Myoinositol	ESI-	179.0547	0.36
Na-Acetylarginine	ESI-	215.0317	0.34
N-Acetylaspartic acid	ESI-	174.0398	0.31
N-Acetylglutamic acid	ESI-	188.0541	0.31
N-Acetylneuraminic acid	ESI-	308.099	0.39
N-Oleoylethanolamine	ESI-	324.29	7.92
Oleamide	ESI-	280.2637	7.97
Oleoyl-L-α-lysophosphatidic acid	ESI-	435.2465	7.19
Orthophosphoric acid	ESI-	96.9704	0.31
Oxypurinol	ESI-	151.0261	0.38
Palmitoylethanolamide	ESI-	298.2744	7.66
P-cresol sulfate	ESI-	187.0064	2.07
P-cresyl glucuronide	ESI-	283.0821	1.79
Phenol red	ESI-	353.0476	2.8
Phenyl sulfate	ESI-	172.9909	1.51
phenylacetylglutamine	ESI-	263.1035	1.78
Phenylalanine	ESI-	164.0709	1.43
proline	ESI-	114.0549	0.44
Pyroglutamic acid	ESI-	128.0358	0.36
S-(5-Adenosy)-L-homocysteine	ESI-	383.1109	1.47
Succinic acid	ESI-	117.0187	0.31
Taurochenodesoxycholic acid	ESI-	498.2877	5.63
Taurocholic acid	ESI-	514.2824	5.12
trans-3-Hydroxy-L-proline	ESI-	130.0488	0.31
trans-9-Octadecenoic acid	ESI-	563.5007	7.19
Tyrosine	ESI-	180.0657	0.47
Uric acid	ESI-	167.0202	0.38
Uridine	ESI-	243.0616	0.46
Ursolic acid	ESI-	455.349	7.27
Valine	ESI-	116.0734	0.4

m/z: detected mass, RT: retention time (minutes).

Annexe 2. List of detected mass and retention time used for lipid ions identification.

Lipid Ion	Mode	Obs m/z	Rt
Cer(d16:1/24:0)+H	ESI+	622.614159	10.05
Cer(d16:1/24:1)+H	ESI+	620.598963	9.26
Cer(d18:1/16:0)+H	ESI+	538.518718	7.52
Cer(d18:1/17:0)+H	ESI+	552.534637	7.95
Cer(d18:1/20:0)+H	ESI+	594.581365	9.17
Cer(d18:1/22:0)+H	ESI+	622.613621	9.85
Cer(d18:1/24:0)+H	ESI+	650.645906	10.49
Cer(d18:1/24:1)+H	ESI+	648.630884	9.81
Cer(d18:1/24:2)+H	ESI+	646.614616	9.27
Cer(d18:2/24:1)+H	ESI+	646.613141	9.31
CerG1(d18:1/16:0)+H	ESI+	700.57225	6.85
CerG1(d18:1/22:0)+H	ESI+	784.665875	9.33
CerG1(d18:1/24:0)+H	ESI+	812.697856	10.02
CerG1(d18:1/24:1)+H	ESI+	810.682137	9.33
CerG1(d18:2/22:0)+H	ESI+	782.651878	8.76
CerG1(d18:2/24:0)+H	ESI+	810.682968	9.52
CerG1(d18:2/24:1)+H	ESI+	808.667568	8.74
CerG2(d16:1/16:0)+H	ESI+	834.593748	5.63
CerG2(d18:1/16:0)+H	ESI+	862.623385	6.53
CerG2(d18:1/24:1)+H	ESI+	972.734889	9.06
CerG2(d18:2/16:0)+H	ESI+	860.607722	5.75
CerG3(d18:1/16:0)+H	ESI+	1024.68042	6.36
CE 15:0	ESI+	628.603	9.85
CE 16:0	ESI+	642.618	12.75
CE 16:1	ESI+	640.603	12.30
CE 17:0	ESI+	656.634	10.50
CE 17:1	ESI+	654.618	9.85
CE 18:1	ESI+	668.634	12.77
CE 18:2	ESI+	666.618	12.36
CE 18:3	ESI+	664.603	11.92
CE 20:2	ESI+	694.650	12.75
CE 20:3	ESI+	692.634	12.40
CE 20:4	ESI+	690.618	12.04
CE 20:5	ESI+	688.603	11.63
CE 22:5	ESI+	716.634	12.08
CE 22:6	ESI+	714.618	11.78
DG(16:0/18:1)+NH4	ESI+	612.557247	8.86
DG(16:0/18:2)+NH4	ESI+	610.540219	8.41
DG(16:0/18:3)+NH4	ESI+	608.524485	7.90
DG(16:0/20:4)+NH4	ESI+	634.540345	8.17
DG(16:0/22:5)+NH4	ESI+	660.555883	8.17

DG(16:0/22:6)+NH4	ESI+	658.539918	7.90
DG(16:1/18:2)+NH4	ESI+	608.524259	7.66
DG(18:0/18:1)+NH4	ESI+	640.589122	9.60
DG(18:1/18:1)+NH4	ESI+	638.571183	8.97
DG(18:1/18:2)+NH4	ESI+	636.555583	8.55
DG(18:1/18:3)+NH4	ESI+	634.539771	7.99
DG(18:1/20:3)+NH4	ESI+	662.571993	8.60
DG(18:1/20:4)+NH4	ESI+	660.555367	8.35
DG(18:1/22:5)+NH4	ESI+	686.571292	8.31
DG(18:1/22:6)+NH4	ESI+	684.555715	8.01
DG(18:2/18:2)+NH4	ESI+	634.539671	7.81
DG(18:2/20:4)+NH4	ESI+	658.539541	7.64
DG(18:2/22:6)+NH4	ESI+	682.539483	7.36
DG(18:3/18:2)+NH4	ESI+	632.524102	7.38
DG(20:5/18:2)+NH4	ESI+	656.52384	7.07
LPC(14:0)+H	ESI+	468.30784	1.45
LPC(15:0)+H	ESI+	482.323373	1.57
LPC(15:1)+H	ESI+	480.307996	1.89
LPC(16:0)+H	ESI+	496.338654	1.78
LPC(16:0e)+H	ESI+	482.359806	1.99
LPC(16:0p)+H	ESI+	480.344607	2.01
LPC(16:1)+H	ESI+	494.323482	1.65
LPC(16:1p)+H	ESI+	478.327939	1.73
LPC(17:0)+H	ESI+	510.354807	1.87
LPC(17:1)+H	ESI+	508.339083	1.67
LPC(18:0)+H	ESI+	524.370823	2.29
LPC(18:0e)+H	ESI+	510.391559	2.52
LPC(18:0p)+H	ESI+	508.375973	2.07
LPC(18:1)+H	ESI+	522.354797	1.93
LPC(18:1p)+H	ESI+	506.36041	2.03
LPC(18:2)+H	ESI+	520.338919	1.67
LPC(18:3)+H	ESI+	518.323452	1.39
LPC(18:4)+H	ESI+	516.3056	1.47
LPC(19:0)+H	ESI+	538.386682	2.58
LPC(19:1)+H	ESI+	536.370678	2.06
LPC(20:0)+H	ESI+	552.403105	2.94
LPC(20:0e)+H	ESI+	538.423896	3.36
LPC(20:0p)+H	ESI+	536.40763	2.62
LPC(20:1)+H	ESI+	550.38748	2.46
LPC(20:1p)+H	ESI+	534.391608	2.16
LPC(20:2)+H	ESI+	548.370042	1.89
LPC(20:3)+H	ESI+	546.354611	1.71
LPC(20:4)+H	ESI+	544.337736	1.73
LPC(20:5)+H	ESI+	542.323439	1.38
LPC(22:0)+H	ESI+	580.433765	3.85
LPC(22:1)+H	ESI+	578.418993	3.00

LPC(22:3)+H	ESI+	574.38488	2.95
LPC(22:4)+H	ESI+	572.369943	1.81
LPC(22:5)+H	ESI+	570.354429	1.60
LPC(22:6)+H	ESI+	568.337079	1.68
LPC(24:0)+H	ESI+	608.465061	4.89
LPC(24:1)+H	ESI+	606.449584	3.88
LPC(26:1)+H	ESI+	634.480981	4.93
LPE(16:0)+H	ESI+	454.291248	1.73
LPE(18:0)+H	ESI+	482.323937	2.15
LPE(18:0p)+Na	ESI+	488.311132	2.60
LPE(18:1)+H	ESI+	480.307072	1.84
LPE(18:2)+H	ESI+	478.292008	1.61
LPE(22:6)+H	ESI+	526.292374	1.48
PC(14:0e/16:0)+H	ESI+	692.558495	7.11
PC(14:0p/20:2)+H	ESI+	742.573181	6.92
PC(15:0/20:4)+H	ESI+	768.552028	6.33
PC(15:0/22:6)+H	ESI+	792.552586	6.05
PC(16:0/15:0)+H	ESI+	720.553138	6.94
PC(16:0/16:0)+H	ESI+	734.56837	7.30
PC(16:0/16:1)+H	ESI+	732.552725	6.67
PC(16:0/17:0)+H	ESI+	748.58409	7.68
PC(16:0/17:1)+H	ESI+	746.568404	7.08
PC(16:0/18:1)+H	ESI+	760.584277	7.49
PC(16:0/18:2)+H	ESI+	758.568536	6.97
PC(16:0/19:1)+H	ESI+	774.600393	7.87
PC(16:0/22:0)+H	ESI+	818.665656	9.58
PC(16:0/22:1)+H	ESI+	816.647202	8.97
PC(16:0/24:2)+H	ESI+	842.662557	9.13
PC(16:0e/18:1)+H	ESI+	746.606253	8.15
PC(16:0e/18:2)+H	ESI+	744.588808	7.42
PC(16:0e/20:4)+H	ESI+	768.588358	7.34
PC(16:0e/22:2)+H	ESI+	800.653088	9.09
PC(16:0e/22:3)+H	ESI+	798.637489	8.58
PC(16:0e/22:6)+H	ESI+	792.589476	7.28
PC(16:0e/24:2)+H	ESI+	828.685866	9.57
PC(16:0p/16:1)+H	ESI+	716.558148	7.20
PC(16:0p/17:1)+H	ESI+	730.574155	7.01
PC(16:0p/18:2)+H	ESI+	742.573646	7.48
PC(16:0p/20:4)+H	ESI+	766.573702	7.34
PC(16:0p/22:6)+H	ESI+	790.573294	6.66
PC(16:0p/24:2)+H	ESI+	826.669725	9.23
PC(16:0p/24:7)+H	ESI+	816.589195	7.19
PC(16:1/16:1)+H	ESI+	730.537784	5.95
PC(16:1/16:2)+H	ESI+	728.522595	5.53
PC(16:1/18:2)+H	ESI+	756.552837	6.25
PC(16:1/18:3)+H	ESI+	754.537928	5.62

PC(16:1/19:1)+H	ESI+	772.584494	7.27
PC(16:1/20:4)+H	ESI+	780.552103	6.19
PC(16:1/20:5)+H	ESI+	778.536595	5.72
PC(16:1/24:2)+H	ESI+	840.649593	8.71
PC(16:1/24:7)+H	ESI+	830.567963	6.13
PC(16:1p/18:2)+H	ESI+	740.558014	6.69
PC(16:1p/19:1)+H	ESI+	756.587625	7.77
PC(16:1p/20:4)+H	ESI+	764.554142	6.50
PC(16:1p/24:2)+H	ESI+	824.654945	8.72
PC(16:1p/24:7)+H	ESI+	814.571096	7.06
PC(16:2/18:3)+H	ESI+	752.521534	5.22
PC(16:2/20:5)+H	ESI+	776.527463	5.56
PC(17:0/20:4)+H	ESI+	796.584017	7.31
PC(17:0/22:6)+H	ESI+	820.584203	7.06
PC(17:1/18:2)+H	ESI+	770.568269	6.74
PC(18:0/16:0)+H	ESI+	762.600482	8.23
PC(18:0/18:1)+H	ESI+	788.615852	8.37
PC(18:0/18:2)+H	ESI+	786.600033	7.71
PC(18:0/19:1)+H	ESI+	802.634148	8.63
PC(18:0/24:1)+H	ESI+	872.712048	10.50
PC(18:0/24:2)+H	ESI+	870.696379	9.98
PC(18:0e/16:0)+H	ESI+	748.62389	8.72
PC(18:0e/18:2)+H	ESI+	772.620937	8.25
PC(18:0e/20:4)+H	ESI+	796.621175	8.25
PC(18:0e/22:5)+H	ESI+	822.636978	8.19
PC(18:0e/22:6)+H	ESI+	820.620533	7.44
PC(18:0e/24:2)+H	ESI+	856.716519	10.22
PC(18:0e/24:6)+H	ESI+	848.653386	8.22
PC(18:0e/24:7)+H	ESI+	846.636886	7.97
PC(18:0p/18:2)+H	ESI+	770.604848	7.69
PC(18:0p/20:4)+H	ESI+	794.606089	8.13
PC(18:0p/22:6)+H	ESI+	818.601338	7.80
PC(18:0p/24:2)+H	ESI+	854.700519	9.69
PC(18:0p/24:7)+H	ESI+	844.619306	8.27
PC(18:1/18:2)+H	ESI+	784.584342	7.14
PC(18:1/20:4)+H	ESI+	808.583846	6.76
PC(18:1/22:6)+H	ESI+	832.584082	6.67
PC(18:1/24:2)+H	ESI+	868.680408	9.25
PC(18:1/24:6)+H	ESI+	860.614603	7.56
PC(18:1/24:7)+H	ESI+	858.599801	6.79
PC(18:1p/22:5)+H	ESI+	818.603069	7.81
PC(18:1p/24:2)+H	ESI+	852.683912	9.15
PC(18:1p/24:7)+H	ESI+	842.602324	7.47
PC(18:2p/16:0)+H	ESI+	742.573364	6.92
PC(18:2p/17:1)+H	ESI+	754.573327	6.98
PC(18:2p/18:0)+H	ESI+	770.604069	7.57

PC(18:2p/19:1)+H	ESI+	782.605338	7.79
PC(18:2p/24:2)+H	ESI+	850.670074	8.61
PC(18:2p/24:6)+H	ESI+	842.602399	7.49
PC(18:4/20:3)+H	ESI+	804.550058	7.17
PC(18:4/20:5)+H	ESI+	800.519478	5.60
PC(19:0/18:2)+H	ESI+	800.616595	8.19
PC(19:0/20:4)+H	ESI+	824.617406	8.15
PC(19:1/18:2)+H	ESI+	798.59991	7.42
PC(20:0/16:0)+H	ESI+	790.631442	8.93
PC(20:0/18:2)+H	ESI+	814.631169	8.39
PC(20:0/20:4)+H	ESI+	838.63223	8.29
PC(20:0/22:4)+H	ESI+	866.664523	8.76
PC(20:0/22:5)+H	ESI+	864.647419	8.51
PC(20:0/22:6)+H	ESI+	862.631596	7.90
PC(20:0e/24:6)+H	ESI+	876.684225	9.14
PC(20:0e/24:7)+H	ESI+	874.670742	8.70
PC(20:0p/24:7)+H	ESI+	872.653635	8.07
PC(20:1/18:2)+H	ESI+	812.616441	8.02
PC(20:1/20:4)+H	ESI+	836.615623	7.63
PC(20:1p/24:2)+H	ESI+	880.716841	10.45
PC(20:1p/24:7)+H	ESI+	870.633853	8.33
PC(20:3/22:6)+H	ESI+	856.582918	6.15
PC(22:4/22:6)+H	ESI+	882.60593	6.73
PC(22:5/22:6)+H	ESI+	880.589823	6.20
PC(24:0/20:4)+H	ESI+	894.696513	9.98
PC(24:1/20:4)+H	ESI+	892.680398	9.25
PE(15:0/15:0)+H	ESI+	664.490294	6.65
PE(16:0/18:1)+H	ESI+	718.537323	7.70
PE(16:0/18:2)+H	ESI+	716.521513	7.02
PE(16:0/20:4)+H	ESI+	740.52167	6.98
PE(16:0/20:5)+H	ESI+	738.505443	6.33
PE(16:0/22:6)+H	ESI+	764.522193	6.65
PE(16:0e/20:4)+H	ESI+	726.542754	7.56
PE(16:0e/22:6)+H	ESI+	750.543038	7.26
PE(16:0p/18:2)+H	ESI+	700.526553	7.53
PE(16:0p/20:3)+H	ESI+	726.540751	7.77
PE(16:0p/20:4)+H	ESI+	724.526601	7.40
PE(16:0p/20:5)+H	ESI+	722.511127	6.82
PE(16:0p/22:4)+H	ESI+	752.558375	7.99
PE(16:0p/22:5)+H	ESI+	750.542215	7.44
PE(16:1/18:2)+H	ESI+	714.505713	6.41
PE(18:0/18:1)+H	ESI+	746.568655	8.49
PE(18:0/18:2)+H	ESI+	744.553412	7.90
PE(18:0/20:4)+H	ESI+	768.553414	7.85
PE(18:0/22:6)+H	ESI+	792.552615	7.57
PE(18:0p/18:1)+H	ESI+	730.573958	8.94

PE(18:0p/18:2)+H	ESI+	728.55801	8.47
PE(18:0p/20:3)+H	ESI+	754.575104	8.59
PE(18:0p/20:4)+H	ESI+	752.558485	8.23
PE(18:0p/22:4)+H	ESI+	780.590978	8.76
PE(18:0p/22:5)+H	ESI+	778.574522	8.56
PE(18:0p/22:6)+H	ESI+	776.559114	8.11
PE(18:1/18:2)+H	ESI+	742.536726	7.12
PE(18:1/20:4)+H	ESI+	766.537338	7.16
PE(18:1/22:6)+H	ESI+	790.537852	6.85
PE(18:1p/16:0)+H	ESI+	702.543339	8.15
PE(18:1p/18:1)+H	ESI+	728.558361	8.29
PE(18:1p/18:2)+H	ESI+	726.542526	7.58
PE(18:1p/20:3)+H	ESI+	752.559165	7.91
PE(18:1p/20:4)+H	ESI+	750.542594	7.61
PE(18:2/20:4)+H	ESI+	764.521289	6.41
PE(18:2p/20:4)+H	ESI+	748.526184	6.88
PE(20:0p/18:1)+H	ESI+	758.606817	9.70
PE(20:0p/18:2)+H	ESI+	756.589173	9.13
PE(20:0p/20:4)+H	ESI+	780.58979	9.09
PE(20:0p/22:5)+H	ESI+	806.605937	9.03
PE(20:0p/22:6)+H	ESI+	804.591818	8.82
PE(20:1/18:2)+Na	ESI+	792.552127	8.11
PI(16:0/18:1)+H	ESI+	837.549041	5.90
PI(16:0/18:2)+H	ESI+	835.532499	5.22
PI(16:0/20:4)+H	ESI+	859.532679	5.20
PI(16:0/22:5)+NH4	ESI+	902.574912	5.20
PI(16:0/22:6)+H	ESI+	883.533268	4.95
PI(17:0/20:4)+NH4	ESI+	890.575061	5.63
PI(18:0/18:1)+NH4	ESI+	882.605559	6.73
PI(18:0/18:2)+H	ESI+	863.562917	6.07
PI(18:0/20:3)+NH4	ESI+	906.606247	6.38
PI(18:0/20:4)+NH4	ESI+	904.590085	6.07
PI(18:0/22:5)+NH4	ESI+	930.60582	6.04
PI(18:0/22:6)+H	ESI+	911.564482	5.85
PI(18:1/18:1)+H	ESI+	863.563593	6.07
PI(18:1/18:2)+H	ESI+	861.548764	5.39
PI(18:1/20:4)+NH4	ESI+	902.575401	5.38
PI(20:1/18:2)+NH4	ESI+	906.605214	6.39
PI(20:1/20:4)+NH4	ESI+	930.606343	6.05
SM(d16:0/15:1)+H	ESI+	661.528071	5.06
SM(d16:0/16:0)+H	ESI+	677.559307	5.87
SM(d16:0/16:1)+H	ESI+	675.543164	5.60
SM(d16:0/17:1)+H	ESI+	689.557856	6.13
SM(d16:0/18:0)+H	ESI+	705.589275	6.79
SM(d16:0/18:1)+H	ESI+	703.573305	6.61
SM(d16:0/18:2)+H	ESI+	701.559301	5.84

SM(d16:0/18:3)+H	ESI+	699.542618	5.18
SM(d16:0/19:1)+H	ESI+	717.59176	6.85
SM(d16:0/20:0)+H	ESI+	733.622317	7.73
SM(d16:0/20:1)+H	ESI+	731.605264	7.55
SM(d16:0/20:2)+H	ESI+	729.589774	6.69
SM(d16:0/20:3)+H	ESI+	727.573382	6.09
SM(d16:0/22:1)+H	ESI+	759.636774	8.45
SM(d16:0/22:2)+H	ESI+	757.621039	7.67
SM(d16:0/22:3)+H	ESI+	755.605266	6.81
SM(d16:0/24:0)+H	ESI+	789.686065	9.41
SM(d16:0/24:1)+H	ESI+	787.670487	9.22
SM(d16:0/24:2)+H	ESI+	785.652945	8.55
SM(d16:0/24:3)+H	ESI+	783.637137	7.74
SM(d16:0/26:1)+H	ESI+	815.700562	9.87
SM(d16:0/26:2)+H	ESI+	813.684109	9.12
SM(d16:0/26:3)+H	ESI+	811.668978	8.57
SM(d16:0/28:2)+H	ESI+	841.71571	10.03
SM(d16:0/28:3)+H	ESI+	839.701386	9.21
SM(d16:0/28:4)+H	ESI+	837.686275	8.63
SM(d16:1/16:1)+H	ESI+	673.528362	4.70
SM(d18:1/12:0)+H	ESI+	647.513008	4.59
SM(d18:1/15:1)+H	ESI+	687.542938	5.22
SM(d18:1/17:1)+H	ESI+	715.573622	6.25
SM(d18:1/22:6)+H	ESI+	775.573673	5.77
SM(d20:0/19:1)+H	ESI+	773.653578	8.88
SM(d20:1/19:1)+H	ESI+	771.637712	8.10
SM(d22:0/19:1)+H	ESI+	801.686235	9.54
SM(d22:1/19:1)+H	ESI+	799.669406	8.86
SM(d22:2/19:1)+H	ESI+	797.653719	8.09
SM(d24:0/19:1)+H	ESI+	829.71726	10.10
So(d18:1)+H	ESI+	300.289233	2.12
TG(10:0/14:0/16:0)+NH4	ESI+	712.646081	11.01
TG(10:0/16:0/16:1)+NH4	ESI+	738.660456	11.07
TG(10:0/16:0/18:2)+NH4	ESI+	764.676095	11.11
TG(10:0/16:0/18:3)+NH4	ESI+	762.661674	10.85
TG(10:0/18:2/18:2)+NH4	ESI+	788.677841	10.75
TG(12:0/14:0/18:2)+NH4	ESI+	764.676628	11.10
TG(12:0/18:2/18:2)+NH4	ESI+	816.708035	11.21
TG(12:0/18:2/18:3)+NH4	ESI+	814.692525	10.83
TG(12:0/18:2/20:4)+NH4	ESI+	840.707991	11.27
TG(13:0/18:2/18:2)+NH4	ESI+	830.722254	11.48
TG(14:0/14:0/18:2)+NH4	ESI+	792.70635	11.56
TG(14:0/14:0/18:3)+NH4	ESI+	790.691379	11.35
TG(14:0/18:2/18:2)+NH4	ESI+	844.737975	11.78
TG(14:0/18:2/18:3)+NH4	ESI+	842.72227	11.45
TG(14:0/18:2/20:4)+Na	ESI+	873.69398	11.64

TG(14:0/18:2/20:5)+Na	ESI+	871.679147	11.22
TG(14:0/18:2/22:6)+NH4	ESI+	892.738763	11.53
TG(14:0/18:3/18:3)+NH4	ESI+	840.709078	10.90
TG(14:0/18:3/22:6)+NH4	ESI+	890.724422	11.05
TG(14:0/20:4/20:4)+NH4	ESI+	892.738778	11.52
TG(14:0/20:4/22:6)+NH4	ESI+	916.739714	11.28
TG(15:0/14:0/18:1)+NH4	ESI+	808.737944	12.21
TG(15:0/14:0/18:2)+NH4	ESI+	806.722321	11.83
TG(15:0/15:0/15:0)+NH4	ESI+	782.722166	12.09
TG(15:0/18:1/18:2)+NH4	ESI+	860.769408	12.25
TG(15:0/18:1/19:1)+NH4	ESI+	876.800993	12.82
TG(15:0/18:1/20:4)+NH4	ESI+	884.769577	12.23
TG(15:0/18:1/22:5)+NH4	ESI+	910.783597	12.16
TG(15:0/18:1/22:6)+NH4	ESI+	908.769758	11.99
TG(15:0/18:2/18:2)+NH4	ESI+	858.753386	11.89
TG(15:0/18:2/18:3)+NH4	ESI+	856.738268	11.54
TG(15:0/18:2/20:4)+NH4	ESI+	882.752762	11.86
TG(15:0/18:2/20:5)+NH4	ESI+	880.7377	11.44
TG(15:0/18:2/22:5)+NH4	ESI+	908.769656	11.80
TG(15:0/18:2/22:6)+NH4	ESI+	906.753924	11.60
TG(15:0/18:3/20:5)+H	ESI+	861.694052	11.54
TG(15:1/18:1/18:1)+NH4	ESI+	860.769198	12.42
TG(15:1/18:1/18:2)+NH4	ESI+	858.753482	12.07
TG(15:1/18:1/19:1)+NH4	ESI+	874.78536	12.47
TG(15:1/18:2/18:2)+NH4	ESI+	856.737919	11.71
TG(16:0/12:0/14:0)+NH4	ESI+	740.675383	11.46
TG(16:0/12:0/16:1)+NH4	ESI+	766.69057	11.51
TG(16:0/12:0/18:3)+NH4	ESI+	790.691148	11.33
TG(16:0/14:0/14:0)+NH4	ESI+	768.706629	11.91
TG(16:0/14:0/15:1)+NH4	ESI+	780.705666	11.81
TG(16:0/14:0/16:0)+NH4	ESI+	796.738138	12.35
TG(16:0/14:0/16:1)+NH4	ESI+	794.72225	11.95
TG(16:0/14:0/18:1)+NH4	ESI+	822.752489	12.50
TG(16:0/14:0/18:2)+NH4	ESI+	820.738448	11.97
TG(16:0/14:0/18:3)+NH4	ESI+	818.722195	11.77
TG(16:0/14:0/20:4)+NH4	ESI+	844.738186	11.99
TG(16:0/14:0/20:5)+Na	ESI+	847.678407	11.63
TG(16:0/14:0/22:6)+NH4	ESI+	868.738501	11.73
TG(16:0/15:0/16:0)+NH4	ESI+	810.754688	12.59
TG(16:0/15:0/18:1)+NH4	ESI+	836.769541	12.60
TG(16:0/15:0/18:2)+NH4	ESI+	834.753697	12.25
TG(16:0/15:1/18:2)+NH4	ESI+	832.737573	12.03
TG(16:0/16:0/16:0)+NH4	ESI+	824.769619	12.68
TG(16:0/16:0/17:0)+NH4	ESI+	838.785397	12.92
TG(16:0/16:0/18:1)+NH4	ESI+	850.785184	12.87
TG(16:0/16:0/18:2)+NH4	ESI+	848.771969	12.88

TG(16:0/16:0/22:5)+NH4	ESI+	898.784725	12.15
TG(16:0/16:1/16:2)+NH4	ESI+	818.723037	12.00
TG(16:0/16:1/18:1)+NH4	ESI+	848.769364	12.52
TG(16:0/16:1/18:2)+NH4	ESI+	846.753462	12.17
TG(16:0/16:1/18:3)+NH4	ESI+	844.737901	11.81
TG(16:0/16:1/22:6)+NH4	ESI+	894.754128	11.87
TG(16:0/17:0/18:1)+NH4	ESI+	864.801465	12.96
TG(16:0/17:0/20:4)+NH4	ESI+	886.785248	12.59
TG(16:0/17:0/22:6)+NH4	ESI+	910.784836	12.39
TG(16:0/17:1/18:1)+NH4	ESI+	862.786037	12.66
TG(16:0/17:1/18:2)+NH4	ESI+	860.770359	12.47
TG(16:0/17:1/19:1)+NH4	ESI+	876.793652	12.45
TG(16:0/17:1/20:4)+NH4	ESI+	884.768908	12.25
TG(16:0/18:1/18:1)+NH4	ESI+	876.801605	12.80
TG(16:0/18:1/18:2)+NH4	ESI+	874.785208	12.53
TG(16:0/18:1/18:3)+NH4	ESI+	872.769637	12.26
TG(16:0/18:1/20:4)+NH4	ESI+	898.785796	12.40
TG(16:0/18:1/22:0)+NH4	ESI+	934.880655	13.79
TG(16:0/18:1/22:1)+NH4	ESI+	932.866034	13.45
TG(16:0/18:1/22:4)+NH4	ESI+	926.817353	12.72
TG(16:0/18:1/22:5)+NH4	ESI+	924.801463	12.36
TG(16:0/18:1/22:6)+NH4	ESI+	922.785102	12.18
TG(16:0/18:1/24:0)+NH4	ESI+	962.911328	14.13
TG(16:0/18:1/24:1)+NH4	ESI+	960.896763	13.79
TG(16:0/18:2/18:2)+NH4	ESI+	872.769429	12.09
TG(16:0/18:2/18:3)+NH4	ESI+	870.753845	11.91
TG(16:0/18:2/20:4)+NH4	ESI+	896.76955	12.05
TG(16:0/18:2/20:5)+NH4	ESI+	894.753766	11.69
TG(16:0/18:2/21:4)+NH4	ESI+	910.784286	12.17
TG(16:0/18:2/22:5)+NH4	ESI+	922.785139	12.01
TG(16:0/18:2/22:6)+NH4	ESI+	920.769591	11.83
TG(16:0/18:3/20:4)+NH4	ESI+	894.754048	11.89
TG(16:0/18:3/22:6)+NH4	ESI+	918.754113	11.56
TG(16:0/22:5/22:6)+NH4	ESI+	970.785172	11.74
TG(16:0/22:6/22:6)+NH4	ESI+	968.76919	11.55
TG(16:0e/16:0/18:1)+NH4	ESI+	836.807857	13.28
TG(16:0e/18:1/18:1)+Na	ESI+	867.779787	13.34
TG(16:0e/18:1/18:2)+NH4	ESI+	860.807264	13.02
TG(16:1/12:0/18:1)+NH4	ESI+	792.70609	11.76
TG(16:1/12:0/18:2)+NH4	ESI+	790.691584	11.15
TG(16:1/14:0/14:0)+Na	ESI+	771.646217	11.52
TG(16:1/14:0/18:1)+NH4	ESI+	820.737606	12.15
TG(16:1/14:0/18:2)+NH4	ESI+	818.722148	11.76
TG(16:1/14:0/18:3)+NH4	ESI+	816.707071	11.39
TG(16:1/14:0/22:6)+NH4	ESI+	866.722786	11.37
TG(16:1/15:0/18:2)+NH4	ESI+	832.73805	11.84

TG(16:1/15:1/18:2)+NH4	ESI+	830.722771	11.65
TG(16:1/16:1/18:2)+Na	ESI+	849.693143	11.81
TG(16:1/18:1/18:1)+NH4	ESI+	874.786909	13.91
TG(16:1/18:1/18:2)+NH4	ESI+	872.772893	12.51
TG(16:1/18:2/18:2)+NH4	ESI+	870.753615	11.73
TG(16:1/18:2/18:3)+NH4	ESI+	868.73841	11.55
TG(16:1/18:2/20:5)+NH4	ESI+	892.738733	11.33
TG(16:1/18:2/22:5)+NH4	ESI+	920.769472	11.65
TG(16:1/18:2/22:6)+NH4	ESI+	918.75365	11.47
TG(16:1/18:3/18:3)+NH4	ESI+	866.724487	11.00
TG(16:2/18:2/18:3)+NH4	ESI+	866.724347	11.19
TG(16:2/18:2/22:6)+NH4	ESI+	916.738938	11.10
TG(17:0/18:1/18:1)+NH4	ESI+	890.818063	13.00
TG(17:0/18:1/18:2)+NH4	ESI+	888.801097	12.64
TG(17:0/18:1/20:4)+NH4	ESI+	912.797497	12.56
TG(17:0/18:1/22:5)+NH4	ESI+	938.817671	12.57
TG(17:0/18:1/22:6)+NH4	ESI+	936.800841	12.38
TG(17:0/18:2/18:2)+NH4	ESI+	886.7853	12.28
TG(17:0/18:2/22:5)+NH4	ESI+	936.800092	12.24
TG(17:0/18:2/22:6)+NH4	ESI+	934.785358	12.06
TG(17:1/18:1/18:1)+NH4	ESI+	888.799835	12.84
TG(17:1/18:1/18:2)+NH4	ESI+	886.784664	12.45
TG(17:1/18:1/19:1)+NH4	ESI+	902.816654	12.87
TG(17:1/18:1/22:5)+NH4	ESI+	936.800268	12.24
TG(17:1/18:2/18:2)+NH4	ESI+	884.769556	12.05
TG(17:1/18:2/18:3)+NH4	ESI+	882.753504	11.67
TG(17:1/18:2/20:4)+NH4	ESI+	908.769439	11.95
TG(17:1/18:2/22:5)+NH4	ESI+	934.785242	11.88
TG(18:0/16:0/16:0)+NH4	ESI+	852.802964	13.09
TG(18:0/16:0/18:1)+NH4	ESI+	878.817277	13.16
TG(18:0/16:0/20:4)+Na	ESI+	905.757226	12.76
TG(18:0/17:0/18:1)+NH4	ESI+	892.834668	13.26
TG(18:0/17:0/20:4)+NH4	ESI+	914.817612	12.92
TG(18:0/18:0/18:1)+NH4	ESI+	906.850521	13.49
TG(18:0/18:0/20:4)+NH4	ESI+	928.833895	13.12
TG(18:0/18:1/18:1)+NH4	ESI+	904.834802	13.30
TG(18:0/18:1/18:2)+NH4	ESI+	902.818251	13.88
TG(18:0/18:1/20:3)+NH4	ESI+	928.834394	13.08
TG(18:0/18:1/20:4)+NH4	ESI+	926.816956	12.90
TG(18:0/18:1/22:5)+NH4	ESI+	952.832696	12.74
TG(18:0/18:1/22:6)+NH4	ESI+	950.817377	12.67
TG(18:0/18:2/20:4)+NH4	ESI+	924.800974	12.76
TG(18:0/20:0/20:4)+H	ESI+	939.8367	13.79
TG(18:0/20:1/20:4)+NH4	ESI+	954.849309	13.10
TG(18:0/20:3/22:5)+NH4	ESI+	976.831036	12.30
TG(18:0/20:4/20:4)+NH4	ESI+	948.801105	12.37

TG(18:0/20:4/22:5)+NH4	ESI+	974.816944	12.33
TG(18:0p/16:0/18:1)+NH4	ESI+	862.824588	13.32
TG(18:1/12:0/14:0)+NH4	ESI+	766.690701	11.52
TG(18:1/14:0/18:2)+NH4	ESI+	846.75431	12.72
TG(18:1/14:0/18:3)+NH4	ESI+	844.738495	12.05
TG(18:1/18:1/18:1)+NH4	ESI+	902.817133	12.93
TG(18:1/18:1/18:2)+NH4	ESI+	900.801379	12.58
TG(18:1/18:1/18:3)+NH4	ESI+	898.785709	12.34
TG(18:1/18:1/19:0)+NH4	ESI+	918.851189	13.33
TG(18:1/18:1/19:1)+NH4	ESI+	916.834595	13.12
TG(18:1/18:1/20:2)+NH4	ESI+	928.831644	12.57
TG(18:1/18:1/20:3)+NH4	ESI+	926.816528	12.60
TG(18:1/18:1/20:4)+NH4	ESI+	924.801122	12.73
TG(18:1/18:1/20:5)+NH4	ESI+	922.785868	11.48
TG(18:1/18:1/22:1)+NH4	ESI+	958.881396	13.49
TG(18:1/18:1/22:3)+NH4	ESI+	954.848776	12.96
TG(18:1/18:1/22:4)+NH4	ESI+	952.832897	12.77
TG(18:1/18:1/22:5)+NH4	ESI+	950.816725	12.56
TG(18:1/18:1/22:6)+NH4	ESI+	948.800365	12.22
TG(18:1/18:1/24:0)+NH4	ESI+	988.926441	14.17
TG(18:1/18:1/24:1)+NH4	ESI+	986.911693	13.84
TG(18:1/18:1/24:6)+NH4	ESI+	976.831679	12.54
TG(18:1/18:2/18:2)+NH4	ESI+	898.785293	12.22
TG(18:1/18:2/18:3)+NH4	ESI+	896.769553	11.87
TG(18:1/18:2/19:1)+NH4	ESI+	914.817149	12.74
TG(18:1/18:2/20:2)+NH4	ESI+	926.816868	12.61
TG(18:1/18:2/20:4)+NH4	ESI+	922.785398	12.05
TG(18:1/18:2/22:1)+NH4	ESI+	956.865656	13.28
TG(18:1/18:2/22:4)+NH4	ESI+	950.8164	12.37
TG(18:1/18:2/22:5)+NH4	ESI+	948.800365	12.19
TG(18:1/18:2/24:1)+NH4	ESI+	984.897112	13.56
TG(18:1/19:1/19:1)+NH4	ESI+	930.846567	12.94
TG(18:1/20:3/20:5)+NH4	ESI+	946.784195	11.67
TG(18:1/20:4/20:4)+NH4	ESI+	946.785165	12.01
TG(18:1/20:4/20:5)+NH4	ESI+	944.769337	11.65
TG(18:1/20:4/22:4)+NH4	ESI+	974.816223	12.31
TG(18:1/20:4/22:5)+NH4	ESI+	972.800675	11.93
TG(18:1/20:4/22:6)+NH4	ESI+	970.784787	11.74
TG(18:1/22:6/22:6)+NH4	ESI+	994.784759	11.58
TG(18:2/18:2/18:2)+NH4	ESI+	896.768796	11.78
TG(18:2/18:2/20:4)+NH4	ESI+	920.769847	11.68
TG(18:2/18:2/22:6)+NH4	ESI+	944.768995	11.48
TG(18:2/20:4/20:4)+NH4	ESI+	944.769875	11.67
TG(18:2/20:4/22:6)+NH4	ESI+	968.769887	11.37
TG(18:2/22:6/22:6)+NH4	ESI+	992.770169	11.19
TG(18:3/18:2/18:2)+NH4	ESI+	894.754021	11.51

TG(18:3/18:2/18:3)+NH4	ESI+	892.739109	11.14
TG(18:3/18:2/20:4)+NH4	ESI+	918.755071	11.29
TG(18:3/18:2/20:5)+NH4	ESI+	916.740474	10.93
TG(18:3/18:2/22:6)+NH4	ESI+	942.754816	11.11
TG(18:3/18:3/18:3)+NH4	ESI+	890.725365	10.66
TG(18:3/20:5/20:5)+H	ESI+	921.695838	10.93
TG(18:3/20:5/22:6)+H	ESI+	947.710337	11.10
TG(18:4/12:0/18:2)+H	ESI+	795.647429	11.16
TG(18:4/14:0/16:0)+Na	ESI+	821.662306	11.41
TG(18:4/14:0/18:2)+NH4	ESI+	840.707989	11.08
TG(18:4/15:0/18:2)+NH4	ESI+	854.723895	11.28
TG(18:4/16:1/18:2)+NH4	ESI+	866.723989	11.18
TG(19:1/18:2/18:2)+NH4	ESI+	912.801668	12.37
TG(20:0/18:1/18:1)+NH4	ESI+	932.866539	13.55
TG(20:0/18:1/18:2)+NH4	ESI+	930.85051	13.30
TG(20:1/17:0/18:1)+NH4	ESI+	918.851074	13.32
TG(20:1/18:1/18:1)+NH4	ESI+	930.850096	13.26
TG(20:1/18:1/18:2)+NH4	ESI+	928.832447	12.89
TG(20:3/18:2/20:4)+NH4	ESI+	946.784585	11.83
TG(20:5/18:2/20:4)+NH4	ESI+	942.755328	11.29
TG(20:5/18:2/20:5)+NH4	ESI+	940.739357	10.85
TG(20:5/18:2/22:6)+NH4	ESI+	966.754958	11.01
TG(20:5/20:5/22:6)+H	ESI+	971.710998	11.01
TG(21:4/18:2/18:2)+NH4	ESI+	934.785891	11.88
TG(22:0/18:2/18:2)+NH4	ESI+	956.866036	13.34
TG(22:4/18:2/20:4)+NH4	ESI+	972.801531	11.93
TG(22:5/18:2/18:2)+NH4	ESI+	946.784837	11.83
TG(24:1/18:2/18:2)+NH4	ESI+	982.880675	13.29
TG(24:4/18:1/18:3)+H	ESI+	959.804288	12.98
TG(4:0/16:0/16:0)+NH4	ESI+	656.583559	9.94
TG(4:0/16:0/18:1)+NH4	ESI+	682.598987	10.04
TG(6:0/16:0/16:0)+NH4	ESI+	684.614507	10.56
TG(6:0/16:0/18:2)+NH4	ESI+	708.61543	10.11
TG(8:0/16:0/18:2)+NH4	ESI+	736.646195	10.62
TG(8:0/16:0/20:4)+H	ESI+	743.616492	11.11
Cer(d18:1/16:0)-H	ESI-	536.504505	7.52
Cer(d18:1/17:0)-H	ESI-	550.519977	7.96
Cer(d18:1/22:0)-H	ESI-	620.59763	9.91
Cer(d18:1/24:0)-H	ESI-	648.629776	10.57
Cer(d18:1/24:1)-H	ESI-	646.613759	9.89
Cer(d18:2/22:0)-H	ESI-	618.582181	9.38
Cer(d18:2/24:0)-H	ESI-	646.613679	10.08
Cer(d18:2/24:1)-H	ESI-	644.598115	9.34
dMePE(16:0/18:1)-H	ESI-	744.554737	7.55
dMePE(16:0/18:2)-H	ESI-	742.538793	6.88
dMePE(16:0/22:6)-H	ESI-	790.534455	6.61

dMePE(18:0/18:2)-H	ESI-	770.570521	7.73
dMePE(18:0/20:4)-H	ESI-	794.569814	7.72
dMePE(18:1/18:2)-H	ESI-	768.553443	7.06
dMePE(18:2/18:2)-H	ESI-	766.538308	6.38
dMePE(18:2p/22:6)-H	ESI-	798.544706	7.41
dMePE(20:0p/20:4)-H	ESI-	806.607391	9.79
dMePE(20:1p/20:4)-H	ESI-	804.590893	9.07
dMePE(20:1p/22:6)-H	ESI-	828.591854	8.80
FA(10:0)-H	ESI-	171.13904	1.20
FA(11:0)-H	ESI-	185.15469	1.28
FA(12:0)-H	ESI-	199.1703	1.37
FA(13:0)-H	ESI-	213.18599	1.51
FA(14:0)-H	ESI-	227.20164	1.65
FA(14:1)-H	ESI-	225.18599	1.46
FA(15:0)-H	ESI-	241.21729	1.81
FA(15:1)-H	ESI-	239.20164	1.60
FA(16:0)-H	ESI-	255.23294	2.05
IS-FFA 16:0_d3	ESI-	258.2518	2.05
FA(16:1)-H	ESI-	253.21729	1.74
FA(17:0)-H	ESI-	269.24859	2.33
FA(17:1)-H	ESI-	267.23294	1.93
FA(17:2)-H	ESI-	265.21729	1.65
FA(18:0)-H	ESI-	283.26424	2.68
IS-FFA 18:0_d3	ESI-	286.2831	2.67
FA(18:1)-H	ESI-	281.24859	2.19
FA(18:2)-H	ESI-	279.23294	1.84
FA(18:3)-H	ESI-	277.21729	1.60
FA(18:4)-H	ESI-	275.20164	1.46
FA(19:0)-H	ESI-	297.27989	3.00
FA(20:0)-H	ESI-	311.29554	3.58
FA(20:1)-H	ESI-	309.27989	2.83
FA(20:2)-H	ESI-	307.26424	2.28
FA(20:3)-H	ESI-	305.24859	1.96
FA(20:4)-H	ESI-	303.232525	1.75
FA(20:5)-H	ESI-	301.216975	1.54
FA(20:6)-H	ESI-	299.20164	1.60
FA(21:0)-H	ESI-	325.31119	4.05
FA(22:0)-H	ESI-	339.32684	4.64
FA(22:1)-H	ESI-	337.31119	3.68
FA(22:2)-H	ESI-	335.29554	2.99
FA(22:3)-H	ESI-	333.27989	2.50
FA(22:4)-H	ESI-	331.26424	2.14
FA(22:5)-H	ESI-	329.24859	1.84
FA(22:6)-H	ESI-	327.23294	1.63
FA(23:0)-H	ESI-	353.34249	5.20
FA(24:0)-H	ESI-	367.35814	5.77

FA(24:1)-H	ESI-	365.34249	4.69
FA(24:2)-H	ESI-	363.32684	3.89
FA(24:3)-H	ESI-	361.31119	3.40
FA(24:4)-H	ESI-	359.29554	2.70
FA(24:5)-H	ESI-	357.27989	2.28
FA(24:6)-H	ESI-	355.263812	2.00
FA(25:0)-H	ESI-	381.37379	5.72
FA(26:0)-H	ESI-	395.38944	6.24
FA(26:1)-H	ESI-	393.37384	5.36
LdMePE(16:0)-H	ESI-	480.308967	1.69
LdMePE(18:0)-H	ESI-	508.340617	2.21
LdMePE(18:1)-H	ESI-	506.325328	1.87
LdMePE(18:2)-H	ESI-	504.309025	1.60
LdMePE(19:0)-H	ESI-	522.355633	2.57
LPC(14:0)+CH3COO	ESI-	526.314757	1.46
LPC(15:0)+CH3COO	ESI-	540.330328	1.57
LPC(16:0)+CH3COO	ESI-	554.34612	1.83
LPC(16:1)+CH3COO	ESI-	552.330572	1.53
LPC(17:1)+CH3COO	ESI-	566.346294	1.70
LPC(18:0)+CH3COO	ESI-	582.37713	2.08
LPC(18:1)+CH3COO	ESI-	580.361815	1.95
LPC(18:2)+CH3COO	ESI-	578.346187	1.51
LPC(18:3)+CH3COO	ESI-	576.330518	1.43
LPC(19:0)+CH3COO	ESI-	596.393466	2.58
LPC(20:0)+CH3COO	ESI-	610.408936	2.93
LPC(20:1)+CH3COO	ESI-	608.393219	2.33
LPC(20:2)+CH3COO	ESI-	606.377403	1.94
LPC(20:3)+CH3COO	ESI-	604.361788	1.69
LPC(20:4)+CH3COO	ESI-	602.345902	1.50
LPC(20:5)+CH3COO	ESI-	600.330674	1.41
LPC(22:0)+CH3COO	ESI-	638.440788	3.92
LPC(22:1)+CH3COO	ESI-	636.4248	3.06
LPC(22:5)+CH3COO	ESI-	628.361305	1.61
LPC(22:6)+CH3COO	ESI-	626.346392	1.53
LPC(24:0)+CH3COO	ESI-	666.471696	4.94
LPE(16:0)-H	ESI-	452.277953	1.74
LPE(16:0p)-H	ESI-	436.283002	2.00
LPE(18:0)-H	ESI-	480.309226	2.35
LPE(18:0p)-H	ESI-	464.314433	2.59
LPE(18:1)-H	ESI-	478.293197	1.85
LPE(18:2)-H	ESI-	476.277919	1.62
LPE(20:3)-H	ESI-	502.293692	1.76
LPE(20:4)-H	ESI-	500.278125	1.58
LPE(20:5)-H	ESI-	498.262668	1.44
LPE(22:5)-H	ESI-	526.293455	1.66
LPE(22:6)-H	ESI-	524.277895	1.55

OAHFA(16:0/18:2)-H	ESI-	533.454681	2.05
OAHFA(18:0/20:2)-H	ESI-	589.517319	2.67
OAHFA(18:1/18:0)-H	ESI-	563.503959	2.19
OAHFA(18:1/20:3)-H	ESI-	585.486034	2.20
OAHFA(18:2/18:1)-H	ESI-	559.472515	1.81
OAHFA(18:2/22:6)-H	ESI-	605.454331	1.82
OAHFA(20:4/20:3)-H	ESI-	607.473	1.78
OAHFA(20:4/22:6)-H	ESI-	629.454462	1.75
OAHFA(22:5/18:1)-H	ESI-	609.488539	1.84
OAHFA(22:6/22:5)-H	ESI-	655.472738	1.65
PC(14:0/18:2)+CH3COO	ESI-	788.544741	5.99
PC(14:0/20:4)+CH3COO	ESI-	812.544663	5.89
PC(14:0/22:6)+CH3COO	ESI-	836.544699	5.62
PC(15:0/18:1)+CH3COO	ESI-	804.576942	7.21
PC(15:0/18:2)+CH3COO	ESI-	802.560322	6.46
PC(15:0/20:4)+CH3COO	ESI-	826.560801	6.39
PC(15:0/22:6)+CH3COO	ESI-	850.560407	6.07
PC(16:0/15:0)+CH3COO	ESI-	778.561708	6.96
PC(16:0/16:0)+CH3COO	ESI-	792.57596	7.32
PC(16:0/16:2)+CH3COO	ESI-	788.545074	6.18
PC(16:0/17:1)+CH3COO	ESI-	804.577163	7.24
PC(16:0/18:1)+CH3COO	ESI-	818.59195	7.49
PC(16:0/18:2)+CH3COO	ESI-	816.576111	6.96
PC(16:0/18:3)+CH3COO	ESI-	814.560693	6.46
PC(16:0/20:2)+CH3COO	ESI-	844.607637	7.64
PC(16:0/20:3)+CH3COO	ESI-	842.59149	7.28
PC(16:0/20:4)+CH3COO	ESI-	840.575404	6.92
PC(16:0/20:5)+CH3COO	ESI-	838.560459	6.20
PC(16:0/22:4)+CH3COO	ESI-	868.606363	7.47
PC(16:0/22:5)+CH3COO	ESI-	866.591732	6.78
PC(16:0/22:6)+CH3COO	ESI-	864.576404	6.60
PC(16:0e/18:1)+CH3COO	ESI-	804.612067	8.18
PC(16:0e/18:2)+CH3COO	ESI-	802.597011	7.54
PC(16:0e/20:4)+CH3COO	ESI-	826.597207	7.40
PC(16:0e/22:5)+CH3COO	ESI-	852.611718	7.32
PC(16:0e/22:6)+CH3COO	ESI-	850.59641	7.18
PC(16:0p/16:0)+CH3COO	ESI-	776.58114	7.88
PC(16:0p/18:1)+CH3COO	ESI-	802.596437	8.00
PC(16:0p/18:2)+CH3COO	ESI-	800.581045	7.35
PC(16:0p/20:3)+CH3COO	ESI-	826.595744	7.59
PC(16:0p/20:4)+CH3COO	ESI-	824.581317	7.23
PC(16:0p/20:5)+CH3COO	ESI-	822.56551	6.66
PC(16:0p/22:4)+CH3COO	ESI-	852.611931	7.84
PC(16:0p/22:5)+CH3COO	ESI-	850.596818	7.23
PC(16:0p/22:6)+CH3COO	ESI-	848.581319	7.01
PC(16:1/18:2)+CH3COO	ESI-	814.55962	6.08

PC(16:1/22:6)+CH3COO	ESI-	862.559674	5.97
PC(16:2/18:2)+CH3COO	ESI-	812.544706	5.56
PC(17:0/18:1)+CH3COO	ESI-	832.607818	7.91
PC(17:0/18:2)+CH3COO	ESI-	830.592067	7.35
PC(17:0/20:3)+CH3COO	ESI-	856.607317	7.59
PC(17:0/20:4)+CH3COO	ESI-	854.591578	7.33
PC(17:0/22:6)+CH3COO	ESI-	878.591864	6.98
PC(17:1/18:2)+CH3COO	ESI-	828.576419	6.61
PC(18:0/16:0)+CH3COO	ESI-	820.607355	8.21
PC(18:0/18:1)+CH3COO	ESI-	846.622793	8.36
PC(18:0/18:2)+CH3COO	ESI-	844.607778	7.82
PC(18:0/20:1)+CH3COO	ESI-	874.654396	9.01
PC(18:0/20:2)+CH3COO	ESI-	872.63827	8.38
PC(18:0/20:3)+CH3COO	ESI-	870.623081	8.05
PC(18:0/20:4)+CH3COO	ESI-	868.607958	7.69
PC(18:0/20:5)+CH3COO	ESI-	866.591657	7.14
PC(18:0/22:4)+CH3COO	ESI-	896.638249	8.33
PC(18:0/22:5)+CH3COO	ESI-	894.623097	7.69
PC(18:0/22:6)+CH3COO	ESI-	892.607729	7.51
PC(18:0e/18:2)+CH3COO	ESI-	830.625979	8.25
PC(18:0e/22:4)+CH3COO	ESI-	882.659673	8.73
PC(18:0e/22:5)+CH3COO	ESI-	880.643172	8.19
PC(18:0e/22:6)+CH3COO	ESI-	878.627329	8.02
PC(18:0p/18:1)+CH3COO	ESI-	830.628719	8.82
PC(18:0p/18:2)+CH3COO	ESI-	828.61217	7.59
PC(18:0p/20:4)+CH3COO	ESI-	852.612079	8.09
PC(18:0p/22:6)+CH3COO	ESI-	876.611947	7.86
PC(18:1/18:2)+CH3COO	ESI-	842.591512	7.10
PC(18:1/20:3)+CH3COO	ESI-	868.606965	7.32
PC(18:1/20:4)+CH3COO	ESI-	866.591673	6.96
PC(18:1/20:5)+CH3COO	ESI-	864.57597	6.42
PC(18:1/22:5)+CH3COO	ESI-	892.607786	6.95
PC(18:1/22:6)+CH3COO	ESI-	890.591918	6.74
PC(18:1p/18:1)+CH3COO	ESI-	828.611552	8.11
PC(18:1p/20:4)+CH3COO	ESI-	850.596615	7.41
PC(18:1p/22:6)+CH3COO	ESI-	874.596844	7.15
PC(18:2/18:2)+CH3COO	ESI-	840.576314	6.38
PC(18:2/20:4)+CH3COO	ESI-	864.576184	6.24
PC(18:2/22:6)+CH3COO	ESI-	888.575719	5.94
PC(18:2p/20:4)+CH3COO	ESI-	848.580298	6.73
PC(18:3/18:2)+CH3COO	ESI-	838.559994	5.79
PC(19:0/18:2)+CH3COO	ESI-	858.622948	8.18
PC(19:0/19:0)+CH3COO	ESI-	876.670005	9.59
PC(19:1/18:2)+CH3COO	ESI-	856.607031	7.41
PC(20:0/20:4)+CH3COO	ESI-	896.6388	8.51
PC(20:0e/20:4)+CH3COO	ESI-	882.659014	9.00

PC(20:0p/20:4)+CH3COO	ESI-	880.643429	8.24
PC(20:1/18:2)+CH3COO	ESI-	870.621675	7.87
PC(20:1/22:6)+CH3COO	ESI-	918.623723	7.51
PC(20:1p/20:4)+CH3COO	ESI-	878.627435	8.19
PC(20:2/20:4)+CH3COO	ESI-	892.607153	7.14
PC(20:5/18:2)+CH3COO	ESI-	862.560634	5.60
PC(22:5/18:2)+CH3COO	ESI-	890.591737	6.26
PE(15:0/15:0)-H	ESI-	662.476785	6.67
PE(16:0/16:1)-H	ESI-	688.491077	6.90
PE(16:0/18:1)-H	ESI-	716.52371	7.66
PE(16:0/18:2)-H	ESI-	714.507947	7.04
PE(16:0/18:3)-H	ESI-	712.492137	6.37
PE(16:0/20:2)-H	ESI-	742.540125	7.75
PE(16:0/20:3)-H	ESI-	740.524118	7.28
PE(16:0/20:4)-H	ESI-	738.505867	6.91
PE(16:0/20:5)-H	ESI-	736.492226	6.34
PE(16:0/22:4)-H	ESI-	766.538932	7.55
PE(16:0/22:5)-H	ESI-	764.523774	7.37
PE(16:0/22:6)-H	ESI-	762.50828	6.75
PE(16:0e/20:4)-H	ESI-	724.528674	7.61
PE(16:0p/18:1)-H	ESI-	700.528787	8.18
PE(16:0p/18:2)-H	ESI-	698.513431	7.49
PE(16:0p/18:3)-H	ESI-	696.495932	6.91
PE(16:0p/20:3)-H	ESI-	724.528757	7.78
PE(16:0p/20:4)-H	ESI-	722.513362	7.43
PE(16:0p/20:5)-H	ESI-	720.49743	6.80
PE(16:0p/22:4)-H	ESI-	750.544672	8.09
PE(16:0p/22:5)-H	ESI-	748.528418	7.37
PE(16:0p/22:6)-H	ESI-	746.513064	7.19
PE(16:1p/20:4)-H	ESI-	720.497876	6.98
PE(17:0/18:2)-H	ESI-	728.525037	7.47
PE(17:0/22:6)-H	ESI-	776.523512	7.12
PE(18:0/18:1)-H	ESI-	744.555291	8.48
PE(18:0/18:2)-H	ESI-	742.539368	7.96
PE(18:0/20:3)-H	ESI-	768.555004	8.14
PE(18:0/20:4)-H	ESI-	766.537699	7.78
PE(18:0/22:4)-H	ESI-	794.569766	8.36
PE(18:0/22:5)-H	ESI-	792.554744	8.15
PE(18:0/22:6)-H	ESI-	790.539327	7.57
PE(18:0e/22:6)-H	ESI-	776.559645	8.13
PE(18:0p/18:1)-H	ESI-	728.56021	8.95
PE(18:0p/18:2)-H	ESI-	726.544315	8.43
PE(18:0p/20:3)-H	ESI-	752.559625	8.59
PE(18:0p/20:4)-H	ESI-	750.54442	8.41
PE(18:0p/20:5)-H	ESI-	748.528892	7.73
PE(18:0p/22:4)-H	ESI-	778.576063	8.79

PE(18:0p/22:5)-H	ESI-	776.56053	8.53
PE(18:0p/22:6)-H	ESI-	774.544449	7.98
PE(18:1/18:1)-H	ESI-	742.539911	7.79
PE(18:1/18:2)-H	ESI-	740.524051	7.26
PE(18:1/20:4)-H	ESI-	764.523662	7.19
PE(18:1/22:5)-H	ESI-	790.537723	7.13
PE(18:1/22:6)-H	ESI-	788.523303	6.88
PE(18:1p/16:0)-H	ESI-	700.528869	8.16
PE(18:1p/18:1)-H	ESI-	726.544323	8.31
PE(18:1p/18:2)-H	ESI-	724.528425	7.74
PE(18:1p/20:3)-H	ESI-	750.544203	7.91
PE(18:1p/20:4)-H	ESI-	748.528961	7.55
PE(18:1p/20:5)-H	ESI-	746.513035	7.01
PE(18:1p/22:5)-H	ESI-	774.543982	7.61
PE(18:1p/22:6)-H	ESI-	772.528451	7.25
PE(18:2/18:2)-H	ESI-	738.508467	6.50
PE(18:2/20:4)-H	ESI-	762.508	6.39
PE(18:2p/18:2)-H	ESI-	722.513118	7.00
PE(18:2p/22:6)-H	ESI-	770.512361	6.57
PE(20:0p/18:2)-H	ESI-	754.575434	9.13
PE(20:0p/20:4)-H	ESI-	778.575965	9.04
PE(20:0p/22:6)-H	ESI-	802.57612	8.79
PE(20:1p/20:4)-H	ESI-	776.559937	8.34
PE(20:1p/22:6)-H	ESI-	800.560187	8.08
PG(18:0/18:1)-H	ESI-	775.548757	6.86
PG(18:0/18:2)-H	ESI-	773.534471	6.30
PG(18:1/18:2)-H	ESI-	771.518024	5.21
PI(16:0/16:1)-H	ESI-	807.502996	5.05
PI(16:0/18:1)-H	ESI-	835.534311	5.85
PI(16:0/18:2)-H	ESI-	833.519189	5.22
PI(16:0/20:3)-H	ESI-	859.534575	5.53
PI(16:0/20:4)-H	ESI-	857.518742	5.16
PI(16:0/22:4)-H	ESI-	885.550058	5.76
PI(16:0/22:5)-H	ESI-	883.53407	5.17
PI(16:0/22:6)-H	ESI-	881.51888	4.92
PI(17:0/18:2)-H	ESI-	847.533775	5.66
PI(17:0/20:4)-H	ESI-	871.534262	5.62
PI(18:0/16:1)-H	ESI-	835.534298	6.00
PI(18:0/16:2)-H	ESI-	833.518824	5.44
PI(18:0/18:1)-H	ESI-	863.565482	6.75
PI(18:0/18:2)-H	ESI-	861.549679	6.15
PI(18:0/18:3)-H	ESI-	859.532989	5.75
PI(18:0/20:3)-H	ESI-	887.565147	6.38
PI(18:0/20:4)-H	ESI-	885.549989	6.20
PI(18:0/20:5)-H	ESI-	883.534812	5.52
PI(18:0/22:4)-H	ESI-	913.580912	6.66

PI(18:0/22:5)-H	ESI-	911.566026	6.45
PI(18:0/22:6)-H	ESI-	909.553624	5.79
PI(18:0p/20:4)-H	ESI-	869.555226	5.81
PI(18:1/18:1)-H	ESI-	861.549506	6.00
PI(18:1/18:2)-H	ESI-	859.534331	5.35
PI(18:1/20:4)-H	ESI-	883.534555	5.36
PI(18:2/18:2)-H	ESI-	857.518641	4.69
PI(19:0/20:4)-H	ESI-	899.565582	6.54
PS(14:0/22:1)-H	ESI-	788.544812	5.98
PS(15:0/22:1)-H	ESI-	802.560442	6.39
PS(16:0/22:0)-H	ESI-	818.591729	7.54
PS(16:0/22:1)-H	ESI-	816.574202	6.80
PS(16:0/24:2)-H	ESI-	842.591682	7.12
PS(17:0/22:1)-H	ESI-	830.591751	7.37
PS(18:0/22:1)-H	ESI-	844.607552	7.86
PS(18:0/24:2)-H	ESI-	870.622603	8.04
PS(18:1/22:0)-H	ESI-	844.607325	7.68
PS(20:0/18:2)-H	ESI-	814.558823	6.07
PS(21:0/18:2)-H	ESI-	828.576111	7.00
PS(21:0/22:5)-H	ESI-	878.59297	7.44
PS(22:0/20:3)-H	ESI-	868.607348	7.33
PS(22:0/20:4)-H	ESI-	866.592001	6.96
PS(22:0/22:5)-H	ESI-	892.607805	6.95
PS(22:0/22:6)-H	ESI-	890.591957	6.76
PS(22:2/18:2)-H	ESI-	838.56056	6.29
PS(24:0/18:2)-H	ESI-	870.62192	7.87
PS(24:1/20:4)-H	ESI-	892.607631	7.13
PS(24:3/16:0)-H	ESI-	840.576195	6.75
PS(24:3/17:0)-H	ESI-	854.59157	7.30
PS(24:3/18:0)-H	ESI-	868.607783	7.68
PS(24:3/18:1)-H	ESI-	866.591585	6.94
PS(24:3/18:2)-H	ESI-	864.576121	6.23
PS(24:4/16:0)-H	ESI-	838.560579	6.20
PS(24:4/18:0)-H	ESI-	866.591822	7.12
PS(24:4/18:1)-H	ESI-	864.576093	6.38
SM(d16:0/15:1)+CH3COO	ESI-	719.534939	5.09
SM(d16:0/16:0)+CH3COO	ESI-	735.565306	5.89
SM(d16:0/16:1)+CH3COO	ESI-	733.550211	5.57
SM(d16:0/17:1)+CH3COO	ESI-	747.566079	6.05
SM(d16:0/18:1)+CH3COO	ESI-	761.581279	6.58
SM(d16:0/18:2)+CH3COO	ESI-	759.566118	5.90
SM(d16:0/19:1)+CH3COO	ESI-	775.597	6.98
SM(d16:0/20:1)+CH3COO	ESI-	789.612668	7.57
SM(d16:0/20:2)+CH3COO	ESI-	787.597231	6.70
SM(d16:0/20:3)+CH3COO	ESI-	785.581751	5.98
SM(d16:0/22:1)+CH3COO	ESI-	817.644211	8.46

SM(d16:0/22:2)+CH3COO	ESI-	815.629151	7.61
SM(d16:0/24:1)+CH3COO	ESI-	845.675839	9.13
SM(d16:0/24:2)+CH3COO	ESI-	843.660173	8.55
SM(d16:0/24:3)+CH3COO	ESI-	841.64463	7.68
SM(d16:0/24:4)+CH3COO	ESI-	839.628496	6.97
SM(d16:0/26:1)+CH3COO	ESI-	873.707726	10.00
SM(d16:0/26:2)+CH3COO	ESI-	871.690859	9.22
SM(d16:0/26:3)+CH3COO	ESI-	869.675728	8.57
SM(d16:0/26:4)+CH3COO	ESI-	867.658243	7.77
SM(d16:0/26:5)+CH3COO	ESI-	865.644088	7.35
SM(d16:0/28:3)+CH3COO	ESI-	897.707013	9.26
SM(d16:0/28:4)+CH3COO	ESI-	895.691365	8.65
SM(d16:1/16:1)+CH3COO	ESI-	731.534627	4.75
SM(d16:1/22:0)+CH3COO	ESI-	817.644247	8.49
SM(d18:0/16:0)+CH3COO	ESI-	763.596617	6.82
SM(d18:1/12:0)+CH3COO	ESI-	705.51877	4.63
SM(d18:1/16:0)+CH3COO	ESI-	761.581367	6.64
SM(d18:1/17:1)+CH3COO	ESI-	773.581074	6.22
SM(d18:1/18:0)+CH3COO	ESI-	789.612919	7.52
SM(d18:1/18:1)+CH3COO	ESI-	787.596975	6.81
SM(d18:1/19:0)+CH3COO	ESI-	803.628547	7.92
SM(d18:1/19:1)+CH3COO	ESI-	801.613181	7.22
SM(d18:1/22:0)+CH3COO	ESI-	845.675981	9.14
SM(d18:2/20:0)+CH3COO	ESI-	815.629548	7.61
SM(d18:2/24:2)+CH3COO	ESI-	867.659566	7.81
SM(d20:0/18:1)+CH3COO	ESI-	817.644783	8.55
SM(d20:0/19:1)+CH3COO	ESI-	831.66047	8.90
SM(d20:1/16:1)+CH3COO	ESI-	787.597252	6.72
SM(d20:1/19:1)+CH3COO	ESI-	829.6446	8.09
SM(d22:0/18:1)+CH3COO	ESI-	845.675688	9.08
SM(d22:0/18:2)+CH3COO	ESI-	843.661517	8.71
SM(d22:0/19:1)+CH3COO	ESI-	859.691132	9.53
SM(d22:1/16:0)+CH3COO	ESI-	817.644684	8.28
SM(d22:1/18:1)+CH3COO	ESI-	843.660314	8.54
SM(d22:1/19:1)+CH3COO	ESI-	857.675612	8.90
SM(d22:1/20:2)+CH3COO	ESI-	869.675409	8.43
SM(d22:2/19:1)+CH3COO	ESI-	855.659992	8.09

Obs m/z: detected mass, Rt: retention time (minutes).

BIBLIOGRAPHY

[1] S. G. Oliver, M. K. Winson, D. B. Kell, and F. Baganz, "Systematic functional analysis of the yeast genome," *Trends Biotechnol.*, vol. 16, no. 9, pp. 373–378, Sep. 1998.

[2] O. Fiehn, "Metabolomics--the link between genotypes and phenotypes," *Plant Mol. Biol.*, vol. 48, no. 1–2, pp. 155–171, Jan. 2002.

[3] J. C. D'Auria and J. Gershenzon, "The secondary metabolism of Arabidopsis thaliana: growing like a weed," *Curr. Opin. Plant Biol.*, vol. 8, no. 3, pp. 308–316, Jun. 2005, doi: 10.1016/j.pbi.2005.03.012.

[4] A. R. Fernie, "The future of metabolic phytochemistry: larger numbers of metabolites, higher resolution, greater understanding," *Phytochemistry*, vol. 68, no. 22–24, pp. 2861–2880, Dec. 2007, doi: 10.1016/j.phytochem.2007.07.010.

[5] E. Pichersky and D. R. Gang, "Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective," *Trends Plant Sci.*, vol. 5, no. 10, pp. 439–445, Oct. 2000.

[6] H. Tweeddale, L. Notley-McRobb, and T. Ferenci, "Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ('metabolome') analysis," *J. Bacteriol.*, vol. 180, no. 19, pp. 5109–5116, Oct. 1998.

[7] K. Dettmer, P. A. Aronov, and B. D. Hammock, "Mass spectrometry-based metabolomics," *Mass Spectrom Rev*, vol. 26, no. 1, pp. 51–78, Feb. 2007, doi: 10.1002/mas.20108.

[8] J. K. Nicholson, J. C. Lindon, and E. Holmes, "'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data," *Xenobiotica*, vol. 29, no. 11, pp. 1181–1189, Nov. 1999, doi: 10.1080/004982599238047.

[9] O. Fiehn, "Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks," *Comp. Funct. Genomics*, vol. 2, no. 3, pp. 155–168, 2001, doi: 10.1002/cfg.82.

[10] R. H. Barton, J. K. Nicholson, P. Elliott, and E. Holmes, "High-throughput 1H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study," *Int J Epidemiol*, vol. 37 Suppl 1, pp. i31-40, Apr. 2008, doi: 10.1093/ije/dym284.

[11] J. M. Fonville, A. D. Maher, M. Coen, E. Holmes, J. C. Lindon, and J. K. Nicholson, "Evaluation of full-resolution J-resolved 1H NMR projections of biofluids for metabonomics information retrieval and biomarker identification," *Anal. Chem.*, vol. 82, no. 5, pp. 1811–1821, Mar. 2010, doi: 10.1021/ac902443k.

[12] E. J. Want *et al.*, "Global metabolic profiling of animal and human tissues via UPLC-MS," *Nat Protoc*, vol. 8, no. 1, pp. 17–32, Jan. 2013, doi: 10.1038/nprot.2012.135.

[13] O. Beckonert *et al.*, "High-resolution magic-angle-spinning NMR spectroscopy for metabolic profiling of intact tissues," *Nature Protocols*, vol. 5, no. 6, pp. 1019–1032, Jun. 2010, doi: 10.1038/nprot.2010.45.

[14] E. Martineau, I. Tea, G. Loaëc, P. Giraudeau, and S. Akoka, "Strategy for choosing extraction procedures for NMR-based metabolomic analysis of mammalian cells," *Anal Bioanal Chem*, vol. 401, no. 7, pp. 2133–2142, Oct. 2011, doi: 10.1007/s00216-011-5310-y.

[15] C. A. Sellick, R. Hansen, G. M. Stephens, R. Goodacre, and A. J. Dickson, "Metabolite extraction from suspension-cultured mammalian cells for global

metabolite profiling," *Nat Protoc*, vol. 6, no. 8, pp. 1241–1249, Jul. 2011, doi: 10.1038/nprot.2011.366.
[16] M. Verma and H. N. Banerjee, "Metabolomic Approaches in Cancer Epidemiology," *Diseases*, vol. 3, no. 3, pp. 167–175, Aug. 2015, doi: 10.3390/diseases3030167.
[17] N. Koen, I. Du Preez, and D. T. Loots, "Metabolomics and Personalized Medicine," *Adv Protein Chem Struct Biol*, vol. 102, pp. 53–78, 2016, doi: 10.1016/bs.apcsb.2015.09.003.
[18] M. Jacob, A. L. Lopata, M. Dasouki, and A. M. Abdel Rahman, "Metabolomics toward personalized medicine," *Mass Spectrom Rev*, Oct. 2017, doi: 10.1002/mas.21548.
[19] H. Gu *et al.*, "1H NMR metabolomics study of age profiling in children," *NMR Biomed*, vol. 22, no. 8, pp. 826–833, Oct. 2009, doi: 10.1002/nbm.1395.
[20] R. Kaddurah-Daouk, B. S. Kristal, and R. M. Weinshilboum, "Metabolomics: a global biochemical approach to drug response and disease," *Annu. Rev. Pharmacol. Toxicol.*, vol. 48, pp. 653–683, 2008, doi: 10.1146/annurev.pharmtox.48.113006.094715.
[21] S. Cacciatore and M. Loda, "Innovation in metabolomics to improve personalized healthcare," *Ann. N. Y. Acad. Sci.*, vol. 1346, no. 1, pp. 57–62, Jun. 2015, doi: 10.1111/nyas.12775.
[22] C. Stella *et al.*, "Susceptibility of human metabolic phenotypes to dietary modulation," *J. Proteome Res.*, vol. 5, no. 10, pp. 2780–2788, Oct. 2006, doi: 10.1021/pr060265y.
[23] A. Scalbert *et al.*, "The food metabolome: a window over dietary exposure," *Am. J. Clin. Nutr.*, vol. 99, no. 6, pp. 1286–1308, Jun. 2014, doi: 10.3945/ajcn.113.076133.
[24] S. Nalbantoglu, "Metabolomics: Basic Principles and Strategies," *Molecular Medicine*, Aug. 2019, doi: 10.5772/intechopen.88563.
[25] "Metabolome Analysis: An Introduction | Wiley," *Wiley.com*. [Online]. Available: https://www.wiley.com/en-us/Metabolome+Analysis%3A+An+Introduction-p-9780471743446. [Accessed: 07-Jan-2020].
[26] S. K. Bhattacharya, Ed., *Metabolomics: Methods and Protocols*. Humana Press, 2019.
[27] G. Theodoridis, H. G. Gika, and I. Wilson, Eds., *Metabolic Profiling: Methods and Protocols*. Humana Press, 2018.
[28] A. Nordström, E. Want, T. Northen, J. Lehtiö, and G. Siuzdak, "Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics," *Anal. Chem.*, vol. 80, no. 2, pp. 421–429, Jan. 2008, doi: 10.1021/ac701982e.
[29] W. R. Wikoff *et al.*, "Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, no. 10, pp. 3698–3703, Mar. 2009, doi: 10.1073/pnas.0812874106.
[30] N. Vinayavekhin and A. Saghatelian, "Untargeted metabolomics," *Curr Protoc Mol Biol*, vol. Chapter 30, p. Unit 30.1.1-24, Apr. 2010, doi: 10.1002/0471142727.mb3001s90.
[31] C. H. Johnson, J. Ivanisevic, and G. Siuzdak, "Metabolomics: beyond biomarkers and towards mechanisms," *Nat. Rev. Mol. Cell Biol.*, vol. 17, no. 7, pp. 451–459, 2016, doi: 10.1038/nrm.2016.25.
[32] L. D. Roberts, A. L. Souza, R. E. Gerszten, and C. B. Clish, "Targeted metabolomics," *Curr Protoc Mol Biol*, vol. Chapter 30, p. Unit 30.2.1-24, Apr. 2012, doi: 10.1002/0471142727.mb3002s98.
[33] E. Dudley, M. Yousef, Y. Wang, and W. J. Griffiths, "Targeted metabolomics and mass spectrometry," *Adv Protein Chem Struct Biol*, vol. 80, pp. 45–83, 2010, doi: 10.1016/B978-0-12-381264-3.00002-3.

[34] A. Ribbenstedt, H. Ziarrusta, and J. P. Benskin, “Development, characterization and comparisons of targeted and non-targeted metabolomics methods,” *PLoS One*, vol. 13, no. 11, Nov. 2018, doi: 10.1371/journal.pone.0207082.
[35] J. Zhou, H. Liu, Y. Liu, J. Liu, X. Zhao, and Y. Yin, “Development and Evaluation of a Parallel Reaction Monitoring Strategy for Large-Scale Targeted Metabolomics Quantification,” *Anal. Chem.*, vol. 88, no. 8, pp. 4478–4486, Apr. 2016, doi: 10.1021/acs.analchem.6b00355.
[36] Y. Li *et al.*, “A novel approach to transforming a non-targeted metabolic profiling method to a pseudo-targeted method using the retention time locking gas chromatography/mass spectrometry-selected ions monitoring,” *J Chromatogr A*, vol. 1255, pp. 228–236, Sep. 2012, doi: 10.1016/j.chroma.2012.01.076.
[37] S. Chen *et al.*, “Pseudotargeted metabolomics method and its application in serum biomarker discovery for hepatocellular carcinoma based on ultra high-performance liquid chromatography/triple quadrupole mass spectrometry,” *Anal. Chem.*, vol. 85, no. 17, pp. 8326–8333, Sep. 2013, doi: 10.1021/ac4016787.
[38] M. M. Gessel, J. L. Norris, and R. M. Caprioli, “MALDI imaging mass spectrometry: spatial molecular analysis to enable a new age of discovery,” *J Proteomics*, vol. 107, pp. 71–82, Jul. 2014, doi: 10.1016/j.jprot.2014.03.021.
[39] Y. Fujimura and D. Miura, “MALDI Mass Spectrometry Imaging for Visualizing In Situ Metabolism of Endogenous Metabolites and Dietary Phytochemicals,” *Metabolites*, vol. 4, no. 2, pp. 319–346, May 2014, doi: 10.3390/metabo4020319.
[40] G. J. Patti, O. Yanes, and G. Siuzdak, “Innovation: Metabolomics: the apogee of the omics trilogy,” *Nat. Rev. Mol. Cell Biol.*, vol. 13, no. 4, pp. 263–269, Mar. 2012, doi: 10.1038/nrm3314.
[41] K. Tanaka *et al.*, “Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry,” *Rapid Communications in Mass Spectrometry*, vol. 2, no. 8, pp. 151–153, 1988, doi: 10.1002/rcm.1290020802.
[42] T. R. Northen *et al.*, “Clathrate nanostructures for mass spectrometry,” *Nature*, vol. 449, no. 7165, pp. 1033–1036, Oct. 2007, doi: 10.1038/nature06195.
[43] J. M. Wiseman, D. R. Ifa, Q. Song, and R. G. Cooks, “Tissue imaging at atmospheric pressure using desorption electrospray ionization (DESI) mass spectrometry,” *Angew. Chem. Int. Ed. Engl.*, vol. 45, no. 43, pp. 7188–7192, Nov. 2006, doi: 10.1002/anie.200602449.
[44] M. L. Kraft, P. K. Weber, M. L. Longo, I. D. Hutcheon, and S. G. Boxer, “Phase separation of lipid membranes analyzed with high-resolution secondary ion mass spectrometry,” *Science*, vol. 313, no. 5795, pp. 1948–1951, Sep. 2006, doi: 10.1126/science.1130279.
[45] A. Buck *et al.*, “High-resolution MALDI-FT-ICR MS imaging for the analysis of metabolites from formalin-fixed, paraffin-embedded clinical tissue samples,” *J. Pathol.*, vol. 237, no. 1, pp. 123–132, Sep. 2015, doi: 10.1002/path.4560.
[46] E. R. Amstalden van Hove, D. F. Smith, and R. M. A. Heeren, “A concise review of mass spectrometry imaging,” *J Chromatogr A*, vol. 1217, no. 25, pp. 3946–3954, Jun. 2010, doi: 10.1016/j.chroma.2010.01.033.
[47] R. A. Dixon *et al.*, “Applications of metabolomics in agriculture,” *J. Agric. Food Chem.*, vol. 54, no. 24, pp. 8984–8994, Nov. 2006, doi: 10.1021/jf061218t.
[48] H. Tian, S. M. Lam, and G. Shui, “Metabolomics, a Powerful Tool for Agricultural Research,” *Int J Mol Sci*, vol. 17, no. 11, Nov. 2016, doi: 10.3390/ijms17111871.
[49] C. B. Clish, “Metabolomics: an emerging but powerful tool for precision medicine,” *Cold Spring Harb Mol Case Stud*, vol. 1, no. 1, Oct. 2015, doi: 10.1101/mcs.a000588.

[50] Y. Sandlers, “The future perspective: metabolomics in laboratory medicine for inborn errors of metabolism,” *Transl Res*, vol. 189, pp. 65–75, 2017, doi: 10.1016/j.trsl.2017.06.005.
[51] J. C. Lindon, E. Holmes, and J. K. Nicholson, “Metabonomics in pharmaceutical R&D,” *FEBS J.*, vol. 274, no. 5, pp. 1140–1151, Mar. 2007, doi: 10.1111/j.1742-4658.2007.05673.x.
[52] D. S. Wishart, “Applications of metabolomics in drug discovery and development,” *Drugs R D*, vol. 9, no. 5, pp. 307–322, 2008.
[53] D. S. Wishart, “Emerging applications of metabolomics in drug discovery and precision medicine,” *Nat Rev Drug Discov*, vol. 15, no. 7, pp. 473–484, 2016, doi: 10.1038/nrd.2016.32.
[54] M. Mamas, W. B. Dunn, L. Neyses, and R. Goodacre, “The role of metabolites and metabolomics in clinically applicable biomarkers of disease,” *Arch. Toxicol.*, vol. 85, no. 1, pp. 5–17, Jan. 2011, doi: 10.1007/s00204-010-0609-6.
[55] E. P. Rhee and R. E. Gerszten, “Metabolomics and cardiovascular biomarker discovery,” *Clin. Chem.*, vol. 58, no. 1, pp. 139–147, Jan. 2012, doi: 10.1373/clinchem.2011.169573.
[56] Biomarkers Definitions Working Group, “Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework,” *Clinical Pharmacology & Therapeutics*, vol. 69, no. 3, pp. 89–95, Mar. 2001, doi: 10.1067/mcp.2001.113989.
[57] N. C. for B. Information, U. S. N. L. of M. 8600 R. Pike, B. MD, and 20894 Usa, *How does the liver work?* Institute for Quality and Efficiency in Health Care (IQWiG), 2016.
[58] Z. M. Younossi, A. B. Koenig, D. Abdelatif, Y. Fazel, L. Henry, and M. Wymer, “Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes,” *Hepatology*, vol. 64, no. 1, pp. 73–84, 2016, doi: 10.1002/hep.28431.
[59] W. R. Kim *et al.*, “OPTN/SRTR 2017 Annual Data Report: Liver,” *Am. J. Transplant.*, vol. 19 Suppl 2, pp. 184–283, Feb. 2019, doi: 10.1111/ajt.15276.
[60] L. Castera, M. Friedrich-Rust, and R. Loomba, “Noninvasive Assessment of Liver Disease in Patients With Nonalcoholic Fatty Liver Disease,” *Gastroenterology*, Jan. 2019, doi: 10.1053/j.gastro.2018.12.036.
[61] E. B. Tapper and R. Loomba, “Noninvasive imaging biomarker assessment of liver fibrosis by elastography in NAFLD,” *Nat Rev Gastroenterol Hepatol*, vol. 15, no. 5, pp. 274–282, 2018, doi: 10.1038/nrgastro.2018.10.
[62] H. K. Drescher, S. Weiskirchen, and R. Weiskirchen, “Current Status in Testing for Nonalcoholic Fatty Liver Disease (NAFLD) and Nonalcoholic Steatohepatitis (NASH),” *Cells*, vol. 8, no. 8, Aug. 2019, doi: 10.3390/cells8080845.
[63] E. Cleveland, A. Bandy, and L. B. VanWagner, “Diagnostic challenges of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis,” *Clin Liver Dis (Hoboken)*, vol. 11, no. 4, pp. 98–104, Apr. 2018, doi: 10.1002/cld.716.
[64] J. M. Pappachan, S. Babu, B. Krishnan, and N. C. Ravindran, “Non-alcoholic Fatty Liver Disease: A Clinical Update,” *J Clin Transl Hepatol*, vol. 5, no. 4, pp. 384–393, Dec. 2017, doi: 10.14218/JCTH.2017.00013.
[65] A. Safaei *et al.*, “Metabolomic analysis of human cirrhosis, hepatocellular carcinoma, non-alcoholic fatty liver disease and non-alcoholic steatohepatitis diseases,” *Gastroenterol Hepatol Bed Bench*, vol. 9, no. 3, pp. 158–173, 2016.
[66] J. M. Mato, C. Alonso, M. Noureddin, and S. C. Lu, “Biomarkers and subtypes of deranged lipid metabolism in non-alcoholic fatty liver disease,” *World J.*

Gastroenterol., vol. 25, no. 24, pp. 3009–3020, Jun. 2019, doi: 10.3748/wjg.v25.i24.3009.
[67] C. Alonso *et al.*, "Metabolomic Identification of Subtypes of Nonalcoholic Steatohepatitis," *Gastroenterology*, vol. 152, no. 6, pp. 1449-1461.e7, 2017, doi: 10.1053/j.gastro.2017.01.015.
[68] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J Clin*, vol. 68, no. 6, pp. 394–424, 2018, doi: 10.3322/caac.21492.
[69] A. Nicholson *et al.*, "The clinical effectiveness and cost-effectiveness of the PROGENSA® prostate cancer antigen 3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and economic evaluation," *Health Technol Assess*, vol. 19, no. 87, pp. i–xxxi, 1–191, Oct. 2015, doi: 10.3310/hta19870.
[70] G. Zadra and M. Loda, "Metabolic Vulnerabilities of Prostate Cancer: Diagnostic and Therapeutic Opportunities," *Cold Spring Harb Perspect Med*, Dec. 2017, doi: 10.1101/cshperspect.a030569.
[71] K. L. Penney *et al.*, "mRNA expression signature of Gleason grade predicts lethal prostate cancer," *J. Clin. Oncol.*, vol. 29, no. 17, pp. 2391–2396, Jun. 2011, doi: 10.1200/JCO.2010.32.6421.
[72] C. O. Madu and Y. Lu, "Novel diagnostic biomarkers for prostate cancer," *J Cancer*, vol. 1, pp. 150–177, Oct. 2010.
[73] R. J. Hendriks, I. M. van Oort, and J. A. Schalken, "Blood-based and urinary prostate cancer biomarkers: a review and comparison of novel biomarkers for detection and treatment decisions," *Prostate Cancer Prostatic Dis.*, vol. 20, no. 1, pp. 12–19, Mar. 2017, doi: 10.1038/pcan.2016.59.
[74] P. Sharma, K. Zargar-Shoshtari, and J. M. Pow-Sang, "Biomarkers for prostate cancer: present challenges and future opportunities," *Future Sci OA*, vol. 2, no. 1, p. FSO72, Mar. 2016, doi: 10.4155/fso.15.72.
[75] G. Lucarelli *et al.*, "Metabolomic profiling for the identification of novel diagnostic markers in prostate cancer," *Expert Rev. Mol. Diagn.*, vol. 15, no. 9, pp. 1211–1224, 2015, doi: 10.1586/14737159.2015.1069711.
[76] N. Gómez-Cebrián, A. Rojas-Benedicto, A. Albors-Vaquer, J. A. López-Guerrero, A. Pineda-Lucena, and L. Puchades-Carrasco, "Metabolomics Contributions to the Discovery of Prostate Cancer Biomarkers," *Metabolites*, vol. 9, no. 3, Mar. 2019, doi: 10.3390/metabo9030048.
[77] J. E. McDunn, S. M. Stirdivant, L. A. Ford, and R. L. Wolfert, "Metabolomics and its Application to the Development of Clinical Laboratory Tests for Prostate Cancer," *EJIFCC*, vol. 26, no. 2, pp. 92–104, Mar. 2015.
[78] A. Sreekumar *et al.*, "Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression," *Nature*, vol. 457, no. 7231, pp. 910–914, Feb. 2009, doi: 10.1038/nature07762.
[79] R. S. Kelly, M. G. Vander Heiden, E. Giovannucci, and L. A. Mucci, "Metabolomic Biomarkers of Prostate Cancer: Prediction, Diagnosis, Progression, Prognosis, and Recurrence," *Cancer Epidemiol. Biomarkers Prev.*, vol. 25, no. 6, pp. 887–906, 2016, doi: 10.1158/1055-9965.EPI-15-1223.
[80] B. J. Trock, "Application of metabolomics to prostate cancer," *Urol. Oncol.*, vol. 29, no. 5, pp. 572–581, Oct. 2011, doi: 10.1016/j.urolonc.2011.08.002.
[81] A. Sreekumar *et al.*, "Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression," *Nature*, vol. 457, no. 7231, pp. 910–914, Feb. 2009, doi: 10.1038/nature07762.

[82] A. R. Lima, M. de L. Bastos, M. Carvalho, and P. Guedes de Pinho, “Biomarker Discovery in Human Prostate Cancer: an Update in Metabolomics Studies,” *Transl Oncol*, vol. 9, no. 4, pp. 357–370, Aug. 2016, doi: 10.1016/j.tranon.2016.05.004.
[83] A. P. Khan *et al.*, “The role of sarcosine metabolism in prostate cancer progression,” *Neoplasia*, vol. 15, no. 5, pp. 491–501, May 2013, doi: 10.1593/neo.13314.
[84] E. A. Struys, A. C. Heijboer, J. van Moorselaar, C. Jakobs, and M. A. Blankenstein, “Serum sarcosine is not a marker for prostate cancer,” *Ann. Clin. Biochem.*, vol. 47, no. Pt 3, p. 282, May 2010, doi: 10.1258/acb.2010.009270.
[85] C. Pérez-Rambla, L. Puchades-Carrasco, M. García-Flores, J. Rubio-Briones, J. A. López-Guerrero, and A. Pineda-Lucena, “Non-invasive urinary metabolomic profiling discriminates prostate cancer from benign prostatic hyperplasia,” *Metabolomics*, vol. 13, no. 5, p. 52, 2017, doi: 10.1007/s11306-017-1194-y.
[86] S. O. Diaz *et al.*, “Second trimester maternal urine for the diagnosis of trisomy 21 and prediction of poor pregnancy outcomes,” *J. Proteome Res.*, vol. 12, no. 6, pp. 2946–2957, Jun. 2013, doi: 10.1021/pr4002355.
[87] G. Quintás, N. Portillo, J. C. García-Cañaveras, J. V. Castell, A. Ferrer, and A. Lahoz, “Chemometric approaches to improve PLSDA model outcome for predicting human non-alcoholic fatty liver disease using UPLC-MS as a metabolic profiling tool,” *Metabolomics*, vol. 8, no. 1, pp. 86–98, Feb. 2012, doi: 10.1007/s11306-011-0292-5.
[88] E. J. Want *et al.*, “Global metabolic profiling procedures for urine using UPLC-MS,” *Nat Protoc*, vol. 5, no. 6, pp. 1005–1018, Jun. 2010, doi: 10.1038/nprot.2010.50.
[89] O. Beckonert *et al.*, “Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts,” *Nat Protoc*, vol. 2, no. 11, pp. 2692–2703, 2007, doi: 10.1038/nprot.2007.376.
[90] A. Alonso, S. Marsal, and A. Julià, “Analytical methods in untargeted metabolomics: state of the art in 2015,” *Front Bioeng Biotechnol*, vol. 3, p. 23, 2015, doi: 10.3389/fbioe.2015.00023.
[91] W. B. Dunn, I. D. Wilson, A. W. Nicholls, and D. Broadhurst, “The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans,” *Bioanalysis*, vol. 4, no. 18, pp. 2249–2264, Sep. 2012, doi: 10.4155/bio.12.204.
[92] M. H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. Wiley, 2008.
[93] D. Antcliffe and A. C. Gordon, “Metabonomics and intensive care,” *Crit Care*, vol. 20, p. 68, Mar. 2016, doi: 10.1186/s13054-016-1222-8.
[94] G. A. N. Gowda and D. Raftery, Eds., *NMR-Based Metabolomics: Methods and Protocols*. Humana Press, 2019.
[95] “Understanding NMR Spectroscopy, 2nd Edition | NMR Spectroscopy / MRI / Imaging | Analytical Chemistry | Chemistry | Subjects | Wiley,” *Wiley.com*. [Online]. Available: https://www.wiley.com/en-us/Understanding+NMR+Spectroscopy%2C+2nd+Edition-p-9780470746080. [Accessed: 08-Oct-2019].
[96] “NMR Data Processing | NMR Spectroscopy / MRI / Imaging | Analytical Chemistry | Chemistry | Subjects | Wiley,” *Wiley.com*. [Online]. Available: https://www.wiley.com/en-us/NMR+Data+Processing-p-9780471039006. [Accessed: 08-Oct-2019].
[97] J. H. Gross, *Mass Spectrometry: A Textbook*, 2nd ed. Berlin Heidelberg: Springer-Verlag, 2011.
[98] H. Kienitz and F. Aulinger, *Massenspektrometrie*. Weinheim/Bergstrasse: Verlag Chemie, 1968.

[99] P. Frycák, R. Husková, T. Adam, and K. Lemr, “Atmospheric pressure ionization mass spectrometry of purine and pyrimidine markers of inherited metabolic disorders,” *J Mass Spectrom*, vol. 37, no. 12, pp. 1242–1248, Dec. 2002, doi: 10.1002/jms.389.
[100] M. S. B. Munson and F. H. Field, “Chemical Ionization Mass Spectrometry. I. General Introduction,” *J. Am. Chem. Soc.*, vol. 88, no. 12, pp. 2621–2630, Jun. 1966, doi: 10.1021/ja00964a001.
[101] A. E. Ashcroft, *Ionization Methods in Organic Mass Spectrometry*. 1997.
[102] M. Karas and F. Hillenkamp, “Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons,” *Anal. Chem.*, vol. 60, no. 20, pp. 2299–2301, Oct. 1988, doi: 10.1021/ac00171a028.
[103] H. Awad, M. M. Khamis, and A. El-Aneed, “Mass Spectrometry, Review of the Basics: Ionization,” *Applied Spectroscopy Reviews*, vol. 50, no. 2, pp. 158–175, Feb. 2015, doi: 10.1080/05704928.2014.954046.
[104] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, “Electrospray ionization for mass spectrometry of large biomolecules,” *Science*, vol. 246, no. 4926, pp. 64–71, Oct. 1989, doi: 10.1126/science.2675315.
[105] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, “Electrospray ionization–principles and practice,” *Mass Spectrometry Reviews*, vol. 9, no. 1, pp. 37–70, 1990, doi: 10.1002/mas.1280090103.
[106] P. H. Dawson, “Quadrupole mass analyzers: Performance, design and some recent applications,” *Mass Spectrometry Reviews*, vol. 5, no. 1, pp. 1–37, 1986, doi: 10.1002/mas.1280050102.
[107] M. Balcerzak, “An overview of analytical applications of time of flight-mass spectrometric (TOF-MS) analyzers and an inductively coupled plasma-TOF-MS technique,” *Anal Sci*, vol. 19, no. 7, pp. 979–989, Jul. 2003.
[108] R. E. March, “Ion Trap Mass Spectrometers,” in *Encyclopedia of Spectroscopy and Spectrometry*, J. C. Lindon, Ed. Oxford: Elsevier, 1999, pp. 1000–1009.
[109] A. M. Haag, “Mass Analyzers and Mass Spectrometers,” *Adv. Exp. Med. Biol.*, vol. 919, pp. 157–169, 2016, doi: 10.1007/978-3-319-41448-5_7.
[110] A. El-Aneed, A. Cohen, and J. Banoub, “Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers,” *Applied Spectroscopy Reviews*, vol. 44, no. 3, pp. 210–230, Apr. 2009, doi: 10.1080/05704920902717872.
[111] R. A. Yost and C. G. Enke, “Selected ion fragmentation with a tandem quadrupole mass spectrometer,” *J. Am. Chem. Soc.*, vol. 100, no. 7, pp. 2274–2275, Mar. 1978, doi: 10.1021/ja00475a072.
[112] Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman, and R. G. Cooks, “The Orbitrap: a new mass spectrometer,” *Journal of Mass Spectrometry*, vol. 40, no. 4, pp. 430–443, 2005, doi: 10.1002/jms.856.
[113] null Badman, null Patterson, null Wells, null Santini, and null Cooks, “Differential non-destructive image current detection in a fourier transform quadrupole ion trap,” *J Mass Spectrom*, vol. 34, no. 8, pp. 889–894, Aug. 1999, doi: 10.1002/(SICI)1096-9888(199908)34:8<889::AID-JMS848>3.0.CO;2-C.
[114] S. Medhe, “Mass Spectrometry: Detectors Review,” *Chemical and Biomolecular Engineering*, vol. 3, no. 4, p. 51, Oct. 2018, doi: 10.11648/j.cbe.20180304.11.
[115] V. Craige Trenerry and S. J. Rochfort, “9.16 - Natural Products Research and Metabolomics,” in *Comprehensive Natural Products II*, H.-W. (Ben) Liu and L. Mander, Eds. Oxford: Elsevier, 2010, pp. 595–628.
[116] D. S. Wishart, “Advances in metabolite identification,” *Bioanalysis*, vol. 3, no. 15, pp. 1769–1782, Aug. 2011, doi: 10.4155/bio.11.155.

[117] D. S. Wishart, “Quantitative metabolomics using NMR,” *TrAC Trends in Analytical Chemistry*, vol. 27, no. 3, pp. 228–237, Mar. 2008, doi: 10.1016/j.trac.2007.12.001.
[118] A. Zhang, H. Sun, P. Wang, Y. Han, and X. Wang, “Modern analytical techniques in metabolomics analysis,” *Analyst*, vol. 137, no. 2, pp. 293–300, Jan. 2012, doi: 10.1039/c1an15605e.
[119] W. B. Dunn, N. J. C. Bailey, and H. E. Johnson, “Measuring the metabolome: current analytical technologies,” *Analyst*, vol. 130, no. 5, pp. 606–625, May 2005, doi: 10.1039/b418288j.
[120] E. M. Weaver and A. B. Hummon, “Imaging mass spectrometry: from tissue sections to cell cultures,” *Adv. Drug Deliv. Rev.*, vol. 65, no. 8, pp. 1039–1055, Jul. 2013, doi: 10.1016/j.addr.2013.03.006.
[121] I. Tkác, G. Oz, G. Adriany, K. Uğurbil, and R. Gruetter, “In vivo 1H NMR spectroscopy of the human brain at high magnetic fields: metabolite quantification at 4T vs. 7T,” *Magn Reson Med*, vol. 62, no. 4, pp. 868–879, Oct. 2009, doi: 10.1002/mrm.22086.
[122] J. Balog *et al.*, “Intraoperative tissue identification using rapid evaporative ionization mass spectrometry,” *Sci Transl Med*, vol. 5, no. 194, p. 194ra93, Jul. 2013, doi: 10.1126/scitranslmed.3005623.
[123] P. A. Vorkas *et al.*, “Untargeted UPLC-MS Profiling Pipeline to Expand Tissue Metabolome Coverage: Application to Cardiovascular Disease,” *Anal. Chem.*, vol. 87, no. 8, pp. 4184–4193, Apr. 2015, doi: 10.1021/ac503775m.
[124] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, “NMRPipe: a multidimensional spectral processing system based on UNIX pipes,” *J. Biomol. NMR*, vol. 6, no. 3, pp. 277–293, Nov. 1995.
[125] G. S. Rule and T. K. Hitchens, *Fundamentals of Protein NMR Spectroscopy*. Springer Netherlands, 2006.
[126] A. Shimizu, M. Ikeguchi, and S. Sugai, “Appropriateness of DSS and TSP as internal references for (1)H NMR studies of molten globule proteins in aqueous media,” *J. Biomol. NMR*, vol. 4, no. 6, pp. 859–862, Nov. 1994, doi: 10.1007/BF00398414.
[127] “Modern NMR Techniques for Synthetic Chemistry,” *CRC Press*. [Online]. Available: https://www.crcpress.com/Modern-NMR-Techniques-for-Synthetic-Chemistry/Fisher/p/book/9781466592247. [Accessed: 04-Jul-2019].
[128] F. Savorani, G. Tomasi, and S. B. Engelsen, “icoshift: A versatile tool for the rapid alignment of 1D NMR spectra,” *J. Magn. Reson.*, vol. 202, no. 2, pp. 190–202, Feb. 2010, doi: 10.1016/j.jmr.2009.11.012.
[129] F. van den Berg, G. Tomasi, and N. Viereck, “Warping: Investigation of NMR Pre-processing and Correction,” in *Magnetic Resonance in Food Science*, 2005, pp. 131–138.
[130] *Magnetic Resonance in Food Science*. 2005.
[131] F. Gan, G. Ruan, and J. Mo, “Baseline correction by improved iterative polynomial fitting with automatic threshold,” *Chemometrics and Intelligent Laboratory Systems*, vol. 82, no. 1, pp. 59–65, May 2006, doi: 10.1016/j.chemolab.2005.08.009.
[132] T. De Meyer *et al.*, “NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm,” *Anal. Chem.*, vol. 80, no. 10, pp. 3783–3790, May 2008, doi: 10.1021/ac7025964.
[133] C. D. Broeckling, I. R. Reddy, A. L. Duran, X. Zhao, and L. W. Sumner, “MET-IDEA: data extraction tool for mass spectrometry-based metabolomics,” *Anal. Chem.*, vol. 78, no. 13, pp. 4334–4341, Jul. 2006, doi: 10.1021/ac0521596.

[134] A. L. Duran, J. Yang, L. Wang, and L. W. Sumner, “Metabolomics spectral formatting, alignment and conversion tools (MSFACTs),” *Bioinformatics*, vol. 19, no. 17, pp. 2283–2293, Nov. 2003, doi: 10.1093/bioinformatics/btg315.
[135] M. Katajamaa and M. Oresic, “Processing methods for differential analysis of LC/MS profile data,” *BMC Bioinformatics*, vol. 6, p. 179, Jul. 2005, doi: 10.1186/1471-2105-6-179.
[136] C. A. Smith, E. J. Want, G. O’Maille, R. Abagyan, and G. Siuzdak, “XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification,” *Anal. Chem.*, vol. 78, no. 3, pp. 779–787, Feb. 2006, doi: 10.1021/ac051437y.
[137] Y. Tikunov *et al.*, “A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles,” *Plant Physiol.*, vol. 139, no. 3, pp. 1125–1137, Nov. 2005, doi: 10.1104/pp.105.068130.
[138] M. Katajamaa and M. Oresic, “Data processing for mass spectrometry-based metabolomics,” *J Chromatogr A*, vol. 1158, no. 1–2, pp. 318–328, Jul. 2007, doi: 10.1016/j.chroma.2007.04.021.
[139] F. Giacomoni *et al.*, “Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics,” *Bioinformatics*, vol. 31, no. 9, pp. 1493–1495, May 2015, doi: 10.1093/bioinformatics/btu813.
[140] O. Hrydziuszko and M. R. Viant, “Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline,” *Metabolomics*, vol. 8, no. 1, pp. 161–174, Jun. 2012, doi: 10.1007/s11306-011-0366-4.
[141] R. Di Guida *et al.*, “Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling,” *Metabolomics*, vol. 12, p. 93, 2016, doi: 10.1007/s11306-016-1030-9.
[142] J. Xia and D. S. Wishart, “Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst,” *Nat Protoc*, vol. 6, no. 6, pp. 743–760, Jun. 2011, doi: 10.1038/nprot.2011.319.
[143] A. Craig, O. Cloarec, E. Holmes, J. K. Nicholson, and J. C. Lindon, “Scaling and normalization effects in NMR spectroscopic metabonomic data sets,” *Anal. Chem.*, vol. 78, no. 7, pp. 2262–2267, Apr. 2006, doi: 10.1021/ac0519312.
[144] F. Dieterle, A. Ross, G. Schlotterbeck, and H. Senn, “Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics,” *Anal. Chem.*, vol. 78, no. 13, pp. 4281–4290, Jul. 2006, doi: 10.1021/ac051632c.
[145] B. M. Bolstad, R. A. Irizarry, M. Astrand, and T. P. Speed, “A comparison of normalization methods for high density oligonucleotide array data based on variance and bias,” *Bioinformatics*, vol. 19, no. 2, pp. 185–193, Jan. 2003, doi: 10.1093/bioinformatics/19.2.185.
[146] D. Amaratunga and J. Cabrera, “Analysis of Data From Viral DNA Microchips,” *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1161–1170, Dec. 2001, doi: 10.1198/016214501753381814.
[147] S. C. Hicks and R. A. Irizarry, “When to use Quantile Normalization?,” *bioRxiv*, p. 012203, Dec. 2014, doi: 10.1101/012203.
[148] O. M. Kvalheim, Frode. Brakstad, and Yizeng. Liang, “Preprocessing of analytical profiles in the presence of homoscedastic or heteroscedastic noise,” *Anal. Chem.*, vol. 66, no. 1, pp. 43–51, Jan. 1994, doi: 10.1021/ac00073a010.
[149] R. A. van den Berg, H. C. J. Hoefsloot, J. A. Westerhuis, A. K. Smilde, and M. J. van der Werf, “Centering, scaling, and transformations: improving the biological

information content of metabolomics data," *BMC Genomics*, vol. 7, p. 142, Jun. 2006, doi: 10.1186/1471-2164-7-142.
[150] B. P. Durbin, J. S. Hardin, D. M. Hawkins, and D. M. Rocke, "A variance-stabilizing transformation for gene-expression microarray data," *Bioinformatics*, vol. 18 Suppl 1, pp. S105-110, 2002, doi: 10.1093/bioinformatics/18.suppl_1.s105.
[151] L. Eriksson, T. Byrne, E. Johansson, J. Trygg, and C. Vikström, *Multi- and Megavariate Data Analysis Basic Principles and Applications*. Umetrics Academy, 2013.
[152] L. Eriksson, T. Byrne, E. Johansson, J. Trygg, and C. Vikström, *Multi- and Megavariate Data Analysis Basic Principles and Applications*. Umetrics Academy, 2013.
[153] "A User's Guide to Principal Components | Wiley," *Wiley.com*. [Online]. Available: https://www.wiley.com/en-us/A+User%27s+Guide+to+Principal+Components-p-9780471471349. [Accessed: 30-Nov-2019].
[154] S. Wold, M. Sjöström, and L. Eriksson, "Partial Least Squares Projections to Latent Structures (PLS) in Chemistry," in *Encyclopedia of Computational Chemistry*, American Cancer Society, 2002.
[155] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, Oct. 2001, doi: 10.1016/S0169-7439(01)00155-1.
[156] H. S. Tapp and E. K. Kemsley, "Notes on the practical utility of OPLS," *TrAC Trends in Analytical Chemistry*, vol. 28, no. 11, pp. 1322–1327, Dec. 2009, doi: 10.1016/j.trac.2009.08.006.
[157] O. Svensson, T. Kourti, and J. F. MacGregor, "An investigation of orthogonal signal correction algorithms and their characteristics," *Journal of Chemometrics*, vol. 16, no. 4, pp. 176–188, 2002, doi: 10.1002/cem.700.
[158] T. Verron, R. Sabatier, and R. Joffre, "Some theoretical properties of the O-PLS method," *Journal of Chemometrics*, vol. 18, no. 2, pp. 62–68, 2004, doi: 10.1002/cem.847.
[159] R. Ergon, "PLS post-processing by similarity transformation (PLS + ST): a simple alternative to OPLS," *Journal of Chemometrics*, vol. 19, no. 1, pp. 1–4, 2005, doi: 10.1002/cem.899.
[160] B. Worley and R. Powers, "Multivariate Analysis in Metabolomics," *Curr Metabolomics*, vol. 1, no. 1, pp. 92–107, 2013, doi: 10.2174/2213235X11301010092.
[161] S. Wiklund *et al.*, "Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models," *Anal. Chem.*, vol. 80, no. 1, pp. 115–122, Jan. 2008, doi: 10.1021/ac0713510.
[162] J. Trygg and S. Wold, "Orthogonal projections to latent structures (O-PLS)," *Journal of Chemometrics*, vol. 16, no. 3, pp. 119–128, 2002, doi: 10.1002/cem.695.
[163] E. Szymańska, E. Saccenti, A. K. Smilde, and J. A. Westerhuis, "Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies," *Metabolomics*, vol. 8, no. Suppl 1, pp. 3–16, Jun. 2012, doi: 10.1007/s11306-011-0330-3.
[164] M. Greiner, D. Pfeiffer, and R. D. Smith, "Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests," *Prev. Vet. Med.*, vol. 45, no. 1–2, pp. 23–41, May 2000.
[165] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, vol. 44, no. 3, pp. 837–845, Sep. 1988.

[166] M. H. Zweig and G. Campbell, “Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine,” *Clin. Chem.*, vol. 39, no. 4, pp. 561–577, Apr. 1993.
[167] F. J. W. M. Dankers, A. Traverso, L. Wee, and S. M. J. van Kuijk, “Prediction Modeling Methodology,” in *Fundamentals of Clinical Data Science*, P. Kubben, M. Dumontier, and A. Dekker, Eds. Cham: Springer International Publishing, 2019, pp. 101–120.
[168] M. W. Kirschner, “The meaning of systems biology,” *Cell*, vol. 121, no. 4, pp. 503–504, May 2005, doi: 10.1016/j.cell.2005.05.005.
[169] K. Suhre *et al.*, “Human metabolic individuality in biomedical and pharmaceutical research,” *Nature*, vol. 477, no. 7362, pp. 54–60, Sep. 2011, doi: 10.1038/nature10354.
[170] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe, “KEGG for integration and interpretation of large-scale molecular data sets,” *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D109-114, Jan. 2012, doi: 10.1093/nar/gkr988.
[171] T. Jewison *et al.*, “SMPDB 2.0: big improvements to the Small Molecule Pathway Database,” *Nucleic Acids Res.*, vol. 42, no. Database issue, pp. D478-484, Jan. 2014, doi: 10.1093/nar/gkt1067.
[172] H. Ma *et al.*, “The Edinburgh human metabolic network reconstruction and its functional analysis,” *Mol Syst Biol*, vol. 3, p. 135, Sep. 2007, doi: 10.1038/msb4100177.
[173] T. Kelder *et al.*, “WikiPathways: building research communities on biological pathways,” *Nucleic Acids Res*, vol. 40, no. Database issue, pp. D1301–D1307, Jan. 2012, doi: 10.1093/nar/gkr1074.
[174] R. Caspi *et al.*, “The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases,” *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D742-753, Jan. 2012, doi: 10.1093/nar/gkr1014.
[175] Q. Wu *et al.*, “Altered Lipid Metabolism in Recovered SARS Patients Twelve Years after Infection,” *Scientific Reports*, vol. 7, no. 1, pp. 1–12, Aug. 2017, doi: 10.1038/s41598-017-09536-z.
[176] W. B. Dunn *et al.*, “Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry,” *Nat. Protocols*, vol. 6, no. 7, pp. 1060–1083, Jun. 2011, doi: 10.1038/nprot.2011.335.
[177] H. Gika and G. Theodoridis, “Sample preparation prior to the LC-MS-based metabolomics/metabonomics of blood-derived samples,” *Bioanalysis*, vol. 3, no. 14, pp. 1647–1661, Jul. 2011, doi: 10.4155/bio.11.122.
[178] E. Hashimoto, K. Tokushige, and J. Ludwig, “Diagnosis and classification of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis: Current concepts and remaining challenges,” *Hepatol. Res.*, vol. 45, no. 1, pp. 20–28, Jan. 2015, doi: 10.1111/hepr.12333.
[179] D. E. Kleiner *et al.*, “Design and validation of a histological scoring system for nonalcoholic fatty liver disease,” *Hepatology*, vol. 41, no. 6, pp. 1313–1321, Jun. 2005, doi: 10.1002/hep.20701.
[180] E. K. Spengler and R. Loomba, “Recommendations for Diagnosis, Referral for Liver Biopsy, and Treatment of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis,” *Mayo Clin. Proc.*, vol. 90, no. 9, pp. 1233–1246, Sep. 2015, doi: 10.1016/j.mayocp.2015.06.013.
[181] C. Estes, H. Razavi, R. Loomba, Z. Younossi, and A. J. Sanyal, “Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease,” *Hepatology*, vol. 67, no. 1, pp. 123–133, 2018, doi: 10.1002/hep.29466.

[182] S. A. Polyzos and C. S. Mantzoros, “An update on the validity of irisin assays and the link between irisin and hepatic metabolism,” *Metab. Clin. Exp.*, vol. 64, no. 9, pp. 937–942, Sep. 2015, doi: 10.1016/j.metabol.2015.06.005.
[183] S. A. Polyzos, J. Kountouras, and C. Zavos, “Nonalcoholic fatty liver disease: the pathogenetic roles of insulin resistance and adipocytokines,” *Curr. Mol. Med.*, vol. 9, no. 3, pp. 299–314, Apr. 2009, doi: 10.2174/156652409787847191.
[184] S. A. Polyzos, J. Kountouras, G. Deretzi, C. Zavos, and C. S. Mantzoros, “The emerging role of endocrine disruptors in pathogenesis of insulin resistance: a concept implicating nonalcoholic fatty liver disease,” *Curr. Mol. Med.*, vol. 12, no. 1, pp. 68–82, Jan. 2012, doi: 10.2174/156652412798376161.
[185] M. Doulberis, G. Kotronis, D. Gialamprinou, J. Kountouras, and P. Katsinelos, “Non-alcoholic fatty liver disease: An update with special focus on the role of gut microbiota,” *Metab. Clin. Exp.*, vol. 71, pp. 182–197, 2017, doi: 10.1016/j.metabol.2017.03.013.
[186] C. Leung, L. Rivera, J. B. Furness, and P. W. Angus, “The role of the gut microbiota in NAFLD,” *Nat Rev Gastroenterol Hepatol*, vol. 13, no. 7, pp. 412–425, 2016, doi: 10.1038/nrgastro.2016.85.
[187] L. Hoyles *et al.*, “Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women,” *Nat. Med.*, vol. 24, no. 7, pp. 1070–1080, Jul. 2018, doi: 10.1038/s41591-018-0061-3.
[188] N. M. Delzenne and L. B. Bindels, “Microbiome metabolomics reveals new drivers of human liver steatosis,” *Nat. Med.*, vol. 24, no. 7, pp. 906–907, Jul. 2018, doi: 10.1038/s41591-018-0126-3.
[189] P. Bedossa *et al.*, “Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients,” *Hepatology*, vol. 56, no. 5, pp. 1751–1759, Nov. 2012, doi: 10.1002/hep.25889.
[190] P. Giraudeau, V. Silvestre, and S. Akoka, “Optimizing water suppression for quantitative NMR-based metabolomics: a tutorial review,” *Metabolomics*, vol. 11, no. 5, pp. 1041–1055, Oct. 2015, doi: 10.1007/s11306-015-0794-7.
[191] H. Y. Carr and E. M. Purcell, “Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments,” *Phys. Rev.*, vol. 94, no. 3, pp. 630–638, May 1954, doi: 10.1103/PhysRev.94.630.
[192] E. A. Howe, R. Sinha, D. Schlauch, and J. Quackenbush, “RNA-Seq analysis in MeV,” *Bioinformatics*, vol. 27, no. 22, pp. 3209–3210, Nov. 2011, doi: 10.1093/bioinformatics/btr490.
[193] M. N. Triba *et al.*, “PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters,” *Mol Biosyst*, vol. 11, no. 1, pp. 13–19, Jan. 2015, doi: 10.1039/c4mb00414k.
[194] E. Fahy *et al.*, “A comprehensive classification system for lipids,” *J. Lipid Res.*, vol. 46, no. 5, pp. 839–861, May 2005, doi: 10.1194/jlr.E400004-JLR200.
[195] E. Fahy *et al.*, “Update of the LIPID MAPS comprehensive classification system for lipids,” *J. Lipid Res.*, vol. 50 Suppl, pp. S9-14, Apr. 2009, doi: 10.1194/jlr.R800095-JLR200.
[196] T. Cajka and O. Fiehn, “LC-MS-Based Lipidomics and Automated Identification of Lipids Using the LipidBlast In-Silico MS/MS Library,” *Methods Mol. Biol.*, vol. 1609, pp. 149–170, 2017, doi: 10.1007/978-1-4939-6996-8_14.
[197] J. P. Koelmel *et al.*, “LipidMatch: an automated workflow for rule-based lipid identification using untargeted high-resolution tandem mass spectrometry data,” *BMC Bioinformatics*, vol. 18, no. 1, p. 331, Jul. 2017, doi: 10.1186/s12859-017-1744-3.

[198] J. E. Kyle *et al.*, "LIQUID: an-open source software for identifying lipids in LC-MS/MS-based lipidomics data," *Bioinformatics*, vol. 33, no. 11, pp. 1744–1746, Jun. 2017, doi: 10.1093/bioinformatics/btx046.
[199] D. R. Matthews, J. P. Hosker, A. S. Rudenski, B. A. Naylor, D. F. Treacher, and R. C. Turner, "Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man," *Diabetologia*, vol. 28, no. 7, pp. 412–419, Jul. 1985, doi: 10.1007/bf00280883.
[200] O. J. Dunn, "Multiple Comparisons Using Rank Sums," *Technometrics*, vol. 6, no. 3, pp. 241–252, 1964, doi: 10.2307/1266041.
[201] A. Cano and C. Alonso, "Deciphering non-alcoholic fatty liver disease through metabolomics," *Biochem. Soc. Trans.*, vol. 42, no. 5, pp. 1447–1452, Oct. 2014, doi: 10.1042/BST20140138.
[202] A. J. Hackstadt and A. M. Hess, "Filtering for increased power for microarray data analysis," *BMC Bioinformatics*, vol. 10, p. 11, Jan. 2009, doi: 10.1186/1471-2105-10-11.
[203] K.-A. Lê Cao, S. Boitard, and P. Besse, "Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems," *BMC Bioinformatics*, vol. 12, p. 253, Jun. 2011, doi: 10.1186/1471-2105-12-253.
[204] H. Chun and S. Keleş, "Sparse partial least squares regression for simultaneous dimension reduction and variable selection," *J R Stat Soc Series B Stat Methodol*, vol. 72, no. 1, pp. 3–25, Jan. 2010, doi: 10.1111/j.1467-9868.2009.00723.x.
[205] S. C. Kalhan *et al.*, "Plasma metabolomic profile in nonalcoholic fatty liver disease," *Metab. Clin. Exp.*, vol. 60, no. 3, pp. 404–413, Mar. 2011, doi: 10.1016/j.metabol.2010.03.006.
[206] A. M. Mondul, S. C. Moore, S. J. Weinstein, E. D. Karoly, J. N. Sampson, and D. Albanes, "Metabolomic analysis of prostate cancer risk in a prospective cohort: The alpha-tocolpherol, beta-carotene cancer prevention (ATBC) study," *Int. J. Cancer*, vol. 137, no. 9, pp. 2124–2132, Nov. 2015, doi: 10.1002/ijc.29576.
[207] S. Koutros *et al.*, "Prospective evaluation of serum sarcosine and risk of prostate cancer in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial," *Carcinogenesis*, vol. 34, no. 10, pp. 2281–2285, Oct. 2013, doi: 10.1093/carcin/bgt176.
[208] D. P. Ankerst, M. Liss, D. Zapata, J. Hoefler, I. M. Thompson, and R. J. Leach, "A case control study of sarcosine as an early prostate cancer detection biomarker," *BMC Urol*, vol. 15, p. 99, Oct. 2015, doi: 10.1186/s12894-015-0095-5.
[209] S. Hercberg *et al.*, "The SU.VI.MAX Study: a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals," *Arch. Intern. Med.*, vol. 164, no. 21, pp. 2335–2342, Nov. 2004, doi: 10.1001/archinte.164.21.2335.
[210] S. Hercberg *et al.*, "A primary prevention trial using nutritional doses of antioxidant vitamins and minerals in cardiovascular diseases and cancers in a general population: the SU.VI.MAX study--design, methods, and participant characteristics. SUpplementation en VItamines et Minéraux AntioXydants," *Control Clin Trials*, vol. 19, no. 4, pp. 336–351, Aug. 1998, doi: 10.1016/s0197-2456(98)00015-4.
[211] W. H. Organization, "International classification of diseases and related health problems, 10^<th> revision," *http://www.who.int/classifications/apps/icd/icd10online*, 2007.
[212] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[213] W. J. Youden, “Index for rating diagnostic tests,” *Cancer*, vol. 3, no. 1, pp. 32–35, Jan. 1950, doi: 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3.
[214] C. Chatfield, “Exploratory data analysis,” *European Journal of Operational Research*, vol. 23, no. 1, pp. 5–13, Jan. 1986, doi: 10.1016/0377-2217(86)90209-2.
[215] S. Sperandei, “Understanding logistic regression analysis,” *Biochem Med (Zagreb)*, vol. 24, no. 1, pp. 12–18, 2014, doi: 10.11613/BM.2014.003.
[216] “Prostate-Specific Antigen Testing: Overview, Physiologic Characteristics of PSA, Other Prostate Cancer Markers,” Dec. 2019.
[217] R. C. Bone *et al.*, “Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine,” *Chest*, vol. 101, no. 6, pp. 1644–1655, Jun. 1992, doi: 10.1378/chest.101.6.1644.
[218] K. R. Ludwig and A. B. Hummon, “Mass spectrometry for the discovery of biomarkers of sepsis,” *Mol Biosyst*, vol. 13, no. 4, pp. 648–664, Mar. 2017, doi: 10.1039/c6mb00656f.
[219] K.-M. Kaukonen, M. Bailey, D. Pilcher, D. J. Cooper, and R. Bellomo, “Systemic inflammatory response syndrome criteria in defining severe sepsis,” *N. Engl. J. Med.*, vol. 372, no. 17, pp. 1629–1638, Apr. 2015, doi: 10.1056/NEJMoa1415236.
[220] D. C. Angus and R. S. Wax, “Epidemiology of sepsis: an update,” *Crit. Care Med.*, vol. 29, no. 7 Suppl, pp. S109-116, Jul. 2001, doi: 10.1097/00003246-200107001-00035.
[221] G. S. Martin, “Sepsis, severe sepsis and septic shock: changes in incidence, pathogens and outcomes,” *Expert Rev Anti Infect Ther*, vol. 10, no. 6, pp. 701–706, Jun. 2012, doi: 10.1586/eri.12.50.
[222] A. Lever and I. Mackenzie, “Sepsis: definition, epidemiology, and diagnosis,” *BMJ*, vol. 335, no. 7625, pp. 879–883, Oct. 2007, doi: 10.1136/bmj.39346.495880.AE.
[223] R. J. Langley and H. R. Wong, “Early Diagnosis of Sepsis: Is an Integrated Omics Approach the Way Forward?,” *Mol Diagn Ther*, vol. 21, no. 5, pp. 525–537, 2017, doi: 10.1007/s40291-017-0282-z.
[224] R. J. Langley *et al.*, “Integrative ‘omic’ analysis of experimental bacteremia identifies a metabolic signature that distinguishes human sepsis from systemic inflammatory response syndromes,” *Am. J. Respir. Crit. Care Med.*, vol. 190, no. 4, pp. 445–455, Aug. 2014, doi: 10.1164/rccm.201404-0624OC.
[225] M. Ferrario *et al.*, “Mortality prediction in patients with severe septic shock: a pilot study using a target metabolomics approach,” *Sci Rep*, vol. 6, p. 20391, Feb. 2016, doi: 10.1038/srep20391.
[226] M. Eckerle *et al.*, “Metabolomics as a Driver in Advancing Precision Medicine in Sepsis,” *Pharmacotherapy*, vol. 37, no. 9, pp. 1023–1032, Sep. 2017, doi: 10.1002/phar.1974.

ABSTRACT

Metabolomics is the science designed to comprehensively study the metabolome, the repertoire of small molecule metabolites, which gives a comprehensive snapshot of the physiological state of the biofluid, extracts or cells studied. Measuring metabolites by using metabolomics is a key complementary to genome, transcriptome and proteome studies, which may improve our understanding of how genetics, environment, the microbiome, disease, drug exposure, diet, and lifestyle influence the phenotype. One of important application of metabolomics in clinical research is the discovery of novel biomarkers. The present PhD thesis focus on biomarkers discovery by applying metabolomics, the objectives were: (1) by using NMR and UPLC-HRMS based metabolomic and lipidomic profiling, to identify novel plasma biomarkers, if any, which characterize the different stage of Non-alcoholic fatty liver disease (NAFLD), and (2) by combining UPLC-HRMS based untargeted metabolomics with epidemiology approach, to identify plasma biomarkers which associated with the risk of developing prostate cancer (PCa) within the following decade.

Keywords: Metabolomics, NMR, LC-MS, multivariate analysis, prostate cancer, NAFLD, biomarkers

RESUME

La métabolomique consiste en l'étude approfondie du métabolome, qui correspond à l'ensemble des métabolites présent dans un organisme. Le métabolome donne un aperçu de l'état physiologique de l'organisme, de l'extrait ou des cellules étudiées. La mesure des métabolites par l'approche métabolomique est un complément important aux études sur le génome, le transcriptome et le protéome, qui peut améliorer notre compréhension sur comment la génétique, l'environnement, le microbiome, les maladies, l'exposition aux médicaments, l'alimentation et le mode de vie influencent le phénotype. L'une des applications importantes de la métabolomique en recherche clinique est la découverte de nouveaux biomarqueurs. La présente thèse porte sur la partie découverte de biomarqueurs par métabolomique. Deux études sont réalisées :
(1) la première utilise l'approche métabolomique et lipidomique basée sur la RMN et l'UPLC-HRMS, pour identifier de nouveaux biomarqueurs plasmatiques qui caractérisent les différents stades de la stéatose hépatique non alcoolique (NAFLD)
(2) la seconde combine la métabolomique non ciblée basée sur UPLC-HRMS avec une approche épidémiologique pour identifier les biomarqueurs plasmatiques associés au risque de développer un cancer de la prostate (PCa) au cours de la décennie suivante.

Mots-clés : Métabolomique, RMN, LC-MS, analyses multivariées, cancer prostate, NAFLD, biomarqueurs