UNIVERSIDAD DE LA RIOJA

TESIS DOCTORAL

Título
Multiplatform metabolome profiling to identify specific signatures and biomarkers in blood samples: untargeted approach

Autor/es
Kateryna Tkachenko

Director/es
Consuelo Pizarro Millán y José María González Sáiz

Facultad
Facultad de Ciencia y Tecnología

Titulación

Departamento
Química

Curso Académico

UNIVERSIDAD DE LA RIOJA

TESIS DOCTORAL 2023

Programa de Doctorado en Química

MULTIPLATFORM METABOLOME PROFILING TO IDENTIFY SPECIFIC SIGNATURES AND BIOMARKERS IN BLOOD SAMPLES: UNTARGETED APPROACH

Kateryna Tkachenko

Directora: Consuelo Pizarro Millán
Director: José María González Saiz

Dra. CONSUELO PIZARRO MILLÁN, Catedrática de Química Analítica del Departamento de Química de la Universidad de la Rioja y

Dr. JOSÉ MARÍA GONZÁLEZ SÁIZ, Catedrático de Ingeniería Química del Departamento de Química de la Universidad de la Rioja

HACEN CONSTAR:

Que la memoria titulada "Multiplatform metabolome profiling to identify specific signatures and biomarkers in blood samples: untargeted approach" realizada por Kateryna Tkachenko en el Departamento de Química de la Universidad de La Rioja y bajo su inmediata dirección, reúne las condiciones exigidas para optar al grado de Doctor en Química.

Logroño, a 24 de febrero de 2023

Fdo: Consuelo Pizarro Millán   Fdo: José Mará González Sáiz

Abstract

One of the significant challenges in identifying effective therapy in many chronic and neurodegenerative diseases is the need for reliable biomarkers. Thus, new point-of-care diagnostics tools are essential for unambiguously distinguishing diseased patients from healthy ones providing results in rapid time.

In this doctoral thesis, an untargeted metabolomics approach based on high-throughput analytical techniques such as vibrational spectroscopy and liquid chromatography-mass spectrometry (LC-MS) was evaluated in different studies related to the field of health and disease. Thus, this doctoral thesis's main objective is to provide an objective diagnosis of disorders such as Parkinson's, Alzheimer's, Amyotrophic lateral sclerosis and Metabolic Syndrome. Different studies were performed to obtain a metabolic profile of healthy and diseased patients. Thus, to obtain specific metabolomic fingerprinting multiple analytical strategies coupled with multivariate strategies were tested and combined in order to exploit their respective strengths and drawbacks. Therefore, distinct mid-infrared metabolic fingerprints in the diseases mentioned above were investigated for patient stratification and to guide an accurate and early differential diagnosis. In addition, UPLC-MS analysis successfully complemented vibrational spectroscopy, providing excellent patient discrimination based on specific blood biomarkers.

The obtained results are very promising, giving place to the new hypothesis about disease pathogenesis and possible involved metabolic pathways that should be validated by a further targeted and multidisciplinary approach.

Resumen

Uno de los desafíos significativos en la identificación de terapias efectivas en muchas enfermedades crónicas y neurodegenerativas es la necesidad de biomarcadores confiables. Por lo tanto, nuevas herramientas de diagnóstico en los puntos de atención son esenciales para distinguir sin ambigüedades a los pacientes enfermos de los sanos, proporcionando resultados rápidos.

En esta tesis doctoral, se evaluó un enfoque de metabolómica no dirigida basado en técnicas analíticas de alto rendimiento, como la espectroscopia vibracional y la cromatografía líquida-espectrometría de masas (LC-MS), en diferentes estudios relacionados con el campo de la salud y la enfermedad. Por lo tanto, el objetivo principal de esta tesis doctoral es desarrollar estrategias para el diagnóstico objetivo de trastornos como la enfermedad de Parkinson, la enfermedad de Alzheimer, la Esclerosis lateral amiotrófica y el Síndrome metabólico. Se llevaron a cabo diferentes estudios para obtener un perfil metabólico de pacientes sanos y enfermos. De este modo, para obtener un perfil metabólico específico, se probaron y combinaron múltiples estrategias analíticas acopladas con estrategias multivariadas con el fin de explotar sus respectivas fortalezas y debilidades. Por lo tanto, se investigó la existencia de huellas metabólicas distintas en el infrarrojo medio en las enfermedades mencionadas anteriormente, para la estratificación de pacientes y guiar un diagnóstico diferencial preciso y temprano. Además, el análisis UPLC-MS complementó con éxito la espectroscopía vibracional, proporcionando una excelente discriminación de pacientes basada en biomarcadores específicos de la sangre.

Los resultados obtenidos son muy prometedores, dando lugar a nuevas hipótesis sobre la patogénesis de las enfermedades y las posibles vías metabólicas involucradas, que deberían ser validadas mediante un enfoque multidisciplinario y dirigido.

CONTENTS

Abbreviations

AD	Alzheimer's disease
ALS	Amyotrophic lateral sclerosis
Ala	Alanine
ATR	Attenuated total reflection
AUC	Area under the curve
CE	Capillary electrophoresis
Cer	Ceramide
CNS	Central nervous system
Cr	Creatinine
CSF	Cerebrospinal fluid
CV	Cross-validation
EMSC	Extended multiple scatter correction
FA	Fatty acids
FDR	False discovery rate
FIR	Far infrared region
FTIR	Fourier transform infrared spectroscopy
FWHM	Full width half maximum
GC	Gas spectrometry
GL	Glycerolipids
Glu	Glutamate
GP	Glycerophospholipids
HC	Healthy controls
HCA	Hierarchical clustering analysis
HMDB	Human Metabolome Data Base
IR	Infrared
kNN	K-nearest neighbours
KS	Kennard-Stone
LC	Liquid chromatography
LDA	Linear discriminant analysis
LLE	Liquid liquid extraction

LOO	Leave one out cross validation
LOOCV	Leave one out cross validation
LysPC	Lysophosphatidilcoline
LysPE	Lysophosphatidiletanolamine
MA	Mooving average
MetS	Metabolic syndrome
MIR	Mid-infrared region
MS	Mass spectrometry
MSC	Multiplicative scatter correction
MUFA	Monounsaturated fatty acids
NIR	near infrared region
NMR	nuclear magnetic resonance
OPLS-DA	orthogonal partial least square discriminant analysis
PC	phosphatidylcholine
PCA	Principle component analysis
PD	Parkinson's disease
PDD	Parkinson-related dementia
PDI	Parkinson's initial stage
PE	Phosphatidiletanolamine
PK	Polyketides
PLS-DA	Partial least square-discriminant analysis
POS	Reactive oxygen species
PR	Prenol lipids
PS	Phosphoserine
PUFA	Polyunsaturated fatty acids
QC	Quality control
Q-TOF	Quadrupole time of flight
RBC	Red blood cells
Rc	Category c rate
RF	Random forest
ROC	Receiver operating curves
ROS	Reactive oxygen species

RS	Raman spectroscopy
SELECT	Stepwise orthogonalization of predictors
SG	Savitsky-Golay
SIMCA	Soft independent modelling by class analogy
SL	Saccharolipids
SM	Sfingomyelin
SNV	Standard normal variate
SP	Sterol lipids
SPE	Solid phase extraction
TOF	Time of flight
TR	Total rate
U(H)PLC	Ultra-high performance liquid chromatography
UA	Uric acid
UFA	Unsaturated fatty acids
VIP	Variable in projection

Chapter 1

Introduction

1 Chapter 1. Introduction

1.1 Presentation of the thesis

Nowadays the magnitude of chronic diseases in the last years created an urgent need for point-of-care diagnostics. Chronic diseases are characterised by long-lasting and, in most cases, rapidly progressive disease which strikes mostly ageing patients. Since the ageing population is increasing, these diseases have increased in prevalence worldwide. Among the chronic diseases, metabolic disorders and neurodegenerative disorders should be outlined. Precisely, this PhD dissertation focuses on studying Metabolic syndrome (MetS) and the most common motor and cognitive neuro disorders such as Parkinson's (PD), Alzheimer's (AD) and Amyotrophic lateral sclerosis (ALS).

The commensurate prevalence of MetS represents a significant socio-economic problem worldwide since it is often associated with an increased risk of diabetes and cardiovascular disease. Patients affected by MetS are diagnosed through clinical indicators which reveal global physiological disturbances and not specific metabolic changes. Often these clinical indicators require more standardisation and a more accurate collection of all parameters to perform a more conclusive diagnosis. In addition, people affected by metabolic syndrome are often misdiagnosed due to the heterogeneity of clinical factors.

Whereas the conventional diagnosis of PD, AD and ALS remains essentially clinical based on the subjective observations of clinicians and confirmatory electromyography tests. In addition, the symptoms may appear only late in the disease course; therefore, there could be a significant gap between the first clinical visit and the acclaimed diagnosis, compromising a patient's survival rates. Consequently, it is necessary to develop an objective diagnostic tool, which would be characterised by high sensitivity, accuracy, and objectivity, allowing the prevention at an early stage.

All these disorders differ in their pathogenesis. Nevertheless, they share a commune point; no specific biomarkers exist for their diagnosis.

Many advances have been reached in genetic studies. Thus, many diseases are diagnosed based on invasive tests, and genetic screening is carried out. It should be said that the results of these tests are not immediate, delaying the diagnosis from days to sometimes weeks due to the complicated processing of these tests and the fact that many national health systems are saturated. Indeed, what does it mean for a patient with a very short life prognosis to have to wait months for a result? In such cases, it would be vital to have a tool that could provide a real-time answer enabling one to classify a patient into a specific category and, two, predict his status: positive or negative, diseased or healthy, at an early or advanced stage. This way, a patient could start treatment and take measures according to his health status.

Considering that the disease initiation or progression encompasses a set of metabolic changes and disarrangements, it would be possible to approach the disease's diagnosis from a holistic, functional perspective. In recent years, investigating the metabolomic changes in patients' profiles has been very important and is already highly exploited in many studies.

The metabolomic studies are conducted in two principal ways: targeted and untargeted. Thus, an untargeted strategy is undertaken without a preliminary hypothesis about metabolites that should be determined. Therefore, herein, only untargeted studies were performed aiming to extract a diversity of metabolites that do not belong to the same biological class or biochemical pathway. This approach creates big data sets; therefore, multivariate statistical analyses are applied to obtain valuable, meaningful information.

In this context, vibrational spectroscopy relying on infrared absorption, such as Fourier transform infrared (FTIR) spectroscopy, has become a promising methodological approach in biomedical applications thanks to advances in instrumentation and chemometrics. Thus, this vibrational technique provides a rapid ''spectroscopic fingerprinting'' of the sample to effectively explore disease processes and find early signs of disease onset. In addition, it requires a small sample volume; it is rapid, non-invasive, low-cost and easy to use—all these characteristics promise an accessible diagnostic platform.

Moreover, mass-mass spectrometry (MS/MS) remains the gold standard in metabolomic studies since it is a suitable technique to analyse and provide the exact mass value of almost any biological molecule susceptible to ionisation. For this reason, besides the FTIR-based approach coupled with chemometric techniques, blood samples from the same cohort of patients were submitted to hyphenated liquid chromatography associated with the MS technique to deepen the information obtained by FTIR studies. The research of specific biomarkers is still an ambitious challenge in metabolomic studies; therefore, the results obtained from LC-MS analysis should confirm the significance of the extracted spectra variables by FTIR approach and give insights about possible biomarkers involved in metabolic pathways, discriminating patients with different metabolic profiles and disease stages.

1.2 OBJECTIVES

Given this perspective, the global objective of the present doctoral thesis is focused on evaluating how the metabolomic profile can be altered in patients with the above-cited diseases and how these metabolic changes, recorded as an infrared (IR) signature, can be associated with pathological conditions, thus facilitating early diagnosis, treatment and follow-up of the disease. In addition, the results obtained by MS analysis are hypothesis-generating, giving place to further targeted analysis for absolute quantification of found biomarkers.

The global objective can be broken down into the following specific objectives.

1) Evaluating the potential of FTIR combined with a classification approach to detect spectra markers indicative of Metabolic syndrome. Afterwards, the classification's performance on reduced spectra fingerprints and measured clinical parameters will be compared.

2) To propose an efficient sequential classification strategy using MIR spectra variables to discriminate between three main categories of patients: PD patients from subjects with AD and healthy controls (HC). Moreover, the discrimination ability of this strategy in discriminating patients based on the progression of the disease will be tested. In addition, tentative biochemical reasoning of the selected spectral variables will be performed.

3) Another objective regarding Parkinson's disease is based on lipid extraction of PD plasma samples performing UPLC-MS/MS untargeted analysis of plasma samples to identify specific lipid biomarkers responsible for separating diseased groups such as PD and AD from healthy controls.

4) Metabolic profiling using untargeted FTIR and ATR-FTIR spectroscopy will be performed for the discrimination of patients with ALS and other motor-related neuro disorders, comparing them to healthy controls.

5) Final specific objective is focused in obtaining a specific lipid biomarker responsible of the onset of ALS disease and/or ALS progression by performing untargeted LC-MS/MS analysis.

1.3 STRUCTURE OF THE THESIS

The doctoral thesis is presented as a traditional doctoral thesis in accordance with the regulations of the University of La Rioja, passed by the Governing Council on 10th May 2022.

This doctoral thesis is divided into seven main chapters: the introduction and general objectives, theoretical foundations, methodology, and three chapters with the results obtained during these years of research. Thus, the general conclusion represents the final chapter.

Each chapter is divided into sections and subsections, and each chapter has its own bibliography, numbered independently. For better visualisation of each chapter, they are also divided by colours.

1) The introduction gathers the fundamental aspects of this thesis to provide an overview of how it was carried out and the main reasons that conducted us to study determined diseases and use specific analytical instrumentation. This chapter also includes the general and specific objectives that underlie the development of this doctoral thesis.

2) This chapter encompasses all the theoretical parts that serve to understand and justify the development of the current dissertation thesis; herein, the main sections are dedicated to the studies of metabolomics and lipidomics, justifying

the latest findings made in these fields. In addition, explanations for using blood as the main matrix for metabolomic studies will be found. Furthermore, the sections regarding the instrumental techniques mainly used in this thesis, their main advantages and drawbacks, are described. Finally, this chapter's last section and subsections are devoted to the chemometrics part, describing the most performed univariate and multivariate statistical methodologies in the studies presented in this thesis.

3) This chapter encompasses the methodological part, enclosing the details on the methods and instruments of the analytical workflow that underlie the results of this doctoral thesis. Thus, details on sample collection and experimental parameters are provided.

The following chapters represent the so-called ''core'' of this thesis, enclosing the results obtained over the years (some of which have been published in international scientific journals, while others are under review). Therefore, the results of this thesis follow according to the specific disease studied. Each chapter is dedicated to a specific disease, encompassing the disease's background and the principal aim, methods, results, and discussions.

4) This chapter is dedicated to Metabolic syndrome, describing a dual classification approach to obtain rapid discrimination of patients with Metabolic Syndrome using FTIR spectroscopy.

5) Herein, studies dedicated to Parkinson's disease are collected. Thus, the principal results obtained using vibrational spectroscopy and liquid chromatography-mass spectrometry have been encompassed.

6) This chapter is addressed to studies on Amyotrophic lateral sclerosis. Thus, the results obtained with FTIR, ATR-FTIR and UPLC-Q-TOF -MS are discussed.

7) The main conclusions are reassumed in this last chapter, discussing if the pre-established aims were achieved.
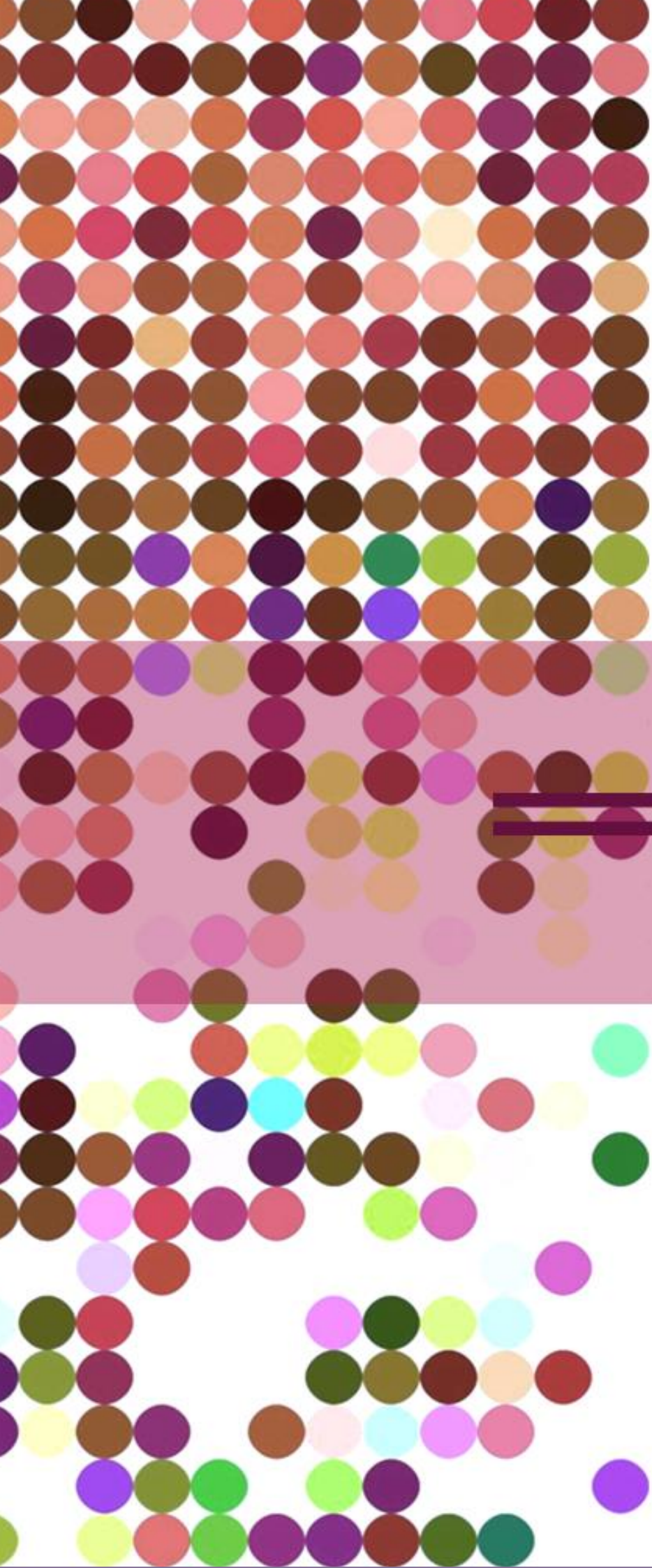
Chapter 2

Justification and theoretical bases

2 CHAPTER 2. JUSTIFICATION AND THEORETICAL BASES

2.1 METABOLOMICS: STATE OF THE ART

For decades, all scientific studies have focused on obtaining a deeper understanding of the pathological mechanisms of complex diseases. Techniques are developing fast, with the continual introduction of new instruments provided within powerful software. The main driving force in the path of these studies is the concept of the so-called "omics" sciences. This dynamic concept includes sciences such as genomics, transcriptomics, proteomics, metabolomics and metabonomic [1]. Analysis at the ''omics '' level is becoming increasingly popular, quickly acquiring more power and significance, resulting in widespread use. Metabo-'omics' is one of the most recent ''omics'' sciences and, as the name suggests, provides a comprehensive characterisation of the metabolite component in the biological system. Many authors use these terms interchangeably to define the same scientific strategies and processes. Metabolomics is more concerned with the phenotype of organisms than genomics, transcriptomics, and proteomics; thus, minimal gene and protein expression changes will be reflected in metabolites in a specific disease state. Several changes that cannot be reflected in these two fields can be detected in metabolomics [2].

Today, a fragile line separates these concepts of metabolomics and metabonomic: metabolomics aims to determine all metabolites in the cell, tissue or biofluid to make the so-called ''metabolic profile''; while metabonomic is a multiparametric response of the living system to genetic or pathophysiological changes [3]. Regardless of the term applied, this type of ''omics'' casts light on the organism's biochemistry, mainly responsible for characterising the composition of a wide range of metabolites, from organic to inorganic and elemental species. The human metabolome system currently comprises thousands of endogenous metabolites ranging from high to low molecular weight, and from hydrophilic to hydrophobic, in various biological matrices, such as cells, tissue, organ or biofluids. The identified compounds in the human body may be consulted in free and publicly accessible databases [4].

Considering that metabolome composition reflects the current status of the organism and metabolites are the end products of cellular regulatory processes required for homeostasis, growth, activity, maintenance and everyday function, the characterisation of metabolic phenotypes represents an attractive diagnostic tool in clinical practice. For this purpose, various disease-related challenges could be addressed in metabolome studies to i) provide insights into the flow of biological information; ii) unravel disease mechanisms; iii) predict biomarkers capable of stratifying patients; iv) disease subtyping and classification based on metabolomic profile.

The metabolomic approach provides insights into biological levels of features determined by systematic or instant metabolic changes. The internal changes in the organism, when passed from ''balanced'' status to ''non-balanced'', are detectable and generally correlated with disease initiation or progression. For this reason, the primary purpose of current studies in the field of metabolomics is to identify new metabolic biomarkers and metabolic pathways capable of deflating the disease process, differentiate patients' rational therapies and highlight future directions in clinical diagnosis. Moreover, classifying samples into subtypes could improve our understanding of disease and lead to adopting more suitable interventions and treatments for patients with different subtypes. Therefore, metabolomics is frequently applied to many diseases, from neonatal disorders to diabetes, cancer, and neurodegenerative disorders [6]– [13]. In addition, the application of metabolomics in the human organism is comprehensive, ranging from the context of cytology and histopathology [13–15] to the discovery of biomarkers in biofluids [2,16,17]. Often metabolomics is combined with other ''omics'' disciplines, such as proteomics or transcriptomics, to integrate information on the disorder studied and provide a deeper insight into the organism [18].

2.1.1 Lipidomics

In this era of omics sciences, lipidomics is a relatively young subdiscipline of metabolomics but an equally important field of biomedical research. Lipids are one of the prominent exponents of biological molecules in the human body, and the number of distinct chemical lipid entities is estimated between 10.000 and 100.000. Advancements in analytical technology and the development of new integrated tools

for understanding the biological roles of lipids dive into the emerging field of lipidomics. More studies focus on defining human plasma lipidome to quantify lipids and establish novel analytical methodologies [18–24]. Lipid analysis is now a regular part of every patient's blood test. Thus, the lipid profile of human plasma reveals the vast structural diversity of lipids that fall into main categories: fatty acyls (FA), glycerolipids (GL), glycerophospholipids (GP), sphingolipids (SP), sterol lipids (ST) and prenol lipids (PR), saccharolipids (SL) and polyketides (PK) [25,26]. These lipid categories are divided into main classes and subclasses based on the molecular structure and sub-structure Figure 2-1.
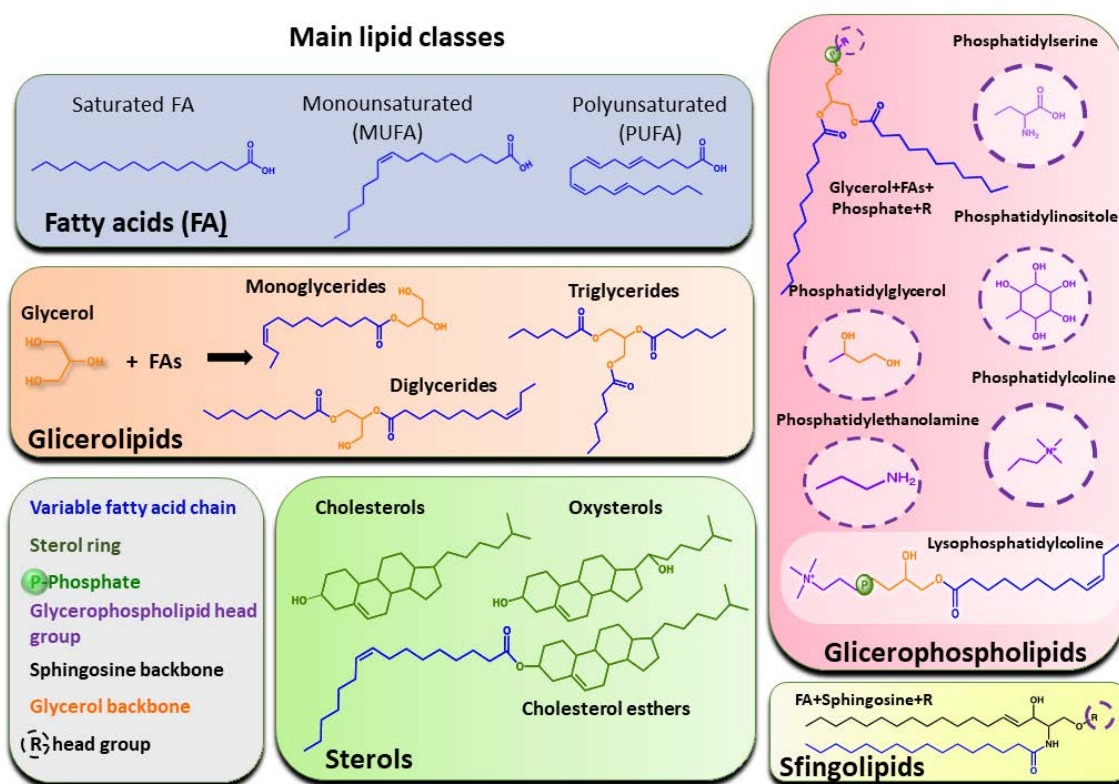
Figure 2-1. Structural classification of the main lipid classes in biological systems.

Many studies are directed towards exploring lipidomic changes and understanding the pathological mechanism behind various diseases [27–31]. Different families of lipids are implicated in a wide variety of biological processes and have a wide range of chief functions in human health and disease. Lipids play a crucial role in multiple positions in the organism: energy homeostasis, membrane structure and dynamics, storage and provision of energy, cell signalling and hormonal regulation. The systems-level scale analysis in lipidomics is displayed in Figure 2-2.
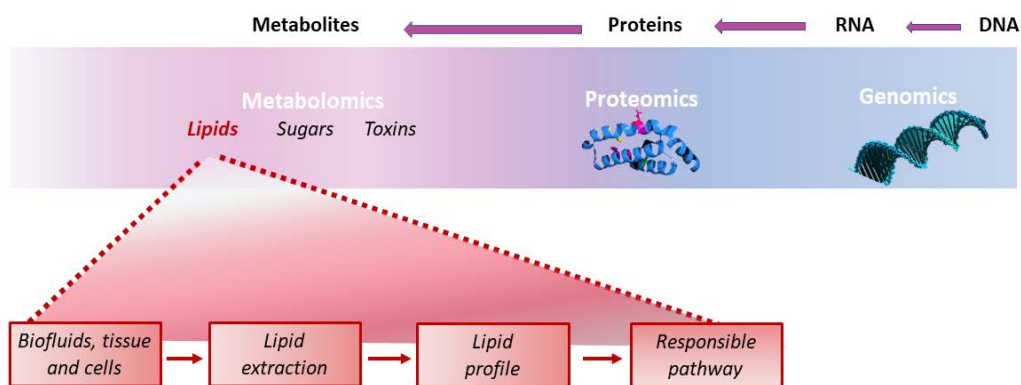
Figure 2-2. The scheme of different "omics" disciplines and the system-level scale analyses in lipidomic studies.

Thus, imbalances in lipid metabolism may contribute to diverse phenotypes and disease states. In addition, multiple known and novel drugs are made for targeting lipid metabolic and signalling pathways, such as those commonly used in the case of the inflammation process as cyclooxygenase inhibitors or those used in metabolic diseases, like statins, to decrease cholesterol levels. Moreover, lipids play a critical role in the structure of the central nervous system, particularly at the cell membrane level, not to say that they are responsible for membrane fluidity and the transmission of electrical signals [32].

Since lipids are the main constituents of cellular membranes and more complex molecules, changes in lipids mean changes in membrane fluidity. Thus, lipids alteration may lead to a cascade of processes, such as ligand-receptor signal transduction or membrane trafficking, influencing cell functions and survival. It was shown that the enrichment of sphingolipids and cholesterols, as well as content in polyunsaturated fatty acids (PUFA), directly determines membrane fluidity and movement of membrane proteins in lipid rafts [33,34], whereas phosphatidylcholine (PC) intermediates membrane fusion, and ceramides are potent regulators of cellular growth and death [35]. Thus, ceramide synthesis and metabolism are highly studied as promising targets for cancer therapy.

Gangliosides are abundant in the central nervous system (CNS) and involve multiple essential functions such as cell-cell recognition, signal transduction, synaptic transmission, cognition and oligodendrocyte differentiation. Changes in their levels have been associated with various neurodegenerative disorders, namely Huntington's disease, Alzheimer's and Parkinson's, and amyotrophic lateral sclerosis. Meanwhile, sterols and derivates are implicated in synaptic formation, axonal growth, signal transduction, learning and memory.

Given this perspective, it seems reasonable that defects and abnormalities in lipid synthesis and metabolism are involved in the onset and development of many disorders and diseases, most of all those that affect CNS. The brain requires a constant source of metabolites to maintain its functions and contains the second-highest lipid concentration in the human body after adipose tissue. Therefore, it is becoming increasingly evident that many neurological disorders involve altered lipid metabolism. The brain's energy requirement derives from the oxidation of fatty acids, entirely taking part in astrocytes [28].

Thus, for example, PUFAs play an essential role in the membrane structure and exert additional functions on cell signalling, particularly neuroinflammation and regulation of energetic metabolism. Furthermore, PUFA can be converted into active molecules, presenting anti-inflammatory or neuroprotective effects. Therefore, it was demonstrated that the omega-3 fatty acids derived from PUFA have an anti-inflammatory and neuroprotective effect in neurodegenerative diseases such as Parkinson's and Alzheimer's [36].

Imbalanced protein-protein and protein-lipid interaction also give place to the above-cited neuro disorders. Thus, aberrant cholesterol metabolism is linked to Alzheimer's [37]. Likewise, aggregates of beta-amyloid protein on the membrane surface and peptide aggregation are promoted by interaction with gangliosides [38,39]. Synucleins bind to fatty acids and lipids and regulate their oligomerisation [40]. Mutation in alpha-synuclein is associated with familial cases of early onset of Parkinson's' disease [41].

Likewise, aberrant accumulation of ceramides[42], precursor molecules of the sphingolipid's metabolism, which could be converted to sphingomyelin and

gangliosides, is commonly considered toxic. In addition, they mediate neuron death by oxidative stress and apoptosis in patients with neurodegenerative diseases [32].

Reactive oxygen species (ROS) lead to cellular damage by modification of the protein, DNA and lipid inventory [35]. Thus, determining oxidative events in membrane lipids and identifying the lipid species involved may provide critical information on the molecular mechanisms of ROS damage. Furthermore, slight or chronic imbalance of the levels of altered lipids contributes to altered cellular function. Therefore, it could lead to the onset of the pathology. Consequently, it is often proposed that lipids profiles are used as prognostic or diagnostic markers since they correlate with the physiological condition of the organism.

All these shreds of evidence suggest a strong correlation between lipid metabolism and motor neuron degeneration. Moreover, changes in lipid metabolism are frequently observed not only in patients with neuro disorders but in chronic infectious states such as HIV, metabolic disorders [43], cardio pathologies [44–46] and cancers[47–51].

2.1.2 Targeted and untargeted metabolomics

Metabolite detection and identification easily represent the two main aims of metabolomics research. Thus, the metabolite profiling/fingerprinting methods and biomarkers identification could be distinguished into targeted, untargeted and semi-targeted strategies. Targeted metabolic profiling requires certain preliminary assumptions and knowledge about the compounds, and it is usually based on the analysis of concentrations or spectra intensities of a small number of known groups of compounds, generally of the most abundant metabolites, which belong to standard physiochemical classes or biochemical pathways [52]. Since GC/LC-MS instruments are prone to batch and time drift effects due to instrument sensitivity and intensity changes, targeted methods could correct batch and drift effects by including labelled internal standards. However, this step would also require more sample preparation. Usually, the generated data would require a more straightforward statistical method. When the study aims to evaluate a specific metabolic pathway, targeted metabolomic is particularly recommended because it enables the creation of quantitative metabolite databases for diverse populations, allowing results to be compared with published literature, therefore, the performance of more sensitive and accurate detection [53].

14

Nevertheless, the targeted strategy presents certain risks because if the preliminary assumption regarding the compound searched was uncertain, the whole study would result in so-called ''much effort for nothing''. Many current targeted methods, discussed below, focus on enabling large-scale metabolic profiling, even if the untargeted process would still conserve the priority of maintaining a more extensive data set.

Compared to the previous strategy, the untargeted process does not require preliminary hypotheses regarding the metabolites that should be determined. Metabolic fingerprinting, as an untargeted approach, employs qualitative and quantitative holistic analyses with sufficient scope to identify as many metabolites as possible in the sample. Furthermore, untargeted studies aim to extract diverse metabolites that do not belong to the same biological class or biochemical pathway from a single sample. Usually, relative quantification is performed, and no metabolite concentration is reported. Due to its metabolome coverage, it is often the most applied approach for discovering potential disease biomarkers. However, the metabolite's quantity limitation depends on the sample type, sample extraction method, and analytical technique used. In addition, this approach is challenging and time-consuming since it requires metabolite annotation/identification; typically, some but not all metabolites are annotated or identified. Usually, after an untargeted strategy, new hypotheses are generated.

For example, using non-targeted metabolomics is also essential to obtain new endogenous up- or down-regulated metabolites, showing that metabolic pathways are impacted by food intake, demonstrating that not only the exogenous metabolites are relevant [54]. Plasma metabolomics-based studies do not usually aim to identify a specific metabolite but rather to comparens and ''fingerprints'' of metabolite changes resulting from disease progression or severity. Untargeted metabolomics would subsequently generate hypotheses relevant to biological pathways, disease mechanisms or drug response, ideally creating inputs and targets for new biomarker discovery. However, this approach generates larger data sets and would consequently require the subsequent use of multivariate statistical analysis. Recently, a semi-targeted approach was introduced in metabolomics studies [55]. As can be deducted, the semi-directed strategy lies between the two processes described above. This approach investigates

from ten to hundreds of metabolites at a time. For example, multiple metabolites are quantified (instead of one metabolite at a time, as in the targeted approach), applying a single specific calibration curve. This strategy is hypothesis generating and hypothesis testing. However, one method does not exclude another; all of theme complementary and used sequentially. The main differences between the two strategies are reassumed in Figure 2-3. Further ore, recent studies highlighted the promising contribution of analytical techniques to all strategies. Thus, Savolainen et al. [56] developed a combined method to acquire targeted and non-targeted metabolomic profiling in an accurate, reproducible and reliable way. Given the overall premise regarding metabolomics, the following sections will describe accessible authenticated instruments and procedures mainly applied during this doctoral thesis. Our findings are based principally on non-targeted metabolomics method development, focusing primarily on the techniques such as FTIR and UPLC for the metabolic profiling of human blood samples.

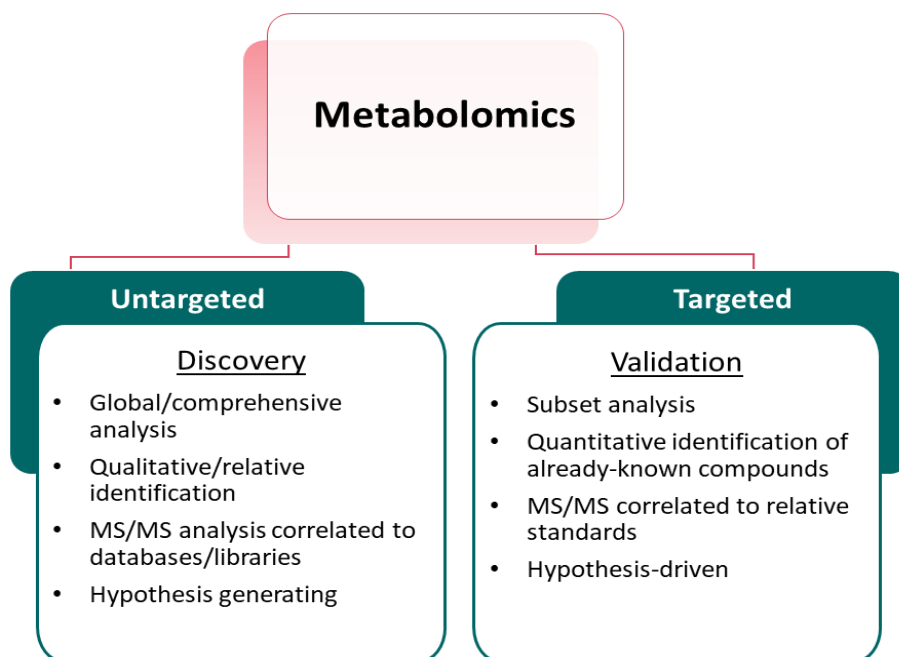**Figure 2-3**. Principal aims of targeted and untargeted strategies in metabolomic-based studies.

2.1.3   Blood as the main biofluid for metabolomics studies

Modern metabolomics consists of comprehensively identifying and quantifying metabolites in complex biological samples. For instance, the type of biological sample selected is critical, as it would define the whole metabolomic study.

A suitable analytical technique can be applied to each type of biological matrix. Sometimes focusing only on a single matrix may entail the loss of relevant information originating from another physiological pathway, which can provide significantly greater coverage of the system's metabolome. A unique multi-matrix platform could be of great relevance and utility. Still, it cannot always be provided with different sample types and not every instrument can be adopted for multi-matrix analyses. Biofluids, namely blood, serum, plasma, saliva, urine, cerebrospinal fluid, tears, and exhaled biofluids, are analysed separately due to their relatively easy collection and abundant biochemical information; they are analysed independently as diagnostic resources. Therefore, selecting an appropriate biofluid requires careful consideration, as it may impact the design and results of metabolic phenotyping studies. Due to its manageability and easy sample collection and preparation, blood is often the most suitable sample type. Together with plasma, it is the second most frequently applied biofluid in metabolic studies after urine [57]. Much metabolomic data has been generated from studies on blood since it is a uniform biofluid that is usually unaffected by factors such as diet or diurnal cycles, or fluid consumption, as occurs with urine and which could confound and interfere with the analysis [53,58–66]. As a small quantity of sample is needed, it appears ideal in studies incorporating metabolic approaches. Thus, there is an increasing interest in using ''liquid biopsy'' based on blood collection. Since blood biopsies are less invasive than tissue biopsies and hence amenable to the serial collection, providing crucial molecular information in real-time makes it more appropriate than other sample types (e.g., tissues), which require more preparation and are more invasive procedures [67]. Some limitations of blood could be associated with cases where venous access in patients is problematic or when personal aversion to blood sampling is present.

For centuries, illnesses were believed to be due to '' lousy blood''. As early as the times of Hippocrates, antically one of the most influential doctors, everybody believed that the human body was composed of 4 biofluids (yellow bile, black bile, urine, and blood). These biofluids were applied in the qualitative diagnosis of pathological state: so,. Soy yellow urine was an indicator of high urea or bilirubin concentration, whereas excessively dark blood was associated with low oxygenation. Moreover, the ancients recognised the importance of blood for thermoregulation and knew that this biofluid integrates many tissues in the human body. Today's knowledge is much more easily

accessible and has acquired greater speed. Blood is known to carry much biochemical information ''to'' and ''from'' tissues; it is a vector of multiple substances and compounds, especially metabolites. As both polar and non-polar metabolites are present in this type of matrix, a wide range of biochemical information is currently available [68] when compared with, for example, urine, which contains principally water-soluble waste metabolites, making this type of matrix limited and less suitable for the analysis of chemical functionalities. Blood composition could be divided into three main parts: 55% plasma, 1-2% buffy coat and 45% of formed elements Figure 2-4.
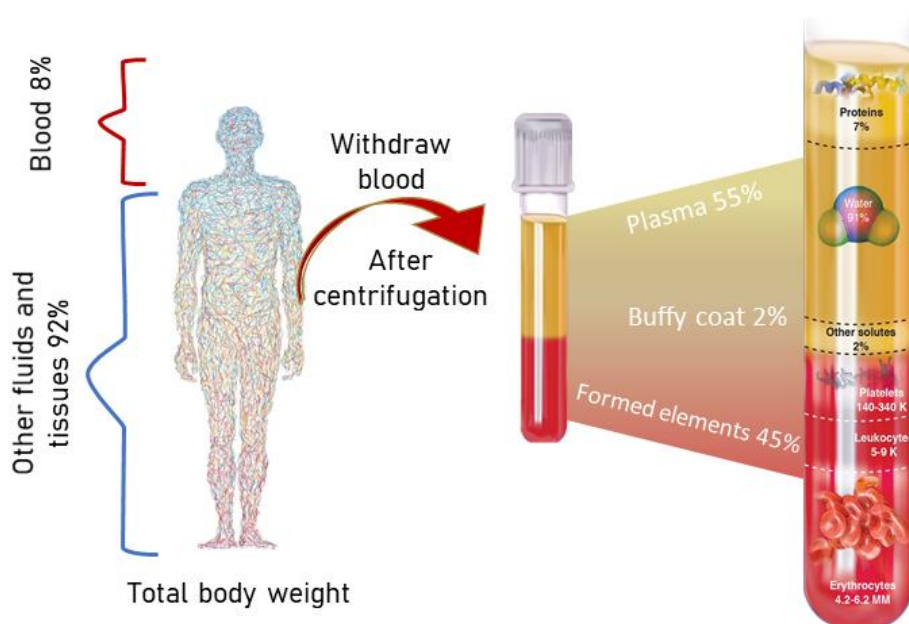
Figure 2-4. The scheme of different ''omics '' disciplines and the system-level scale analyses in lipidomic studies

Careful consideration of the potential impact of sample type on the design results of metabolic phenotyping studies should be considered. Plasma is a liquid volume of whole blood. It is mainly used for analytical purposes because the principal composition comprises water, lipids, sugars, amino acids, metabolites and higher molecular mass compounds such as proteins, DNA and RNA. Moreover, the plasma component of blood contains proteins and metabolites, including those that leak into the bloodstream from damaged cells after cellular injury; for this reason, post-mortem blood samples have been the most frequently analysed in metabolomic forensic analysis [69].

Usually, primary differences among blood sample types are explained by inter-individual variation and require individual correction techniques to detect differences attributed to sample type [70]. Furthermore, metabolites that discriminated serum from plasma included glycerophospholipids, lipoproteins, and energy pathway metabolites, while amino acids, glycerophospholipids, and energy pathway metabolites distinguished plasma samples with variable platelet content. It is evident that metabolomic analyses do not focus solely on endogenous compounds but comprise exogenously-derived metabolites, such as xenobiotics and their products deriving from phase 1 and 2 metabolisms [71]. For this purpose, plasma is usually highly suited for pharmacokinetic drug monitoring (e.g., direct oral anticoagulant quantification [17]. Nevertheless, compared to other biofluids, such as urine, blood is less prone to the influence of diet and diurnal variation of endogenous and exogenous metabolite levels, which do not contain discriminant power of pathological relevance [53].

So far, the application of plasma metabolomics is highly suited for accessing and understanding various illnesses, from diabetes to cancer [18,72–75]. Thus, cancer patients' blood lipid levels were associated with cancer pathogenesis and progression [51]. All these studies focused on identifying biomarkers in human plasma, providing encouraging results in terms of instrumental performance.

2.1.4 Metabolome measuring in human diseases.

''Omics'' technologies have promoted the understanding of the complex gene-environment interactions that impact health. Thus, clinical metabolic profiling evaluates gene-environment interactions that are usually connected to disease risk factors. Localised metabolic, physical or histological perturbations in the human body result in system-level changes in the organism. These changes are detectable by profiling the whole metabolome in the biological samples. The choice of a suitable analytical technique is influenced by many factors, principally due to variations in metabolite levels and their physiochemical diversity in blood samples. The ideal system would be precise, accurate and easy to use, with easily interpretable biological results. Despite the limitations mentioned above, metabolomic science fulfils most of these requisites. Moreover, the maturity of metabolomics has stimulated a worldwide ''boom'' in

metabolomics applications, capturing the principal directions of emerging metabolomic technologies from micro-organism engineering to plant-based applications [76,77].

Therefore, numerous analytical platforms are commonly suited to this matrix type and are widely applied for targeted and untargeted metabolomics. Hyphenated techniques such as liquid chromatography-mass spectrometry (LC/MS), gas chromatography-mass spectrometry (GC/MS), or capillary electrophoresis-mass spectrometry accomplish the metabolic profiling strategy, providing a detailed chromatographic profile of samples and measuring a wide range of metabolites. Otherwise, so-called ''high-throughput'' analytical platforms such as nuclear magnetic resonance (NMR) spectroscopy or vibrational spectroscopy, namely techniques such as Fourier-transform infrared spectroscopy (FTIR) and Raman spectroscopy (RS), belong to the ''fingerprinting'' methods. These platforms are generally faster and non-destructive, producing a comprehensive profile of metabolite spectra [78–86].

2.1.5 Principal limits of metabolomics studies

Unlike this, there is a growing interest in extracting biomarkers from biofluids. However, this involves some hurdles, considering that most metabolites detected in a complex biological mixture, such as plasma, tend to be ambiguous or completely unknown. Like every technique or approach, metabolomics is also affected by certain limitations, which should be considered and improved. For example, multiple parallel processes coincide in the organism, making the resulting target molecule or molecular signature ambiguous. Generally, the influence of individual phenotypic response on external or internal disturbances such as drug treatment, gender, diet, age or environmental stimuli, and genetic-related factors, could significantly affect metabolic balance in the human body, therefore making the metabolomic approach essential in disease detection. For example, the human body and human hare exposed to environmental factors such as Pb, which significantly affects metabolic profiling [87]. Not to mention that the increasingly ageing population, predominantly influenced by cancer and chronic diseases, make metabolomic analyses even more laborious and susceptible to errors, hence the need for superior diagnostic capabilities [88,89]. In addition, and interestingly enough, structural similarities or variables in the solution behaviour of metabolites may cause complications during the analysis. For this purpose,

a versatile and robust analytical technique is required. Each analytical method applied in the metabolomics exhibits unique and favourable characteristics, but which approaches does a better job? Which tecmethodovides a deeper understanding of the disease and is better adjusted to modern scientific requirements regarding sensitivity, reproducibility and accuracy.

2.2 ANALYTICAL INSTRUMENTS IN METABOLOMIC STUDIES

Due to the large number of chemical compounds and their diversity in physical-chemical characteristics, a universal technique cannot be selected to approach any problem in metabolomics. Instead, a wide variety of techniques exist and are used separately or combined to provide detection in metabolomic studies. Therefore, each technique individually provides determined characteristics that should be considered when selecting one or more.

The research group's experience in infrared spectroscopy and increased sensitivity that can provide mass spectrometry were decisive in developing this doctoral thesis. Thus, the following section will highlight the advantages and limits of MS and FTIR techniques in blood-based metabolomics studies.

2.2.1 Mass spectrometry

Current literature on plasma metabolomics highlights mass spectrometry as one of the main analytical techniques used in the ''omics'' field. MS provides information on measured compounds based on their mass-to-charge ratio (*m/z*) [90]. This analytical tool consists of three parts: an ion source, a mass analyser and a detector Figure 2-5.
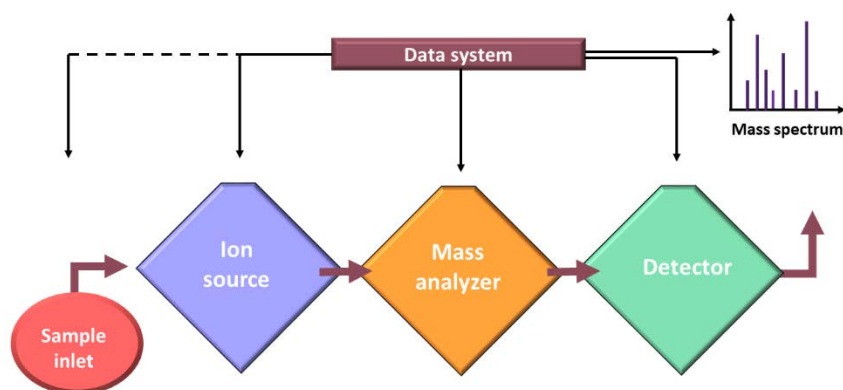
Figure 2-5. Schematic representation of mass spectrometry instrument. The mass analyser is the second component of a typical mass spectrometer sandwiched between the ion source and the ion detector. Its primary function is to separate ions based on their mass-to-charge ratio (m/z) so that they can be detected and analysed.

Before detection, each biological sample must be ionised, and the electron ionisation forms gaseous compounds, separating a multitude of ions present in the sample. Therefore, the MS instrument has many advantages since it can detect a wide range of metabolites in a minimal concentration range, from millimolar to nanomolar. Furthermore, due to its broad metabolic coverage, high sensitivity, mass resolving power and mass accuracy, MS suits perfect for the metabolic profiling of blood samples.

MS performance could be improved by selecting the appropriate instrumental or technical variants; thus, an ionisation approach could increase metabolite coverage. For example, electrospray ionisation (ESI) is the most preferable and utilised technique in biofluid analyses due to its ability to generate many ions. At the same time, selecting an adequate mass analyser could overcome the resolution limitations of the chosen separation technique.

Technological advances in high-resolution tandem mass spectrometry (MS/MS) have undeniably facilitated biomarker identification.

Unfortunately, many spectra still need to be assigned during analysis. MS fragmentation presents certain fundamental limitations, considering the complexity of metabolomic samples. Direct injection rarely provides accurate mass information. Even if biological samples could be analysed by direct injection, mass spectrometry-based

metabolomic studies are generally preceded by a separation step within a liquid or gas column. This combination typically uses hyphenated techniques through coupling two or even three analytical methods, mainly combining one chromatographic and one spectroscopic technique. The integrated system's basic workflow consists of separated metabolites deriving from interaction with adsorbent material inside the column that enters the spectroscopic instrument through an interphase [91]. The need for coupling arises from the existence of more complex problems, such as a need for efficient diagnostic and prognostic biomarkers in complex human illnesses. Nowadays, the coupling of MS is limited not only to chromatographic techniques, but it also complements high-resolution proton spectroscopy (^{1}H NMR) [78,92–94], GC [56,84,95–98] or capillary electrophoresis (CE) [6,99–104]. Coupled systems help overcome many limitations a single instrument faces, providing more comprehensive information and enabling access to enhanced opportunities for resolving complex biological mixtures such as blood.

2.2.2   Understanding of LC/MS and its use in metabolomic and lipidomic studies

Liquid chromatography hyphenated to mass spectrometry (LC-MS) is currently the most widely used and well-integrated technique in metabolomic studies. LC-MS-based metabolomics has revolutionised the study of small molecules and our understanding of drug discovery[105], chronic disease [106,107] and biomarker identification [108–110]. However, an untargeted disease phenotyping, searching for biomarkers that can give direct insight into the disease process, could be very frustrating. Therefore, the LC-MS approach is considered a good compromise regarding the speed of analysis, metabolites coverage and sample throughput. In addition, LC-MS is suitable for both qualitative and quantitative analytical strategies of metabolic phenotyping. Considering its ease of use, HPLC-MS with ESI in positive and negative modes are the most applied techniques for untargeted metabolic phenotyping.

Considering that the LC-MS approach was used to obtain results in this doctoral thesis, some experimental workflow steps will be deepened, dwelling on points that could create a real bottleneck in the analysis, adding bias to results or even nullifying them.

2.2.2.1 *UPLC*

The HPLC platform has been surpassed by the ultra-high-performance U(H)P LC-MS technique. Compared to the HPLC standard platform, UPLC operates with a smaller particle size of stationary phase (≤2 µM) and higher pressure, offering significantly improved chromatographic efficiency and higher resolution separations. The reduction of peak width also increases sensitivity compared to conventional HPLC. Because of the higher sensitivity and speed, UPLC has a better peak resolution, reducing the ion suppression problem, which is very important in complex mixture separation [111]. Careful selection of the column and LC conditions also provides rapid analysis and high sensitivity in metabolome coverage. A combination of reversed-phase (RF) and hydrophobic interaction (HILIC)-LC are utilised in metabolomics studies to monitor metabolites with different polarities.

2.2.2.2 *Mass analyser*

In addition to the higher resolution and sensitivity, frequent technical updates make liquid chromatography one step forward, among other techniques, to perform metabolic profiling analysis. When conventional coupling with ESI instrumentation results is insufficient to achieve confidence in identification, LC and MS are combined with a second analyser to ulteriorly improve sensitivity, mass accuracy, and isotope abundance accuracy [112].

Among the most frequently used mass analysers in untargeted metabolic phenotyping time-of-flight (TOF), hybrid quadrupole/time-of-flight (Q-TOF), and Orbitraps should be highlighted. These mass analysers provide achievable high mass resolutions, typically up to 50,000 full-width half maximum (FWHM) for TOF and Q-TOF mass analysers and a scan rate of up to 100 Hz [52].

The Q-TOF hybrid configuration allows ions of a specific *m/z* ratio (ideally related to a single metabolite) to be isolated in the quadrupole mass analyser that can only measure ions with nominal mass. The TOF, on the other hand, propels the ions entering the analyser by applying an electric potential orthogonally and accelerating them. In this way, by using equal energy to all ions, which will turn into kinetic energy, those with lower mass reach higher speeds and arrive at the detector first, showing a shorter flight time. Therefore, the time it takes an ion to cross the flight tube is measured, and both

small and large ions must arrive at the detector, making the equipment only capable of working in full scan mode. In addition, due to its characteristics, the TOF presents a better resolution for ions with a higher m/z ratio (Figure 2-6).
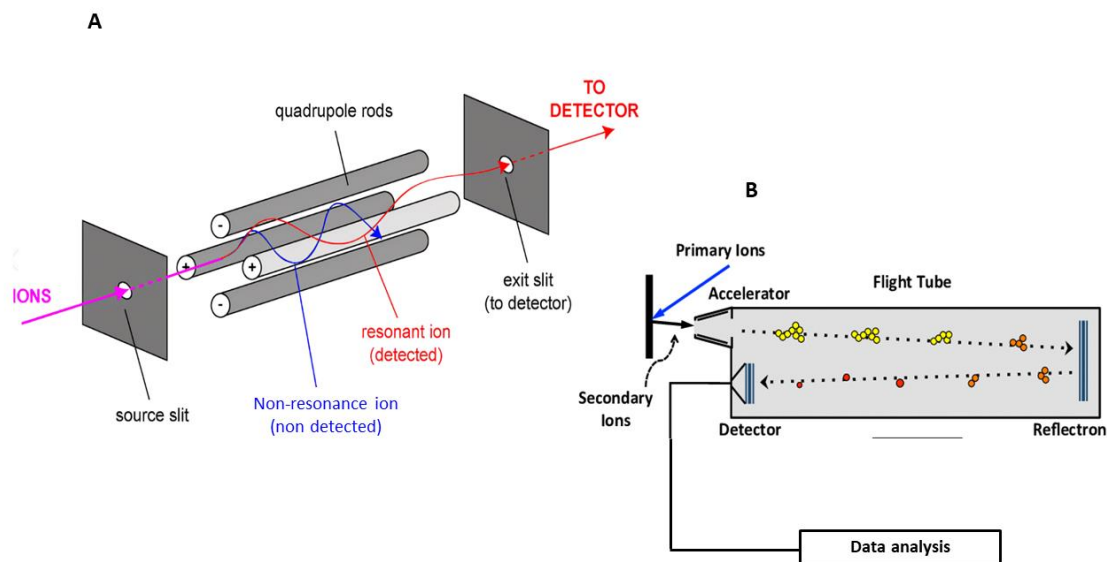
Figure 2-6: Scheme of ion behaviour in principal analyser used in metabolomic studies. **A)** The quadrupole analyser comprises four parallel cylindrical or hyperbolic metallic bars, to which a direct current and a radio frequency voltage are applied two by two. With the application of these potentials, ions entering one part of the quadrupole are intermittently attracted by one or the other bars so that only one m/z ratio can traverse it without colliding. Ions that traverse the quadrupole describe a helical trajectory without colliding (resonant ions) and finally exit the quadrupole, which. At the same time, they do not meet the m/z ratio; they collide before reaching it (non-resonant ions). **B)** Schematic representation of TOF analyser.

The hybrid instruments such as Q-TOF have, in addition to the first (quadrupole) and second analyser (time of flight), another "intermediate" analyser (initially a quadrupole, but currently it can be a hexapole or other designs) that acts as a collision cell. Thus, ions passing through the first quadrupole collide with an inert gas in this cell. Depending on the collision energy being applied, therefore, fragmentation can occur in a collision cell operated between the quadrupole and TOF mass analysers. Finally, they reach the TOF, obtaining a complete high-resolution spectrum of all ions (*full scan*).

This means that we work by acquiring two sequential functions, one of low energy and one of high energy; in the first acquisition (function), the ions do not fragment, and we can observe the pseudo molecular ion ([M+H]+ or [M-H]-). In contrast, the fragments

generated from all ions that pass through the quadrupole filter are observed in the second function. Because chromatographic peaks last a few seconds (in UHPLC around 6s) and simultaneous acquisitions are performed several times per second (about 0.3s per acquisition), in a 6-second peak, we can have 20 acquisition points, 10 of which will correspond to the low energy function (unfragmented ion) and 10 points of high energy (fragments), so we have information about the intact molecule and structural information in a single analysis with sufficient points to define the chromatographic peak. This mode of operation is called full-scan MS/MS analysis. However, higher scan frequencies allow many scans to be collected across a chromatographic peak, including complete scan data acquisition and data-dependent analysis. For this reason, the main advantages of the Q-TOF analyser, compared with other analysers, are the high sensitivity, the increased scanning speed and the high resolution, enabling a determination of the exact mass of the ions in the detector. Furthermore, it is incredibly advantageous for the structure elucidation process to use tandem mass spectrometry (MS/MS) made possible by hybrid instruments such as the Q-TOF. This allows us to isolate the ion of interest through the first quadrupole, fragment it at different collision energies, and then analyse its product ions to clarify the compound's structure without incurring errors in assigning fragments to the observed m/z relationships.

In this type of analysis, a huge amount of data is obtained, making bioinformatics necessary to extract the information required for metabolic analysis. In addition, it also provides isotopic pattern information, which is extremely useful in establishing the correct elemental composition of the observed ions.

Considering that this hybrid analyser is often combined with a UPLC instrument of separation, these technological advances maximise the benefits of UPLC-based metabolome profiling analysis.

2.2.2.3 Matrix

The use of biological matrices in metabolomic studies is pretty frequent since they are easier to treat as they are composed of a large percentage of water. Meanwhile, urine is a cleaner sample which requires dilution before chromatographic analysis; in contrast, plasma is a complex matrix rich in proteins. Thus, adequate bio-analytical procedures must be first applied to remove endogenous compounds and

isolate/concentrate the analytes of interest during biomarker detection [17]. This step often requires using organic solvents and different instrumentation, increasing times and costs. In this sense, matrices such as blood (serum or plasma) are treated to remove the present macromolecules (e.g., DNA and proteins) with acetonitrile (ACN), methanol, or ethanol, avoiding possible interferences in the metabolic analysis, which focuses on lower molecular weight molecules. Once deproteinised and centrifuged, the matrix can be analysed directly by liquid chromatography or undergo different treatments to adapt it to other separation techniques. Since lipids are involved in many cellular processes in the organism, lipid disbalance and alteration give place to multiple diseases; therefore, LC-MS lipidomic studies have gained popularity in recent years. Thus, for example, for lipid profiling in plasma/serum extracts, the mobile phase condition is usually studied and optimised to maximise the resolution of these analytes [113]. However, since the blood metabolome is widely extended, specific metabolic ranges must still be completed and undiscovered. Thus, serum might be more suited for small molecule analyses because the protein content is lower. Therefore, competing ionisation is reduced, and overall sensitivity increases [35]. However, the comparison of non-targeted serum and plasma studies revealed more peptide/protein fragments in serum, which resulted in additional disturbing effects during ionisation [114]. Implementing ulterior matrix type would provide a comprehensive overview of disorders not well reflected only in blood. Thus, a multi-matrix platform approach is already highly implemented by utilising polar metabolic fingerprinting of plasma, urine and faeces in UPLC-MS studies [115]. An acceptable performance criterion of these novel methods highlighted the merits of performing a multi-matrix platform for disease-related biomarker detection and potential pathway elucidation.

2.2.2.4 Extraction

LC-MS-based lipidomic analyses typically start with extracting the lipids from the biological samples, followed by LC separation. The extraction methods in lipidomic studies should meet the primary requisites, namely, be fast and reproducible. In addition, samples are often available in minimal amounts; thus, the main advantage of LC-MS-based analysis is that they require a small sample. Usually, for the analysis of lipids in the biofluids such as plasma or serum, the samples are collected if 10-100μL.

Several sample-preparation methods can be applied to biological samples, such as liquid-liquid extraction (LLE), solid-phase extraction (SPE), and organic solvent precipitation, to improve the overall lipid coverage. Most of the cited sample-preparation methods still rely on the extraction procedure that Folch et al. proposed.[116] years ago, which is based on a mixture of chloroform/MeOH (2:1, v/v) for extraction. This method is obsolete and is replaced by the procedure proposed by Matyash et al. [117], which involves the addition of MeOH and MTBE (1.5:5, v/v) to the sample. The phase separation is induced by water addition. Therefore, the organic phase containing the lipidic part will be situated in the upper layer, and its collection is much more simplified, minimising dripping losses. In addition, chloroform is listed as a toxic and cancerogenic chemical. Thus, utilising the MTBE method became preferable in laboratory practice, reducing significant health risks and environmental problems.

2.2.2.5 Column and mobile phase

Metabolite separation could be improved by appropriate column selection. Many commercially available columns are usually 100% compatible with the aqueous phase, maximising the retention of polar compounds. However, many parameters in this step offer the usual advantages and drawbacks; on the one hand, the column length could provide better separation performance; on the other, it would increase the time and costs of metabolomic studies. Besides the column, an adequate mobile phase regarding flow rate and separation capabilities should be considered. Indeed, the separation and detection of blood metabolites could be affected by many parameters, from the type of sample injection responsible for a resolution to gradient elution that could drastically improve or deteriorate the entire analytical flow. Thus, the mobile phase's acidification permits avoiding significant baseline disruption. Moreover, lipidomic studies implicate MeOH, which is known to facilitate lipid elution, avoiding its accumulation in the column, which may further contribute to the minimisation of matrix effects [118]. Nevertheless, LC-MS untargeted metabolomic studies require using other solvents such as ACN, which could not be considered" environmentally friendly", involving all the problems related to its use. Nevertheless, many protocols are standardised, entirely green and cost-reduced protocols in LC-MS-based metabolomics studies are still lacking.

2.2.2.6 Quality control

Untargeted metabolomics studies perform the metabolic profiling of hundreds and thousands of metabolites in the sample, where these analytes' nature has not been discovered previously or is partially hypothesised. Thus, the major problem in these studies is that adding internal standards for each compound could be more technically and economically feasible. Therefore, quality control (QC) samples are implemented in untargeted metabolic profiling to ensure the validity of analytical data. QC is used to demonstrate the goodness and accuracy of analytical performance and that results are reported correctly. The general approach of ''pooled'' QC samples, initially proposed by Sangster et al.[119], where aliquots of samples are combined to provide a ''mean'' sample, which is used to ensure the acceptable repeatability/reproducibility of the analysis. Therefore, features detected in the QCs are grouped based on the comprehensive CV, which acceptance criterion in untargeted studies is a maximum of 30% CV [120]. Different procedures can generate pooled QC samples, but the same storage and sample preparation protocol must be used to guarantee analytical precision.

2.2.2.7 LC-MS data processing

The LC-MS metabolomic analysis generates a massive amount of data. Each biological or QC sample analysed generates a single raw data file containing information related to retention time, *m/z*, and intensity and metadata listing instrument variables and operating parameters. The raw data format and the metadata included depend on the instrument platform. Thus, besides all the previous steps, which can increase time and cost and reduce the accuracy of the analysis, raw data processing is another challenging task in metabolomic studies. Converting data from all raw data files into a single data matrix requires using a range of software. Modifying this data format to a universal format should usually be applied before raw data processing, typically NetCDF with scientific instrument company software or mzML with ProteoWizard [19]. Modern instruments are provided with software that directly converts all the files of measured samples, but this process takes time.

2.2.2.8 *Compound identification*

The advances in mass accuracy and isotopic abundance accuracy are essential to reduce compound candidates. In addition, a series of MS^2, MS^3 and MS^n derive from a sequential fragmentation of selected ions, thus, containing important structure information such as feasible elemental composition and identification of core molecules, which allows the logical tracing of structure [106]. Therefore, a precursor ion mass (MS^1), a fragment mass (MS^2), and chromatographic retention time in combination with a chemical standard are used to confirm a specific compound in a peak unambiguously.

There are still two unsolved issues in LC-MS-based metabolomic studies i) currently, none of the available technologies is capable of providing a comprehensive analysis of all structurally diverse classes of metabolites in a single separation [121], ii) the structural elucidation of specific metabolites is limited to reduced coverage of spectral libraries. Thus, many discriminative features/compounds cannot be found in any spectral library or are stereoisomers. Furthermore, the improvements in informatics and growth of spectral libraries such as Human Metabolome DataBase (HMDB)[122], METLIN (https://metlin.scripps.edu), MassBANK[123], LipidMaps[124] still present some gaps, namely pure reference standards are non-available for many compounds. Therefore, metabolomics analysis could not consider the most potentially useful compounds or endogenously expressed unidentified metabolites. Thus, the main issue in LC-MS-based studies is that structural elucidation of unknown metabolites is intricate due to the incompleteness of measured MS/MS data in public or commercial databases[106]. Meanwhile, on the one hand, LC-MS can avoid the chemical derivatisation step typically requested during GC-MS analyses; on the other, non-targeted LC-MS-based metabolomic biomarker identification could present specific bottlenecks in terms of spectral library utility, which is much more limited compared to, for example, GC–MS [125].

2.2.3   Vibrational techniques

2.2.3.1 *Fourier transform infrared spectroscopy and its use in metabolomic studies*

Over the last decade, vibrational techniques have been highly implemented for biomedical applications demonstrating high specificity and sensitivity for disease

classification. Vibrational approaches, such as FTIR spectroscopy, are classical methods that provide non-invasive structural information about compounds.

Fourier-transformed infrared spectroscopy is an interesting diagnostic tool. Its great potential has already been demonstrated in a myriad of studies in different research areas [126–137]. The infrared approach is auspicious compared with other techniques, as it provides a faster, more accessible, low-cost diagnostic platform, thus fulfilling current clinical necessities. Furthermore, based on the Lambert-Beer law's selective absorbance of specific infrared wavelengths, where absorbance is directly proportional to concentration, it can accurately identify and quantify multiple compounds [138].

The electromagnetic spectrum of the IR region covers a range of wavelengths from 10 to 12800 cm^{-1} and is divided into three principal zones: far IR (FIR) from 10-400 cm^{-1}, mid-IR (MIR) from 400-4000 cm^{-1} and near IR (NIR) from 4000-12800 cm^{-1}.

 From the bio-analytical point of view, the most critical region is the MIR spectral region. The spectrum is acquired when the sample is placed in an IR beam's path [139]. The biological sample's absorption spectrum derives from the capacity of organic chemical bonds present in the model to undergo transitions in the MIR region. Molecules have naturally occurred frequencies of rotation and vibration since they are not rigid sticks (e.g., stretching or bending). Based on the chemical bond vibrations of samples, MIR provides a wavenumber-absorbance intensity, typically called ''fingerprint spectra''.

The metabolic fingerprints can be extracted from different matrices, such as tissues, cells or biofluids. Therefore, blood-based IR spectroscopy, whose use is widespread. Blood could be considered an ideal biofluid for IR analyses, as it is homogeneous and readily available. For this purpose, FTIR spectroscopy lends itself very well to clinical routine. In addition, the significant advantage of FTIR spectroscopy resides in its sensitivity to small changes in the composition of an ensemble of known and unknown biomarkers of a complex fluid such as blood, which provide a unique FTIR signature that can be used as a ''barcode'' of the disease.

The driving force behind this transition from traditional laboratory analysis is the need for diagnostic information in a timelier manner, not forgetting additional arguments such as costs, safety and environmental considerations.

All these characteristics make FTIR instrumentation one step forward in clinical implementation for diagnostic purposes. Vibrational spectroscopy has been generally described as an easy-to-use, non-invasive, non-destructive and reagent-free method, significantly reducing analysis costs [140]. Furthermore, FTIR enables the simultaneous determination of multiple constituents through a single analysis; thus, it is ideally suited for determining the structural features of proteins in biofluids. Independently of the initial aim of the study, FTIR spectroscopy is twofold in terms of the purpose of analysis, allowing both quantitative and qualitative applications.

In addition, current FTIR instruments are small and portable, with easy installation (e.g., no purge of nitrogen is required). Not less important, the instruments are very affordable and intuitive in their utilisation, requiring minimum user training preparation. Modern instrument configurations such as Attenuated Total Reflection (ATR) spectrometers are becoming very popular in point-of-care diagnosis. In ATR spectroscopy, the sample (in order of a few microliters) is deposited directly onto the surface of an ATR crystal (typically Zn-SE) with a high refractive index compared to the sample Figure 2-7. Therefore, the deposition system on ATR crystal is easily cleaned after each examination, compared to standard analysis in a cuvette. Nevertheless, on the one hand, ATR allows for performing multiple and reproducible analyses with minimum effort; on the other, the measured sample could not be recuperated after each measurement.
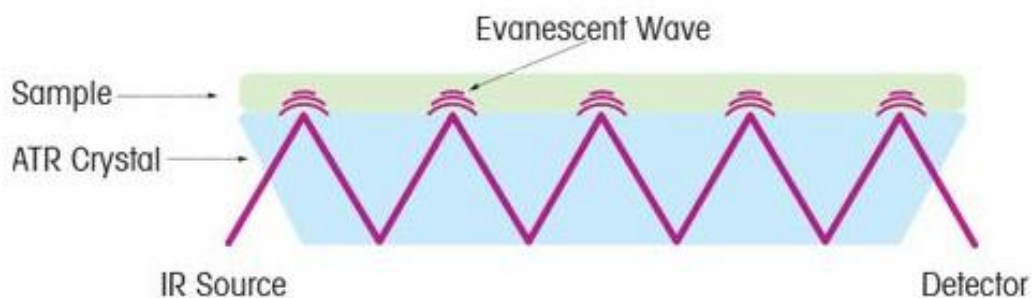
Figure 2-7. Diagram of Attenuated Total Reflectance accessory (ATR), in which the infrared radiation passes through the optical material- ATR crystal. The main advantage of using this accessory is that the measurement is very simple, just by placing a drop on the accessory window (the drop size does not matter).

It should be said that IR analyses are often affected by spectral perturbation, such as dispersion effects due to water content in the samples. Sometimes, the experimental factors (e.g., temperature) could add ulterior spectral artefacts. Nevertheless, these limitations are easily resolvable with opportune informatics techniques and compensated with little or any sample preparation time.

2.2.3.2 ATR-FTIR instrument

ATR-FTIR spectroscopy platform is highly used to detect spectral biomarkers in different malignancies and disorders. Moreover, it allows the delivery of rapid information, which is particularly useful in cases when a result is urgently required. The utility of this analytical platform has been demonstrated in many fields, from studies investigating saliva in rodents [141] to studies of a wide range of cancers [142,143] such as meningioma, colon, tissues or even ovarian cancer [144]. Considering that cancer has a high mortality rate, the correct identification and early diagnosis of cancer would significantly improve success in this field, enabling early-stage therapeutic intervention. However, defining the chemical phenotype in affected human subjects often requires laborious and time-consuming analysis. Therefore, searching for the best analytical technique to predict cancer incidence, severity and progression continue. Some progress has already been made, but other branches remain uncovered and must be elucidated [145].

In addition, the potential of the ATR technique is often reinforced by the results in accuracy. Thus, Bury et al. [146] performed a study in which ATR-spectroscopy analysed

various brain tumours from blood plasma. Their study demonstrated the potential of ATR-FTIR for detecting patients with primary metastatic brain tumours, with up to 100% accuracy for high-grade glioma vs low-grade glioma, achieving 88-100% accuracy for meningiomas. Their results are auspicious and could improve future clinical workflows.

Further research has demonstrated the potential of ATR-FTIR as a spectroscopic diagnostic tool, confirming that it can detect and quantify malaria parasite infection in wet packer red blood cells (RBC) samples [147]. Furthermore, untargeted metabolic fingerprinting strategies based on FTIR were used to exploit the disturbance of metabolic patterns and biomarker candidates of childhood obesity [148] field is critical to safeguard more fragile populations, as obesity could lead to several severe complications in the future, namely cardiovascular diseases or type 2 diabetes mellitus.

All these studies highlighted the main advantages of ATR:

1. Minimal sample preparation: ATR requires minimal sample preparation, making it a simple and easy-to-use technique.

2. Wide range of sample types: ATR can analyse a wide range of sample types, including solids, liquids, and gases.

3. High sensitivity: ATR has high sensitivity, allowing it to detect small amounts of analyte.

4. Minimal interference: ATR measurements are minimally affected by water vapour, carbon dioxide, and other substances that interfere with FTIR measurements.

Non-destructive analysis: ATR is a non-destructive technique, allowing for further study of the sample.

As we can see, the development of medical diagnosis based on vibrational spectroscopy is based on two steps: i) identification of the collected biomarkers and ii) classification of spectral biomarkers. Usually, the spectral wavenumbers used for biomarker identification and elucidation are considered variables. Thus, these untargeted analyses are coupled with chemometric techniques to reduce the spectral variables (from hundreds to approximately a dozen) and minimise the influence of noisy

34

and redundant information. As a last step, biochemical assignment of the selected important essential performed, which should be further validated by other targeted studies and with the utilisation of additional analytical platforms.

Despite its potential, FTIR spectroscopy is not currently used in clinical practice because most available studies were proof of concept studies.

This analytical method does not purport to resolve the question of the final diagnosis. Still, it could help screen blood samples of multiple disorders, providing spectral biomarkers for initial objective diagnosis or monitoring of the disease progression. Given this perspective, the FTIR technique was used in this dissertation thesis to evaluate spectra biomarkers of normal vs diseased samples, identifying spectral biomarkers responsible for metabolic-specific changes.

2.3 ANALYTICAL FLOW

To ensure the efficient handling of large data sets, each analytical platform follows an analytical workflow that includes a sequence of steps. Defining a standard workflow to obtain valuable and conclusive results and avoid errors is crucial. This process usually starts with the experimental design and with the validation of results. Thus, the most widely reported and applied steps include sample collection, sample pre-treatment, metabolite detection and data treatment. All of this process will be discussed throughout this chapter. The selection of steps depends on the type of analytical method used in the study and the kind of sample to be analysed, i.e., liquid or solid. Most of the required steps are similar regardless of the chosen analytical platform and could reveal unintended sources of variation in analytical yield. Thus, decisions about sample preparation, pre-treatment or experimental conditions to be used during the analysis could significantly alter the quality of the analytical results and increase challenges during metabolomic analysis [8]. Moreover, these analytical steps are often time-consuming, laborious, and error-prone. Due to the diversity of blood components and the need to improve the coverage of various metabolites, an effective and suitable protocol for sample storage and pre-treatment should be selected. Figure 2-8 provides an overview of the main steps in analytical workflow in blood metabolomics-based

studies, describing the most resilient analytical steps independently of the analytical platform.
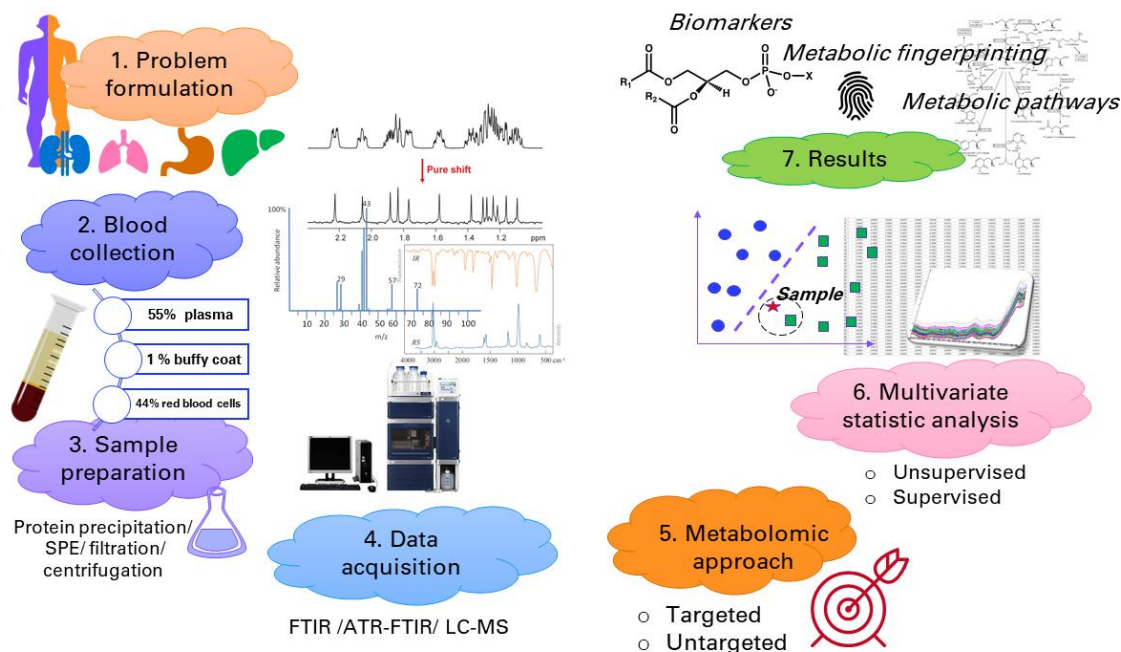
Figure 2-8. Schematic representation of the analytical workflow of the results obtained in this doctoral thesis*.

2.3.1   Sample design

The metabolomic workflow begins with sample design. The samples in the model should be "well distributed" to uniformly represent all potential sources of variability in the population of interest or at least the most important ones. Since sample design is often underestimated in the analytical routines, this study emphasised its importance, as it plays a crucial role in analytical workflow and could be considered the cornerstone of nearly every research study. Unfortunately, most scientific publications attach little importance to this step, and sample collection results are often unrepresentative, making the whole research worthless. Generally speaking, the analytical model can only be developed using appropriate standards, and clearly, these same standards must be well distributed over the concentration range of interest. Therefore, it is good practice in classification problems to be provided with something comparable in concept to a set of standards for developing the model.

Two strategic aspects of this concept must be highlighted: firstly, it could be a preliminary step to sample acquisition to allow a design of a representative population sample, thus achieving not only the whole population in similar proportions of subgroups but contemporary the whole sample set large enough to provide meaningful information and avoid bias; and secondly, sample design could be seen as a subsequent step of sample collection, selecting previously collected samples for further statistical steps. We have discussed this topic in previous studies, in which we provided detailed explanations of the main principles of developed techniques for appropriate sample selection [149].

Since this method represents a crucial part of statistics, Forina et al. [150] exhaustively described the theoretical details of the method, evaluating that a good sampling design selects the samples for calibration or regression with a uniform distribution. A subset of samples representative of the population of all possible samples can be chosen by applying different methodologies. The strategy to be adopted could be straightforward or, on the very contrary, complex; therefore, many techniques for uniform design are proposed. Having available the necessary information (corresponding to the most important sources of variability), a good sampling design can be obtained using methodologies such as Kennard-Stone (KS) or Pugwain's method for uniform design. Thus, the K-S design can get one or more sets of samples and can be applied separately for each class (60% for training, 20% for validation and 20% for prediction) to maximise the minimum Euclidean distances between selected and unselected samples [151]. Likewise, the K-S sample selection algorithm could be used in IR analysis, dividing spectra into training (70%) and test (30%) sets using, where the training set is required for constructing classification models. In contrast, the test set is often used for final model evaluation [152]. Other methodologies are applied when it is necessary to consider other unknown or less important factors, which encompass different types of variability in the population data set.

2.3.2   Sample collection and preparation

Long-term experiences have shown that basic steps, like sample analysis, could lead to undesired variations even at the first view. Once the patients are selected, sample collection is the next important step. Often, it is widely underestimated, leading to

analytical errors that could be not fixable anymore. Independently of the chosen analytical tool, the necessary precaution should be taken before proceeding with any analytical step to avoid ulterior bottlenecks in laboratory workflows. Thus, concepts of good laboratory practice, such as randomised sample order or batch acquisition, are often underestimated, therefore, could be a source of internal variations too.

To start, it is essential to have a thorough understanding of the type of samples to be worked with, as the more information we have about the sample type, the less chance there will be of making errors that compromise the results. We must ensure the samples have common characteristics (e.g., not mixing plasma with serum); otherwise, uncontrolled differences could mask differences in the compound of interest. Thus, sample homogeneity allows us to work with many samples and obtain more robust results.

In addition, since blood is readily available, uniform and homeostatic biofluid, biological replicates should still be accurately considered to obtain high data quality and reliability of the studies. It is important to perform a universal and straightforward sample treatment, or if not possible, perform several simple treatments to cover the broadest possible range of polarities in terms of compounds. Therefore, plasma samples should be collected similarly throughout the study and stored under the same condition. Usually, blood samples are stored at -80 °C to stop any metabolic activity which could continue in the cells [52]. The plasma samples tend to have specific stability in room temperature conditions; repeat freeze-thaw cycles should be avoided because they could lead to ulterior variance [153]. For example, the study of Lovergne et al. [127] observed that the sample drying at room temperature resulted in a difference in the sample residual water content that was dependent on the relative environmental humidity; they concluded that additional devices such as purging systems could be implemented better to control the ambient air humidity for plasma based FTIR analyses. When the serum is the matrix of choice, a clotting process could be involved during sample collection, and centrifugation should be executed to separate the liquid plasma from red and white blood cells. In this case, when the analysis is performed by liquid chromatography, often, anticoagulants are required to avoid the clotting process. In addition, sample preparation workflow generally involves an extraction of the

metabolites based on the metabolite under investigation and the method used for the analyses. Compared to cells and tissues, plasma samples would not require lyophilisation, which in other matrices is necessary to prevent peak overlapping. Usually, the first step in plasma studies involves the precipitation of proteins. Some techniques require only drying samples for IR, while others require LLE [63] by adding solvents to precipitate higher molecular mass biochemicals such as proteins. Sample drying could be avoided in LC-MS when an appropriate solvent is selected, decreasing the analysis time. When unextracted blood samples are used, it could lead to false positive results making the differences in lipoprotein particle composition look important [53]. In addition, comparing the sample preparation protocols for urine, plasma and tissue samples showed that plasma sample metabolic profiling is relatively similar to specimen preparation of urinary metabolic profiling but without the need for incubation with urease enzymes [84]. To highlight another advantage of plasma samples, the metabolite stability of plasma in the storage condition is less susceptible to changes than urine, which has more possibility of bacterial contamination [154]. Thus, the sensitivity of the UHPLC plasma profiling could be enhanced significantly by applying a novel approach to sample preparation [71]. For example, plate removal and extraction with different solid phase extraction (SPE) phases to concentrate plasma samples for the following direct injection of concentrated plasma demonstrated to improve the detection of less abundant xenometabolites in UHPLC untargeted metabolomic studies. Moreover, continuous improvements in instrument automation could reduce costs and time associated with sample preparation.

2.4   CHEMOMETRIC ANALYSIS

Once data are collected, they must be interpreted and transformed into valuable information. Most of the time, analytical techniques generate spectra, and the resultant metabolite data interpretation relies on the application of chemometrics. The term ''chemometric'', according to the definition of the Chemometrics Society, is "the chemical discipline that uses mathematical and statistical methods to design or select optimal procedures and experiments, and to provide maximum chemical information by analysing chemical data". This term was introduced in the early 1970s by Svante Wold. It is now an indispensable discipline in analytical studies, especially in metabolomics.

Almost two decades ago, Workman [155] summarised the critical advantages of chemometric applications, such as speed in obtaining high-quality information in real time from data. Furthermore, the author mentioned that chemometrics promises to improve measurements and knowledge of existing processes. Moreover, last but not least, it is economical, which is why chemometrics became an integral part of metabolic profiling.

As spectroscopic chemistry generates large and complex datasets, a fast and accurate statistical tool is required to deal with the complexity and volume of data generated in metabolomic studies and to interpret the relative metabolic changes, for example, at different stages of a disease. Therefore, the analytical step in chemometrics is recommended and essential. This chapter will discuss the strengths and drawbacks of each analytical technique. In addition, the role that takes chemometrics in the analytical workflow will be highlighted, summarising the most frequently applied chemometrics methods in blood-based metabolomic studies performed in this thesis.

The chemometric data analysis workflow in metabolomic studies consists of various steps, such as data pre-processing and unfolding, discriminant class-model generation and finally-validation. Figure 2-9 provides a comprehensive insight into the main aims of different chemometric stages, suitable for targeted and untargeted blood metabolomic analysis.
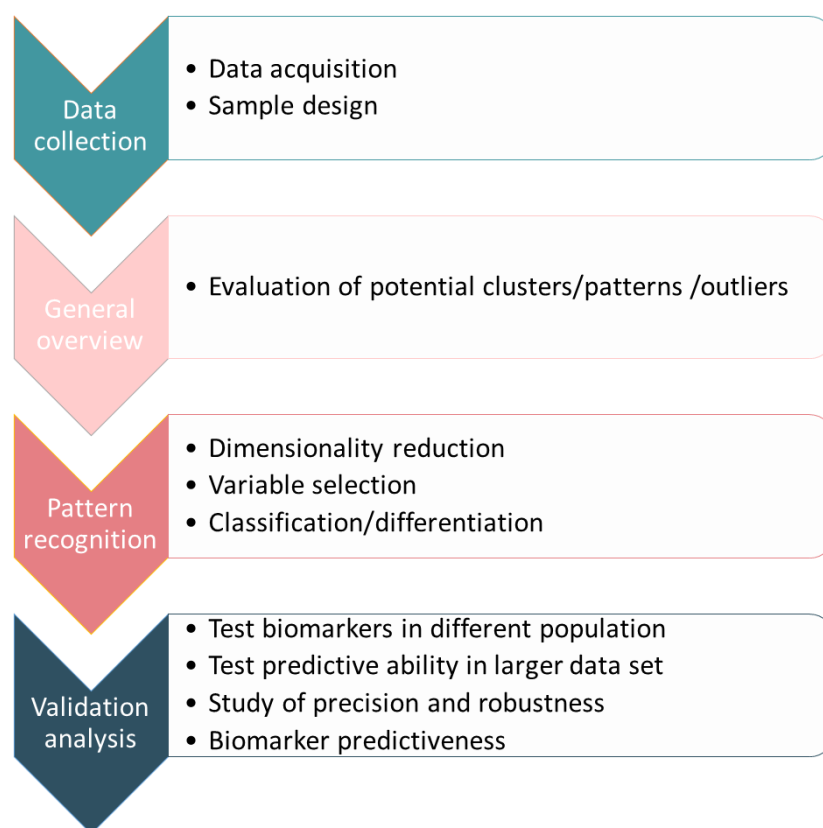
Figure 2-9. Overview of main steps of chemometrics analysis applied for discovering biomarkers in blood-based metabolomic studies.

2.4.1   Pre-processing step

Metabolomic studies are performed with many samples, expressed in different formats, such as ''ppm'', ''retention time'' o ''wavelengths'', relative to the intensities of spectra or concentrations. Long years of studies have demonstrated that generated raw omics data could not be processed directly before a proper pre-processing step. So, once the raw data is acquired, it goes through data storage, conversion and import. Once these steps have been completed, the data are ready for pre-processing. Applying pre-treatment techniques to spectroscopic and chromatographic signals minimises variations in the measurement conditions since the data obtained from these platforms may show fluctuations due to temperature changes or modifications in the optical length causing overlapping bands. Such variations can also be observed as scattering, shifts between signals, or irrelevant information in the spectra or chromatograms. As mentioned, blood contains heterogeneous molecules, some in the form of long-chain lipids or proteins. For example, during the analysis, they could appear as baseline distortion, thus precluding and impacting metabolite quantification [156] [52]. Suppose a suitable algorithm is not applied to correct initial spectral distortion. In that case, these

41

minor and innocent variations will introduce significant bias in the subsequent multivariate analysis, obscuring valuable information relating to biomarkers.

Various pre-processing tests were studied and used in this doctoral thesis to correct specific issues associated with spectral data acquisition and build robust and accurate models. Thus, these methods can be categorised as signal correction methods and classical pre-processing methods [157].

2.4.1.1 Signal correction methods

In spectroscopic data, the different perturbations can be addressed to light scattering effects, which induce a photon loss (additive effect) and an increased path length (multiplicative effect). Almost all of the pre-processing methods indicated below are ''row-wise'' and the most widely applied include:

Normalisation methods such as multiplicative scatter correction (MSC) and extended multiplicative scatter correction (EMSC), fit each spectrum to a reference spectrum; standard normal variate (SNV) is extensively applied to correct a non-desired intensity variation between objects. Normalisation could be done to a particular peak, for example, to avoid variations in signal intensities attributed to experimental sources or ion intensities in the case of LC-MS analysis or to the area (area normalisation).

Baseline correction removes experimental and instrumental artefacts; this method can be used to eliminate background noise in IR spectra by subtracting the lowest value of each spectrum from all variables,

 Smoothing is also used to correct random noise through different available digital filters (e.g., *Gaussian, Savitsky-Golay, and Moving Average*). Thus, **Savitzky-Golay smoothing** is a digital filtering technique that uses a type of moving window in which a set of adjacent data points are averaged. Still, unlike other smoothing techniques, it also fits a polynomial function to the data within the window. Once the polynomial is fitted to the data within the window, the function's value at the window's centre is used as the smoothed value for that point. This process is repeated for each position of the window along the signal, resulting in a smoothed signal. Likewise, the **Moving average** uses the average of neighbouring points to calculate the new value. The window moves along the data set, and the average value is recalculated for each point in the series. The window size is an important parameter in this method, and it can be chosen based on

the desired level of smoothing or noise reduction. A larger window size will result in a smoother data set, but it may also introduce some lag or delay in detecting changes in the data. A smaller window size will provide more sensitivity to changes but may also result in more variability or noise in the data.

Derivatives are generally used to correct for additive effects in spectroscopy (e.g., *Savitsky-Golay,* which smooths and enhances the noise).

Alignment methods are often applied to chromatographic data by means of signal comparison. It is required that the peak corresponding to the same compound does not show any variation in retention time in the different spectra or replicates. To solve such variations, an alignment pre-treatment is usually applied to the chromatographic data. One of the most relevant techniques for correcting the shift between chromatograms consists of making shifts concerning specific chromatogram signals, using them as an internal reference.

2.4.1.2   Classical pre-processing methods

Scaling is applied to avoid higher signals having a more significant influence than smaller signals and spectral alignment to correct local signal shifts (typically observed across the retention time axis of MS data). Thus, *Row Autoscaling or Standard Normal Variate (SNV)* treatment consists of centring the column and normalising it. *Pareto scaling* consists of adjusting each variable's magnitude to equalise the noise level in all variables. Therefore, it is used when noise is expected to be proportional to the square root of the standard deviation of the variables.

Mean centring is the most common pre-processing in projection methods such as PCA or PLS, which will be discussed below. It is applied to centre the subspace to the barycentre of the original data set for better data visualisation.

Each technique could require the application of more than one pre-processing step. Indeed, in the case of hyphenated approaches, data processing proceeds through multiple stages, including filtering and feature detection (or peak picking), for the removal of noise and accurate peak detection and peak matching across samples; alignment and normalisation for the correction of drifts in retention time and intensity between samples, respectively [158]. Furthermore, MS-based datasets are usually affected by chemical noise, which is typically induced by molecules in buffers and

solvents and can be especially strong at the beginning and the end of elution. This type of noise causes a shift in the baseline in the intermediate mass range of LC-MS, and the noise filtering of LC-MS data could be more challenging compared to other analytical techniques [125].

Once the data are improved by an adequate pre-processing method, ensuring the robustness and accuracy of the subsequent steps, multiple univariate and multivariate techniques are performed. Conventionally, this step leads to metabolomic feature extraction, which removes outliers and irrelevant information and reduces data dimensionality.

2.4.2 Chemometric methods

One of the main aims of applying chemometric algorithms is to identify the relationship between chemo-metrically characterised objects. Thus, pattern recognition, classification and class modelling methods are used to pursue this scope. Pattern recognition methods depend on a priori knowledge of the system under study. Thus, having a series of objects that belong to different categories already known, it is necessary to establish a classification model that allows the classification of future unknown objects into one of the categories provided by the initial objects. Usually, the chemometric techniques could be divided into two main categories- univariate and multivariate.

The main statical methods performed in this doctoral thesis to analyse and interpret the biological results will be described hereunder.

2.4.2.1 *Univariate statistic*

The univariate approach and multivariate analysis study variation and test the statistical significance of parameters and variables in metabolomic studies. The standard statistical tests applied in the univariate method are divided into *parametric and non-parametric tests*.

Parametric tests, such as *Student's t-test* or *ANOVA*, assuming that normality is verified, seek to identify differences among two or more groups. Usually, the Student's t-test and the false discovery rate (FDR) are applied to determine significant differences ($p < .05$, $q < 0.01$) in metabolite expression and pathways between diseased and control

samples. In addition, they are used to show correlation and display significantly altered levels of possible biomarkers between groups [159]. The *Kolmogorov-Smirnov* normality test often confirms the two tests' output. Usually, heatmaps based on Student's t-test or ANOVA are performed for better visual interpretation to indicate each biomarker's contribution to group separation Figure 2-10.
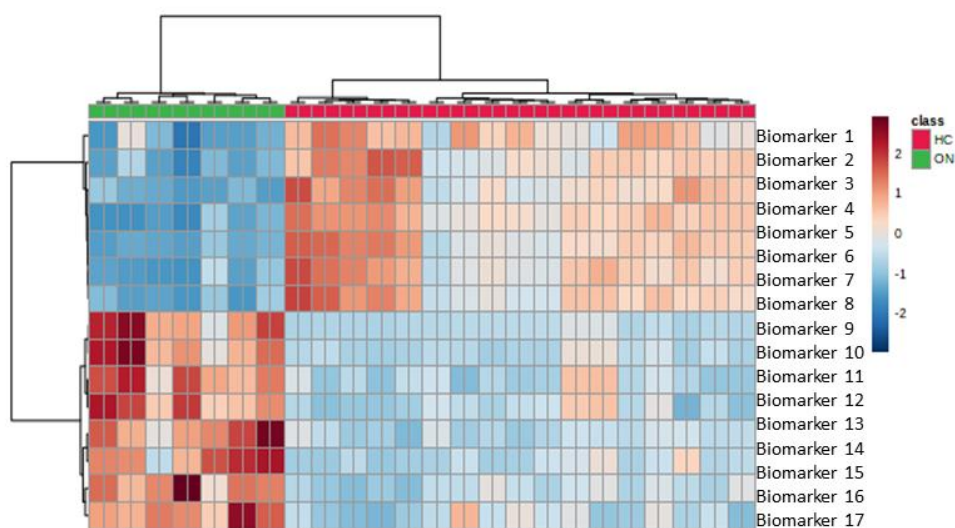
Figure 2-10. Schematic representation of heatmaps.

Non-parametric tests, such as the *Mann-Whitney U or Kruskal-Wallis tests*, are applied when the data normality cannot be assumed. This test is usually used for metabolite peaks to find a significant difference between case and control (p-value <0.05), indicating changes in several metabolic pathways involved in the disease under study. In addition, in metabolomic studies, the *Bonferroni correction* is often applied to minimise the probability of at least one false positive in many metabolic features. Thus, for example, when searching for metabolic biomarkers for differentiation of diseased/controls, after Bonferroni correction, the defined significance level could reduce or confirm the initially suspicious metabolites [160].

One main advantage of the univariate statistic is that it is easy to use and interpret. In addition, many problems could be directly solved by applying only the univariate statistic, such as finding disturbed metabolic patterns or the severity of the disease [161]. Therefore, many metabolomics studies obtained valuable knowledge about plasma biomarkers and potential therapeutic targets associated with infection by

applying the univariate approach to the problem. Furthermore, these findings serve as an essential resource for further research into the pathogenesis of the disease under study.

However, high-dimensional data, so typically attributable to modern metabolomic studies, have placed the univariate approach slightly on the back burner. The so-called ''rule of one'', changing just one feature at a time, does not enable the visualisation of large-scale datasets. Moreover, considering the features independently does not yield satisfactory analytical results regarding data interpretability.

2.4.2.2 Multivariate statistics

Since the differences between groups are rarely known in advance, using multivariate techniques is essential to highlight these differences. Moreover, the multivariate approach considers that generated analytical datasets could be affected by substantial correlation among features, variable collinearity or interactions, thus revealing their relationship pattern.

Many different multivariate methods have been described in the literature for data modelling. As mentioned before, one of the primary steps during multivariate analysis is extracting features from the obtained data set. Feature extraction methods can be divided into supervised methods (samples are allocated into pre-established classes) and unsupervised methods (no prior assumption is made on the samples).

Due to the large number of features involved in the study that could correspond to spectral wavelengths or registered peak intensities, the analyst applies methods to reduce the dimensionality of features to produce a smaller number of variables with statistical significance.

This section provides an overview of the methods that were widely applied to perform this doctoral thesis, such as principal component analysis (PCA), selection of predictors, linear discriminant analysis, partial least squares-discriminant analysis (PLS-DA), etc. [162–166]. A schematic representation of these methods is reported in Figure 2-11.
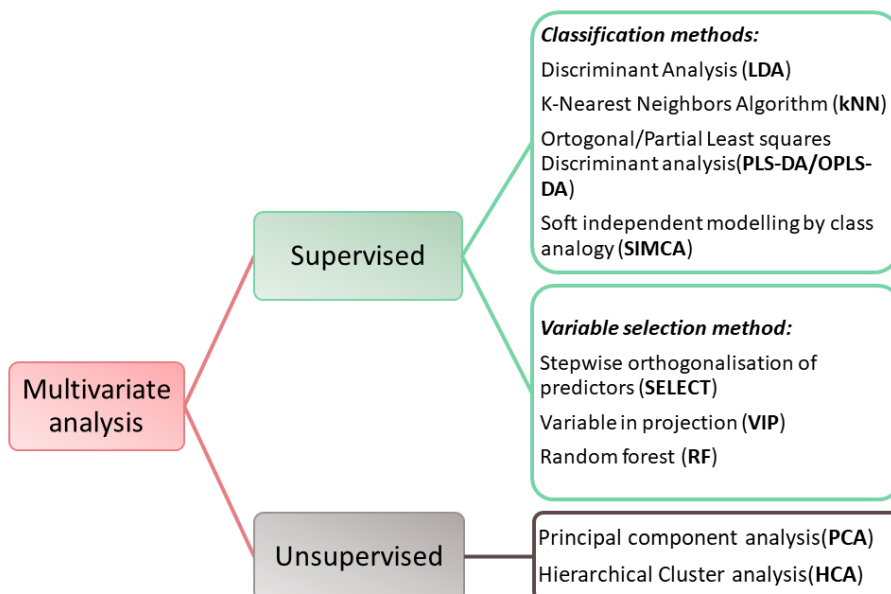
Figure 2-11. Summary of the most applied multivariate technique in this doctoral thesis.

2.4.2.2.1   Unsupervised methods

PCA is one of the most widely applied unsupervised methods, based on decomposing the data matrix into principal components and generating plots of scores and loadings, respectively. The score vectors represent the projection of each sample. The loading vectors correspond to the individual contribution of each variable. The first principal component typically explains most of the data variance. The generated plots offer a quick insight into exploring the data, providing a preliminary global overview of the metabolomic dataset. In addition, the PCA method provides simultaneous detection of a broad range of features to detect trends, groups or outliers Figure 2-12 [167]. Moreover, PCA analysis is often used to confirm the analytical system's index of stability and repeatability [9].

Sometimes, it is even used to suggest good instrumental precision and confirm correlation and similarities in the features [115]. PCA has been used in several studies to suggest or confirm good and satisfactory analytical results. PCA outputs are often applied for PLS-DA and OSC-PLS-DA models.
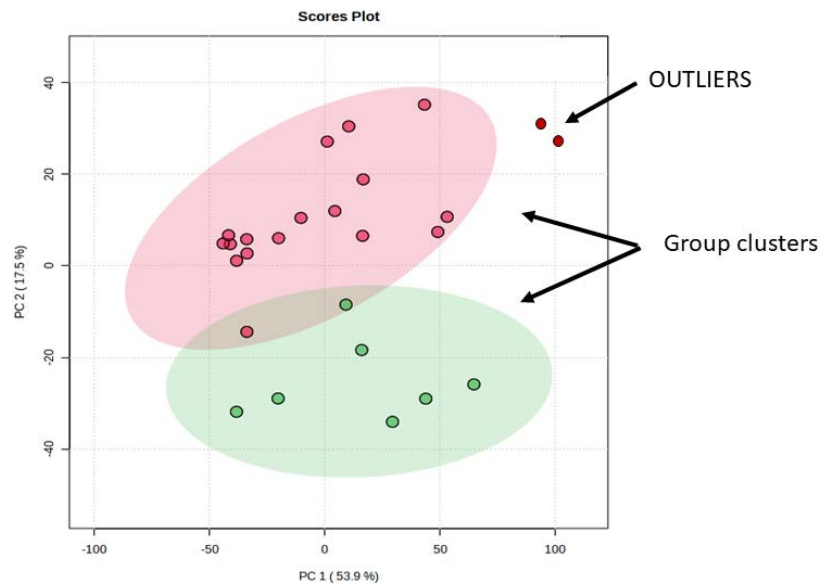
Figure 2-12: Example of the PCA score plot showing group segregation and sample outliers.

When no apparent clustering is possible, **hierarchical cluster analysis (HCA)** could be helpful. Another unsupervised tool for data visualisation and clustering, based principally on the Euclidean or Mahalanobis distance, this method measures the degree of membership of each sample for each of the subgroups, providing dendrograms [168] Figure 2-13. In most cases, the dendrogram is constructed using *Pearson's correlation* and average linkage algorithm to show a relationship between normal and diseased samples, thus, providing classification and differentiations between the patient groups [169]. The tree-like structure usually indicates which samples are joined in which cluster as a function of the distance between them. The major disadvantage is the difficulty interpreting data, especially when there are too many samples or variables.
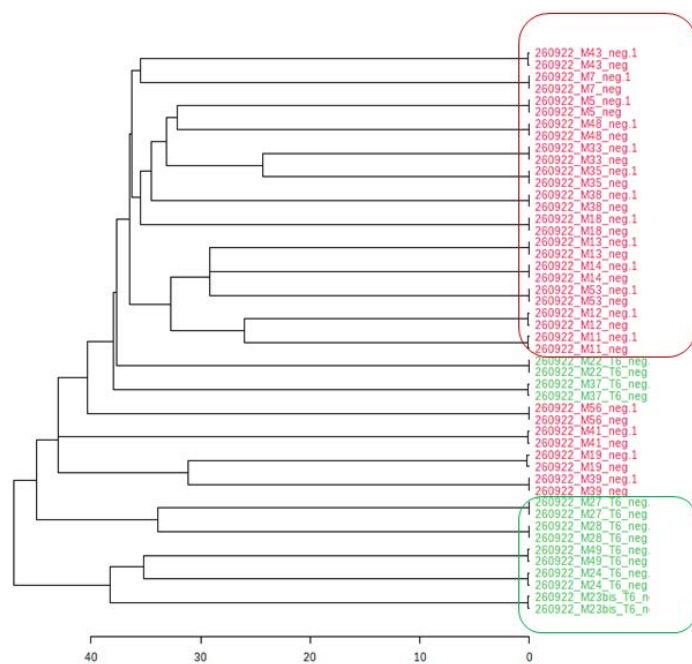
Figure 2-13: Example of how can appear hierarchical clustering analysis with representative groups clustering.

Usually, to interpret classification results, the receiver operating curves (ROC) validate the discriminating power of the compounds responsible for each classification. ROC curve analysis is widely considered the most objective and statistically valid method for biomarker performance evaluation, searching for a classifier that can maintain a high rate of true positives and a low rate of false positives. Thus, the area under the curve (AUC) with values (>70, *p-value* < 0.05) allowed the evaluation of the sensitivity and specificity of each compound to be considered as a relevant biomarker Figure 2-14. The higher the AUC value, the better the classification performance of the features in the specific classification. In addition, this method is helpful if it should be decided between different parts, thus, better performance and prediction.
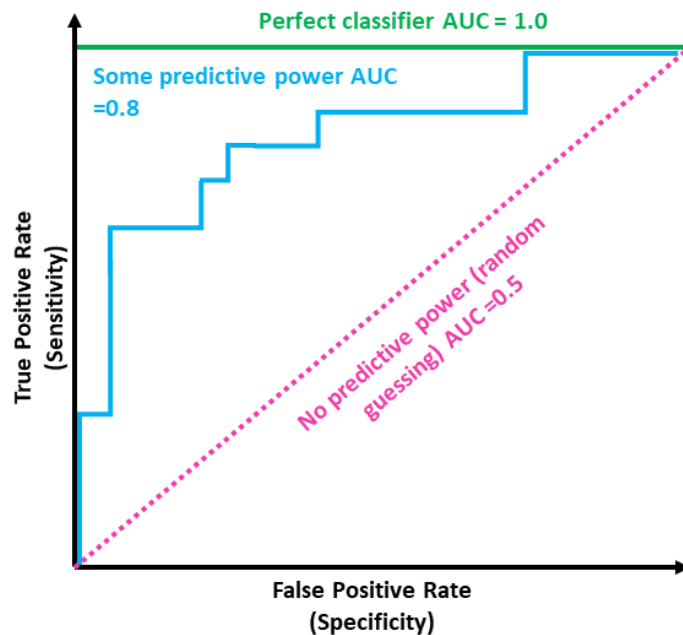
Figure 2-14: Schematic representation of AUC curves. The AUC lines represent the performance of possible biomarker candidates in binary classification. The pink traced line represents random guessing classifier. On the contrary, the green line indicates the perfect classifier with zero error rates, namely a 100% rate of true positive and a 0% rate of false positive. In most cases, all biomarkers are placed somewhere between these two lines. Therefore, the ''perfect'' biomarker should be placed close to the left-high corner, procuring more predictive power.

2.4.2.2.1.1 *Variable selection*

In this doctoral thesis, all metabolic profiling studies utilised the untargeted strategy for biomarkers detection, evaluating potential metabolites as best as possible without prior knowledge. Thus, the obtained data sets were characterised by containing an excessive amount of irrelevant information, specifically irrelevant variables, that could probably deteriorate and confuse the subsequent analyses. Therefore, a chemometric step, including variable selection, is imperative in these studies. Thus, variable selection methods perform contemporary: variable elimination to avoid overfitting and extracting important features from the data set. Numerous variable selection methods are proposed in the literature and are widely applied during the analysis using different analytical techniques. Two central problems must be considered: which criterion to use for optimisation of the number of variables; and which is the best strategy to operate according to the aim of the study [170]. Is the best method to search for the best subset by comparing all possible models through forward selection to find the single best

variable? Otherwise, is the strategy based on backward elimination to delete uninteresting variables one at a time?

A **<u>stepwise orthogonalisation of predictors (SELECT)</u>** method could be applied to the whole spectra dataset so that the information responsible for the successful discrimination between categories would be compressed into reduced variables subset, evaluating disease-specific signature/features/bands in samples. To assess significant metabolomic signatures in analysed samples, the SELECT method was widely applied in this doctoral thesis to develop a reliable classification that could discriminate between patients in different stages of disease progression.

Searching for the optimal combinations of variables to overcome very optimistic results from a prediction standpoint is a good practice for validating the predictors with an independent dataset. For the same purpose, variable selection methods are combined or followed through other strategies, which shall be discussed below.

2.4.2.2.2 Supervised techniques

While unsupervised techniques do not a priori support data knowledge, supervised methods generally do to predict, discriminate and classify new data. Generally, the combination of PCA and LDA has often been applied in plasma metabolomic studies [120].

Much like PCA, **<u>linear discriminant analysis (LDA)</u>** is a feature reduction method which maximises the ratio of between-class variance and minimises the percentage of within-class variance based on the Mahalanobis distance [172] or instead removes redundant and dependent features by transforming them from a higher dimensional space to a space with lower dimensions [173]. The terms ''classification'' or ''discrimination'' are widely applied for biomarker discovery in metabolomics studies. Thus, the LDA approach is used to identify a linear transformation that discriminates between different classes in the data set. LDA methods are broadly applied for patient stratification and differentiation, different biochemical signatures based on the degree of disease [174].

<u>SIMCA</u> (Soft independent modelling of class analogies) is a type of multivariate analysis that combines PCA and Discriminant analysis to classify data into different groups based on their properties [175,176]. Thus, classification methods such as LDA are

based on developing classification rules and delimiters between classes, whereas in class models, significance limits are built for the specified classes. These limits define the membership parameters for each class; thus, an unknown sample can be classified as not belonging to any defined categories because it is not included in any of its class spaces. Coomans's plots are built in SIMCA to better visualise group differentiation (Figure 2-15).
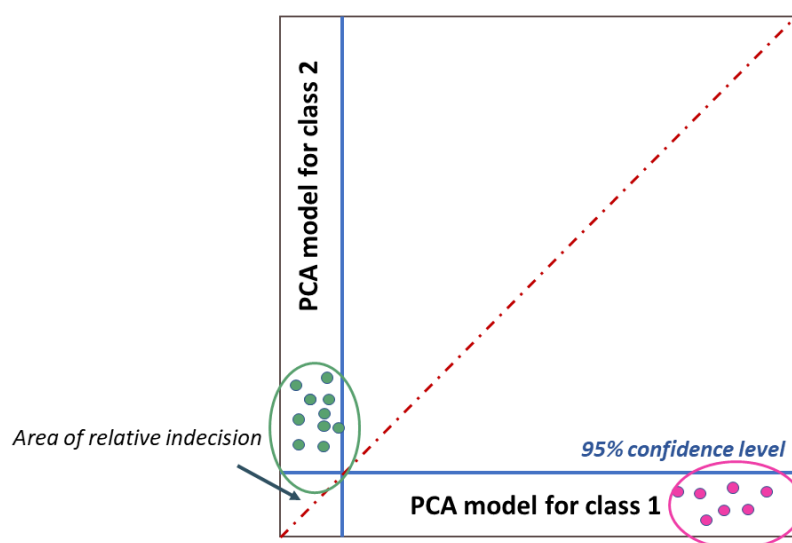
Figure 2-15. The two axes represent the first two principal components of the data. Each point on the plot represents an observation in the data set. The colour or shape of the point is used to indicate which group the observation belongs to. The location of a point on the plot indicates its similarity to other observations in the data set.

Another method based on the Euclidean and Mahalanobis distance is the **K nearest neighbours (KNN)** method. This method is much simpler than the one described above. A new sample would be classified by calculating the distance from each sampling training set; once a K nearest one is found, the unknown sample ranked in the group with the most members amongst these neighbours. This method has the advantage of not assuming the groups' shapes [177].

In terms of supervised methods, one of the most widely applied in chemometrics is **Partial Least Squares (PLS).** Similar in concept to PCA, this multivariate linear regression method finds the relations between predictors, **X**, and responses containing diverse sources of variation (data X and response Y), by maximising the covariance of their latent

variables, helping to understand which variables are more correlated to the response. PLS is a simple and flexible approach because it can handle incomplete and noisy data with many variables (and observations). Therefore, PLS regression and LDA make it an even more robust approach for analysing data for complicated problems such as biomarker discovery. Partial least squares discriminant analysis (PLS-DA) is another well-established chemometric method that projects the high dimensional data into a low dimensional space intending to capture the most data variance. Still, in this case, the projection direction is only computed on the X data without referring to the experimental condition Y [171]. Therefore, it should not be seen as a regression method but as a classification tool [178]. One advantage of PLS-DA lies in its ability to handle highly collinear data. Moreover, PLS-DA can provide excellent insights into the cause of discrimination by checking the behaviour of variables. Thus, while the separation between two groups could not be apparent in PCA plots, PLS-DA models are usually built to obtain maximum separation between the groups, revealing a significant cluster of metabolites as discriminative markers of patient status [8]. Therefore, **Q2** is an estimate of the model's predictive ability and is calculated via cross-validation (CV). Therefore, good predictions will have a high Q2. However, it is possible to have a negative Q2, meaning your model is not predictive or overfitted. There are two critical measures in PLS-DA for biomarker identification: variable importance in projection (VIP) and the weighted sum of absolute regression coefficients.

The recent modification of PLS-DA is the **Orthogonal partial least squares-discriminant analysis (OPLS-DA)** method, another multivariate linear method of choice in blood metabolomics. It provides a better interpretation of relevant variables than PLS by decomposing the data into so-called ''predictive'' information related to the response Y. Compared to PLS-DA, it explains which variables have the most excellent discriminatory power [84]. Therefore, it is often applied to identify metabolic biomarkers correlated with a continuous variable. Usually, Z-plots that combine the covariance and correlation loading profiles are used to visualise the variable influence in the OPLS-DA model. This corresponds to combining the contribution, or magnitude (covariance), with the effect and reliability (correlation) for the model variables concerning model component scores.

However, although PCA and OPLSDA are the gold standards for binary classification, these discriminant analyses are known to generate models that might overfit the data. On the other hand, machine learning approaches are more suited for analysing metabolomics data. Thus, **a random forest (RF)** is a machine-learning approach to select the best-performing features/compounds per pairwise comparison based on the lowest mean values for minimum depths in the trees (no. of trees = 100/1000/2000/5000) and the frequencies found in trees [179]. Minimum depth indicates how early in decision trees a possible biomarker is involved. Thus lower the minimum depth value, the better it is. On the contrary, higher frequencies at lower nodes indicate that some features/compounds effectively classify the different sample groups.

Heatmaps are used to visualise each technique's classification ability to visually visualise the data table. Thus, each coloured cell on the map corresponds to a concentration value in the data matrix. Usually, samples are displayed in rows and features/compounds in columns. This graphical method is often applied in liquid chromatography metabolomic studies analysis to identify samples or features that are unusually high/low.

2.4.2.2.3 Validation

However, one more step is required to ensure the statistical performance of chemometric models, namely validation. This procedure is necessary to obtain a measure of the error, specifically to evaluate the prediction error and the prediction power of the model. The proof of the model indicates the predictive ability of the model to classify samples that have not been used in the construction of the model. The most frequently used method creates two sample subsets: the training and the evaluation or test set. The regression model would be developed with the objects in the training set. The prediction error is evaluated on the things of the test set as the standard deviation of the prediction error on the test set [150]. Thus, once plasma metabolic patterns with potentially favourable characteristics to distinguish patient profiles have been found, further statistical validation is highly recommended. Considering that metabolomic studies often include a reduced number of samples, when compared with the number of variables, testing the classification ability of the model becomes complicated. An appropriate validation strategy should be applied to metabolomic data to avoid the

introduction of bias, generally based on cross-validation procedures, such as full cross-validation or leave-one-out cross-validation (LOOCV). In this last case, many models are built as samples are analysed; therefore, many interactions will be generated, allowing us to obtain models with more extraordinary predictive ability but with higher processing time. The level of validation and the validation method could confirm the results obtained and enable more in-depth large-scale research at a later stage. Furthermore, the proof of the model's classification ability could help cancel the clinical acceptance boundaries.

2.5 REFERENCES

1. Jan, S.; Ahmad, P. *Introducing Metabolomics*; 2019; ISBN 9780128148723.

2. Wang, W.; Liu, X.; Wu, J.; Kang, X.; Xie, Q.; Sheng, J.; Xu, W.; Liu, D.; Zheng, W. Plasma Metabolite Profiling Reveals Potential Biomarkers of Giant Cell Tumor of Bone by Using NMR-Based Metabolic Profiles: A Cross-Sectional Study. *Medicine (United States)* **2019**, *98*, doi:10.1097/MD.0000000000017445.

3. Nicholson, J.K. Metabonomics a Generic Platform for the Investigation of Drug Toxicity , Altered Gene Function and Disease Processes. *Drug Discovery World Summer* **2004**, 23–28.

4. HMDB Human Metabolome Database.

5. Hertel, J.; Harms, A.C.; Heinken, A.; Baldini, F.; Thinnes, C.C.; Glaab, E.; Vasco, D.A.; Pietzner, M.; Stewart, I.D.; Wareham, N.J.; et al. Integrated Analyses of Microbiome and Longitudinal Metabolome Data Reveal Microbial-Host Interactions on Sulfur Metabolism in Parkinson's Disease. *Cell Rep* **2019**, *29*, 1767--1777.e8, doi:10.1016/j.celrep.2019.10.035.

6. García-Pérez, I.; Vallejo, M.; García, A.; Legido-Quigley, C.; Barbas, C. Metabolic Fingerprinting with Capillary Electrophoresis. *J Chromatogr A* **2008**, *1204*, 130–139, doi:10.1016/j.chroma.2008.07.025.

7. Hertel, J.; Harms, A.C.; Heinken, A.; Baldini, F.; Thinnes, C.C.; Glaab, E.; Vasco, D.A.; Pietzner, M.; Stewart, I.D.; Wareham, N.J.; et al. Integrated Analyses of Microbiome and Longitudinal Metabolome Data Reveal Microbial-Host Interactions on Sulfur Metabolism in Parkinson's Disease. *Cell Rep* **2019**, *29*, 1767-1777.e8, doi:10.1016/j.celrep.2019.10.035.

8. Liu, R.; Chou, J.; Hou, S.; Liu, X.; Yu, J.; Zhao, X.; Li, Y.; Liu, L.; Sun, C. Evaluation of Two-Step Liquid-Liquid Extraction Protocol for Untargeted Metabolic Profiling of Serum Samples to Achieve Broader Metabolome Coverage by UPLC-Q-TOF-MS. *Anal Chim Acta* **2018**, *1035*, 96–107, doi:10.1016/j.aca.2018.07.034.

9. Fanos, V.; Noto, A.; Caboni, P.; Pintus, M.C.; Liori, B.; Dessì, A.; Mussap, M. Urine Metabolomic Profiling in Neonatal Nephrology. *Clin Biochem* **2014**, *47*, 708–710, doi:10.1016/j.clinbiochem.2014.05.020.

10. Shao, Y.; Le, W. Recent Advances and Perspectives of Metabolomics-Based Investigations in Parkinson's Disease. *Mol Neurodegener* **2019**, *14*, 1–12, doi:10.1186/s13024-018-0304-2.

11. Elmonem, M.A.; Abdelazim, A.M. Novel Biomarkers for Lysosomal Storage Disorders: Metabolomic and Proteomic Approaches. *Clinica Chimica Acta* **2020**, *509*, 195–209, doi:10.1016/j.cca.2020.06.028.

12. Orczyk-Pawilowicz, M.; Jawien, E.; Deja, S.; Hirnle, L.; Zabek, A.; Mlynarz, P. Metabolomics of Human Amniotic Fluid and Maternal Plasma during Normal Pregnancy. *PLoS One* **2016**, *11*, doi:10.1371/journal.pone.0152740.

13. ELSTON, C.W.; ELLIS, I.O. Pathological Prognostic Factors in Breast Cancer. I. The Value of Histological Grade in Breast Cancer: Experience from a Large Study with Long-term Follow-up. *Histopathology* **1991**, doi:10.1111/j.1365-2559.1991.tb00229.x.

14. Junker, S.; Krumbholz, G.; Frommer, K.W.; Rehart, S.; Steinmeyer, J.; Rickert, M.; Schett, G.; Müller-Ladner, U.; Neumann, E. Differentiation of Osteophyte Types in Osteoarthritis – Proposal of a Histological Classification. *Joint Bone Spine* **2016**, doi:10.1016/j.jbspin.2015.04.008.

15. Cho, K.R.; Shih, I.M. Ovarian Cancer. *Annual Review of Pathology: Mechanisms of Disease* 2009.

16. Goldman, J.G.; Andrews, H.; Amara, A.; Naito, A.; Alcalay, R.N.; Shaw, L.M.; Taylor, P.; Xie, T.; Tuite, P.; Henchcliffe, C.; et al. Cerebrospinal Fluid, Plasma, and Saliva in the BioFIND Study: Relationships among Biomarkers and Parkinson's Disease Features. *Movement Disorders* **2018**, *33*, 282–288, doi:10.1002/mds.27232.

17. Gouveia, F.; Bicker, J.; Gonçalves, J.; Alves, G.; Falcão, A.; Fortuna, A. Liquid Chromatographic Methods for the Determination of Direct Oral Anticoagulant Drugs in Biological Samples: A Critical Review. *Anal Chim Acta* **2019**, *1076*, 18–31, doi:10.1016/j.aca.2019.03.061.

18. Rombouts, C.; de Spiegeleer, M.; van Meulebroek, L.; de Vos, W.H.; Vanhaecke, L. Validated Comprehensive Metabolomics and Lipidomics Analysis of Colon Tissue and Cell Lines. *Anal Chim Acta* **2019**, *1066*, 79–92, doi:10.1016/j.aca.2019.03.020.

19. Massey, K.A.; Nicolaou, A. Lipidomics of Oxidized Polyunsaturated Fatty Acids. *Free Radic Biol Med* **2013**, *59*, 45–55, doi:10.1016/j.freeradbiomed.2012.08.565.

20. Quehenberger, O.; Armando, A.M.; Brown, A.H.; Milne, S.B.; Myers, D.S.; Merrill, A.H.; Bandyopadhyay, S.; Jones, K.N.; Kelly, S.; Shaner, R.L.; et al. Lipidomics Reveals a Remarkable Diversity of Lipids in Human Plasma1. *J Lipid Res* **2010**, *51*, 3299–3305, doi:10.1194/jlr.M009449.

21. Peña-Bautista, C.; Álvarez-Sánchez, L.; Roca, M.; García-Vallés, L.; Baquero, M.; Cháfer-Pericás, C. Plasma Lipidomics Approach in Early and Specific Alzheimer's Disease Diagnosis. *J Clin Med* **2022**, *11*, doi:10.3390/jcm11175030.

22. Xu, T.; Hu, C.; Xuan, Q.; Xu, G. Recent Advances in Analytical Strategies for Mass Spectrometry-Based Lipidomics. *Anal Chim Acta* **2020**, *1137*, 156–169, doi:10.1016/J.ACA.2020.09.060.

23. Chen, J.; Liu, C.; Ye, S.; Lu, R.; Zhu, H.; Xu, J. UPLC-MS/MS-based Plasma Lipidomics Reveal a Distinctive Signature in Systemic Lupus Erythematosus Patients. *MedComm (Beijing)* **2021**, *2*, 269, doi:10.1002/MCO2.67.

24. Castellanos, D.B.; Martín-Jiménez, C.A.; Rojas-Rodríguez, F.; Barreto, G.E.; González, J. Brain Lipidomics as a Rising Field in Neurodegenerative Contexts: Perspectives with Machine Learning Approaches. *Front Neuroendocrinol* 2021, *61*.

25. Fahy, E.; Subramaniam, S.; Brown, H.A.; Glass, C.K.; Merrill, A.H.; Murphy, R.C.; Raetz, C.R.H.; Russell, D.W.; Seyama, Y.; Shaw, W.; et al. A Comprehensive Classification System for Lipids. *J Lipid Res* **2005**, *46*, 839–861, doi:10.1194/jlr.E400004-JLR200.

26. Fahy, E.; Subramaniam, S.; Murphy, R.C.; Nishijima, M.; Raetz, C.R.H.; Shimizu, T.; Spener, F.; van Meer, G.; Wakelam, M.J.O.; Dennis, E.A. Update of the LIPID MAPS Comprehensive Classification System for Lipids. *J Lipid Res* **2009**, *50*, doi:10.1194/jlr.R800095-JLR200.

27. Leishman, E.; Kunkler, P.E.; Hurley, J.H.; Miller, S.; Bradshaw, H.B. *Bioactive Lipids in Cancer, Inflammation and Related Diseases: Acute and Chronic Mild Traumatic Brain Injury Differentially Changes Levels of Bioactive Lipids in the CNS Associated with Headache*; 2019; Vol. 1161; ISBN 9783030216368.

28. Honn, K. v; Zeldin, D.C. *Advances in Experimental Medicine and Biology 1161 The Role of Bioactive Lipids in Cancer, Infl Ammation and Related Diseases*;

29. Haslam, R.P.; Feussner, I. Green Light for Lipid Fingerprinting. *Biochim Biophys Acta Mol Cell Biol Lipids* **2017**, *1862*, 782–785, doi:10.1016/j.bbalip.2017.04.005.

30. Bestard-Escalas, J.; Garate, J.; Maimó-Barceló, A.; Fernández, R.; Lopez, D.H.; Lage, S.; Reigada, R.; Khorrami, S.; Ginard, D.; Reyes, J.; et al. Lipid Fingerprint Image Accurately Conveys Human Colon Cell Pathophysiologic State: A Solid Candidate as Biomarker. *Biochim Biophys Acta Mol Cell Biol Lipids* **2016**, *1861*, 1942–1950, doi:10.1016/j.bbalip.2016.09.013.

31. Vanherle, S.; Haidar, M.; Irobi, J.; Bogie, J.F.J.; Hendriks, J.J.A. Extracellular Vesicle-Associated Lipids in Central Nervous System Disorders. *Adv Drug Deliv Rev* 2020, *159*, 322–331.

32. Schmitt, F.; Hussain, G.; Dupuis, L.; Loeffler, J.P.; Henriques, A. A Plural Role for Lipids in Motor Neuron Diseases: Energy, Signaling and Structure. *Front Cell Neurosci* 2014, *8*.

33. Lang, F. Mechanisms and Significance of Cell Volume Regulation. *J Am Coll Nutr* **2007**, *26*, 613S-623S, doi:10.1080/07315724.2007.10719667.

34. Lingwood, D.; Simons, K. Lipid Rafts as a Membrane-Organizing Principle. *Science (1979)* 2010, *327*, 46–50.

35. Wenk, M.R. Lipidomics: New Tools and Applications. *Cell* 2010, *143*, 888–895.

36. Arsenault, D.; Julien, C.; Tremblay, C.; Calon, F. DHA Improves Cognition and Prevents Dysfunction of Entorhinal Cortex Neurons in 3xTg-AD Mice. *PLoS One* **2011**, *6*, doi:10.1371/journal.pone.0017397.

37. Gnanalingham, K.K.; Byrne, E.J.; Thornton, A.; Sambrook, M.A.; Bannister, P. Motor and Cognitive Function in Lewy Body Dementia: Comparison with Alzheimer's and Parkinson's Diseases. *J Neurol Neurosurg Psychiatry* **1997**, *62*, 243–252, doi:10.1136/jnnp.62.3.243.

38. Chiti, F.; Dobson, C.M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress over the Last Decade. *Annu Rev Biochem* **2017**, *86*, 27–68, doi:10.1146/ANNUREV-BIOCHEM-061516-045115.

39. Martial, B.; Lefèvre, T.; Auger, M. Understanding Amyloid Fibril Formation Using Protein Fragments: Structural Investigations via Vibrational Spectroscopy and Solid-State NMR. *Biophys Rev* **2018**, *10*, 1133–1149, doi:10.1007/s12551-018-0427-2.

40. Meade, R.M.; Fairlie, D.P.; Mason, J.M. Alpha-Synuclein Structure and Parkinson's Disease. *Mol Neurodegener* **2019**, *14*, 1–14.

41. Fayyad, M.; Salim, S.; Majbour, N.; Erskine, D.; Stoops, E.; Mollenhauer, B.; El-Agnaf, O.M.A. Parkinson's Disease Biomarkers Based on α-Synuclein. *J Neurochem* **2019**, *150*, 626–636, doi:10.1111/JNC.14809.

42. Cutler, R.G.; Pedersen, W.A.; Camandola, S.; Rothstein, J.D.; Mattson, M.P. Evidence That Accumulation of Ceramides and Cholesterol Esters Mediates Oxidative Stress - Induced Death of Motor Neurons in Amyotrophic Lateral Sclerosis. *Ann Neurol* **2002**, *52*, 448–457, doi:10.1002/ana.10312.

43. Mili, N.; Paschou, S.A.; Goulis, D.G.; Dimopoulos, M.-A.; Lambrinoudaki, I.; Psaltopoulou, T. Obesity, Metabolic Syndrome, and Cancer: Pathophysiological and Therapeutic Associations. *Endocrine* **2021**, *74*, 478–497, doi:10.1007/s12020-021-02884-x.

44. Niwa, K. Metabolic Syndrome and Coronary Artery Disease in Adults with Congenital Heart Disease. *Cardiovasc Diagn Ther* **2021**, *11*, 563–576, doi:10.21037/cdt-20-781.

45. Ho, J.E.; Larson, M.G.; Ghorbani, A.; Cheng, S.; Chen, M.H.; Keyes, M.; Rhee, E.P.; Clish, C.B.; Vasan, R.S.; Gerszten, R.E.; et al. Metabolomic Profiles of Body Mass Index in the Framingham Heart Study Reveal Distinct Cardiometabolic Phenotypes. *PLoS One* **2016**, doi:10.1371/journal.pone.0148361.

46. Li, C.; Zhang, Z.; Peng, Y.; Gao, H.; Wang, Y.; Zhao, J.; Pan, C. Plasma Neutrophil Gelatinase-Associated Lipocalin Levels Are Associated with the Presence and Severity of Coronary Heart Disease. *PLoS One* **2019**, *14*, e0220841, doi:10.1371/JOURNAL.PONE.0220841.

47. Esposito, K.; Chiodini, P.; Capuano, A.; Bellastella, G.; Maiorino, M.I.; Giugliano, D. Metabolic Syndrome and Endometrial Cancer: A Meta-Analysis. *Endocrine* **2014**, *45*, 28–36, doi:10.1007/s12020-013-9973-3.

48. Esposito, K.; Chiodini, P.; Colao, A.; Lenzi, A.; Giugliano, D. Metabolic Syndrome and Risk of Cancer: A Systematic Review and Meta-Analysis. *Diabetes Care* **2012**, *35*, 2402–2411, doi:10.2337/DC12-0336.

49. Renehan, A.G.; Tyson, M.; Egger, M.; Heller, R.F.; Zwahlen, M. Body-Mass Index and Incidence of Cancer: A Systematic Review and Meta-Analysis of Prospective Observational Studies. *The Lancet* **2008**, *371*, 569–578, doi:10.1016/S0140-6736(08)60269-X.

50. Ros-Mazurczyk, M.; Jelonek, K.; Marczyk, M.; Binczyk, F.; Pietrowska, M.; Polanska, J.; Dziadziuszko, R.; Jassem, J.; Rzyman, W.; Widlak, P. Serum Lipid Profile Discriminates Patients with Early Lung Cancer from Healthy Controls. *Lung Cancer* **2017**, *112*, 69–74, doi:10.1016/j.lungcan.2017.07.036.

51. Munir, R.; Usman, H.; Hasnain, S.; Smans, K.; Kalbacher, H.; Zaidi, N. Atypical Plasma Lipid Profile in Cancer Patients: Cause or Consequence? *Biochimie* **2014**, *102*, 9–18, doi:10.1016/j.biochi.2014.03.010.

52. Sussulini, A. Erratum to: Chapters 1 and 11 of Metabolomics: From Fundamentals to Clinical Applications. *Adv Exp Med Biol* 2017.

53. Serkova, N.J.; Standiford, T.J.; Stringer, K.A. The Emerging Field of Quantitative Blood Metabolomics for Biomarker Discovery in Critical Illnesses. *Am J Respir Crit Care Med* 2011.

54. Medina, S.; Ferreres, F.; García-Viguera, C.; Horcajada, M.N.; Orduna, J.; Savirón, M.; Zurek, G.; Martínez-Sanz, J.M.; Gil, J.I.; Gil-Izquierdo, A. Non-Targeted Metabolomic Approach Reveals Urinary Metabolites Linked to Steroid Biosynthesis Pathway after Ingestion of Citrus Juice. *Food Chem* **2013**, *136*, 938–946, doi:10.1016/j.foodchem.2012.09.004.

55. Reisz, J.A.; Zheng, C.; D'Alessandro, A.; Nemkov, T. Untargeted and Semi-Targeted Lipid Analysis of Biological Samples Using Mass Spectrometry-Based Metabolomics. In *Methods in Molecular Biology*; Humana Press Inc., 2019; Vol. 1978, pp. 121–135.

56. Savolainen, O.I.; Sandberg, A.-S.; Ross, A.B. A Simultaneous Metabolic Profiling and Quantitative Multimetabolite Metabolomic Method for Human Plasma Using Gas-Chromatography Tandem Mass Spectrometry. **2015**, doi:10.1021/acs.jproteome.5b00790.

57. Bury, D.; Morais, C.L.M.; Paraskevaidi, M.; Ashton, K.M.; Dawson, T.P.; Martin, F.L. Spectral Classification for Diagnosis Involving Numerous Pathologies in a Complex Clinical Setting: A Neuro-Oncology Example. *Spectrochim Acta A Mol Biomol Spectrosc* **2019**, *206*, 89–96, doi:10.1016/j.saa.2018.07.078.

58. Wikoff, W.R.; Anfora, A.T.; Liu, J.; Schultz, P.G.; Lesley, S.A.; Peters, E.C.; Siuzdak, G. Metabolomics Analysis Reveals Large Effects of Gut Microflora on Mammalian Blood Metabolites. *Proc Natl Acad Sci U S A* **2009**, *106*, 3698–3703, doi:10.1073/pnas.0812874106.

59. Bicalho, B.; Holovati, J.L.; Acker, J.P. Phospholipidomics Reveals Differences in Glycerophosphoserine Profiles of Hypothermically Stored Red Blood Cells and Microvesicles. *Biochim Biophys Acta Biomembr* **2013**, *1828*, 317–326, doi:10.1016/j.bbamem.2012.10.026.

60. Investigation of Human Blood Plasma SamplePreparation for Performing Metabolomics UsingUltrahigh Performance Liquid ChromatographyMass Spectrometry.

61. Li, J.; Gu, C.; Zhu, M.; Li, D.; Chen, L.; Zhu, X. Correlations between Blood Lipid, Serum Cystatin C, and Homocysteine Levels in Patients with Parkinson's Disease. *Psychogeriatrics* 2020, *20*, 180–188.

62. Setoyama, D.; Yoshino, A.; Takamura, M.; Okada, G.; Iwata, M.; Tsunetomi, K.; Ohgidani, M.; Kuwano, N.; Yoshimoto, J.; Okamoto, Y.; et al. Personality Classification Enhances Blood Metabolome Analysis and Biotyping for Major Depressive Disorders: Two-Species Investigation. *J Affect Disord* **2021**, *279*, 20–30, doi:10.1016/j.jad.2020.09.118.

63. Bruce, S.J.; Tavazzi, I.; Parisod, V.; Rezzi, S.; Kochhar, S.; Guy, P.A. Investigation of Human Blood Plasma Sample Preparation for Performing Metabolomics Using Ultrahigh Performance Liquid Chromatography/Mass Spectrometry. *Anal Chem* **2009**, doi:10.1021/ac8024569.

64. Bernini, P.; Bertini, I.; Luchinat, C.; Nincheri, P.; Staderini, S.; Turano, P. Standard Operating Procedures for Pre-Analytical Handling of Blood and Urine for Metabolomic Studies and Biobanks. *J Biomol NMR* **2011**, *49*, 231–243, doi:10.1007/s10858-011-9489-1.

65. Jim Enez-Jim Enez, F.J.; Alonso-Navarro, H.; Garc Ia-Mart, E.; Ag Undez, J.A.G. Cerebrospinal and Blood Levels of Amino Acids as Potential Biomarkers for Parkinson's Disease: Review and Meta-Analysis. *Eur J Neurol* **2020**, *2020*, 2336–2347, doi:10.1111/ene.14470.

66. Greco, A.; Chiesa, M.R.; da Prato, I.; Romanelli, A.M.; Dolciotti, C.; Cavallini, G.; Masciandaro, S.M.; Scilingo, E.P.; del Carratore, R.; Bongioanni, P. Using Blood Data for the Differential Diagnosis and Prognosis of Motor Neuron Diseases: A New Dataset for Machine Learning Applications. *Sci Rep* **2021**, *11*.

67. Nikanjam, M.; Kato, S.; Kurzrock, R. Liquid Biopsy: Current Technology and Clinical Applications. *J Hematol Oncol* 2022, *15*.

68. de Paepe, E.; van Meulebroek, L.; Rombouts, C.; Huysman, S.; Verplanken, K.; Lapauw, B.; Wauters, J.; Hemeryck, L.Y.; Vanhaecke, L. A Validated Multi-Matrix Platform for Metabolomic Fingerprinting of Human Urine, Feces and Plasma Using Ultra-High Performance Liquid-Chromatography Coupled to Hybrid Orbitrap High-Resolution Mass Spectrometry. *Anal Chim Acta* **2018**, doi:10.1016/j.aca.2018.06.065.

69. Castillo-Peinado, L.S.; Luque de Castro, M.D. Present and Foreseeable Future of Metabolomics in Forensic Analysis. *Anal Chim Acta* 2016.

70. Kaluarachchi, M.; Boulangé, C.L.; Karaman, I.; Lindon, J.C.; Ebbels, T.M.D.; Elliott, P.; Tracy, R.P.; Olson, N.C. A Comparison of Human Serum and Plasma Metabolites Using Untargeted 1 H NMR Spectroscopy and UPLC-MS. *Metabolomics* **2018**, *14*, doi:10.1007/s11306-018-1332-1.

71. David, A.; Abdul-Sada, A.; Lange, A.; Tyler, C.R.; Hill, E.M. A New Approach for Plasma (Xeno)Metabolomics Based on Solid-Phase Extraction and Nanoflow Liquid Chromatography-Nanoelectrospray Ionisation Mass Spectrometry. *J Chromatogr A* **2014**, *1365*, 72–85, doi:10.1016/j.chroma.2014.09.001.

72. Dudzik, D.; Zorawski, M.; Skotnicki, M.; Zarzycki, W.; Kozlowska, G.; Bibik-Malinowska, K.; Vallejo, M.; García, A.; Barbas, C.; Ramos, M.P. Metabolic Fingerprint of Gestational Diabetes Mellitus. *J Proteomics* **2014**, *103*, 57–71, doi:10.1016/j.jprot.2014.03.025.

73. Rodrigues, D.; Monteiro, M.; Jerónimo, C.; Henrique, R.; Belo, L.; Bastos, M. de L.; Guedes de Pinho, P.; Carvalho, M. Renal Cell Carcinoma: A Critical Analysis of Metabolomic Biomarkers Emerging from Current Model Systems. *Translational Research* **2017**, *180*, 1–11, doi:10.1016/j.trsl.2016.07.018.

74. Maitra, I.; Morais, C.L.M.; Lima, K.M.G.; Ashton, K.M.; Date, R.S.; Martin, F.L. Attenuated Total Reflection Fourier-Transform Infrared Spectral Discrimination in Human Bodily Fluids of Oesophageal Transformation to Adenocarcinoma. *Analyst* **2019**, *144*, 7447–7456, doi:10.1039/c9an01749f.

75. Suman, S.; Sharma, R.K.; Kumar, V.; Sinha, N.; Shukla, Y. Metabolic Fingerprinting in Breast Cancer Stages through 1H NMR Spectroscopy-Based Metabolomic Analysis of Plasma. *J Pharm Biomed Anal* **2018**, *160*, 38–45, doi:10.1016/j.jpba.2018.07.024.

76. Kamphorst, J.J.; Lewis, I.A. Editorial Overview: Recent Innovations in the Metabolomics Revolution. *Curr Opin Biotechnol* **2017**, *43*, iv–vii, doi:10.1016/j.copbio.2017.01.005.

77. Li, Y.; Shen, Y.; Yao, C. liang; Guo, D. an Quality Assessment of Herbal Medicines Based on Chemical Fingerprints Combined with Chemometrics Approach: A Review. *J Pharm Biomed Anal* **2020**, *185*, 113215, doi:10.1016/j.jpba.2020.113215.

78. Emwas, A.H.; Roy, R.; McKay, R.T.; Tenori, L.; Saccenti, E.; Nagana Gowda, G.A.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; et al. Nmr Spectroscopy for Metabolomics Research. *Metabolites* **2019**, *9*, doi:10.3390/metabo9070123.

79. Türker-Kaya, S.; Huck, C.W. A Review of Mid-Infrared and near-Infrared Imaging: Principles, Concepts and Applications in Plant Tissue Analysis. *Molecules* **2017**, *22*, doi:10.3390/molecules22010168.

80. Walsh, M.J.; Reddy, R.K.; Bhargava, R.; Member, A. Label-Free Biomedical Imaging With Mid-IR Spectroscopy. **2012**, *18*, 1502–1513.

81. Vinaixa, M.; Schymanski, E.L.; Neumann, S.; Navarro, M.; Salek, R.M.; Yanes, O. Mass Spectral Databases for LC/MS- and GC/MS-Based Metabolomics: State of the Field and Future Prospects. *TrAC - Trends in Analytical Chemistry* **2016**, *78*, 23–35, doi:10.1016/j.trac.2015.09.005.

82. Alonso, A.; Marsal, S.; Julià, A. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Front Bioeng Biotechnol* 2015, *3*.

83. Barth, A. Infrared Spectroscopy of Proteins. *Biochim Biophys Acta Bioenerg* **2007**, *1767*, 1073–1101, doi:10.1016/j.bbabio.2007.06.004.

84. Pasikanti, K.K.; Ho, P.C.; Chan, E.C.Y. Gas Chromatography/Mass Spectrometry in Metabolic Profiling of Biological Fluids. *J Chromatogr B Analyt Technol Biomed Life Sci* 2008, *871*, 202–211.

85. Mitchell, A.L.; Gajjar, K.B.; Theophilou, G.; Martin, F.L.; Martin-Hirsch, P.L. Vibrational Spectroscopy of Biofluids for Disease Screening or Diagnosis: Translation from the Laboratory to a Clinical Setting. *J Biophotonics* **2014**, *7*, 153–165, doi:10.1002/jbio.201400018.

86. Dunn, W.B.; Bailey, N.J.C.; Johnson, H.E. Measuring the Metabolome: Current Analytical Technologies. *Analyst* 2005, *130*, 606–625.

87. Kelly, R.S.; Bayne, H.; Spiro, A.; Vokonas, P.; Sparrow, D.; Weiss, S.T.; Schwartz, J.; Nassan, F.L.; Lee-Sarwar, K.; Huang, M.; et al. Metabolomic Signatures of Lead Exposure in the VA Normative Aging Study. *Environ Res* **2020**, *190*, 110022, doi:10.1016/j.envres.2020.110022.

88. Frenzel, S.; Wittfeld, K.; Habes, M.; Klinger-König, J.; Bülow, R.; Völzke, H.; Grabe, H.J. A Biomarker for Alzheimer's Disease Based on Patterns of Regional Brain Atrophy. *Front Psychiatry* **2020**, *10*, 1–9, doi:10.3389/fpsyt.2019.00953.

89. Carmona, P.; Molina, M.; Calero, M.; Bermejo-Pareja, F.; Martínez-Martín, P.; Toledano, A. Discrimination Analysis of Blood Plasma Associated with Alzheimer's Disease Using Vibrational Spectroscopy. *Journal of Alzheimer's Disease* **2013**, doi:10.3233/JAD-122041.

90. Deda, O.; Gika, H.; Raikos, N.; Theodoridis, G. *GC-MS-Based Metabolic Phenotyping*; Elsevier Inc., 2018; ISBN 9780128122938.

91. Nagajyothi, S.; Swetha, Y.; Neeharika, J.; Suresh, P. V; Ramarao, N. *Hyphenated Techniques-A Comprehensive Review*; 2017;

92. Gowda, G.A.N.; Raftery, D. *NMR-Based Metabolomics Methods and Protocols Methods in Molecular Biology 2037*;

93. Silva, R.A.; Pereira, T.C.S.; Souza, A.R.; Ribeiro, P.R. 1H NMR-Based Metabolite Profiling for Biomarker Identification. *Clinica Chimica Acta* 2020, *502*.

94. Emwas, A.H.; Roy, R.; McKay, R.T.; Ryan, D.; Brennan, L.; Tenori, L.; Luchinat, C.; Gao, X.; Zeri, A.C.; Gowda, G.A.N.; et al. Recommendations and Standardization of Biomarker Quantification Using NMR-Based Metabolomics with Particular Focus on Urinary Analysis. *J Proteome Res* 2016.

95. Garcia, A.; Barbas, C. Gas Chromatography-Mass Spectrometry (GC-MS)-Based Metabolomics. *Methods Mol Biol* **2011**, doi:10.1007/978-1-61737-985-7_11.

96. Seeley, J. v.; Seeley, S.K. Multidimensional Gas Chromatography: Fundamental Advances and New Applications. *Anal Chem* **2013**, *85*, 557–578, doi:10.1021/ac303195u.

97. Fiehn, O. Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Curr Protoc Mol Biol* **2016**, doi:10.1002/0471142727.mb3004s114.

98. Chan, Q.; Caruso, J.A. Plasma-Based Gas Chromatography Detectors. In *Gas Chromatography*; 2012 ISBN 9780123855404.

99. Ramautar, R. Capillary Electrophoresis-Mass Spectrometry for Clinical Metabolomics. In *Advances in Clinical Chemistry*; 2016 ISBN 9780128046890.

100. Zhang, W.; Segers, K.; Mangelings, D.; van Eeckhaut, A.; Hankemeier, T.; vander Heyden, Y.; Ramautar, R. Assessing the Suitability of Capillary Electrophoresis-Mass Spectrometry for Biomarker Discovery in Plasma-Based Metabolomics. *Electrophoresis* **2019**, *40*, 2309–2320, doi:10.1002/elps.201900126.

101. Zhang, W.; Hankemeier, T.; Ramautar, R. Next-Generation Capillary Electrophoresis–Mass Spectrometry Approaches in Metabolomics. *Curr Opin Biotechnol* **2017**, *43*, 1–7, doi:10.1016/j.copbio.2016.07.002.

102. Ramautar, R.; Demirci, A.; Jong, G.J. de Capillary Electrophoresis in Metabolomics. *TrAC - Trends in Analytical Chemistry* **2006**, doi:10.1016/j.trac.2006.02.004.

103. Sugimoto, M.; Wong, D.T.; Hirayama, A.; Soga, T.; Tomita, M. Capillary Electrophoresis Mass Spectrometry-Based Saliva Metabolomics Identified Oral, Breast and Pancreatic Cancer-Specific Profiles. *Metabolomics* **2010**, doi:10.1007/s11306-009-0178-y.

104. Ramautar, R.; Somsen, G.W.; de Jong, G.J. CE-MS for Metabolomics: Developments and Applications in the Period 2016–2018. *Electrophoresis* 2019.

105. Chen, C.; Gonzalez, F.J.; Idle, J.R. LC-MS-BASED METABOLOMICS IN DRUG METABOLISM.

106. Cui, L.; Lu, H.; Lee, Y.H. Challenges and Emergent Solutions for LC-MS/MS Based Untargeted Metabolomics in Diseases. *Mass Spectrom Rev* 2018, *37*, 772–792.

107. Li, J.; Che, N.; Xu, L.; Zhang, Q.; Wang, Q.; Tan, W.; Zhang, M. LC-MS-Based Serum Metabolomics Reveals a Distinctive Signature in Patients with Rheumatoid Arthritis. *Clinical Rheumatology 2018 37:6* **2018**, *37*, 1493–1502, doi:10.1007/S10067-018-4021-6.

108. Becker, S.; Kortz, L.; Helmschrodt, C.; Thiery, J.; Ceglarek, U. LC-MS-Based Metabolomics in the Clinical Laboratory. *J Chromatogr B Analyt Technol Biomed Life Sci* 2012, *883–884*, 68–75.

109. Guan, S.; Jia, B.; Chao, K.; Zhu, X.; Tang, J.; Li, M.; Wu, L.; Xing, L.; Liu, K.; Zhang, L.; et al. UPLC-QTOF-MS-Based Plasma Lipidomic Profiling Reveals Biomarkers for Inflammatory Bowel Disease Diagnosis. *J Proteome Res* **2020**, *19*, 600–609, doi:10.1021/acs.jproteome.9b00440.

110. Cao, Y.; He, K.; Cheng, M.; Si, H.Y.; Zhang, H.L.; Song, W.; Li, A.L.; Hu, C.J.; Wang, N. Two Classifiers Based on Serum Peptide Pattern for Prediction of HBV-Induced Liver Cirrhosis Using MALDI-TOF MS. *Biomed Res Int* **2013**, doi:10.1155/2013/814876.

111. Zhao, Y.Y.; Cheng, X.L.; Vaziri, N.D.; Liu, S.; Lin, R.C. UPLC-Based Metabonomic Applications for Discovering Biomarkers of Diseases in Clinical Chemistry. *Clin Biochem* **2015**, *47*, 16–26, doi:10.1016/j.clinbiochem.2014.07.019.

112. Gika, H.G.; Theodoridis, G.A.; Plumb, R.S.; Wilson, I.D. Current Practice of Liquid Chromatography-Mass Spectrometry in Metabolomics and Metabonomics. *J Pharm Biomed Anal* **2014**, *87*, 12–25, doi:10.1016/j.jpba.2013.06.032.

113. Knittelfelder, O.L.; Weberhofer, B.P.; Eichmann, T.O.; Kohlwein, S.D.; Rechberger, G.N. A Versatile Ultra-High Performance LC-MS Method for Lipid Profiling. *J Chromatogr B Analyt Technol Biomed Life Sci* **2014**, *951–952*, 119–128, doi:10.1016/j.jchromb.2014.01.011.

114. Becker, S.; Kortz, L.; Helmschrodt, C.; Thiery, J.; Ceglarek, U. LC-MS-Based Metabolomics in the Clinical Laboratory. *J Chromatogr B Analyt Technol Biomed Life Sci* 2012, *883–884*, 68–75.

115. De Paepe, E.; Van Meulebroek, L.; Rombouts, C.; Huysman, S.; Verplanken, K.; Lapauw, B.; Wauters, J.; Hemeryck, L.Y.; Vanhaecke, L. A Validated Multi-Matrix Platform for Metabolomic Fingerprinting of Human Urine, Feces and Plasma Using Ultra-High Performance Liquid-Chromatography Coupled to Hybrid Orbitrap High-Resolution Mass Spectrometry. *Anal Chim Acta* **2018**, *1033*, 108–118, doi:10.1016/j.aca.2018.06.065.

116. FOLCH, J.; LEES, M.; SLOANE STANLEY, G.H. A Simple Method for the Isolation and Purification of Total Lipides from Animal Tissues. *J Biol Chem* **1957**, *226*, 497–509, doi:10.1016/S0021-9258(18)64849-5.

117. Matyash, V.; Liebisch, G.; Kurzchalia, T. v.; Shevchenko, A.; Schwudke, D. Lipid Extraction by Methyl-Tert-Butyl Ether for High-Throughput Lipidomics. *J Lipid Res* **2008**, *49*, 1137–1146, doi:10.1194/JLR.D700041-JLR200.

118. Tulipani, S.; Mora-Cubillos, X.; Jaúregui, O.; Llorach, R.; García-Fuentes, E.; Tinahones, F.J.; Andres-Lacueva, C. New and Vintage Solutions To Enhance the Plasma Metabolome Coverage by LC-ESI-MS Untargeted Metabolomics: The Not-So-Simple Process of Method Performance Evaluation. *Anal. Chem* **2015**, *87*, 39, doi:10.1021/ac503031d.

119. Sangster, T.; Major, H.; Plumb, R.; Wilson, A.J.; Wilson, I.D. A Pragmatic and Readily Implemented Quality Control Strategy for HPLC-MS and GC-MS-Based Metabonomic Analysis. *Analyst* **2006**, *131*, 1075–1078, doi:10.1039/B604498K.

120. Gika, H.G.; Wilson, I.D. Global Metabolic Profiling for the Study of Alcohol-Related Disorders. *Bioanalysis* 2014.

121. Gika, H.; Virgiliou, C.; Theodoridis, G.; Plumb, R.S.; Wilson, I.D. Untargeted LC/MS-Based Metabolic Phenotyping (Metabonomics/Metabolomics): The State of the Art. *J Chromatogr B Analyt Technol Biomed Life Sci* 2019, *1117*, 136–147.

122. Human Metabolome Database Available online: https://hmdb.ca/ (accessed on 24 January 2023).

123. MassBank | Database | Search Available online: http://www.massbank.jp/Search (accessed on 25 January 2023).

124. LIPID MAPS Available online: https://lipidmaps.org/ (accessed on 24 January 2023).

125. Yi, L.; Dong, N.; Yun, Y.; Deng, B.; Ren, D.; Liu, S.; Liang, Y. Chemometric Methods in Data Processing of Mass Spectrometry-Based Metabolomics: A Review. *Anal Chim Acta* **2016**, *914*, 17–34, doi:10.1016/j.aca.2016.02.001.

126. Khalil, S.K.H.; Azooz, M. a; Division, P. Application of Vibrational Spectroscopy in Identification of the Composition of the Urinary Stones. *Journal of Applied Science Research* **2007**, *3*, 387–391.

127. Lovergne, L.; Lovergne, J.; Bouzy, P.; Untereiner, V.; Offroy, M.; Garnotel, R.; Thiéfin, G.; Baker, M.J.; Sockalingum, G.D. Investigating Pre-Analytical Requirements for Serum and Plasma Based Infrared Spectro-Diagnostic. *J Biophotonics* **2019**, *12*, 1–12, doi:10.1002/jbio.201900177.

128. Gajjar, K.; Heppenstall, L.D.; Pang, W.; Ashton, K.M.; Trevisan, J.; Patel, I.I.; Llabjani, V.; Stringfellow, H.F.; Martin-Hirsch, P.L.; Dawson, T.; et al. Diagnostic Segregation of Human Brain Tumours Using Fourier-Transform Infrared and/or Raman Spectroscopy Coupled with Discriminant Analysis. *Analytical Methods* **2013**, *5*, 89–102, doi:10.1039/c2ay25544h.

129. Kaznowska, E.; Depciuch, J.; Łach, K.; Kołodziej, M.; Koziorowska, A.; Vongsvivut, J.; Zawlik, I.; Cholewa, M.; Cebulski, J. The Classification of Lung Cancers and Their Degree of Malignancy by FTIR, PCA-LDA Analysis, and a Physics-Based Computational Model. *Talanta* **2018**, *186*, 337–345, doi:10.1016/j.talanta.2018.04.083.

130. Spalding, K.; Bonnier, F.; Bruno, C.; Blasco, H.; Board, R.; Benz-de Bretagne, I.; Byrne, H.J.; Butler, H.J.; Chourpa, I.; Radhakrishnan, P.; et al. Enabling Quantification of Protein Concentration in Human Serum Biopsies Using Attenuated Total Reflectance – Fourier Transform Infrared (ATR-FTIR) Spectroscopy. *Vib Spectrosc* **2018**, *99*, 50–58, doi:10.1016/j.vibspec.2018.08.019.

131. Wang, X.; Wu, Q.; Li, C.; Zhou, Y.; Xu, F.; Zong, L.; Ge, S. A Study of Parkinson's Disease Patients' Serum Using FTIR Spectroscopy. *Infrared Phys Technol* **2020**, *106*, 103279, doi:10.1016/j.infrared.2020.103279.

132. Mateus Pereira de Souza, N.; Hunter Machado, B.; Koche, A.; Beatriz Fernandes da Silva Furtado, L.; Becker, D.; Antonio Corbellini, V.; Rieger, A. Detection of Metabolic Syndrome with ATR-FTIR Spectroscopy and Chemometrics in Blood Plasma. *Spectrochim Acta A Mol Biomol Spectrosc* **2023**, *288*, 122135, doi:10.1016/J.SAA.2022.122135.

133. Banerjee, A.; Gokhale, A.; Bankar, R.; Palanivel, V.; Salkar, A.; Robinson, H.; Shastri, J.S.; Agrawal, S.; Hartel, G.; Hill, M.M.; et al. Rapid Classification of COVID-19 Severity by ATR-FTIR Spectroscopy of Plasma Samples. *Cite This: Anal. Chem* **2021**, *93*, 10391–10396, doi:10.1021/acs.analchem.1c00596.

134. Eikje, N.S. Diabetic Interstitial Glucose in the Skin Tissue by Atr-Ftir Spectroscopy versus Capillary Blood Glucose. *J Innov Opt Health Sci* **2010**, *3*, 81–90, doi:10.1142/S1793545810000903.

135. el Khoury, Y.; Collongues, N.; de Sèze, J.; Gulsari, V.; Patte-Mensah, C.; Marcou, G.; Varnek, A.; Mensah-Nyagan, A.G.; Hellwig, P. Serum-Based Differentiation between Multiple Sclerosis and Amyotrophic Lateral Sclerosis by Random Forest Classification of FTIR Spectra. *Analyst* **2019**, *144*, 4647–4652, doi:10.1039/c9an00754g.

136. Ahmed, S.S.S.J.; Santosh, W.; Kumar, S.; Thanka Christlet, T.H. Neural Network Algorithm for the Early Detection of Parkinson's Disease from Blood Plasma by FTIR Micro-Spectroscopy. *Vib Spectrosc* **2010**, *53*, 181–188, doi:10.1016/j.vibspec.2010.01.019.

137. Roy, S.; Perez-Guaita, D.; Bowden, S.; Heraud, P.; Wood, B.R. Spectroscopy Goes Viral: Diagnosis of Hepatitis B and C Virus Infection from Human Sera Using ATR-FTIR Spectroscopy. *Clinical Spectroscopy* **2019**, *1*, 100001, doi:10.1016/j.clispe.2020.100001.

138. Finlayson, D.; Rinaldi, C.; Baker, M.J. Is Infrared Spectroscopy Ready for the Clinic? *Anal Chem* 2019.

139. Kochan, K.; Bedolla, D.E.; Perez-Guaita, D.; Adegoke, J.A.; Chakkumpulakkal Puthan Veettil, T.; Martin, M.; Roy, S.; Pebotuwa, S.; Heraud, P.; Wood, B.R. Infrared Spectroscopy of Blood. *Appl Spectrosc* 2021, *75*, 611–646.

140. Christensen, D.; Rudiether, A.; Kochan, K.; Peacuterez-Guaita, D.; Wood, B. Whole-Organism Analysis by Vibrational Spectroscopy. *Annual Review of Analytical Chemistry* **2019**, *12*, 89–108, doi:10.1146/annurev-anchem-061318-115117.

141. Caixeta, D.C.; Aguiar, E.M.G.; Cardoso-Sousa, L.; Coelho, L.M.D.; Oliveira, S.W.; Espindola, F.S.; Raniero, L.; Crosara, K.T.B.; Baker, M.J.; Siqueira, W.L.; et al. Salivary Molecular Spectroscopy: A Sustainable, Rapid and Non-Invasive Monitoring Tool for

Diabetes Mellitus during Insulin Treatment. *PLoS One* **2020**, *15*, doi:10.1371/journal.pone.0223461.

142. Lilo, T.; Morais, C.L.M.; Ashton, K.M.; Pardilho, A.; Davis, C.; Dawson, T.P.; Gurusinghe, N.; Martin, F.L. Spectrochemical Differentiation of Meningioma Tumours Based on Attenuated Total Reflection Fourier-Transform Infrared (ATR-FTIR) Spectroscopy. *Anal Bioanal Chem* **2020**, *412*, 1077–1086, doi:10.1007/s00216-019-02332-w.

143. Song, C.L.; Vardaki, M.Z.; Goldin, R.D.; Kazarian, S.G. Fourier Transform Infrared Spectroscopic Imaging of Colon Tissues: Evaluating the Significance of Amide I and C–H Stretching Bands in Diagnostic Applications with Machine Learning. *Anal Bioanal Chem* **2019**, *411*, 6969–6981, doi:10.1007/s00216-019-02069-6.

144. Owens, G.L.; Gajjar, K.; Trevisan, J.; Fogarty, S.W.; Taylor, S.E.; Da Gama-Rose, B.; Martin-Hirsch, P.L.; Martin, F.L. Vibrational Biospectroscopy Coupled with Multivariate Analysis Extracts Potentially Diagnostic Features in Blood Plasma/Serum of Ovarian Cancer Patients. *J Biophotonics* **2014**, *7*, 200–209, doi:10.1002/jbio.201300157.

145. Breast Cancer Statistics | World Cancer Research Fund Available online: https://www.wcrf.org/dietandcancer/cancer-trends/breast-cancer-statistics (accessed on 21 January 2021).

146. Bury, D.; Morais, C.L.M.; Paraskevaidi, M.; Ashton, K.M.; Dawson, T.P.; Martin, F.L. Spectral Classification for Diagnosis Involving Numerous Pathologies in a Complex Clinical Setting: A Neuro-Oncology Example. *Spectrochim Acta A Mol Biomol Spectrosc* **2019**, *206*, 89–96, doi:10.1016/j.saa.2018.07.078.

147. Martin, M.; Perez-Guaita, D.; Andrew, D.W.; Richards, J.S.; Wood, B.R.; Heraud, P. The Effect of Common Anticoagulants in Detection and Quantification of Malaria Parasitemia in Human Red Blood Cells by ATR-FTIR Spectroscopy. *Analyst* **2017**, *142*, 1192–1199, doi:10.1039/c6an02075e.

148. Zeng, M.; Liang, Y.; Li, H.; Wang, M.; Wang, B.; Chen, X.; Zhou, N.; Cao, D.; Wu, J. Plasma Metabolic Fingerprinting of Childhood Obesity by GC/MS in Conjunction with Multivariate Statistical Analysis. *J Pharm Biomed Anal* **2010**, *52*, 265–272, doi:10.1016/j.jpba.2010.01.002.

149. Pizarro Millán, C.; Forina, M.; Casolino, C.; Leardi, R. Extraction of Representative Subsets by Potential Functions Method and Genetic Algorithms. *Chemometrics and Intelligent Laboratory Systems* **1998**, *40*, 33–52, doi:10.1016/S0169-7439(97)00080-4.

150. Forina, M.; Lanteri, S.; Casale, M. *Multivariate Calibration*; 2007; Vol. 1158; ISBN 0103532684.

151. Santos, M.C.D.; Nascimento, Y.M.; Araújo, J.M.G.; Lima, K.M.G. ATR-FTIR Spectroscopy Coupled with Multivariate Analysis Techniques for the Identification of DENV-3 in Different Concentrations in Blood and Serum: A New Approach. *RSC Adv* **2017**, *7*, 25640–25649, doi:10.1039/c7ra03361c.

152. Paraskevaidi, M.; Morais, C.L.M.; Freitas, D.L.D.; Lima, K.M.G.; Mann, D.M.A.; Allsop, D.; Martin-Hirsch, P.L.; Martin, F.L. Blood-Based near-Infrared Spectroscopy for the Rapid Low-Cost Detection of Alzheimer's Disease. *Analyst* **2018**, doi:10.1039/c8an01205a.

153. Pizarro, C.; Arenzana-Rámila, I.; Pérez-del-Notario, N.; Pérez-Matute, P.; González-Sáiz, J.M. Thawing as a Critical Pre-Analytical Step in the Lipidomic Profiling of Plasma Samples: New Standardized Protocol. *Anal Chim Acta* **2016**, *912*, 1–9, doi:10.1016/j.aca.2016.01.058.

154. Lauridsen, M.B.; Bliddal, H.; Christensen, R.; Danneskiold-Samsøe, B.; Bennett, R.; Keun, H.; Lindon, J.C.; Nicholson, J.K.; Dorff, M.H.; Jaroszewski, J.W.; et al. H NMR Spectroscopy-Based Interventional Metabolic Phenotyping: A Cohort Study of Rheumatoid Arthritis Patients., doi:10.1021/pr1002774.

155. Workman, J. The State of Multivariate Thinking for Scientists in Industry: 1980-2000. In Proceedings of the Chemometrics and Intelligent Laboratory Systems; 2002.

156. Virel, A.; Dudka, I.; Laterveer, R.; Bjerkén, S. af 1H NMR Profiling of the 6-OHDA Parkinsonian Rat Brain Reveals Metabolic Alterations and Signs of Recovery after N-Acetylcysteine Treatment. *Molecular and Cellular Neuroscience* **2019**, *98*, 131–139, doi:10.1016/j.mcn.2019.06.003.

157. Barbosa-Cánovas, G. v; Board, A.; Candoğan, K.; Hartel, R.W.; Kokini, J.; Mccarthy, M.; Niranjan, K.; Peleg, M.; Rahman, S.; Rao, M.A. *Food Engineering Series Series Editor*;

158. González-Domínguez, R.; García-Barrera, T.; Gómez-Ariza, J.L. Combination of Metabolomic and Phospholipid-Profiling Approaches for the Study of Alzheimer's Disease. *J Proteomics* **2014**, *104*, 37–47, doi:10.1016/j.jprot.2014.01.014.

159. Lin, S.; Yue, X.; Wu, H.; Han, T. li; Zhu, J.; Wang, C.; Lei, M.; Zhang, M.; Liu, Q.; Xu, F. Explore Potential Plasma Biomarkers of Acute Respiratory Distress Syndrome (ARDS) Using GC–MS Metabolomics Analysis. *Clin Biochem* **2019**, *66*, 49–56, doi:10.1016/j.clinbiochem.2019.02.009.

160. Schmerler, D.; Neugebauer, S.; Ludewig, K.; Bremer-Streck, S.; Brunkhorst, F.M.; Kiehntopf, M. Targeted Metabolomics for Discrimination of Systemic Inflammatory Disorders in Critically Ill Patients. *J Lipid Res* **2012**, *53*, 1369–1375, doi:10.1194/jlr.P023309.

161. Wu, D.; Shu, T.; Yang, X.; Song, J.-X.; Zhang, M.; Yao, C.; Liu, W.; Huang, M.; Yu, Y.; Yang, Q.; et al. Plasma Metabolomic and Lipidomic Alterations Associated with COVID-19. *Natl Sci Rev* **2020**, doi:10.1093/nsr/nwaa086.

162. Cook, R.D. Principal Components, Sufficient Dimension Reduction, and Envelopes. *Annu Rev Stat Appl* **2018**, *5*, 533–559, doi:10.1146/annurev-statistics-031017-100257.

163. M. Forina, S. Lanteri, C. Armanino, M.C.C. Oliveros, C.C. V-Parvus 2003.

164. Forina, M.; Lanteri, S.; Oliveros, M.C.C.; Millan, C.P. Selection of Useful Predictors in Multivariate Calibration. *Anal Bioanal Chem* **2004**, *380*, 397–418, doi:10.1007/s00216-004-2768-x.

165. Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. In Proceedings of the Chemometrics and Intelligent Laboratory Systems; 2001.

166. Palomino-Vasco, M.; Mora-Diez, N.M.; Rodríguez-Cáceres, M.I.; Acedo-Valenzuela, M.I.; Alcaraz, M.R.; Goicoechea, H.C. Exploring the Potential of Combining Chemometric

Approaches to Model Non-Linear Multi-Way Data with Quantitative Purposes – A Case Study. *Anal Chim Acta* **2021**, *1141*, 63–70, doi:10.1016/j.aca.2020.10.039.

167. Gu, H.; Pan, Z.; Xi, B.; Asiago, V.; Musselman, B.; Raftery, D. Principal Component Directed Partial Least Squares Analysis for Combining Nuclear Magnetic Resonance and Mass Spectrometry Data in Metabolomics: Application to the Detection of Breast Cancer. *Anal Chim Acta* **2011**, doi:10.1016/j.aca.2010.11.040.

168. Forina, M.; Armanino, C.; Raggio, V. Clustering with Dendrograms on Interpretation Variables. *Anal Chim Acta* **2002**, *454*, 13–19, doi:10.1016/S0003-2670(01)01517-3.

169. Ahmed, S.S.S.J.; Santosh, W.; Kumar, S.; Thanka Christlet, T.H. Neural Network Algorithm for the Early Detection of Parkinson's Disease from Blood Plasma by FTIR Micro-Spectroscopy. *Vib Spectrosc* **2010**, *53*, 181–188, doi:10.1016/j.vibspec.2010.01.019.

170. Ortmayr, K.; Causon, T.J.; Hann, S.; Koellensperger, G. Trends in Analytical Chemistry Increasing Selectivity and Coverage in LC-MS Based Metabolome Analysis. *Trends in Analytical Chemistry* **2016**, *82*, 358–366, doi:10.1016/j.trac.2016.06.011.

171. Worley, B.; Halouska, S.; Powers, R. Utilities for Quantifying Separation in PCA/PLS-DA Scores Plots. *Anal Biochem* **2013**, *433*, 102–104, doi:10.1016/j.ab.2012.10.011.

172. Forina, M.; Oliveri, P.; Casale, M. Complete Validation for Classification and Class Modeling Procedures with Selection of Variables and/or Additional Computed Variables. *Chemometrics and Intelligent Laboratory Systems* **2010**, *102*, 110–122, doi:10.1016/j.chemolab.2010.04.011.

173. Tharwat, A.; Gaber, T.; Ibrahim, A.; Hassanien, A.E. Linear Discriminant Analysis: A Detailed Tutorial. *AI Communications* **2017**, *30*, 169–190, doi:10.3233/AIC-170729.

174. Sreedhar, H.; Varma, V.K.; Gambacorta, F. V.; Guzman, G.; Walsh, M.J. Infrared Spectroscopic Imaging Detects Chemical Modifications in Liver Fibrosis Due to Diabetes and Disease. *Biomed Opt Express* **2016**, *7*, 2419, doi:10.1364/boe.7.002419.

175. Flåten, G.R.; Grung, B.; Kvalheim, O.M. A Method for Validation of Reference Sets in SIMCA Modelling. *Chemometrics and Intelligent Laboratory Systems* **2004**, *72*, 101–109, doi:10.1016/J.CHEMOLAB.2004.03.003.

176. Frisvad, J.C. Classification I SIMCA What Is Classification ? SIMCA Is a Supervised Method. **2010**, 1–14.

177. Lavine, B. A User-Friendly Guide to Multivariate Calibration and Classification , Tomas Naes, Tomas Isakson, Tom Fearn and Tony Davies, NIR Publications, Chichester, 2002, ISBN 0-9528666-2-5, £45.00. . *J Chemom* **2003**, doi:10.1002/cem.815.

178. Lindon, J.C.; Nicholson, J.K.; Holmes, E. *The Handbook of Metabonomics and Metabolomics*; 2007; ISBN 9780444528414.

179. Area-Gomez, E.; Larrea, D.; Yun, T.; Xu, Y.; Hupf, J.; Zandkarimi, F.; Chan, R.B.; Mitsumoto, H. Lipidomics Study of Plasma from Patients Suggest That ALS and PLS Are Part of a Continuum of Motor Neuron Disorders. *Sci Rep* **2021**, *11*, 13562, doi:10.1038/s41598-021-92112-3.

Chapter 3

Methodology

3   CHAPTER 3. METHODOLOGY

3.1   CHEMICALS AND REAGENTS

Ultrapure water, used to prepare all the aqueous solutions, was obtained from a Milli-Q system (Milipore, Bedford, MA, USA). LC–MS grade acetonitrile (ACN), isopropyl alcohol (IPA), ammonium formate, high-performance liquid chromatography (HPLC) grade methanol and methyl tert-butyl ether (MTBE) were supplied by Aldrich Chemie (Steinheim, Germany).

3.2   METHODS OF SAMPLE PREPARATION

3.2.1   Sample collection and storage

3.2.1.1   *Samples of patients with Parkinson, Alzheimer and Metabolic syndrome*

Plasma samples were provided by la Rioja Blood Bank, after the submission for diagnostic testing in the respective Department in the Centre for Biological Research of La Rioja (CIBIR) in Spain. Venous blood samples were drawn via antecubital venepuncture from each subject in a sitting position. Becton Dickinson (BD) Vacutainer plastic blood collection tubes were used, with clot activator and K2EDTA for serum and plasma separation, respectively. Blood was processed to obtain serum and plasma by centrifugation at 2200 rpm for 15 min at 4·C. As soon as the samples were delivered to the laboratory, they were frozen and stored at -80°C until further use.

3.2.1.2   *Samples of patients with Amyotrophic lateral sclerosis*

The blood samples from Niguarda Ca'Granda Hospital in Milan (Italy) were obtained after first centrifuging and each Pax tube containing the blood sample for 10 minutes at 3000-5000 rpm and incubated for 2 hours at room temperature, following the recommendations of the commercial kit [1]. The supernatant samples recovered from this first centrifugation were used to perform the analysis in this study. In addition, these samples were preserved at -80 °C for further use.

3.2.2   Lipid extraction of plasma samples

3.2.2.1 MTBE-US-assisted lipid extraction method

The samples left to defrost in the fridge for 8 hours by night was submitted to the extraction the following day.

Lipid extraction was performed according to a protocol tested and optimised in our research group [2]. According to this MTBE-US-assisted lipid extraction method, 5 µL of Milli-Q water was added to a 10 µL aliquot of human blood plasma. Then, 20 µL of methanol was added to precipitate proteins by vortex-mixing for 2 min. Then, 250 µl de MTBE was added and dispersed by immersing the mixture in an ultrasonic water bath supplied by ATU Ultrasonidos (Valencia, Spain). The ultrasound frequency and power were 40 kHz and 100 W, respectively. The temperature was set at 15·C, and the time was adjusted to 30 min. Once USAE was performed, 25 µL of MilliQ water was added to the mixture. The organic phase was separated by centrifugation at 3000 rpm for 10 min at 10 ·C in an Eppendorf 5403 Refrigerated Centrifuge (Hettich, Tuttlingen, Germany).

The lipid extracts in the upper phase were diluted five times with injection solvent before being collected and poured into an autosampler vial.

To check the methodology's performance and test its precision, quality samples (QC) were processed in the same manner as the actual samples and were inserted regularly throughout the analytical run.

3.3   ANALYTICAL METHODS

3.3.1   Analytical instruments

3.3.1.1   Analysis of samples by FT-IR

The ABB FT-IR MB3000 Fourier transform infrared spectrophotometer (Zurich, Switzerland) was used to obtain the metabolic profiles of plasma samples of patients with Parkinson's, Alzheimer's disorders and Metabolic syndrome.

The instrumental components of the ABB FT-IR MB3000 spectrophotometer are as follows in Figure 3-1.

The removable cell for liquid samples from PerkinElmerSpecac (Cell Ommi, Specac Ltd., United Kingdom) used to obtain the spectra consists of two CaF2 windows separated by a 50-micron Mylar spacer, which allows fixing the optical path.

Image of the equipment and instrumentation used: a) FT-IR spectrophotometer MB3000. b) Specacr liquid cell. c) Components of the Specac liquid cell.
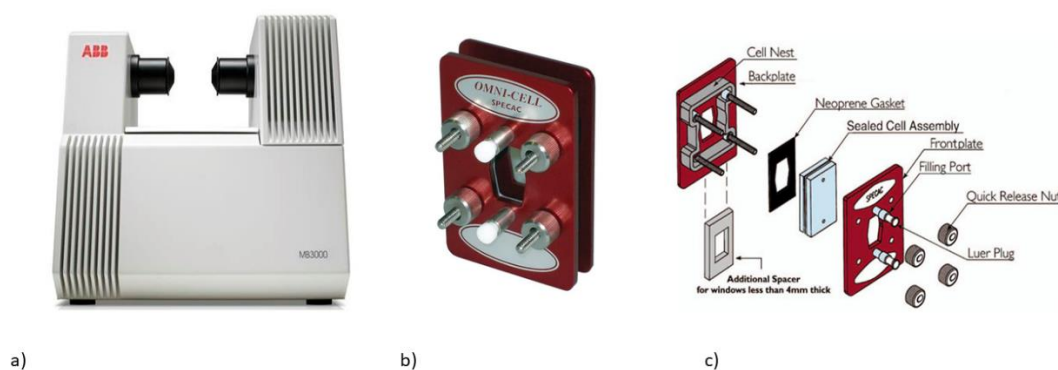
Figure 3-1. Principal parts of the infrared spectroscopy used for the analysis: a) FT-IR spectrophotometer MB3000. b) Specac liquid cell. c) Components of the Specac liquid cell.

The procedure carried out for obtaining the metabolic profiles is as follows:

1. Open the N2 source. This source must remain open while performing spectroscopic measurements to purge the equipment, eliminating atmospheric water vapour and CO2.

2. Turn on the Horizon MBTM program and set the acquisition parameters for the method (Table 4.3.1). The following measurements were taken at a temperature of 23.0 ± 1.0°C.

3. Perform a reference by placing the cuvette with a Milli-Q water sample. The reference should be performed before each sample measurement to check that the measurement conditions are stable and that the data obtained are not disturbed by temperature or other factors.

4. Disassemble the Specac cuvette and add 20 μL of the plasma sample between the CaF2 windows. Each sample should be measured in triplicate to minimise possible experimental errors.

5. Record the IR spectrum of the analysed sample between 400-4000 cm^{-1} in case of analysis of samples relative to Metabolic syndrome and Parkinson's disease and 1000-1500cm^{-1} in case of ALS samples. To perform a new spectrum, it is necessary to disassemble the PerkinElmer cuvette and clean both the CaF2 windows and the Mylar spacer with distilled water. These components should be dried using tissue paper.

During the analysis, it was necessary to use control samples (QC) to check the performance and reproducibility of the methodology, inserting them regularly during the analyses. The control samples were processed the same way as the real samples.

3.3.1.1.1 ATR-FTIR spectrometer

A compact Spectrum Two FT-IR spectrometer was used for the integrative analysis of ALS blood samples. The same procedure of the analytical flow was followed as described above for the ABB instrument, except that this instrument does not require an N2 source, which is the main advantage in everyday analysis and portability (Figure 3-2).

Moreover, the moving platform allowed the analyses on the diamond crystal for spectral acquisition. Thus, 20 μL volume was deposited onto the ATR crystal. Therefore, ATR-FTIR spectra were collected between 1500-1000 cm^{-1} with 16 scans and a spectral resolution of 2 nm. The samples were randomised to minimise other influences, such as noise and instrumental variability. The spectra for each sample were collected in triplicate. The ATR crystal was cleaned with distilled water, and a new background spectrum was acquired to take into account ambient changes before the next sampling.
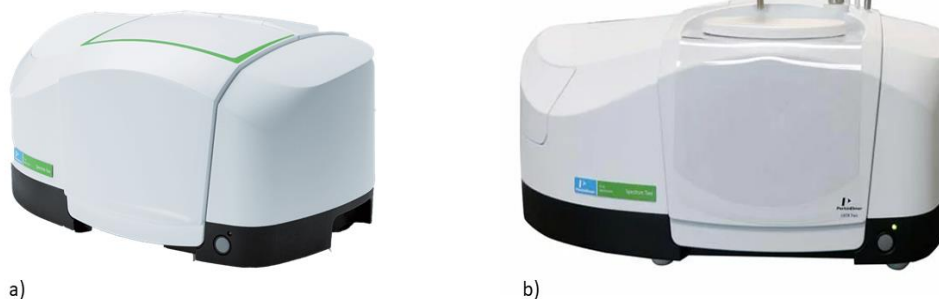
Figure 3-2. Spectrum Two™ spectrometer image: a) General representation, b) ATR accessorised.

3.3.1.2 LC conditions and experiments

3.3.1.2.1 Liquid Chromatography-Mass Spectrometry

To determine plasma lipid profiles, a Waters Acquity UPLC chromatography system (Milford, MA, USA), equipped with a Waters Acquity HSS T3 100 × 2.1 (i.d.) mm 1.8 µm particle size column and a Waters VanGuard precolumn of the same material, coupled to a Microtof-Q (Q-TOF) mass spectrometer (Bruker Daltonik GmbH, Germany) with an electrospray interface (ESI) was used (Figure 3-3). To ensure the quality and stability of samples, the temperature of the autosampler was maintained at 5°C, and the column at 55°C. A mass spectrometer was operated in both positive and negative modes. Chromatographic and mass spectrometry data were acquired using Data Analysis software Version 4.0 (Bruker Daltonik GmbH, Germany). 2 µL samples were injected. Elution was performed using the gradient mobile consisting of phase A (acetonitrile–water mixture (60:40, v/v) with 10 mM ammonium formate) and phase B (acetonitrile–isopropanol mixture (10:90, v/v) with 10 mM ammonium formate). UPLC separation was performed using a linear gradient that increased from 40% to 100% B within 10 min and was held at 100% B for an additional 2 min. Finally, it increased from 0% to 60% A within 3.5 min. The total run time was 15.5 min. The flow rate was 0.4 mL*min^{-1}, and the injection volume was 10 µL.

Mass spectrometry data were acquired using a Waters Synapt XS HDMS (Waters Corp, Milford, USA) set to collect the data in continuum format using electrospray ionisation (ESI) in positive ionisation mode (ESI+) and negative ionisation mode (ESI−), over the mass range of m/z 50–2000. The capillary and sampling cone were set to 1.75 kV and 40 V, respectively, with the source temperature set to 120 °C and the desolvation temperature to 500 °C. Gas flow rates were set at 800 L/h for the desolvation gas and 50 L/h for the cone gas, and the nebuliser gas was fixed at 6 bars. The mass spectrometer was set to acquire in resolution mode with a scan time of 0.4 s. Fragment ion information was acquired using a collision energy ramp from 20 to 50 V.

Lockmass correction was achieved by infusing leucine enkephalin at 10 µL/min through a lock spray probe and acquired every 30 s; for positive mode, [M + H]+ = 556.2771, and negative mode, [M − H]− = 554.2615. The data were collected using MassLynx V 4.2 (Waters Corp., Milford, USA).

The diseased samples and controls were alternated concerning run, avoiding batch effect. Moreover, QC samples were inserted regularly throughout the analytical run (after every 20 real samples) to check the methodology's performance and test its precision.

Figure 3-3. UPLC-MS instrument coupled to Q/TOF mass analyser.

3.3.1.2.2 LC-MS data analysis

The acquired raw mass data were imported to Progenesis QI software (Waters Corporation, Milford, MA, USA) for peak detection, alignment, retention time correction and normalisation.

Repeatability and intermediate precision were determined using quality-control samples (QCs) to check the precision of the method by computing the Relative Standard Deviation (RSD) or Coefficient of Variation (CV) of intra-day and inter-day analyses.

MS-derived lipid identification was based on the mass match of lipids with available online databases: the Human Metabolome Data Base (HMDB) (http://hmdb.ca)[3,4], LIPID MAPS (http://lipidMAPS.org) [5], METLIN (http://metlin.scripps.edu) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/). These databases are complementary to each other, in such a way that it is strongly recommended the combined use of all of them to compile a list of potential hits as exhaustive as possible with identifications of a high probability of being correct. Lipid metabolites were manually identified based on their exact masses, specific fragment and/or neutral losses [6]. The allowable mass error in database searching was adjusted to 5 ppm for the attribution of the precursor ion. Thus, an attempted assignment of possible features to specific compounds was performed.

3.4 SOFTWARE AND PROGRAMMES FOR DATA ANALYSIS

The main informatics software used to perform data analysis is summarised.

-The Unscrambler v 11.0 AspenTech Ltd (Berkshire, United Kingdom);

- V-PARVUS 2011: An Extendable Package of Programs for Data Explorative Analysis, Classification and Regression Analysis. Dipartimento di Chimica e Tecnologie Farmaceutiche ed Alimentari, Genova, Italia;

- MATLAB, High-performance numeric computation and visualization software, Version 6.5 for Windows, The MathWorks, Massachusetts, United States;

-The R Project for Statistical Computing, with the implementation of CAMERA and XCMS. Software Open Access;

-Galaxy server (http://www.**usegalaxy**.org);

-Progenesis™ QI software undertakes the spectral alignment, consistent peak picking across all runs, normalisation of the total compound abundance as well as compound quantification;

-The generated matrices were subsequently analyzed using MetaboAnalyst 5.0 (http://www.metaboanalyst.ca/), a comprehensive free and publicly accessible platform for metabolomics analysis that allows for applying univariate and multivariate methods.

3.5 References

1.   Preanalytix.Com: PAXgene - Specimen Collection & Processing Available online: https://www.preanalytix.com/ES?cHash=66aa4e4012b1785c4e7ec8d0b38b573f (accessed on 28 November 2022).

2.   Pizarro, C.; Arenzana-Rámila, I.; Pérez-Del-Notario, N.; Pérez-Matute, P.; González-Sáiz, J.M. Plasma Lipidomic Profiling Method Based on Ultrasound Extraction and Liquid Chromatography Mass Spectrometry. *Anal Chem* **2013**, *85*, 12085–12092, doi:10.1021/AC403181C.

3.   Wishart, D.S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A.C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; et al. HMDB: The Human Metabolome Database. *Nucleic Acids Res* **2007**, doi:10.1093/nar/gkl923.

4.   Wishart, D.S.; Guo, A.C.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res* **2022**, *50*, D622–D631, doi:10.1093/NAR/GKAB1062.

5.   Fahy, E.; Subramaniam, S.; Murphy, R.C.; Nishijima, M.; Raetz, C.R.H.; Shimizu, T.; Spener, F.; van Meer, G.; Wakelam, M.J.O.; Dennis, E.A. Update of the LIPID MAPS Comprehensive Classification System for Lipids. *J Lipid Res* **2009**, *50*, doi:10.1194/jlr.R800095-JLR200.

6.   Han, Xianlin. *Lipidomics : Comprehensive Mass Spectrometry of Lipids*; ISBN 9781118893128.

Chapter 4

Metabolic syndrome

Abstract

In this chapter, the potential of Fourier-transform infrared spectroscopy combined with chemometric tools to detect spectra markers indicative of metabolic syndrome was studied. Considering that MetS is a complex of interrelated risk factors for cardiovascular disease and diabetes, new point of-care diagnostics tools are highly requested to provide results quickly. Around 105 plasma samples were collected and divided into two groups according to the presence of at least three of the five clinical parameters used for MetS diagnosis. A dual classification approach was studied based on selecting the most important spectral variables and classification methods, linear discriminant analysis (LDA) and SIMCA class modelling, respectively. The same classification methods were applied to measured clinical parameters at our disposal. Thus, the classification's performance of reduced spectra fingerprints and measured clinical parameters were compared. Both approaches achieved excellent discrimination results among groups, providing almost 100% accuracy. Nevertheless, SIMCA class modelling showed higher classification performance between MetS and no MetS on IR-reduced variables compared to clinical ones. Finally, the potential of this method to be used as a supportive diagnostic or screening tool in clinical routine. was discussed.

Resumen

En este capítulo se estudió el potencial de la espectroscopía infrarroja por transformada de Fourier combinada con herramientas quimiométricas para detectar marcadores espectrales indicativos del síndrome metabólico. Teniendo en cuenta que el MetS es un conjunto de factores de riesgo interrelacionados para enfermedades cardiovasculares y diabetes, se necesitan nuevas herramientas de diagnóstico en el punto de atención para proporcionar resultados rápidamente. Se recolectaron alrededor de 105 muestras de plasma y se dividieron en dos grupos según la presencia de al menos tres de los cinco parámetros clínicos utilizados para el diagnóstico del MetS. Se estudió un enfoque de clasificación dual basado en la selección de las variables espectrales más importantes y los métodos de clasificación, análisis discriminante lineal (ADL) y modelado de clase SIMCA, respectivamente. Se aplicaron los mismos métodos de clasificación a los parámetros clínicos medidos a nuestra disposición. Por lo tanto, se comparó el rendimiento de la clasificación de huellas dactilares de espectros reducidos y parámetros clínicos medidos. Ambos enfoques lograron excelentes resultados de discriminación entre grupos, proporcionando una precisión cercana al 100%. Sin embargo, el modelado de clase SIMCA mostró un mayor rendimiento de clasificación entre el MetS y no MetS en variables IR reducidas en comparación con las clínicas. Finalmente, se discutió el potencial de este método para ser utilizado como una herramienta de diagnóstico o detección de apoyo en la práctica clínica habitual.

4 Chapter 4. Metabolic Syndrome

4.1 Dual Classification Approach for the Rapid Discrimination of Metabolic Syndrome by FTIR

4.1.1 Introduction

The high prevalence of non-communicable diseases (NCD) in adults is reflected in increased costs for public health systems worldwide [1]. Among these NCD, metabolic syndrome (MetS) plays a significant role. MetS is often associated with an increased risk of diabetes and cardiovascular disease, resulting in increased incidence of morbidity and mortality and reduced quality of life [2–6]. Thus, the commensurate prevalence of metabolic syndrome burdens national health expenditure, representing a significant socio-economic problem, particularly in low- and middle-income countries [7–10]. However, MetS is a multifactorial disorder accompanied by conflicting opinions on its definition [11–13]. In particular, many different definitions have been proposed to describe MetS in adults. The main discrepancies were associated with inclusion and exclusion criteria adopted according to the World Health Organization (WHO), National Cholesterol Education Program (NCEP), Adult Treatment Panel III (ATPIII), and International Diabetes Federation (IDF). Finally, in 2009, the definition for metabolic syndrome was harmonised [14]: MetS is a disease formed by metabolic and vascular abnormalities, namely insulin resistance (IR), visceral adiposity, atherogenic dyslipidaemia, and oxidative and endothelial dysfunction. These risk factors easily predispose hyperglycaemia and hypertension, atherosclerotic vascular diseases and viral infection [15–18].

Given the complex and intertwined nature of MetS, it would be utopian to think that a single biomarker could define it unambiguously. Thus, parameters concerned around central obesity (waist circumference (WC)), hypertension (blood pressure), atherogenic dyslipidaemia (small low-density lipoprotein (LDL) and levels of high-density lipoprotein (HDL) cholesterol), and insulin resistance (fasting glucose levels) are usually measured to evaluate MetS diagnosis [19]. Due to the heterogeneity of these factors, people affected by metabolic syndrome are three times more likely to suffer acute myocardial

87

infarction, cerebrovascular events, diabetes, or stroke. In addition, they have higher mortality rates [20]. Besides the economic impact, misdiagnosis or tardive diagnosis could lead not only to inefficient treatment outcomes but even to significant dysfunctions such as cancer [21,22]. Thus, early and proper diagnosis plays a crucial role in delaying the pathology's onset or progression as much as possible and improving a patient's condition.

Today, MetS diagnosis is based on several steps such as measuring metabolic markers of insulin resistance and other indices of metabolic syndrome (triglycerides, HDL cholesterol levels, and blood glucose) that are obtainable from routine clinical biochemistry laboratories, whereas blood pressure is measured in primary care [23]. The collection and analysis of samples also entails a waiting time for laboratory results and additional time for a new medical consultation. Although the proposed definition of MetS shares some common features, the clinical diagnosis lacks standardisation. On that basis, it was proposed that individuals showing a combination of any three out of these five simple clinical criteria were likely to be characterised by insulin resistance. Prospective analyses have also shown that any combination of these factors was predictive of an increased risk of both type 2 diabetes and cardiovascular disease. First, it is still challenging to identify a unified criteria for MetS applicable across all ethnicities. In addition, the contribution of each parameter seems to have different importance based on the evaluation adopted in each clinical environment (e.g., diagnosis focussed on glucose tolerance instead of obesity cut-offs). Moreover, there is variation in the cut-off values of diagnostic inclusion criteria (≥140/90 mmHg according to WHO vs. ≥130/85 mmHg according to ATP III for blood pressure). The application of MetS diagnosis in clinical practice could also be compromised, since most patient registries have missing data, limiting a study's accuracy or leading to false-positive results. In addition, measurements such as WC, one of the predominant parameters for defining MetS, are not always feasible in patients because the diagnosis can often be limited by the patient's inability to perform a complete physical examination.

Given these perspectives, the need for standardised clinical diagnostic tools and protocols becomes imperative in the prevention and diagnosis of MetS. For this reason, analysing global metabolic profiles instead of disparate clinical measurements could be

essential in shedding light on MetS disarrangements. A multifactorial and complex pathology such as MetS seems to require an approach from a holistic functional perspective, so an analysis of metabolic profiles reflecting the global clinical status of a patient could represent a suitable alternative.

By now, metabolomics plays a key role as a powerful analytical tool that has been widely applied to investigate plenty of disorders and disarrangements [24–26]. Metabolomics analysis has the potential to discover biomarkers and allow for the detection of a wide range of metabolites. In recent years, there has been a great interest in extracting biomarkers from biofluids and, considering that blood is a biofluid containing numerous valuable metabolic information, it seems that it in particular, it appropriately reflects metabolic changes and disarrangements during disease initiation or progression [27–29]. In this context, techniques based on vibrational spectroscopy are particularly suitable as sample preparation is simple, non-invasive, rapid, and low-cost [30]. Therefore, the Fourier transformed infrared spectroscopy (FTIR) technique has been established as a reliable analytical tool in metabolomic-based studies [31–36]. Moreover, another significant advantage resides in the fact that FTIR is ideally suitable for acquose matrices such as blood [37,38]; the instrument requires the collection of only one blood sample, with little or almost null pre-treatment. In this study, we proposed an FTIR-based method that investigates many components at a time, which are registered as spectral signatures. The development of a chemometric strategy capable of extrapolating the most significant infrared (IR) signatures plays a crucial role in this study, since each spectrum is unique for every patient and reflects their metabolic status. Non-targeted metabolomic studies, such as the one presented here, aim to extract the metabolic signatures instead of individual biomarkers with limited potential, and this permits the classification of patients according to their molecular patterns, reflecting clinical/pathological conditions such as MetS or no MetS.

This method could greatly support clinicians, capturing the complexity of the MetS metabolic profile when the clinical indicators are missing or lacking sufficient discriminative power, revealing the globality of physiological disturbances. We do not want to underestimate the importance of clinical diagnosis at any time. Still, our main

aim is to propose an alternative analytical strategy that could be of great diagnostic relevance and support, limiting the time and cost of clinical measurements.

4.1.2   Methods

4.1.2.1   Study Population

A total of 105 plasma samples from anonymous donors were recruited from Infectious Disease Area, Center for Biomedical Research of La Rioja (Logroño,Spain). This study was approved by the Committee for Ethics in Drug Research in La Rioja (CEImLAR) (23 April 2013, reference number 121) and a written informed consent was achieved from all participants. The patients were evaluated by the NCEP-ATP-III scale and, if eligible, were assigned to a metabolic syndrome category. MetS was defined as the concomitant presence of at least three of the following risk factors: elevated TGL (≥150 mg/dL), low concentrations of the fraction HDL cholesterol (<50 levels mg/dL in women or <40 mg/dL levels in men), increased WC (≥88 cm in women or ≥102 cm in men), elevated blood pressure (>130/85 mmHg), and elevated fasting glucose (>110 mg/dL or diabetes) [39]. Thus, the patients were divided into two groups by the criteria of MetS: 19 patients tested as MetS positive and 86 as MetS negative. The patients enrolled in this study were also characterised by the presence of viral load through serological evidence of HIV or co-infection of HIV/HCV. A correct distribution between patients with and without infection in both categories has been ensured to not introduce bias in future models developed for diagnosing MetS.

4.1.2.2   Sample Collection

Once drawn, the venous blood samples were centrifuged at 2200 g for 15 min at 4 °C and the obtained plasma were transferred into a clean Eppendorf tube. Aliquots of 200 µL of each sample were stored at −80 °C until the day of the analysis. Before FTIR measurements, plasma samples were defrosted during the night according to the optimised ultrasound-based protocol for lipidomic analyses developed in our research group [40].

4.1.3 Instrumentation

FTIR spectroscopy measurements were performed by a Spectrum-One ABB Miracle Type MB3000 FT-IR Spectrophotometer as described in <u>Chapter 3.3.1 Analysis of samples by FT-IR</u>.

4.1.4 Data Analysis

After data acquisition, the processing and computational analysis of raw metabolic data was performed using Unscrambler (version X 11.0, Camo ASA, Oslo, Norway), V-Parvus (version PARVUS2011, Michele Forina, Genoa, Italy), and Matlab (MATLAB 9.4 R2018a). Two different regions of the mid-IR spectrum were analysed: the first region examined was the biochemical "fingerprint region" at 1500–1050 cm^{-1}, and the second was a higher region at 2950–2700 cm^{-1}. Remaining wavenumber ranges, as they were affected by signal saturation effects caused mainly by strong water absorptions or noise, were removed, and not considered for further analysis. Given the high dimensionality of biological spectral data, many disturbing factors influence the spectral data acquisition, such as random noise, baseline distortions, or light scattering. Thus, the pre-processing step is imperative in analysis to reduce these factors. To compensate for instrumental artefacts and sample to sample variations, different pre-processing methods were evaluated individually or in combination to minimise the adulterant-unrelated variability, namely derivatives (e.g., Savitzky–Golay (S–G) first and second derivatives), standard normal variate (SNV), and extended multiplicative scatter correction (EMSC). Thus, better resolution of overlapping peaks and decreased scatter effects were ensured after applying the combination (S–G) smoothing and SNV.

The entire data set was split into two independent subsets to develop and validate the classifications proposed: a training set with 95 samples (used to optimise and develop the classification rules and models) and a test set with ten samples (never used in the construction of the classification but to evaluate their actual predictive ability). The test set used was the same for all methods applied and classifications developed. As a result, the smoothed and normalised output tables were always centred before additional multivariate analysis and classification algorithms.

4.1.5   Results and Discussion

After careful pre-processing, FTIR measurements were submitted for further multivariate analysis. Thus, five measured clinical variables and a total of 838 spectra variables over the wavelength ranges of 1583–1050 cm^{-1} and 2973–2700 cm^{-1} collected from 105 patients were included. The two main categories of this study were patients with and without metabolic syndrome, i.e., MetS and no MetS, respectively.

4.1.5.1   Descriptive Statistics

Herein, an analysis was performed based on the distribution of five clinical parameters. It should be noted that one of the most critical clinical measurements, waist circumference, was not included in this study because most patients had missing data in the clinical register. Therefore, only parameters that were available for all patients have been used for the further comparative classification step. Thus, the descriptive statistics were calculated to analyse the distribution of clinical data in a box and whisker plot (**Figure 4-1**). The plot shows that TGL values seem to have more influence and variability between the two categories of patients; indeed, MetS patients have significantly higher values ranging from a minimum of 33 to 338 (mg/dL). The general distribution trend indicates that MetS patients also have slightly higher diastolic and systolic blood pressure values and glucose levels, whereas HDL values are lower, ranging from 25 to 95 (mg/dL). **Table 4-1** shows the ranges of the collected values with the respective medians between the two categories.

Distribution of clinical values in MetS and no MetS
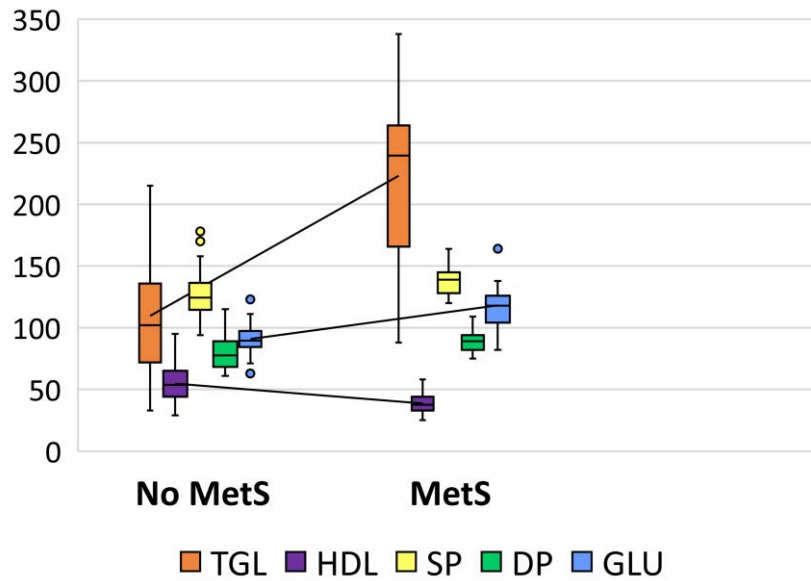
Figure 4-1. Box and whisker plot showing the distribution of clinical values levels in patients with MetS and no MetS. The line located in the middle of the box represents the median and is used to better visualise the differences between clinical parameters: triglycerides (TGL) levels are displayed in orange *(■); high density lipoprotein (HDL) in violet (■); systolic pressure (SP) in yellow (■); diastolic pressure (DP) in green (■); and glucose (GLU) in blue (■).

Table 4-1. The distribution of the clinically measured parameters in MetS and no MetS patients expressed in mg/dL and in mmHg.

Clinical Parameters	MetS			No MetS		
	Max	Min	Mean	Max	Min	Mean
Systolic blood pressure	174	120	136	178	94	126
Diastolic blood pressure	109	75	87	115	61	79
Triglycerides	338	88	242	215	33	109
HDL	58	25	37	95	29	55
Glucose	164	82	114	123	63	91

4.1.5.2 Exploratory Analysis with PCA

An unsupervised pattern recognition method based on principal component analysis (PCA) was performed for the initial data overview and to investigate any possible clustering of samples based on five collected clinical parameters and 838 spectral variables, respectively.

The PCA score plot of clinical parameters, with 50.46% of explained variance by PC1, displays evident clustering according to known categories, delimitated by the parallel to the bisector of the second quadrant (Figure 4-2). Whereas PCA performed on pre-treated IR spectra accounted for 83.12% of explained variability on the PC1, evidenced by very subtle clustering between known categories (Figure 4-3).

94
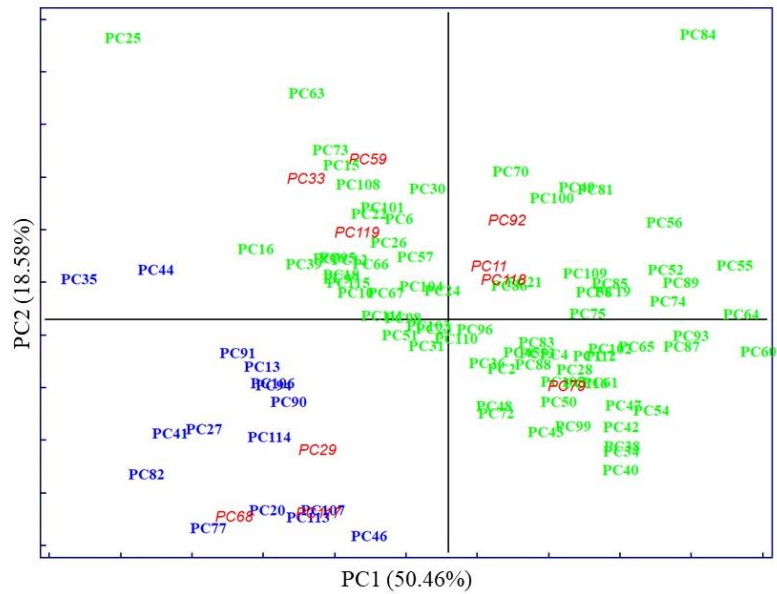
Figure 4-2. Scores for the plasma samples on the first two principal components explaining the variability in the dataset of five measured clinal parameters. The samples are labelled according to their specific pathology: no MetS (■)*,* MetS (■), and external test samples (■).
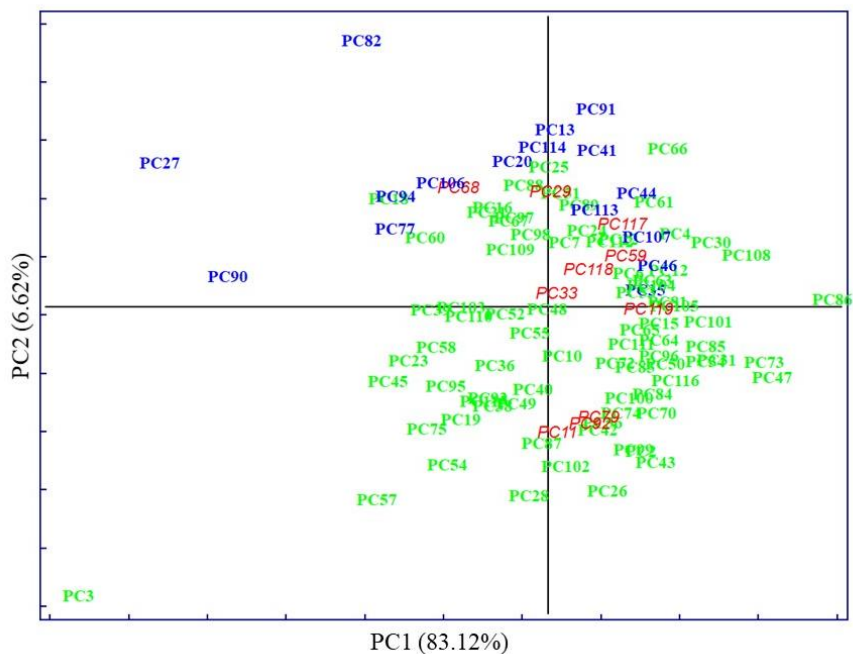
Figure 4-3. Scores for the plasma samples on the first two principal components explaining the variability in the IR spectral dataset. The samples are labelled according to their specific pathology: no MetS (■), MetS (■), and external test samples (■).

In both cases, the first PCs explained most of the data's variability. The distribution of samples in principal component space suggests that it only seems possible to address subsequent, direct discrimination in the case of analysis of clinical parameters. Thus, parameters such as TGL and GLU majorly contributed to the segregation of no MetS from MetS and the values of HDL contributed to the separation of MetS from no MetS, as was shown in preliminary analysis by descriptive statistics. No evident clustering among the two main categories was observed performing PCA on spectral variables; only a few outliers were determined and excluded from further analysis. The high degree of overlapping features among the two classes was expected, as most blood components are common in all individuals. This also indicates the need to perform a selection of relevant spectral variables, closely related to clinicopathological parameters of prognostic importance in MetS. Therefore, other chemometric strategies were used to investigate and highlight metabolomic differences in metabolic syndrome using IR spectra.

4.1.5.3   Supervised Techniques

The selection of variables in tandem with classification methods to extract reduced IR fingerprints that reflect the metabolic profiles of patients for a potential MetS diagnosis was studied. Therefore, a dual approach was applied based on a classification method on the one hand and a class modelling method on the other.

For its part, discriminant techniques focus on the differences between samples belonging to different categories, dividing the multidimensional space into as many subregions as the number of the considered classes. As a result of this work principle, every tested sample would always be assigned to one of the predefined categories, even in the case where an analysed sample truly belongs to a class not considered in the study. Regarding the above, it makes good sense to evaluate the application of a discriminant classification strategy in a two-class (binary) classification problem such as the one addressed in this paper. In particular, linear discriminant analysis (LDA), the most widely used classification algorithm, was used.

On the other hand, in contrast to class discrimination, class modelling approaches exploit similarities among inter-category samples to construct an individual model for every class independently from the others. Consequently, the developed class models

may not entirely cover the original multivariate space. This fact opens the door to different assignment scenarios depending on whether a sample falls clearly into a single class region (so that it is assigned to that) or if it falls in overlapping regions (leading to a confusing classification in multiple classes), and, finally, when a sample falls outside every class model constructed (predicted as member of none of the considered categories). Therefore, due to their specific properties, modelling techniques, such as soft independent modelling by class analogy (SIMCA), are suitable for classification problems in which the emphasis is placed on a particular class of interest, as may be the case here with the MetS category.

4.1.5.3.1 SELECT-LDA

Considering that IR data presents high dimensionality, eliminating the futile features due to noise and identifying the relevant and important variables to be applied in the following classification steps was imperative. Thus, the stepwise orthogonalization of predictors (SELECT) algorithm [39,40] was prioritised among other variable selection techniques since it enabled us to optimise discrimination by simultaneously performing feature selection and classification. Moreover, thanks to its stepwise decorrelation procedure, SELECT also avoids the presence of redundant information in the subset of selected significant predictors. In addition, it has previously demonstrated its accurate prediction ability in selecting the most important variable for the discrimination of pathological status [41,42]. Thus, SELECT was applied to extract the most significant wavenumbers from the IR dataset, providing input features for a further dual-classification approach. Based on the commonly established rule, the number of training objects selected was always at least three times greater than the number of finally selected wavenumbers. An in-depth study of the literature is encouraged to understand the algorithm's rules [43].

4.1.5.3.2 LDA on Clinical Parameters

LDA is a well-known and extensively applied powerful supervised chemometric classification technique [44]. Based on LDA classification rules, the objects are always classified in one of the predefined classes.

LDA of five clinical parameters, built by leave one out (LOO) cross-validation, was performed to evaluate the feasibility of this classification methodology to differentiate between MetS and no MetS patients. Excellent discrimination among categories was achieved, providing a 100% level of correctly classified samples for no MetS subjects and patients with metabolic syndrome, respectively. Satisfactory external prediction performances ranging from 98.73% to 100% were achieved for both categories (within one no MetS subject classified as MetS), respectively (Table 4-2). Furthermore, a clear interclass separation achieved between these main categories can also be visually appreciated in the corresponding discriminative histogram (Figure 4-4). This classification performance was almost predictable since the PCA results already showed a clear clustering between the two groups.

Table 4-2. Results of LDA classification performance on clinical parameters

Clinical Parameters	Classification (%)	External Prediction (%)	Total Rate (%)
MetS	100	100	100
No MetS	100	98.73 (1)[1]	99.36
Total rate	100	98.94	99.47

[1]The one corresponds to one misclassified subject in cross-validation

.
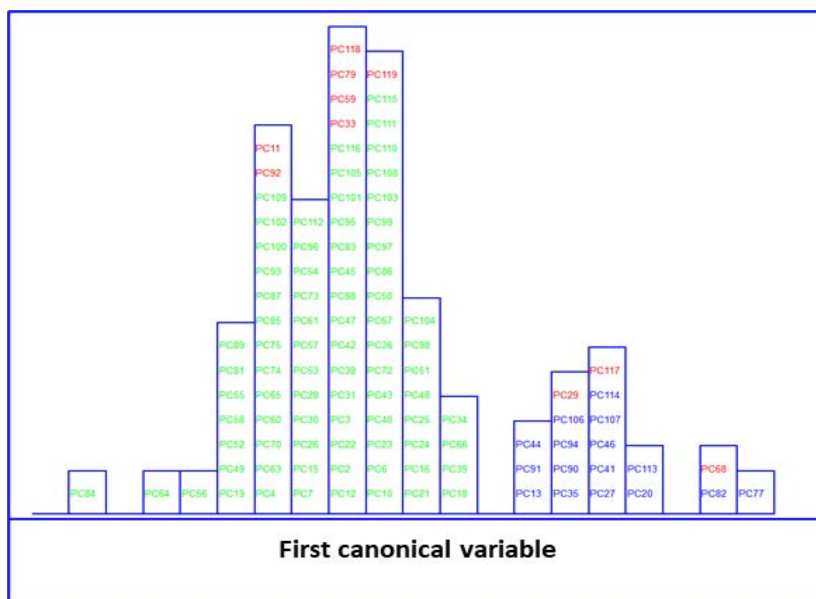
Figure 4-4. Histogram of the first canonical variable for the discrimination of MetS (■) and no MetS (■) patients within included (■) test set, after performing LDA in the stratification approach based on clinical parameters (y-axis indicates the maximum discrimination power between categories).

The object belonging to the category MetS which was classified as no MetS was characterised by the following clinical parameters: 213 mg/mL of TGL, 76 mg/mL of HDL, 139 mmHg of SP, 83 mmHg of DP, and 102 mg/mL of GLU. As we can see, two out of five parameters have increased values, and the DP parameter is very close to the cut-off value, which is 85 mmHg based on the NCEP-ATP-III scale. Thus, this patient might instead be classified as MetS positive, presenting almost three out of five clinical parameters with augmented values. In addition, as we said above, the TGL parameter has a major contribution, among other parameters, to MetS classification. Thus, the plausible explanation could be that this subject, who has greater values of TGL, is more likely to be classified as MetS by LDA rather than no MetS. However, as we highlighted before, the eligibility criteria can be very insidious and create confusion and misassignment, worsening and delaying the patients' well-being.

4.1.5.3.3 SELECT-LDA on IR Wavenumbers

Likewise, LDA on the IR dataset, containing 838 wavenumbers, was also performed. Before LDA analysis, as explained above, SELECT was applied to extract those predictor variables correlated with the discrimination between categories here considered. Therefore, based on the SELECT rules, 20 selected spectra variables were decorrelated

from other signals and used for LDA. The 20 selected features showed an outstanding classification performance and the results were higher in performance than LDA results on clinical parameters, achieving 100% in classification and external prediction, respectively. The results of the SELECT LDA performance are displayed in Table 4-3. The suitability of the classification strategy applied to reduced IR plasma signatures can be visually appreciated in Figure 4-5. A discriminative histogram shows a clear group separation on the first canonical variable.

Table 4-3. Results of SELECT LDA classification performance on 20 IR selected spectral variables.

| Clinical Parameters	Classification (%)	External Prediction (%)	Total Rate (%)
MetS	100	100	100
No MetS	100	100	100
Total rate	100	100	100
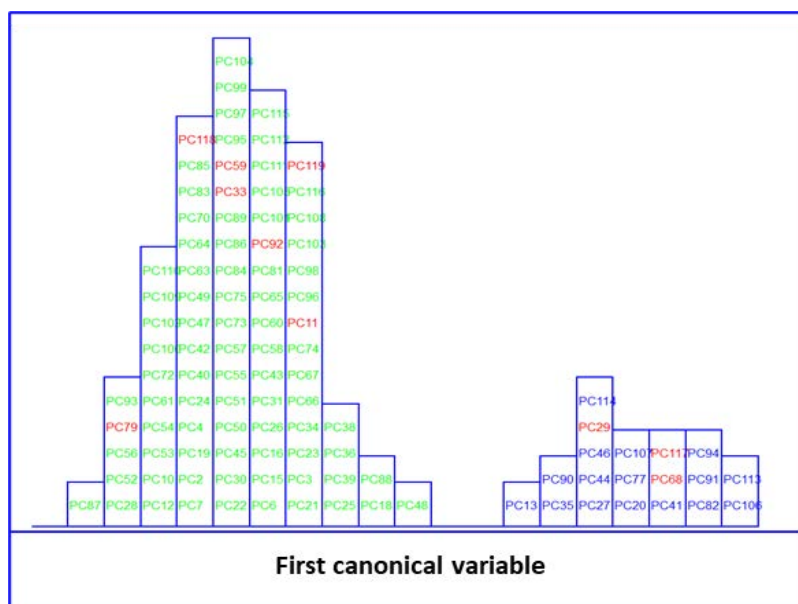
Figure 4-5. Histogram of the first canonical variable for the discrimination of MetS (■) and no MetS (■) patients within the included (■) test set, after performing SELECT-LDA in the stratification approach based on 20 IR variables (y-axis indicates the maximum discrimination power between categories).

4.1.5.4   SIMCA

In an attempt to go one step further in this classification strategy, it was decided to build optimised class models based on clinical parameters and the subset of reduced IR signatures selected by SELECT. SIMCA often outperforms other classification methods,

where a new sample will always be classified in one of the predefined categories. Classification methods such as LDA are based on the development of classification rules and delimiters between classes, whereas in class models, significance limits are built for the specified classes. These limits define the membership parameters for each class; thus, an unknown sample can be classified as not belonging to any defined categories because it is not included in any of its class spaces. SIMCA class modelling uses the number of true/false positives and negatives and statistics, showing the ability of a classification model to recognise class members (*sensitivity* or true positive rate) and showing how good the model is for identifying strangers (*specificity* or true negative rate). Moreover, SIMCA class modelling is often used to describe the class structure of the data set, requiring little or no prior assumptions to build the model.

On applying SIMCA, independent PCA modelling is performed for each class; each sample is fitted in a PCA model to check the separation between classes [41]. This model uses the optimal number of principal components that best describes and groups an individual class. This model can then be used to classify new samples whose class is unknown. The principal components are obtained usually using the NIPALS (non-iterative partial least squares) algorithm after separate autoscaling of the data. Finally, the models built for the different classes are compared by studying their differences and analogies [42]. Each class is modelled independently; thus, it is sensitive to the quality of the data used to generate the principal component models for each class in the training set (at a 5% significance level).

4.1.5.4.1 SIMCA on Clinical Parameters

Herein, SIMCA modelling was performed on five clinical parameters (Table 4-4). A class modelling of five clinical parameters of MetS was built using 4PCs for the inner space of classes, achieving satisfactory results in both internal prediction (LOO) and external prediction 98.95%. SIMCA builds a mathematical model of the category with its principal components and a sample is accepted by the specific category if its distance to the model is not significantly different from the class residual standard deviation. The results of SIMCA modelling can be visually appreciated by a Cooman's Plot, representing the samples' distances against each of the two models. The Cooman's plots were built considering a 95% confidence level to define the class space and the unweighted

augmented distance. This diagram is an effective visual representation that directly indicates the quality of the model constructed with the magnitude of the distance between categories. Thus, the distances to the principal component models and SIMCA approximation in a two-class problem for the class of MetS and no MetS are plotted in Figure 4-6. No clear outliers were observed, but several samples that fall into the joint space of both categories belong mainly to the MetS category. This relatively large number of samples plotted in the class-space common (overlapping) to the two models representing MetS and no MetS patients, as well as the considerable amount of no MetS samples located near their class boundary, suggest potential specificity problems associated with this classification approach based on clinical parameters. Therefore, the distribution of some samples from the MetS category in the area of relative indecision (small left quadrant) could be due to the unequivocal diagnostic parameters defining metabolic syndrome. In fact, these patients have three out of five altered parameters not necessarily similar. In addition, some parameters may be much less marked than others, confounding the decision about their location inside the model.

Table 4-4. The values of discriminant and modelling powers of clinical parameters after SIMCA class-modelling.

Clinical parameters	Discriminant power	Modelling power	
		Category MetS	Category No MetS
Systolic blood pressure	1.99	0.70	0.73
Diastolic blood pressure	2.01	0.70	0.73
Triglycerides	2.18	0.94	0.96
HDL	2.34	0.79	0.94
Glucose	2.36	0.84	0.97
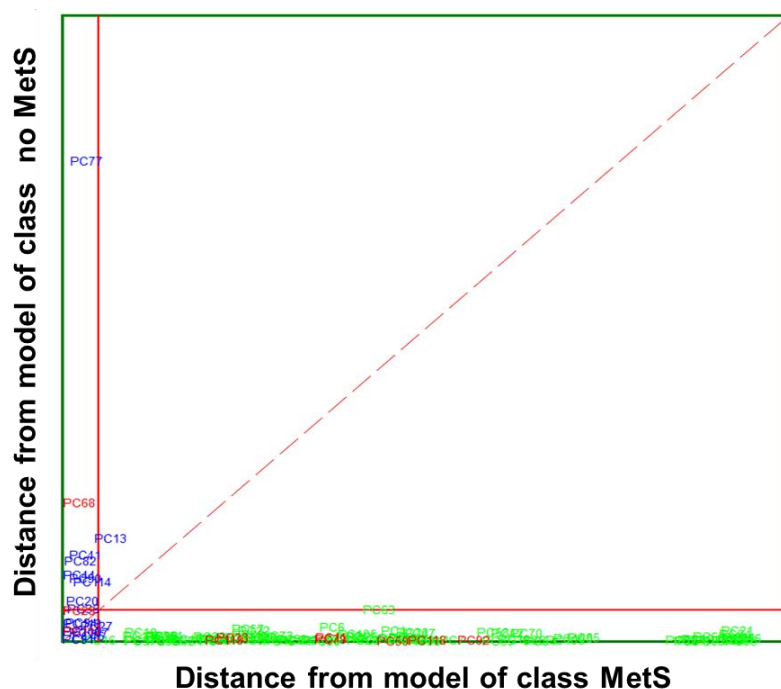
Figure 4-6. Cooman's plot displaying the results obtained by applying SIMCA class-modelling to clinical parameters: MetS (■) and no MetS (■) patients within the included (■) test set. The red solid line indicates a confidence level for class space at 95%. The red dashed line indicates equal class distance.

The data modelling power (MP) and discriminatory power (DP) of the SIMCA class modelling of clinical parameters are presented in Table 4-5. The MP describes how well a variable helps each principal component to model variation in the data, and discriminatory power (DP) describes how well a variable helps each principal component model to classify samples in a training set. The first detail that can be noticed is that, comparably, the MP in no MetS is consistently higher for all parameter pairs. This was expected as the distribution of the values of clinical parameters for each class of patients was significantly different. Nevertheless, the values of TGL have the highest modelling power in both MetS and no MetS categories, with values of 0.94 and 0.96, respectively. This ability of TGL to discriminate between the two groups is justified by previous studies, as metabolic syndrome patients should have significantly higher TGL values. This difference in modelling power is especially remarkable by the measured glucose (0.97 vs. 0.84) and HDL (0.94 vs. 0.79). In addition, clinical parameters such as glucose and HDL also showed significant discriminant power, with values of 2.63 and 2.58, respectively. These two parameters are also perfectly in line with the data collected

from our patients. The MetS group is characterised by high glucose and low HDL values. These same parameters are often responsible for the presence or future development of comorbidities in patients such as diabetes, cardiac disease, and obesity. Other clinical parameters seem to contribute less to the principal component models; indeed, no significant difference was observed in the values distribution of SP or DP between the two categories.

4.1.5.4.2 SELECT-SIMCA on IR Wavenumbers

The best recognition ability (percentage of the samples in training set correctly classified during the modelling step) afforded by SIMCA was achieved by only ten of 20 previously selected wavenumbers by SELECT, providing 98.94% in classification and 95.79% in external prediction, respectively. Interestingly, eight out of ten selected wavenumbers belong to the ''fingerprint region'', which reflects the production of characteristic perturbations in the metabolome and other such variations. The absorption pattern in this area is highly complex; that same inherent complexity makes it unique for each sample and reflects its pathophysiological status. Thus, eight of the selected IR spectral wavenumbers may reflect the current status of the organism and could be directly correlated with the presence or absence of the disease. The results of SIMCA performance applied to clinical variables and to reduced number of IR spectral variables are summarised in Table 5.

Table 4-5. The results of SIMCA class-modelling performance on clinical parameters and ten selected IR spectral variables.

Variables	Classification (%)	LOO (%)	CV Efficiency (%)	Efficiency Forced Model (%)	Total Rate (%)
5 clinical measurements	98.59	97.18	87.05	95.68	100
10 IR selected wavenumbers	97.18	94.37	87.92	97.86	100

A Cooman's plot is presented to show discrimination between the two MetS categories of IR variables (Figure 4-7), where the distance to the PC models for MetS and no MetS are displayed. Compared to the Cooman's plot of clinical parameters, it is observed that there is better separation and discrimination between categories. The Cooman's plot showed a high degree of interclass specificity and a patently clear separation between class models, with a significant improvement from the models constructed from available clinical parameters to those constructed from IR variables. The no MetS patients appear evidently segregated and concentrated forming a dense cluster at large distances from the model of MetS class. Likewise, the vast majority of MetS samples fall clearly and univocally into their class region, far from the class limit for the no MetS model. Furthermore, the single MetS sample located in the inconclusive classification region is virtually placed above the membership threshold.
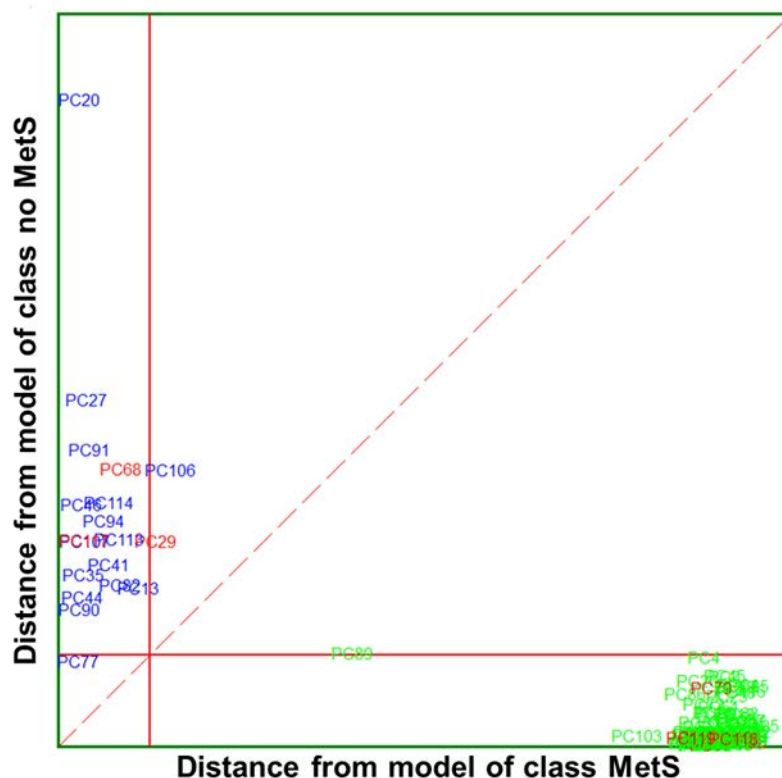
Figure 4-7. Cooman's plot displaying the results obtained by applying the SELECT-SIMCA class-modelling to ten selected IR signals: MetS (■) and no MetS (■) patients within included (■) test set. The red solid line indicates a confidence level for class space at 95%. The red dashed line indicates equal class distance.

From ten selected wavenumbers, the highest discriminant power (5.87) was obtained by the 1133.09 cm^{-1} spectra variable from the ''fingerprint region'' (Table 4-6), followed by 4.31 for 1557.40 cm^{-1} and 4.29 for 2948.94 cm^{-1} from the higher spectral region. The average discriminant power for IR variables is higher compared to DP values obtained with SIMCA modelling of clinical parameters, indicating the increased suitability of the method compared to those using values obtained from clinical measurements. Likewise, the contribution of IR variables to the model variation was of major strength compared to clinical parameters. Thus, all the selected variables contributed equally to marking the difference between MetS and no MetS with an MP equal to 1.00. Furthermore, the distance between classes was 5.19, significantly higher than in the case of SIMCA class modelling applied to clinical parameters (4.26). These results highlight that the proposed method outperformed in accuracy and specificity of the evaluation parameters used in clinical practice. Since the clinical diagnosis of metabolic syndrome lacks standardisation, the results of the obtained model capacity could greatly support clinical

decisions, for example, in terms of exclusion and inclusion evaluation criteria for MetS discrimination.

Table 4-6. Discriminative and modelling powers of ten selected spectra variables after SELECT-SIMCA class modelling.

Wavenumber (cm⁻¹)	Discriminant Power	Modelling Power	
		Category MetS	Category No MetS
2860.22	3.77		
1423.36	4.23		
1562.22	3.66		
1578.61	3.75		
1108.98	3.70		
1316.32	3.64	1.00	1.00
2948.94	4.29		
1557.40	4.31		
1133.09	5.86		
1247.85	3.58		

Our principal aim was to obtain optimal segregation between patients without additional clinical, physical, or ethnic data, and this goal was achieved.

4.1.6   Biochemical Reasoning of Ten Extracted Signals

Herein, we presented a simple, non-invasive, low-cost FTIR-based method for rapid discrimination between MetS and no MetS patients. The use of FTIR spectroscopy is gaining momentum for diagnosis of multiple disorders, from infectious diseases such as hepatitis C and B viruses or malaria to cancers [47–-53]. Due to its ease of use and portability, the potential for using FTIR techniques in clinical environments is within reach. Our strategy extracted the metabolic signatures, instead of individual biomarkers with limited potential, that permit the classification of patients according to molecular patterns. Thus, the FTIR technique provided an overview of spectral changes associated with lipid, protein, or carbohydrate metabolisms.

Ten out of twenty previously selected wavenumbers showed higher discriminant power than clinical parameters. Thus, among these, influential bands at 1578.61, 1562.22, and 1557.40 cm⁻¹ could be assigned to [δ (N-H) + ν (C-H)] of the amide II region of proteins. These discriminative signals may suggest some link with HDL lipoproteins,

which showed significant influence among five clinical factors for the classification of MetS and no MetS subjects. Likewise, the higher absorbance in peaks at 2860.22 cm^{-1} and 2948.94 cm^{-1} could be attributed to CH3 and CH2 sym. stretching of lipids or carbohydrates, which is perfectly congruent with the formulated theories about MetS impairments and their possible implication in the disease. Moreover, as discussed above, TGL and GLU levels seemed to have more influence and variability between the two categories of patients; thus, these attempted assignments properly reflect the actual situation of the patient's metabolism. In addition, the variable at 1133.09 cm^{-1} could be associated with stretching C-O/C-O(H) of carbohydrates or proteins, since it was already shown that the parameters such as glucose or HDL have remarkable modelling and discriminant powers compared to other measured factors.

In this study, the selected spectral biomarkers perfectly reflect the clinical reality of the patient's metabolic profile. Thus, the explanation of the most significant spectral bands confirms the potential of FTIR spectroscopy to deal with such a complex disorder as MetS.

4.1.7 Conclusions

We firmly believe that this alternative analytical strategy could be of great diagnostic relevance and support for clinicians, limiting the time and cost of MetS diagnosis. Moreover, the evaluation of the metabolic profile captures the globality of physiological disturbances, whereas clinical indicators often lack sufficient discriminative power. The results indicate the possibility of rapid application of this strategy to screen for patients with metabolic syndrome. The LDA classifications and SIMCA developed models demonstrated that the spectral variables could provide the same discriminative results as measured clinical parameters. Therefore, why take five measurements when one measurement could provide the same classification ability, greatly stratifying categories of patients? The proposed FTIR method is quick, simple, and non-invasive, and it could be perfectly implemented for large scale-analysis in clinical routines. The principal limitation of this study resides in the relatively tiny sample size at our disposal. In addition, this is a cross-sectional study; therefore, no data on confounding factors (such as gender, age, or diet) were routinely included. The results of a more extensive data set

would be required to strengthen the validity of the adopted classification strategy and lead to a firmer conclusion.

4.2 REFERENCES

1. Saklayen, M.G. The Global Epidemic of the Metabolic Syndrome. *Curr Hypertens Rep* **2018**, *20*, doi:10.1007/S11906-018-0812-Z.

2. Esposito, K.; Chiodini, P.; Capuano, A.; Bellastella, G.; Maiorino, M.I.; Giugliano, D. Metabolic Syndrome and Endometrial Cancer: A Meta-Analysis. *Endocrine* **2014**, *45*, 28–36, doi:10.1007/s12020-013-9973-3.

3. Mili, N.; Paschou, S.A.; Goulis, D.G.; Dimopoulos, M.-A.; Lambrinoudaki, I.; Psaltopoulou, T. Obesity, Metabolic Syndrome, and Cancer: Pathophysiological and Therapeutic Associations. *Endocrine* **2021**, *74*, 478–497, doi:10.1007/s12020-021-02884-x.

4. Esposito, K.; Chiodini, P.; Colao, A.; Lenzi, A.; Giugliano, D. Metabolic Syndrome and Risk of Cancer: A Systematic Review and Meta-Analysis. *Diabetes Care* **2012**, *35*, 2402–2411, doi:10.2337/DC12-0336.

5. Alexandra, K.; Konstantinos, I.; Konstantinos, S.; Alexandros, S.; Michalis, D.; Vasilios, A.; Katsimardou, A.; Imprialos, K.; Stavropoulos, K.; Sachinidis, A.; et al. Hypertension in Metabolic Syndrome: Novel Insights. *Curr Hypertens Rev* **2019**, *16*, 12–18, doi:10.2174/1573402115666190415161813.

6. Isomaa, B.; Almgren, P.; Tuomi, T.; Forsén, B.; Lahti, K.; Nissén, M.; Taskinen, M.R.; Groop, L. Cardiovascular Morbidity and Mortality Associated with the Metabolic Syndrome. *Diabetes Care* **2001**, *24*, 683–689, doi:10.2337/diacare.24.4.683.

7. Federspil, G.; Nisoli, E.; Vettor, R. A Critical Reflection on the Definition of Metabolic Syndrome. *Pharmacol Res* **2006**, *53*, 449–456, doi:10.1016/J.PHRS.2006.03.008.

8. Abebe, S.M.; Demisse, A.G.; Alemu, S.; Abebe, B.; Mesfin, N. Magnitude of Metabolic Syndrome in Gondar Town, Northwest Ethiopia: A Community-Based Cross-Sectional Study. *PLoS One* **2021**, *16*, doi:10.1371/journal.pone.0257306.

9. Motuma, A.; Gobena, T.; Roba, K.T.; Berhane, Y.; Worku, A. Metabolic Syndrome Among Working Adults in Eastern Ethiopia. *Diabetes Metab Syndr Obes* **2020**, *13*, 4941–4951, doi:10.2147/DMSO.S283270.

10. Misra, A.; Khurana, L. The Metabolic Syndrome in South Asians: Epidemiology, Determinants, and Prevention. *Metab Syndr Relat Disord* **2009**, *7*, 497–514, doi:10.1089/met.2009.0024.

11. Huang, P.L. A Comprehensive Definition for Metabolic Syndrome. *DMM Disease Models and Mechanisms* **2009**, *2*, 231–237, doi:10.1242/DMM.001180.

12. Punthakee, Z.; Goldenberg, R.; Katz, P. Definition, Classification and Diagnosis of Diabetes, Prediabetes and Metabolic Syndrome. *Can J Diabetes* **2018**, *42 Suppl 1*, S10–S15, doi:10.1016/J.JCJD.2017.10.003.

13. Alberti, K.G.M.M.; Zimmet, P.; Shaw, J. Metabolic Syndrome--a New World-Wide Definition. A Consensus Statement from the International Diabetes Federation. *Diabet Med* **2006**, *23*, 469–480, doi:10.1111/J.1464-5491.2006.01858.X.

14.   KG Alberti, R.E.S.G.P.Z.J.C.K.D.J.F.W.J.C.L.S.S. Harmonizing the Metabolic Syndrome: A Joint Interim Statement of the International Diabetes Federation Task Force on Epidemiology and Prevention. *Circulation* **2009**, *120*, 1640–1645, doi:10.1161/circulationaha.109.192644.

15.   Reddy, P.; Leong, J.; Jialal, I. Amino Acid Levels in Nascent Metabolic Syndrome: A Contributor to the pro-Inflammatory Burden. *J Diabetes Complications* **2018**, *32*, 465–469, doi:10.1016/j.jdiacomp.2018.02.005.

16.   Syndrome, M. Crossm Metabolic Syndrome and Viral Pathogenesis : Lessons From. **2016**.

17.   O'Neill, S.; O'Driscoll, L. Metabolic Syndrome: A Closer Look at the Growing Epidemic and Its Associated Pathologies. *Obesity Reviews* **2015**, *16*, 1–12, doi:10.1111/obr.12229.

18.   Lee, Y.H.; Pratley, R.E. The Evolving Role of Inflammation in Obesity and the Metabolic Syndrome. *Curr Diab Rep* **2005**, *5*, 70–75, doi:10.1007/s11892-005-0071-7.

19.   Bovolini, A.; Garcia, J.; Andrade, M.A.; Duarte, J.A. Metabolic Syndrome Pathophysiology and Predisposing Factors. *Int J Sports Med* **2021**, *42*, 199–214, doi:10.1055/a-1263-0898.

20.   Fanta, K.; Daba, F.B.; Asefa, E.T.; Chelkeba, L.; Melaku, T. Prevalence and Impact of Metabolic Syndrome on Short-Term Prognosis in Patients with Acute Coronary Syndrome: Prospective Cohort Study. *Diabetes Metab Syndr Obes* **2021**, *14*, 3253–3262, doi:10.2147/DMSO.S320203.

21.   Wiklund, P.K.; Pekkala, S.; Autio, R.; Munukka, E.; Xu, L.; Saltevo, J.; Cheng, S.; Kujala, U.M.; Alen, M.; Cheng, S. Serum Metabolic Profiles in Overweight and Obese Women with and without Metabolic Syndrome. *Diabetology & Metabolic Syndrome 2014 6:1* **2014**, *6*, 1–9, doi:10.1186/1758-5996-6-40.

22.   Esposito, K.; Chiodini, P.; Capuano, A.; Bellastella, G.; Maiorino, M.I.; Rafaniello, C.; Panagiotakos, D.B.; Giugliano, D. Colorectal Cancer Association with Metabolic Syndrome and Its Components: A Systematic Review with Meta-Analysis. *Endocrine* **2013**, *44*, 634–647, doi:10.1007/s12020-013-9939-5.

23.   Shao, Y.; Le, W. Recent Advances and Perspectives of Metabolomics-Based Investigations in Parkinson's Disease. *Mol Neurodegener* **2019**, *14*, 1–12, doi:10.1186/s13024-018-0304-2.

24.   González-Domínguez, R.; García-Barrera, T.; Gómez-Ariza, J.L. Combination of Metabolomic and Phospholipid-Profiling Approaches for the Study of Alzheimer's Disease. *J Proteomics* **2014**, *104*, 37–47, doi:10.1016/j.jprot.2014.01.014.

25.   Alonso, A.; Marsal, S.; Julià, A. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Front Bioeng Biotechnol* 2015, *3*.

26.   Spalding, K.; Bonnier, F.; Bruno, C.; Blasco, H.; Board, R.; Benz-de Bretagne, I.; Byrne, H.J.; Butler, H.J.; Chourpa, I.; Radhakrishnan, P.; et al. Enabling Quantification of Protein Concentration in Human Serum Biopsies Using Attenuated Total Reflectance –

Fourier Transform Infrared (ATR-FTIR) Spectroscopy. *Vib Spectrosc* **2018**, *99*, 50–58, doi:10.1016/j.vibspec.2018.08.019.

27.	Gika, H.G.; Wilson, I.D. Global Metabolic Profiling for the Study of Alcohol-Related Disorders. *Bioanalysis* 2014.

28.	Serkova, N.J.; Standiford, T.J.; Stringer, K.A. The Emerging Field of Quantitative Blood Metabolomics for Biomarker Discovery in Critical Illnesses. *Am J Respir Crit Care Med* 2011.

29.	Finlayson, D.; Rinaldi, C.; Baker, M.J. Is Infrared Spectroscopy Ready for the Clinic? *Anal Chem* 2019.

30.	Lovergne, L.; Lovergne, J.; Bouzy, P.; Untereiner, V.; Offroy, M.; Garnotel, R.; Thiéfin, G.; Baker, M.J.; Sockalingum, G.D. Investigating Pre-Analytical Requirements for Serum and Plasma Based Infrared Spectro-Diagnostic. *J Biophotonics* **2019**, *12*, 1–12, doi:10.1002/jbio.201900177.

31.	Maitra, I.; Morais, C.L.M.; Lima, K.M.G.; Ashton, K.M.; Date, R.S.; Martin, F.L. Attenuated Total Reflection Fourier-Transform Infrared Spectral Discrimination in Human Bodily Fluids of Oesophageal Transformation to Adenocarcinoma. *Analyst* **2019**, *144*, 7447–7456, doi:10.1039/c9an01749f.

32.	Spalding, K.; Bonnier, F.; Bruno, C.; Blasco, H.; Board, R.; Benz-de Bretagne, I.; Byrne, H.J.; Butler, H.J.; Chourpa, I.; Radhakrishnan, P.; et al. Enabling Quantification of Protein Concentration in Human Serum Biopsies Using Attenuated Total Reflectance – Fourier Transform Infrared (ATR-FTIR) Spectroscopy. *Vib Spectrosc* **2018**, *99*, 50–58, doi:10.1016/j.vibspec.2018.08.019.

33.	Roy, S.; Perez-Guaita, D.; Bowden, S.; Heraud, P.; Wood, B.R. Spectroscopy Goes Viral: Diagnosis of Hepatitis B and C Virus Infection from Human Sera Using ATR-FTIR Spectroscopy. *Clinical Spectroscopy* **2019**, *1*, 100001, doi:10.1016/j.clispe.2020.100001.

34.	Kaznowska, E.; Depciuch, J.; Łach, K.; Kołodziej, M.; Koziorowska, A.; Vongsvivut, J.; Zawlik, I.; Cholewa, M.; Cebulski, J. The Classification of Lung Cancers and Their Degree of Malignancy by FTIR, PCA-LDA Analysis, and a Physics-Based Computational Model. *Talanta* **2018**, *186*, 337–345, doi:10.1016/j.talanta.2018.04.083.

35.	Perez-Guaita, D.; Garrigues, S.; de la, M. Infrared-Based Quantification of Clinical Parameters. *TrAC - Trends in Analytical Chemistry* 2014, *62*, 93–105.

36.	Wang, X.; Wu, Q.; Li, C.; Zhou, Y.; Xu, F.; Zong, L.; Ge, S. A Study of Parkinson's Disease Patients' Serum Using FTIR Spectroscopy. *Infrared Phys Technol* **2020**, *106*, 103279, doi:10.1016/j.infrared.2020.103279.

37.	Spalding, K.; Bonnier, F.; Bruno, C.; Blasco, H.; Board, R.; Benz-de Bretagne, I.; Byrne, H.J.; Butler, H.J.; Chourpa, I.; Radhakrishnan, P.; et al. Enabling Quantification of Protein Concentration in Human Serum Biopsies Using Attenuated Total Reflectance – Fourier Transform Infrared (ATR-FTIR) Spectroscopy. *Vib Spectrosc* **2018**, *99*, 50–58, doi:10.1016/j.vibspec.2018.08.019.

38.   Baioumi, A.Y.A.A. Comparing Measures of Obesity: Waist Circumference, Waist-Hip, and Waist-Height Ratios. In *Nutrition in the Prevention and Treatment of Abdominal Obesity*; Elsevier, 2019; pp. 29–40.

39.   Pizarro, C.; Arenzana-Rámila, I.; Pérez-del-Notario, N.; Pérez-Matute, P.; González-Sáiz, J.M. Thawing as a Critical Pre-Analytical Step in the Lipidomic Profiling of Plasma Samples: New Standardized Protocol. *Anal Chim Acta* **2016**, *912*, 1–9, doi:10.1016/j.aca.2016.01.058.

40.   Forina, M.; Lanteri, S.; Oliveros, M.C.C.; Millan, C.P. Selection of Useful Predictors in Multivariate Calibration. *Anal Bioanal Chem* **2004**, *380*, 397–418, doi:10.1007/s00216-004-2768-x.

41.   Pizarro, C.; Esteban-Díez, I.; Arenzana-Rámila, I.; González-Sáiz, J.M. Discrimination of Patients with Different Serological Evolution of HIV and Co-Infection with HCV Using Metabolic Fingerprinting Based on Fourier Transform Infrared. *J Biophotonics* **2018**, *11*, 1–12, doi:10.1002/jbio.201700035.

42.   Pizarro, C.; Esteban-Díez, I.; Arenzana-Rámila, I.; González-Sáiz, J.M. Discrimination of Patients with Different Serological Evolution of HIV and Co-Infection with HCV Using Metabolic Fingerprinting Based on Fourier Transform Infrared. *J Biophotonics* **2018**, *11*, 1–12, doi:10.1002/jbio.201700035.

43.   Pizarro, C.; Esteban-Díez, I.; Espinosa, M.; Rodríguez-Royo, F.; González-Sáiz, J.M. An NMR-Based Lipidomic Approach to Identify Parkinson's Disease-Stage Specific Lipoprotein-Lipid Signatures in Plasma. *Analyst* **2019**, *144*, 1334–1344, doi:10.1039/c8an01778f.

44.   Tkachenko, K.; Espinosa, M.; Esteban-Díez, I.; González-Sáiz, J.M.; Pizarro, C. Extraction of Reduced Infrared Biomarker Signatures for the Stratification of Patients Affected by Parkinson's Disease: An Untargeted Metabolomic Approach. *Chemosensors* **2022**, *10*, doi:10.3390/chemosensors10060229.

45.   Cocchi, M.; Biancolillo, A.; Marini, F. Chemometric Methods for Classification and Feature Selection. In *Comprehensive Analytical Chemistry*; Elsevier B.V., 2018; Vol. 82, pp. 265–299 ISBN 9780444640444.

46.   M. Forina, S. Lanteri, C. Armanino, M.C.C. Oliveros, C.C. V-Parvus 2003.

47.   Forina, M.; Oliveri, P.; Casale, M. Complete Validation for Classification and Class Modeling Procedures with Selection of Variables and/or with Additional Computed Variables. *Chemometrics and Intelligent Laboratory Systems* **2010**, *102*, 110–122, doi:10.1016/j.chemolab.2010.04.011.

48.   Brown, S.; Tauler, R.; Walczak, B. *Comprehensive Chemometrics*; 2010; ISBN 9780444527011.

49.   van der Greef, J.; Smilde, A.K. Symbiosis of Chemometrics and Metabolomics: Past, Present, and Future. *J Chemom* **2005**, *19*, 376–386, doi:10.1002/cem.941.

Chapter 5

Parkinson's disease

Abstract

In this chapter, Parkinson's disease is investigated. The reported results are referred principally to two primary studies conducted on PD -using FTIR spectroscopy and UPLC-MS. Thus, in the first section, the potency of FTIR coupled with a chemometric strategy based on a three-step classification approach in the stratification of Parkinson's patients is discussed. Thus, the disease was effectively classified and differentiated from the control group and other impairments such as Alzheimer's dementia. Spectral signatures in human plasma have been successfully identified for differentiation between patient categories by selecting significant wavenumbers closely related to PD pathogenesis and metabolic biomarkers. In addition, the discrimination results in both sub-classification problems, succeeding in the stratification of patients with different PD stage progression profiles and those with different dementia type profiles are reported. All the speculations about the involvement of selected bands in the pathogenesis of PD are immensely reasonable. They seem to be confirmed by the second section of the results performed by the UPLC-MS untargeted approach. Thus, this chapter's second section focuses on the specific lipid biomarkers identified for PD discrimination from healthy controls and AD patients.

Resumen

En este capítulo se investiga la enfermedad de Parkinson. Los resultados conseguidos se refieren principalmente a dos estudios realizados en PD, utilizando espectroscopía FTIR y UPLC-MS, respectivamente. Así, en la primera sección, se discute la potencia de FTIR acoplado con una estrategia quimiométrica basada en un enfoque de clasificación a tres pasos en la estratificación de pacientes con Parkinson. Por lo tanto, la categoría enfermedad de Parkinson se clasificó de manera efectiva y se diferenció del grupo de control y otras afecciones como la demencia de Alzheimer. Se identificaron con éxito firmas espectrales en plasma humano para la diferenciación entre categorías de pacientes mediante la selección de números de onda significativos estrechamente relacionados con la patogénesis de PD y biomarcadores metabólicos. Además, se muestran los resultados de discriminación en ambos problemas de subclasificación, logrando la estratificación de pacientes con diferentes perfiles de progresión de etapas de PD y aquellos con diferentes perfiles de tipo de demencia. Todas las especulaciones sobre la contribución de bandas seleccionadas en la patogénesis de PD son inmensamente razonables y pueden ser confirmadas por la segunda sección de los resultados realizados por el enfoque no dirigido de UPLC-MS. Así, la segunda sección de este capítulo se centra en los biomarcadores específicos de lípidos identificados para la discriminación de PD, de controles sanos y pacientes con AD

5   Chapter 5. Parkinson's disease

5.1   Extraction of reduced infrared biomarker signatures for the stratification of patients affected by Parkinson's disease: an untargeted metabolomic approach

5.1.1   Introduction

A considerable segment of the ageing population worldwide suffers from Parkinson's disease, the second most prevalent progressive neurological disorder after Alzheimer's dementia [1–3]. The risk factors for PD are complex and likely interconnected, so the onset of PD is thought to be caused by a combination of genetic predisposition and environmental influences (exposome). Unfortunately, there is still no standard treatment for this disorder. Most of the currently available therapeutic options are focused on treating and mitigating the symptoms, i.e., on palliative care, but, curative treatment is not yet a medical reality, with its consequent huge impact on morbidity and mortality.

All findings towards detecting the disorder's pathogenesis suggest various metabolites that are modified in PD, from a significant role attributed to α-synuclein [4–6] and Lewy's body [7] to exosome factors that are involved [8–10]. Thus, Parkinson's disease is often associated with defects in lipids metabolism particularly in the central nervous system [11,12]. It was reported that oxidative stress in the substantia nigra at the time of death in advanced Parkinson's disease manifests in increased lipid peroxidation [13]. Blood-based biomarkers are also widely studied because they can be more accurate than clinical observations regarding dopamine deficiency effects, such as bradykinesia, rigidity and tremors. In addition, it was suggested that changes in cholesterol levels and cholesterol derivatives might indirectly be related to the onset of PD.

Meanwhile, higher uric acid (UA) levels were associated with a decreased risk of disorder, and ratio acid uric/creatinine (UA/Cr) was evaluated as a predictive factor for a slower disease progression [14,15]. Elevated blood glucose levels were directly related

to a longer duration of PD and a higher score of dysautonomia in moderate to advanced PD patients [16]. Interestingly, increased concentration of histamine in the nervous system and putamen of PD patients were reported from different research groups [17, 18]. Despite a large shred of evidence, no specific biomarker was approved for diagnosis or prognostic purposes of this progressive multisystem condition [19–24]. Thus, the conventional diagnosis of Parkinson's remains essentially clinical based on the subjective observations of clinicians. Nowadays, none of the available clinical tests has been proven to comply with high sensitivity, accuracy, and objectivity for PD detection. In addition, the instruments applied in the clinical environment are expensive and unwieldy.

Nevertheless, the correct diagnosis identifying PD at the early stage is crucial. This is because once the patient appears with clinical symptoms, the damage in the brain is already irreversible. For this reason, the diagnosis must be made as soon as possible to avoid further extensive neuronal loss [25].

In this context, the attention to specific biomarkers shifted to explore the production of perturbations in the metabolome and such variations in studies of Parkinson's disease [26–29]. Untargeted metabolomic studies, such as that here presented, aim to extract the metabolic fingerprints of analysed samples to classify them according to biological status or origin based on these unique and individual molecular patterns.

For more than ten years, the vibrational approach, nuclear magnetic resonance (NMR) [30, 31] spectroscopy and a wide range of highly sensitive mass spectrometry (MS) [32,33] based methods have been proven to be a valuable ally for the evaluation and classification of normal and pathological samples [34–39], especially for Parkinson's detection [40–42]. However, despite the large number of studies found in the literature that support the emerging potential of metabolic fingerprinting in clinics, a still challenging bottleneck of these types of studies is their actual translation to clinical practice. From a technical point of view, breaking down barriers to clinical translation depends on advances in measurement technology. Being these instruments rapid, non-invasive, non-destructive, reliable and easy to use, they are perfect candidates as high-throughput screening techniques for fingerprinting. In this context, both vibrational and NMR spectroscopy remains the prime analytical choice because they perfectly fulfil the

required characteristics of analytical instrument, making large-scale or routine studies much more feasible than with MS-based applications [43].

Likewise, exhaustive metabolomic fingerprinting research should not be limited to a unique analytical platform but rather should test and combine multi-analytical strategies in order to exploit their respective strengths and overcome their weaknesses. To this end, Fourier-transform infrared (FT-IR) and NMR spectroscopies can successfully complement each other.

In previous work from our research group [44] we evaluated the potential of using a non-targeted lipidomic approach to extract NMR-based signatures for the clinical differential diagnosis and stratification of PD. To complete and complement this work, we now proposed a metabolomic fingerprint classification strategy also aimed at extracting disease/stage-specific panels of infrared markers but expanding the search to a broader range of molecular species (not restricted to lipid compounds) thanks to the inherent ability of FT-IR to provide biochemical information holistically.

Moreover, IR is particularly suitable for analysing human biofluids, such as blood, that are easily obtainable and reflect several physiological functions of the body [29, 45, 46]. The blood is the primary carrier of metabolites throughout the entire organism, and it is composed of a variety of biological materials, mainly proteins, lipids, sugars. All these are active in the infrared range, and each biomolecule is determined by its unique structure. The changes in their chemical structure can be investigated simultaneously instead of studying isolated molecules, providing a metabolic signature for PD. Thus, the spectrum recorded by FT-IR from a biological sample generates a unique IR spectral signature, reflecting its specificity. Furthermore, the IR spectral modes of plasma may reflect the current status of the organism and could be directly correlated with the presence or absence of the disease. For these reasons, IR is highly exploited to identify possible spectra biomarkers associated with Parkinson's differentiation [47].

5.1.2 Aim of the study

Given these perspectives, in this study, the direction of research shifts towards investigating the existence of distinct mid-infrared metabolic fingerprints in PD-related diseases, which would drive PD patient stratification and would guide an accurate and

early differential diagnosis. Thus, an untargeted metabolomics approach (FT-MIR application in the fingerprint region coupled with multivariate data analysis) was used to reveal spectroscopic biomarker signatures that define patient subgroups for the clinical diagnosis and classification of PD at different stages of the disease. PD at the initial stage (PDI) should be differentiated and not confused with developed PD-related dementia (PDD). Ideally, an accurate analytical tool should be able to differentiate Parkinson's disease from other neurological impairments such as Alzheimer's disease. In most cases, subjects affected with PD share a common profile accumulating abnormally aggregated proteins with Alzheimer's disease. The two main chemometric strategies employed in this work are based on the initial selection of discriminant variables and the subsequent development of a linear discriminant analysis (LDA) classification. One of our primary goals was to achieve accurate discrimination between plasma samples of patients affected by PD from subjects affected by AD and from healthy control individuals, confirming that infrared signatures can be associated with metabolomic changes based on different pathological conditions. The second major objective was to obtain accurate discrimination between two PD subgroups, identifying reliable wavenumber predictors for disease stage differentiation. However, we also decided to deepen the problem and apply the developed classification rules to a new problem. The differentiation between patients with Parkinson's developed dementia and those affected by Alzheimer's dementia was also studied. Within this method, we aimed to reveal that LDA stepwise wavenumber selection may extract significant bio-spectroscopic markers, enable objective diagnosis and make possible the differentiation of such a multifactorial disease as Parkinson's.

Considering the potential clinical translation advantages derived from adopting an IR fingerprint-based classification, the cost-effectiveness and relative ease of access to IR portable devices should be pointed out, which could be ideal for point-of-care testing, primary health care or wherever required.

5.1.3   Materials and Methods

5.1.3.1   Study population

The present study involved 97 patients whose plasma samples were submitted to the Molecular Neurobiology Unit in the Centre for Biomedical Research of La Rioja (CIBIR) from the Neurology Department of San Pedro Hospital (Logroño, Spain). The patient cohort was sub-grouped into different classes depending on the type of neurodegenerative pathology. A total of 41 patients were classified as PDI, and nine were considered to have PDD. There were 23 patients in the group with AD and 24 healthy controls (HC), belonging to the family environment and being alike in age to PD patients enrolled in the study. Ethical approval was granted by the Research Ethics Committee of the Hospital San Pedro of La Rioja, and individual informed consent was obtained from all those taking part.

5.1.3.2   Collection and handling of plasma samples

Plasma samples were provided by La Rioja Blood Bank. Venous blood samples were drawn via antecubital venepuncture from each subject in a sitting position. Becton Dickinson (BD) Vacutainer® plastic blood collection tubes with K2EDTA were used for plasma separation. Blood was processed to obtain plasma by centrifugation at 2200 g for 15 min at 4 °C. Recently delivered plasma samples to the laboratory were immediately frozen and then stored in Eppendorf tubes as aliquots of 200 μL each at −80 °C until further use. Prior to FTIR analysis, plasma thawing was performed according to an optimised ultrasound-based protocol for lipidomic analyses recently developed in our research group [48].

5.1.3.3   Instrument

Plasma samples were measured by Spectrum-One ABB Miracle Type MB3000 FT-IR Spectrophotometer (Zurich, Switzerland). Each plasma sample (25 μL) was manually spotted onto a CaF_2 windows liquid cell PerkinElmer (Omni Cell, Specac Ltd., Orpington, UK) with a 50 μm Mylar spacer. The sections were recorded in the medium infrared range. Data points in the range of 4000-300 cm^{-1} were collected with a resolution of 2 cm^{-1}, and 32 scans were accumulated and averaged. All measurements were made in triplicate. A mean spectrum was subsequently obtained from the replicates recorded for

each plasma sample. The sample temperature was maintained at 23.0 ºC ± 1.0 ºC while recording the signals, and a constant nitrogen purge was applied to remove atmospheric water vapor and CO_2. Data analysis was performed using the Horizon MBTM program. During the analysis, it was necessary to use quality control (QC) samples to check the methodology's performance and test its reproducibility. Therefore, QC samples were processed similarly to the actual samples and were inserted regularly.

5.1.3.4 Pre-processing of spectra

FT-MIR dataset was processed and analysed with Parvus [49] and Unscrambler 11 chemometric software package (version 11.0, Camo Software, Oslo, Norway). Pre-processing is imperative in analysing high-dimensionality biological spectral data; it corrects many problems with spectral data acquisition such as random noise, baseline distortions or light scattering. Spectra were then cut to include the biochemical ''fingerprint region'' between 1490-1155 cm^{-1}; other regions were excluded from further analysis as non-informative zones. Often spectral wavenumbers have solid correlations and, therefore, highly amenable [50]. Finally, MIR-spectra data was submitted to Extended Multiple Scatter Correction (EMSC) pre-treatment. This method was preferable as a pre-processing step allowing for resolution of overlapping peaks, and it showed better results in decreasing scatter effects. The data were always centered before multivariate analysis.

5.1.4 Data analysis

5.1.4.1 PCA

Chemometric methods are increasingly applied to obtain meaningful and reliable information from the registered spectra, enabling their characterisation and enhancing process understanding. Principal component analysis (PCA) is one of the most useful preliminary steps for the exploratory analysis of a large number of correlated features such as FT-MIR spectral data. This unsupervised data analysis step was performed to reduce the dimensionality of our complex spectral dataset into a few Principal Components (PCs), still preserving the majority of information and capturing relevant sources of data variability. However, this step alone was not enough to allow a clear separation of our data. However, it helped to identify outliers. Furthermore, it

highlighted the need to resort to a two-step sequential classification strategy based on applying an efficient variable selection technique to find the informative features that would allow successful prediction.

Before setting any classification approach, an important parameter should be considered: the study's objective, the number of categories, and the particular requirements of a sample assigned to a specific class. The proposed classification strategy was seen as a sequence of two consecutive classification problems in this study. The approach used for classification first concerns a generic classification of the disease and then specific discrimination of PD to distinguish the two different stages of the disease. Thus, for a first global classification, three distinct categories have been defined (Parkinson's disease, Alzheimer's disease and control subjects) so that patients with PD at an early stage and those with PD-related dementia were considered as one class (Parkinson's disease). Preliminary investigations have shown that this approach considerably improves the performance of the classification method. Then, after reaching a clear differentiation between the above three previously designated classes, PD has been divided into two classes to further classify these patients according to the severity of the disease. Besides, a strategy to differentiate two types of dementia affecting patients with separate neurodegenerative disorders was investigated and applied.

5.1.4.2   Classification methods

Classification methods aim to compute classification rules to assign objects to one of the categories of the problem being studied (unlike class modelling techniques, there is no possibility of non-classification), the objects are always classified in one of the predefined classes [51].

The linear discriminant analysis (LDA) classification [52, 53], well known and extensively applied powerful supervised chemometric classification technique, enabled us to discriminate between prespecified subgroups. A stepwise orthogonalization of predictors (SELECT) optimised the discrimination performance by selecting the most significant wavenumbers for a reliable classification between patients' classes to use later as an input for a stepwise LDA. Therefore, to gain a clear understanding of the

procedure carried out by SELECT, a careful reading of the original paper is particularly recommended [49].

One crucial step in LDA-based variable selection is choosing the number of variables to be used in classification development. The optimal complexity of each classification developed was assessed by internal cross-validation. To control and avoid possible overfitting, the maximum number of variables to be retained by SELECT was limited so that the number of training objects would be at least three times greater than the number of finally selected wavenumbers. Therefore, essential variables were selected and decorrelated with other variables based on the maximum correlation weight. This step is essential when dealing with big data dimensionality, eliminating the futile features due to noise and identifying the relevant and important variables to be applied in the following steps. Cross-validation (CV) was used to optimize classifications, whereas external validation to evaluate prediction ability, respectively. The quality of the discriminant rules derived was evaluated according to several parameters:

– Total classification (prediction) rate (TR)

$$TR = \frac{\sum_c m_{cc}}{N}$$

– Category c rate (R_c)

$$R_c = \frac{m_{cc}}{N_c}$$

These equations were applied in both classification and (internal/external) prediction, where m_{cc} is the number of correct classifications (predictions) for a certain category c, $\sum_c \boldsymbol{m_{cc}}$ is the total number of correct classifications (predictions), N_c is the number of classifications (predictions) for a certain category c, and N is the total number of classifications (total predictions). Note that N_c is not always equal to the number of samples belonging to class c, just as N does not always represent the total number of objects, since, for instance, an object can be classified several times during cross-validation. Scores plots for the LDA canonical variables, were also used to analyse the goodness of the results.

5.1.5 Limit of the study

A primary limitation of the current work relies on the relatively small number of cases available for subsequent infrared analysis. Further research, entailing the analysis of a larger patients' database, would be required to confirm the proposed classification strategy's validity and generalisation ability based on reduced IR signatures. However, the validation methodology followed in this study, which was based on both internal cross-validation (to develop optimal classifications) and external validation (to assess the actual predictive performance of the classifications constructed), served to prevent overfitting and ensure the reliability of the results obtained despite the limited number of samples in our cohort.

5.1.6 Results and Discussion

Using plasma-based vibrational spectroscopy, we achieved results with significant discrimination relevance at multiple levels. FT-MIR spectroscopy coupled with chemometric strategy has been proven to detect differences between pathological patients and healthy subjects and between different stages of neurodegenerative disease. As contributors to the discrimination between patients, several selected wavenumbers appeared to be of particular interest. Thus, we performed a tentative biochemical assignment of the selected signatures for each classification step. It should be made clear that the tentative band assignment and its speculative nature was performed to attribute a possible chemical reasoning to the reduced fingerprint bands responsible for each stratification. Therefore, the attempt band assignment was not, at any time, the primary objective of this study. However, chemical reasoning could be a starting point in future targeted-based studies.

Thus, numerous selected wavenumbers in the fingerprint region were associated with key molecules, from carbohydrates to nucleic acids. Our results are reassuring and utterly consistent with the formulated theories about PD pathogenesis and metabolic biomarkers. Furthermore, the spectral absorptions implicated in the discrimination between disease principal groups and subgroups were quite similar. Sometimes, the exact wavenumber was associated with a different absorption-type but always in line with the possible contribution found in the literature.

5.1.7 FT-MIR spectral profiles

The fingerprint region (the segment of the IR spectrum below 1500 cm^{-1}) is a complex area that comprises many bands arising from the fundamental vibrations of many significant biomolecules overlapping each other. Many compounds generate a characteristic pattern of absorption that is unique to a sample [47]. Different vibrational modes in molecules ''functional groups'' create a very complex pattern of absorptions that contain a huge amount of valuable information about the metabolic changes that occur during the onset and progression of the disease [54, 55]. Only one spectral zone between 1490 cm^{-1} and 1155 cm^{-1} was considered for the analysis and was shown in the corresponding plots Figure 5-1. Thus, it seems likely that changes in band profiles in this region could be related to alteration in the metabolism induced by neurodegenerative diseases.
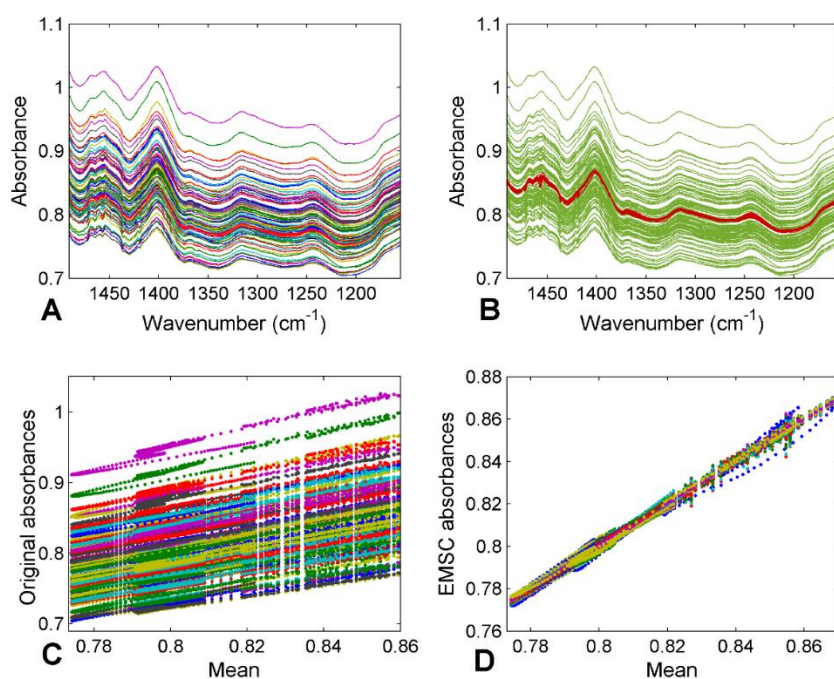
Figure 5-1. FT-IR spectra of plasma samples before (**A**) and after (**B**) EMSC pre-treatment, and the corresponding plots of scattering effects between signals before (**C**) and after (**D**) correction.
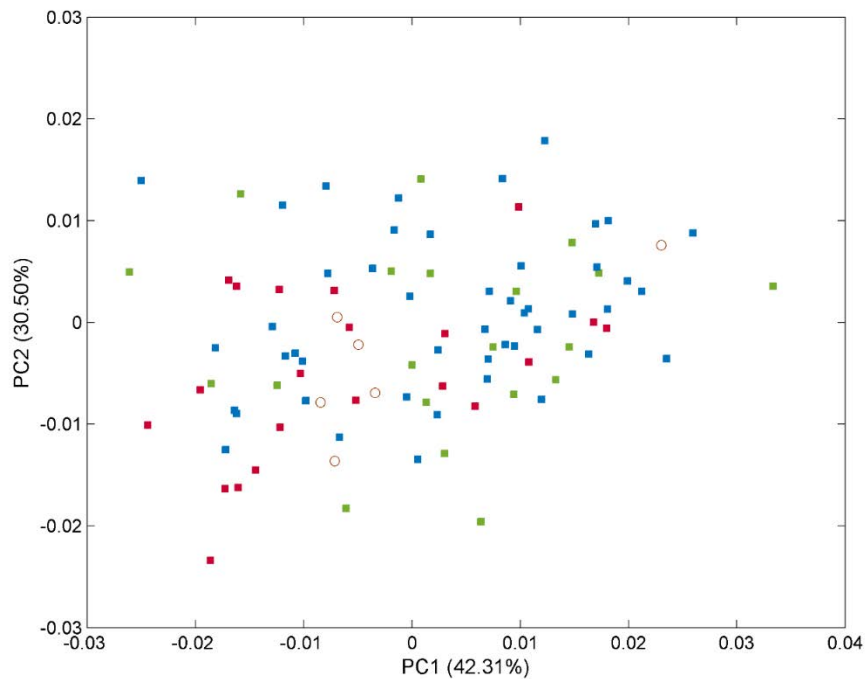
Figure 5-2. Scores for the plasma samples from the 97 patients (full data set) on the first 2 principal components explaining the variability in the IR spectral data. The samples are labelled according to their specific pathology: healthy control (■), PD (■), AD (■) subjects and external test samples (O).

5.1.7.1 *Global classification: 3-class approach for discriminating patients with PD, AD and healthy controls*

Based on the results of the exploratory analyses, considering the classification problem studied here as a unique step approach was rejected as it would lead to unreliable results Figure 5-2. For this aim, a multivariate classification approach based on a multiple-step sequential classification has been studied and applied to a patient population.

Firstly, three separate classes were defined; the group of Parkinson's patients at the early stage was combined with patients accompanied by dementia. Thus, the first global classification approach concerns healthy subjects, PD (PDI + PDD) and AD. The best solution for the global classification problem was obtained when the SELECT-LDA method was applied, so the information responsible for the successful discrimination between PD, AD and the control group was compressed from 340 to only 30 variables

131

(conforming to a reduced IR signature of disease status) with almost 100% of total correct assignment rates in classification and prediction (Table 5-1). Classification rates broken down per category revealed excellent discrimination among classes, providing a 100% level of correctly classified samples for control subjects and patients with Alzheimer's, respectively, and 99.79% correct assignment rates in the case of patients with Parkinson's disorder. Satisfactory internal prediction performances ranging from 86.36 to 95.24 % were achieved for the various categories, but the real strong point consisted in external prediction. The prediction performance of the SELECT-LDA classification was developed and optimized using ten cancellation groups for CV and a test subset of six samples, distributed randomly in the following way: two for the control group, two for PD and two for AD. As a result, all six test samples were correctly classified, achieving 100%, therefore confirming the reliability of the developed classification strategy based on the reduced IR fingerprint extracted wavenumbers (Table 5-2). Furthermore, a clear interclass separation reached between the three classes of patients can also be visually appreciated in the corresponding plot of LDA score differences (Figure 5-3).

Table 5-1. Percentages of correctly classified samples/patients in both classification and internal/external validation corresponding to the SELECT-LDA performed when addressing the global classification approach.

3-class approach SELECT-LDA: global classification with 30 variables			
	Classification %	**Prediction (CV 10) %**	**External Prediction %**
PD (I+D)	100.00	86.36	100.00
AD	99.79	89.58	100.00
HC	100.00	95.24	100.00
Total rate	99.89	90.11	100.00

Table 5-2. Discriminant wavenumbers (in order of selection) corresponding to the SELECT-LDA classification developed from IR spectra of plasma samples in the global approach.

| 1st-step differentiation approach: 30 biomarkers			
3 global categories (PD, AD and HC) differentiation			
Selection order	Wavenumber(cm⁻¹)	Selection order	Wavenumber(cm⁻¹)
1	1489.9008	16	1336.5712
2	1171.6696	17	1289.3187
3	1316.3201	18	1335.6069
4	1377.0734	19	1253.6382
5	1319.6382	20	1182.2773
6	1312.4628	21	1203.4927
7	1284.4970	22	1474.4714
8	1271.9606	23	1302.8194
9	1266.1746	24	1455.1847
10	1215.0647	25	1294.1404
11	1156.2402	26	1334.6425
12	1159.1332	27	1438.7909
13	1443.1286	28	1477.3644
14	1286.4257	29	1424.3259
15	1288.3544	30	1403.1105
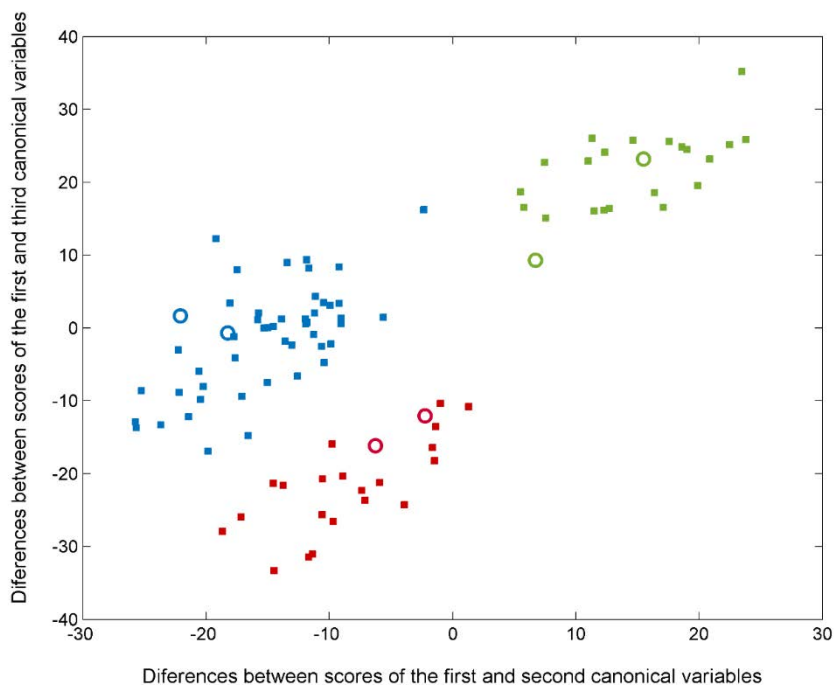
Figure 5-3: Plot of the differences between discriminant scores for plasma samples after performing SELECT-LDA in the global classification approach. The 3 categories considered were labelled as: healthy control (■), PD (■), AD (■) subjects. Test samples are displayed as unfilled circles (O).

The patient profile of three main categories seems to be unequivocally distinguished by the score difference between the first and third canonical variables. Likewise, the maximum difference between healthy subjects and two groups affected by neurodegenerative disorders (PD and AD) is evidenced by the score difference between the first and third canonical variables, suggesting that the two disease-carrying groups do not share a similar profile with the healthy group. In particular, it can be observed that there is a double contribution of both differences between scores; thus, two classes characterized by pathology are separated by an imaginary axis coinciding with the delimiter.

5.1.7.2 *Chemical reasoning about possible contribution of selected wavenumbers in the global classification*

In this first classification step, several selected wavenumbers appear to be of particular interest; thus, we performed an attempt biochemical assignment of the selected variables. Different bands comparing PD, AD and HC were selected between 1150-1000 cm^{-1}, which is usually associated with oxidative stress [39]. Therefore, it could be attributed to different levels of damage caused by free radicals. Many studies have shown that UA plays an essential role as an antioxidant reagent, mainly as the urate in the human body. Furthermore, increased levels of this endogenous compound are linked with reduced brain damage caused by reactive oxygen species (ROS) [14, 15]. The topic of oxidative stress joins many other shreds of evidence that will be discussed further on.

Close to this region, bands at 1156 and 1159 cm^{-1}, or 1302 cm^{-1} could be attributed to bending (CH) of Ala. Likewise, selected signatures between 1411-1400 cm^{-1} could correspond to symmetric stretching (ṽ sym) (COO-) and ṽ sym. (C=O) of Ala or Glu. Various successful breakthroughs have shown metabolite variations in blood samples of PD. Significant variations of glutamate (Glu) and alanine (Ala) in cerebrospinal fluid were also observed in PD and have been broadly targeted [27].

Discriminant signatures around 1215/1253 cm^{-1} due to asymmetric stretching (ṽ ass) (P=O) of phospholipids; stretching vibration (COO-) of fatty acids at 1403 cm^{-1} and bending vibrations (δ) (CH$_2$) at 1455, 1474, 1477, 1489 cm^{-1} of lipids, could be attributed to controversial theories made about the role of cholesterol, lipids and proteins in

Parkinson's disease [56]. Furthermore, different research findings supported the associations with pathological interactions of alpha-synuclein in PD. Thus, the selected protein vibrations also seem to play a key role in patients' stratification [6]. Multiple bands, such as 1312, 1403, and 1455 cm^{-1}, were assigned to lipids perfectly congruent with previous findings.

In this classification strategy, one of the bands was assigned to histidine, namely the signature at 1215 cm^{-1}, proving some evidence found by Picca et al. [10] where higher concentrations of 3-methyl-histidine, citrulline, and serine were determined in control participants. At the same time, the band at 1182 cm^{-1} could correspond to bending (C-OH) and stretching (C=O) of serine for the differentiation of three main categories of patients. The role of serine is of great importance, as it takes part in different metabolic pathways, including the generation of phosphatidylserine and phosphoserine, where both play an important role in the function of the neurodegenerative system [57]. In addition, other studies suggested that histamine could potentially affect neuronal survival and participate in neurotrophic processes aimed to re-establish damaged brain functions [17,18]. From the analysis of this first-step classification, different wavenumbers resulted significantly in the region between 1360-1220 cm^{-1}. Therefore, we could speculate that these spectral markers could be attributed to amide III-band stretching and bending and corresponding to histamine.

Many of the selected spectral biomarkers hold the potential to be linked with carbohydrates. For this purpose, it seems particularly important that many influential bands in Parkinson's classification were also attributed to ''sugars''. Thus, from 30 spectral variables, bands at 1156 cm^{-1} and 1159 cm^{-1} could be due to stretching (CO-O-C) of carbohydrates. The region 1330-1220 cm^{-1} seems to be rich in methylene stretching of carbohydrates residues, and therefore, the band at 1424 cm^{-1} could correspond to polysaccharides. Other studies had already evidenced a substantial contribution to the separation between metabolite profiles of unmedicated PD patients and controls of alternative metabolites, such as myoinositol, sorbitol, citrate, acetate, succinate and pyruvate [27].

5.1.7.3 PD stratification: 2-class approach for discriminating between patients with early-stage PD and PD-related dementia

After the differentiation among three classes of patients was obtained, we carried out a sub-analysis, comparing the group of patients affected by Parkinson's disease depending on the disease's progress. Therefore, a single undifferentiated group of Parkinson's was split up into two distinct classes: PD at an early stage and PD related dementia. The classification strategy followed the same classification rules as described ahead [52]. However, in compliance with cited before rules, the number of discriminant variables had changed that training objects were at least three times greater than the number of final selected wavenumbers. Full CV and a test subset of 3 samples, selected randomly as two for the PDI group and one for PDD, were performed to optimise and validate classifications, always on autoscaled data. In this case, 100% in classification and 100% in prediction (internal and external) were achieved (Table 5-3). Herein, the number of features was reduced from 340 to only 15 variables, which makes our approach even more remarkable (Table 5-4). A discriminative histogram is reported in Figure 5-4, which shows a clear class separation on the first canonical variable. This feature proves an outstanding performance of the method and its reliability and good performance of the classification strategy based on reduced MIR–plasma signatures.

Table 5-3. Percentages of samples/patients classified correctly in both classification and internal/external validation in the SELECT-LDA performed when addressing the classification approach for PD stratification

2-class approach SELECT-LDA: Parkinson's differentiation with 15 variables			
	Classification %	**Prediction (LOO) %**	**External Prediction %**
PDI	100.00	100.00	100.00
PDD	100.00	100.00	100.00
Total rate	100.00	100.00	100.00
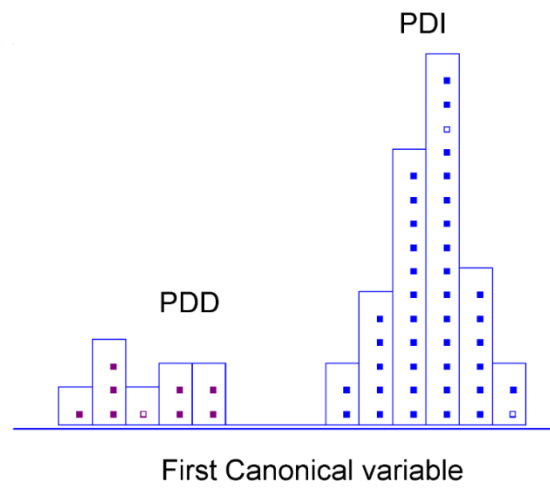
Figure 5-4. Histogram of canonical variable for the discrimination of PD early stage (■) and PD-related dementia (■) patients after performing SELECT-LDA in the PD stratification approach.

Table 5-4. Discriminant wavenumbers (in order of selection) corresponding to the SELECT-LDA classification developed from IR spectra of plasma samples for the differentiation of PD progression stage.

2nd-step differentiation approach: 15 biomarkers	
2 categories (PDI and PDD) differentiation	
Selection order	**Wavenumber (cm-1)**
1	1294.1404
2	1292.2117
3	1437.8266
4	1435.8979
5	1443.6126
6	1475.4357
7	1297.0334
8	1476.4001
9	1342.3572
10	1170.7052
11	1171.6696
12	1226.6368
13	1445.5413
14	1224.7081
15	1214.1004

5.1.7.4 Chemical reasoning about possible contribution of selected wavenumbers in the second-step classification

Thus, among 15 relevant wavenumbers, bands discriminating PDI and PDD at 1214, 1224, and 1226 cm-1 were associated with uric acid ring vibrations, presumably explaining the different levels of brain damage during the disease progression stage.

The selection of different wavenumbers at 1443, 1445, 1475 and 1476 cm-1 due to lipid structures could suggest oxidative stress in the substantia nigra at the time of death in advanced Parkinson's disease manifests in terms of increased lipid peroxidation, superoxide dismutase activity, and zinc levels [13]. In addition, previous studies have already performed the analysis with infrared spectroscopy defining bands due to methylene deformation of lipids or methyl bending of lipids, able to differentiate PD and controls [42]. Thus, our results prove the importance of this region.

Different studies highlighted the role of phosphoethanolamine, which is the head group of different lipids, including phosphatidylethanolamine, lysophosphatidyl-ethanolamine, and sphingomyelin. Multiple functions of this molecule in the body and in PD patients was evidenced [58]. Thus, it makes sense that many of the selected

spectral markers were assigned to phosphate groups and lipids. Beyond phosphoethanolamine, a circulating amino acid signature encompassing higher amino acids levels was found in older people with PD.

Different research groups reported that impairment in oxidative stress was directly linked with an elevation of plasma sorbitol concentrations in drug-naive patients. As observed, the critical role of oxidative stress in PD is undeniably evident. Several evidence lines have also been found for dysregulation in glucose metabolism in moderate to advanced PD patients [16]. The reduced concentrations of alanine, lactic acid and glucose were detected and correlated with affected glucose metabolism [24].

Similarly, the selected band at 1297 cm^{-1} due to bending (CH$_2$) of Ala could have a double significance. Considering that carnosine is a dipeptide of alanine and histidine, which have an antioxidant function in PD, as was already observed before, the decreased levels in alanine could also justify the decrease in the levels of carnosine [59]. Thus, considering the significance of some bands associated with histidine, it is congruent to assume the possible correspondence of these bands to carnosine. Therefore, our findings could be related to the role of a biomarker of carnosine in PD.

5.1.7.5 Dementia type differentiation: 2-class approach for discriminating between patients with PD-related dementia and Alzheimer's dementia.

The last sub-classification problem was examined in greater depth to directly discriminate between dementia associated with both pathologies: Parkinson-related dementia and Alzheimer's dementia. The analytical approach was based on the same classification strategy described before, following the rule of extraction of the truly discriminant variables by SELECT. Ten cancellation groups for the CV and a subset of 2 random samples (each for one category) was performed to optimise and validate classifications. The achieved results were auspicious, so the success rate for both categories was 100% in classification, 87.50% and 100% in internal prediction and 100% of success rate in external prediction, respectively (Table 5-5). Given the results obtained per category, it is essential to highlight that AD patients were perfectly discriminated from patients with PD-related dementia (100% correct assignments in both classification and CV). Focusing solely on the numeric value of the obtained percentages made it more challenging to accurately classify PDD subjects into their real category. It should be

considered that the slight decrease in internal validation performance observed for this latter class is related to the wrong assignment of a single sample. However, the limited number of available patients for this category indeed contributed to maximising the influence of this deviation. The results can be visually appreciated in Figure 5-5. Histogram of canonical variable for the discrimination of PD related dementia (■) and Alzheimer's dementia (■) patients after performing SELECT-LDA in the PD dementia type direct differentiation approach. Only ten selected variables were needed to provide a remarkable stratification between two different types of Dementia (Table 5-6).
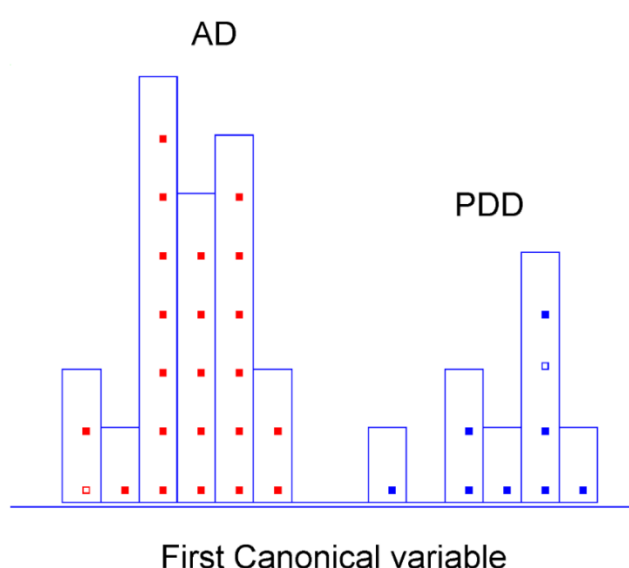
Figure 5-5. Histogram of canonical variable for the discrimination of PD related dementia (■) and Alzheimer's dementia (■) patients after performing SELECT-LDA in the PD dementia type direct differentiation approach.

Table 5-5. Percentages of samples/patients classified correctly in both classification and internal/external validation in the SELECT-LDA performed when addressing the classification approach for dementia type differentiation.

2-class approach SELECT-LDA: Dementia differentiation with 15 variables		
	Classification %	**Prediction (LOO) %**	**External Prediction %**
PDI	100.00	87.50	100.00
AD	100.00	100.00	100.00
Total rate	100.00	96.67	100.00

Table 5-6. Percentages of samples/patients classified correctly in both classification and internal/external validation in the SELECT-LDA performed when addressing the classification approach for dementia type differentiation.

3rd-step differentiation approach: 10 biomarkers	
2 categories (PDD and AD) differentiation	
Selection order	**Wavenumber (cm^{-1})**
1	1340.4286
2	1487.0078
3	1488.9365
4	1187.0990
5	1277.7466
6	1380.9307
7	1188.0633
8	1379.0020
9	1382.8594
10	1402.1461

5.1.7.6   *Chemical reasoning about possible contribution of selected wavenumbers in the third-step classification*

To distinguish between spectrochemical profile of patients with PDD and AD, the discriminant band at 1340 cm^{-1} due to bending (CH$_2$) was selected as the most important. It was assigned to the absorption of collagen [60]. The involvement of determined vibrations attributable to collagen is entirely reasonable. Furthermore, it was shown that the brain's neurons are the source of the specific type of collagen (collagen VI). Thus, an increased level of this collagen in the brain could have a protective function against AD [61]. Overall, other studies hypothesized that the progressive degradation of nerve cells coming with Alzheimer's disease duration could potentially undermine their ability to produce collagen [62].

Moreover, in spectroscopic sub-signatures that could further differentiate between PDD and AD another region at 1402 cm^{-1} was associated with stretching vibration ($\tilde{v}$) (C=C) of UA, which was discussed and justified above. Multiple signatures attributed to stretching vibrations of metile and methylene of lipids and phospholipids at 1379, 1380, 1382, 1402, 1487 cm^{-1} and bending (CH$_2$) at 1488 cm^{-1} have also been selected. Likewise, for the role of band at 1377 cm^{-1} in the global classification, in this classification sub-

problem, we also speculated that bands at 1379, 1380 and 1382 cm^{-1} could correspond to amino acids and among these, proline [63].

Supplementary data (Table S 5-7, Table S 5-8, Table S 5-9) summarize all relevant absorption bands for each classification type, including the identification of the bond vibrations involved and the respective biochemical assignments.

The main highlight of this untargeted FTIR-based metabolomics study was focused on discovering if spectroscopic signatures can differentiate PD from other neurodegenerative conditions with shared symptoms, but not on the analysis of the specific contribution of each FTIR reduced fingerprint component. Nevertheless, the provided biochemical reasoning about the contribution of specific bands in the differentiation of patients seems to be ideally in line with our results, reinforcing the suitability of our classification strategy.

5.1.8   Conclusions

There is still no standard robust approach for the objective diagnosis of Parkinson's disease; this field remains underexplored and poorly understood. The reported results highlighted the potency of the adopted chemometric strategy, based on a three-step classification approach in the stratification of Parkinson's patients; the disease was effectively classified and differentiated from the control group and other impairments such as Alzheimer's dementia. Spectral signatures in human plasma have been successfully identified for differentiation between patient categories by selecting significant wavenumbers closely related to PD pathogenesis and metabolic biomarkers. Moreover, the rapid, high-throughput and relatively inexpensive method provided optimal discrimination results in both sub-classification problems, succeeding in the stratification of patients with different PD stage progression profiles and those with different dementia type profiles, respectively. The reported untargeted metabolomic approach seems to deal significantly with the necessity of developing an alternative screening method to distinguish patient profiles, thus taking vibrational spectroscopy one step forward towards clinical implementation. All the speculations made about the involvement of selected bands in the pathogenesis of PD are immensely reasonable, and their role is perfectly justifiable for patient stratification. A primary limitation of the current work relies on the relatively small number of available plasma samples,

especially for one subgroup (patients with PD-related dementia), preventing more general conclusions. Further investigation is required; a more significant number of enrolled patients could strengthen the validity of the proposed classification strategies as an objective diagnosis of PD.

Supplementary data of Chapter 5.1

Table S 5-7. Biochemical tentative assignments of the most discriminant wavenumbers selected by SELECT-LDA in the global classification approach aimed at differentiating between PD, AD, and healthy patients.

Spectral Regions in Literature	Peak Position (cm^{-1}) ± 1	Tentative Band Assignment	Contributions
~1155	**1156, 1159**	$\tilde{v}$ sym. (CO-O-C)	Carbohydrates
~1185–1120	**1171, 1182**	$\tilde{v}$ (C-C) and (O-P-O); (C-O) ring vibrations	Nucleic acid "sugars"
~1225	**1215**	$\tilde{v}$ asym. (O-P-O)	Nucleic acids, phospholipids
~1250–1220	**1253**	$\tilde{v}$ sym. (P = O) of the PO2 groups	Nucleic acids, phospholipids
~1360–1220	**1266, 1271, 1284, 1286, 1288, 1289, 1294, 1302, 1312, 1316, 1319, 1335, 1336, 1377**	$\tilde{v}$ (C-C) and (C-O) $\tilde{v}$ (C-N) and C-(NO_2)) $\tilde{v}$ sym. (PO$_2$), predominantly $\tilde{v}$ (C-N) with significant contributions from $\tilde{v}$ (CH$_2$) of carbohydrate residues, δ (CH$_2$)	Amide III band, proteins
~1370	**1377**	sym. def. CH$_3$ and sym. def. CH$_2$	Proteins, amino acids (cytosine, guanine, proline) lipids, phospholipids
~1400	**1403**	$\tilde{v}$ (C = O) of (COO) group	Fatty acids and amino acids
~1420	**1424**	$\tilde{v}$ sym. (COO), δ asym. (CH$_2$)	Polysaccharides
~1455–1450	**1455**	δ asym. (CH$_3$) and (CH$_2$) modes	Proteins, lipids
~1490–1470	**1474, 1477, 1489**	δ (CH$_2$)	Lipids

Table S 5-8. Biochemical tentative assignments of the most discriminant wavenumbers selected by SELECT-LDA in the classification approach aimed at stratifying PD patients.

Spectral Regions in Literature	Peak Position (cm^{-1}) $\pm$ 1	Tentative Band Assignment	Contributions
~1185–1120	1170, 1171	$\tilde{v}$ (C-C) and (O-P-O); (C-O) ring vibrations	Nucleic acid "sugars"
~1233 ~1225	1214, 1224, 1226	$\tilde{v}$ asym. (O-P-O)	Nucleic acids; phospholipids; uric ring vibrations
~1360–1220	1292, 1294, 1297, 1342	$\tilde{v}$ (C-N) and C-(NO_2)); $\tilde{v}$ sym. (PO_2) predominantly $\tilde{v}$ (C-N) with significant contributions from $\tilde{v}$ (CH_2) of carbohydrate residues; δ (CH_2)	Amide III band; proteins; collagen
~1420	1435, 1437	$\tilde{v}$ sym. (COO); δ (CH_2)	Polysaccharides
~1455–1450	1443, 1445	δ asym. (CH_3) and (CH_2) modes	Proteins; lipids
~1490–1470	1475, 1476	δ (CH_2)	Lipids

Table S 5-9. Biochemical tentative assignments of the most discriminant wavenumbers selected by SELECT-LDA in the classification approach aimed at achieving direct discrimination between dementia associated with both analysed pathologies (PD-related dementia and Alzheimer's dementia).

Spectral Regions in Literature	Peak Position (cm^{-1}) $\pm$ 1	Tentative Band Assignment	Contributions
~1185–1120	1187, 1188	$\tilde{v}$ (C-C) and (O-P-O) (C-O) ring vibrations	Nucleic acid "sugars"
~1360–1220	1277 1340	$\tilde{v}$ (C-C) and (C-O) $\tilde{v}$ (C-N) and (C-(NO_2)) $\tilde{v}$ sym. (PO$_2$) $\tilde{v}$ (C-N) with significant contributions from $\tilde{v}$ (CH$_2$) of carbohydrate residues, δ (CH$_2$)	Amide III band; proteins; collagen
~1370	1379, 1380, 1382	sym. def. CH$_3$ and sym. def. CH$_2$	Proteins; amino acids (cytosine, guanine, proline) Lipids; phospholipids
~1405–1400	1402	$\tilde{v}$ (C = O) of (COO) group $\tilde{v}$ (C = C)	Fatty acids; amino acids; (aspartate, glutamate) Uric acid
~1490–1470	1487, 1488	δ (CH$_2$)	Lipids

5.1.9   References

1.   Mhyre, T. R.; Boyd, J. T.; Hamill, R. W.; Maguire-Zeiss, K. A. Parkinson's Disease. Subcell. Biochem. 2012, 65, 389–455. https://doi.org/10.1007/978-94-007-5416-4_16.

2.   Selkoe, D. J.; Peter J Lansbury, J. Alzheimer's Disease Is the Most Common Neurodegenerative Disorder. 1999.

3.   Rizek, P.; Kumar, N.; Jog, M. S. An Update on the Diagnosis and Treatment of Parkinson Disease. CMAJ. Canadian Medical Association November 1, 2016, pp 1157–1165. https://doi.org/10.1503/cmaj.151179.

4.   Meade, R. M.; Fairlie, D. P.; Mason, J. M. Alpha-Synuclein Structure and Parkinson's Disease. Mol. Neurodegener. 2019, 14 (1), 1–14.

5.   Rocha, E. M.; De Miranda, B.; Sanders, L. H. Alpha-Synuclein: Pathology, Mitochondrial Dysfunction and Neuroinflammation in Parkinson's Disease. Neurobiol. Dis. 2018, 109, 249–257. https://doi.org/10.1016/j.nbd.2017.04.004.

6.   Galvagnion, C. The Role of Lipids Interacting with α-Synuclein in the Pathogenesis of Parkinson's Disease. J. Parkinsons. Dis. 2017, 7 (3), 433–450. https://doi.org/10.3233/JPD-171103.

7.   Gnanalingham, K. K.; Byrne, E. J.; Thornton, A.; Sambrook, M. A.; Bannister, P. Motor and Cognitive Function in Lewy Body Dementia: Comparison with Alzheimer's and Parkinson's Diseases. J. Neurol. Neurosurg. Psychiatry 1997, 62 (3), 243–252. https://doi.org/10.1136/jnnp.62.3.243.

8.   Tofaris, G. K. A Critical Assessment of Exosomes in the Pathogenesis and Stratification of Parkinson's Disease. J. Parkinsons. Dis. 2017, 7 (4), 569–576. https://doi.org/10.3233/JPD-171176.

9.   D'Andrea, G.; Pizzolato, G.; Gucciardi, A.; Stocchero, M.; Giordano, G.; Baraldi, E.; Leon, A. Different Circulating Trace Amine Profiles in De Novo and Treated Parkinson's Disease Patients. Sci. Rep. 2019, 9 (1), 1–11. https://doi.org/10.1038/s41598-019-42535-w.

10.   Picca, A.; Calvani, R.; Landi, G.; Marini, F.; Biancolillo, A.; Gervasoni, J.; Persichilli, S.; Primiano, A.; Urbani, A.; Bossola, M.; Bentivoglio, A. R.; Cesari, M.; Landi, F.; Bernabei, R.; Marzetti, E.; Lo Monaco, M. R. Circulating Amino Acid Signature in Older People with Parkinson's Disease: A Metabolic Complement to the EXosomes in PArkiNson Disease (EXPAND) Study. Exp. Gerontol. 2019, 128 (August), 110766. https://doi.org/10.1016/j.exger.2019.110766.

11.   Trošt, M.; Perovnik, M.; Pirtošek, Z. Correlations of Neuropsychological and Metabolic Brain Changes in Parkinson's Disease and Other α-Synucleinopathies. Front. Neurol. 2019, 10 (November), 1–10. https://doi.org/10.3389/fneur.2019.01204.

12.   The Role of Lipids Interacting with $α$-Synuclein in the Pathogenesis of Parkinson's Disease. J. Parkinsons. Dis. 2017, 7 (3), 433–450. https://doi.org/10.3233/JPD-171103.

13. Jenner, P.; Dexter, D. T.; Sian, J.; Schapira, A. H. V.; Marsden, C. D. Oxidative Stress as a Cause of Nigral Cell Death in Par-kinson's Disease and Incidental Lewy Body Disease. Ann. Neurol. 1992, 32 (S1), S82–S87. https://doi.org/10.1002/ana.410320714.

14. Yu, Z.; Zhang, S.; Wang, D.; Fan, M.; Gao, F.; Sun, W.; Li, Z.; Li, S. The Significance of Uric Acid in the Diagnosis and Treatment of Parkinson Disease. Medicine (United States). Lippincott Williams and Wilkins November 1, 2017. https://doi.org/10.1097/MD.0000000000008502.

15. Zhong, L.-L.; Song, Y.-Q.; Tian, X.-Y.; Cao, H.; Ju, K.-J. Level of Uric Acid and Uric Acid/Creatinine Ratios in Correlation with Stage of Parkinson Disease. Medicine (Baltimore). 2018, 97 (26), e10967. https://doi.org/10.1097/MD.0000000000010967.

16. Marques, A.; Dutheil, F.; Durand, E.; Rieu, I.; Mulliez, A.; Fantini, M. L.; Boirie, Y.; Durif, F. Glucose Dysregulation in Par-kinson's Disease: Too Much Glucose or Not Enough Insulin? Park. Relat. Disord. 2018, 55, 122–127. https://doi.org/10.1016/j.parkreldis.2018.05.026.

17. Anichtchik, O. V.; Peitsaro, N.; Anichtchik, O. V.; Rinne, J. O.; Kalimo, H.; Kalimo, H.; Panula, P. Distribution and Modulation of Histamine H3 Receptors in Basal Ganglia and Frontal Cortex of Healthy Controls and Patients with Parkinson's Disease. Neurobiol. Dis. 2001, 8 (4), 707–716. https://doi.org/10.1006/nbdi.2001.0413.

18. Shan, L.; Swaab, D. F.; Bao, A. M. Neuronal Histaminergic System in Aging and Age-Related Neurodegenerative Disorders. Exp. Gerontol. 2013, 48 (7), 603–607. https://doi.org/10.1016/j.exger.2012.08.002.

19. Hertel, J.; Harms, A. C.; Heinken, A.; Baldini, F.; Thinnes, C. C.; Glaab, E.; Vasco, D. A.; Pietzner, M.; Stewart, I. D.; Wareham, N. J.; Langenberg, C.; Trenkwalder, C.; Krüger, R.; Hankemeier, T.; Fleming, R. M. T.; Mollenhauer, B.; Thiele, I. Integrated Analyses of Microbiome and Longitudinal Metabolome Data Reveal Microbial-Host Interactions on Sulfur Metabolism in Parkinson's Disease. Cell Rep. 2019, 29 (7), 1767-1777.e8. https://doi.org/10.1016/j.celrep.2019.10.035.

20. Kataoka, H.; Sugie, K. Serum Adiponectin Levels between Patients with Parkinson's Disease and Those with PSP. Neurol. Sci. 2020, 41 (5), 1125–1131. https://doi.org/10.1007/s10072-019-04216-4.

21. Maass, F.; Michalke, B.; Willkommen, D.; Leha, A.; Schulte, C.; Tönges, L.; Mollenhauer, B.; Trenkwalder, C.; Rückamp, D.; Börger, M.; Zerr, I.; Bähr, M.; Lingor, P. Elemental Fingerprint: Reassessment of a Cerebrospinal Fluid Biomarker for Par-kinson's Disease. Neurobiol. Dis. 2020, 134 (September 2019), 104677. https://doi.org/10.1016/j.nbd.2019.104677.

22. Espay, A. J.; Kalia, L. V.; Gan-Or, Z.; Williams-Gray, C. H.; Bedard, P. L.; Rowe, S. M.; Morgante, F.; Fasano, A.; Stecher, B.; Kauffman, M. A.; Farrer, M. J.; Coffey, C. S.; Schwarzschild, M. A.; Sherer, T.; Postuma, R. B.; Strafella, A. P.; Singleton, A. B.; Barker, R. A.; Kieburtz, K.; Olanow, C. W.; Lozano, A.; Kordower, J. H.; Cedarbaum, J. M.; Brundin, P.; Standaert, D. G.; Lang, A. E. Disease Modification and Biomarker Development in Parkinson Disease: Revision or Reconstruction? Neurology 2020, 94 (11), 481–494. https://doi.org/10.1212/WNL.0000000000009107.

23. Santaella, A.; Kuiperij, H. B.; Van Rumund, A.; Esselink, R. A. J.; Van Gool, A. J.; Bloem, B. R.; Verbeek, M. M. Inflammation Biomarker Discovery in Parkinson's Disease and Atypical Parkinsonisms. BMC Neurol. 2020, 20 (1), 1–8. https://doi.org/10.1186/s12883-020-1608-8.

24. Öhman, A.; Forsgren, L. NMR Metabonomics of Cerebrospinal Fluid Distinguishes between Parkinson's Disease and Controls. Neurosci. Lett. 2015, 594, 36–39. https://doi.org/10.1016/j.neulet.2015.03.051.

25. Obeso, J. A.; Stamelou, M.; Goetz, C. G.; Poewe, W.; Lang, A. E.; Weintraub, D.; Burn, D.; Halliday, G. M.; Bezard, E.; Przedborski, S.; Lehericy, S.; Brooks, D. J.; Rothwell, J. C.; Hallett, M.; DeLong, M. R.; Marras, C.; Tanner, C. M.; Ross, G. W.; Langston, J. W.; Klein, C.; Bonifati, V.; Jankovic, J.; Lozano, A. M.; Deuschl, G.; Bergman, H.; Tolosa, E.; Rodriguez-Violante, M.; Fahn, S.; Postuma, R. B.; Berg, D.; Marek, K.; Standaert, D. G.; Surmeier, D. J.; Olanow, C. W.; Kordower, J. H.; Calabresi, P.; Schapira, A. H. V.; Stoessl, A. J. Past, Present, and Future of Parkinson's Disease: A Special Essay on the 200th Anniversary of the Shaking Palsy. Movement Disorders. John Wiley and Sons Inc. September 1, 2017, pp 1264–1310. https://doi.org/10.1002/mds.27115.

26. Ohmichi, T.; Mitsuhashi, M.; Tatebe, H.; Kasai, T.; Ali El-Agnaf, O. M.; Tokuda, T. Quantification of Brain-Derived Extracellular Vesicles in Plasma as a Biomarker to Diagnose Parkinson's and Related Diseases. Park. Relat. Disord. 2019, 61 (November 2018), 82–87. https://doi.org/10.1016/j.parkreldis.2018.11.021.

27. Ahmed, S. S.; Santosh, W.; Kumar, S.; Christlet, H. T. T. Metabolic Profiling of Parkinson's Disease: Evidence of Biomarker from Gene Expression Analysis and Rapid Neural Network Detection. J. Biomed. Sci. 2009, 16 (1), 1–12. https://doi.org/10.1186/1423-0127-16-63.

28. Shao, Y.; Le, W. Recent Advances and Perspectives of Metabolomics-Based Investigations in Parkinson's Disease. Mol. Neu-rodegener. 2019, 14 (1), 1–12. https://doi.org/10.1186/s13024-018-0304-2.

29. Figura, M.; Kuśmierska, K.; Bucior, E.; Szlufik, S.; Koziorowski, D.; Jamrozik, Z.; Janik, P. Serum Amino Acid Profile in Patients with Parkinson's Disease. PLoS One 2018, 13 (1). https://doi.org/10.1371/journal.pone.0191670.

30. Dashti, H.; Westler, W. M.; Tonelli, M.; Wedell, J. R.; Markley, J. L.; Eghbalnia, H. R. Spin System Modeling of Nuclear Magnetic Resonance Spectra for Applications in Metabolomics and Small Molecule Screening. Anal. Chem. 2017, 89 (22), 12201–12208.

31. Antcliffe, D.; Jiménez, B.; Veselkov, K.; Holmes, E.; Gordon, A. C. Metabolic Profiling in Patients with Pneumonia on Intensive Care. 2017. https://doi.org/10.1016/j.ebiom.2017.03.034.

32. Bereman, M. S.; Kirkwood, K. I.; Sabaretnam, T.; Furlong, S.; Rowe, D. B.; Guillemin, G. J.; Mellinger, A. L.; Muddiman, D. C. Metabolite Profiling Reveals Predictive Biomarkers and the Absence of β-Methyl Amino-l-Alanine in Plasma from Indi-viduals Diagnosed with Amyotrophic Lateral Sclerosis. J. Proteome Res. 2020, 19 (8), 3276–3285.

33.   Ros-Mazurczyk, M.; Jelonek, K.; Marczyk, M.; Binczyk, F.; Pietrowska, M.; Polanska, J.; Dziadziuszko, R.; Jassem, J.; Rzyman, W.; Widlak, P. Serum Lipid Profile Discriminates Patients with Early Lung Cancer from Healthy Controls. Lung Cancer 2017, 112 (November 2016), 69–74. https://doi.org/10.1016/j.lungcan.2017.07.036.

34.   Yao, L.; Lyu, N.; Chen, J.; Pan, T.; Yu, J. Joint Analyses Model for Total Cholesterol and Triglyceride in Human Serum with Near-Infrared Spectroscopy. Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 2016, 159, 53–59. https://doi.org/10.1016/j.saa.2016.01.022.

35.   Khalil, S. K. H.; Azooz, M. a; Division, P. Application of Vibrational Spectroscopy in Identification of the Composition of the Urinary Stones. J. Appl. Sci. Res. 2007, 3 (5), 387–391.

36.   Selvaraju, R.; Raja, A.; Thiruppathi, G. FT-Raman Spectral Analysis of Human Urinary Stones. Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 2012, 99, 205–210. https://doi.org/10.1016/j.saa.2012.09.004.

37.   Roy, S.; Perez-Guaita, D.; Bowden, S.; Heraud, P.; Wood, B. R. Spectroscopy Goes Viral: Diagnosis of Hepatitis B and C Virus Infection from Human Sera Using ATR-FTIR Spectroscopy. Clin. Spectrosc. 2019, 1, 100001. https://doi.org/10.1016/j.clispe.2020.100001.

38.   Lilo, T.; Morais, C. L. M.; Ashton, K. M.; Pardilho, A.; Davis, C.; Dawson, T. P.; Gurusinghe, N.; Martin, F. L. Spectrochemical Differentiation of Meningioma Tumours Based on Attenuated Total Reflection Fourier-Transform Infrared (ATR-FTIR) Spectroscopy. Anal. Bioanal. Chem. 2020, 412 (5), 1077–1086. https://doi.org/10.1007/s00216-019-02332-w.

39.   Carmona, P.; Molina, M.; Calero, M.; Bermejo-Pareja, F.; Martínez-Martín, P.; Toledano, A. Discrimination Analysis of Blood Plasma Associated with Alzheimer's Disease Using Vibrational Spectroscopy. J. Alzheimer's Dis. 2013. https://doi.org/10.3233/JAD-122041.

40.   Ahmed, S. S. S. J.; Santosh, W.; Kumar, S.; Thanka Christlet, T. H. Neural Network Algorithm for the Early Detection of Parkinson's Disease from Blood Plasma by FTIR Micro-Spectroscopy. Vib. Spectrosc. 2010, 53 (2), 181–188. https://doi.org/10.1016/j.vibspec.2010.01.019.

41.   Li, S.; Liu, J.; Li, G.; Zhang, X.; Xu, F.; Fu, Z.; Teng, L.; Li, Y.; Sun, F. Near-Infrared Light-Responsive, Pramipexole-Loaded Biodegradable PLGA Microspheres for Therapeutic Use in Parkinson's Disease. Eur. J. Pharm. Biopharm. 2019, 141 (May), 1–11. https://doi.org/10.1016/j.ejpb.2019.05.013.

42.   Wang, X.; Wu, Q.; Li, C.; Zhou, Y.; Xu, F.; Zong, L.; Ge, S. A Study of Parkinson's Disease Patients' Serum Using FTIR Spec-troscopy. Infrared Phys. Technol. 2020, 106 (December 2019), 103279. https://doi.org/10.1016/j.infrared.2020.103279.

43.   Perez-Guaita, D.; Garrigues, S.; de la, M. Infrared-Based Quantification of Clinical Parameters. TrAC - Trends Anal. Chem. 2014, 62, 93–105. https://doi.org/10.1016/j.trac.2014.06.012.

44. Pizarro, C.; Esteban-Díez, I.; Espinosa, M.; Rodríguez-Royo, F.; González-Sáiz, J. M. An NMR-Based Lipidomic Approach to Identify Parkinson's Disease-Stage Specific Lipoprotein-Lipid Signatures in Plasma. Analyst 2019, 144 (4), 1334–1344. https://doi.org/10.1039/c8an01778f.

45. Lawton, M.; Baig, F.; Toulson, G.; Morovat, A.; Evetts, S. G.; Ben-Shlomo, Y.; Hu, M. T. Blood Biomarkers with Parkinson's Disease Clusters and Prognosis: The Oxford Discovery Cohort. Mov. Disord. 2020, 35 (2), 279–287. https://doi.org/10.1002/mds.27888.

46. Zhao, H. W.; Lin, J.; Wang, X. B.; Cheng, X.; Wang, J. Y.; Hu, B. L.; Zhang, Y.; Zhang, X.; Zhu, J. H. Assessing Plasma Levels of Selenium, Copper, Iron and Zinc in Patients of Parkinson's Disease. PLoS One 2013, 8 (12), 1–10. https://doi.org/10.1371/journal.pone.0083060.

47. Barth, A. Infrared Spectroscopy of Proteins. Biochim. Biophys. Acta - Bioenerg. 2007, 1767 (9), 1073–1101. https://doi.org/10.1016/j.bbabio.2007.06.004.

48. Pizarro, C.; Arenzana-Rámila, I.; Pérez-del-Notario, N.; Pérez-Matute, P.; González-Sáiz, J. M. Thawing as a Critical Pre-Analytical Step in the Lipidomic Profiling of Plasma Samples: New Standardized Protocol. Anal. Chim. Acta 2016, 912, 1–9. https://doi.org/10.1016/j.aca.2016.01.058.

49. M. Forina, S. Lanteri, C. Armanino, M.C.C. Oliveros, C. C. V-Parvus 2003. Dip. Chimica e Tecnologie Farmaceutiche ed Alimentari, University of Genova,.

50. Tabora, J. E.; Domagalski, N. Multivariate Analysis and Statistics in Pharmaceutical Process Research and Development. Annual Review of Chemical and Biomolecular Engineering. Annual Reviews Inc. June 7, 2017, pp 403–426. https://doi.org/10.1146/annurev-chembioeng-060816-101418.

51. Worsfold, P. J. Chemometrics: A Textbook (Data Handling in Science and Technology, Vol. 2). Anal. Chim. Acta 1989, 225. https://doi.org/10.1016/s0003-2670(00)84639-5.

52. Forina, M.; Oliveri, P.; Casale, M. Complete Validation for Classification and Class Modeling Procedures with Selection of Variables and/or with Additional Computed Variables. Chemom. Intell. Lab. Syst. 2010. https://doi.org/10.1016/j.chemolab.2010.04.011.

53. Casale, M.; Sáiz Abajo, M. J.; González Sáiz, J. M.; Pizarro, C.; Forina, M. Study of the Aging and Oxidation Processes of Vinegar Samples from Different Origins during Storage by Near-Infrared Spectroscopy. In Analytica Chimica Acta; 2006. https://doi.org/10.1016/j.aca.2005.10.063.

54. Bury, D.; Morais, C. L. M.; Paraskevaidi, M.; Ashton, K. M.; Dawson, T. P.; Martin, F. L. Spectral Classification for Diagnosis Involving Numerous Pathologies in a Complex Clinical Setting: A Neuro-Oncology Example. Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 2019, 206, 89–96. https://doi.org/10.1016/j.saa.2018.07.078.

55. Mitchell, A. L.; Gajjar, K. B.; Theophilou, G.; Martin, F. L.; Martin-Hirsch, P. L. Vibrational Spectroscopy of Biofluids for Disease Screening or Diagnosis: Translation

from the Laboratory to a Clinical Setting. J. Biophotonics 2014, 7 (3–4), 153–165. https://doi.org/10.1002/jbio.201400018.

56. Tofaris, G. K. A Critical Assessment of Exosomes in the Pathogenesis and Stratification of Parkinson's Disease. J. Parkinsons. Dis. 2017, 7 (4), 569–576. https://doi.org/10.3233/JPD-171176.

57. Rocha, E. M.; De Miranda, B.; Sanders, L. H. Alpha-Synuclein: Pathology, Mitochondrial Dysfunction and Neuroinflammation in Parkinson's Disease. Neurobiol. Dis. 2018, 109, 249–257. https://doi.org/10.1016/j.nbd.2017.04.004.

58. Picca, A.; Calvani, R.; Landi, G.; Marini, F.; Biancolillo, A.; Gervasoni, J.; Persichilli, S.; Primiano, A.; Urbani, A.; Bossola, M.; Bentivoglio, A. R.; Cesari, M.; Landi, F.; Bernabei, R.; Marzetti, E.; Lo Monaco, M. R. Circulating Amino Acid Signature in Older People with Parkinson's Disease: A Metabolic Complement to the EXosomes in PArkiNson Disease (EXPAND) Study. Exp. Gerontol. 2019, 128 (October), 110766. https://doi.org/10.1016/j.exger.2019.110766.

59. Zhao, J.; Shi, L.; Zhang, L. R. Neuroprotective Effect of Carnosine against Salsolinol-Induced Parkinson's Disease. Exp. Ther. Med. 2017, 14 (1), 664–670. https://doi.org/10.3892/etm.2017.4571.

60. Lilo, T.; Morais, C. L. M.; Ashton, K. M.; Pardilho, A.; Davis, C.; Dawson, T. P.; Gurusinghe, N.; Martin, F. L. Spectrochemical Differentiation of Meningioma Tumours Based on Attenuated Total Reflection Fourier-Transform Infrared (ATR-FTIR) Spectroscopy. Anal. Bioanal. Chem. 2020, 412 (5), 1077–1086. https://doi.org/10.1007/s00216-019-02332-w.

61. Cheng, J. S.; Dubal, D. B.; Kim, D. H.; Legleiter, J.; Cheng, I. H.; Yu, G.-Q.; Tesseur, I.; Wyss-Coray, T.; Bonaldo, P.; Mucke, L. Collagen VI Protects Neurons against Aβ Toxicity. https://doi.org/10.1038/nn.2240.

62. Paraskevaidi, M.; Morais, C. L. M.; Lima, K. M. G.; Snowden, J. S.; Saxon, J. A.; Richardson, A. M. T.; Jones, M.; Mann, D. M. A.; Allsop, D.; Martin-Hirsch, P. L.; Martin, F. L. Differential Diagnosis of Alzheimer's Disease Using Spectrochemical Analysis of Blood. Proc. Natl. Acad. Sci. U. S. A. 2017. https://doi.org/10.1073/pnas.1701517114.

63. Jim Enez-Jim Enez, F. J.; Alonso-Navarro, H.; Garc Ia-Mart, E.; Ag Undez, J. A. G. Cerebrospinal and Blood Levels of Amino Acids as Potential Biomarkers for Parkinson's Disease: Review and Meta-Analysis. Eur. J. Neurol. 2020, 2020, 2336–2347. https://doi.org/10.1111/ene.14470.

5.2 IDENTIFICATION OF LIPIDOMIC TRAITS IN PLASMA SAMPLES FOR THE DISCRIMINATION OF PARKINSON'S DISEASE: UPLC-MS UNTARGETED APPROACH

5.2.1 Introduction

Nowadays, the prevalence of progressive neurodegenerative disease constitutes to be a socio-economic burden [1]. Thus, Parkinson's disease is one of the most complex progressive neurodegenerative disorders which manifest with a broad range of motor and non-motor symptoms [2]. It is the second most common age-related neurodegenerative disorder after Alzheimer's (AD). Such as this last one, PD is characterised by the accumulation of intracellular protein aggregates, Lewy bodies, composed primarily of the protein alpha-synuclein [3] For this reason, PD is both, a cerebral amyloid disease and the most common synucleinopathy [4]. In recent years, the interest of the scientific community in finding specific biomarkers of this neurodegenerative disease has grown substantially [5]. Despite insight derived from causative genetic mutations, which explain only a small proportion of cases the remaining 90% are due to non-genetic factors and are apparently sporadic [6]. The exact pathogenetic mechanism underlying this disease is still poorly understood.

Despite all the advances in genetics and neuroimaging, the PD diagnosis remains essentially clinical, based on subjective observations of clinicians. The most critical challenge in clinical practice stands in the inability to make a definitive diagnosis at the early stages predicting the disease progression. The sign and symptoms appear on a later stage of the disease, when the neurodegenerative process has started and is irreversible. In addition, even when the new diagnostic criteria are correctly applied, the misdiagnosis rate is still high [7], precluding the intervention at the early stage of the disease. The false discovery rates regard the more frequent presence in elderly population of conditions such as essential tremor, cognitive impairment due to AD or progressive supranuclear palsy. In addition, the clinical frame could be ulteriorly complicated by the increasing incidence of co-morbidities. Therefore, the molecular mechanism leading to neurodegenerations remain elusive.

Among currently hypotheses, the complex convergence of genetic and environmental factors, such as exposure to heavy metals, smoking or dietary habits were proposed to

play an important role in PD pathogenesis[8]. Many studies highlighted the role of oxidative stress and its elevated implication in protein misfolding, death of neuronal cells, lipid peroxidation- all mechanism at the basis of neuron degeneration[9]. Increasing evidence has demonstrated that brain and cognitive aging are accompanied by peripheral metabolic perturbations [10]. Since the metabolites are the end points of multiple interactions and processes that happens in the organism, metabolomics is an increasingly recognized tool for investigating altered metabolic profiles of patients. Within the metabolome, lipids are involved in the important biological functions, including structure of cell membranes, energy storage and signalling. Thus, many studies reported that glycerophospholipids, sphingolipids and ceramides exert important biological roles in the central nervous system (CNS), such as signal transduction, apoptosis and structural neuronal maintenance [11]. In addition, changes in total ceramide molecular species and even changes in ceramide acyl chain length can affect membrane order or membrane lipid peroxidation [12]. Recently, it was shown that phospholipids and sphingolipids are highly concentrated in membrane lipids rafts (MLRs) . These specialized plasma membrane microdomains are integral to regulating intracellular trafficking and signal transduction [13]. Therefore, alterations in the composition of MLRs have been reported in PD. Likewise, it was shown that Alzheimer's patients are characterized by the decreased levels of sphingomyelins and increased levels of ceramides due to sphingomyelin hydrolysis. Given this perspective, it is evident that dysregulated metabolism of the lipids in the brain may lead to functional neurological disorders.

Metabolomics studies have already proven its great potential coupled with high-throughput techniques to perform metabolic profiling to evaluate significantly discriminant biomarkers between healthy and diseased groups [14–20] that may contribute to neurodegeneration. In addition, since blood is readily available and easy sample, compared to cerebrospinal fluid that could create complications during the collection step, many studies predilate to investigate blood biomarkers.

Thus, herein, we performed an untargeted metabolomic analysis using UPLC-MS/MS technique on plasma samples to investigate possible biomarkers responsible for changes and disarrangement in PD pathogenesis. Lipid extraction was performed to select only possible lipid species involved in PD or AD pathogenetic mechanism. Many studies focus

only on comparing metabolic profile and healthy controls. In this study we included AD patients, since misdiagnosis between Parkinson's related dementia and Alzheimer dementia still occurs.

5.2.2 Materials and Methods

5.2.2.1 Chemicals

Ultrapure water, used to prepare all the aqueous solutions, was obtained from the Milli-Q system (Milipore, Bedford, MA, USA). LC-MS grade acetonitrile (ACN), isopropyl alcohol (IpA), ammonium formate, high-performance liquid chromatography (HPLC) grade methanol and MTBE were supplied by Aldrich Chemie (Steinheim, Germany).

5.2.2.2 Study population

The blood from overnight fasting subjects was collected at The Molecular Neurobiology Department at the Biomedical Research Center of La Rioja (CIBIR), and the plasma of each participant was obtained. The study was conducted with x patients who, following the evolutionary stages established by Hoehn and Yahr: x patients have early PD-stage. They have recently been diagnosed with the disease solely based on their characteristic symptoms; x patients with AD Alzheimer's disease and x controls (CO), who belong to the family environment and are of a similar age to patients with a group of Parkinson's Disease enrolled in this study (Table 5-10). The study was conducted according to the guidelines of the Helsinki Declaration and was approved by the Ethics Committee of San Pedro Hospital. Written informed consent was obtained from each participant involved in this study. Any additional information was included in this study.

Table 5-10. Distribution of patient population.

	Gender	N·	Age distribution	Age
Controls	Male	9	60-80	69±1
	Female	16	43-76	59±1
PD	Male	29	46-82	67±1
	Female	16	59-85	67±1
AD	Male	6	68-81	76±1
	Female	21	60-85	73±1

5.2.2.3   Collection and handling of plasma samples

Blood samples were collected in Becton Dickinson (BD) Vacutainer plastic tubes with K2EDTA for plasma separation. Thus, plasma fraction was obtained by centrifugation at 2200 g for 15 min at 4·C. All samples as aliquots of 200 μl were frozen and stored in Eppendorf tubes at -80·C until further use.

5.2.2.4   Lipid extraction

Before lipid extraction, plasma thawing was performed according to the ultrasound-assisted extraction (USAE) protocol for lipidomic analysis [21]. Thus, the samples left to defrost in the fridge for 8 hours by night were submitted to the extraction the following day.

According to the MTBE-US-assisted lipid extraction method, 5 μL of Milli-Q water was added to a 10 μL aliquot of human blood plasma. Then, 20 μL of methanol was added to precipitate proteins by vortex-mixing for 2 min. Then, 250 μl de MTBE was added and dispersed by immersing the mixture in an ultrasonic water bath supplied by ATU Ultrasonidos (Valencia, Spain). The ultrasound frequency and power were 40 kHz and 100 W, respectively. The temperature was set at 15·C, and the time was adjusted to 30 min. Once USAE was performed, 25 μL of Milli-Q water was added to the mixture. Finally, the organic phase was separated by centrifugation at 3000 rpm for 10 min at 10 ·C in an Eppendorf 5403 Refrigerated Centrifuge (Hettich, Tuttlingen, Germany).

The lipid extracts in the upper phase were diluted five times with injection solvent before being collected and poured into an autosampler vial. Quality samples (QC) were processed similarly to the real samples.

5.2.2.5 Liquid chromatography-Mass Spectrometry

To determine plasma lipid profiles, a Waters Acquity UPLC chromatography system (Milford, MA, USA), equipped with a Waters Acquity HSS T3 100 × 2.1 (i.d.) mm 1.8 µm particle size column and a Waters VanGuard precolumn of the same material, coupled to a Microtof-Q (Q-TOF) mass spectrometer (Bruker Daltonik GmbH, Germany) with an electrospray interface (ESI) was used. To ensure the quality and stability of samples, the temperature of the autosampler was maintained at 5°C, and the column at 55°C. A mass spectrometer was operated in both positive and negative modes. Chromatographic and mass spectrometry data were acquired using Data Analysis software Version 4.0 (Bruker Daltonik GmbH, Germany). 2 µL samples were injected. Elution was performed using the gradient mobile consisting phase A (acetonitrile–water mixture (60:40, v/v) with 10 mM ammonium formate) and phase B (acetonitrile–isopropanol mixture (10:90, v/v) with 10 mM ammonium formate). UPLC separation was performed using a linear gradient that increased from 40% to 100% B within 10 min, and was held at 100% B for an additional 2 min. Finally, it increased from 0% to 60% A within 3.5 min. The total run time was 15.5 min. The flow rate was set at 0.4 mL*min^{-1}, and the injection volume was 10 µL.

Mass spectrometry data were acquired using a Waters Synapt XS HDMS (Waters Corp, Milford, USA) set to collect the data in continuum format using electrospray ionisation (ESI) in positive ionisation mode (ESI+) and negative ionisation mode (ESI−), over the mass range of m/z 50–2000. Capillary and sampling cone were set to 1.75 kV and 40 V, respectively, with the source temperature set to 120 °C and the desolvation temperature 500 °C. Gas flow rates were set at 800 L/h for the desolvation gas and 50 L/h for the cone gas, and the nebuliser gas was fixed at 6 bars. The mass spectrometer was set to acquire in resolution mode with a scan time of 0.4 s. Fragment ion information was acquired using a collision energy ramp from 20 to 50 V.

Lockmass correction was achieved by infusing leucine enkephalin at 10 µL/min through a lockspray probe and acquired every 30 s; for positive mode, [M + H]+ = 556.2771, and negative mode, [M − H]− = 554.2615. The data were collected using MassLynx V 4.2 (Waters Corp., Milford, USA).

The diseased samples and controls were alternated concerning run, avoiding batch effect. Moreover, QC samples were inserted regularly throughout the analytical run

157

(after every 20 real samples) to check the methodology's performance and test its precision.

5.2.2.6 Lipidomic data processing

The acquired raw mass data were imported to Progenesis QI software (Waters Corporation, Milford, MA, USA) for peak detection, alignment, retention time correction and normalisation. Considering a more significant number of variables than the sample size, dimensionality reduction was imperative. The unsupervised principal component analysis (PCA) was performed on generated pre-processed data to monitor the stability of the study and observe the samples' separation and exclude the presence of outliers. Thus, QC samples were used to monitor the analytical performance of the UPLC and the system's stability. The high degree of aggregation of the QC samples in the PCA model was an instrumental stability and reproducibility index.

Lipid metabolites were manually identified based on their exact masses, specific fragment and/or neutral losses [22] . A maximum error of 5mDa was defined for the attribution of the precursor ion. The statistically significant metabolites were identified by the following databases: LIPID MAPS (https://lipidmaps.org) and HMDB (https://hmdb.ca). Thus, an attempted assignment of possible features to specific compounds was performed.

5.2.2.7 Statistical analysis for biomarker analysis

The generated matrices were subsequently analyzed using MetaboAnalyst 5.0 (http://www.metaboanalyst.ca/), a comprehensive free and publicly accessible platform for metabolomics analysis that allows for applying univariate and multivariate methods. One-way parametric ANOVA followed by Tukey's post-test ($p < 0.05$) was used to identify significantly altered lipid species. The Benjamin-Hochberg-based false discovery rate (FDR) was used for multiple testing corrections, with p FDR < 0.05 considered statistically significant. In addition, the fold change (log2FC) analysis was also performed by comparing the mean intensity.

Thus, the supervised partial least square-discriminant analysis (PLS-DA) and the orthogonal partial least square-discriminant analysis (OPLS-DA) analysis were applied on normalized and Pareto-scaled matrix data to perform binary classification: i) CO vs PD, ii)

PD vs AD and iii) CO vs AD. Differential lipid biomarkers were defined as variable importance projection (VIP) > 1. Each model was validated by Q2Y (predictive variation) and R2Y (explained variation) parameters based on 10-fold cross-validation (10-FC) and leave one out cross-validation (LOOCV).

The receiver operating curves (ROC) were obtained to validate the discriminating power of the compounds responsible for each classification. ROC curve analysis is widely considered to be the most objective and statistically valid method for biomarker performance evaluation. Thus, the area under the curve (AUC) values (>70, *p-value* < 0.05) allowed the evaluation of the sensitivity and specificity of each compound to be considered as a relevant biomarker.

5.2.3   Results

To investigate how lipids are affected in diseased patients, initial global lipidome analysis of controls samples, PD and AD diseased patients was performed. To discover specific metabolites, an untargeted lipidome profiling was conducted. After data processing and normalization, 555 features/compounds were detected in negative ion mode in plasma samples.

Univariate analysis performed by one-way ANOVA yielded 88 altered lipid molecular species resulted altered when comparing all groups. The 15 first most discriminant features with tentative identification of lipids are displayed in heatmap plot (Figure 5.6). Of note, tiny clustering is observed between the groups; most controls are dispersed between two diseased groups, it could be explained by the age of healthy patients, and the possible alteration of lipidomic profile similar to those patients affected by PD or AD.

In addition, the PLS-DA was performed (R2=0.95 and relatively high Q2=0.49), a score plot of the PLS-DA showed a relatively clear group clustering according to component 1 (9.8%), Figure 5.7. As revealed by the top 20 identified features selected based on the VIP score, the previous lipid trend is repeated.

Thus, the alterations in several molecular species seems to belong to sphingolipids and glycerophospholipids, triglycerides and fatty acids metabolism, and the alteration in such lipid species are evidenced by both univariate and multivariate analyses.
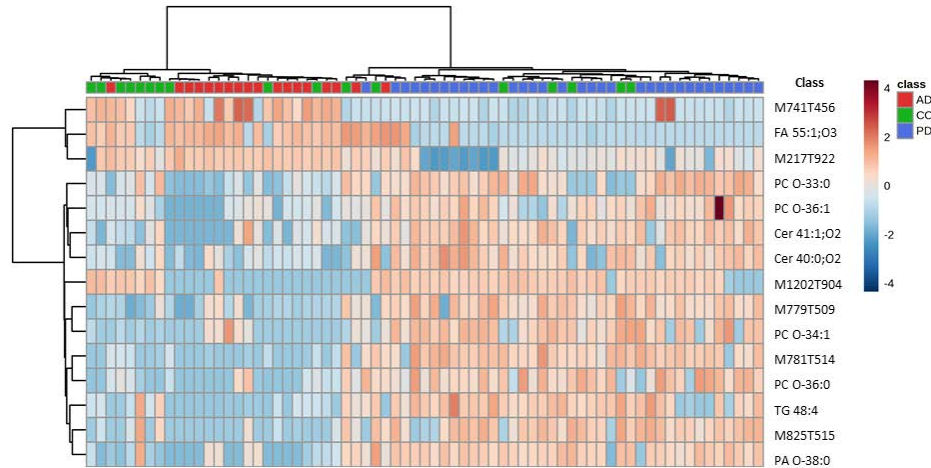
Figure 5.6. Heatmap plot displaying patient clustering (control samples (CO), Parkinson and Alzheimer patients (PD and AD), respectively) according to one-way ANOVA followed by Tukey's post-test (p< 0.05; FDR-adjusted). Identified lipid species are shown in rows, while samples are displayed in columns, according to cluster analysis (clustering based on Euclidean distance and Ward clustering algorithm. Each coloured cell on the heatmap plot corresponds to values above(red) or below (blue) the mean normalised peak intensity for a given compound. Abbreviations: Cer: ceramides; FA: fatty acids; PA: phosphatidic acids; PC: phosphatidylcholines; TG: Triglycerides.

Figure 5.7. Multivariate analysis of plasma lipidome from healthy controls (CO), Parkinson's (PD) and Alzheimer's (AD) diseased patients. (A) Score plot of partial least square-discriminant analysis (PLS-DA); (B) Top 20 identified lipid compounds according

to component 1 values of the PLS-DA model. Abbreviations: Cer: ceramides; DG: Diacylglycerols; FA: fatty acids; PA: phosphatidic acids; PC: phosphatidylcholines.
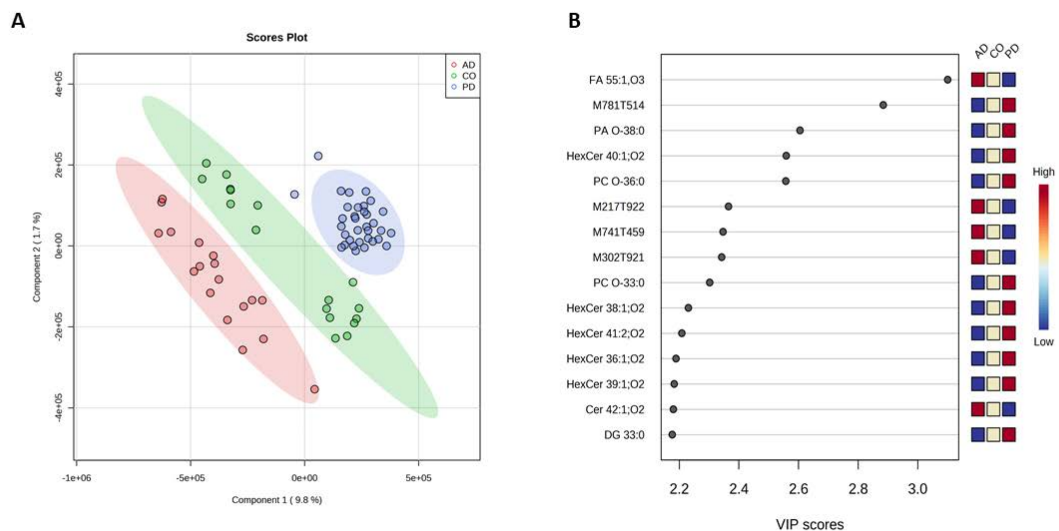
Unfortunately, some features that contribute to group differentiation remained unidentified. But those that we tentatively identified are perfectly in line with previous data, revealing the differences between PD and AD lipid profiles. Thus, we decided to examine specific differences between the groups, performing a pairwise comparison.

5.2.3.1 Pairwise comparison

To explore the metabolic differences between groups, three binary classifications were performed using the supervised OPLS-DA method, which maximises the distance between groups and identifies essential variables to the classification based on the VIP score. Each separation was validated by a cross-validation algorithm showing relatively high R2 and Q2 values, and increased performance from components 1 to 3 was observed. As shown in Figure 5.8, all the groups are spatially segregated, one from each o. Thus, the control group were separated from diseased groups, CO vs PD (R2Y =0.69 and Q2=0.32); CO vs AD (R2Y =0.62 and Q2=0.32), respectively. In addition, PD and AD patients were also significantly separated (R2Y =0.74, Q2=0.59).
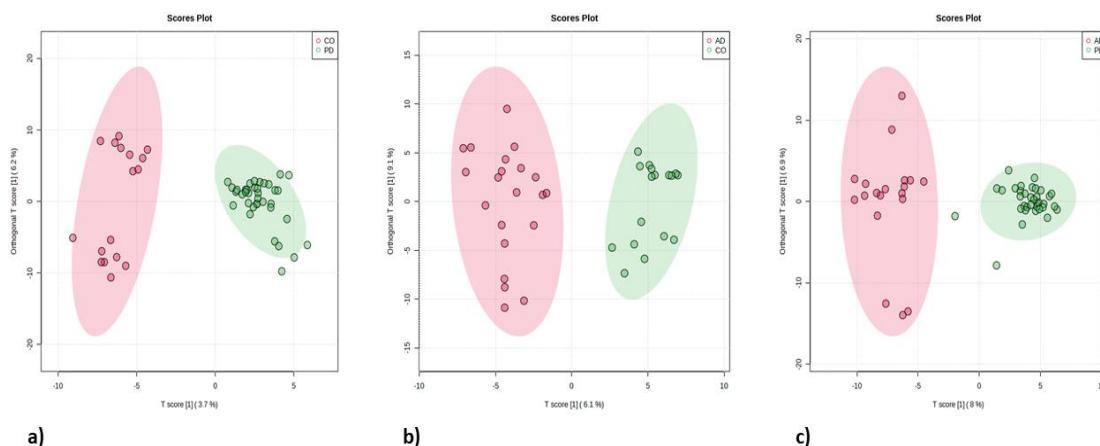
Figure 5.8. Pairwise comparison by orthogonal partial least-squared discriminant analysis (OPLS-DA) score plot of the UPLS-MS/MS data in ESI (-) mode. Before statistical analysis the data were Pareto-scaled. Score plot of the OPLS-DA revealing a clear segregation of groups a) healthy controls and Parkinson's patients, b) controls and Alzheimer patients, and c) Alzheimer's and Parkinson's patients.

According to the score OPLS-DA plots, the control and Alzheimer's groups were less homogeneous than the Parkinson groups. The diverse co-morbidities of the patients in these groups can explain this heterogeneity.

A good separation between PD and AD samples can be observed in the dendrogram obtained by performing the hierarchical clustering (HC) algorithm (Figure 5.9), with a Euclidean distance measure (in the scale distance from 1 to 60) and the Ward clustering algorithm. As observed, some overlapping samples are present, indicating the similarities in metabolic profile between Parkinson's patients and Alzheimer patients.
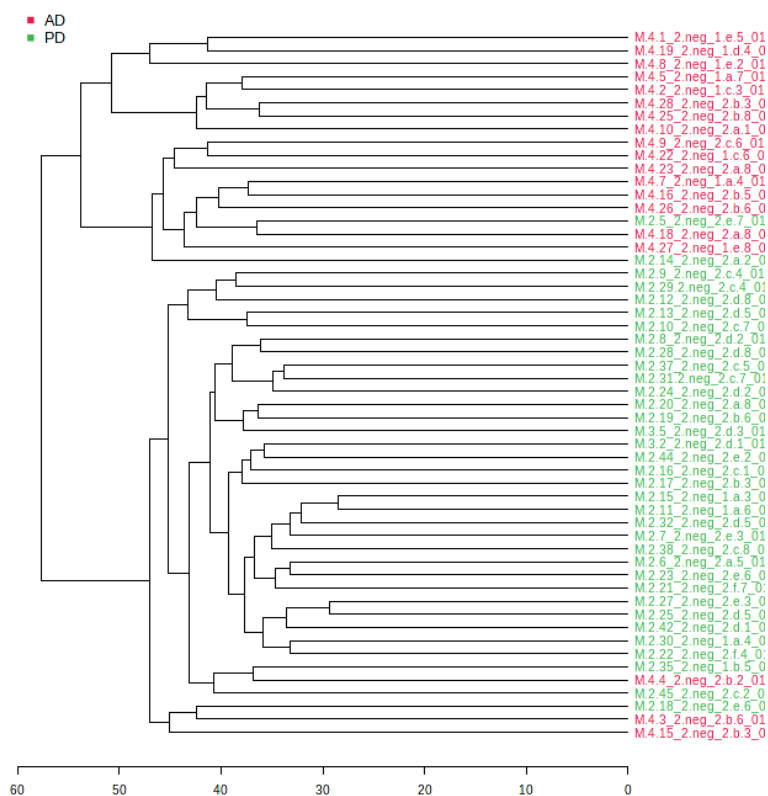
Figure 5.9. Hierarchical clustering dendrogram of samples using the Euclidian distance measure and the Ward clustering algorithm.

Differences in lipidome alterations evidenced by OPLS-DA models need to be interpreted cautiously since univariate statistics yielded fewer specimens (Table 5-11). Therefore, an attempted assignment was also performed for these compounds based on the minimum m/z error and more probable neutral loss.

These results reinforce the previous one, comparing three groups and confirming the involvement of sphingolipid and glycerophospholipid metabolism in PD and AD diseases.

162

Table 5-11. The information on selected biomarker panels.

M/Z	+/M/Z	NAME	ION	CATEGORY	T.STAT	P.VALUE	LOG2(FC)	VIP
PD VS AD								
782.6632	0.0117	HexCer 40:1;O2	[M-H]$^-$	Sphingolipids	-4.57	0.00003	-1.4582	2.43
764.5892	0.0081	PC O-32:0	[M+Formate]$^-$	Glycerophospholipids	-5.46	0.00000	-1.7798	2.21
312.7256	0.0103	CoA 7:1;O4	[M-3H]$^{3-}$	Fatty Acyls	4.20	0.00011	1.258	2.24
CO VS AD								
811.6900	0.0202	SM 43:2;O2	[M-CH$_3$]$^-$	Phosphosphingolipids	-3.65	0.00083	-1.3194	2.00
764.5892	0.0081	PC O-32:0	[M+Formate]$^-$	Glycerophosphocholines	-3.54	0.00114	-1.3693	1.91
CO VS PD								
782.6633	0.0117	HexCer 40:1;O2	[M-H]$^-$	Sphingolipids	-4.25	0.00009	-1.3118	2.63
301.7632	0.0494	LPC 24:1	[M-2H]$^{2-}$	Ceramides	3.85	0.00032	1.1784	2.54

Thus, ceramide HexCer (40:1;O2) compared significantly in both pairwise comparisons, namely the differentiation of Parkinson patients and controls and Parkinson and Alzheimer subjects. Of note, most samples of the Parkinson's group displayed increased levels of HexCer (40:1; O2) when compared to CO or AD. On the other hand, the putative biomarker sphingomyelin (SM) 43:2; O2, the ceramide biochemical precursor, was under-expressed in the AD group compared to controls. Likewise, the phosphatidylcholine (PC) O-32:0 is the most noticeable change (*p-value* < 0.05) decreased in AD patients when compared to PD or controls. Interestingly, PD groups showed increased levels of PC (O-32:0) compared to the AD group and reduced levels of LPC 24:1 when compared to controls. Of note, the putative compound CoA 7:1;O4 resulted in discriminative between AD and PD patients, suggesting the underlying metabolic mechanism that involve higher expression of this compound in AD but not in AD pathogenesis. Therefore, major lipidome changes in the plasma are linked to altered sphingolipid and phospholipid metabolism, as were shown previously comparing the changes in plasma of these compounds between groups are shown in Figure 5.10.
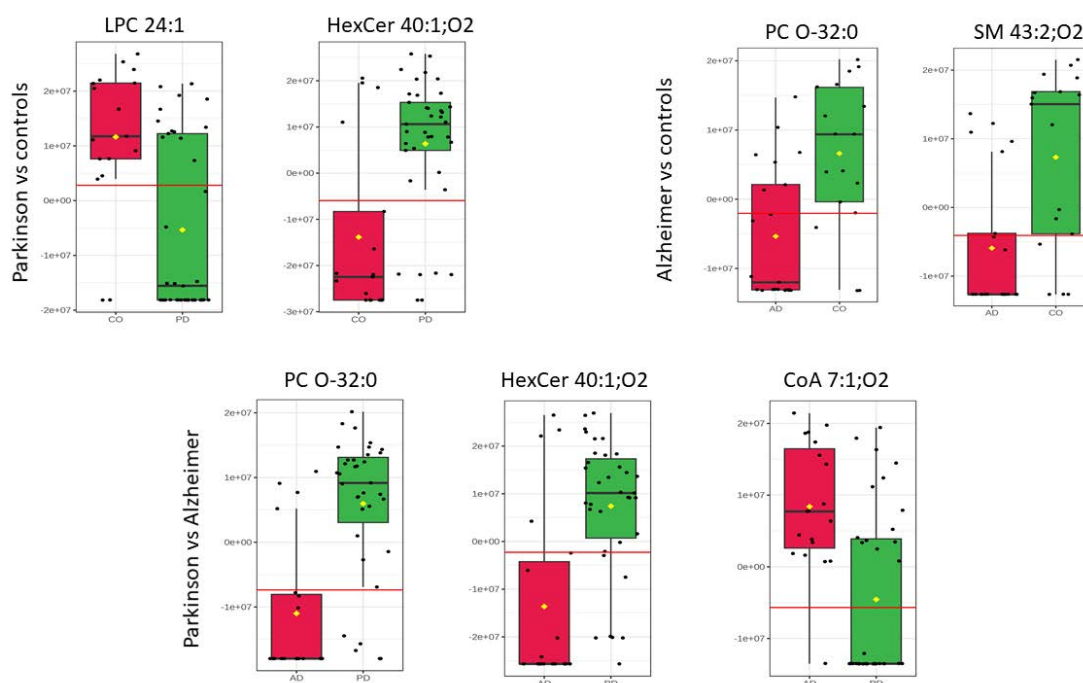
Figure 5.10. Box and whiskers plot illustrating increased and decreased intensity levels of p altered content of lipid biomarkers, identified in plasma by comparing a) healthy controls and Parkinson's patients, b) controls and Alzheimer patients, and c) Alzheimer's

and Parkinson's patients. Medians, interquartile ranges (boxes), minimal and maximal values (whiskers) and missing values (black dots) are displayed.

5.2.3.2 Predictive performance of lipid biomarkers by AUC

The AUC of ROC curves of the potential lipid biomarkers were calculated to validate the discriminating power of the compounds responsible for separating controls from diseased patients and to distinguish groups with neurogenerative disorders. Higher values of AUC close to 1 indicate higher prediction. Significant AUC values (above 70) were observed in all biomarkers of pair comparisons. Thus, the diagnostic performance of potential biomarkers is displayed in Figure 5.11.

Figure 5.11. The diagnostic performance of identified lipids via AUC curves for comparison between a) healthy controls and Parkinson's patients, b) controls and Alzheimer patients, and c) Alzheimer's and Parkinson's patients are indicated. The AUC, 95% CI of each biomarker's sensitivity (true positive rate) and specificity (false positive rate) is displayed.

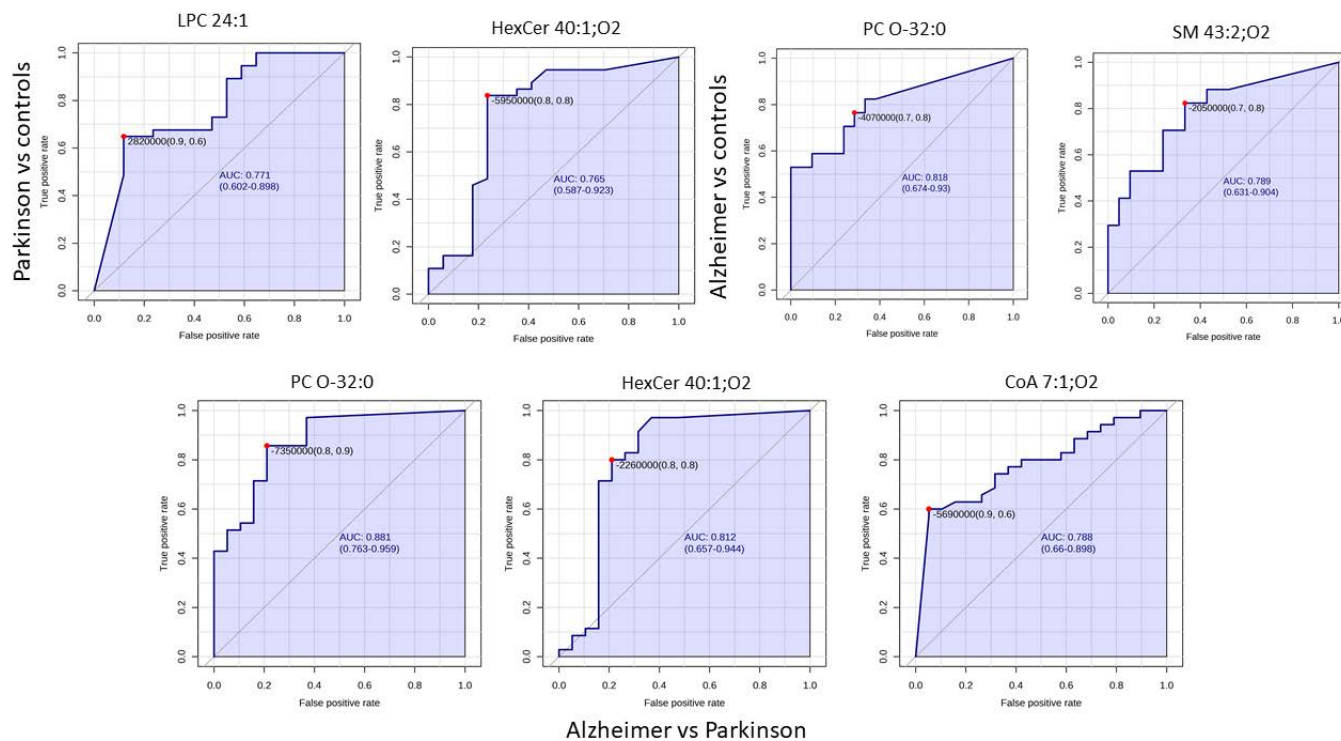The classification performance of the AUC models for each biomarker panel are summarized in Table 5-12.

Table 5-12. Results of diagnostic performance of the putative metabolite panel in pairwise comparison: Alzheimer's and Parkinson's patients; controls and Alzheimer patients; healthy controls and Parkinson's patients.

Name	AUC	Sensitivity	Specificity
PD vs AD			
HexCer 40:1;O2	0.806	80.00%	78.00%
PC O-32:0	0.874	77.10%	78.90%
CoA 7:1;O4	0.782	60.00%	94.70%
CO vs AD			
SM 43:2;O2	0.787	70.00%	76.20%
PC O-32:0	0.808	70.00%	76.20%
CO vs PD			
Hex Cer 40:1;O2	0.765	78.40%	76.50%
LPC 24:1	0.768	62.20%	88.20%

As note, all the AUC models showed relatively high diagnostic performance. Nevertheless, the biomarker panel to discriminate between Parkinson's and Alzheimer's disorder reached the highest AUC values for PC O-32:0 (0.874) among other comparative models. Moreover, this metabolite obtained good diagnostic performance in the classification of CO and AD groups (70% and 76.20 % in sensitivity and specificity, respectively) confirming its importance in AD pathogenesis. In addition, the biomarker CoA 7:1; O4 from the panel PD vs AD, showed the highest specificity values (94.70%)

among all other discriminative lipid metabolites, suggesting its important contribution for PD and AD segregation.

These results confirm that identified specimens, mainly choline-dependent phospholipids and ceramides might have a great promise to be considered as biomarkers for.

5.2.4 Discussion

The role of lipids in CNS and their implementation in multiple cellular mechanisms, membrane fluidity control, synapse stabilisation and transmission of electrical signals is widely known. Thus, the alteration of lipid metabolism in the CNS is addressed to neurodegenerative diseases and disorders, such as Parkinson's disease, Alzheimer's disease and many other injuries of CNS [11].

We evaluated and compared the lipidome profile of patients affected by neurodegenerative diseases such as Parkinson's and Alzheimer's. To the best of our knowledge, this is the first study focusing on both neuro disorders. Usually, very little importance is addressed because PD and AD present overlapping symptoms and could share similar metabolic profiles, which lead to misdiagnosis. Most previous studies focused only on the difference between one neurodegenerative disorder and normal samples and in lucky cases, on the progression of the same disease. Herein, the power of advanced untargeted metabolomic analyses based on UPLC-QTOF-MS was applied to gain insight into PD and AD mechanisms and specific plasma lipid biomarkers. Blood-based biomarkers are still not routinely implemented in clinical practice but may be helpful since there is less risk of complication in older patients than with CSF sampling. By performing binary classification, the metabolites with high missingness or known drug metabolites were excluded as non-informative. Therefore, only a few statistically relevant metabolites were identified and considered prognostic. In addition, we performed a tentative assignment of the compounds. Thus, significant differences in patients' profiles were found in the sphingolipid and glycerophospholipid categories. Interestingly, two of the identified compounds (HexCer 40:1; O2 and PC O-32:0) were

repeated in different classifications, reinforcing their significance as a discriminative biomarker of PD and AD.

Herein, it was observed that choline-based metabolites could be involved in the pathogenesis of Parkinson's and Alzheimer's diseases. Thus, lower phosphatidylcholine (PC) levels in AD patients were observed by performing both binary classifications: AD vs CO and AD vs PD, respectively. The decreased plasma PC levels in AD patients have been described previously. The targeted study by Mapstone et al.[23] showed that the pre-clinical group of AD patients would have a depletion of PC metabolites in the future. Likewise, Kim et al. [24] identified a panel of three PCs that were decreased in plasma of (younger and older) AD participants compared to normal controls. This evidence, joined with our findings, reinforces the theory that peripheral lipids are implicated in AD pathology.

On the other hand, PC's spontaneous hydrolysis or enzymatic degradation is known to be responsible for lysophosphatidylcholine (LPC) generation [25]. It was shown that PC and LPC mechanism alterations are implicated in many diseases [26–28], including neurological disorders such as PD and AD. Recently, the study of Miletić Vukajlović et al. [29] investigated the association of PC/LPC ratio and the stages or disease progression of Parkinson, confirming that PD patients had elevated PC/LPC ratios regardless of the stage or duration of the disease. Thus, they assumed that oxidative stress might induce modifications in enzymes responsible for the conversion of the investigated species. In addition, López de Frutos et al. [30] have found elevated serum PC levels with simultaneously decreased LPC levels, rendering an increased PC/LPC ratio. These findings are perfectly congruent with the obtained results in our study. Thus, PC levels were increased in the PD group when comparing PD and AD patients. Meanwhile, the LPC levels were decreased, comparing PD to normal controls. Globally, it could confirm the theory about elevated plasma PC/LPC ratio in Parkinson's disorder.

Similarly, to the controversial presence of phosphocholines in PD and AD patients, we also found the different levels of ceramide derivates between neurodegenerative disorders. Several lines of evidence implicate various sphingolipids in neuronal signalling and toxicity [31,32]. Thus, SM is involved in signal transduction and regulating inflammatory processes, such as response to oxidative stress [33]. Hydrolysis of SM

produces ceramides (Cer), which are known to mediate the relationship between $A\beta$ pathology and neurodegeneration. Therefore, the perturbation in SM/Cer homeostasis might contribute to neurodegeneration and the relationship between SM and aspects of neurodegeneration is widely explored in AD [34]. The lipidomic plasma analysis of Baloni et al. identified the SM ratio as a strong intermediate trait for sphingolipid dysregulation in AD. In addition, the study of Huo et al. [14] demonstrated a decreased level of SM in the AD cohort. Therefore, the AD group has shown a reduced sphingomyelin level herein compared to healthy controls. Our findings are perfectly consistent with the literature evidence confirming the relationship between sphingomyelins and AD cognitive impairment.

Meanwhile, decreased sphingolipids contribute to AD disorders; on the contrary, increased levels of ceramide species (e.g. mohesylceramides or lactosylceramides) contribute to PD pathophysiology [35]. Changes in sphingolipids metabolism and those associated with PD are usually linked to genetic conformation involved in PD impairment. For example, mutations in the GBA gene, encoding for the lysosomal enzyme glucocerebrosidase, catalyse the synthesis of ceramide from glucosylceramides (GlcCer), leading to the accumulation of GlcCer in brain and blood [36]. Thus, Cer: GlcCer ratio alterations could contribute to alpha-synuclein accumulation in glial cells [37]. Possibly, the identified ceramide HexCer (40:1;O2) showed higher PD levels than controls or AD patients. The accumulation of ceramides leads to neurotoxicity; thus, it is expected to find the same levels of ceramides in both neuro disorders. Nevertheless, it was shown that an increase in ceramide levels is detected during the initial stage of dementia and decreases afterwards during the course of AD [38].

Interestingly, among three discriminant compounds identified to discriminate AD from PD patients, CoA 7:1;O4 biomarker was evaluated. Acetyl-CoA plays a key role in the proper functioning of the cell, such as glycolysis, fatty acid synthesis or TCA cycle in mitochondria. Thus, dysregulation and deficit in mitochondrial metabolism are often linked to cognitive dysfunction in AD patients, meanwhile, increased acetyl-CoA levels provide neuroprotection [39]. Therefore, further studies should be done to completely understand and justify the contribution of this metabolite to neurodegenerative

pathologies. Nevertheless, these preliminary results provided meaningful lipidomic information about PD and AD plasma profiles.

Although only some specific PD and AD biomarkers were identified using the untargeted lipidomic approach, the current findings are perfectly congruent with previously reported studies. Therefore, they could be the standpoint for hypothesis-driven targeted analysis. Nevertheless, certain limitations should be taken into consideration. Firstly, a larger sample size and a more homogenous cohort of participants must confirm the results. Furthermore, the effects of genetic background, nutrition, and other factors that might impact blood lipid composition should have been taken into account. In addition, many evaluated features remained unidentified, indicating that the endogenous metabolites remain poorly understood.

5.2.5   Conclusions

An UPLC-MS/MS approach was performed to study the plasma lipidomic profile of patients with neuro disorders, PD and AD respectively. Similar investigations confirmed the lipid alterations presented in this study. Furthermore, they showed statistically significant difference, thus could be a prominent biomarker to the differentiation between AD or PD and healthy controls, and between two neuro disorders that often share similar symptoms, leading to clinical misdiagnosis. Thus, significant changes in sphingolipids and glycerophospholipids in blood composition were observed responsible for patient differentiation, which could be useful for diagnostics and may lead to promising new therapeutic targets. Further, targeted and longitudinal studies are required.

5.2.6   References

[1]   E. Ray Dorsey *et al.*, "Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016," *Lancet Neurol*, vol. 17, no. 11, pp. 939–953, Nov. 2018, doi: 10.1016/S1474-4422(18)30295-3.

[2]   T. Li and W. Le, "Biomarkers for Parkinson's Disease: How Good Are They?," *Neurosci Bull*, vol. 36, no. 2, pp. 183–194, Feb. 2020, doi: 10.1007/S12264-019-00433-1.

[3]   T. R. Mhyre, J. T. Boyd, R. W. Hamill, and K. A. Maguire-Zeiss, "Parkinson's disease," *Subcell Biochem*, vol. 65, pp. 389–455, May 2012, doi: 10.1007/978-94-007-5416-4_16.

[4]   S. M. A. Zella *et al.*, "Emerging Immunotherapies for Parkinson Disease," *Neurology and Therapy*, vol. 8, no. 1. 2019. doi: 10.1007/s40120-018-0122-z.

[5]   L. Parnetti *et al.*, "CSF and blood biomarkers for Parkinson's disease," *Lancet Neurol*, vol. 18, no. 6, pp. 573–586, Jun. 2019, doi: 10.1016/S1474-4422(19)30024-9.

[6]   G. Alves, K. F. Pedersen, and K. F. Pedersen, "Epidemiology of Parkinson ' s disease Epidemiology of Parkinson ' s disease," *J Neurol*, vol. 5, no. August 2015, pp. 525–535, 2008, [Online]. Available: http://dx.doi.org/10.1016/j.neurol.2015.09.012

[7]   G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, and G. Logroscino, "Accuracy of clinical diagnosis of Parkinson disease," *Neurology*, vol. 86, no. 6, pp. 566–576, Feb. 2016, doi: 10.1212/WNL.0000000000002350.

[8]   E. Deas *et al.*, "Alpha-synuclein oligomers interact with metal ions to induce oxidative stress and neuronal death in Parkinson's disease," *Antioxid Redox Signal*, vol. 24, no. 7, pp. 376–391, 2016, doi: 10.1089/ars.2015.6343.

[9]   Z. Yu *et al.*, "The significance of uric acid in the diagnosis and treatment of Parkinson disease," *Medicine (United States)*, vol. 96, no. 45. Lippincott Williams and Wilkins, Nov. 01, 2017. doi: 10.1097/MD.0000000000008502.

[10]   A. Picca *et al.*, "Circulating amino acid signature in older people with Parkinson's disease: A metabolic complement to the EXosomes in PArkiNson Disease (EXPAND) study," *Exp Gerontol*, vol. 128, no. August, p. 110766, 2019, doi: 10.1016/j.exger.2019.110766.

[11]   R. S. Yadav and N. K. Tiwari, "Lipid Integration in Neurodegeneration: An Overview of Alzheimer's Disease," *Molecular Neurobiology*, vol. 50, no. 1. Humana Press Inc., pp. 168–176, Oct. 02, 2014. doi: 10.1007/s12035-014-8661-5.

[12]   S. K. Abbott *et al.*, "Altered ceramide acyl chain length and ceramide synthase gene expression in Parkinson's disease," *Movement Disorders*, vol. 29, no. 4, pp. 518–526, 2014, doi: 10.1002/MDS.25729.

[13]   T. Moll, J. N. G. Marshall, N. Soni, S. Zhang, J. Cooper-Knock, and P. J. Shaw, "Membrane lipid raft homeostasis is directly linked to neurodegeneration," *Essays in*

Biochemistry, vol. 65, no. 7. Portland Press Ltd, pp. 999–1011, Dec. 01, 2021. doi: 10.1042/EBC20210026.

[14]   Z. Huo *et al.*, "Metabolic Profiling of Cognitive Aging in Midlife," *Front Aging Neurosci*, vol. 12, Nov. 2020, doi: 10.3389/FNAGI.2020.555850/FULL.

[15]   A. Nicholls, G. Theodoridis, and I. D. Wilson, "Global metabolic profiling in health and disease," *Global Metabolic Profiling: Clinical Applications*, pp. 2–5, 2014, doi: 10.4155/EBO.13.339.

[16]   W. Han, S. Sapkota, R. Camicioli, R. A. Dixon, and L. Li, "Profiling novel metabolic biomarkers for Parkinson's disease using in-depth metabolomic analysis," *Mov. Disord.*, vol. 32, no. 12, pp. 1720–1728, Dec. 2017, doi: 10.1002/mds.27173.

[17]   Y. Shao *et al.*, "Comprehensive metabolic profiling of Parkinson's disease by liquid chromatography-mass spectrometry", doi: 10.1186/s13024-021-00425-8.

[18]   "Alzheimer plasma metabolomic."

[19]   C. Peña-Bautista, L. Álvarez-Sánchez, M. Roca, L. García-Vallés, M. Baquero, and C. Cháfer-Pericás, "Plasma Lipidomics Approach in Early and Specific Alzheimer's Disease Diagnosis.," *J Clin Med*, vol. 11, no. 17, Aug. 2022, doi: 10.3390/jcm11175030.

[20]   R. Casanova *et al.*, "Blood metabolite markers of preclinical Alzheimer's disease in two longitudinally followed cohorts of older individuals," *Alzheimer's and Dementia*, vol. 12, no. 7, pp. 815–822, Jul. 2016, doi: 10.1016/J.JALZ.2015.12.008.

[21]   C. Pizarro, I. Arenzana-Rámila, N. Pérez-Del-Notario, P. Pérez-Matute, and J. M. González-Sáiz, "Plasma lipidomic profiling method based on ultrasound extraction and liquid chromatography mass spectrometry," *Anal Chem*, vol. 85, no. 24, pp. 12085–12092, Dec. 2013, doi: 10.1021/AC403181C.

[22]   Xianlin. Han, *Lipidomics : comprehensive mass spectrometry of lipids*.

[23]   M. Mapstone *et al.*, "Plasma phospholipids identify antecedent memory impairment in older adults," *Nat Med*, vol. 20, no. 4, p. 415, 2014, doi: 10.1038/NM.3466.

[24]   M. Kim *et al.*, "Association between Plasma Ceramides and Phosphatidylcholines and Hippocampal Brain Volume in Late Onset Alzheimer's Disease," *Journal of Alzheimer's Disease*, vol. 60, no. 3, p. 809, 2017, doi: 10.3233/JAD-160645.

[25]   D. Patel and S. N. Witt, "Ethanolamine and Phosphatidylethanolamine: Partners in Health and Disease," *Oxidative Medicine and Cellular Longevity*, vol. 2017. Hindawi Limited, 2017. doi: 10.1155/2017/4829180.

[26]   Y. Xu *et al.*, "Unfolding the pathophysiological role of bioactive lysophospholipids.," *Curr Drug Targets Immune Endocr Metabol Disord*, vol. 3, no. 1, pp. 23–32, 2003, doi: 10.2174/1568005310303010023.

[27]   Q. Tian, B. A. Mitchell, M. Zampino, and L. Ferrucci, "Longitudinal associations between blood lysophosphatidylcholines and skeletal muscle mitochondrial function," *Geroscience*, vol. 44, no. 4, pp. 2213–2221, Aug. 2022, doi: 10.1007/s11357-022-00548-w.

[28]   A. Grzelczyk and E. Gendaszewska-Darmach, "Novel bioactive glycerol-based lysophospholipids: New data-New insight into their function," *Biochimie*, vol. 95, no. 4, pp. 667–679, Apr. 2013, doi: 10.1016/j.biochi.2012.10.009.

[29]   J. Miletić Vukajlović *et al.*, "Increased plasma phosphatidylcholine/lysophosphatidylcholine ratios in patients with Parkinson's disease," *Rapid Communications in Mass Spectrometry*, vol. 34, no. 4, Feb. 2020, doi: 10.1002/rcm.8595.

[30]   L. López de Frutos *et al.*, "Serum Phospholipid Profile Changes in Gaucher Disease and Parkinson's Disease," *Int J Mol Sci*, vol. 23, no. 18, Sep. 2022, doi: 10.3390/ijms231810387.

[31]   S. T. Ngo, "Lipids: Key players in central nervous system cell physiology and pathology," *Seminars in Cell and Developmental Biology*, vol. 112. Elsevier Ltd, pp. 59–60, Apr. 01, 2021. doi: 10.1016/j.semcdb.2021.02.003.

[32]   E. Fahy *et al.*, "A comprehensive classification system for lipids," *J Lipid Res*, vol. 46, no. 5, pp. 839–861, 2005, doi: 10.1194/jlr.E400004-JLR200.

[33]   F. Schmitt, G. Hussain, L. Dupuis, J. P. Loeffler, and A. Henriques, "A plural role for lipids in motor neuron diseases: Energy, signaling and structure," *Frontiers in Cellular Neuroscience*, vol. 8, no. FEB. Frontiers Research Foundation, Feb. 20, 2014. doi: 10.3389/fncel.2014.00025.

[34]   A. Morrow *et al.*, "Cerebrospinal fluid sphingomyelins in Alzheimer's disease, neurodegeneration, and neuroinflammation," *J Alzheimers Dis*, vol. 90, no. 2, p. 667, Sep. 2022, doi: 10.3233/JAD-220349.

[35]   M. Alaamery *et al.*, "Role of sphingolipid metabolism in neurodegeneration," *Journal of Neurochemistry*, vol. 158, no. 1. John Wiley and Sons Inc, pp. 25–35, Jul. 01, 2021. doi: 10.1111/jnc.15044.

[36]   L. C. Guedes *et al.*, "Serum lipid alterations in GBA-associated Parkinson's disease," *Parkinsonism Relat Disord*, vol. 44, pp. 58–65, Nov. 2017, doi: 10.1016/J.PARKRELDIS.2017.08.026.

[37]   N. Plotegher, L. Bubacco, E. Greggio, and L. Civiero, "Ceramides in Parkinson's Disease: From Recent Evidence to New Hypotheses," *Front Neurosci*, vol. 13, no. APR, 2019, doi: 10.3389/FNINS.2019.00330.

[38]   P. Katsel, C. Li, and V. Haroutunian, "Gene expression alterations in the sphingolipid metabolism pathways during progression of dementia and Alzheimer's disease: A shift toward ceramide accumulation at the earliest recognizable stages of Alzheimer's disease?," *Neurochem Res*, vol. 32, no. 4–5, pp. 845–856, Apr. 2007, doi: 10.1007/S11064-007-9297-X/FIGURES/3.

[39]   A. Currais *et al.*, "Elevating acetyl-CoA levels reduces aspects of brain aging," *Elife*, vol. 8, Nov. 2019, doi: 10.7554/eLife.47866.

Chapter 6

Amyotrophic lateral sclerosis

Abstract

This chapter summarises all the studies performed on Amyotrophic lateral sclerosis disease. ALS is characterised by progressive muscle weakness, with rapid progression and fatal outcomes only a few years after diagnosis. In addition, the tardive appearance of ALS clinical symptoms delays the diagnosis and appropriate treatment. Moreover, misdiagnosis often happens with other motor neuro disorders. Here, we investigated the potential of both FTIR and ATR-FTIR spectroscopy as rapid tools for discriminating ALS patients from controls and ALS progression. Within this method, we aimed to reveal that the adopted analytical strategy may extract significant bio-spectroscopic markers, providing excellent group separation and achieving high classification accuracy. It is well-known that lipids exert various functions in the central nervous system, including roles in cell structure, synaptic transmission, and multiple metabolic processes. Thus, a non-targeted lipidomic approach using UPLC-MS/MS was performed to unravel alterations in one or more lipid specimens and metabolic pathways where they are possibly involved in ALS disease.

Resumen

Este capítulo resumen todos los estudios realizados sobre la enfermedad de Esclerosis Lateral Amiotrófica (ELA). La ELA se caracteriza por una debilidad muscular progresiva, con una rápida progresión y resultados fatales solo unos pocos años después del diagnóstico. Además, la aparición tardía de los síntomas clínicos de la ELA retrasa el diagnóstico y el tratamiento apropiado. Par no decir que a menudo ocurre un diagnóstico erróneo con otros trastornos neurológicos motores. En este trabajo, investigamos el potencial de la espectroscopía FTIR y ATR-FTIR como herramientas rápidas para discriminar pacientes con ELA de los controles y para evaluar la progresión de la ELA. Con este método, pretendíamos demostrar que la estrategia analítica adoptada puede extraer marcadores bio-espectroscópicos significativos, proporcionando una excelente separación de grupos y logrando una alta precisión de clasificación. Es bien sabido que los lípidos ejercen diversas funciones en el sistema nervioso central, incluyendo roles en la estructura celular, la transmisión sináptica y múltiples procesos metabólicos. Por lo tanto, se realizó un enfoque lipidómico no dirigido utilizando UPLC-MS/MS para descubrir alteraciones en una o más especies de lípidos y en las vías metabólicas en las que posiblemente están involucrados en la enfermedad de la ELA.

6 CHAPTER 6. AMYOTROPHIC LATERAL SCLEROSIS

6.1 EMERGING FTIR-CHEMOMETRIC APPROACH FOR ALS PATIENTS' DISCRIMINATION BASED ON SELECTED SPECTRA BIOMARKERS

6.1.1 Introduction

Amyotrophic lateral sclerosis (ALS) is a rare progressive neurodegenerative disease that affects upper and/or lower motor neurons [1]. Progressive muscle weakness and atrophy are typical ALS hallmarks with fatal outcomes (3-5 years after the symptoms' onset) due to respiratory muscle paralysis [2]. Different pathogenic mechanisms showed to account for the deleterious effect of ALS, such as genetic mutations, e.g., the gene encoding the protein superoxide dismutase 1 (SOD1)[3] that are present in a significant percentage of familial cases or mutations in the fused in sarcoma (FUS) gene [4], the presence of mutant proteins such as TDP-43, aberrant RNA metabolism and protein aggregation, and environmental influences. In addition, there is increasing evidence indicating that dysregulation of lipid homeostasis is involved in the neurodegenerative and neuroinflammatory disorders such as ALS [5]. Lipids play a crucial role in the central nervous system (CNS), especially in astrocytes, where they are involved in energy generation, membrane fluidity and cell to cell signalling. Thus, alteration in lipid metabolism, e.g., dysregulation in sphingolipids and glycosphingolipid's structure and metabolism in astrocytes, contribute to pathogenic mechanism in neurodegenerative disorders[6–8].

Moreover, among different lipids present in neuro-motor cells, cholesterol is of particulate interest, regulating cell membrane flexibility. Many emerging evidences showed that higher serum cholesterol levels may prolong the survival in ALS patients [9]. Nevertheless, no specific biomarkers exist for ALS diagnosis.

Currently, ALS diagnosis is still challenging; the process relies essentially on clinical assessment of symptoms, physical examination, and confirmatory electromyography tests [10]. In addition, the arrangement of symptoms displayed by patients during the course of the disease reflects the progressive loss of motor neurons. Thus, patients with ALS share some overlapping features with other commonly known neurodegenerative

disorders such as Parkinson's, Alzheimer or other neuro-disorders (ON), namely, muscular dystrophies such as Myotonic Dystrophy or Becker's Muscular Dystrophy [11–16]. This issue makes their differentiation a challenging and time-consuming task. In addition, the misdiagnosis could occur not only because of the resemblance between the symptoms with other neuro disorders but either because of clinical and pathophysiological heterogeneity between ALS phenotypes [17].

Symptoms may appear only late in the disease course; therefore, there could be a significant gap between the first clinical visit and the acclaimed diagnosis. For this reason, tardive diagnosis and misdiagnosis may compromise a patient's survival rates because once the onset of the symptoms happens (functional involvement by weakness, wasting or spasticity), the neuron degeneration that has occurred is already irreversible. In addition, the delay period in ALS diagnosis prevents opportune treatment and patient prognostication in this way. Therefore, the universal and objective measure of disease progression is crucial and would benefit patient well-being.

In this context, metabolomic approaches investigating changes in the whole metabolome are gaining momentum. Investigating the metabolomic profile may reveal novel dysfunctional pathways suitable for therapeutic targeting. Since metabolites are the final product of the cellular process, they are thought of as a reflection of possible cellular anomalies and dysregulations [18]. In addition, metabolomics is particularly suitable for studying easily accessible matrices such as saliva, urine or blood to provide early diagnosis and to define clinical subgroups at the metabolomics level [19].

Metabolomics studies usually are coupled with computational methods such as machine learning techniques, which are becoming highly helpful for the illness monitoring process [20,21]. The need for early and reliable differential diagnosis of subgroups of ALS patients or ON disorders that share some common features is imperative for starting advanced specific therapies as soon as possible.

For this reason, herein, we performed an FTIR-based approach coupled with chemometric techniques on a small volume of blood supernatant to effectively explore disease processes and find early signs indicative of ALS onset. FTIR presents multiple

advantages for this type of analysis since it is rapid, non-invasive, reagent-free and cost-affordable [22].

Thus, besides being an established tool in analytical chemistry, FTIR spectroscopy is an emerging tool in the differential diagnosis between diseased and healthy patients and is always one step forward to incorporating it into clinical reality [23–35].

Changes and disarrangements in the organism are reflected in metabolite composition in such a complex fluid as blood, thus, providing a unique FTIR signature that can be used as a "barcode" of the disease. Herein, we sought to identify the spectral differences, readily detectable through FTIR, associated with healthy and diseased patients or specific points of the ALS pathological course. This study focuses on the detection of certain spectra biomarkers related to amyotrophic lateral sclerosis progression stage, which can be used to discriminate between affected patients from the controls and from patients affected by ON.

6.1.2 Experimental section

6.1.2.1 *Study approval*

The study protocol has been approved by the local ethics committee (Comité de Etica de la Investigación de la Comunidad de Aragón or CEICA) (CP-CI PI18/078). Blood samples from patients were obtained with written informed consent prior to inclusion in the study to publication of their case details, which has been conducted according to Declaration of Helsinki principles, and according to the Directive 2004/23/EC of the European Parliament and of the Council and to the institutional ethical committees of Niguarda Ca'Granda Hospital (approval N· 636-122015;23-12-2015). Participants were identified by number, not by name.

6.1.2.2 *Experimental design*

This study included a total of 76 blood samples of a cohort of participants from Niguarda Ca'Granda Hospital in Milan (Italy) matched for age and gender, whenever possible. The samples analysed in this study were obtained after first centrifuging each Pax tube containing the blood sample for 10 minutes at 3000-5000 rpm and incubated for 2 hours at room temperature, following the recommendations of the commercial kit [36]. The supernatant samples recovered from this first centrifugation were used to

perform the analysis in this study. In addition, these samples were preserved at -80 °C for further use. A total of 35 ALS patients, 34 healthy controls and 7 patients with other neuropathies (ON) were included. The ALS group was divided in two subgroups, 19 patients that were collected at first diagnosis ALS (T0) of which 16 were obtained as samples after 6 months of diagnosis/treatment ALS (T6). The ALS group included familial ALS (fALS) cases due to quadruple mutation in the ALS susceptibility genes SOD1/TDP43/FUS/c9orf72. Similarly, considering the availability of the samples at T6 in the same participants, the healthy control group was divided into 19 participants at T0 of which 15 were obtained as samples at T6. Regarding the ON group, only samples of 7 participants collected at T6 were at our disposal. The ON participants were affected by: Becker's Muscular Dystrophy, Extrapyramidal syndrome, Facioscapulohumeral Muscular Dystrophy and Myotonic Dystrophy.

6.1.2.3   FTIR spectroscopy

Spectra were obtained using Spectrum Two FT-IR spectrometers (PerkinElmer) and recorded in the region 1500-1000 cm^{-1} spectral range with a resolution of 2 cm^{-1} as described in Chapter 3.3.1.1 Analysis of samples by FT-IR .

6.1.2.4   Data analysis

The averaged spectra were processed with Unscrambler 11 chemometric software package (version 11.0, Camo Software, Oslo, Norway) and a Parvus software[37], a free and open-source machine learning source developed at the University of Genova, Italy. In order to correct problems that can arise during the spectral data acquisition such as random noise, baseline distortions or light scattering, extended multiple scatter correction (EMSC) was applied. Afterwards the Savitzky-Golay normalization, performing second derivates with nine points was applied. The application of efficient variable selection technique, such as SELECT, to find the informative features that would allow successful prediction was studied. SELECT is an important variable selection method that select essential variables and decorrelate them from other variables based on the maximum correlation weight. Therefore, only really relevant and important variables are extracted and used in the subsequent classification steps. Thus, after performed the data dimensionality reduction, eliminating futile spectra variable, linear discriminant analysis (LDA) approach was performed [38,39]. This two-step strategy has

been widely applied in differential diagnosis of other disorders performed by our group [40–42] ; thus, a careful reading of the theory is encouraged for its comprehension. A leave-one-out (LOO) cross validation (CV) was used in every classification step to optimise classifications, and external validation to evaluate prediction ability.

A small test set, due to the small number of samples in some categories were used to calculate model prediction ability. The evaluation of the model is conducted based on accuracy, sensitivity and selectivity of the model towards each group of patients.

6.1.3 Results

This study focuses on the detection of certain spectra biomarkers related to ALS progression stage, which can be used as input features to discriminate affected patients from the controls and from patients affected by other neuropathies. With this aim, a dataset of healthy control and diseased blood samples were collected and the IR spectra obtained by SELECT algorithm was used as predictor variables to create the following LDA classification models: i) ALS versus controls versus ON, ii) ALS (T0) versus ALS (T6) versus ON iii) ALS (T0 and T6) versus controls, iv) ALS T0 versus HC (T0), v) only ALS progression stage (T0 versus T6).

All the spectra were registered in the fingerprint region, the so-called ''barcode'' of the metabolic changes. However, our previous studies and experience showed that other regions are usually non-informative. Thus, to reduce the analysis time, all the spectra were registered in the 1500-1000 cm^{-1} region where the absorption of proteins, carbohydrates, lipids and nucleic bases occurs; it is even more suitable for clinical reality. Therefore, all the analysis was performed in this region.

6.1.3.1 ALS(T6) versus HC(T6) versus ON

For this first classification, 33 subjects divided into three main groups, ALS (T6), healthy subjects (T6) and ON, were included. In order to obtain as much as possible congruent and reliable classificatory results, only patients collected after six months of the disease progression were selected for this classification. The SELECT strategy was performed to investigate the most important variables that contribute to patient group differentiation. Herein, 11 spectra biomarkers were needed for optimal LDA separation between the three groups of blood.

The full cross-validation was applied on the second derivate of the training set's spectra, and the resulting prediction matrices are displayed in Table 6-1. Among 29 test samples, none were misclassified, leading to a performance of 100% in sensitivity, specificity and precision. Therefore, healthy patients could not be confused with having one of two pathologies: ALS or other neuropathies. Furthermore, the excellent specificity of the model towards three groups reflects the high ability of FTIR spectra in detecting ALS samples and discriminating them from other motor-related pathologies. Thus, the two medical conditions are perfectly separated.

The prediction performance with four spectra in the test set was performed using the model built here on the training test. The overall performance obtained in external prediction gave as good results as in classification found for the training set, 100%, respectively. The classification ability to discriminate between tree groups could be visually appreciated in Figure 6.1. The patient profile of these patients' groups seems to be unequivocally separated by the score difference of canonical variables. Likewise, the score differences between the first and second canonical variable (x-axis) and the score difference between the first and third canonical variables (y-axis) seems to have the same contribution in separating clearly healthy subjects and pathologic subject between them.

Table 6-1. Prediction matrices and percentages of correctly classified samples in both classification and internal/external validation corresponding to the SELECT-LDA performed to discriminate between HC, ALS and ON patients.

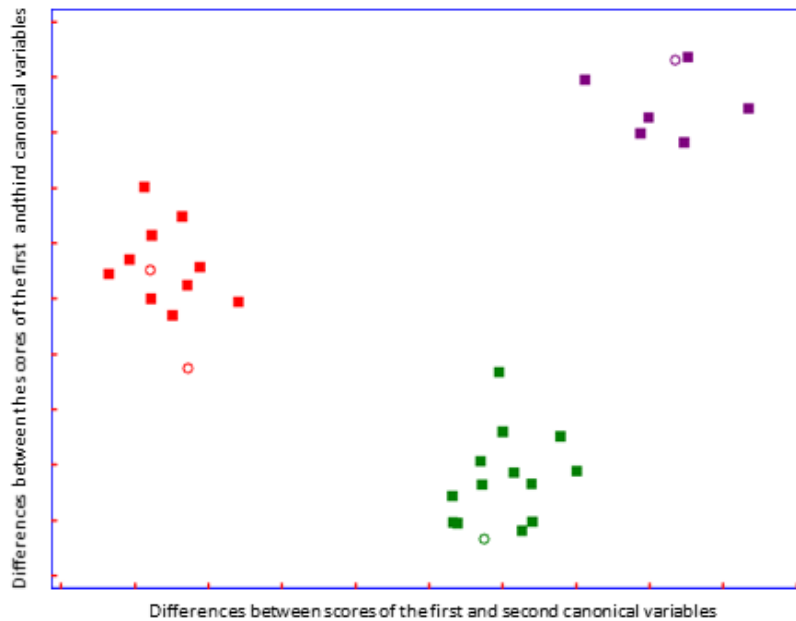| Category						
Pathology	ALS	HC	ON	Sensitivity%	Specificity %	Precision %
Training set						
ALS	10	0	0			
HC	0	13	0			
ON	0	0	6	100	100	100
Test set						
ALS	2	0	0			
HC	0	1	0			
ON	0	0	1			

Figure 6.1. Plot of the differences between discriminant scores for supernatant blood samples after performing SELECT-LDA in the first FTIR-based classification approach. The three groups of samples considered are labelled as: healthy control (■), ALS (■), and ON (■) subjects. Test samples are displayed as unfilled circles (○).

6.1.3.2 ALS (T0) versus ALS (T6) versus ON

Herein, the classification of 40 recorded spectra of recently diagnosed ALS patients (T0), patients after six months (T6) of diagnosis and patients with other neurodegenerative motor pathologies was performed. For this discrimination, 16 variables previously selected by the SELECT algorithm were used. The training and test sets yielded the prediction matrices shown in Table 6-2.

Table 6-2. Prediction matrices and percentages of correctly classified samples in both classification and internal/external validation corresponding to the SELECT-LDA performed to discriminate between ALS T0, ALS T6 and ON.

	Category						
	Pathology	ALS T0	ALS T6	ON	Sensitivity%	Specificity %	Precision %
Training set	ALS T0	16	0	0			
	ALS T6	2	12	0	100	94.44	100
	ON	0	0	6			
Test set	ALS T0	2	0	0			
	ALS T6	0	1	0	100	100	100
	ON	0	0	1			

The total LDA classification ability yielded 100%, meanwhile, the overall predictive ability of the model showed 94.44 %. Compared to previous optimal separation performed to distinguish patients based on the advancement of the disease, herein, among 14 ALS (T6) patients, two subjects were misclassified as patients with recently diagnosed ALS (T0). These results could reflect some issues due to the heterogeneity of ALS phenotype, and confirmed that the progression of the disease could have been more manifest in terms of metabolic evolution, as it has been previously reported [17]. Nevertheless, a test set of 4 spectra not included in the training set showed perfect prediction ability, and all patients were ideally classified. A good group separation can be appreciated in Figure 6-2. The three groups of patients affected by similar neurodegenerative disturbs are perfectly separated by the score difference between the first and third canonical variables (y-axis). Interestingly, it can be observed that the two ALS disease-carrying groups are almost not separated on the x-axis, but both ALS groups are greatly distinguished from patients carrying other neuro disorders, suggesting that ALS groups are characterised by changes in the metabolic profile that do not describe patients with ON.
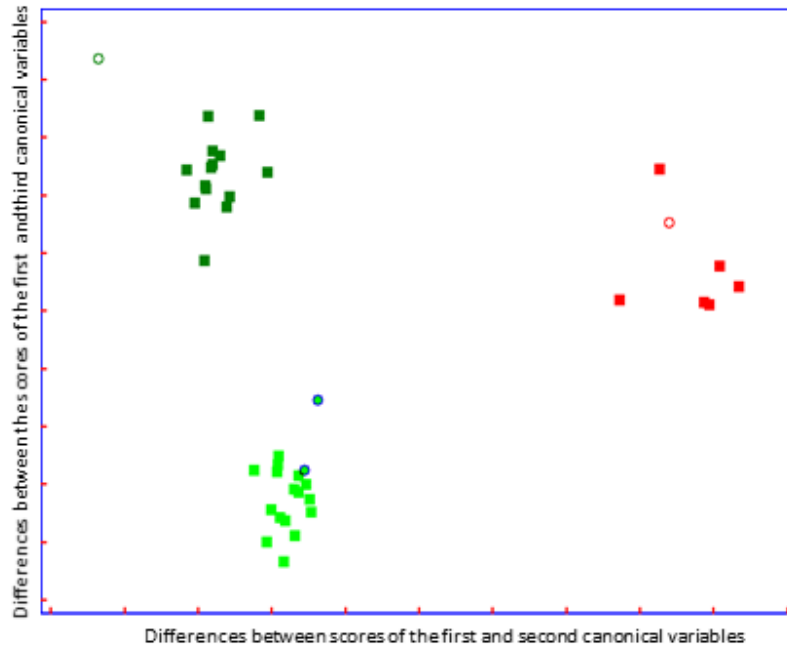
Figure 6-2. Plot of the differences between discriminant scores for supernatant blood samples after performing SELECT-LDA to discriminate between three groups of samples: ALS T0 (■), ALS T6 (■), and ON (■) subjects. Test samples are displayed as unfilled circles (○).

6.1.3.3 ALS (T0) versus ALS (T6) versus HC (T6)

Based on excellent discrimination ability of developed classification rules, we decided to test the performance of LDA in discriminate between patients affected by ALS but in different stage of the disease and healthy controls. Thus, 47 of second-derivate spectra, five of which was used as a test set was included. As in previous classification, 16 spectra variables were selected as important and decorrelated for this classification step. The resulting prediction matrix after a full cross-validation is showed below (Table 6-3).

Table 6-3. Prediction matrices and percentages of correctly classified samples in both classification and internal/external validation corresponding to the SELECT-LDA performed to discriminate between ALS T0, ALS T6 and HC.

	Pathology	ALS T0	ALS T6	HC	Sensitivity%	Specificity %	Precision %
Category							
	ALS T0	15	0	0			
Training set	ALS T6	0	14	0	100	97.67	100
	HC	1	0	1			
	ALS T0	2	0	3			
Test set	ALS T6	0	2	0	100	100	100
	HC	0	0	1			

Among 29 samples belonging to ALS group none were misclassified, leading an outstanding performance in sensitivity, specificity and precision. Only one of healthy subjects was confused as having initial stage of ALS, leading to 97.67% of specificity. Nevertheless, the overall performance obtained in external prediction provided excellent results, 100%, respectively. In addition, excellent specificity of the model towards three groups and the discrimination ability of selected FTIR variables to discriminate between affected patients and healthy one could be appreciated in Table 6-3. Thus, all of conditions are perfectly separated. Herein, the discrimination between states of ALS could be observed on both, x-and y-axes. Likewise, HC patients seem isolated from ALS patients on the score difference between the first and second canonical variables, and between the first and third, respectively.
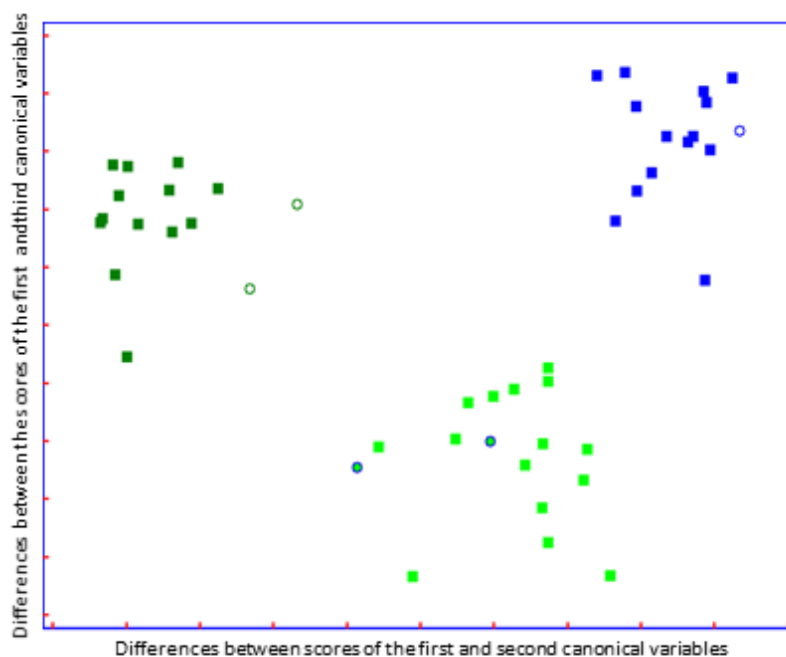
Figure 6-3. Plot of the differences between discriminant scores of blood samples after performing SELECT-LDA to discriminate between three groups of samples: ALS T0 (■), ALS T6 (■), and HC (■) subjects. Test samples are displayed as circles (O).

6.1.3.4 ALS T0 versus HC (T0)

Once the differentiation among three classes of patients was performed, we performed sub-classifications to discriminate directly between healthy controls and patients at the initial stage of ALS (T0) and between the ALS progression stage. These types of classification are the most exciting and helpful in the clinical setting to obtain quick screening and diagnosis without recurring laborious and time-consuming clinical tests. Therefore, patients at the initial ALS stage and controls whose blood was collected at the same time as those of ALS (T0) were included in this first sub-classification. In compliance with the rules followed before, the number of discriminant variables changed so that training objects were at least three times greater than the number of final selected wavenumbers. LOO CV was performed on autoscaled data. A test subset of five samples selected randomly, four for the HC group and one for ALS (T0), was used to optimise and validate classifications. Herein, 100% in classification and 100% in prediction (internal and external) were achieved (Table 4). Herein, the number of features was reduced from 500 to only ten variables, proving this classification strategy's outstanding performance. A discriminative histogram is displayed in Table 6-4, showing

a clear class separation on the first canonical variable, which represents the direction with the maximum discrimination power, namely the maximum Fisher ratio (the ratio between the interclass and the intraclass variance).

Table 6-4. Prediction matrices and percentages of samples classified correctly in both classification and internal/external validation in the SELECT-LDA performed when addressing the classification approach to discriminate between ALS at initial stage and healthy controls.

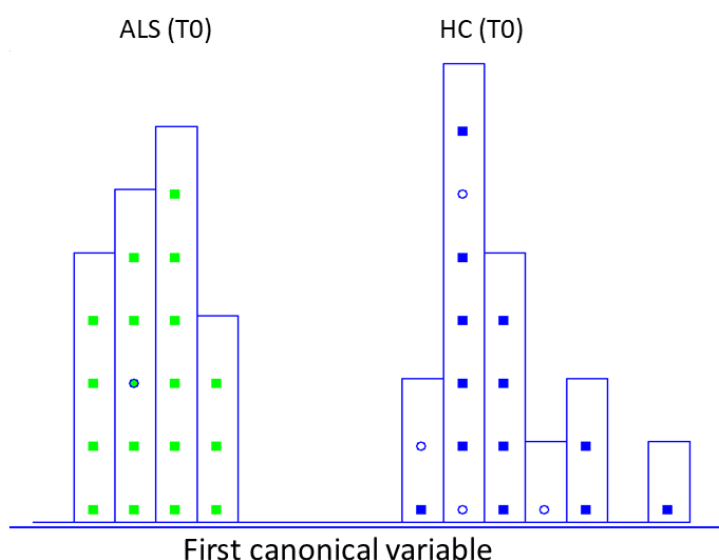| Category						
	Pathology	ALS T0	HC	Sensitivity%	Specificity %	Precision %
Training set	ALS T0	13	0	100	100	100
	HC	0	17			
Test set	ALS T0	4	0	100	100	100
	HC	0	1			

Figure 6-4. Histogram of the first canonical variable for the discrimination of ALS patients at the initial stage (T0) (■) and healthy controls at the same time of blood collection (T0) (■) after performing SELECT-LDA (y-axis indicates the maximum discrimination power between categories). Test samples are displayed as circles (O).

6.1.3.5   ALS T0 versus ALS T6

Based on excellent classification results and perfect group separation obtained utilising three groups of patients, we wanted to test FTIR's ability to distinguish only between

ALS patients in different disease stages. The classification strategy followed the same classification rules as described ahead. Thus, only nine wavenumbers were retained to create this additional classification step, based on applying SELECT as a variable selection method in tandem with LDA. LOO CV and a test subset of five samples selected randomly, three for the ALS (T0) group and two for ALS (T6), were performed to optimise and validate classifications, always on autoscaled data. In this case, optimal accuracy, specificity and sensitivity were achieved, using only nine discriminative spectra bands to remarkably separate between ALS patients in different stage of the disease. The prediction matrix is displayed in Table 6-5. In this sub-classification, a clear group separation between ALS patients before and after six months of disease progression is visually appreciated by the histogram in Figure 5, where the direction with the maximum discrimination power, namely the first canonical variable, displays good within-class separation. This feature proves an outstanding performance of the method and its reliability and good performance of the classification strategy based on reduced spectra signatures.

Table 6-5. Prediction matrices and percentages of samples classified correctly in both classification and internal/external validation in the SELECT-LDA performed when addressing the classification approach for ALS progress disease stratification.

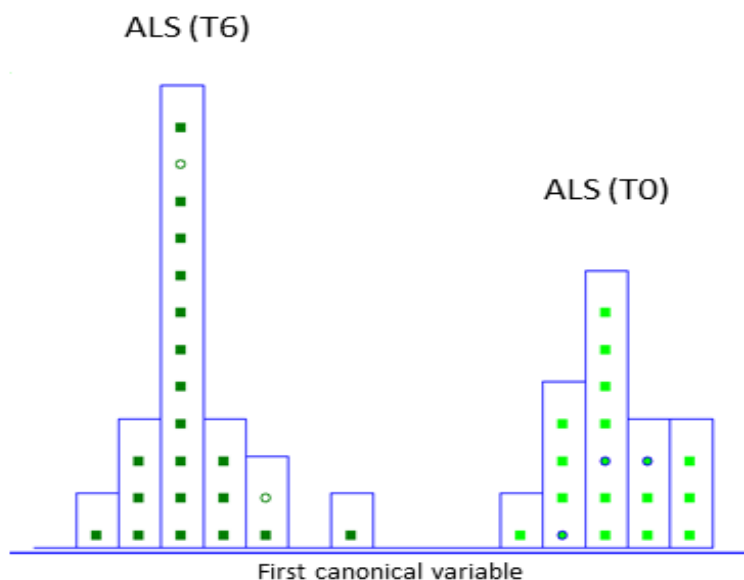| Category						
	Pathology	ALS T0	ALS T6	Sensitivity%	Specificity %	Precision %
Training set	ALS T0	15	0	100	100	100
	ALS T6	0	20			
Test set	ALS T0	3	0	100	100	100
	ALS T6	0	2			

Figure 6.5. Histogram of the first canonical variable for the discrimination of ALS patients at diagnosis time (T0) (■) and after six month of disease progression (T6) (■) after performing SELECT-LDA (y-axis indicates the maximum discrimination power between categories). Test samples are displayed as circles (O).

6.1.3.6 Discriminative spectra and biochemical reasoning

The discriminative spectra biomarkers selected in each classification problem are summarized in Table 6-6. As shown, some overlapping peaks resulted significantly in different types of classification. Therefore, we performed a comparative analysis of the selected peaks that coincided among different classifications. Interestingly, some bands coincided in distinguishing ALS from other pathological conditions, independently of its progression stage, namely spectra signatures at 1355,1356.5 cm^{-1}. Therefore, the selected peaks at 1006, 1007.5, 1011, 1012.5, and 1200.5 cm^{-1} contribute to the discrimination of ALS patients at the initial disease stage (T0). Likewise, spectroscopic signatures at 1054, 1054.5, 1057.5, 1060, 1060.5, 1073.5, 1076 cm^{-1} discriminate specifically subjects in the advanced ALS (T6) stage. In addition, many of the selected peak positions have the same significance and contribution to patients' separation. All relevant bands and their respective biochemical assignments, including the possible identification of bond vibration are reported in Table 6-7

Table 6-6. Discriminant wavenumbers (in order of selection) corresponding to the SELECT-LDA classification developed from FTIR spectra of supernatant blood samples in every classification step.

ALS vs HC vs ON 11 variables	ALS T0 vs ALS T6 VS HC 16 variables	ALS T0 vs ALS T6 vs ON 16 variables	ALS T0 vs ALS T6 9 variables	ALS T0 vs HC T0 10 variables
1321.5	1321.5	1308	1110	1099
1302.5	1060.5	1043	1425.5	1273.5
1060	1355	1011	1308	1465.5
1443	1425.5	1231.5	1049	1011
1182	1345.5	1060	1241	1171.5
1420.5	1159.5	1411	1200.5	1037.5
1073.5	1097	1473.5	1376.5	1007.5
1171	1354.5	1364	1068.5	1200.5
1090	1088	1432	1076	1116.5
1154.5	1032	1049.5		1455
1051.5	1054	1062		
	1259.5	1471		
	1273	1356.5		
	1006	1012.5		
	1057.5	1349		
	1405.5	1170.5		

Table 6-7. Biochemical tentative assignments of the most discriminant wavenumbers selected by SELECT-LDA in each classification step.

Spectral regions in literature	Selected discriminant bands	Attempt band assignment	Biochemical reasoning
~1010-1080	1006, 1007.5, 1011, 1012.5, 1032,1037.5, 1043, 1049, 1049.5, 1054, 1054.5, 1057.5, 1060, 1062, 1065.5, 1068.5, 1076, 1073.5,	$\tilde{v}$ sym. (PO$_2$)	Phospholipids Nucleic acids
	1090, 1097, 1088, 1110, 1099, 1116.5		
~1185-1120	1154.5,1159.5, 1170.5, 1171.5, 1182, 1200.5	$\tilde{v}$ (C-C) and (O-P-O) (C-O) ring vibrations $\tilde{v}$ sym. (CO-O-C)	Nucleic acid ''sugars'' Carbohydrates
~1350-1220	1231.5, 1241, 1259.5, 1273, 1273.5, 1302, 1302.5, 1308, 1321.5, 1345.5, 1349	$\tilde{v}$ (C-C) and (C-O) $\tilde{v}$ (C-N) and (C-(NO_2)) $\tilde{v}$ sym. (PO$_2$) $\tilde{v}$ (C-N) with significant contributions from $\tilde{v}$ (CH$_2$) of carbohydrate residues, δ (CH$_2$)	Amide III-band; Proteins;
~1360-1380	1354.5, 1355, 1356.5, 1364, 1376.5	sym. def. CH$_3$ and sym. def. CH$_2$	Proteins; amino acids Lipids; Phospholipids
~1405-1400	1405.5	$\tilde{v}$ (C=O) of (COO)-group $\tilde{v}$ (C=C)	Fatty acids; Amino acids; (aspartate, glutamate)
	1411, 1420.5, 1425.5, 1432, 1443, 1455, 1465.5		
~1490-1470	1473.5, 1471	δ (CH$_2$)	Lipids

6.1.4 Discussion

FTIR spectroscopy is a promising diagnostic tool that is highly applied to distinguish diseased samples from normal ones, showing high sensitivity, accuracy and specificity results. The main advantage of this technique is that minimum or any sample pre-treatment is required. Thus, it can provide the response rapidly. In addition, it is user-friendly and is suited for analysing different biological samples, such as blood, which are easily collected in clinical settings. Moreover, it should be outlined that sampling blood from patients is less invasive than sampling cerebrospinal fluid (CSF), which is usually collected to perform tests in ALS. Considering the molecular complexities of blood composition, applying the chemometric technique is imperative. Thus, statistical and mathematical algorithms are usually applied to extract chemo-physical information from the generated spectral data. Previous studies performed *in vitro* in mutant-derived S. *cerevisiae* suggested aggregation propensities related to G93A and G85R mutations using FTIR spectroscopy [43]. These mutations are linked to familial cases of ALS, indicating that they could form oligomeric aggregates, probably with molecular intermediates, to finally coalesce into large insoluble aggregates, which could enhance toxicity. However, few studies have shed light on molecular disturbances using FTIR spectroscopy *in vitro* or *in vivo* ALS models [44]. Therefore, this study analysed non-invasive biological samples to characterising better ALS patients concerning healthy controls and neuro-disorders participants. The designed chemometric strategy in this study exhibited optimal classification accuracy (~100%), identifying the most important spectral biomarkers that majorly contribute to patients' variability. Our method based on different classification steps showed excellent discriminative results and optimal prediction accuracy, 100%, for each classification. We can speculate that the two patients in the ALS category (T6) were misclassified as those with recently diagnosed ALS, because their metabolic profile has not yet undergone the changes typically characterizing advanced stage of ALS. Moreover, it should be considered that since these patients received the diagnosis it is more likely, that they started a treatment, thus, in these specific patients, the treatment response could contribute in decelerating the progression of ALS. This possible explanation to justify misclassified ALS patients

actually confirms everything that has been said so far, that the sooner the disease is diagnosed, the sooner the treatment begins, improving the patient's well-being. In addition, another plausible reasoning regards the different ALS genotypes, it could just be that these two subjects belong to distinct disease phenotype, and therefore, based on their genotype they can manifest slower progression of ALS. This speculation could be the basis for further fascinating investigation to discriminate between patients with different genetic background or patient on treatment versus patients without treatment.

On the other hand, the healthy person who has been misclassified as taking part of patients with recently diagnosed ALS, perhaps should already undergo some monitoring. Maybe this misclassification is due to some similarities in patients' metabolic profile, characterising neurodegeneration in the central nervous system. Thus, these findings are also useful to deep in the problem and the collection of exhaustive information completing patient's dataset would confirm FTIR ability to predict possible disease initiation. In fact, the most useful ability among all, to prevent neurodegeneration and start as soon as possible the treatment.

The main aim of this work was to determine the most discriminant wavenumbers based on the SELECT-LDA classification approach in order to: differentiate between different categories of patients, investigate if it is possible to segregate ALS directly at initial stage from healthy controls, and if this method is sensitive to discriminate between ALS progression stage. As we could have noticed, the same wavenumbers appeared significant in different classification steps. Therefore, a comparative analysis of the most significant band extracted in each classification step allowed us to identify those that majorly contribute to ALS discrimination. The last sub-classification performance was highly relevant to compare the discriminative signatures because of the involvement of only two categories; the decision was even more remarkable. We attempted to assign numerous selected wavenumbers in the fingerprint region that were associated with different compounds emerging as biomarkers of ALS. In line with our results, we want to discuss some evidence already found in the literature that reinforces these theories.

Thus, the wavenumbers at 1007.5 cm^{-1} selected to discriminate between ALS T0 and HC T0 seems to have the same contribution and significance as 1006 cm^{-1} chosen in the second classification step for the differentiation of ALS (T0) vs ALS (T6) vs HC (T6). Considering the coexistence of only one group in these two classifications, it is clear that this band contributes to the discrimination of only patients with ALS (T0). Based on the same reasoning and comparing other classification steps, discriminant bands at 1011, and 1076,1200.5 cm^{-1} indicate patients affected by ALS at the early stage of the disease. The discriminant signatures that belong to the infrared region of ~1080-1010 cm^{-1} can be attributed to symmetric stretching (v˜ ass) (P = O) of phospholipids and nucleic acids. The role of phospholipids is of great importance in the CNS. Thus, levels of sphingomyelin and long-chain triglycerides in the CSF of ALS patients have been shown to be correlated with the progression of ALS in the function of the neurodegenerative system[7].

Likewise, spectroscopic signatures at 1054, 1054.5, 1057.5, 1060, 1060.5, 1073.5,1076 and 1425.5 cm^{-1} discriminate specifically subjects in the advanced ALS (T6) stage. For example, among these wavenumbers, the signatures at 1060 cm^{-1} were selected in three classifications, namely to distinguish ALS (T6) vs HC (T6) vs ON, ALS (T0) vs ALS (T6) vs HC (T6) and ALS (T0) vs ALS (T6) vs ON. These bands also belong to the same infrared region and have the same biochemical significance. They are also associated with symmetric stretching (v˜ ass) (P = O) of phospholipids and nucleic acids. Likewise, the band at 1425.5 cm^{-1} used to discriminate between ALS (T0) vs ALS (T6) and between ALS (T0) vs ALS (T6) vs HC (T6), seems to have the same biochemical significance as band at 1420.5 cm^{-1} selected only for the discrimination of ALS (T6) vs HC (T6) vs ON. Thus, they seem to contribute exclusively to discriminate patients after six months of ALS progression.

Interestingly, several selected bands contributed to the discrimination of ALS as a global category, independently of its progression stage. Thus, the variable at 1355 cm^{-1} that contributes to segregation between ALS (T0) vs ALS (T6) vs HC (T6) was also found significant to differentiate between the categories of ALS progression and ON. Since these bands resulted significant in classifications where patients with different ALS advancement were involved, it is deductible that this band is specific for nothing else

but ALS condition. Likewise, discriminant spectral biomarker at 1170.5,1171.5 cm^{-1} was selected as relevant in three classification steps, namely to discriminate three global categories of patients at T6, to separate ALS (T6) from ALS (T0) from healthy controls and to discriminate ALS (T0) from HC (T0). These bands could be attributed to the stretching (CO-O-C) of carbohydrates. Therefore, evidence is emerging that impairment of the glycosphingolipid's metabolism may promote multiple pathological mechanisms, such as dysregulation of the signalling pathway, inflammation, and oxidative stress, correlated with ALS initiation and progression. Moreover, we speculated that bands at 1355 and 1356.5 cm^{-1} due to stretching vibrations of metile and methylene of lipids and phospholipids might have the same chemical and physical significance and contribute as spectral biomarkers in distinguishing ALS disorder. These wavenumbers have the potential to be linked to lipids, such as cholesterol. Thus, it was shown that cholesterol is fundamental for lipid raft formation, glucose transport, and inflammatory signalling, and it regulates cell membrane flexibility through interactions with nearby phospholipids. In addition, from the increasing evidence emerged that cholesterol levels are lower in ALS patients.

Different research groups reported that impairment in lipid homeostasis and oxidative stress impact progression and resolution of neurodegenerative disorders such as ALS [5]. The astrocyte storage of lipids has a protective role in CNS; thus, alterations might lead to inflammation, signalling, oxidative stress and mitochondrial energy generation in neurons. Moreover, in recent findings was shown that high glucose exposure in astrocytes can lead to increased glycogen storage but at the expense of decreased mitochondrial and glycolytic capacity when subsequently metabolically stressed [6]. Thus, it makes sense that many of the remanent spectral signatures were assigned to lipids, carbohydrates and phosphate groups of phospholipids.

The emphasis in this non-targeted metabolomic study was based on searching for disease and stage-specific spectroscopic signatures capable of differentiating ALS from other neuro-disorders conditions with shared symptoms. A tentative band assignment was performed, speculating the individual contribution of each significant infrared variable. Nevertheless, further dedicated and targeted studies could have an additional contribution and significance, elucidating the potential role of each spectra biomarker.

6.1.5 Conclusions

In this study, the power of FTIR spectroscopy coupled with classification techniques in discriminating the spectra of healthy controls and patients with ALS or other neuro-disorders were demonstrated. Different combinations of patient's groups were studied to confirm the discrimination ability of infrared technique, yielding very high precision and specificity towards groups. All the healthy patients were greatly separated, ensuring the power of the built model in preventing the misdiagnosis of non-diseased patients as ones having neuromuscular disorders. In addition, no ALS were confused with patients affected by other neuro-disorders leading the specificity values of 100% for both pathologies. Test set for all of performed classifications showed excellent prediction ability with any misclassification. We also obtained optimal discrimination results between patients of ALS during different progression stage of the disease, that could be due to started treatment as well as the ulterior changes occurred in patients' metabolome. We compared the discriminative feature that overlapped in all classifications. We are convinced that these results may take FTIR one step forward to its application in clinical setting, for preventing misdiagnosis of ALS and help patient well-being. Its utilization is fast, cost-effective and non-invasive, since it is perfectly suitable for biological biofluids. The overall quality of the classification's rules could be improved by the collection of mayor samples and information about patients' phenotypes, treatment and possible comorbidities

6.1.6 References

(1)   Xu, R. S.; Yuan, M. Considerations on the Concept, Definition, and Diagnosis of Amyotrophic Lateral Sclerosis. *Neural Regen Res* **2021**, *16* (9), 1723–1729. https://doi.org/10.4103/1673-5374.306065.

(2)   Area-Gomez, E.; Larrea, D.; Yun, T.; Xu, Y.; Hupf, J.; Zandkarimi, F.; Chan, R. B.; Mitsumoto, H. Lipidomics Study of Plasma from Patients Suggest That ALS and PLS Are Part of a Continuum of Motor Neuron Disorders. *Sci Rep* **2021**, *11* (1), 13562. https://doi.org/10.1038/s41598-021-92112-3.

(3)   Sun, S.; Sun, Y.; Ling, S. C.; Ferraiuolo, L.; McAlonis-Downes, M.; Zou, Y.; Drenner, K.; Wang, Y.; Ditsworth, D.; Tokunaga, S.; Kopelevich, A.; Kaspar, B. K.; Lagier-Tourenne, C.; Cleveland, D. W. Translational Profiling Identifies a Cascade of Damage Initiated in Motor Neurons and Spreading to Glia in Mutant Sod1-Mediated ALS. *Proc Natl Acad Sci U S A* **2015**, *112* (50), E6993–E7002. https://doi.org/10.1073/PNAS.1520639112/-/DCSUPPLEMENTAL.

(4)   Caballero-Hernandez, D.; Toscano, M. G.; Cejudo-Guillen, M.; Garcia-Martin, M. L.; Lopez, S.; Franco, J. M.; Quintana, F. J.; Roodveldt, C.; Pozo, D. The "Omics" of Amyotrophic Lateral Sclerosis. *Trends Mol Med* **2016**, *22* (1), 53–67. https://doi.org/10.1016/j.molmed.2015.11.001.

(5)   Vanherle, S.; Haidar, M.; Irobi, J.; Bogie, J. F. J.; Hendriks, J. J. A. Extracellular Vesicle-Associated Lipids in Central Nervous System Disorders. *Advanced Drug Delivery Reviews*. Elsevier B.V. January 1, 2020, pp 322–331. https://doi.org/10.1016/j.addr.2020.04.011.

(6)   Lee, J. A.; Hall, B.; Allsop, J.; Alqarni, R.; Allen, S. P. Lipid Metabolism in Astrocytic Structure and Function. *Seminars in Cell and Developmental Biology*. Elsevier Ltd April 1, 2021, pp 123–136. https://doi.org/10.1016/j.semcdb.2020.07.017.

(7)   Bouscary, A.; Quessada, C.; René, F.; Spedding, M.; Turner, B. J.; Henriques, A.; Ngo, S. T.; Loeffler, J. P. Sphingolipids Metabolism Alteration in the Central Nervous System: Amyotrophic Lateral Sclerosis (ALS) and Other Neurodegenerative Diseases. *Seminars in Cell and Developmental Biology*. Elsevier Ltd April 1, 2021, pp 82–91. https://doi.org/10.1016/j.semcdb.2020.10.008.

(8)   Leal, A. F.; Suarez, D. A.; Echeverri-Peña, O. Y.; Albarracín, S. L.; Alméciga-Díaz, C. J.; Espejo-Mojica, Á. J. Sphingolipids and Their Role in Health and Disease in the Central Nervous System. *Advances in Biological Regulation*. Elsevier Ltd August 1, 2022. https://doi.org/10.1016/j.jbior.2022.100900.

(9)   Dorst, J.; Kühnlein, P.; Hendrich, C.; Kassubek, J.; Sperfeld, A. D.; Ludolph, A. C. Patients with Elevated Triglyceride and Cholesterol Serum Levels Have a Prolonged Survival in Amyotrophic Lateral Sclerosis. *J Neurol* **2011**, *258* (4), 613–617. https://doi.org/10.1007/S00415-010-5805-Z/TABLES/2.

(10)   Goutman, S. A.; Hardiman, O.; Al-Chalabi, A.; Chió, A.; Savelieff, M. G.; Kiernan, M. C.; Feldman, E. L. Recent Advances in the Diagnosis and Prognosis of Amyotrophic Lateral Sclerosis. *Lancet Neurol* **2022**, *21* (5), 480–493. https://doi.org/10.1016/S1474-4422(21)00465-8.

(11)   Santaella, A.; Kuiperij, H. B.; van Rumund, A.; Esselink, R. A. J.; van Gool, A. J.; Bloem, B. R.; Verbeek, M. M. Inflammation Biomarker Discovery in Parkinson's Disease and Atypical Parkinsonisms. *BMC Neurol* **2020**, *20* (1), 1–8. https://doi.org/10.1186/s12883-020-1608-8.

(12)   Chung, E. J.; Babulal, G. M.; Monsell, S. E.; Cairns, N. J.; Roe, C. M.; Morris, J. C. Clinical Features of Alzheimer Disease with and without Lewy Bodies. *JAMA Neurol* **2015**, *72* (7). https://doi.org/10.1001/jamaneurol.2015.0606.

(13)   Becker, N.; Moore, S. A.; Jones, K. A. The Inflammatory Pathology of Dysferlinopathy Is Distinct from Calpainopathy, Becker Muscular Dystrophy, and Inflammatory Myopathies. *Acta Neuropathol Commun* **2022**, *10* (1), 1–10. https://doi.org/10.1186/s40478-022-01320-z.

(14)   Salari, N.; Fatahi, B.; Valipour, E.; Kazeminia, M.; Fatahian, R.; Kiaei, A.; Shohaimi, S.; Mohammadi, M. Global Prevalence of Duchenne and Becker Muscular Dystrophy: A Systematic Review and Meta-Analysis. *J Orthop Surg Res* **2022**, *17* (1), 1–12. https://doi.org/10.1186/s13018-022-02996-8.

(15)   Ripolone, M.; Velardo, D.; Mondello, S.; Zanotti, S.; Magri, F.; Minuti, E.; Cazzaniga, S.; Fortunato, F.; Ciscato, P.; Tiberio, F.; Sciacco, M.; Moggio, M.; Bettica, P.; Comi, G. P. Muscle Histological Changes in a Large Cohort of Patients Affected with Becker Muscular Dystrophy. *Acta Neuropathol Commun* **2022**, *10* (1), 1–12. https://doi.org/10.1186/s40478-022-01354-3.

(16)   Fujino, H.; Saito, T.; Takahashi, M. P.; Takada, H.; Nakayama, T.; Imura, O.; Matsumura, T. Quality of Life and Subjective Symptom Impact in Japanese Patients with Myotonic Dystrophy Type 1. *BMC Neurol* **2022**, *22* (1), 1–7. https://doi.org/10.1186/s12883-022-02581-w.

(17)   Amyotrophic Lateral Sclerosis (ALS) Pathway. **2020**, 2020.

(18)   Jan, S.; Ahmad, P. *Introducing Metabolomics*; 2019. https://doi.org/10.1016/b978-0-12-814872-3.00001-1.

(19)   Alonso, A.; Marsal, S.; Julià, A. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Frontiers in Bioengineering and Biotechnology*. 2015. https://doi.org/10.3389/fbioe.2015.00023.

(20)   Madsen, R.; Lundstedt, T.; Trygg, J. Chemometrics in Metabolomics-A Review in Human Disease Diagnosis. *Anal Chim Acta* **2010**, *659* (1–2), 23–33. https://doi.org/10.1016/j.aca.2009.11.042.

(21)   Gowda, G. A. N.; Zhang, S.; Gu, H.; Asiago, V.; Shanaiah, N.; Raftery, D. Metabolomics-Based Methods for Early Disease Diagnostics. *Expert Review of Molecular Diagnostics*. September 2008, pp 617–633. https://doi.org/10.1586/14737159.8.5.617.

(22)   Beć, K. B.; Grabska, J.; Huck, C. W. Biomolecular and Bioanalytical Applications of Infrared Spectroscopy – A Review. *Anal Chim Acta* **2020**, *1133*, 150–177. https://doi.org/10.1016/j.aca.2020.04.015.

(23)   Finlayson, D.; Rinaldi, C.; Baker, M. J. Is Infrared Spectroscopy Ready for the Clinic? *Analytical Chemistry*. 2019. https://doi.org/10.1021/acs.analchem.9b02280.

(24)   Perez-Guaita, D.; Garrigues, S.; de la, M. Infrared-Based Quantification of Clinical Parameters. *TrAC - Trends in Analytical Chemistry*. Elsevier B.V. November 1, 2014, pp 93–105. https://doi.org/10.1016/j.trac.2014.06.012.

(25)   Kaznowska, E.; Depciuch, J.; Łach, K.; Kołodziej, M.; Koziorowska, A.; Vongsvivut, J.; Zawlik, I.; Cholewa, M.; Cebulski, J. The Classification of Lung Cancers and Their Degree of Malignancy by FTIR, PCA-LDA Analysis, and a Physics-Based Computational Model. *Talanta* **2018**, *186* (November 2017), 337–345. https://doi.org/10.1016/j.talanta.2018.04.083.

(26)   Lilo, T.; Morais, C. L. M.; Ashton, K. M.; Pardilho, A.; Davis, C.; Dawson, T. P.; Gurusinghe, N.; Martin, F. L. Spectrochemical Differentiation of Meningioma Tumours Based on Attenuated Total Reflection Fourier-Transform Infrared (ATR-FTIR) Spectroscopy. *Anal Bioanal Chem* **2020**, *412* (5), 1077–1086. https://doi.org/10.1007/s00216-019-02332-w.

(27)   Spalding, K.; Bonnier, F.; Bruno, C.; Blasco, H.; Board, R.; Benz-de Bretagne, I.; Byrne, H. J.; Butler, H. J.; Chourpa, I.; Radhakrishnan, P.; Baker, M. J. Enabling Quantification of Protein Concentration in Human Serum Biopsies Using Attenuated Total Reflectance – Fourier Transform Infrared (ATR-FTIR) Spectroscopy. *Vib Spectrosc* **2018**, *99* (August), 50–58. https://doi.org/10.1016/j.vibspec.2018.08.019.

(28)   Wang, X.; Wu, Q.; Li, C.; Zhou, Y.; Xu, F.; Zong, L.; Ge, S. A Study of Parkinson's Disease Patients' Serum Using FTIR Spectroscopy. *Infrared Phys Technol* **2020**, *106* (December 2019), 103279. https://doi.org/10.1016/j.infrared.2020.103279.

(29)   Tomasid, R. C.; Sayat, A. J.; Atienza, A. N.; Danganan, J. L.; Ramos, M. R.; Fellizar, A.; Notarteid, K. I.; Angeles, L. M.; Bangaoilid, R.; Santillan, A.; Albanoid, P. M. Detection of Breast Cancer by ATR-FTIR Spectroscopy Using Artificial Neural Networks. **2022**. https://doi.org/10.1371/journal.pone.0262489.

(30)   Banerjee, A.; Gokhale, A.; Bankar, R.; Palanivel, V.; Salkar, A.; Robinson, H.; Shastri, J. S.; Agrawal, S.; Hartel, G.; Hill, M. M.; Srivastava, S. Rapid Classification of COVID-19 Severity by ATR-FTIR Spectroscopy of Plasma Samples. *Cite This: Anal. Chem* **2021**, *93*, 10391–10396. https://doi.org/10.1021/acs.analchem.1c00596.

(31)   Eikje, N. S. Diabetic Interstitial Glucose in the Skin Tissue by Atr-Ftir Spectroscopy versus Capillary Blood Glucose. *J Innov Opt Health Sci* **2010**, *3* (2), 81–90. https://doi.org/10.1142/S1793545810000903.

(32)   Sitnikova, V. E.; Kotkova, M. A.; Nosenko, T. N.; Kotkova, T. N.; Martynova, D. M.; Uspenskaya, M. v. Breast Cancer Detection by ATR-FTIR Spectroscopy of Blood Serum and Multivariate Data-Analysis. *Talanta* **2020**, *214* (October 2019), 120857. https://doi.org/10.1016/j.talanta.2020.120857.

(33)   el Khoury, Y.; Collongues, N.; de Sèze, J.; Gulsari, V.; Patte-Mensah, C.; Marcou, G.; Varnek, A.; Mensah-Nyagan, A. G.; Hellwig, P. Serum-Based Differentiation between Multiple Sclerosis and Amyotrophic Lateral Sclerosis by Random Forest Classification of FTIR Spectra. *Analyst* **2019**, *144* (15), 4647–4652. https://doi.org/10.1039/c9an00754g.

(34)   Roy, S.; Perez-Guaita, D.; Bowden, S.; Heraud, P.; Wood, B. R. Spectroscopy Goes Viral: Diagnosis of Hepatitis B and C Virus Infection from Human Sera Using ATR-FTIR

Spectroscopy. *Clinical Spectroscopy* **2019**, *1* (January), 100001. https://doi.org/10.1016/j.clispe.2020.100001.

(35) Khanmohammadi, M.; Ghasemi, K.; Garmarudi, A. B. Genetic Algorithm Spectral Feature Selection Coupled with Quadratic Discriminant Analysis for ATR-FTIR Spectrometric Diagnosis of Basal Cell Carcinoma via Blood Sample Analysis. *RSC Adv* **2014**. https://doi.org/10.1039/c4ra04965a.

(36) *preanalytix.com: PAXgene - Specimen Collection & Processing*. https://www.preanalytix.com/ES?cHash=66aa4e4012b1785c4e7ec8d0b38b573f (accessed 2022-11-28).

(37) M. Forina, S. Lanteri, C. Armanino, M.C.C. Oliveros, C. C. V-Parvus 2003. Dip. Chimica e Tecnologie Farmaceutiche ed Alimentari, University of Genova,. http://www.parvus.unige.it.

(38) Forina, M.; Lanteri, S.; Oliveros, M. C. C.; Millan, C. P. Selection of Useful Predictors in Multivariate Calibration. *Anal Bioanal Chem* **2004**, *380* (3 SPEC.ISS.), 397–418. https://doi.org/10.1007/s00216-004-2768-x.

(39) Cocchi, M.; Biancolillo, A.; Marini, F. Chemometric Methods for Classification and Feature Selection. In *Comprehensive Analytical Chemistry*; Elsevier B.V., 2018; Vol. 82, pp 265–299. https://doi.org/10.1016/bs.coac.2018.08.006.

(40) Tkachenko, K.; Espinosa, M.; Esteban-Díez, I.; González-Sáiz, J. M.; Pizarro, C. Extraction of Reduced Infrared Biomarker Signatures for the Stratification of Patients Affected by Parkinson's Disease: An Untargeted Metabolomic Approach. *Chemosensors* **2022**, *10* (6). https://doi.org/10.3390/chemosensors10060229.

(41) Pizarro, C.; Esteban-Díez, I.; Arenzana-Rámila, I.; González-Sáiz, J. M. Discrimination of Patients with Different Serological Evolution of HIV and Co-Infection with HCV Using Metabolic Fingerprinting Based on Fourier Transform Infrared. *J Biophotonics* **2018**, *11* (3), 1–12. https://doi.org/10.1002/jbio.201700035.

(42) Pizarro, C.; Esteban-Díez, I.; Espinosa, M.; Rodríguez-Royo, F.; González-Sáiz, J. M. An NMR-Based Lipidomic Approach to Identify Parkinson's Disease-Stage Specific Lipoprotein-Lipid Signatures in Plasma. *Analyst* **2019**, *144* (4), 1334–1344. https://doi.org/10.1039/c8an01778f.

(43) Oztug Durer, Z. A.; Cohlberg, J. A.; Dinh, P.; Padua, S.; Ehrenclou, K.; Downes, S.; Tan, J. K.; Nakano, Y.; Bowman, C. J.; Hoskins, J. L.; Kwon, C.; Mason, A. Z.; Rodriguez, J. A.; Doucette, P. A.; Shaw, B. F.; Valentine, J. S. Loss of Metal Ions, Disulfide Reduction and Mutations Related to Familial ALS Promote Formation of Amyloid-Like Aggregates from Superoxide Dismutase. *PLoS One* **2009**, *4* (3), e5004. https://doi.org/10.1371/JOURNAL.PONE.0005004.

(44) Ami, D.; Duse, A.; Mereghetti, P.; Cozza, F.; Ambrosio, F.; Ponzini, E.; Grandori, R.; Lunetta, C.; Tavazzi, S.; Pezzoli, F.; Natalello, A. Tear-Based Vibrational Spectroscopy Applied to Amyotrophic Lateral Sclerosis. *Anal Chem* **2021**, *93* (51), 16995–17002. https://doi.org/10.1021/acs.analchem.1c02546.

6.2 SPECTROCHEMICAL DIFFERENTIATION OF ALS ONSET AND PROGRESSION BASED ON ATR-FTIR SPECTROSCOPY: PRELIMINARY STUDY

In this section, preliminary studies based on ATR-FTIR using the same ALS patient cohort as described above is reported. Considering the simple, fast and cost-effective advantages of ATR-FTIR instrumentation. Herein, ATR-FTIR's capability to extract relevant spectral features associated with ALS onset and progression and classify them in a predictive fashion was investigated.

6.2.1 Data Analysis

After data acquisition, the processing and computational analysis of raw metabolic data was performed using Unscrambler (version X 11.0, Camo ASA, Oslo, Norway), V-Parvus (version PARVUS2011, Michele Forina, Genoa, Italy). Based on previous knowledge and experience about higher spectral regions where noise, water absorption and saturation are usually verified. only the mid-IR spectrum was analysed: the biochemical "fingerprint region" at 1500–1050 cm^{-1}. In this way, the analytical time was ulteriorly reduced. Given the high dimensionality of biological spectral data, many disturbing factors, such as random noise, baseline distortions, or light scattering, influence spectral data acquisition. Thus, the pre-processing step is imperative in analysis to reduce these factors. To compensate for instrumental artefacts and sample sample-to-completions, different pre-processing methods were evaluated individually or in combination to minimise the adulterant-unrelated variability. Thus, the normalisation, moving average and Savitzky–Golay (S–G) second derivatives were applied to ATR spectra. Therefore, better resolution of overlapping peaks and decreased scatter effects were ensured.

The entire data set was split into two independent subsets to develop and validate the classifications proposed: a training set and a test set. The test set was different for all methods applied, and classifications were developed, considering that different subsets of patients were studied. As a result, the smoothed and normalised output tables were always centred before additional multivariate analysis and classification algorithms.

6.2.2.1   LDA to discriminate between ALS patients at disease onset and healthy controls

After careful pre-processing, ATR-FTIR measurements were submitted for further multivariate analysis. In this preliminary analysis, two pairwise comparisons were studied, namely, the ALS disease progression, including ALS (T0) and ALS (T6) (a total of 43 objects) and a prognostic comparison, including ALS (T0) and healthy controls (a total of 35 objects), respectively.

As good data analysis practice, unsupervised PCA was performed to search for possible clusters and outliers. Once all the outliers were removed, a stepwise decorrelation procedure, SELECT, was performed, avoiding redundant information in the subset of selected significant predictors (spectral variables). In addition, it has previously demonstrated its accurate prediction ability in selecting the most important variable for the discrimination of pathological status. Thus, SELECT was applied to extract the most significant wavenumbers from the IR dataset, providing input features for further classification and class modelling.

LDA on ten spectra variables that were decorrelated from other variables by SELECT, built by leave one out (LOO) cross-validation, was performed to evaluate the feasibility of this classification methodology to differentiate between the onset of ALS and healthy controls. Excellent discrimination among categories was achieved, providing a 100% level of correctly classified samples for diseased subjects and controls. In addition, excellent external prediction performances of 100% were achieved using a total of five patients in the test set distributed randomly for both categories (within no misclassified patients), respectively (Table **6-8 Table 6-8.** Results of LDA classification performance on 10 selected variables**Table 4-2**). Furthermore, a clear interclass separation achieved between these main categories can also be visually appreciated in the corresponding discriminative histogram (Figure 6.6). Nevertheless, the excellent separation on the first canonical variable, displaying the control group, also indicates it slight heterogeneity.

Table 6-8. Results of LDA classification performance on 10 selected variables to distinguish between ALS at disease onset and controls.

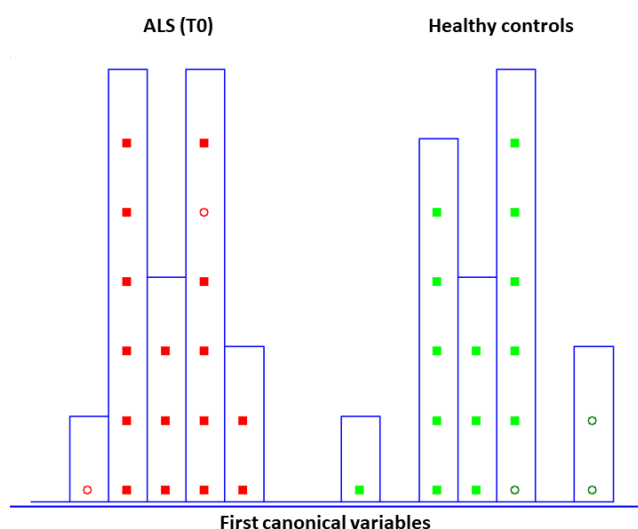| Group	Classification (%)	Prediction (%)	External Prediction (%)	Total Rate (%)
ALS (T0)	100	100	100	100
Controls	100	100	100	100
Total rate	100	100	100	100

Figure 6.6. Histogram of the first canonical variable for the discrimination of ALS (T0) (■) and healthy controls (■) patients (y-axis indicates the maximum discrimination power between categories).

6.2.2.2   LDA to discriminate between ALS disease progression

Likewise, LDA on the IR dataset to distinguish the ALS progression stage, built by 10 cancellation groups for CV was also performed. Before LDA analysis, as explained above, SELECT was applied to extract those predictor variables correlated with the discrimination between categories here considered. Therefore, based on the SELECT rules, 12 selected spectra variables were decorrelated from other wavenumbers and used for LDA classification. The 12 selected spectra biomarkers showed an outstanding classification performance of 100%, and the results were slightly lower in prediction ability, 97.37%, respectively, compared to 100% obtained in the previous classification. Thus, one object from the ALS (T6) category was classified as ALS (T0). The results of the SELEC-LDA performance are displayed in Table 6-9. The classification strategy's

suitability for reduced IR plasma signatures can be visually appreciated in Figure 6.7. A discriminative histogram shows a clear group separation on the first canonical variable.

Table 6-9. Results of LDA classification performance on 12 selected variables

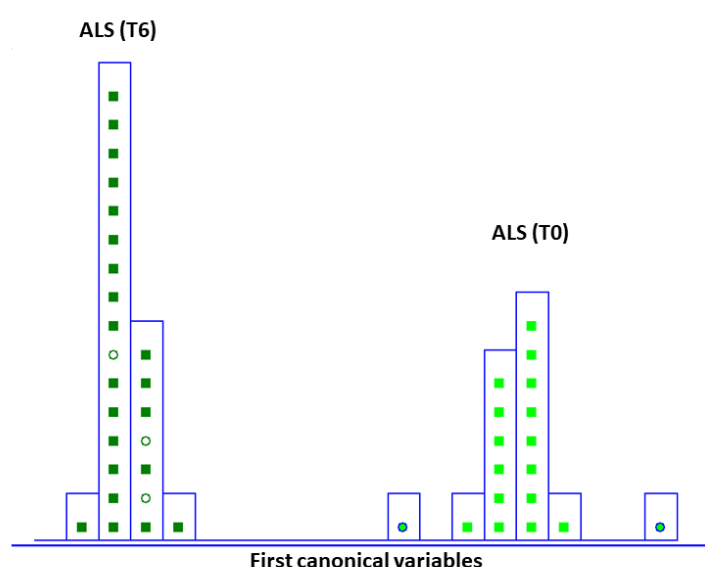| Group	Classification (%)	Prediction (%)	External Prediction (%)	Total Rate (%)
ALS (T0)	100	100	100	100
ALS (T6)	100	95.45	100	100
Total rate	100	97.37	100	99.1

Figure 6.7. Histogram of the first canonical variable for the discrimination of ALS (T0) (■) and ALS at 6 months follow-up patients (■) (y-axis indicates the maximum discrimination power between categories).

6.2.2.3 SIMCA class modelling to discriminate patients' groups

In an attempt to go one step further in this classification strategy, it was decided to build optimised class models based on clinical parameters and the subset of reduced IR signatures selected by SELECT. SIMCA often outperforms other classification methods, where a new sample will always be classified in one of the predefined categories. Classification methods such as LDA are based on the development of classification rules and delimiters between classes, whereas in class models, significance limits are built for the specified classes (SIMCA class modelling was previously described in Chapter 2.4.2.2.2 Supervised techniques).

Herein, SIMCA modelling was performed on previously selected spectral variables, 10 for the classification problem ALS (T0) vs HC, and 12 to discriminate ALS (T0) from ALS (T6) Table 6-10. A class modelling of 10 variables achieved satisfactory results in both internal prediction (LOO) and external prediction 80.00%. Considering this is a preliminary study, the external prediction showed lower accuracy due to fewer external predictors. Thus, only two objects for ALS (T0) and 3 for the control group were used to test the prediction ability, whereas one of the control groups was misclassified as having the disease. These results should be cautiously retested utilising a more significant number of samples, but still, yet, the results are auspicious.

Likewise, to model ALS (T0) vs ALS (T6) the same number of external predictors were used and the equally low external prediction was achieved, due to a low number of objects used as a test set. SIMCA builds a mathematical model of the category with its principal components. The specific category accepts a sample if its distance to the model is not significantly different from the class residual standard deviation. The results of SIMCA modelling for ALS (T0) and healthy controls can be visually appreciated by a Cooman's Plot Figure 6.8, representing the samples' distances against each of the two models. The Cooman's plots were built considering a 95% confidence level to define the class space and the unweighted augmented distance. This diagram is an effective visual representation that directly indicates the quality of the model constructed with the magnitude of the distance between categories. Thus, the distances to the principal component models and SIMCA approximation in a two-class problem for the class of diseased ALS (T0) at onset and ALS (T6) at 6-month follow-up are plotted in Figure 6.9.
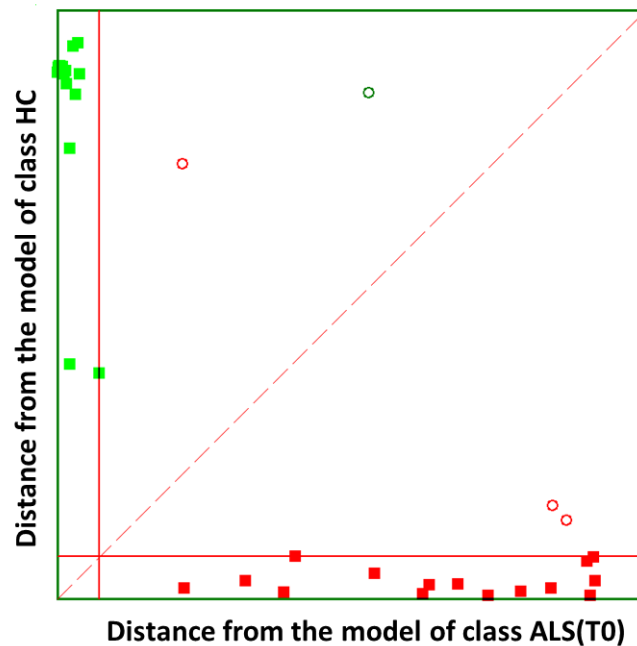
Figure 6.8. Cooman's plot displaying the results obtained by applying SIMCA class-modelling to 10 IR spectral variables: ALS (T0) (■) and HC (■) patients within the included (○) external test set. The red solid line indicates a confidence level for class space at 95%. The red dashed line indicates equal class distance.

As can be observed in the plot, some objects of the external set are outside the model. Nevertheless, Cooman's plot shows high interclass specificity and a patently clear separation between class models. The healthy controls appear evidently segregated and concentrated, forming a dense cluster at large distances from the model of the diseased ALS patient's class. Likewise, it is observable that ALS (T0) patients are forming an apparent clustering. Still, most samples fall clearly and univocally into their class region, far from the class limit for the HC model. Furthermore, as said above, the single ALS (T0) sample located in the inconclusive classification region is virtually placed above the membership threshold, pertinent to the other class.
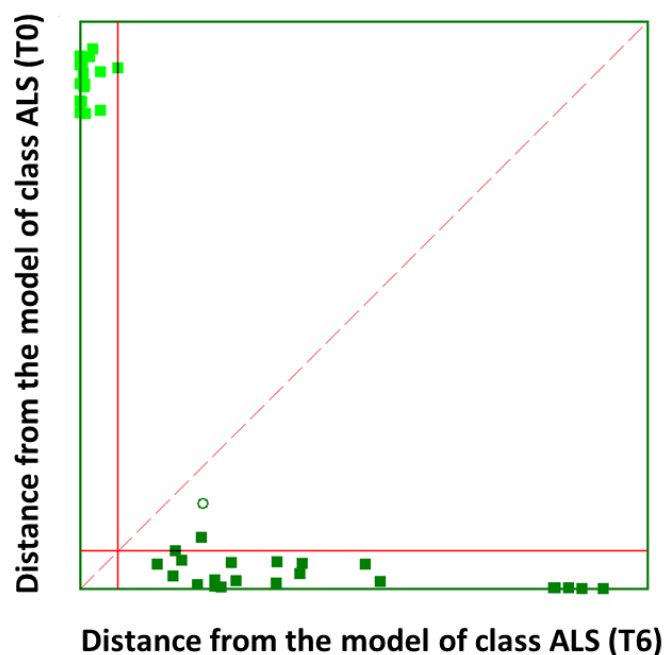
Figure 6.9. Cooman's plot displaying the results obtained by applying SIMCA class-modelling to 10 IR spectral variables: ALS (T0) (■) and ALS T6 (■) patients within the included (○) external test set. The red solid line indicates a confidence level for class space at 95%. The red dashed line indicates equal class distance.

In a two-class problem for the class of ALS (T0) and ALS (T6), one sample that falls into the joint space of both categories belonging mainly to the ALS (T6) category, was observed. Compared to the o previous problem class model, the distribution of some samples from the ALS (T6) class outside the model could indicate that their metabolic profile may be much less marked than others, confounding the decision about their location inside the model. Nevertheless, their location is still very close to the confidence interval of 95%. Moreover, this class modelling showed much higher discriminant power performance but lower modelling power than the previous one. (Table 6-10) As can be seen in the plot, the ALS (T0) class ALS (T0) are perfectly segregated in the upper left corner, very far from the ALS (T6) class.

Table 6-10. The results of SIMCA class-modelling performance of two differentiation problems

Differentiation problem	Classification (%)	External prediction (%)	Efficiency (%)	Efficiency forced model (%)	Total Rate (%)
ALS (T0) vs HC *10 variables*	100	80	100	100	95.00
ALS(T0) vs ALS(T6) *12 variables*	100	80	97.33	100	94.33

The values of modelling and discrimination power of each IR-selected variable for specific class modelling problem were calculated Table 6-11 and Table 6-12. Thus, the MP describes how well a variable helps each principal component model variation in the data, and discriminatory power (DP) describes how well a variable helps each principal component model classify samples in a training set. Thus, the selected variable of both class modelling showed very high values of discriminant power, indicating the contribution of each one in class separation.

Of note, both classification problems showed the apparent clustering of patient categories. Meanwhile, it was observed that some samples fell into the opposite class region and slightly outside their class model. Nevertheless, it should be outlined that any sample category was located in the area of relative indecision (small left quadrant), indicating explicit separation of categories based on the diseased status and disease progression.

Table 6-11. Discriminative and modelling powers of ten selected spectra variables obtained in SIMCA class modelling of healthy controls and patients at ALS onset

Spectra variable	Discriminant power	Modelling power	
		Category HC	Category ALS(T0)
1185	5.46		
1335.5	5.38		
1180	5.36		
1355	5.35		
1310.5	5.34	1.00	1.00
1234	5.34		
1400.5	5.32		
1475	5.32		
1033.5	5.31		
1449.5	5.27		

Table 6-12. Discriminative and modelling powers of twelve selected spectra variables obtained in SIMCA class modelling to compare ALS disease progression (ALS (T0) vs ALS(T6))

Spectra variable	Discriminant power	Modelling power	
		Category ALS (T0)	Category ALS(T6)
1478	13.74	0.74	0.95
1372	53.60	0.32	0.82
1308.5	54.37	0.85	0.91
1492	58.44	0.77	0.74
1344.5	63.87	0.65	0.51
1361	45.25	0.90	0.60
1380	50.41	0.92	0.91
1105.5	59.86	0.87	0.58
1381.5	44.15	0.76	0.56
1104	38.24	0.92	0.95
1304	45.87	0.85	0.65
1193.5	18.41	0.90	0.91

6.2.3 Conclusion

Our principal goal was to test the ATR-FTIR coupled with already known and widely used chemometrics strategy's ability to obtain optimal segregation between patients without additional clinical, physical, or ethnic data, and this goal was achieved. The perfect discrimination between diseased ALS and healthy patients was obtained based only on metabolic fingerprinting using a few spectral variables. In addition, it showed high sensitivity and specificity in distinguishing patients at different stages of the

disease. These preliminary results showed that ATR-FTIR measurement shares the same benefits as the standard FTIR, leading to reliable sample segregation with high internal and external prediction accuracy. Nevertheless, this method outperforms the FTIR cuvette measurements since it does not require cuvette handling, increasing its suitability for determining routine parameters in clinical reality. The ''shotgun'' analysis on small blood drops that provide an answer in real-time about your disease status, is the type of instrument that should be implemented in a clinical environment. Further studies utilising a larger sample size will take ATR-FTIR one step forward as a diagnostic /screening tool for ALS elucidation.

6.3 UPLC-QTOF-MS BASED LIPIDOMIC BLOOD PROFILING REVEALS BIOMARKER OF AMYOTROPHIC DISEASE PROGRESSION AND ITS DIFFERENTIATION FROM ANOTHER RELATED MOTOR NEURO DISEASE

6.3.1 Introduction

As was introduced in the previous section when performing studies on FTIR, the clinical profile and the pathogenetic mechanism of ALS patients have been the subject of considerable research and investigation; however, reliable and readily accessible molecular biomarkers, such as those obtained from patients' biofluids, are still lacking to provide accurate ALS diagnosis. In addition, ALS is often confounded with other neuro diseases (ON) with a similar clinical profile. For example, in patients with Parkinson's or Alzheimer's, Myotonic and Becker (BMD) muscular dystrophies are also evaluated by progressive muscular weakness and cognitive impairments [1].

Numerous scientific findings claim that only a small per cent of ALS cases are caused by genetic factors (familial or sporadic gene mutations); thus, the remanent cases have unknown causes and could be characterised by heterogeneous factors such as gene/environment interactions or dysregulated metabolic conditions (e.g., aberrant protein aggregation, oxidative stress, altered lipid and RNA metabolism). Given this perspective, novel ALS biomarkers may arise from studying the different metabolic pathways.

Lipids exert multiple functions, such as membrane fluidity and structure, molecular signalling, and mediation of inflammatory responses, and also contribute to nervous system maintenance[2,3]. Therefore, alteration in any of these processes directly affects the lipidome of the cell, tissue and biofluids surrounding them [4–6]. The central nervous system (CNS) is characterised by the presence of a wide variety of lipids, and lipidomic alterations, especially to molecular ageing or in response to increased reactive oxygen species (ROS), could contribute to the onset of neurodegeneration [7]. Thus, lipid dysregulation in CNS and circulation in ALS patients may be clinically associated with disease severity[8,9]. A great deal of evidence has shown that ALS patients report dysfunctional lipid pathways, reinforcing lipid role in ALS pathogenesis[10]. Therefore, patients with elevated blood levels of triacylglycerides (TAGs) and low/high-density

lipoprotein (LDL/HDL) have a better prognosis by nearly a year [11]. Likewise, elevated arachidonic acid and polyunsaturated fatty acids (PUFA) could contribute to motor neuron dysfunction and death in patients. Increased fatty acids (FA) oxidation and decreased subcutaneous fat stores are other metabolic features correlated with ALS even prior to motor degeneration. Recently, the causal mutation in *SPTLC1,* a gene responsible for sphingolipid biosynthesis, has also been associated with ALS pathogenesis. This gene encodes for the activity of serine palmitoyltransferase (SPT), the multisubunit enzyme that catalyses the initial and rate-limiting step in sphingolipid biosynthesis requires close homeostatic regulation to prevent cell toxicity due to the excess sphingolipids. Thus, SPT alterations are linked to neurodegeneration [12]. Moreover, the beneficial effect of a high-fat diet was severally confirmed by various research groups delaying disease onset and extending life expectancy in ALS patients [13]. In addition, the accumulation of ceramides, arachidonic acid, and lysophosphatidylcholine (lysoPC) seems very important to motor neurons [14–16].

All these findings suggest that many efforts have been made to understand the disease mechanism and discover biomarkers for ALS diagnosis. Furthermore, these findings evidence that alteration in lipid metabolism may drive ALS pathogenesis and thus could serve as a biomarker target. Lipidomics is a sub-discipline of metabolomics, and in the last few years, it has been widely used to detect lipid alteration in biological systems [17–23].

Herein, we performed UPLC untargeted lipidomic approach to evaluate potential lipid biomarkers for diagnosing ALS. Given the non-invasive nature, blood samples were used to reveal circulating lipid biomarkers. Differential lipid metabolites were identified comparing healthy controls and ALS patients, ALS and other motor-related neuro disorders and ALS progression stage. Univariate and multivariate analyses were performed to validate the identified biomarkers to assess their diagnostic performance of ALS disease.

6.3.2   Material and method

6.3.2.1   Study population and sample collection

The sample cohort described previously (Chapter 6.1.2, Experimental design) was included to perform UPLC-MS analysis. Thus, 76 blood samples were included from Niguarda Ca'Granda Hospital in Milan (Italy). The participants were matched for age and gender whenever possible.

A total of 35 ALS patients, 34 healthy controls and seven patients with other neuropathies (ON) were included. The ALS group was divided into two subgroups; 19 patients were collected at first diagnosis of ALS (T0), of which 16 were obtained as samples after six months of diagnosis/treatment ALS (T6). The ALS group included familial (fALS) cases due to quadruple mutation in the ALS susceptibility genes SOD1/TDP43/FUS/c9orf72. Similarly, considering the availability of the samples at T6 in the same participants, the healthy control group was divided into 19 participants at T0, of which 15 were obtained as samples at T6. Regarding the ON group, only samples of 7 participants were at our disposal. The ON participants were affected by: Becker's Muscular Dystrophy, Extrapyramidal syndrome, Facioscapulohumeral Muscular Dystrophy and Myotonic Dystrophy.

6.3.2.2   Lipidomic analysis

The chemicals used to perform this study are the same described in (Chapter 3.1, Chemicals and reagents). Thus, the lipidomic analysis is based on a previously described methodology, performing the MTBE-US-assisted lipid extraction method (Chapter 3.2.2, MTBE-US-assisted lipid extraction method).

6.3.2.3   UPLC-Q/TOF Analysis

LC analysis was performed on a Waters ACQUITY I-Class UPLC system (Waters Corp, Milford, USA) equipped with a Waters Acquity HSS T3 100 × 2.1 (i.d.) mm 1.8 μm particle size column and a Waters VanGuard precolumn of the same material, using the same condition as described in (Chapter 3.3.1.2.2 Liquid Chromatography-Mass Spectrometry).

6.3.2.4 Data processing and statistical analysis

Progenesis QI software (Waters Corporation, Milford, MA, USA) processed the acquired mass data for peak detection, alignment and normalisation. The generated data with the information of retention time, accurate molecular mass, and MS^E data were transferred into EZinfo 2.0 software (Waters Corporation, Milford, MA, USA). Considering a bigger number of variables than the sample size, dimensionality reduction was imperative. The unsupervised principal component analysis (PCA) was performed for the initial data overview. QC samples were used to monitor the analytical performance of the UPLC and the system's stability. Thus, the high degree of aggregation of the QC samples in the PCA model was an instrumental stability and reproducibility index. Different samples were detected as outliers and excluded from further analysis.

6.3.2.5 Feature annotation

To select the most reliable identity of the lipid when more than one option was available, the following criteria were followed: the most probable ionisation adduct based on the composition of the mobile phase, the minor m/z error of the exact mass (5 ppm maximum tolerance), highest fragmentation score and most probable isotopic distribution.

6.3.2.6 Data analysis

The generated matrices were subsequently analysed using MetaboAnalyst 5.0 (a comprehensive platform for metabolomics analysis that allows the application of univariate and multivariate methods, enrichment analysis and pathway analysis.

Then, supervised PLS-DA and OPLS-DA analyses, the most used methods for binary classification, were performed on the data set to achieve the maximum separation between the following groups considered in this study: ALS T0 vs ALS T6, ALS vs CO, ALS vs ON, HC vs ON Considering that the sample size was relatively small; thus the results could be due to overfitting, the models were validated by Q2Y (predictive variation) and R2Y (explained variation) parameters based on 10-fold cross-validation and leave one out cross validation (LOOCV). The variable importance projection (VIP ≥ 1.0) from the peaks intensity, fold change (FC) analysis by comparing the mean intensity of lipid metabolites, and false discovery rate (FDR) corrected *p-value* was applied to screen the

discriminant biomarkers. However, to reject the hypothesis that the results obtained by PLS-DA and OPLS-DA were obtained by chance or by overfitting, a random forest (RF) machine learning approach (Number of trees =5000) was performed in an attempt to find non-linear patterns in lipid specimens that can explain variation between groups.

6.3.3   Results

To investigate the involvement of lipids in amyotrophic lateral sclerosis, a global lipidome analysis of ALS at the onset and ALS at six months follow-up, other motor neuro disorder cases and their respective age and gender-matched control group were performed. Therefore, the lipidome analyses of blood supernatant samples revealed different lipid species when performing pairwise comparisons. Many of identified discriminant features belonging to known metabolites of commonly used drug treatments, or diets were not considered significant biomarkers. By comparing different models in pairwise classifications, many features still need to be identified. Nevertheless, all the classification performed by PLS-DA and OPLS-SA were validated and provided significant separation between groups in positive and negative ionisation mode analyses.

. The lipidome enrichment analysis revealed the majority of identified lipids belonging mainly to fatty acyls, glycolipid and glycerophospholipid families (Figure 6.10)
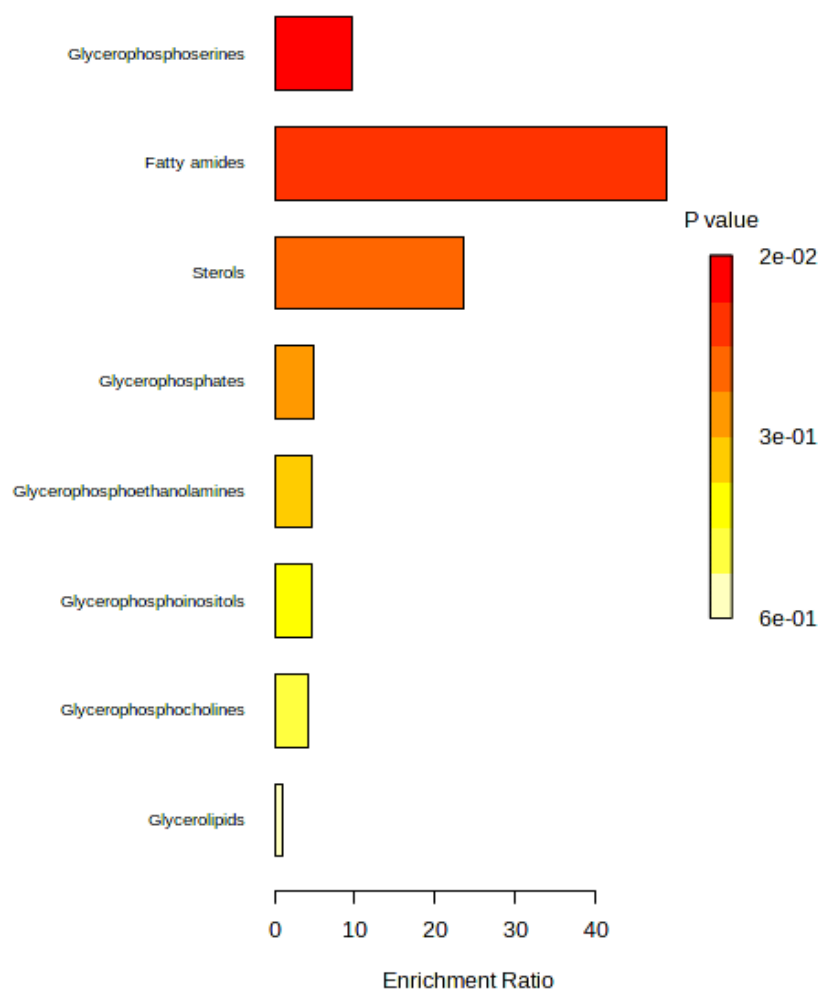
Figure 6.10. Main differential classes of lipids that contribute to patient differentiation in various pairwise comparisons.

Different groups combination was studied and tested by PLS-DA, OPLS-DA and RF models in positive and negative ionization modes. Thus, a total of 12 pairwise comparisons were performed. Therefore, in Figure 6.11 we selected a subset of lipid metabolites that contributed the most to the classification between groups.
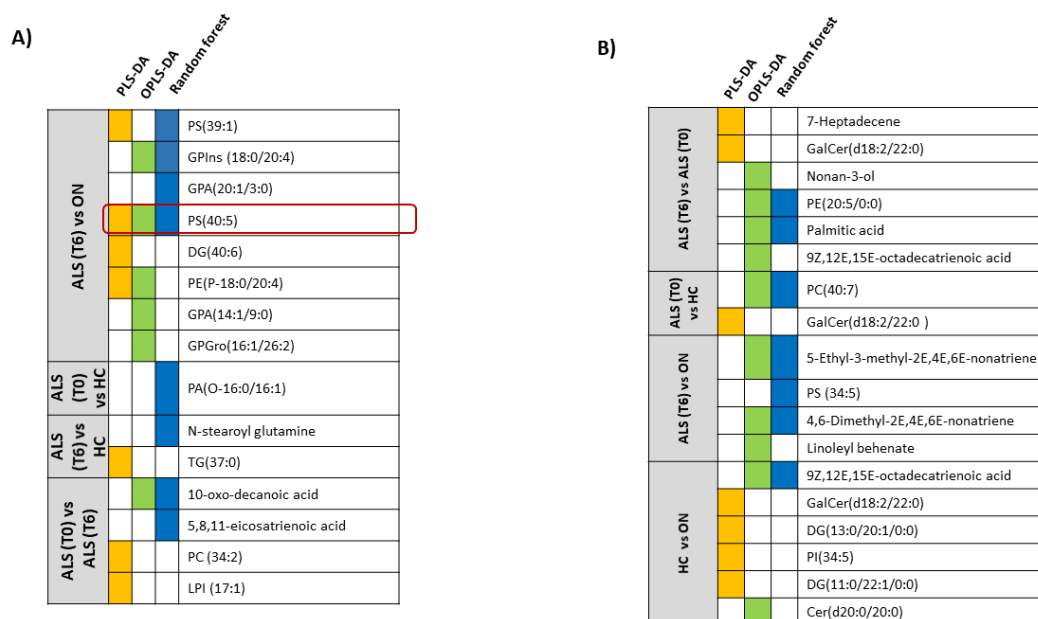
Figure 6.11. The lipid features were selected and prioritized by different statistical models. Each line of the table represents a lipid species colored by its selection in specific models, thus: yellow for selection by PLS-DA, green by OPLS-DA and blue by RF. A) results relative to molecular findings detected in negative ionisation mode B) in positive ionisation mode.

There can be several reasons for incongruences in compound selection using PLS-DA, OPLS-DA, and RF algorithms. One possible reason is that the three models use different criteria for selecting important variables, and this can lead to differences in the selected variables. For example, PLS-DA and OPLS-DA are linear regression-based models considering the covariance between the variables and the response. Still, OPLS-DA has an additional step of orthogonalisation that removes any variation in the X matrix that is not correlated with the response. On the other hand, random forest is a tree-based model that uses a different criterion to evaluate variable importance, such as the mean decrease in a mean decrease accuracy.

Moreover, the sample size and variability within the samples can also affect the selection of variables by the models. Small sample sizes at our disposal led to instability in the model selection and resulted in different variables being selected by different models.

To address incongruences in the compound selection, it is essential to thoroughly examine and compare the results from all three models and identify the variables

consistently selected by multiple models. As we can observe, PS (40:5) is the unique biomarker selected by all three models to discriminate between ALS at six months of follow-up and ON cases. Interestingly, another phosphoserine derivate PS (39:1), was confirmed by two models (PLS-DA and RF) to discriminate between ALS(T6) and ON in negative ionisation mode. Of note, the contribution of PS (34:5) was evaluated for the same group separation but in a positive mode. It seems evident the importance of PS metabolism to discriminate between ALS cases and other related motor disorders.

Multivariate analysis evidenced the contribution of several individual molecules belonging to the class of glycerophospholipids in patients' differentiation: PE (20:5/0:0), PC (34:2) and LPI (17:1) seem to contribute to distinguishing ALS progression; likewise, PE(P-18:0/20:4), GPIns(18:0/20:4) and GPA(20:1/3:0) contribute for the segregation between ON and ALS (T6). Of note, the untargeted analyses revealed PA(O-16:0/16:1) and PC (40:7) as prognostic lipid specimens of ALS onset.

As shown in Figure 6.11, the discrimination between groups is also linked to alterations in the levels of different fatty acyls and their conjugates, such as sphingolipids and diglycerols (DGs).

6.3.3.1 Diagnostic model

Additionally, we performed a validation study by ROC (AUC) curves to confirm the validity of the selected variables by RF and, at least by another model, to exclude the impact of overfitting. It is also essential to consider the biological relevance of the selected variables and their potential involvement in the disease or condition being studied. Therefore, AUC curves for each discriminant lipid species from each set of pairwise classifications in positive and negative ionisation modes were performed., As summarised in Table 6-13 and Table 6-14, there were four metabolic biomarkers for distinguishing ALS disease severity, three prognostic lipid species for ALS onset, six putative biomarkers for discriminating between ALS at six months follow-up and other motor neuro disorders and one putative prognostic biomarker of ON.

As can be observed, all four putative lipid species to discriminate between ALS (T0) and ALS (T6) showed AUC ≥ 0.9, and equally high sensitivity and specificity, near 100%, respectively. Thus, ALS (T6) patients showed higher levels of palmitic and 10-oxo-decanoic-acids. Meanwhile, lower levels of 5,8,11-eicosatrienoic acid and PE (20:5/0:0).

The AUC values of lipid metabolites for distinguishing ALS (T6) and ON displayed results higher than 0.743. In addition, PS (40:5) showed the best AUC=0.876 to predict the status of motor neuro disease (ALS versus ON). Besides PS (40:5), the most prominent changes in PS (36:1) and GPIns(18:0/20:4) were all increased in ON patients when compared to ALS (T6). Interestingly, the lipids belonging to fatty esters class 4,6-Dimethyl-2,4,6-nonatriene and 5-Ethyl-3-methyl-2,4,6-nonatriene were found decreased in ON patients compared to ALS cases.

The phosphatidic acid PA(O-16:0/16:1), which are decreased in ALS patients at early stage compared to controls, showed the best prediction performance (AUC=0.771) among three putative metabolites. The 9,12,15-octadecatrienoic acid, which are significantly decreased in ON patients compared to controls showed not only the most prominent changes (log2FC=2.7863, p=$1.5928E^{-15}$), but also the best AUC=1 performance.

Table 6-13. The information on the selected biomarker panels by two statistical models in negative ionization mode.

Lipid specimen	AUC	T-test	Log2 FC
ALS (T0) vs ALS (T6)			
10-oxo-decanoic-acid	0.977	3.7801E-16	-1.219
5,8,11-eicosatrienoic acid	0.991	4.0801E-10	1.235
ALS (T6) vs ON			
GPIns(18:0/20:4)	0.8	1.5967E-4	-0.540
PS(36:1)	0.743	5.8594E-4	-0.267
GPA(14:1/9:0)	0.848	4.2943E-5	0.433
PS(40:5)	0.876	2.5815E-6	-0.432
ALS (T0) vs HC			
PA(O-16:0/16:1)	0.771	5.3499E-6	-1.138
ALS (T6) vs HC			
N-stearoyl glutamine	0.722	0.015152	0.138

Table 6-14. The information on the selected biomarker panels by two statistical models in positive ionization mode.

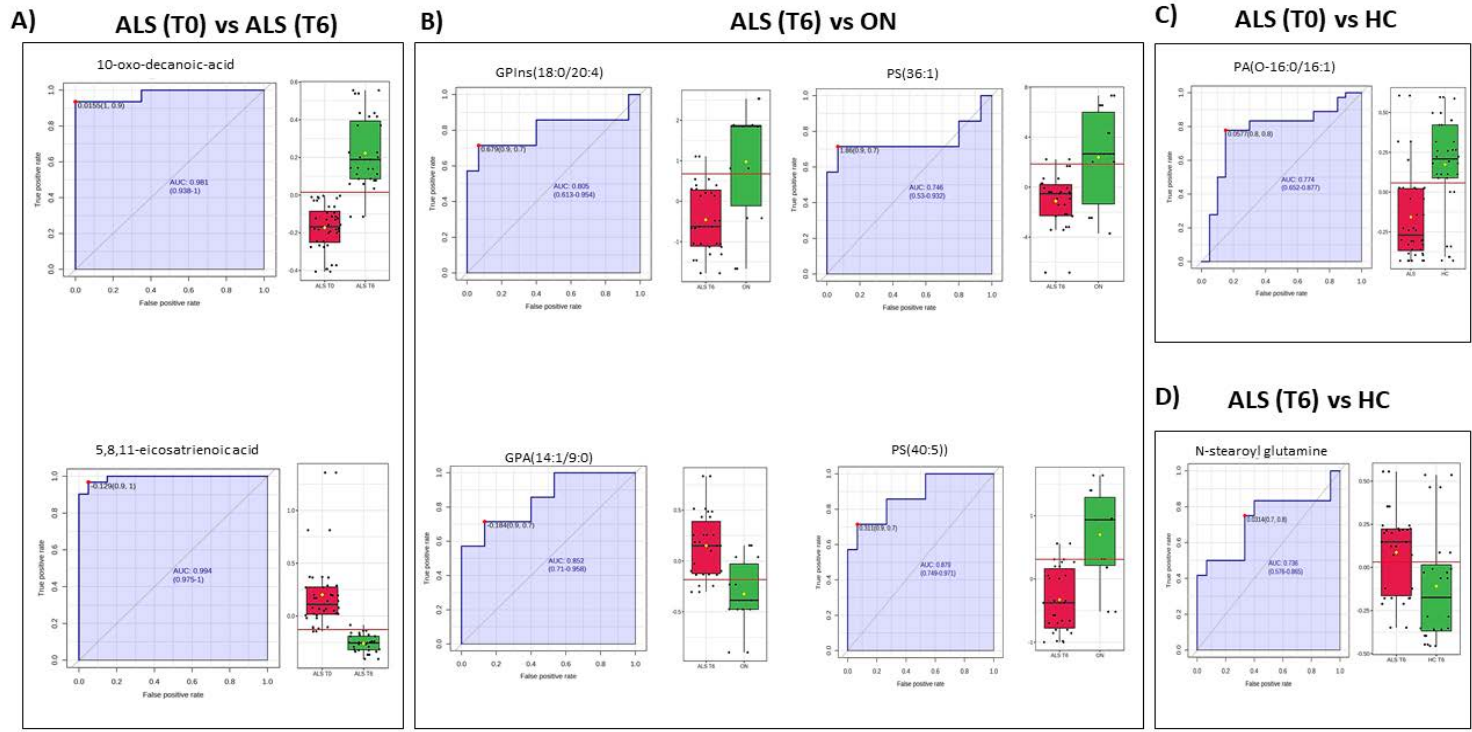| Lipid specimen	AUC	T-test	Log2 FC
ALS (T0) vs ALS (T6)			
Palmitic acid	0.915	7.4305E-10	-0.319
PE (20:5/0:0)	0.936	5.0447E-9	1.634
ALS (T6) vs ON			
4,6-Dimethyl-2,4,6-nonatriene	0.8	0.0032111	0.310
5-Ethyl-3-methyl-2,4,6-nonatriene	0.847	9.2229E-4	0.536
ALS (T0) vs HC			
PC (40:7)	0.711	2.1023E-4	-0.429
HC vs ON			
9,12,15-octadecatrienoic acid	1.0	1.5928E-15	2.786

.

Figure 6.12. Diagnostic performance and prediction of the most discriminant identified biomarkers. The AUC curves and respective box plots for comparison between A) ALS at the onset and ALS after a 6-month follow-up; B) ALS after a 6-month follow-up, and other motor neuro disorders. C) ALS at onset and healthy controls; D) advanced ALS and healthy controls

227

Figure 6.13. Diagnostic performance and prediction of the most discriminant identified biomarkers. The AUC curves and respective box plots for comparison between A) ALS at the onset and ALS after a 6-month follow-up; B) ALS after a 6-month follow-up, and other motor neuro disorders. C) ALS at onset and healthy controls; D) advanced ALS and healthy control.
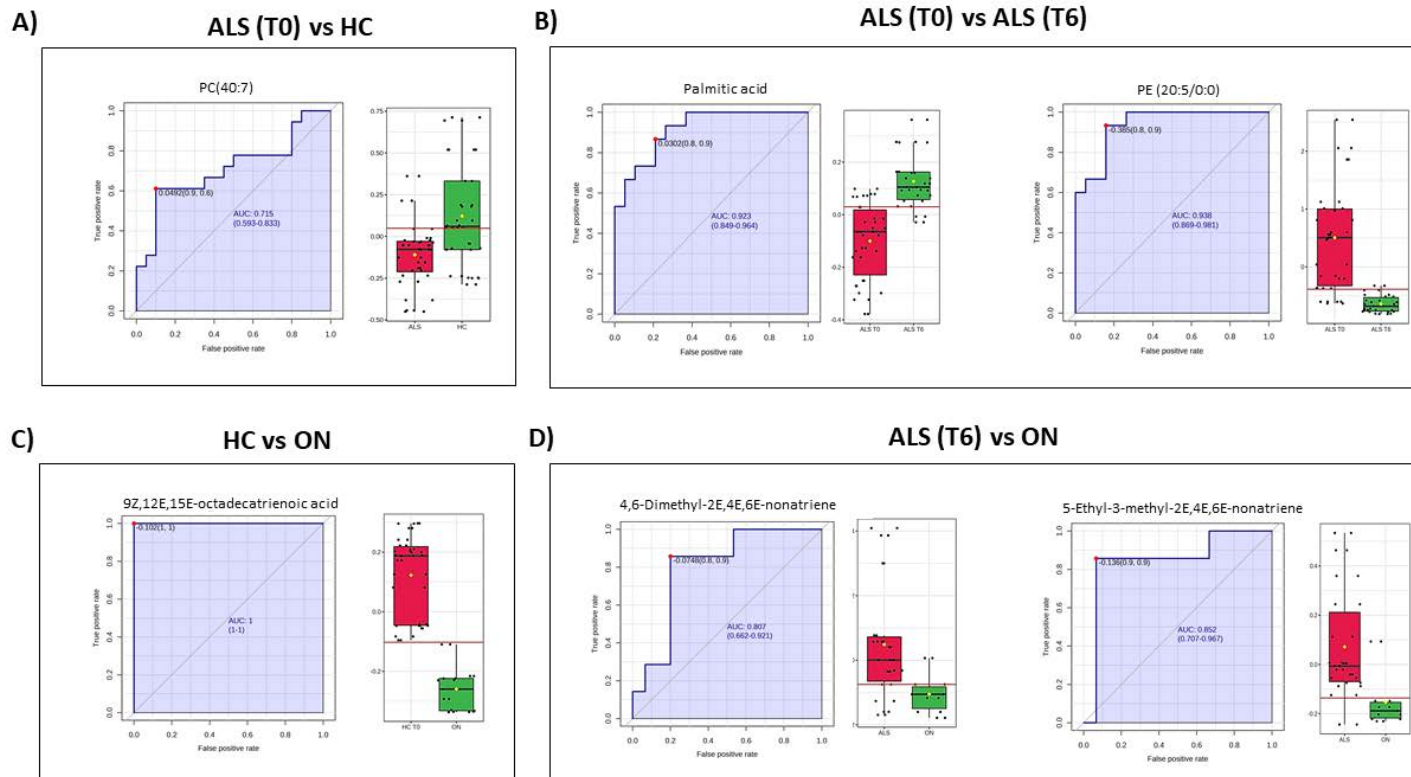
6.3.4 Discussion

In this study, we broadly searched lipid blood biomarkers in ALS and other neuro disorders, performing a comprehensive lipidomic approach. Lipids are the most abundant metabolites in peripheral blood and are usually reported to be easily accessible indicators of metabolic dysfunction.

To our knowledge, this is the first lipidomic study of the supernatant blood of patients with ALS, which aimed to identify lipid profiles for diagnosing and predicting disease evolution. In addition, ALS lipid profiles were compared to other neuro disorders, and we tried to evaluate predictive biomarkers that can explain its underlying pathologic mechanisms. Several statistical models were used to evaluate the most discriminant biomarkers of ALS patients and controls and patients affected by other confounding neurodegenerative diseases. Multiple pairwise comparisons between patient groups have led to optimal discrimination results. Furthermore, all the models led to optimal group separation. However, many metabolites possibly contributing to patients' segregation were unidentified, strengthening that ALS metabolic alterations are broad and poorly defined.

The present work's findings revealed that the principal lipid specimens (fatty acyls glycerolipids and glycerophospholipids) contribute to the greatest extent of ALS differentiation. Therefore, almost all of them are congruent with most of the findings of the previous ALS metabolomics studies.

It is widely proposed that hypermetabolism, defined as an abnormally elevated level of resting energy expenditure, negatively impacts the course of ALS disease and persists over time [24–26]. This situation often leads to reduced fat depots, and this alteration may be associated with increased TGs in plasma. Therefore, abnormalities in skeletal muscle energy metabolism are accompanied by a defect in energy homeostasis exhibited as decreased glycolysis, increased fatty acid beta-oxidation in skeletal muscles, and decreased subcutaneous fat stores [27–29]. Since hypermetabolism is a prognostic factor for ALS, it is perfectly plausible to find altered fatty acyls, fatty esters and fatty acids lipid specimens involved in fatty acid biosynthesis, fatty acid degradation/elongation or fat metabolism, which could be due to augmented energy expenditure in ALS progression. Therefore, the tentatively assigned discriminant lipids distinguishing ALS progression (ALS (T0) vs ALS (T6)), such as 7-Heptadecene, Palmitic

acid, C18:3n-3,6,9, 10-oxo-decanoic acid showed higher levels in cases of 6 months follow up ALS. Altogether, adjustments in lipid metabolism probably respond to the high energy demands of the disease and modulate ALS phenotypes [30]. In addition, Henriques et al.[31] discussed how the energetic metabolism influences the degree of unsaturation and length of fatty acids in lipids. ALS disease has been associated with changes in lipid metabolism, including unsaturated fatty acids (UFAs). Studies have shown that ALS patients have altered levels of UFAs, including increased arachidonic acid (AA) and decreased docosahexaenoic acid (DHA) [32].

Moreover, polyunsaturated fatty acids (PUFAs) are essential in inflammation and oxidative stress. Thus, it was shown that AA levels are increased in the spinal cord and brain tissues of patients with ALS [33,34]. This increase in AA levels may contribute to the development and progression of ALS, as AA can produce inflammatory mediators such as prostaglandins and leukotrienes, which can lead to oxidative stress and neuroinflammation. Thus, one of the most consistent changes achieved through different statistical models is the PUFA (5,8,11-eicosatrienoic acid) metabolite which arose as a suitable biomarker candidate (AUC=0.994) to discriminate the ALS progression stage. Our findings confirm the hypothesis that the changes in the proportions of fatty acids in blood lipids are a source of novel, reliable markers to help monitor disease severity in ALS patients. PLS-DA model, in ALS(T6) versus control comparison, highlighted that blood triglyceride (TG 12:0/12:0/13:0) showed diminished values in samples from ALS (T6), suggesting possible increased consumption of this lipid or its decreased production. Nevertheless, our results disagree with increased levels of specific TG species in plasma found by Sol et al. Further studies are needed to elucidate the role of TG in ALS diagnosis fully.

Many glycerophospholipids, major membrane constituents, and ceramides specimens are differential between ALS and non-ALS cases. Concerning ALS differentiation from healthy controls, PC (40:7) was selected by two statistical models to distinguish ALS (T0) and CO groups. The decreased PC (40:7) is not in accordance with previously reported results, where the levels of PC specimens were found to be increased in CSF or plasma samples of ALS patients. Nevertheless, other lipid specimens concerning ALS progression, PE (25:0/0:0), as well as previously cited 5,8,11-eicosatrienoic acid, also showed decreased levels after six months of disease

progression [35]. Considering that synaptic activity is compromised in ALS patients, PC synthesis via CDP-choline, which requires choline, a pyrimidine and PUFAs as the precursors for optimal stimulation pathway, could justify the decreased levels of these lipid species. In addition, data from many sources indicate that low PE can occur with age and is often linked as a factor to other neurodegenerative diseases such as Parkinson's disease [36].

Interestingly, increased levels of ceramide derivates were found to characterise ALS patients and ON. Several lines of evidence implicate various sphingolipids in neuronal signalling and toxicity[2,37]. Since SM is involved in signal transduction and regulating inflammatory processes, such as response to oxidative stress [38], the derivates of SM hydrolysis-ceramides (Cer) are known to be involved in neurodegenerative processes [34] [14]. Therefore, the perturbation in SM/Cer homeostasis might contribute to neurodegeneration. Our findings are consistent with the literature evidence confirming the relationship between sphingomyelins and cognitive impairment. Our data revealed that some sphingolipids, particularly galactosyl ceramides GalCer (d18:2/22:0), significantly increased during the ALS progression. In addition, the same ceramide species were selected to discriminate between ALS(T0) and CO, whereas expected showed increased levels in the diseased group. Thus, strengthening the fact that ceramide accumulation characterises the ALS onset and its progression. Interestingly, the same ceramide species resulted in an increase in ON cases compared to the control group. These results confirm previous findings indicating the involvement of ceramides and neuro disease development and highlight the potential use of SM and ceramide as biomarkers in ALS [39–41]. Nevertheless, only one model confirmed these results and should be interpreted cautiously.

One of our principal goals was to discriminate between ALS patients and other confounding neuro disorders. Since clinical evaluation lack of sensitivity leads to misdiagnosis and tardive treatment, herein, we performed different pairwise comparisons to evaluate biomarkers for the group of patients carrying other neurodegenerative phenotypes. Interestingly, among identified compounds, phosphoserine (PS (40:5)) was discriminant by comparing ALS (T6) group and ON patients and was validated by all three statistical models. Moreover, other PS (39:1) metabolite was also identified by comparing ALS after six months of disease progression

and ON and validated by two statistical models. In addition, by comparing the same groups in positive ionisation mode, another PS specimen (PS (34:5) was statistically significant. Thus, patients belonging to ON phenotypes in all three classifications displayed increased PS biomarkers levels compared to the ALS (T6) group. It is well known that PS, like other phospholipids, have different important functions, such as contributing to membrane fluidity and signalling in cell membranes (cit Agrawal). Clinical trials suggest that PS may have applications for the prevention and therapy of cognitive disorders [43]s.[44]. PS decarboxylation leads to phosphatidylethanolamine and subsequently to phosphatidylcholine. In the case of pathological conditions, PS signalling may be dysregulated. Since PS is a precursor for PC and PE, it seems reasonable that altered PS metabolism leads to decreased PC and PE in ALS patients, which were obtained in our results.

We acknowledge that our study presents different limitations. Thus, most of the lipid specimens that resulted significantly in these comparisons were unidentified, suggesting that many poorly explored metabolites still could explain ALS mechanisms. These results should be interpreted cautiously; even so, all the statistical models were validated. Only one of the identified lipids was discriminant in all three models, namely PS (40:5), which distinguished between ALS at six months of disease progression and patients with other neuro disorders. Nevertheless, biologically meaningful lipid species were identified, and their contribution to patients' differentiation is justifiable and confirmed by previous studies. Additionally, our study is longitudinal, thus, we were able to access temporal changes in a metabolic profile highlighting earlier versus later alterations in disease developments. Moreover, our study includes samples from patients with other neurological disorders, which allowed additional comparisons and identification of the discriminant metabolites that could help avoid future misdiagnosis.

6.3.5 Conclusions

To objectively track ALS disease progression and its onset, reliable and readily accessible biomarkers, such as those from patient blood, are highly needed. This untargeted metabolomic approach found evidence of previously established and identified lipid species and novel emerging metabolites. Due to small cohorts of patients,

no conclusions could be formally drawn about the potential prognostic effects of the identified biomarkers. Nevertheless, our findings suggest the contribution of alterations in UFAs, PUFAs, FAs and glycerolipid metabolism to the pathogenesis of ALS and other neuro disorders. Here we found that the changes in the proportions of fatty acids from blood lipids are a source of novel, reliable markers to help monitor ALS disease severity. We hypothesised that PS metabolites could be a trait of other neuro disorders.

These results could be a promising stand point for new targeted studies. The reported blood lipid profiles of ALS need to be completed, since no associations with clinical variables have been controlled; therefore, the further investigation is needed.

6.3.6 References

1.   Salari, N.; Fatahi, B.; Valipour, E.; Kazeminia, M.; Fatahian, R.; Kiaei, A.; Shohaimi, S.; Mohammadi, M. Global Prevalence of Duchenne and Becker Muscular Dystrophy: A Systematic Review and Meta-Analysis. *J Orthop Surg Res* **2022**, *17*, 1–12, doi:10.1186/s13018-022-02996-8.

2.   Ngo, S.T. Lipids: Key Players in Central Nervous System Cell Physiology and Pathology. *Semin Cell Dev Biol* 2021, *112*, 59–60.

3.   Schmitt, F.; Hussain, G.; Dupuis, L.; Loeffler, J.P.; Henriques, A. A Plural Role for Lipids in Motor Neuron Diseases: Energy, Signaling and Structure. *Front Cell Neurosci* 2014, *8*.

4.   Caballero-Hernandez, D.; Toscano, M.G.; Cejudo-Guillen, M.; Garcia-Martin, M.L.; Lopez, S.; Franco, J.M.; Quintana, F.J.; Roodveldt, C.; Pozo, D. The "Omics" of Amyotrophic Lateral Sclerosis. *Trends Mol Med* **2016**, *22*, 53–67, doi:10.1016/j.molmed.2015.11.001.

5.   Amyotrophic Lateral Sclerosis (ALS) Pathway. **2020**, 2020.

6.   FernÁndez-Eulate, G.; Ruiz-Sanz, J.I.; Riancho, J.; Zufiría, M.; GereÑu, G.; FernÁndez-TorrÓn, R.; Poza-Aldea, J.J.; Ondaro, J.; Espinal, J.B.; GonzÁlez-ChinchÓn, G.; et al. A Comprehensive Serum Lipidome Profiling of Amyotrophic Lateral Sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener* **2020**, *21*, 252–262, doi:10.1080/21678421.2020.1730904.

7.   Chaves-Filho, A.B.; Pinto, I.F.D.; Dantas, L.S.; Xavier, A.M.; Inague, A.; Faria, R.L.; Medeiros, M.H.G.; Glezer, I.; Yoshinaga, M.Y.; Miyamoto, S. Alterations in Lipid Metabolism of Spinal Cord Linked to Amyotrophic Lateral Sclerosis. *Sci Rep* **2019**, *9*, 1–14, doi:10.1038/s41598-019-48059-7.

8.   Dupuis, L.; Corcia, P.; Fergani, A.; Gonzalez De Aguilar, J.-L; Bonnefont-Rousselot, D.; Bittar, R.; Seilhean, P.D.; Hauw, J.-J.; Lacomblez, L.; Loeffler, J.-P.; et al. *Dyslipidemia Is a Protective Factor in Amyotrophic Lateral Sclerosis*; 2008;

9.   Wuolikainen, A.; Moritz, T.; Marklund, S.L.; Antti, H.; Andersen, P.M. Disease-Related Changes in the Cerebrospinal Fluid Metabolome in Amyotrophic Lateral Sclerosis Detected by GC/TOFMS. *PLoS One* **2011**, *6*, doi:10.1371/journal.pone.0017947.

10.   Agrawal, I.; Lim, Y.S.; Ng, S.Y.; Ling, S.C. Deciphering Lipid Dysregulation in ALS: From Mechanisms to Translational Medicine. *Transl Neurodegener* 2022, *11*.

11.   Dorst, J.; Kühnlein, P.; Hendrich, C.; Kassubek, J.; Sperfeld, A.D.; Ludolph, A.C. Patients with Elevated Triglyceride and Cholesterol Serum Levels Have a Prolonged Survival in Amyotrophic Lateral Sclerosis. *J Neurol* **2011**, *258*, 613–617, doi:10.1007/s00415-010-5805-z.

12.   Mohassel, P.; Donkervoort, S.; Lone, M.A.; Nalls, M.; Gable, K.; Gupta, S.D.; Foley, A.R.; Hu, Y.; Saute, J.A.M.; Moreira, A.L.; et al. Childhood Amyotrophic Lateral Sclerosis Caused by Excess Sphingolipid Synthesis. *Nat Med* **2021**, *27*, 1197–1204, doi:10.1038/s41591-021-01346-1.

13.   Dorst, J.; Schuster, J.; Dreyhaupt, J.; Witzel, S.; Weishaupt, J.H.; Kassubek, J.; Weiland, U.; Petri, S.; Meyer, T.; Grehl, T.; et al. Effect of High-Caloric Nutrition on Serum Neurofilament Light Chain Levels in Amyotrophic Lateral Sclerosis. *J Neurol Neurosurg Psychiatry* 2020, *91*, 1007–1009.

14.   Dodge, J.C.; Treleaven, C.M.; Pacheco, J.; Cooper, S.; Bao, C.; Abraham, M.; Cromwell, M.; Sardi, S.P.; Chuang, W.L.; Sidman, R.L.; et al. Glycosphingolipids Are Modulators of Disease Pathogenesis in Amyotrophic Lateral Sclerosis. *Proc Natl Acad Sci U S A* **2015**, *112*, 8100–8105, doi:10.1073/pnas.1508767112.

15.   Henriques, A.; Croixmarie, V.; Priestman, D.A.; Rosenbohm, A.; Dirrig-Grosch, S.; D'Ambra, E.; Huebecker, M.; Hussain, G.; Boursier-Neyret, C.; Echaniz-Laguna, A.; et al. Amyotrophic Lateral Sclerosis and Denervation Alter Sphingolipids and Up-Regulate Glucosylceramide Synthase. *Hum Mol Genet* **2015**, *24*, 7390–7405, doi:10.1093/hmg/ddv439.

16.   Cutler, R.G.; Pedersen, W.A.; Camandola, S.; Rothstein, J.D.; Mattson, M.P. Evidence That Accumulation of Ceramides and Cholesterol Esters Mediates Oxidative Stress - Induced Death of Motor Neurons in Amyotrophic Lateral Sclerosis. *Ann Neurol* **2002**, *52*, 448–457, doi:10.1002/ana.10312.

17.   Sethi, S.; Brietzke, E. Recent Advances in Lipidomics: Analytical and Clinical Perspectives. *Prostaglandins Other Lipid Mediat* **2017**, *128–129*, 8–16, doi:10.1016/j.prostaglandins.2016.12.002.

18.   Rampler, E.; Abiead, Y. el; Schoeny, H.; Rusz, M.; Hildebrand, F.; Fitz, V.; Koellensperger, G. Recurrent Topics in Mass Spectrometry-Based Metabolomics and Lipidomics—Standardization, Coverage, and Throughput. *Anal Chem* **2021**, *93*, 519, doi:10.1021/ACS.ANALCHEM.0C04698.

19.   Hu, T.; Zhang, J.L. Mass-Spectrometry-Based Lipidomics. *J Sep Sci* 2018, *41*, 351–372.

20.   Răchieriu, C.; Eniu, D.T.; Moiş, E.; Graur, F.; Socaciu, C.; Socaciu, M.A.; Hajjar, N. al Lipidomic Signatures for Colorectal Cancer Diagnosis and Progression Using Uplc-qtof-esi+ms. *Biomolecules* **2021**, *11*, 1–17, doi:10.3390/biom11030417.

21.   Massey, K.A.; Nicolaou, A. Lipidomics of Oxidized Polyunsaturated Fatty Acids. *Free Radic Biol Med* **2013**, *59*, 45–55, doi:10.1016/j.freeradbiomed.2012.08.565.

22.   Wu, D.; Shu, T.; Yang, X.; Song, J.-X.; Zhang, M.; Yao, C.; Liu, W.; Huang, M.; Yu, Y.; Yang, Q.; et al. Plasma Metabolomic and Lipidomic Alterations Associated with COVID-19. *Natl Sci Rev* **2020**, doi:10.1093/nsr/nwaa086.

23.   Peña-Bautista, C.; Álvarez-Sánchez, L.; Roca, M.; García-Vallés, L.; Baquero, M.; Cháfer-Pericás, C. Plasma Lipidomics Approach in Early and Specific Alzheimer's Disease Diagnosis. *J Clin Med* **2022**, *11*, doi:10.3390/jcm11175030.

24.   Ferri, A.; Coccurello, R. What Is "Hyper" in the ALS Hypermetabolism? *Mediators Inflamm* **2017**, *2017*, doi:10.1155/2017/7821672.

25.   Ahmed, R.M.; Dupuis, L.; Kiernan, M.C. Paradox of Amyotrophic Lateral Sclerosis and Energy Metabolism. *J Neurol Neurosurg Psychiatry* **2018**, *89*, 1013–1014, doi:10.1136/jnnp-2018-318428.

26.   Jésus, P.; Fayemendy, P.; Nicol, M.; Lautrette, G.; Sourisseau, H.; Preux, P.M.; Desport, J.C.; Marin, B.; Couratier, P. Hypermetabolism Is a Deleterious Prognostic Factor in Patients with Amyotrophic Lateral Sclerosis. *Eur J Neurol* **2018**, *25*, 97–104, doi:10.1111/ene.13468.

27.   Anakor, E.; Duddy, W.J.; Duguez, S. The Cellular and Molecular Signature of ALS in Muscle. *J Pers Med* **2022**, *12*, doi:10.3390/jpm12111868.

28.   Bouteloup, C.; Desport, J.C.; Clavelou, P.; Guy, N.; Derumeaux-Burel, H.; Ferrier, A.; Couratier, P. Hypermetabolism in ALS Patients: An Early and Persistent Phenomenon. *J Neurol* **2009**, *256*, 1236–1242, doi:10.1007/s00415-009-5100-z.

29.   Desport, J.C.; Torny, F.; Lacoste, M.; Preux, P.M.; Couratier, P. Hypermetabolism in ALS: Correlations with Clinical and Paraclinical Parameters. *Neurodegener Dis* **2006**, *2*, 202–207, doi:10.1159/000089626.

30.   FernÁndez-Eulate, G.; Ruiz-Sanz, J.I.; Riancho, J.; Zufiría, M.; GereÑu, G.; FernÁndez-TorrÓn, R.; Poza-Aldea, J.J.; Ondaro, J.; Espinal, J.B.; GonzÁlez-ChinchÓn, G.; et al. A Comprehensive Serum Lipidome Profiling of Amyotrophic Lateral Sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener* **2020**, *21*, 252–262, doi:10.1080/21678421.2020.1730904.

31.   Henriques, A.; Blasco, H.; Fleury, M.C.; Corcia, P.; Echaniz-Laguna, A.; Robelin, L.; Rudolf, G.; Lequeu, T.; Bergaentzle, M.; Gachet, C.; et al. Blood Cell Palmitoleate-Palmitate Ratio Is an Independent Prognostic Factor for Amyotrophic Lateral Sclerosis. *PLoS One* **2015**, *10*, doi:10.1371/JOURNAL.PONE.0131512.

32.   Bjornevik, K.; Zhang, Z.; O'Reilly, É.J.; Berry, J.D.; Clish, C.B.; Deik, A.; Jeanfavre, S.; Kato, I.; Kelly, R.S.; Kolonel, L.N.; et al. Prediagnostic Plasma Metabolomics and the Risk of Amyotrophic Lateral Sclerosis. *Neurology* **2019**, *92*, E2089–E2100, doi:10.1212/WNL.0000000000007401.

33.   Fitzgerald, K.C.; O'Reilly, É.J.; Falcone, G.J.; McCullough, M.L.; Park, Y.; Kolonel, L.N.; Ascherio, A. Dietary ω-3 Polyunsaturated Fatty Acid Intake and Risk for Amyotrophic Lateral Sclerosis. *JAMA Neurol* **2014**, *71*, 1102–1110, doi:10.1001/jamaneurol.2014.1214.

34.   Veldink, J.H.; Kalmijn, S.; Groeneveld, G.J.; Wunderink, W.; Koster, A.; de Vries, J.H.M.; van der Luyt, J.; Wokke, J.H.J.; van den Berg, L.H. Intake of Polyunsaturated Fatty Acids and Vitamin E Reduces the Risk of Developing Amyotrophic Lateral Sclerosis. *J Neurol Neurosurg Psychiatry* **2007**, *78*, 367–371, doi:10.1136/JNNP.2005.083378.

35.   Blasco, H.; Veyrat-Durebex, C.; Bocca, C.; Patin, F.; Vourc'H, P.; Kouassi Nzoughet, J.; Lenaers, G.; Andres, C.R.; Simard, G.; Corcia, P.; et al. Lipidomics Reveals Cerebrospinal-Fluid Signatures of ALS. *Sci Rep* **2017**, *7*, doi:10.1038/s41598-017-17389-9.

36.   Patel, D.; Witt, S.N. Ethanolamine and Phosphatidylethanolamine: Partners in Health and Disease. *Oxid Med Cell Longev* 2017, *2017*.

37.   Fahy, E.; Subramaniam, S.; Brown, H.A.; Glass, C.K.; Merrill, A.H.; Murphy, R.C.; Raetz, C.R.H.; Russell, D.W.; Seyama, Y.; Shaw, W.; et al. A Comprehensive Classification System for Lipids. *J Lipid Res* **2005**, *46*, 839–861, doi:10.1194/jlr.E400004-JLR200.

38.   Schmitt, F.; Hussain, G.; Dupuis, L.; Loeffler, J.P.; Henriques, A. A Plural Role for Lipids in Motor Neuron Diseases: Energy, Signaling and Structure. *Front Cell Neurosci* 2014, *8*.

39.   Morrow, A.; Panyard, D.J.; Deming, Y.K.; Jonaitis, E.; Dong, R.; Vasiljevic, E.; Betthauser, T.J.; Kollmorgen, G.; Suridjan, I.; Bayfield, A.; et al. Cerebrospinal Fluid Sphingomyelins in Alzheimer's Disease, Neurodegeneration, and Neuroinflammation. *J Alzheimers Dis* **2022**, *90*, 667, doi:10.3233/JAD-220349.

40.   Chaves-Filho, A.B.; Pinto, I.F.D.; Dantas, L.S.; Xavier, A.M.; Inague, A.; Faria, R.L.; Medeiros, M.H.G.; Glezer, I.; Yoshinaga, M.Y.; Miyamoto, S. Alterations in Lipid Metabolism of Spinal Cord Linked to Amyotrophic Lateral Sclerosis. *Sci Rep* **2019**, *9*, doi:10.1038/s41598-019-48059-7.

41.   Leal, A.F.; Suarez, D.A.; Echeverri-Peña, O.Y.; Albarracín, S.L.; Alméciga-Díaz, C.J.; Espejo-Mojica, Á.J. Sphingolipids and Their Role in Health and Disease in the Central Nervous System. *Adv Biol Regul* 2022, *85*.

42.   Czubowicz, K.; Jęśko, H.; Wencel, P.; Lukiw, W.J.; Strosznajder, R.P. The Role of Ceramide and Sphingosine-1-Phosphate in Alzheimer's Disease and Other Neurodegenerative Disorders. *Mol Neurobiol* **2019**, *56*, 5436–5455, doi:10.1007/S12035-018-1448-3.

43.   Tabatabaie, L.; Klomp, L.W.; Berger, R.; de Koning, T.J. L-Serine Synthesis in the Central Nervous System: A Review on Serine Deficiency Disorders. *Mol Genet Metab* 2010, *99*, 256–262.

44.   Holeček, M. Serine Metabolism in Health and Disease and as a Conditionally Essential Amino Acid. *Nutrients* **2022**, *14*, doi:10.3390/NU14091987.
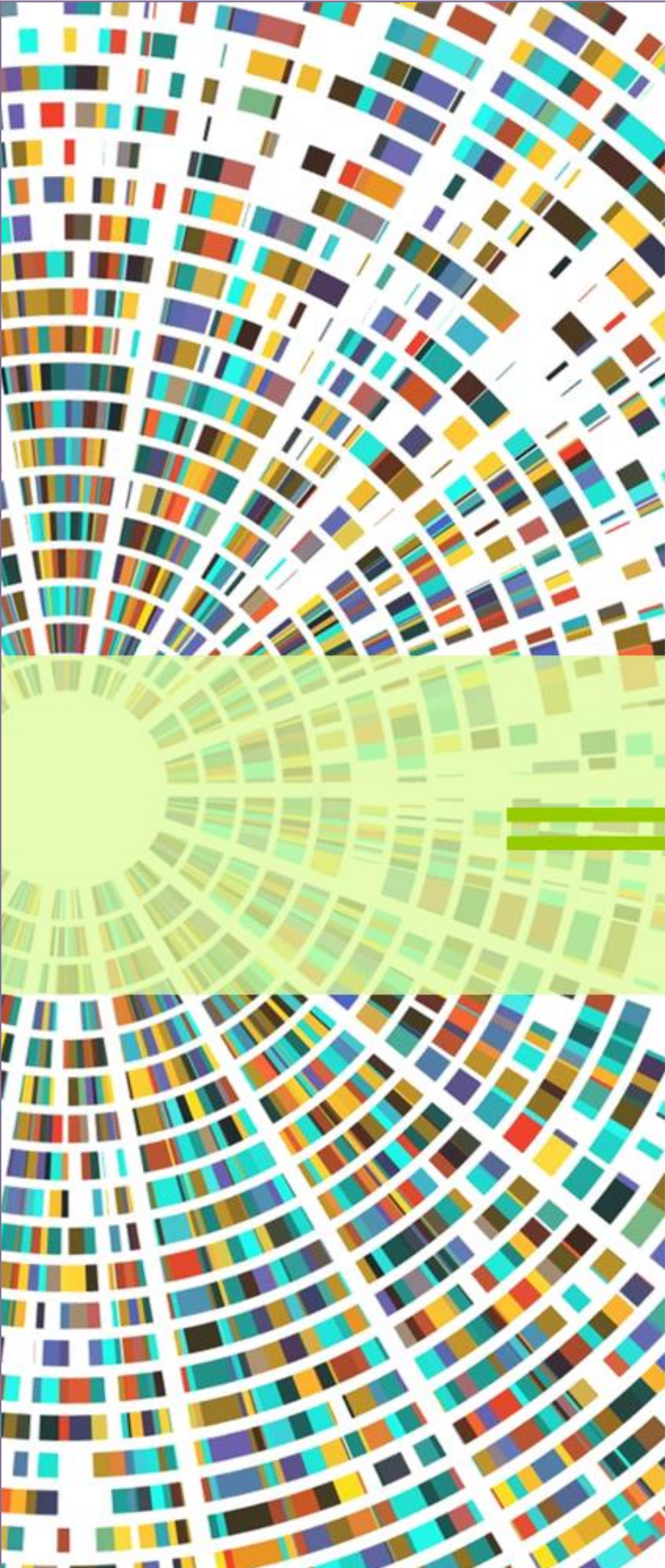
Chapter 7

Conclusions

7   CHAPTER 7: CONCLUSIONS

This final chapter gathers the general and specific conclusions of this thesis. At the same time, each previous chapters collect more detailed conclusions of each study.

Considering the need for standardised clinical diagnostic tools and protocols for the prevention and diagnosis of many diseases, analysing global metabolic profiles instead of disparate clinical measurements could be essential to shed light on the disarrangements of various diseases. Therefore, the diseases such as MetS, PD, AD and ALS were greatly approached from a holistic functional perspective performing analysis of metabolic profiles that could reflect the global clinical patient status.

Non-targeted metabolomic studies based on FTIR spectroscopy aimed to extract the metabolic signatures instead of individual biomarkers, permitting the classification of patients according to their molecular patterns. Thus, developing a chemometric strategy capable of extrapolating the most significant infrared signatures was crucial in this doctoral thesis. Each spectrum is unique for every patient, reflecting clinical/pathological conditions. Summarising:

- The LDA classifications and SIMCA-developed models demonstrated that the spectral variables obtained by FTIR could provide the same discriminative results as measured clinical parameters for MetS discrimination.

- A three-step classification approach revealed spectroscopic biomarker signatures that define patient subgroups for the clinical diagnosis and classification of PD at different stages of the disease. PD at the initial stage (PDI) was significantly differentiated and not confused with developed PD-related dementia (PDD), providing optimal classification results in both sub-classification problems. In addition, this approach successfully stratified patients with different PD stage progression profiles and those with different dementia-type profiles. All the speculations about the involvement of selected bands in the pathogenesis of PD are immensely reasonable, and their role is perfectly justifiable for patient stratification.

239

- Different combinations of patient groups were studied to confirm the discrimination ability of the infrared technique in ALS classification, yielding very high precision and specificity towards groups. All the healthy patients were significantly separated, confirming the power of the built model in preventing the misdiagnosis of non-diseased patients as ones having neuromuscular disorders. In addition, no ALS were confused with other neuro-disorder cases leading highest specificity values for both pathologies.

- The potential of ATR-FTIR spectroscopy was investigated as a rapid tool for discriminating ALS patients from controls and ALS disease progression. This preliminary study's analysis based on LDA classification and SIMCA class modelling provided excellent group separation. Therefore, both ALS at onset and after six months of follow-up and ALS at the disease onset from controls were discriminated with almost 100% accuracy by LDA and SIMCA, highlighting the great potential of the instrument to be implemented for fast screening purposes.

It is well-known that lipids exert various functions in the central nervous system, including roles in cell structure, synaptic transmission, and multiple metabolic processes. Thus, herein, a non-targeted lipidomic approach using UPLC-MS/MS was performed to unravel alterations in one or more lipid specimens and metabolic pathways where they are possibly involved.

- Statistically significant lipid compounds belonging to glycerophospholipids and were identified in UHPLC-MS untargeted approach, which could be a prominent biomarker to the differentiation between AD or PD and healthy controls.

- UPLC-MS metabolomic approach found evidence of previously established lipid compounds and identified novel emerging metabolites for ALS discrimination., Our results revealed discriminant traits of ALS prognosis and ALS progression suggesting the contribution of alterations in UFAs, PUFAs, FAs and glycerolipid metabolism to the pathogenesis of ALS and other neuro disorders.

The principal limitation of these studies resides in the relatively reduced sample size at our disposal for some patient. Nevertheless, it should be highlighted that it is extremely

challenging to collect samples for some diseases such as Amyotrophic lateral sclerosis. As note, we conducted a longitudinal study on ALS cohort, and compared group of patients at different stage of disease progression. In our knowledge, very limited studies have performed longitudinal studies on ALS. Therefore, the adopted analytical strategy could be converted in universal diagnostic methodology, by increasing the patient dataset. In addition, the routinely included confounding factors, such as dietary habits, medications, comorbidities and race, would suggest possible solutions in future studies and strengthen the validity of the discussed classification strategies.

Thus, cost-effectiveness and relative ease of access and IR devices' portability take vibrational spectroscopy one step towards clinical implementation, which could be ideal for point-of-care testing, primary health care, or wherever required.

Exhaustive metabolomic fingerprinting research should not be limited to a unique analytical platform but rather should test and combine multiple analytical strategies in order to exploit their respective strengths and overcome their weaknesses. In this doctoral thesis, FTIR spectroscopy and UPLC-MS successfully complemented each other, providing excellent patient discrimination based on spectral and on specific blood biomarkers.